**Determining Protein Interaction Specificity of Native and Designed bZIP Family Transcription Factors**

by

Aaron W. Reinke

B.S. Biochemistry and Molecular Biology
University of California, Davis, 2005

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2012



Signature of Author:_____
Department of Biology
February 6, 2012

Certified by: _____
Amy Keating
Associate Professor of Biology
Thesis Supervisor

Accepted by:_____
Robert T. Sauer
Salvador E. Luria Professor of Biology
Co-Chair, Biology Graduate Committee

**Determining Protein Interaction Specificity of Native and Designed bZIP Family Transcription Factors**

by

Aaron W. Reinke

Submitted to the Department of Biology
on February 6, 2012 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biology at the Massachusetts Institute of Technology

## ABSTRACT

Protein-protein interactions are important for almost all cellular functions. Knowing which proteins interact with one another is important for understanding protein function as well as for being able to disrupt their interactions. The basic leucine-zipper transcription factors (bZIPs) are a class of eukaryotic transcription factors that form either homodimers or heterodimers that bind to DNA in a site-specific manner. bZIPs are similar in sequence and structure, yet bZIP protein-protein interactions are specific, and this specificity is important for determining which DNA sites are bound. bZIP proteins have a simple structure that makes them experimentally tractable and well suited for developing models of interaction specificity. While current models perform well at being able to distinguish interactions from non-interactions, they are not fully accurate or able to predict interaction affinity.

Our current understanding of protein interaction specificity is limited by the small number of large, high-quality interaction data sets that can be analyzed. For my thesis work I took a biophysical approach to experimentally measure the interactions of many native and designed bZIP and bZIP-like proteins in a high-throughput manner. The first method I used involved protein arrays containing small spots of bZIP-derived peptides immobilized on glass slides, which were probed with fluorescently labeled candidate protein partners. To improve upon this technique, I developed a solution-based FRET assay. In this experiment, two different dye-labeled versions of each protein are purified and mixed together at multiple concentrations to generate binding curves that quantify the affinity of each pair-wise interaction.

Using the array assay, I identified novel interactions between human proteins and virally encoded bZIPs, characterized peptides designed to bind specifically to native bZIPs, and measured the interactions of a large set of synthetic bZIP-like coiled coils. Using the solution-based FRET assay, I quantified the bZIP interaction networks of five metazoan species and observed conservation as well as rewiring of interactions throughout evolution. Together, these studies have identified new interactions, created peptide reagents, identified sequence determinants of interaction specificity, and generated large amounts of interaction data that will help in the further understanding of bZIP protein interaction specificity.

Thesis Supervisor: Amy Keating
Title: Associate Professor of Biology

**ACKNOWLEDGEMENTS**

I would like to thank the following people that helped make this work possible:

My advisor, Amy Keating, for giving me the freedom to be able to go in the directions I found most interesting and providing advice, guidance, and support along the way. She has also been instrumental in helping me improve my ability to both perform and communicate science.

My thesis committee members, Rick Young and Dennis Kim, for providing advice and challenging me to think how my work fits into a larger picture. Marian Walhout for coming to the defense.

Members of the Keating lab, past and present, for always being helpful, providing advice, and creating a fun environment in which to do experiments. Gevorg Grigoryan, Scott Chen, Judy Baek, and Orr Ashenberg who were a pleasure to collaborate with. Jen Kaplan for reading my thesis.

Bob Grant for teaching me what I know about X-ray crystallography.

Members of the Baker, Kim, Laub, Sauer, and Schwartz labs for being generous with both equipment and advice.

Ted Powers for having me in his lab as an undergraduate and teaching me how to be a scientist. Karen Wedaman for showing me there is more fun to be had in lab than just washing dishes.

Friends and classmates for providing ample reasons to take a break from lab.

My family for their support and encouragement.

Steph, for being my cohort.

## TABLE OF CONTENTS

## LIST OF FIGURES AND TABLES

**CHAPTER 1:**
**An introduction to the study of protein-protein interactions**



**CHAPTER 2:**
**Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays**



**CHAPTER 3:**
**Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays**



**CHAPTER 4:**
**A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering**

**CHAPTER 5:**
**Conservation and rewiring of bZIP protein-protein interaction networks**



**APPENDIX A:**
**Supplementary Information for "Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays"**

**APPENDIX B:**

**Supplementary Information for "Design of protein-interaction specificity affords selective bZIP-binding peptides"**

**APPENDIX C:**

**Supplementary Information for "A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering"**



**APPENDIX D:**

**Design of peptide inhibitors that bind the bZIP domain of Epstein-Barr virus protein BZLF1**

# Chapter 1

# An introduction to the study of protein-protein interactions

Protein-protein interactions are essential for most cellular functions. Thus understanding which proteins interact with each other is necessary for understanding how cells work. The problem of how each protein is able to interact with a specific set of partners is complex. It is estimated that 74,000–200,000 interactions occur among the ~25,000 proteins encoded by the human genome (Venkatesan, et al. 2009). This huge amount of interactions is further complicated by the fact that protein-protein interactions have a diverse set of properties. Interaction interfaces are structurally varied in nature and can either be mediated through domain-domain interactions or by domains binding to short peptide regions. While some interactions are stable, many interactions are dynamic and of lower affinity. Some proteins interact with few partners, but some interact with many (Han, et al. 2004). All of these factors combine to make it difficult to know which proteins interact with each other.

There are many goals in studying protein-protein interactions. The first is to identify which interactions occur. This is often a first step in understanding the function of a protein, because knowing which proteins it interacts with gives insight into a protein's functional role. Large data sets of interactions can also be used to determine interaction network structure (Han, et al. 2004). As this is a critical goal, a number of techniques have been developed for measuring interactions on a large scale. A second goal in studying protein-protein interactions is to identify the functional significance of the interactions. This is often attempted by knocking out or knocking down a gene of interest for one or both partners and assaying the phenotypic effect. Unfortunately this removes all interactions of the knocked out gene. A more focused approach is

to generate mutants that specifically disrupt an interaction without compromising the entire function of the protein (Dreze, et al. 2009).

In addition to identifying interactions and determining their functions, there is a need to understand biophysically how proteins interact. This understanding is important for being able to generate models that describe the relationship between sequence and interaction properties. There are several practical uses of such models. Models can be used to predict interactions from protein sequence alone (Chen, et al. 2008). This can be useful for identifying unknown interactions important for human biology, and also for predicting interactions from the increasingly large number of genomes being sequenced. Models that could predict what effect mutations have on binding affinity and specificity would be useful, especially for understanding the basis of disease. An ability to accurately model interactions could also support the design of proteins with specific interaction properties, such as peptides designed to specifically disrupt interactions (Grigoryan, et al. 2009).

Two general approaches exist for measuring protein-protein interactions on a large scale. One involves measurements that are done using full-length proteins, either *in vivo* in the organism of interest or in yeast. These approaches have the advantage of being able to be applied on a proteome-wide scale. A complementary set of approaches are those that rely on domain-based *in vitro* measurement techniques. In these approaches, large domain families are selected and representative domains are cloned. These domains are then expressed, purified, and tested against a number of potential interaction partners using a variety of different experimental techniques. These methods can quantify large numbers of similar interactions, generating the

type of data that is the most useful for modeling interactions. The most widely used techniques and the advantages and disadvantages of each approach are discussed below.

*Proteome-wide methods for the study of protein interactions*

Three main experimental techniques have been shown to be useful on a proteome-wide scale for measuring protein-protein interactions (Figure 1.1). 1) In the yeast two-hybrid (Y2H) assay, one protein is fused to an activator domain and the other to a DNA-binding domain. Yeast expressing both constructs display transcriptional reporter activity if the two proteins interact. Several versions of the assay exist, but the most common relies on the GAL4 transcription factor driving a variety of selectable reporter genes (Rajagopala and Uetz. 2011). 2) Protein fragment complementation assays (PCA) involve a reporter protein that is split into two fragments, with the N-terminal fragment fused to one of the proteins being tested and the C-terminal fragment fused to the other. When a pair of proteins interacts, the protein activity of the split reporter is reconstituted. The most commonly used split protein in yeast is a mutant version of dihydrofolate reductase, which allows for selection using the drug methotrexate (Michnick, et al. 2011). 3) Affinity purifications followed by mass spectrometry (AP/MS) involves fusing each protein to an affinity tag that is then used to purify the protein along with any other proteins that are associated with it. Isolated protein complexes are then digested into peptides using proteases such as trypsin, and the identity of the peptides is determined using MS/MS. Many different tags exist for doing purification, with the most common being tandem affinity purification tags that allow for two rounds of purification to eliminate background binding (Gavin, et al. 2011).

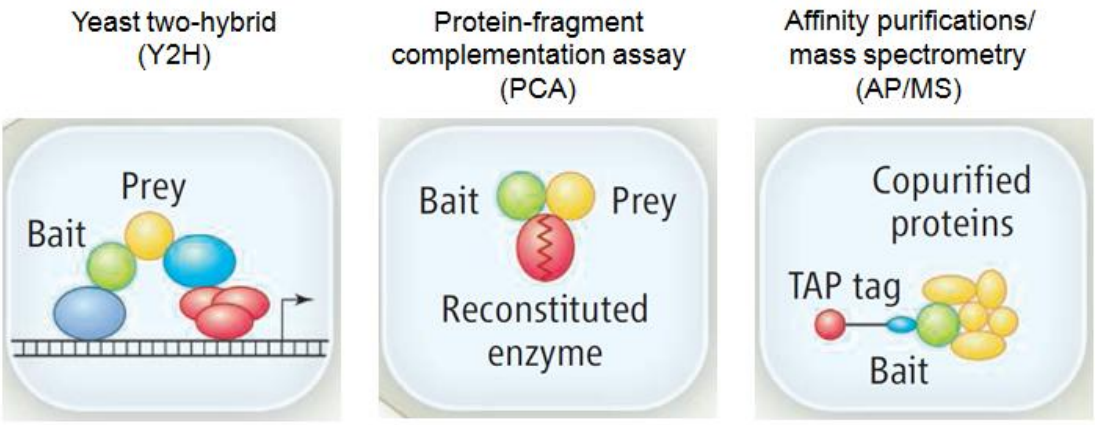



Figure 1.1. Proteome-wide methods for measuring protein-protein interactions. Modified from (Jensen and Bork. 2008).

The first attempts to map interactions on a proteome-wide scale were done using Y2H applied first to T7 bacteriophage, followed by other viruses as well as partial attempts in *H. pylori*, *S. cerevisiae*, *C. elegans*, and *D.melanogaster* (McCraith, et al. 2000, Uetz, et al. 2000, Rain, et al. 2001, Flajolet, et al. 2000, Ito, et al. 2001, Ito, et al. 2000, Giot, et al. 2003, Li, et al. 2004, Walhout, et al. 2000). These initial studies were followed by an improvement in the methodology and throughput of the assay, which was subsequently applied to several bacteria, more complex organisms such as human and Arabidopsis, and higher-coverage versions of the *C. elegans* and yeast interaction maps (Stelzl, et al. 2005, Titz, et al. 2008, Rual, et al. 2005, Parrish, et al. 2007, Simonis, et al. 2009, Yu, et al. 2008). Y2H was the first technology that allowed interactions to be measured on a large scale, and this approach revealed the size and connectedness of interaction networks. Y2H suffers from a high false negative rate, however, with as few as 10% of true interactions being detected; this resulted in little overlap of interactions in initial studies (Yu, et al. 2008). Low assay sensitivity in Y2H has been addressed both by measuring every potential interaction in an array format, using all possible combinations of N-terminal and C-terminal fusion constructs, and by measuring protein fragments in addition

to full-length proteins (Xin, et al. 2009, Boxem, et al. 2008, Chen, et al. 2010). Even when using multiple Y2H versions in an array format, 20% of a gold set of interactions still could not be detected, likely because of the requirement for proteins to be expressed and localized and to interact as fusion proteins in the yeast nucleus (Chen, et al. 2010). While much effort has been made to prevent assay false positives, interactions can nevertheless be detected between proteins that may never interact physiologically, due to never being co-expressed or co-localized.

PCA was first used on a proteome-wide scale to map interactions in *S. cerevisiae* (Tarassov, et al. 2008). While so far less used than Y2H, PCA has several advantages. Interactions can be measured under the endogenous promoter with native localization in living cells. The data generated also provide some topological information, as the maximum distance the two fused halves can be from one another is 80 Å. A drawback is that only the interactions that occur under the cellular conditions measured can be observed. In the study by Tarassov et al., measurements were done under only one condition and thus likely missed interactions from proteins that were not expressed or differentially localized. False positives can arise in PCA due to the split fragments bringing proteins together that otherwise wouldn't interact. Additional versions of PCA based on fluorescence or luminescence have the potential to detect interactions *in vivo* as well as to provide cellular and subcellular localization information (Michnick, et al. 2011).

AP/MS was first applied on a proteome-wide scale to map interactions in yeast. In two pilot studies and then in two subsequent studies, the vast majority of the ~6,000 yeast proteins were tagged and over 1/3 of purifications were successful (Ho, et al. 2002, Krogan, et al. 2006, Gavin, et al. 2002, Gavin, et al. 2006). This technique has also been applied to *E. coli*, *M. pneumonia*, *D.melanogaster,* and human interactions (Malovannaya, et al. 2011, Guruharsha, et

al. 2011, Kuhner, et al. 2009, Hu, et al. 2009, Arifuzzaman, et al. 2006, Butland, et al. 2005). AP/MS, like PCA, has the advantage of being able to detect interactions *in vivo*, but suffers from only detecting interactions under the conditions they are assayed under. Quantitative approaches hold promise for comparing between different conditions and cell states (Bantscheff, et al. 2007). The AP/MS approach suffers from potential false negatives, including interactions that are transient, have fast off rates, or are lost during the isolation and washing procedure. False positives are also a problem, and these can arise both from highly expressed non-specifically binding proteins, as well from disruption of cellular substructure that can allow differentially sublocalized proteins to interact.

A main difficulty in this approach is engineering organisms to express the tagged proteins of interest. Proteins fused to an affinity tag under an endogenous promoter are preferred because overexpression of a protein can lead to false positive interactions (Ho, et al. 2002). Only in yeast and recently in *E. coli* has endogenous tagging been possible. Recent methods for cloning large amounts of DNA including regulatory regions will allow for greater coverage in systems such as human cell lines (Poser, et al. 2008, Hutchins, et al. 2010). Antibodies provide a potential way to circumvent using engineered strains. A recent study using a large number of antibodies in human cells identified specific interactions by constraining interactions to be present in reciprocal isolations (Malovannaya, et al. 2011). Making the large numbers of antibodies required to bind to every protein is difficult, though affinity reagents based on other scaffolds hold promise (Boersma and Pluckthun. 2011).

All of these proteome-wide methods are not yet comprehensive. Even in yeast, where all three approaches have been used, there is not yet complete coverage. Y2H applied to yeast has

only mapped ~20% of the estimated total interactions (Yu, et al. 2008). PCA was able to test 93% of genes, but the sensitivity of the assay is not known (Tarassov, et al. 2008). In the two large yeast AP/MS studies, 60% of the proteome was detected, but only 18% of the interactions observed are shared between the two studies (Goll and Uetz. 2006). This lack of complete coverage is due both to the number of proteins that were assayed as well as the sensitivity of the assays. There is also little overlap in the interactions detected by these three methods because each method has biases towards different classes of proteins (Jensen and Bork. 2008). Further improvement to these assays, combined with other potential high-throughput approaches, should allow for even more complete maps of interactions to emerge (Snider, et al. 2010, Kung and Snyder. 2006, Lievens, et al. 2009, Miller, et al. 2009, Petschnigg, et al. 2011).

A major drawback of these approaches is they typically give little structural information on how the interactions occur. In the case of Y2H and PCA, it is likely that the pair of fused proteins is directly mediating the interaction. In the case of AP/MS, complexes of interacting proteins are isolated, and it is typically not known what the direct physical interactions that occur are. Additionally, these methods don't provide information on the regions of proteins mediating the interactions. This type of information could be gained by using Y2H with protein fragments to map minimal interacting domains, or by using AP/MS with crosslinkers of defined length to provide spatial constraints to the regions of proteins that interact (Boxem, et al. 2008, Stengel, et al. 2011).

*Domain-based approaches for studying protein interaction specificity*

As an alternative to mapping interactions of full-length proteins on a proteome-wide scale, much effort has been made to measure the interactions of individual domain families.

Proteins are composed of many different domains, of which at least 70 are known to mediate protein-protein interactions (Letunic, et al. 2012, Pawson and Nash. 2003). Domains can interact with other structured domains or with short peptide regions, and these interactions can be influenced by post-translational modifications such as phosphorylation (Pawson and Nash. 2003). There are several advantages of focusing on domains. Domains alone have been shown to be sufficient to bind to partners independent of the rest of the protein. In fact, proteins often have regulatory regions that can inhibit interactions in the context of the full-length protein. Domains often behave better *in vitro* than full-length proteins. Finally, focusing on domains reduces the complexity of determining where the partner binds.

A collection of different techniques has been shown to be useful for measuring the specificity of protein domains *in vitro*. Several of the most widely used methods are described below. Selection-based techniques such as phage display, yeast display, and ribosome display all work by expressing a protein or peptide that is linked to its genetically encoded message. A large number of different library members, $10^7$ to $10^{14}$, can be expressed at a time, and interactions can be identified by pulling down with the domain of interest or through cell sorting. The selected sequences can then be determined by sequencing the DNA of the binding population. A large advantage of this approach is that only one partner needs to be purified and a very large number of potential binders can be assayed at a time. The drawback of this approach is that it typically only identifies high-affinity binders, missing weak interactions and non-interactions that could be important for understanding binding specificity and function (Shao, et al. 2011, Liu, et al. 2010). Also, libraries are often biased as to which sequences are expressed.

Protein arrays involve printing proteins onto a solid surface. Arrays can be prepared in a 96-well format, where each well contains an identical subarray containing several hundred proteins. The arrays can then be probed with a fluorescently-labeled partner, allowing for many interactions to be measured in parallel. If done at multiple concentrations, quantitative binding affinities can be determined (Jones, et al. 2006). Arrays can also be prepared by synthesizing peptides on cellulose membranes, known as SPOT arrays (Briant, et al. 2009). Both protein and peptide arrays have the advantage that binders from a range of different affinities as well as non-binders can be measured at the same time. Disadvantages include potential artifacts resulting from measuring interactions on a surface, as well as the technical nature of preparing protein arrays.

Solution measurements of protein interactions can be done in high-throughput in 384-well plates using either fluorescence polarization or FRET (Stiffler, et al. 2006). This approach has the advantage of being able to quantify interactions without the issue of potential surface artifacts. The main drawback to this type of approach is that these experiments are often time consuming and costly, which limits the number of potential interactions that can be assayed. High-throughput data processing and curve-fitting is also challenging. Solution methods, protein arrays, and display methods are complementary to one another, and often multiple techniques are used on a domain family to gain a deeper understanding of the determinants of binding specificity, as discussed below.

The binding specificity of several domain families has been investigated in detail. Three of the largest domain families are the PDZ, SH2, and SH3 domains, which have all been studied extensively using high-throughput approaches (Figure 1.2). These families contain many

members, and the individual domains are small in size and experimentally tractable. These domains also all bind short peptides, which can be expressed as random libraries, synthesized on surfaces, or fluorescently labeled. Work on these domains has demonstrated that peptide-binding domains can display a high degree of specificity. This has also to led to the idea that although interactions *in vivo* can be influenced by many cellular effects, such as expression and localization, binding specificity can also be hardwired in protein sequence (Liu, et al. 2010, Stiffler, et al. 2007, Tonikian, et al. 2009).

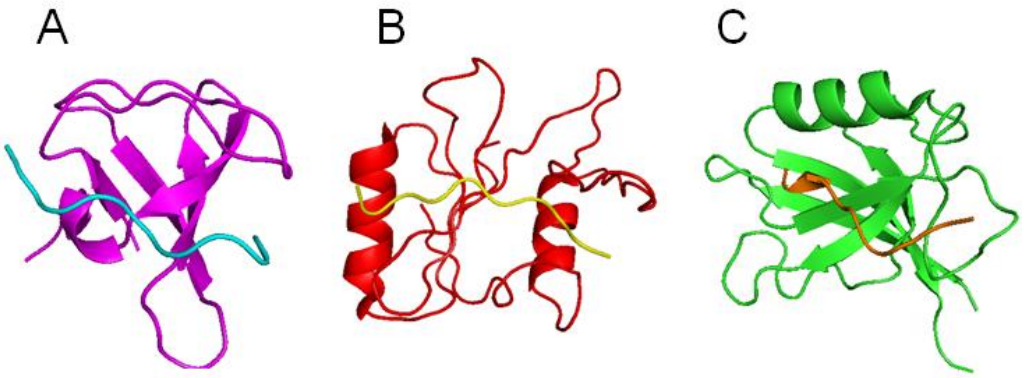



Figure 1.2. Structures of peptide-binding domains in complex with peptides. A) SH3 domain (PDB: 1ABO). B) SH2 domain (PDB: 1D4W). C) PDZ domain (PDB: 1MFG). Figures generated using PyMOL (DeLano Scientific, Palo Alto, CA).

SH3 domains are involved in signaling by binding mainly to multi-proline-containing peptides. The domains consist of ~80 amino-acid residues, and there are 400 SH3 domains in humans and 27 in yeast (Castagnoli, et al. 2004). They were originally divided into two classes, binding either the consensus motif +XXPXXP or PXXPX+ (where X is any residue and + is either arginine or lysine). Cesareni and coworkers expanded on previous studies by measuring the interaction specificity of 25 yeast SH3 domains using phage display, peptide arrays, and Y2H (Tonikian, et al. 2009, Landgraf, et al. 2004, Tong, et al. 2002). These three experimental data

sets were combined into a single model that showed better prediction than any single technique. This demonstrated the usefulness of applying different measurement technologies to the same problem. These experiments also revealed that although the majority of domains did fall into the two specificity classes, within these classes there are many distinct specificities. Further, positions outside of the core binding motif were shown to be important for binding.

SH2 domains are composed of ~100 amino-acid residues and bind to phosphotyrosine-containing peptides. There are 120 SH2 domains in humans, and they are involved in signaling downstream from protein-tyrosine kinases (Liu, et al. 2006). As it is difficult to express phosphorylated peptides, most work on SH2 binding specificity has been performed using protein and peptide arrays. MacBeath and coworkers measured the binding of about 90 SH2 domains against 61 phosphtyrosine peptides {{71 Jones,R.B. 2006}}. The authors printed domains on the surface of glass slides and generated binding curves using fluorescently-labeled peptides. This was the first large scale quantitative affinity study of any binding domain and showed that proteins arrays could be used not just for detecting interactions but for quantifying the strength of the interactions. In another study the specificity of 76 SH2 domains was determined using a version of SPOT arrays where each position was fixed to one amino acid at a time while all other positions except the phosphotyrosine were randomized. These experiments suggested that there were only a limited number of specificity-determining residues on the peptides that were recognized by each domain (Huang, et al. 2008). In an alternative approach, 50 SH2 domains were measured against 192 phosphotyrosine peptides derived from native proteins using SPOT arrays. This revealed that SH2 domains displayed specificity with respect to these peptides and were more specific than previously anticipated. This suggested that

permissive residues alone are not enough to determine binding specificities, and non-permissive residues can be important (Liu, et al. 2010).

PDZ domains are composed of ~80 amino-acid residues and typically bind to short, C-terminal peptides. They are present in all domains of life (~250 domains in human) and are involved in many different cellular signaling processes (Tonikian, et al. 2008). Many different high-throughput experimental approaches have been used to measure their interaction specificity, including protein arrays, SPOT arrays, phage display, Y2H, and fluorescence polarization (Stiffler, et al. 2007, Tonikian, et al. 2008, Wiedemann, et al. 2004, Lenfant, et al. 2010). Two groups have recently measured a large number of interactions using different approaches. MacBeath and coworkers measured the interactions of 85 murine PDZ domains with over 200 peptides. They used a two-stage strategy that involved identifying positive and negative interactions on arrays presenting PDZ domains, and then quantifying the affinity for the positives using fluorescence polarization (Stiffler, et al. 2006, Stiffler, et al. 2007). Sidhu and coworkers profiled binding specificity using phage display with a peptide library that had at least 7 positions randomized. They measured the binding specificity of 82 native PDZ domains from human and *C. elegans*, 83 synthetic domains, and 91 single point mutants (Tonikian, et al. 2008, Ernst, et al. 2009, Ernst, et al. 2010). While initial studies suggested that PDZ domains could be grouped into three different classes of broad specificity, these newer and much more expansive studies have shown PDZ domains to be much more selective and have identified at least 23 distinct specificity clusters. While they do display specificity, each PDZ domain is predicted to interact with ~250 proteins on average (Stiffler, et al. 2007). PDZ domains are also known to interact with internal peptides, as well as to form dimers with other PDZ domains using a distinct interface (Im, et al. 2003). Recently, 157 domains were measured against each other using

protein arrays, and 30% of domains were shown to interact with each other (Chang, et al. 2011). Interpretation of these interactions is difficult, as it is unclear which interface of the PDZ domain is used in mediating the interactions.

The data for PDZ domain binding have been a rich source for development of models to predict binding specificity. Computational modeling was used to predict the binding specificity of 17 PDZ domains analyzed by phage display. On average, half of the positions bound by each domain were predicted well (Smith and Kortemme. 2010). Two groups also developed models of PDZ domain binding using the MacBeath data set of quantitative interactions and non-interactions. Chen et al. trained a novel model on the data and were able to predict new interactions with ~50% accuracy (Chen, et al. 2008). A different machine learning approach on the same data set was able to predict the affinity of a set of single point mutants with a correlation of 0.92 (Shao, et al. 2011). These results indicate clear progress, but while there is now an enormous amount of data, the problem of predicting interactions with high accuracy based on sequence and structure is far from solved.

In summary, domain-based *in vitro* assays provide a reductionist approach that allows for the decoupling of cellular influences, such as expression and localization, and focusing on measuring all interactions that can physically occur. Systematic data sets of both interactions and non-binders can be generated that are useful for developing models of binding specificity. Binding models are useful for predicting interactions in each domain family, as well as for uncovering general principles that govern protein-binding specificity. The domain-based approach is complementary to the proteome-wide approach. Having a deep understanding of the binding specificity of a large number of domains would allow mapping of domain interactions to

the larger proteome-wide datasets. Domain interactions can also be inferred from proteome-wide data sets, which could potentially identify domain interactions that can be interrogated *in vitro* (Deng, et al. 2002).

*bZIPs as a model class of protein-protein interaction*

The basic leucine-zipper transcription factors (bZIPs) are an evolutionally conserved family of eukaryotic transcription factors that are ideal for studying protein-protein interaction specificity. bZIPs bind to DNA site specifically via a basic region. Immediately C-terminal to the DNA-binding residues is a coiled coil that mediates the formation of either homodimers or heterodimers (Figure 1.3A). The bZIP proteins are involved in many different cellular processes and can act as both activators and repressors of transcription (Hirai S, Bourachot B,Yaniv M. 1990, Lai and Ting. 1999). The protein partnering specificity is important, as it can dictate which DNA sites are bound (Hai and Curran. 1991). There are several features that make bZIPs an ideal domain to study protein-protein interaction specificity. They have a simple interaction interface of two alpha helices forming a parallel dimeric coiled coil. Further simplifying the interaction is the repeating-heptad structure, where each position in the heptad can be designated with a letter **abcdefg**. The bZIP coiled coils are thought to interact exclusively with one another, which limits the number of potential interactions to be tested. There are a number of bZIPs in both human and other species, which provides a large collection of sequences for which to map the specificity (Amoutzias, et al. 2007). The coiled coils in bZIPs are typically ~35-50 amino acids long, making them very experimentally tractable. bZIPs are also a model system for understanding coiled-coil interaction specificity more broadly, which is important because coiled coils are predicated to occur in ~10% of proteins in eukaryotic genomes (Liu and Rost. 2001). What is

known about how bZIPs interact, and what the specificity determining features are, is the result of the work of many laboratories and is summarized below.

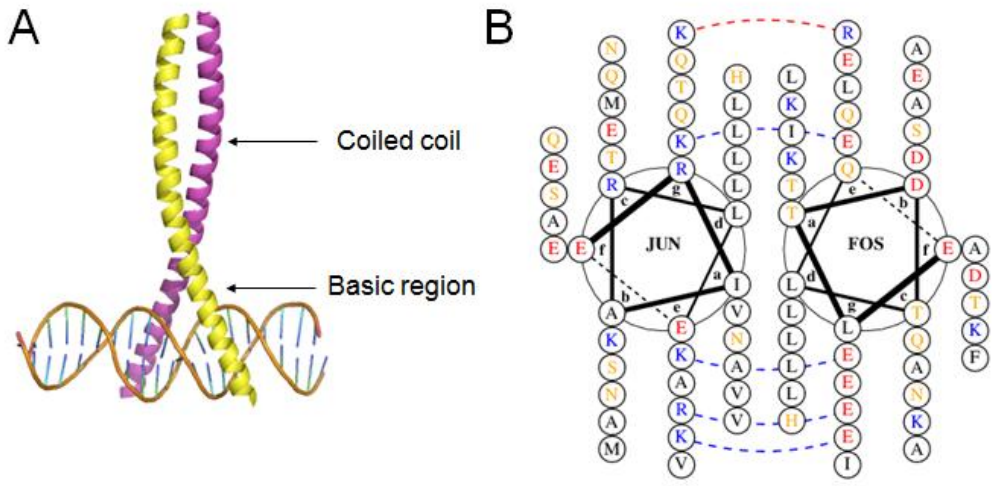



Figure 1.3. Structure of a bZIP coiled coil. A). Crystal structure of a quaternary complex of JUN and FOS bound to double-stranded DNA (PDB: 1FOS). B). Helical wheel diagram of JUN and FOS. Hydrophobic residues, black. Polar residues, yellow. Positively charged residues, blue. Negatively charged residues, red. Attractive **g-e'** electrostatics, blue-dashed lines. Repulsive **g-e'** electrostatics, red-dashed lines. Crystal structure figure created using PyMOL (DeLano Scientific, Palo Alto, CA). Helical wheel diagram generated using DrawCoil 1.0. http://www.gevorggrigoryan.com/drawcoil/)

*Identification and initial characterization of bZIPs*

The founding members of the bZIP family were first discovered and characterized by converging work on oncogenic viruses, yeast transcriptional regulation, and viral enhancer binding proteins. FOS and JUN were both identified first in oncogenic retroviruses and then cloned from human cells (Curran, et al. 1982, van Straaten, et al. 1983, Maki, et al. 1987). GCN4

was identified in yeast as being a positive regulator of amino-acid biosynthesis (Hinnebusch and Fink. 1983). CEBPA was identified from rat livers as a protein that bound to viral enhancers (Landschulz, et al. 1989). Molecular work on these four bZIPs led to a detailed, mechanistic understanding of how bZIPs function. The functional region of GCN4 responsible for DNA binding was narrowed to a 60 amino-acid region (Hope and Struhl. 1986). GCN4 was then shown to bind to palindromic DNA sites as a dimer and form stable complexes even without DNA present (Hope and Struhl. 1987). FOS and JUN were shown to form heterodimers, and it was demonstrated that this association depends on the leucine-zipper domain (Sassone-Corsi, et al. 1988, Turner and Tjian. 1989, Gentz, et al. 1989).

McKnight and coworkers first observed that these four proteins shared a common structural feature that was termed a "leucine zipper," and suggested that these leucine zippers associated as dimers in an antiparallel fashion (Landschulz, et al. 1988). Shortly thereafter disulfide exchange experiments on GCN4 showed that the association was that of a parallel dimer, and the interaction was suggestive of a coiled coil (O'Shea, et al. 1989). Using CEBPA, it was shown that mutations to the leucine zipper prevented both dimerization and DNA binding whereas mutations in the basic region disrupted only DNA binding (Landschulz, et al. 1989). Several groups also made chimeras between different leucine zippers and basic regions. These chimera experiments demonstrated that the leucine zipper was responsible for dimerization, the basic region bound to DNA, and these functions were separable (Agre, et al. 1989, Sellers and Struhl. 1989, Kouzarides and Ziff. 1989). Going even further, two groups showed that that the native leucine zipper could be replaced with either an idealized coiled coil, or a disulfide bond, demonstrating that a dimerized basic region is sufficient for binding to DNA (Talanian, et al. 1990, O'Neil, et al. 1990). Structural models were developed that consisted of bZIPs forming

parallel dimers via the coiled-coil domain, with the basic regions forming a continuous helix that interacted with DNA (O'Neil, et al. 1990, Vinson, et al. 1989). Crystal structures of a homodimer of GCN4 and a heterodimer of JUN and FOS, both bound to DNA, provided experimental evidence in excellent agreement with these models (Ellenberger, et al. 1992, Glover and Harrison. 1995).

*Specificity determinants of bZIP protein-protein interactions*

Two major findings from these studies were that the leucine zipper controlled dimerization specificity and that only certain homodimers and heterodimers could interact (Sellers and Struhl. 1989, Kouzarides and Ziff. 1989). Understanding this specificity became a major research focus. O'Shea and Kim made chimeras of the **bcf** positions (the outside of the helix) and the **adeg** positions (the inside of the helix) between GCN4, FOS and JUN. This experiment showed that specificity was largely influenced by the **adeg** positions. They further showed that just the **eg** positions were sufficient to explain the specificity between these bZIPs, and placing the 8 residues in these positions from FOS and from JUN into GCN4 was sufficient for the specific formation of heterodimers (Figure 1.3B) (O'Shea, et al. 1992). To test the principals of **g-e'** electrostatics, two peptides were designed, one that had glutamates at all **eg** positions and another that had all lysines at these positions. These peptides, termed peptide "Velcro," were show to form very weak homodimers, but when mixed together formed strong heterodimers. (O'Shea, et al. 1993). Using these same principals Vinson and coworkers predicted native bZIPs that would and would not form heterodimers and validated these predications experimentally (Vinson, et al. 1993).

It was later shown that asparagines at **a** positions could also impart specificity, in that they could pair with asparagines at an **a** position on the opposite helix, but not with hydrophobic amino acids such as isoleucine, valine, or leucine (Zeng, et al. 1997, Acharya, et al. 2006, Acharya, et al. 2002). The **a** position has also been observed to be involved in imparting structural specificity, as mutating an asparagine at an **a** position to a hydrophobic amino acid can lead to the formation of oligomers and/or loss of orientation specificity (Harbury, et al. 1993, Lumb and Kim. 1995). Leucine, which is the most common amino acid at **d** positions in native bZIPs, was shown to be the most stabilizing homotypic interaction at the **d** position (Moitra, et al. 1997). Coupling energies of electrostatics of **g-e'** interactions were measured using double mutant alanine thermodynamic cycle analysis (Krylov, et al. 1994). Coupling energies of all pairwise interactions amongst the 10 most common amino acids at the **a** position were also measured (Acharya, et al. 2006). This confirmed the preference for asparagines not to pair with hydrophobic amino acids at **a-a',** with asparagine-isoleucine destabilizing the interaction 1000-fold. In contrast, these measurements showed that **a-a'** interactions with polar amino acids such as lysine or arginine paired with asparagine were favorable. The combination of these rules has been used to predict specificity on a genome-wide basis (Vinson, et al. 2002, Fassler, et al. 2002, Deppmann, et al. 2004). Additionally, **a-g'** and **d-e'** electrostatic interactions have been shown to be important in determining specificity (Grigoryan, et al. 2009, Reinke, et al. 2010).

*Modeling of bZIP protein-protein interactions*

To develop a deeper understating of bZIP interaction specificity, it is necessary to measure a large number of interactions and develop models that can predict them. Using a protein array assay, the majority of human bZIPs were measured against one another, which

demonstrated that bZIPs do indeed display interaction specificity (Newman and Keating. 2003). A large number of GCN4 single and double point mutants were also measured using SPOT arrays, though this data is difficult to interpret due to the structural ambiguity of these interacting complexes (Portwich, et al. 2007).

There have been several efforts to develop models that can accurately predict the binding specificity of bZIPs. Using simple rules based on **g-e'** electrostatics or quantitative coupling energies is only partially able to describe this specificity (Newman and Keating. 2003, Fong, et al. 2004). Using a machine learning approach to derive weights from a database of known coiled-coil interactions, 70% of true strong interactions could be predicted at an 8% false negative rate (Fong, et al. 2004). Arndt and coworkers developed a model based on the Vinson coupling energies and trained it on a set of melting temperatures for FOS and JUN family bZIPs and coiled coils selected to bind to either JUN or FOS. This model also included a term for helix propensity, and slightly outperformed the previous models in predicting the array interactions (Mason, et al. 2006). A structural modeling approach that also included helix stability and machine learning weights for **a-a'**and **d-d'** interactions also performed quite well (Grigoryan and Keating. 2006). While these models perform well in discriminating strong interactions from non-binders, they are not fully accurate at this task. Further, they are unable to perform more challenging tasks such as predicting the affinity of interactions. To improve models it would be useful to have a large, quantitative, and diverse data set. This additional data would be useful both to further benchmark models based on structure, as well as to train machine-learning based approaches.

*Design of synthetic bZIPs*

There has been a long standing interest in designing synthetic coiled coils that can bind to native bZIPs or be used as molecular parts. Vinson and coworkers generated dominant negative inhibitors of bZIPs by appending an acidic extension to a native leucine zipper (A-ZIPs) (Krylov, et al. 1995). They showed that these A-ZIPs would target bZIPs with the same specificity of the fused leucine zipper but with increased affinity. Several studies have demonstrated that A-ZIPs can prevent bZIPs from binding DNA and are useful *in vivo* (Krylov, et al. 1995, Ahn, et al. 1998, Gerdes, et al. 2006, Acharya, et al. 2006, Oh, et al. 2007). Since most human bZIPs interact with at least several other bZIPs, most human bZIPs cannot be targeted specifically using this approach (Newman and Keating. 2003). To attempt to design more stable and specific leucine zippers against either FOS or JUN, PCA in bacteria was used to select synthetic binders out of peptides libraries. While these selected peptides did bind with greater affinity than their native counterparts, they were not very specific for binding to JUN vs. FOS vs. themselves (Mason, et al. 2006). By expressing a competitive off-target peptide, the authors were able to adapt the selections to generate slightly more specific binders (Mason, et al. 2007). It is unclear how useful this approach is, since if the number of potential off-targets is large it would be difficult to apply this to more than several competitors.

The first attempt to reengineer bZIPs with defined specificities for use as molecular parts was that of peptide 'Velcro' (O'Shea, et al. 1993). Additional work has generated pairs of peptides that have a range of affinities as well as pairs that are orthogonal to one another (Moll, et al. 2001, Lai, et al. 2004, Bromley, et al. 2009, Diss and Kennan. 2008a, Diss and Kennan. 2008b). Native and designed synthetic coiled coils have been useful for making artificial

transcription factors, rewiring cellular pathways, and assembling nano-scale fibers (Mapp, et al. 2000, Wolfe, et al. 2003, McAllister, et al. 2008, Bashor, et al. 2008).

*Research approach*

In my thesis work I focused on understanding the specificity of interactions of native and designed bZIP coiled coils using high-throughput measurement techniques. In chapter 2, I describe the measurement of interactions between viral and host bZIPs. Four bZIPs, each encoded by an oncogenic virus, were measured against a representative panel of 33 human bZIPs. Most previously reported interactions were observed and several novel interactions were identified. Two of the viral bZIPs, MEQ and HBZ, interact with multiple human partners and have unique interaction profiles compared to any human bZIP, whereas the other two viral bZIPs, K-bZIP and BZLF1, display homo-specificity. In chapter 2 and appendix D, I describe experimental characterization of inhibitors that can prevent the viral bZIPs MEQ and bZLF1 from binding to DNA. In chapter 3, a novel computational method was used by my collaborator to design peptides that would specifically bind to target human bZIP proteins, yet not interact with other human bZIPs or self-associate. I tested 48 of these designs for their ability to interact specifically with the intended target. Of the 20 human bZIP families targeted, designs for 8 of the families bound the target more tightly than they bound to any other family. This represents the first large-scale computational design and testing of peptides that interact specifically with native targets. In chapter 4 I describe the measured interactions of 48 designed synthetic peptides as well as 7 human bZIPs to generate a 55-member synthetic protein interactome. This interaction network contains many sub-networks consisting of 3 to 6 protein nodes. Of special interest are pairs of interactions that act orthogonally to one another, as these could have many applications

in molecular engineering. I characterized two such sets of orthogonal heterodimers using solution assays and x-ray crystallography. In chapter 5, I quantitatively measured bZIP protein-protein interaction networks for 7 species (five metazoans and two single-cell organisms) using a high-throughput FRET assay. The 5 metazoan species contain a core set of interactions that is invariantly conserved. Interestingly, while all the networks contain this set of core interactions, each species network is diversified, both through rewiring of interactions between conserved proteins as well as the addition of new proteins and interactions. To understand the sequence changes that lead to changes in interactions, several examples of paralogs with different interaction specificities were identified. Mutants containing a small number of sequence changes were observed to largely switch interaction profiles between paralogs. Taken together, these projects have identified many new interactions, generated specific peptide reagents, identified sequence determinants of interaction specificity, and provided large data sets that will be useful for further understanding the specificity of bZIP proteins.

## REFERENCES

1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006 Sep 26;45(38):11324-32.

2. Acharya A, Ruvinov SB, Gal J, Moll JR, Vinson C. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: Amino acids I, V, L, N, A, and K. Biochemistry. 2002 Dec 3;41(48):14122-31.

3. Acharya A, Rishi V, Moll J, Vinson C. Experimental identification of homodimerizing B-ZIP families in homo sapiens. Journal of Structural Biology. 2006;155(2):130-9.

4. Agre P, Johnson PF, McKnight SL. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. Science. 1989 Nov 17;246(4932):922-6.

5. Ahn S, Olive M, Aggarwal S, Krylov D, Ginty DD, Vinson C. A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. Mol Cell Biol. 1998 Feb;18(2):967-77.

6. Amoutzias GD, Veron AS, Weiner J,3rd, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL. One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. Mol Biol Evol. 2007 Mar;24(3):827-35.

7. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang HC, Hirai A, Tsuzuki K, Nakamura S, Altaf-Ul-Amin M, Oshima T, Baba T, Yamamoto N, Kawamura T, Ioka-Nakamichi T, Kitagawa M, Tomita M, Kanaya S, Wada C, Mori H. Large-scale identification of protein-protein interaction of escherichia coli K-12. Genome Res. 2006 May;16(5):686-91.

8. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: A critical review. Anal Bioanal Chem. 2007 Oct;389(4):1017-31.

9. Bashor CJ, Helman NC, Yan S, Lim WA. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. Science. 2008 March 14, 2008;319(5869):1539-43.

10. Boersma YL, Pluckthun A. DARPins and other repeat protein scaffolds: Advances in engineering and applications. Curr Opin Biotechnol. 2011 Dec;22(6):849-57.

11. Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, Mana M, de Lichtervelde L, Mul JD, van de Peut D, Devos M, Simonis N, Yildirim MA, Cokol M, Kao HL, de Smet AS, Wang H, Schlaitz AL, Hao T, Milstein S, Fan C, Tipsword M, Drew K, Galli M, Rhrissorrakrai K, Drechsel D, Koller D, Roth FP, Iakoucheva LM, Dunker AK, Bonneau R, Gunsalus KC, Hill DE, Piano F, Tavernier J, van den Heuvel S, Hyman AA, Vidal M. A protein domain-based interactome network for C. elegans early embryogenesis. Cell. 2008 Aug 8;134(3):534-45.

12. Briant DJ, Murphy JM, Leung GC, Sicheri F. Rapid identification of linear protein domain binding motifs using peptide SPOT arrays. Methods Mol Biol. 2009;570:175-85.

13. Bromley EHC, Sessions RB, Thomson AR, Woolfson DN. Designed alpha-helical tectons for constructing multicomponent synthetic biological systems. J Am Chem Soc. 2009;131(3):928-30.

14. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A. Interaction network containing conserved and essential protein complexes in escherichia coli. Nature. 2005 Feb 3;433(7025):531-7.

15. Castagnoli L, Costantini A, Dall'Armi C, Gonfloni S, Montecchi-Palazzi L, Panni S, Paoluzi S, Santonico E, Cesareni G. Selectivity and promiscuity in the interaction network mediated by protein recognition modules. FEBS Lett. 2004 Jun 1;567(1):74-9.

16. Chang BH, Gujral TS, Karp ES, BuKhalid R, Grantcharova VP, MacBeath G. A systematic family-wide investigation reveals that ~30% of mammalian PDZ domains engage in PDZ-PDZ interactions. Chem Biol. 2011 Sep 23;18(9):1143-52.

17. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G. Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol. 2008 Sep;26(9):1041-5.

18. Chen YC, Rajagopala SV, Stellberger T, Uetz P. Exhaustive benchmarking of the yeast two-hybrid system. Nat Methods. 2010 Sep;7(9):667,8; author reply 668.

19. Curran T, Peters G, Van Beveren C, Teich NM, Verma IM. FBJ murine osteosarcoma virus: Identification and molecular cloning of biologically active proviral DNA. J Virol. 1982 Nov;44(2):674-82.

20. Deng M, Mehta S, Sun F, Chen T. Inferring domain-domain interactions from protein-protein interactions. Genome Res. 2002 Oct;12(10):1540-8.

21. Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, Taparowsky EJ, Vinson C. Dimerization specificity of all 67 B-ZIP motifs in arabidopsis thaliana: A comparison to homo sapiens B-ZIP motifs. Nucleic Acids Res. 2004 Jun 29;32(11):3435-45.

22. Diss ML, Kennan AJ. Simultaneous directed assembly of three distinct heterodimeric coiled coils. Org Lett. 2008a Sep 4;10(17):3797-800.

23. Diss ML, Kennan AJ. Orthogonal recognition in dimeric coiled coils via buried polar-group modulation. Journal of the American Chemical Society. 2008b;130(4):1321-7.

24. Dreze M, Charloteaux B, Milstein S, Vidalain PO, Yildirim MA, Zhong Q, Svrzikapa N, Romero V, Laloux G, Brasseur R, Vandenhaute J, Boxem M, Cusick ME, Hill DE, Vidal M. 'Edgetic' perturbation of a C. elegans BCL2 ortholog. Nat Methods. 2009 Nov;6(11):843-9.

25. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: Crystal structure of the protein-DNA complex. Cell. 1992 Dec 24;71(7):1223-37.

26. Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, Sidhu SS. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol Biosyst. 2010 Oct;6(10):1782-90.

27. Ernst A, Sazinsky SL, Hui S, Currell B, Dharsee M, Seshagiri S, Bader GD, Sidhu SS. Rapid evolution of functional complexity in a domain family. Sci Signal. 2009 Sep 8;2(87):ra50.

28. Fassler J, Landsman D, Acharya A, Moll JR, Bonovich M, Vinson C. B-ZIP proteins encoded by the drosophila genome: Evaluation of potential dimerization partners. Genome Res. 2002 Aug;12(8):1190-200.

29. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, Tiollais P, Transy C, Legrain P. A genomic approach of the hepatitis C virus generates a protein interaction map. Gene. 2000 Jan 25;242(1-2):369-79.

30. Fong J, Keating A, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biology. 2004;5(2):R11.

31. Gavin AC, Maeda K, Kuhner S. Recent advances in charting protein-protein interaction: Mass spectrometry-based approaches. Curr Opin Biotechnol. 2011 Feb;22(1):42-9.

32. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006 Mar 30;440(7084):631-6.

33. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002 Jan 10;415(6868):141-7.

34. Gentz R, Rauscher FJ,3rd, Abate C, Curran T. Parallel association of fos and jun leucine zippers juxtaposes DNA binding domains. Science. 1989 Mar 31;243(4899):1695-9.

35. Gerdes MJ, Myakishev M, Frost NA, Rishi V, Moitra J, Acharya A, Levy MR, Park SW, Glick A, Yuspa SH, Vinson C. Activator protein-1 activity regulates epithelial tumor cell identity. Cancer Res. 2006 Aug 1;66(15):7578-88.

36. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL,Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. A protein interaction map of drosophila melanogaster. Science. 2003 Dec 5;302(5651):1727-36.

37. Glover JN, Harrison SC. Crystal structure of the heterodimeric bZIP transcription factor c-fos-c-jun bound to DNA. Nature. 1995 Jan 19;373(6511):257-61.

38. Goll J, Uetz P. The elusive yeast interactome. Genome Biol. 2006;7(6):223.

39. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009 Apr 16;458(7240):859-64.

40. Grigoryan G, Keating AE. Structure-based prediction of bZIP partnering specificity. J Mol Biol. 2006 Feb 3;355(5):1125-42.

41. Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O, McKillip E, Shah S, Stapleton M, Wan KH, Yu C, Parsa B, Carlson JW, Chen X, Kapadia B, VijayRaghavan K, Gygi SP, Celniker SE, Obar RA, Artavanis-Tsakonas S. A protein complex network of drosophila melanogaster. Cell. 2011 Oct 28;147(3):690-703.

42. Hai T, Curran T. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc Natl Acad Sci U S A. 1991 May 1;88(9):3720-4.

43. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004 Jul 1;430(6995):88-93.

44. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science. 1993 November 26, 1993;262(5138):1401-7.

45. Hinnebusch AG, Fink GR. Positive regulation in the general amino acid control of saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1983 Sep;80(17):5374-8.

46. Hirai S, Bourachot B,Yaniv M. Both jun and fos contribute to transcription activation by the heterodimer. Oncogene. 1990 Jan;5(1):39-46.

47. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry. Nature. 2002 Jan 10;415(6868):180-3.

48. Hope IA, Struhl K. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J. 1987 Sep;6(9):2781-4.

49. Hope IA, Struhl K. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell. 1986 Sep 12;46(6):885-94.

50. Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. Global functional atlas of escherichia coli encompassing previously uncharacterized proteins. PLoS Biol. 2009 Apr 28;7(4):e96.

51. Huang H, Li L, Wu C, Schibli D, Colwill K, Ma S, Li C, Roy P, Ho K, Songyang Z, Pawson T, Gao Y, Li SS. Defining the specificity space of the human SRC homology 2 domain. Mol Cell Proteomics. 2008 Apr;7(4):768-84.

52. Hutchins JR, Toyoda Y, Hegemann B, Poser I, Heriche JK, Sykora MM, Augsburg M, Hudecz O, Buschhorn BA, Bulkescher J, Conrad C, Comartin D, Schleiffer A, Sarov M, Pozniakovsky A, Slabicki MM, Schloissnig S, Steinmacher I, Leuschner M, Ssykor A, Lawo S, Pelletier L, Stark H, Nasmyth K, Ellenberg J, Durbin R, Buchholz F, Mechtler K, Hyman AA, Peters JM. Systematic analysis of human protein complexes identifies chromosome segregation proteins. Science. 2010 Apr 30;328(5978):593-9.

53. Im YJ, Lee JH, Park SH, Park SJ, Rho SH, Kang GB, Kim E, Eom SH. Crystal structure of the shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. J Biol Chem. 2003 Nov 28;278(48):48099-104.

54. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A. 2001 Apr 10;98(8):4569-74.

55. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci U S A. 2000 Feb 1;97(3):1143-7.

56. Jensen LJ, Bork P. Biochemistry. not comparable, but complementary. Science. 2008 Oct 3;322(5898):56-7.

57. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006 Jan 12;439(7073):168-74.

58. Kouzarides T, Ziff E. Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. Nature. 1989 Aug 17;340(6234):568-71.

59. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF. Global landscape of protein complexes in the yeast saccharomyces cerevisiae. Nature. 2006 Mar 30;440(7084):637-43.

60. Krylov D, Olive M, Vinson C. Extending dimerization interfaces: The bZIP basic region can form a coiled coil. EMBO J. 1995 Nov 1;14(21):5329-37.

61. Krylov D, Mikhailenko I, Vinson C. A thermodynamic scale for leucine zipper stability and dimerization specificity: E and g interhelical interactions. EMBO J. 1994 Jun 15;13(12):2849-61.

62. Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castano-Diez D, Chen WH, Devos D, Guell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Bottcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC. Proteome organization in a genome-reduced bacterium. Science. 2009 Nov 27;326(5957):1235-40.

63. Kung LA, Snyder M. Proteome chips for whole-organism assays. Nat Rev Mol Cell Biol. 2006 Aug;7(8):617-22.

64. Lai C, Ting L. Transcriptional repression of human hepatitis B virus genes by a bZIP family member, E4BP4. J. Virol. 1999 April 1, 1999;73(4):3197-209.

65. Lai JR, Fisk JD, Weisblum B, Gellman SH. Hydrophobic core repacking in a coiled-coil dimer via phage display: Insights into plasticity and specificity at a protein-protein interface. Journal of the American Chemical Society. 2004;126(34):10514-5.

66. Landgraf C, Panni S, Montecchi-Palazzi L, Castagnoli L, Schneider-Mergener J, Volkmer-Engert R, Cesareni G. Protein interaction networks by proteome peptide scanning. PLoS Biol. 2004 Jan;2(1):E14.

67. Landschulz WH, Johnson PF, McKnight SL. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science. 1989 Mar 31;243(4899):1681-8.

68. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. Science. 1988 Jun 24;240(4860):1759-64.

69. Lenfant N, Polanowska J, Bamps S, Omi S, Borg JP, Reboul J. A genome-wide study of PDZ-domain interactions in C. elegans reveals a high frequency of non-canonical binding. BMC Genomics. 2010 Nov 26;11:671.

70. Letunic I, Doerks T, Bork P. SMART 7: Recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012 Jan;40(1):D302-5.

71. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M. A map of the interactome network of the metazoan C. elegans. Science. 2004 Jan 23;303(5657):540-3.

72. Lievens S, Vanderroost N, Van der Heyden J, Gesellchen V, Vidal M, Tavernier J. Array MAPPIT: High-throughput interactome analysis in mammalian cells. J Proteome Res. 2009 Feb;8(2):877-86.

73. Liu BA, Jablonowski K, Shah EE, Engelmann BW, Jones RB, Nash PD. SH2 domains recognize contextual peptide sequence information to determine selectivity. Mol Cell Proteomics. 2010 Nov;9(11):2391-404.

74. Liu BA, Jablonowski K, Raina M, Arce M, Pawson T, Nash PD. The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. Mol Cell. 2006 Jun 23;22(6):851-68.

75. Liu J, Rost B. Comparing function and structure between entire proteomes. Protein Sci. 2001 October 1, 2001;10(10):1970-9.

76. Lumb KJ, Kim PS. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. Biochemistry. 1995;34(27):8642-8.

77. Maki Y, Bos TJ, Davis C, Starbuck M, Vogt PK. Avian sarcoma virus 17 carries the jun oncogene. Proc Natl Acad Sci U S A. 1987 May;84(9):2848-52.

78. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G, Kim BJ, Li C, Chen R, Li W, Wang Y, O'Malley BW, Qin J. Analysis of the human endogenous coregulator complexome. Cell. 2011 May 27;145(5):787-99.

79. Mapp AK, Ansari AZ, Ptashne M, Dervan PB. Activation of gene expression by small molecule transcription factors. Proceedings of the National Academy of Sciences of the United States of America. 2000 April 11, 2000;97(8):3930-5.

80. Mason JM, Muller KM, Arndt KM. Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. Biochemistry. 2007 Apr 24;46(16):4804-14.

81. Mason JM, Schmitz MA, Muller KM, Arndt KM. Semirational design of jun-fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. Proc Natl Acad Sci U S A. 2006 Jun 13;103(24):8989-94.

82. McAllister KA, Zou H, Cochran FV, Bender GM, Senes A, Fry HC, Nanda V, Keenan PA, Lear JD, Saven JG, Therien MJ, Blasie JK, DeGrado WF. Using α-helical coiled-coils to design nanostructured metalloporphyrin arrays. Journal of the American Chemical Society. 2008;130(36):11921.

83. McCraith S, Holtzman T, Moss B, Fields S. Genome-wide analysis of vaccinia virus protein-protein interactions. Proc Natl Acad Sci U S A. 2000 Apr 25;97(9):4879-84.

84. Michnick SW, Ear PH, Landry C, Malleshaiah MK, Messier V. Protein-fragment complementation assays for large-scale analysis, functional dissection and dynamic studies of protein-protein interactions in living cells. Methods Mol Biol. 2011;756:395-425.

85. Miller BW, Lau G, Grouios C, Mollica E, Barrios-Rodiles M, Liu Y, Datti A, Morris Q, Wrana JL, Attisano L. Application of an integrated physical and functional screening approach to identify inhibitors of the wnt pathway. Mol Syst Biol. 2009;5:315.

86. Moitra J, Szilak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. Biochemistry. 1997 Oct 14;36(41):12567-73.

87. Moll JR, Ruvinov SB, Pastan I, Vinson C. Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10-15 M. Protein Sci. 2001 March 1, 2001;10(3):649-55.

88. Newman JR, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 Jun 27;300(5628):2097-101.

89. Oh WJ, Rishi V, Orosz A, Gerdes MJ, Vinson C. Inhibition of CCAAT/Enhancer binding protein family DNA binding in mouse epidermis prevents and regresses papillomas. Cancer Res. 2007 February 15, 2007;67(4):1867-76.

90. O'Neil KT, Hoess RH, DeGrado WF. Design of DNA-binding peptides based on the leucine zipper motif. Science. 1990 Aug 17;249(4970):774-8.

91. O'Shea EK, Lumb KJ, Kim PS. Peptide 'velcro': Design of a heterodimeric coiled coil. Curr Biol. 1993 Oct 1;3(10):658-67.

92. O'Shea EK, Rutkowski R, Kim PS. Mechanism of specificity in the fos-jun oncoprotein heterodimer. Cell. 1992 Feb 21;68(4):699-708.

93. O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. Science. 1989 Jan 27;243(4890):538-42.

94. Parrish JR, Yu J, Liu G, Hines JA, Chan JE, Mangiola BA, Zhang H, Pacifico S, Fotouhi F, DiRita VJ, Ideker T, Andrews P, Finley RL,Jr. A proteome-wide protein interaction map for campylobacter jejuni. Genome Biol. 2007;8(7):R130.

95. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. Science. 2003 Apr 18;300(5618):445-52.

96. Petschnigg J, Snider J, Stagljar I. Interactive proteomics research technologies: Recent applications and advances. Curr Opin Biotechnol. 2011 Feb;22(1):50-8.

97. Portwich M, Keller S, Strauss HM, Mahrenholz CC, Kretzschmar I, Kramer A, Volkmer R. A network of coiled-coil associations derived from synthetic GCN4 leucine-zipper arrays. Angew Chem Int Ed Engl. 2007;46(10):1654-7.

98. Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, Pelletier L, Kittler R, Hua S, Naumann R, Augsburg M, Sykora MM, Hofemeister H, Zhang Y, Nasmyth K, White KP, Dietzel S, Mechtler K, Durbin R, Stewart AF, Peters JM, Buchholz F, Hyman AA. BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. Nat Methods. 2008 May;5(5):409-15.

99. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P. The protein-protein interaction map of helicobacter pylori. Nature. 2001 Jan 11;409(6817):211-5.

100. Rajagopala SV, Uetz P. Analysis of protein-protein interactions using high-throughput yeast two-hybrid screens. Methods Mol Biol. 2011;781:1-29.

101. Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J Am Chem Soc. 2010 May 5;132(17):6025-31.

102. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE,

Roth FP, Vidal M. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005 Oct 20;437(7062):1173-8.

103. Sassone-Corsi P, Ransone LJ, Lamph WW, Verma IM. Direct interaction between fos and jun nuclear oncoproteins: Role of the 'leucine zipper' domain. Nature. 1988 Dec 15;336(6200):692-5.

104. Sellers JW, Struhl K. Changing fos oncoprotein to a jun-independent DNA binding protein with GCN4 dimerization specificity by swapping "leucine zippers". Nature. 1989 Sep 7;341(6237):74-6.

105. Shao X, Tan CS, Voss C, Li SS, Deng N, Bader GD. A regression framework incorporating quantitative and negative interaction data improves quantitative prediction of PDZ domain-peptide interaction from primary sequence. Bioinformatics. 2011 Feb 1;27(3):383-90.

106. Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmens I, Hirozane-Kishikawa T, Hao T, Sahalie JM, Venkatesan K, Gebreab F, Cevik S, Klitgord N, Fan C, Braun P, Li N, Ayivi-Guedehoussou N, Dann E, Bertin N, Szeto D, Dricot A, Yildirim MA, Lin C, de Smet AS, Kao HL, Simon C, Smolyar A, Ahn JS, Tewari M, Boxem M, Milstein S, Yu H, Dreze M, Vandenhaute J, Gunsalus KC, Cusick ME, Hill DE, Tavernier J, Roth FP, Vidal M. Empirically controlled mapping of the caenorhabditis elegans protein-protein interactome network. Nat Methods. 2009 Jan;6(1):47-54.

107. Smith CA, Kortemme T. Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. J Mol Biol. 2010 Sep 17;402(2):460-74.

108. Snider J, Kittanakom S, Damjanovic D, Curak J, Wong V, Stagljar I. Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. Nat Protoc. 2010 Jul;5(7):1281-93.

109. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. A human protein-protein interaction network: A resource for annotating the proteome. Cell. 2005 Sep 23;122(6):957-68.

110. Stengel F, Aebersold R, Robinson CV. Joining forces: Integrating proteomics and crosslinking with the mass spectrometry of intact complexes. Mol Cell Proteomics. 2011 Dec 16

111. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. Science. 2007 Jul 20;317(5836):364-9.

112. Stiffler MA, Grantcharova VP, Sevecka M, MacBeath G. Uncovering quantitative protein interaction networks for mouse PDZ domains using protein microarrays. J Am Chem Soc. 2006 May 3;128(17):5913-22.

113. Talanian RV, McKnight CJ, Kim PS. Sequence-specific DNA binding by a short peptide dimer. Science. 1990 Aug 17;249(4970):769-71.

114. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, Malitskaya Y, Vogel J, Bussey H, Michnick SW. An in vivo map of the yeast protein interactome. Science. 2008 Jun 13;320(5882):1465-70.

115. Titz B, Rajagopala SV, Goll J, Hauser R, McKevitt MT, Palzkill T, Uetz P. The binary protein interactome of treponema pallidum--the syphilis spirochete. PLoS One. 2008 May 28;3(5):e2292.

116. Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S, Quondam M, Zucconi A, Hogue CW, Fields S, Boone C, Cesareni G. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science. 2002 Jan 11;295(5553):321-4.

117. Tonikian R, Xin X, Toret CP, Gfeller D, Landgraf C, Panni S, Paoluzi S, Castagnoli L, Currell B, Seshagiri S, Yu H, Winsor B, Vidal M, Gerstein MB, Bader GD, Volkmer R, Cesareni G, Drubin DG, Kim PM, Sidhu SS, Boone C. Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol. 2009 Oct;7(10):e1000218.

118. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS. A specificity map for the PDZ domain family. PLoS Biol. 2008 Sep 30;6(9):e239.

119. Turner R, Tjian R. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. Science. 1989 Mar 31;243(4899):1689-94.

120. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. Nature. 2000 Feb 10;403(6770):623-7.

121. van Straaten F, Muller R, Curran T, Van Beveren C, Verma IM. Complete nucleotide sequence of a human c-onc gene: Deduced amino acid sequence of the human c-fos protein. Proc Natl Acad Sci U S A. 1983 Jun;80(11):3183-7.

122. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabasi AL, Vidal M. An empirical framework for binary interactome mapping. Nat Methods. 2009 Jan;6(1):83-90.

123. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, Bonovich M. Classification of human B-ZIP proteins based on dimerization properties. Mol Cell Biol. 2002 Sep;22(18):6321-35.

124. Vinson CR, Hai T, Boyd SM. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: Prediction and rational design. Genes Dev. 1993 Jun;7(6):1047-58.

125. Vinson CR, Sigler PB, McKnight SL. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science. 1989 Nov 17;246(4932):911-6.

126. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M. Protein interaction mapping in C. elegans using proteins involved in vulval development. Science. 2000 Jan 7;287(5450):116-22.

127. Wiedemann U, Boisguerin P, Leben R, Leitner D, Krause G, Moelling K, Volkmer-Engert R, Oschkinat H. Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. J Mol Biol. 2004 Oct 22;343(3):703-18.

128. Wolfe SA, Grant RA, Pabo CO. Structure of a designed dimeric zinc finger protein bound to DNA. Biochemistry. 2003;42(46):13401-9.

129. Xin X, Rual JF, Hirozane-Kishikawa T, Hill DE, Vidal M, Boone C, Thierry-Mieg N. Shifted transversal design smart-pooling for high coverage interactome mapping. Genome Res. 2009 Jul;19(7):1262-9.

130. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet AS, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabasi AL, Tavernier J, Hill DE, Vidal M. High-quality binary protein interaction map of the yeast interactome network. Science. 2008 Oct 3;322(5898):104-10.

131. Zeng X, Herndon AM, Hu JC. Buried asparagines determine the dimerization specificities of leucine zipper mutants. Proc Natl Acad Sci U S A. 1997 Apr 15;94(8):3673-8.

# CHAPTER 2

# Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays

**Reproduced with permission from:**
Reinke AW, Grigoryan G, Keating AE. Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. Biochemistry. 2010 Mar 9;49(9):1985-97.

**Supporting information:**
This paper included supplemental tables and figures which are in appendix A.

**Collaborator notes:**
Gevorg Grigoryan computationally designed the anti-MEQ peptide.

**ABSTRACT**

Basic-region leucine-zipper transcription factors (bZIPs) contain a segment rich in basic amino acids that can bind DNA, followed by a leucine zipper that can interact with other leucine zippers to form coiled-coil homo- or heterodimers. Several viruses encode proteins containing bZIP domains, including four that encode bZIPs lacking significant homology to any human protein. We investigated the interaction specificity of these four viral bZIPs by using coiled-coil arrays to assess self-associations as well as hetero-interactions with 33 representative human bZIPs. The arrays recapitulated reported viral-human interactions and also uncovered new associations. MEQ and HBZ interacted with multiple human partners and had unique interaction profiles compared to any human bZIPs, whereas K-bZIP and BZLF1 displayed homo-specificity. New interactions detected included HBZ with MAFB, MAFG, ATF2, CEBPG, and CREBZF, and MEQ with NFIL3. These were confirmed in solution using circular dichroism. HBZ can hetero-associate with MAFB and MAFG in the presence of MARE-site DNA, and this interaction is dependent on the basic region of HBZ. NFIL3 and MEQ have different yet overlapping DNA-binding specificities and can form a heterocomplex with DNA. Computational design considering both affinity for MEQ and specificity with respect to other undesired bZIP-type interactions was used to generate a MEQ dimerization inhibitor. This peptide, anti-MEQ, bound MEQ both stably and specifically, as assayed using coiled-coil arrays and circular dichroism in solution. Anti-MEQ also inhibited MEQ binding to DNA. These studies can guide further investigation of the function of viral and human bZIP complexes.

## INTRODUCTION

Many viruses hijack cellular machinery by using viral proteins to interact with host proteins. Viruses can incorporate host protein domains into their genomes for this purpose, as is the case for several viruses that use BCL-2 homologs to prevent apoptosis. Viral host-derived protein domains often make interactions similar to those of their homologues, although these can occur in a misregulated manner (Hardwick and Bellows. 2003). Alternatively, host-derived protein domains can diverge from their cellular counterparts, such that they retain little sequence similarity. In such cases, virus-host protein interactions can be expected to differ markedly from corresponding host-host complexes (Kvansakul, et al. 2008).

The bZIP transcription factors are a large class of proteins found in most eukaryotic organisms. Named for their DNA-binding and dimerization domain, bZIP proteins interact with DNA site-specifically via a region of conserved basic amino acids. Immediately C-terminal to the basic region is the leucine zipper, a coiled coil that mediates the formation of homodimeric or heterodimeric complexes. The dimerization specificity of the leucine zippers allows for combinatorial interactions that can influence DNA binding and thus transcriptional regulation (Daury, et al. 2001, Hai and Curran. 1991). Given the importance of protein partnering specificity for the function of the bZIPs, a high-throughput protein array assay was used to determine the global *in vitro* interaction profiles of most human bZIPs. The coiled-coil microarray assay used for this purpose was shown to identify most reported interactions, and the relative stabilities of interactions measured on the arrays were also shown to agree well with solution measurements (Grigoryan, et al. 2009a, Newman and Keating. 2003).

Proteins containing bZIP domains have been identified in several viruses. Three human bZIP proteins, JUN (cJun), FOS (cFos), and MAF (cMaf), occur in an altered form in oncogenic

avian and murine retroviruses. These homologous viral bZIPs maintain the protein dimerization properties of the human proteins and are oncogenic because of altered regulation (van Straaten, et al. 1983a, Bos, et al. 1989, Kataoka, et al. 1993). Four viral bZIPs that have little homology to human bZIPs have also been identified, and although several interactions with host proteins have been reported, global investigation of the interactions of these proteins with host bZIPs is lacking.

Human T-cell leukemia virus type 1 is a retrovirus that causes adult T-cell leukemia; it encodes the bZIP protein HBZ (reviewed in (Mesnard, et al. 2006)). HBZ has been shown to repress both viral and cellular gene expression. A recent study suggests that in addition to the role of the HBZ protein in disease progression, the mRNA of HBZ promotes proliferation (Satou, et al. 2006). Interactions have been reported between HBZ and many human bZIPs both *in vivo* and *in vitro* including ATF4, JUN, JUNB, JUND, CREB1, and ATF1 (Lemasson, et al. 2007, Thebault, et al. 2004, Basbous, et al. 2003, Gaudray, et al. 2002). HBZ has been shown to form complexes with JUN, JUNB, CREB1, and ATF4 and to prevent these proteins from binding DNA. In contrast, HBZ has been reported to increase the transcriptional activity of JUND (Thebault, et al. 2004).

MEQ is encoded by Marek's disease virus (MDV), an oncogenic herpes virus that infects chickens. The disease is estimated to cost the US poultry industry one billion dollars annually (reviewed in (Nair. 2005)). MEQ has been demonstrated to be largely responsible for the oncogenic properties of MDV (Suchodolski, et al. 2009, Levy, et al. 2005, Brown, et al. 2009). MEQ can self-associate as well as interact with a variety of other bZIPs *in vitro* including: JUN, JUNB, CREB1, ATF1, ATF2, ATF3, FOS, and BATF3 (Suchodolski, et al. 2009, Levy, et al.

2003, Qian, et al. 1996, Qian, et al. 1995). Additionally, MEQ has been shown to bind JUN *in vivo*, and JUN is required for MEQ to transform cells (Levy, et al. 2005, Levy, et al. 2003).

Two gammaherpesviruses are reported to encode bZIP-containing proteins. These viruses are implicated in several proliferative disorders in humans. Epstein-Barr virus encodes BZLF1 (ZEBRA, Zta, Z, EB1) and is associated with Burkitt's lymphoma, Hodgkin's disease, and nasopharyngeal carcinoma. Kaposi's sarcoma-associated herpesvirus encodes K-bZIP (K8, RAP) and is involved in Kaposi's sarcoma, primary effusion lymphoma, and multicentric Castleman's disease (reviewed in (Thomas. 2006, Kutok and Wang. 2006)). These two proteins are positional homologs and also share low sequence similarity with one another (Lin, et al. 1999). BZLF1 is responsible for triggering the switch from latent to lytic infection by binding sites within the viral genome and causing transcriptional activation of many genes. It is also involved in viral replication (Countryman, et al. 1987, Schepers, et al. 1996). Over-expression of K-bZIP does not cause virus reactivation, but K-bZIP is necessary for viral replication as well as the repression and activation of many genes within the viral genome, though not always through direct binding to promoters (Rossetto, et al. 2009, Ellison, et al. 2009). BZLF1 and K-bZIP both interact with many viral and cellular proteins including the human bZIP CEBPA (C/EBPα) (Sinclair. 2003). Interestingly, the interaction with CEBPA for both BZLF1 and K-bZIP is proposed to involve higher order oligomers rather than just dimers (Wu, et al. 2004, Wu, et al. 2003). Recently, the crystal structure of BZLF1 was solved showing that a C-terminal region adjacent to the leucine zipper folds back and stabilizes the coiled-coil structure, significantly stabilizing the homodimer (Petosa, et al. 2006). K-bZIP has been shown to self associate through its leucine zipper, but this homomeric interaction was reported to be one of higher order oligomers (Lin, et al. 1999, Wu, et al. 2003).

Given the importance of both human and viral bZIPs to human health, several strategies have been used to generate inhibitors that can prevent dimerization and/or DNA binding. One approach is to use the leucine zipper of a homodimerizing bZIP as a dominant negative. The utility of this approach has been demonstrated in the context of BZLF1. Using a peptide that consisted of only the leucine zipper of BZLF1, (Hicks, et al. 2003) showed that BZLF1 could be prevented from binding DNA. However, the $EC_{50}$ for the peptide was high micromolar. A possible disadvantage of using native leucine-zipper peptides as inhibitors is that these may associate with bZIPs other than the desired target. Recently, we reported a computational design method for obtaining peptides that interact specifically with leucine zippers and applied it to a range of human targets. Out of the 20 human bZIP families, peptides were designed that successfully interacted with 19. Of these 19, 8 designs bound to their target family stronger than to any other family (Grigoryan, et al. 2009a).

Here we report all pair-wise interactions of four bZIP peptides derived from viral proteins with 33 human bZIP proteins measured using peptide microarrays. We identified several new interactions for both MEQ and HBZ, and these interactions were confirmed using circular dichroism and gel-shift assays. Additionally, we designed a peptide, anti-MEQ, to serve as a MEQ dimerization inhibitor. We demonstrate that this peptide binds specifically to MEQ and can prevent MEQ from binding DNA.

## EXPERIMENTAL METHODS

*Plasmid construction, protein expression and purification*

Human protein constructs used for array experiments have been previously described and are listed in Table A.S1 (Grigoryan, et al. 2009a, Newman and Keating. 2003). Synthetic genes

encoding the leucine zipper regions of HBZ, MEQ, BZLF1, K-bZIP, anti-MEQ and the full bZIP domains of MAFB, HBZ, and MEQ were synthesized using DNAWorks to design primers and a two-step PCR method to anneal them (Hoover and Lubkowski. 2002). The bZIP domains of CREBZF, ATF2 and JUND were cloned from plasmids acquired from Open Biosystems (Gerhard, et al. 2004) and NFIL3, JUN, CEBPG and MAFG were cloned from plasmids obtained from PlasmID (Witt, et al. 2006).These proteins were cloned into modified versions of a pDEST17 vector. Proteins were expressed in RP3098 cells and purified under denaturing conditions using Ni-NTA followed by reverse-phase HPLC as described previously (Newman and Keating. 2003, Grigoryan, et al. 2009b). A tagless version of anti-MEQ used for gel-shift and CD studies was constructed by cloning into pSV282 (Vanderbilt University Medical Center, Center for Structural Biology). The 6XHIS-MBP-anti-MEQ fusion protein was expressed in RP3098 cells by growing a 1 L culture in LB at 37 °C and inducing at 0.5 OD by adding 1 mM IPTG and growing for 4 hours. The fusion protein was purified under native conditions by binding to Ni-NTA resin (Qiagen) and eluted by adding 8 ml buffer (300 mM imidazole, 20 mM TRIS, 500 mM NaCl, 1 mM DTT, pH 7.9). The fusion protein was then dialyzed overnight into TEV cleavage buffer (50 mM TRIS, 50 mM NaCl, 1 mM DTT, 0.5 mM EDTA, pH 7.5) and then cleaved by adding 100 µl TEV protease (1mg/ml) for 3 hours at 18-22 °C. This mixture was then added to Ni-NTA resin and the flow-through collected. The anti-MEQ peptide was further purified using reverse-phase HPLC. The molecular weights of the peptides were confirmed by mass spectrometry. Protein sequences generated for this study are listed in Table A.S1.

*Coiled-coil arrays*

Array experiments were performed as described previously (Grigoryan, et al. 2009b). The average background-corrected fluorescence values for all measurements are listed in Table A.S2**.** Two measures used to report fluorescence intensities in the figures are $S_{array}$ and *arrayscore*. These are defined in references (Grigoryan, et al. 2009b) and (Reinke, et al. 2010), respectively.

*Circular dichroism*

Circular dichroism experiments were performed as described previously (Grigoryan, et al. 2009b). The concentrations used for each experiment are listed in the figure legends. Thermal dentaturations were measured from 0 to 65 °C and all were reversible with all complexes having differences in $T_m$ of less than 3 °C upon refolding. The buffer for CD measurements of MEQ was PBS (potassium phosphate (pH 7.4) and 150 mM KCl) with 1mM DTT. For measurements of HBZ the buffer also included 200 mM GdnHCl and 0.25 mM EDTA.

*Phylogenetic analysis*

An unrooted phylogenetic tree was constructed using the neighbor-joining method for the 53 human and 4 viral bZIP leucine-zipper regions as described previously (Grigoryan, et al. 2009b). For comparison of chicken and human sequences, each human bZIP was used to BLAST the *G. gallus* genome and 41 chicken bZIPs were identified. Leucine-zipper regions were defined as previously reported (Newman and Keating. 2003). Families were defined according to evolutionary conservation and interaction profiles, as in (Newman and Keating. 2003, Amoutzias, et al. 2007).

*Gel-shift assay*

DNA probes for the AP-1, TFIID, and NF-κB sites were obtained from Promega. Other probes were based on literature-defined sequences (MARE (Kataoka, et al. 1994), CAAT (Acharya, et al. 2006a), CRE1 (Levy, et al. 2003), CRE2 (Chen, et al. 1995), MDVORI (Levy, et al. 2003)),

ordered as PAGE-purified oligos (IDT) and then annealed. Probes were end labeled with [γ-$^{32}$P]ATP using PNK (NEB). Proteins were incubated for 3 hours at 18-22 °C in gel-shift buffer (50 mM KCl, 25 mM TRIS pH 8.0, 0.5 mM EDTA, 2.5 mM DTT, 1 mg/ml BSA, 10% (v/v) glycerol, 0.1 mg/ml competitor DNA (Poly (I)·Poly (C) (GE)). Radiolabeled DNA was then added and incubated for 1 hour at 18-22 °C. Radiolabeled DNA was at a final concentration of 0.7 nM, except for experiments in Figure A.S4 where the final concentration was 20 nM. Protein/DNA mixtures were loaded on NOVEX DNA retardation gels (Invitrogen) using 0.5X TBE buffer and run at 200-300V for 15-25 minutes. For complexes involving JUN proteins, the buffer was pre-cooled to 4 °C to prevent complex dissociation. Gels were dried and imaged using a phosphorimaging screen and a Typhoon 9400 imaging system.

*Computational design of anti-MEQ*

Anti-MEQ was designed using CLASSY with the HP/S/Ca energy function as previously reported (Grigoryan, et al. 2009b). 46 human proteins and design homodimerization were used as negative design states.

**RESULTS**

*Four unique bZIPs are encoded by viral genomes*

There are four bZIPs of viral origin described in the literature that are not closely related to any human bZIP. These are MEQ, HBZ, BZLF1, and K-bZIP. To compare these proteins to the human bZIPs, a phylogenetic tree was constructed using the leucine- zipper regions of the four viral proteins as well as 53 human bZIPs (Figure 2.1A). According to this analysis, all four viral bZIPs are quite diverged from human bZIPs. The sequences of the 4 viral bZIPs are aligned to several representative human sequences in Figure 2.1B. Human bZIPs have a highly

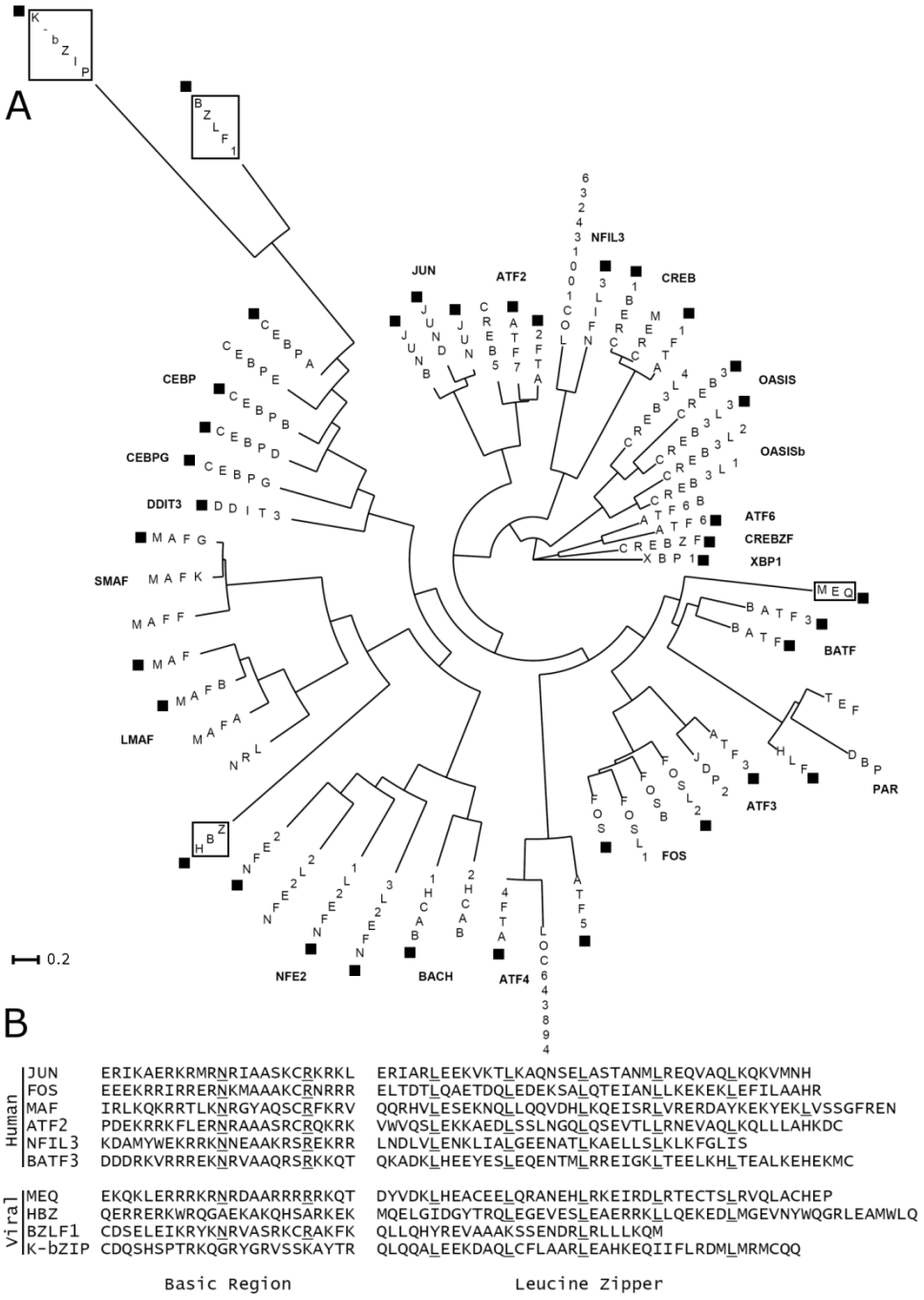



Figure 2.1. Sequence properties of human and viral bZIPs. (A) A phylogenetic tree was inferred by neighbor-joining using only the leucine-zipper region of each of the 53 human bZIPs and the 4 viral bZIPs. Viral sequences are boxed. Proteins used to measure interactions are indicated with a black square. Family names are in bold. The scale bar refers to amino-acid changes per position. (B) Multiple-sequence alignment of viral bZIPs with representative human bZIPs. The following are underlined: Highly conserved basic-region asparagine and arginine residues and conserved leucines in the leucine zippers.

conserved basic region consisting of the motif (R/K)XX(R/K)N(R/K)XAAXX(S/C)RX(R/K)(R/K) (Adya, et al. 1994), and a striking feature of the basic-region alignment is the presence of an invariant asparagine in almost all human bZIPs. The only two human families that do not have this asparagine are CREBZF (ZF, Zhangfei) and DDIT3 (CHOP), which are not known to bind DNA as homodimers but can bind as heterodimers (Hogan, et al. 2006, Ubeda, et al. 1996). An arginine, separated by 8 residues from the conserved asparagine, is strictly conserved in all human bZIPs. Both MEQ and BZLF1 conform well to this conserved motif and include the key asparagine and arginine residues. In contrast, the basic regions from HBZ and K-bZIP poorly match the basic-region motif, and neither contains the key conserved asparagine or arginine. The leucine-zipper regions of human bZIPs are 4-7 heptads long and are characterized by strong conservation of leucine every 7 amino acids. MEQ, HBZ, and to a lesser extent K-bZIP, have mostly canonical leucine-zipper regions. On the other hand, BZLF1 has a very short leucine zipper that is non-canonical, with only one coiled-coil **d**-position leucine (coiled-coil residues are traditionally labeled **a**-**f**, with **a** and **d** largely buried in the core, **e** and **g** on the periphery and **b**, **c** and **f** on the outside of the helical complex). BZLF1 has also been shown to be stabilized by an extended C-terminal region that makes contacts with the coiled coil (Petosa, et al. 2006). These observations are consistent with reports of both MEQ and BZLF1 binding DNA, whereas there is no direct evidence to support binding of DNA by HBZ (Levy, et al. 2003, Petosa, et al. 2006, Hivin, et al. 2006). K-bZIP has been shown to directly bind DNA, though it is not clear whether the bZIP domain is involved (Lefort, et al. 2007).

Unlike HBZ, K-bZIP and BZLF1, which are found in viruses that infect humans, MEQ is encoded by an avian oncovirus. Because of the availability of a large number of human, but not

avian, bZIP clones, we wanted to confirm that human bZIPs could be used as a reasonable substitute for chicken bZIPs. MEQ has been previously reported to interact with both human and mouse bZIP proteins (Levy, et al. 2003). We also compared human bZIP sequences to chicken bZIP sequences and found them to be highly homologous (Figure A.S1). Considering just the coiled-coil interface positions that are most responsible for interactions (**adeg**), 85% of direct orthologues have greater than 90% identity. Additionally, all human bZIP families are conserved between human and chicken, except DDIT3, which is specific to humans.

*Detection of viral-human bZIP interactions*

Interactions between human and viral bZIPs were measured using a previously described protein microarray assay (Newman and Keating. 2003, Grigoryan, et al. 2009b). The leucine-zipper regions of 33 representative human bZIPs were purified and printed onto aldehyde-derivatized glass slides along with leucine zippers from the 4 viral proteins (Table A.S1). All human bZIP families were represented on the arrays except for OASISb, which is very similar in its protein sequences and interaction profiles to OASIS. Each protein was then individually fluorescently labeled and used to probe the arrays at a concentration of ~160 nM, unless otherwise indicated. A total of 8 spots on the surface were used for each measurement. The fluorescence intensity of each spot was corrected for background, and the averages of the 8 values were converted into a score called $S_{array}$, a Z-score like measure, as described previously (Table A.S2) (Grigoryan, et al. 2009b).

As in prior studies using this technique, there were several indications that the data are of good quality. First, interactions observed among human bZIPs (measured simultaneously with the viral-human interaction data) were highly consistent with previously published data (Figure A.S2) (Newman and Keating. 2003). Next, each heteromeric interaction was measured twice,

once when the first protein was on the surface and again when it was in solution. Most interactions were observed in both directions. Further, interactions involving MEQ and HBZ were measured over a large range of concentrations and gave rise to similar interaction profiles (Figure 2.2B). Finally, many interactions observed between viral and human proteins were consistent with prior reports, as discussed below.

Most previously reported interactions involving the viral bZIPs were observed on the array and are indicated by green boxes in Figure 2.2A. Exceptions are boxed with green dotted lines in Figure 2.2A and include the interaction of HBZ with ATF4 and the interaction of MEQ with FOS. However, the HBZ—ATF4 interaction was reported to be weaker with just the leucine-zipper region (as was measured on the arrays) than in context of the entire protein (Gaudray, et al. 2002). Also, the interaction of MEQ with FOS has been shown to be weak compared to other interactions of MEQ (Suchodolski, et al. 2009, Levy, et al. 2003). Several interactions previously reported to not occur were also not observed to interact on the arrays. These include HBZ self interaction, HBZ—FOS, BZLF1—FOS, and BZLF1—JUN (Basbous, et al. 2003, Chang, et al. 1990, Matsumoto, et al. 2005). Both BZLF1 and K-bZIP have been reported to interact with CEBPA, but not as heterodimers (see Discussion).

Many previously unreported interactions were detected for HBZ. New partners included: MAF and MAFB, MAFG, ATF2 and ATF7, CEBPG (C/EBPγ), CREBZF, and ATF3. MEQ was found to interact with NFIL3 (E4BP4) and BACH1. Meq was also found to interact with JUND and ATF7, members of the JUN and ATF2 families that MEQ is known to interact with. Interactions were also observed for MEQ with DDIT3 and NFE2, but these proteins are not conserved between human and chickens. DDIT3 is a member of the one human bZIP family that is not found in chickens. The NFE2 family is conserved in chickens but the human NFE2 protein

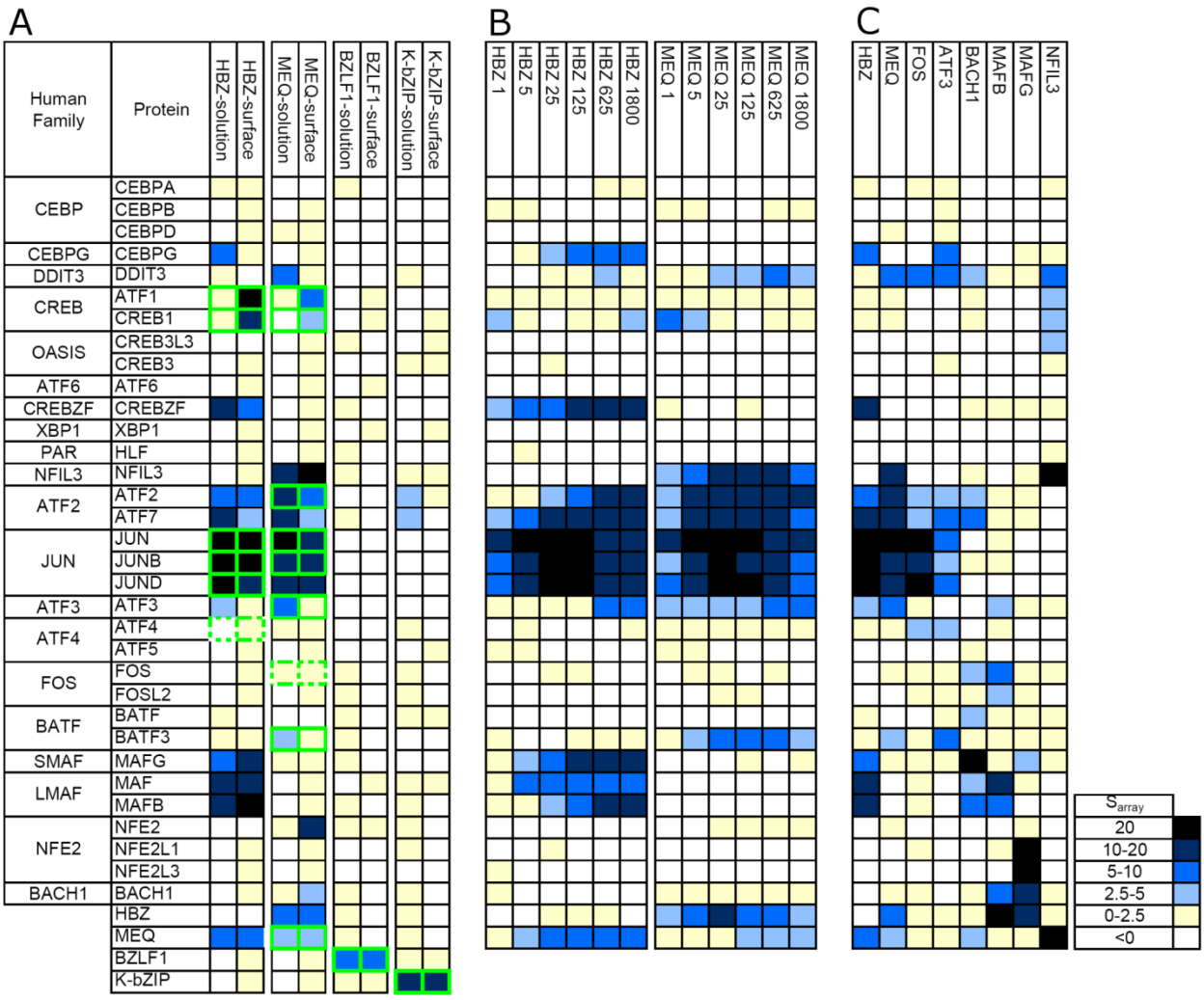



Figure 2.2. Identification of viral bZIP interactions using peptide microarrays. Interactions are displayed as a color map, as indicated by the scale at right. Family names are listed in the first column and individual proteins are listed in the second column. (A) Interactions made by 4 viral proteins, when in solution or on the surface, are shown in columns. The potential interaction partners are in rows. Each heteromeric interaction is shown twice, corresponding to the two different ways it could be measured. Observations in agreement with prior studies are boxed in green; those interactions reported to be weak in the literature are indicated with dashed boxes. Fluorescently labeled proteins were used at 160 nM in a phosphate buffer that included 1M guanidine hydrochloride (GdnHCl), except BZLF1 was used at a probe concentration of 1280 nM and K-bZIP was used at a probe concentration of 640 nM in 2.2 M GdnHCl. (B) MEQ and HBZ interactions at concentrations of 1-1800 nM. (C) Interaction profiles for 6 human proteins used as solution probes are shown for comparison to HBZ and MEQ solution profiles.

does not have a direct ortholog, and the member of the family that does have a conserved ortholog, NFE2L1, is not observed to interact with MEQ.

BZLF1 was observed to self-associate strongly, but not to interact with any human bZIP peptides (Figure 2.2A). A BZLF1 construct with the C-terminal extension (BZLF1CT) also did not interact strongly with any human proteins. This construct gave greater fluorescence signal when probed against itself than against the version containing just the leucine zipper. BZFL1CT also showed strong signal at a lower concentration than BZLF1 with just the leucine zipper. This result is consistent with previous reports documenting stabilities in solution, and further demonstrates the ability of the arrays to accurately report relative affinities (Figure A.S3) (Hicks, et al. 2003, Hicks, et al. 2001). K-bZIP interacted with itself stronger than with any other protein on the arrays. Weak interactions were observed with ATF2 and ATF7 when K-bZIP was in solution, but these interactions were not observed when K-bZIP was on the surface (Figure 2.2A).

A significant result of this experiment is that the leucine-zipper regions of MEQ and HBZ participate in multiple interactions with different human bZIPs, while BZLF1 and K-bZIP display almost exclusive homo-association (Figure 2.2A). Additionally, MEQ and HBZ each interact with a unique combination of partners (Figure 2.2C). HBZ interacts with many of the same proteins as ATF3 and FOS, but is distinguished by many other strong interactions that are not made by these proteins, including interactions with both the small and large MAF families, CEBPG, and CREBZF. MEQ also has a similar profile to human ATF3 and FOS, but additionally interacts strongly with NFIL3.

*Validation of novel interactions of HBZ and MEQ in solution*

To validate novel interactions detected on the protein microarrays we tested associations using circular dichroism (CD). Proteins consisting of the bZIP domain (basic region plus leucine zipper) were used for these experiments (see Methods, Table A.S1). For NFIL3, the chicken and human proteins are identical in this region. For the MAF proteins, the extended homology region that contains an auxiliary DNA binding domain was included (Kerppola and Curran. 1994). We first tested JUN for interaction with HBZ and MEQ. JUN has been reported to interact with both MEQ and HBZ and was also observed to interact with both on the arrays. HBZ and Jun each at 40 μM were mixed together and the CD spectrum was measured (Figure 2.3A). Spectra were also recorded for each protein in isolation. The spectra of each individual protein, as well as the mixture, had minima at 208 and 222 nm, which is characteristic of coiled coils. The observed mean residue ellipticity at 222 nm ($[\theta]_{222}$) is consistent with the expected helical content for these peptides forming coiled coils, as the leucine zipper accounts for ~50% of the sequence (Chen, et al. 1974). The mixture also had increased signal compared to the sum of the individual proteins, indicating a hetero-association. Similar results were observed for MEQ and JUN (Figure 2.3B).

We next tested HBZ against the newly identified partners ATF2, CEBPG, CREBZF, MAFG, and MAFB. Thermal melts monitored by CD were performed with each protein at a concentration of 4 μM and the mixture at 8 μM. Thermal melts were also carried out for HBZ with JUN (Figure 2.3, C-H). Over a large range of temperature all mixtures had increased signal over the sum of the spectra for the individual proteins, confirming the interactions. We then performed thermal melts of MEQ with NFIL3 and with JUN (Figure 2.3I, J). Again the mixture had increased signal over that expected for non-interacting proteins. Two pairs not observed to interact on the arrays, HBZ with NFIL3 and MEQ with MAFB, also were not observed to

interact in solution (Figure 2.3K, L). Thus, all protein pairs tested in solution agreed well with the results of the protein array assay.

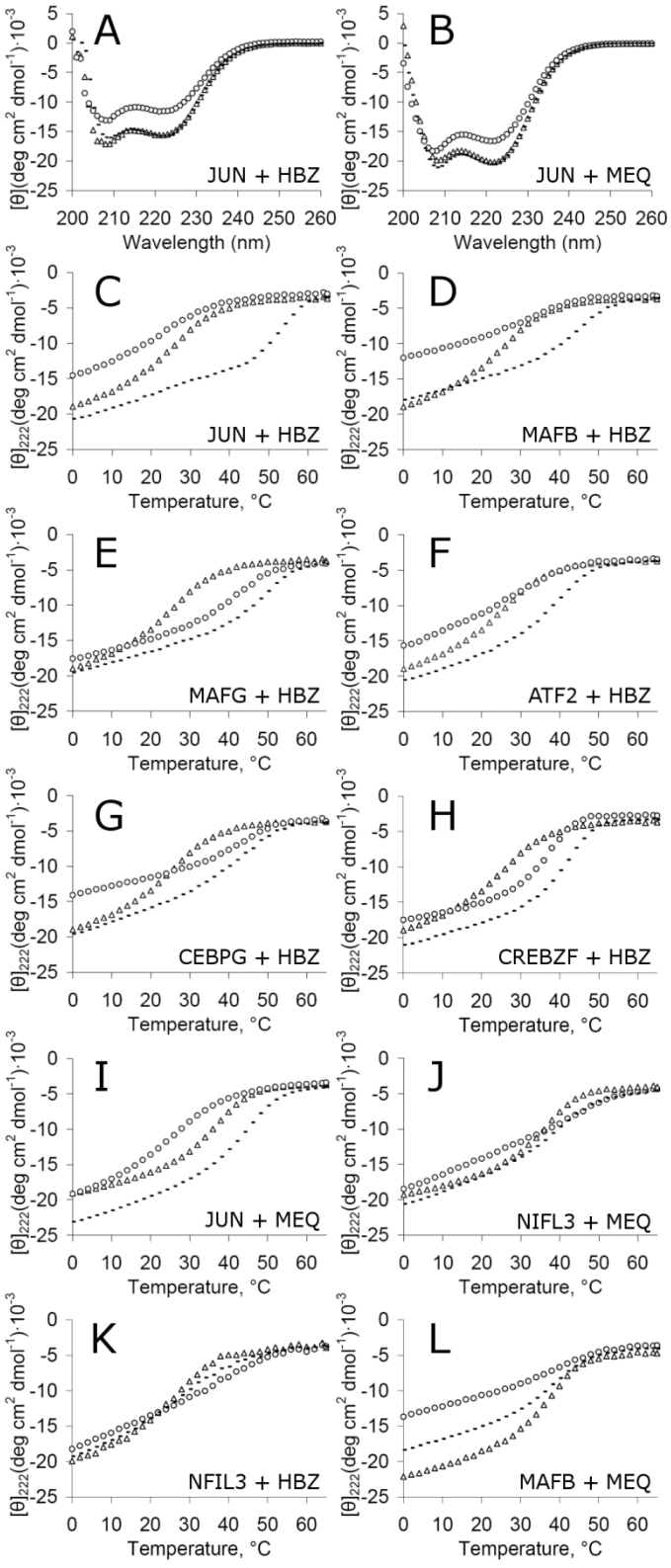


Figure 2.3. Solution measurements of novel interactions for HBZ and MEQ.
(A and B) CD spectra at 40 μM for each protein or 80 μM for each mixture at 25 °C. (A) HBZ (open triangles), JUN (open circles), mixture (dashed line). (B) MEQ (open triangles), JUN (open circles), mixture (dashed line). (C-J) Thermal melts monitored by CD at 4 μM for each protein or 8 μM for each mixture. All mixtures are shown in dashed lines. (C) HBZ (open triangles), JUN (open circles). (D) HBZ (open triangle), MAFB (open circles). (E) HBZ (open triangles), MAFG (open circles). (F) HBZ (open triangles), ATF2 (open circles). (G) HBZ (open triangles), CEBPG (open circles). (H) HBZ (open triangles), CREBZF (open circles). (I) MEQ (open triangles), JUN (open circles). (J) MEQ (open triangles), E4BP4 (open circles). (K) HBZ (open triangles), NIFL3 (open circles). (L) MEQ (open triangles), MAFB (open circles).

*Characterization of HBZ interactions with human proteins in the presence of DNA*

We tested whether HBZ could prevent its human bZIP interaction partners from binding DNA and/or whether heteromeric HBZ complexes could themselves bind DNA. MAFB and MAFG were tested in a gel-shift assay using a MARE site (Kataoka, et al. 1994). Both bound MARE DNA as homodimers at a concentration of 4 nM, but HBZ did not bind even at a 100-fold higher concentration. Surprisingly, when HBZ at increasing concentrations was mixed with a constant concentration of either MAFB or MAFG, an additional shifted band of greater mobility appeared (Figure 2.4A). To determine whether the interaction was dependant on the basic region of HBZ, a leucine-zipper-only version of HBZ, HBZLZ, was mixed with the MAF proteins and the amount of MAF protein bound to DNA was decreased, though a higher concentration of HBZLZ was required. No additional complex was formed (Figure 2.4B). Taken together, this is the first evidence, to our knowledge, that suggests HBZ can directly bind to DNA. We also tested whether ATF2 or CEBPG could bind DNA in complex with HBZ. Both ATF2 and CEBPG at 20 nM were prevented from binding MARE DNA by HBZ, but no additional band was formed (Figure 2.4C). CREBZF was not tested in complex with HBZ on DNA as CREBZF is not known to bind DNA by itself (Hogan, et al. 2006).

MAFG also binds to the AP-1 site and was tested for binding in combination with HBZ. In contrast to the MARE site, HBZ decreased the binding of MAFG to the AP-1 site without any additional shifted bands (Figure 2.4D). Both the AP-1 and the MARE sites contain the core consensus binding site TGA(C/G)TCA. The MAF proteins have an auxiliary binding domain that is responsible for binding flanking residues of the MARE site (Kerppola and Curran. 1994). At the middle of the binding site, position 0, the MARE site we used has a C and the AP-1 site contains a G. To determine if this middle position can affect binding by HBZ:MAFG complexes, the 0 position in MARE was changed to a G and the same position in AP-1 was changed to a C. The mutant sites had similar binding properties to unchanged sites, suggesting that the middle position is not the key element that influences HBZ—MAFG heteroassociation on AP-1 vs. MARE (Figure 2.4E, F). This suggests that bases flanking the core site are important for HBZ binding.

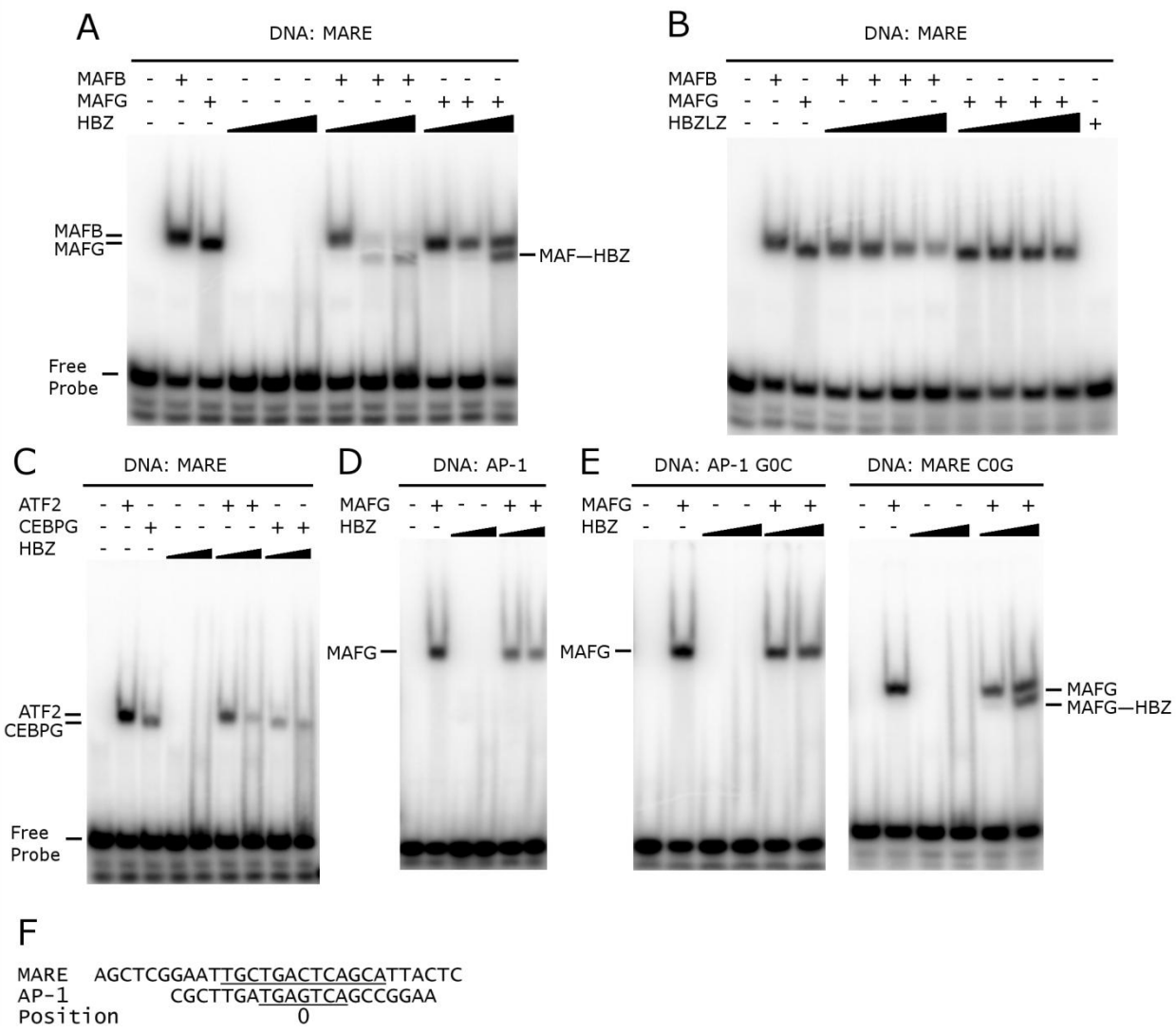



Figure 2.4. Binding of HBZ and human bZIPs to specific DNA sites assessed by gel-shifts. Homo- and hetero- complexes formed on DNA are indicated. (A) HBZ can form heterocomplexes with MAFB or MAFG on MARE DNA sites. The concentration of MAFB was 4 nM, MAFG was 4 nM, and HBZ was 4, 40 or 400 nM. (B) A leucine-zipper-only version of HBZ (HBZLZ) prevents MAFB and MAFG from binding MARE-site DNA. The concentration of MAFB was 4 nM, MAFG was 4 nM, and HBZLZ was at 4, 40, 400, or 4000 nM. HBZLZ incubated alone was at 4000 nM. (C) HBZ prevents ATF2 and CEBPG from binding MARE-site DNA. The concentrations of ATF2 and CEBPG were 20 nM, and HBZ was at 40 or 400 nM. (D) HBZ prevents MAFG from binding AP-1 DNA. The concentration of MAFG was 20 nM, and HBZ was at 40 or 400 nM. (E) AP-1-site variant TGACTCA (G0C) was not sufficient for HBZ to bind DNA with MAFG. The concentration of MAFG was 20 nM with AP-1 G0C and 4 nM with MARE C0G. The concentration of HBZ was 40 or 400 nM. (F) DNA sequences used in gel-shift assays; the consensus site is underlined and position 0 is indicated.

*Characterization of MEQ and NFIL3 binding to DNA*

To determine whether NFIL3 and MEQ could bind DNA as heterodimers we tested several known bZIP-binding sites. AP-1, also known as TRE, is a site bound by JUN and by FOS-JUN heterodimer (Rauscher Iii, et al. 1988). CAAT contains the consensus site for the CEBP family of bZIPs (Oikarinen, et al. 1987). CRE1 and CRE2 are two CRE-like sites that have been previously used in DNA binding studies with MEQ and NFIL3 and are each one change away from the consensus CRE site TGACGTCA (Levy, et al. 2003, Chen, et al. 1995). Also tested was the MDVORI site, which is derived from the origin of replication of Marek's disease virus. MEQ has previously been shown to bind this site as a homodimer (Levy, et al. 2003). In a gel-shift assay neither MEQ nor NFIL3 bound strongly to the negative control sites TFIID or NF-κB at 80 nM. Only MEQ bound strongly to the AP-1 site and only NFIL3 bound strongly to the CAAT site. Both MEQ and NFIL3 bound the CRE1 and CRE2 sites, though NFIL3 bound more strongly. For both of these sites there appeared to be some heterodimer formation, but the predominant species was the NFIL3 homodimer. Interestingly, NFIL3 bound to the MDVORI site, but weaker than MEQ did. The mixture on the MDVORI site was composed of primarily MEQ homodimers (Figure 2.5A).

We also wanted to know if MEQ and NFIL3 could bind a DNA site predominantly as a heterodimer. Previously it has been shown that heterodimer sites can be constructed by taking consensus half-sites for each of two interacting bZIPs (Vinson, et al. 1993). A DNA site was constructed that contains the consensus half-sites for MEQ (ACAC) and NFIL3 (GTAA), referred to in Figure 2.5C and below as MEQ/NFIL3 (Levy, et al. 2003, Cowell, et al. 1992). This site has only two changes from the MDVORI site, at positions +1 and +3. This hybrid site was bound as a homodimer by both MEQ and NFIL3. NFIL3 bound tighter than MEQ, and when

both proteins were mixed, the predominant species bound to the site was the heterodimer, with mobility between the two homodimers (Figure 2.5A).

The MDVORI probe used in Figure 2.5 was 30 base pairs long, and to further probe the nature of a MEQ—NFIL3 interaction we tested whether NFIL3 bound at a similar location as MEQ. Competition gel-shift experiments were performed to test this. MEQ and NFIL3 were individually incubated with radiolabeled MDVORI site and with cold competitor DNA encoding either the MDVORI site or a variant of it. Single-base substitutions were made at 10 consecutive positions in the site. Additionally, 2 double substitutions and a triple-mutant site were constructed. Changes that affected MEQ binding by at least 2-fold were localized toward the 5' half of the MDVORI site (Figure 2.5B). Substitutions that weakened MEQ binding included -4C, -3A, and -1A. The change of -2T strengthened MEQ binding. The double mutant of -3A:-1A decreased binding more than either individual substitution. NFIL3 binding was decreased by the changes -1A, +1C, +2G, and +4C. The substitutions -3A and +3A increased binding of NFIL3. Combining -2T and +1C decreased binding further. The changes of -1A and +2G decreased binding individually, but when both were together in combination with -3A, the triple mutant had no decrease in binding (Figure 2.5B). These 13 altered sites were also tested for direct binding to MEQ and NFIL3 and the results were consistent with the competition binding experiments (Figure A.S4). Additionally, significant heterodimer formation by MEQ and NFIL3 was observed on the +3A site, consistent with results using the hybrid site in Figure 2.5A (Figure A.S4). Several things are apparent from this experiment. First, mutations on the 5' side of the site affect binding to MEQ, while those toward the 3' end, and at the last two positions of the left half-site, affect binding to NFIL3. Second, most DNA base changes have differential effects on

the binding of MEQ and NFIL3, demonstrating they indeed have different binding specificities. Third, MEQ and NFIL3 bind at a similar location on the MDVORI DNA site.

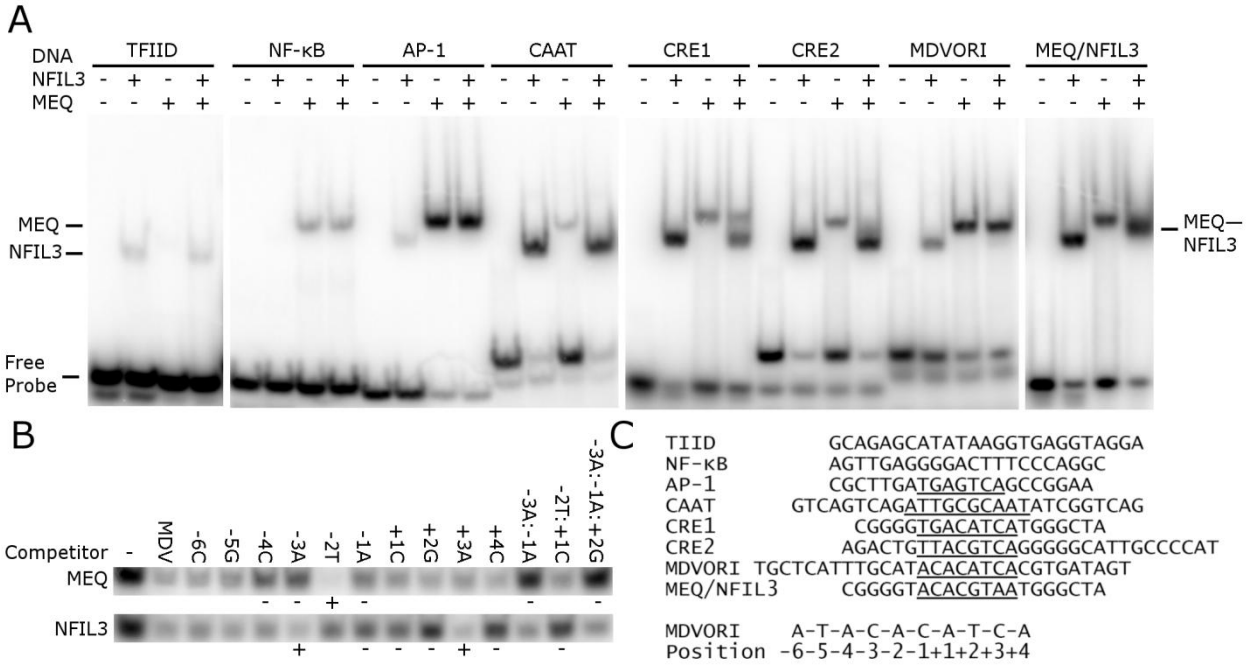



Figure 2.5. MEQ and NFIL3 interact and have different but overlapping DNA-binding specificities. (A) Gel-shift experiments with MEQ and NFIL3. The concentration of MEQ and NFIL3 was 80 nM each or 160 nM total protein for mixtures. Each homodimer and heterodimer is indicated. (B) Competition gel-shift demonstrates that MEQ and NFIL3 bind to similar regions of an MDV probe, but have differing specificities. Each protein was at 80 nM incubated with 0.7 nM radiolabeled MDVORI DNA. Above each lane is listed the mutation or mutations made in cold competitor DNA (400 nM). Three individual experiments were quantified, and those positions that gave ≥ 2-fold changes are indicated (+/- indicate increase/decrease in binding). (C**)** DNA sequences used in gel-shift assays; the consensus site is underlined. The positions of the MDV site are numbered.

*Generation of a specific inhibitor of MEQ dimerization*

Specific inhibitors of bZIP interactions could provide valuable tools for elucidating function and could potentially serve to validate transcription factors as therapeutic targets. To generate a specific peptide to bind MEQ, that could potentially act as a dominant-negative inhibitor, we used a recently-published computational design method called CLASSY (Grigoryan, et al. 2009b). CLASSY was used to automatically design a peptide sequence

predicted to bind MEQ but to have minimal binding to any human bZIP or to itself. The designed 42-residue anti-MEQ peptide was tested for its ability to bind MEQ using bZIP protein arrays including anti-MEQ, MEQ, and a panel of 32 human bZIPs (these included all the human peptides tested previously except DDIT3, which is specific to humans). Anti-MEQ bound MEQ stronger than any human protein (Figure 2.6A). The other proteins that anti-MEQ bound strongest are the ATF2 family proteins ATF2 and ATF7, followed by JUN and BATF3. Even at the highest concentration tested, 2000 nM, anti-MEQ bound MEQ and the ATF2 family proteins preferentially to other bZIPs. Anti-MEQ was not observed to interact with itself on the arrays. These results demonstrate that anti-MEQ is specific for binding to MEQ.

To compare the stability of the anti-MEQ—MEQ complex with that of other MEQ interactions, we probed MEQ against an array including MEQ, anti-MEQ, and the panel of 32 human peptides. MEQ interacted with anti-MEQ as strongly as it interacted with JUN, which was MEQ's strongest interaction partner observed on the array (Figure 2.6A). The three strongest off-target interactions, ATF2, JUN, and BATF3, were also tested in solution against anti-MEQ. ATF2 bound anti-MEQ strongly compared to its strongest interactions on the array. In contrast, JUN and BATF had much weaker interactions with anti-MEQ. Overall these results suggest that anti-MEQ has high affinity for, and is specific for, interaction with MEQ.

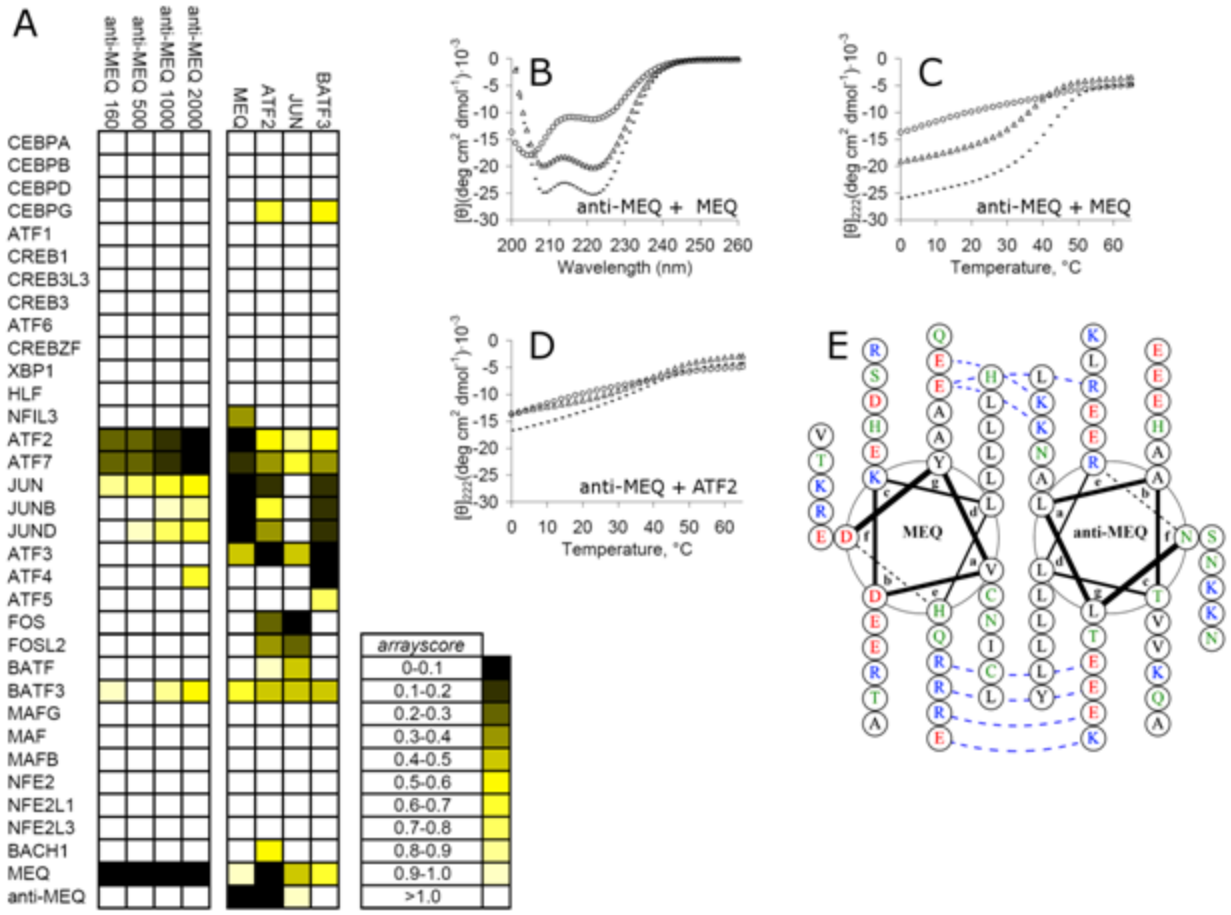



Figure 2.6. Anti-MEQ binds MEQ with high affinity and specificity. (A) Designed peptide anti-MEQ characterized using coiled-coil arrays. Color map of *arrayscore* is shown, with the colors defined in the scale. Left, anti-MEQ at different concentrations (nM) in solution is listed in columns, with proteins printed on the surface in rows. Right, MEQ and 3 human bZIPs tested against anti-MEQ and other proteins printed on the surface. (B) CD spectra at 40 µM for each protein or 80 µM for the mixture taken at 25 °C. MEQ (open triangles), anti-MEQ (open circles), and the mixture (dashed line). (C and D) Thermal melts monitored by CD at 4 µM for each protein or 8 µM for the mixture. (C) MEQ (open triangles), anti-MEQ (open circles), and the mixture (dashed line). (D)ATF2 (open triangles), anti-MEQ (open circles), and the mixture (dashed line). (E) Helical wheel diagram predicted for the interaction of MEQ with anti-MEQ. Interaction is depicted as a parallel dimer where **d-d**, **a-a**, **e-g',** and **a-g'** interactions in each heptad potentially contribute to both stability and specificity. Hydrophobic residues are in black, charged residues are in red/blue, and polar residues are in green. Potential attractive electrostatic interactions are shown in dashed blue lines. Diagram created using DrawCoil 1.0. (http://www.gevorggrigoryan.com/drawcoil/).

The interaction of MEQ and anti-MEQ in solution was studied using a tag-less version of anti-MEQ (see Methods). CD spectra showed that anti-MEQ is mostly unfolded at 40 µM, and when combined with 40 µM MEQ the mixture has more signal than either MEQ or anti-MEQ alone, indicating an interaction. The mixture also has a spectrum characteristic of a coiled coil (Figure 2.6B). Thermal melts of anti-MEQ mixed with both MEQ and ATF2 were performed at 4 µM of each protein and 8 µM total protein for the mixture (Figure 2.6C, D). The temperature of half denaturation ($T_m$) for anti-MEQ was 12.8 °C, for MEQ was 35.2 °C, and for ATF2 was 36.7 °C. The $T_m$ for anti-MEQ in complex with MEQ was 40.5 °C, and thus the hetero-association is more stable than either the homo-association of MEQ or anti-MEQ. The $T_m$ of ATF2 in complex with MEQ was 31.8 °C. JUN, a strong interaction partner for MEQ, has a $T_m$ in complex with MEQ of 41.3 °C. These results are consistent with the array data. A helical wheel diagram depicting the predicted interaction geometry of MEQ and anti-MEQ is shown in Figure 2.6E. The design has leucine residues at 5 consecutive **d**-positions, imparting stability to the complex (Moitra, et al. 1997). It also introduces a complementary asparagine residue to interact with an asparagine at an **a**-position in MEQ; this interaction is known to favor parallel dimer formation (Harbury, et al. 1993). The **e**- and **g**-positions of anti-MEQ are complementary to those in MEQ, and two rather uncommon cysteine residues at **a** positions are predicted to lie across from designed alanine and lysine residues. Lysine at core **a** positions also favors dimer formation (Campbell, et al. 2002). Two lysines at **a** positions are complementary to glutamate residues at **g** position on the opposite helix. These **a-g'** interactions have previously been shown to make important contributions to specificity (Grigoryan, et al. 2009b, Reinke, et al. 2010).

To test whether anti-MEQ could prevent MEQ from binding DNA, 20 nM MEQ was incubated with increasing amounts of anti-MEQ and then radiolabeled MDVORI DNA was

added to the reactions. Anti-MEQ prevented MEQ binding of DNA, with an $IC_{50}$ of less than 500 nM (Figure 2.7A). Binding of MEQ to AP-1 DNA was also inhibited by anti-MEQ (data not shown). The experiment was repeated with 20 nM JUN and the AP-1 DNA site. No decrease in JUN binding was observed even at 12.5 µM anti-MEQ (Figure 2.7B). The strongest off-target interaction for anti-MEQ, ATF2, was also tested. At 20 nM ATF2 no decrease in binding was observed when incubated with anti-MEQ (Figure 2.7C). At 4 nM ATF2, anti-MEQ decreased ATF2 binding, but at higher concentrations than required for preventing MEQ binding (Figure 2.7D). These results show that anti-MEQ can prevent MEQ from binding DNA at a lower concentration than it inhibits its strongest off-target interaction.

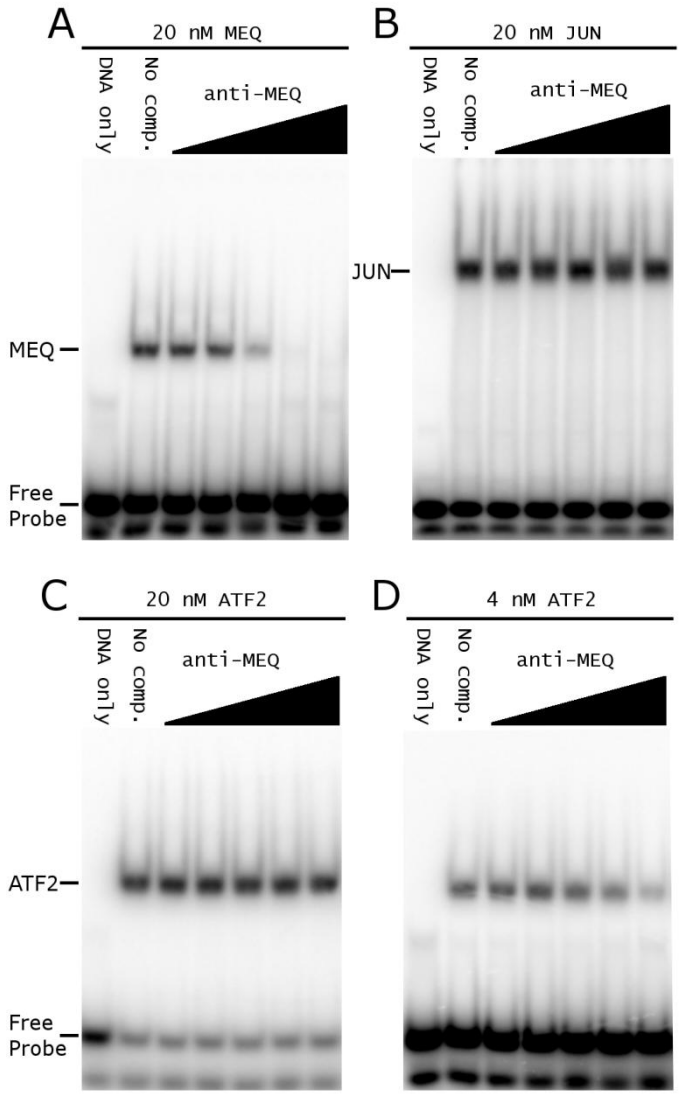



Figure 2.7. Anti-MEQ prevents MEQ from binding DNA. Competition gel-shifts with a constant amount of the indicated protein bound to DNA were titrated with increasing amounts of anti-MEQ. Concentrations of competitor peptide were 0.02, 0.1, 0.5, 2.5, and 12.5 µM. Labeled DNA was present at 0.7nM. (A) 20 nM MEQ with MDVORI DNA. (B) 20 nM JUN with AP-1 DNA. (C) 20 nM ATF2 with CRE2 DNA. (D) 4 nM ATF2 with CRE2 DNA.

## DISCUSSION

The bZIP transcription factors function by forming homodimers or heterodimers with other bZIP proteins. In this context, viruses use bZIP proteins in a number of distinct ways that are illuminated by the systematic study we present here. Several viral bZIPs that have high sequence identity to host homologues maintain the interaction patterns of the host bZIPs. Examples include v-FOS, v-JUN, and v-MAF (Bos, et al. 1989, Kataoka, et al. 1993, van Straaten, et al. 1983b). The viral bZIPs HBZ and MEQ are not closely related to any other bZIPs, and they have distinct interaction profiles compared to the human bZIPs. K-bZIP and BZLF1 are also not highly conserved, but these two viral proteins primarily self-associate.

For MEQ and HBZ, we have uncovered new interactions that suggest possible mechanisms of action of these proteins. The mechanism of HBZ protein function has been somewhat of a mystery. The non-canonical basic region of this protein argues against direct DNA binding, yet HBZ was shown previously to have a strong activation domain, suggesting that it might function to regulate transcription when complexed with human bZIP proteins and DNA (Nair. 2005, Kuhlmann, et al. 2007). Additionally, HBZ has been shown to stimulate the transcriptional activity of JUND. However, data so far have not supported a direct interaction of HBZ itself with DNA. Here we present the first evidence that HBZ can directly bind DNA. HBZ can bind a MARE DNA site with MAFB or MAFG. This binding of DNA is specific and is dependent both on the HBZ basic region and on DNA that flanks the central binding site. While most bZIPs bind a 4-5 base pair half site, different flanking regions around an AP-1 site have previously been shown to have different affinities for binding to JUN/FOS heterodimers (Ryseck and Bravo. 1991). The HBZ ternary complex that we observed on DNA with either MAFB or MAFG was weaker than MAF—DNA complexes. Also the HBZ—MAF complexes occurred at

higher concentrations of HBZ relative to MAF protein. Because the sequence of the HBZ basic region is unique, it may have a distinct DNA-binding specificity, and it remains a possibility that higher affinity sites for HBZ in association with MAF proteins exist.

The MAF proteins belong to two classes: the four large MAF proteins, which contain a transcriptional activation domain, and the three small MAF proteins that don't. The small MAFs interact with the NFE2 and BACH families of bZIPs and play a major role in the response to oxidative stress (Katsuoka, et al. 2005). The large MAF proteins are similar to the JUN proteins in that both are involved in cell growth and proliferation, both are proto-oncogenes and can cause cellular transformation, and both have retroviral homologues (Vogt. 2001, Pouponnot, et al. 2005). They also have been reported to share similar downstream targets in inducing cellular transformation (Kataoka, et al. 2001).

Other proteins we confirmed to interact with HBZ *in vitro* are CEBPG, ATF2, and ZF. The CEBP family of bZIPs is involved in cell growth and differentiation. CEBPG forms heterodimers with CEBP family proteins to repress their transcriptional activity (Parkin, et al. 2002). ATF2 has been shown to stimulate JUN-mediated cellular transformation (Huguier, et al. 1998). CREBZF was identified as having a role in herpes simplex virus gene expression (Lu and Misra. 2000). CREBZF also interacts with ATF4, a known partner of HBZ (Gaudray, et al. 2002, Hogan, et al. 2006). With ATF2 and CEBPG, it remains to be seen whether HBZ heterodimers can bind DNA sites, or if HBZ functions primarily by preventing binding of these partners to target sites. Neither HBZ nor CREBZF have been shown to bind DNA as homodimers, suggesting that they have intrinsically weaker affinities for DNA and together would not be likely to bind DNA as a heterodimer. These newly reported partners, along with other known partners, suggest that HBZ has the potential to impact several different transcriptional pathways.

It was recently shown that MEQ homodimers alone are not sufficient to induce transformation, suggesting that heterodimer formation with other bZIPs is necessary (Suchodolski, et al. 2009). JUN has been shown to be an important MEQ partner, required for cellular transformation mediated by MEQ, but it is likely that other interaction partners are also functionally important. Here we identified a previously unreported bZIP partner, NFIL3, and showed that it can form heterodimers with MEQ on DNA. NFIL3 was first identified as a transcriptional repressor that bound the adenovirus E4 promoter, and was later shown to have an activating role associated with anti-apoptotic activity. NFIL3 is also involved in regulating circadian rhythms (Cowell, et al. 1992, Ikushima, et al. 1997, Doi, et al. 2001). NFIL3 was further shown in the hepatitis B virus to both repress viral gene expression as well as viral replication (Lai and Ting. 1999). In this context, it is interesting that NFIL3 can bind to the MDVORI site as a homodimer. The MDVORI site from the origin of replication of MDV is also situated between a bidirectional promoter that MEQ has been shown to repress as a homodimer (Levy, et al. 2003, Qian, et al. 1996). There may exist functionally significant sites that MEQ and NFIL3 can bind as a heterodimer with greater affinity than either homodimer. It will be important to determine what role NFIL3 has on the MDV life cycle, both alone and in combination with MEQ.

Other novel interactions with bZIP families detected on the arrays but not tested further are HBZ with ATF3, and MEQ with BACH1. Based on the intensity of the fluorescence signal, these interactions are likely to be weaker than the other interactions tested, but they may be significant. Also, numerous interactions that we did not assay are highly likely to occur involving paralogs of the proteins tested here. Although these interactions need to be confirmed experimentally, most bZIP paralogs are highly similar to each other and have been shown to

share similar interaction profiles (Newman and Keating. 2003, Grigoryan, et al. 2009b, Amoutzias, et al. 2007).

An interesting result is that the leucine-zipper regions of BZLF1 and K-bZIP preferentially self-associate. Both proteins are reported to interact with CEBPA, but this pairing was not observed in our array experiments. Basic residues were required for this interaction in the case of BZLF1, and both proteins were observed to interact with CEBPA as higher-order multimers and not as heterodimers (Wu, et al. 2004, Wu, et al. 2003). Our observation that the leucine zippers are not sufficient for these interactions is consistent with those studies. It is not surprising, given the unique structure of BZLF1, that it does not form canonical interactions with human BZIPs.

MDV is the first oncogenic virus for which a vaccine was made available to control the disease, but increasing viral resistance is becoming a real concern to the poultry industry. Further, MDV has proven valuable as a model oncogenic virus, and deletion and knockdown experiments have demonstrated the necessity of MEQ for oncogenic transformation (Levy, et al. 2005). It has also been reported that a virus encoding a MEQ protein that cannot form homodimers or heterodimers has a complete loss of oncogenicity, suggesting that the function of MEQ could be inhibited by preventing MEQ from interacting with bZIPs (Brown, et al. 2009). We showed that a computationally designed anti-MEQ peptide can prevent MEQ from binding DNA in a specific manner, indicating that anti-MEQ could be a useful reagent for studying the role of MEQ on the oncogenic properties of MDV. If necessary to achieve higher affinity, the anti-MEQ peptide could potentially be further stabilized through the addition of an acidic peptide extension that interacts with the basic region, as has been demonstrated for numerous other coiled coils by the Vinson laboratory (Acharya, et al. 2006b).

In summary, we have shown that coiled-coil arrays are a powerful method for broadly surveying the interaction properties of viral bZIP dimerization domains. Comprehensive testing for *in vitro* interactions with all human bZIP families is an important step in exploring the functions of these proteins. Further, we have validated that several newly discovered viral-host complexes can bind to DNA, suggesting a mechanism by which viruses hijack cellular transcriptional control. Determining which of the bZIPs that can associate *in vitro* also interact with functional consequences *in vivo* will be an important next step.

**ACKNOWLEDGEMENTS**

We thank the MIT BioMicro center for arraying instrumentation, J. Fisher for technical assistance in cloning and protein purification and the Laub lab for use of space and equipment. We thank members of the Keating laboratory for comments on the manuscript.

**ABBREVIATIONS**

ATP, adenosine triphosphate; BLAST, basic local alignment search tool; BSA, bovine serum albumin; bZIP, basic-region leucine-zipper transcription factor; CD, circular dichroism; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GdnHCl, guanidine hydrochloride; HPLC, high performance liquid chromatography; IPTG, isopropyl β-D-1-thiogalactopyranoside; MBP, maltose-binding protein; Ni-NTA, nickel-nitrilotriacetic acid; PBS, phosphate buffered saline; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PNK, polynucleotide kinase; TBE, tris/borate/EDTA; TEV, tobacco etch virus; TRIS, tris(hydroxymethyl)aminomethane.

## REFERENCES


1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006a Sep 26;45(38):11324-32.

2. Acharya A, Rishi V, Moll J, Vinson C. Experimental identification of homodimerizing B-ZIP families in homo sapiens. Journal of Structural Biology. 2006b;155(2):130-9.

3. Adya N, Zhao LJ, Huang W, Boros I, Giam CZ. Expansion of CREB's DNA recognition specificity by tax results from interaction with ala-ala-arg at positions 282-284 near the conserved DNA-binding domain of CREB. Proc Natl Acad Sci U S A. 1994 Jun 7;91(12):5642-6.

4. Amoutzias GD, Veron AS, Weiner J,3rd, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL. One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. Mol Biol Evol. 2007 Mar;24(3):827-35.

5. Basbous J, Arpin C, Gaudray G, Piechaczyk M, Devaux C, Mesnard JM. The HBZ factor of human T-cell leukemia virus type I dimerizes with transcription factors JunB and c-jun and modulates their transcriptional activity. J Biol Chem. 2003;278:43620-7.

6. Bos TJ, Rauscher FJ,3rd, Curran T, Vogt PK. The carboxy terminus of the viral jun oncoprotein is required for complex formation with the cellular fos protein. Oncogene. 1989 Feb;4(2):123-6.

7. Brown AC, Smith LP, Kgosana L, Baigent SJ, Nair V, Allday MJ. Homodimerization of the meq viral oncoprotein is necessary for induction of T-cell lymphoma by marek's disease virus. J Virol. 2009 Nov;83(21):11142-51.

8. Campbell KM, Sholders AJ, Lumb KJ. Contribution of buried lysine residues to the oligomerization specificity and stability of the fos coiled coil. Biochemistry. 2002 Apr 16;41(15):4866-71.

9. Chang YN, Dong DL, Hayward GS, Hayward SD. The epstein-barr virus zta transactivator: A member of the bZIP family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. J Virol. 1990 Jul;64(7):3358-69.

10. Chen W, Lewis KS, Chandra G, Cogswell JP, Stinnett SW, Kadwell SH, Gray JG. Characterization of human E4BP4, a phosphorylated bZIP factor. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 1995;1264(3):388-96.

11. Chen YH, Yang JT, Chau KH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. Biochemistry. 1974 Jul 30;13(16):3350-9.

12. Countryman J, Jenson H, Seibl R, Wolf H, Miller G. Polymorphic proteins encoded within BZLF1 of defective and standard epstein-barr viruses disrupt latency. J. Virol. 1987 December 1, 1987;61(12):3672-9.

13. Cowell IG, Skinner A, Hurst HC. Transcriptional repression by a novel member of the bZIP family of transcription factors. Mol. Cell. Biol. 1992 July 1, 1992;12(7):3070-7.

14. Daury L, Busson M, Tourkine N, Casas F, Cassar-Malek I, Wrutniak-Cabello C, Castellazzi M, Cabello G. Opposing functions of ATF2 and fos-like transcription factors in c-jun-mediated myogenin expression and terminal differentiation of avian myoblasts. Oncogene. 2001 Nov 29;20(55):7998-8008.

15. Doi M, Nakajima Y, Okano T, Fukada Y. Light-induced phase-delay of the chicken pineal circadian clock is associated with the induction of cE4bp4, a potential transcriptional repressor of cPer2 gene. Proceedings of the National Academy of Sciences of the United States of America. 2001 July 3, 2001;98(14):8089-94.

16. Ellison TJ, Izumiya Y, Izumiya C, Luciw PA, Kung H. A comprehensive analysis of recruitment and transactivation potential of K-rta and K-bZIP during reactivation of kaposi's sarcoma-associated herpesvirus. Virology. 2009;387(1):76-88.

17. Gaudray G, Gachon F, Basbous J, Biard-Piechaczyk M, Devaux C, Mesnard JM. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J Virol. 2002;76:12813-22.

18. Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, Guyer M, Peck AM, Derge JG, Lipman D, Collins FS, Jang W, Sherry S, Feolo M, Misquitta L, Lee E, Rotmistrovsky K, Greenhut SF, Schaefer CF, Buetow K, Bonner TI, Haussler D, Kent J, Kiekhaus M, Furey T, Brent M, Prange C, Schreiber K, Shapiro N, Bhat NK, Hopkins RF, Hsie F, Driscoll T, Soares MB, Casavant TL, Scheetz TE, Brown-stein MJ, Usdin TB, Toshiyuki S, Carninci P, Piao Y, Dudekula DB, Ko MS, Kawakami K, Suzuki Y, Sugano S, Gruber CE, Smith MR, Simmons B, Moore T, Waterman R, Johnson SL, Ruan Y, Wei CL, Mathavan S, Gunaratne PH, Wu J, Garcia AM, Hulyk SW, Fuh E, Yuan Y, Sneed A, Kowis C, Hodgson A, Muzny DM, McPherson J, Gibbs RA, Fahey J, Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M, Madari A, Young AC, Wetherby KD, Granite SJ, Kwong PN, Brinkley CP, Pearson RL, Bouffard GG, Blakesly RW, Green ED, Dickson MC, Rodriguez AC, Grimwood J, Schmutz J, Myers RM, Butterfield YS, Griffith M, Griffith OL, Krzywinski MI, Liao N, Morin R, Palmquist D, Petrescu AS, Skalska U, Smailus DE, Stott JM, Schnerch A, Schein JE, Jones SJ, Holt RA, Baross A, Marra MA, Clifton S, Makowski KA, Bosak S, Malek J. The status, quality, and expansion of the NIH full-length cDNA project: The mammalian gene collection (MGC). Genome Res. 2004 Oct;14(10B):2121-7.

19. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009a Apr 16;458(7240):859-64.

21. Hai T, Curran T. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc Natl Acad Sci U S A. 1991 May 1;88(9):3720-4.

22. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science. 1993 Nov 26;262(5138):1401-7.

23. Hardwick JM, Bellows DS. Viral versus cellular BCL-2 proteins. Cell Death Differ. 2003 Jan;10 Suppl 1:S68-76.

24. Hicks MR, Balesaria S, Medina-Palazon C, Pandya MJ, Woolfson DN, Sinclair AJ. Biophysical analysis of natural variants of the multimerization region of epstein-barr virus lytic-switch protein BZLF1. J Virol. 2001 Jun;75(11):5381-4.

25. Hicks MR, Al-Mehairi SS, Sinclair AJ. The zipper region of epstein-barr virus bZIP transcription factor zta is necessary but not sufficient to direct DNA binding. J. Virol. 2003 July 15, 2003;77(14):8173-7.

26. Hivin P, Arpin-Andre C, Clerc I, Barbeau B, Mesnard JM. A modified version of a fos-associated cluster in HBZ affects jun transcriptional potency. Nucleic Acids Res. 2006;34:2761-72.

27. Hogan M,R., Cockram G,P., Lu R. Cooperative interaction of zhangfei and ATF4 in transactivation of the cyclic AMP response element. FEBS letters. 2006;580(1):58-62.

28. Hoover DM, Lubkowski J. DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. 2002 May 15;30(10):e43.

29. Huguier S, Baguet J, Perez S, van Dam H, Castellazzi M. Transcription factor ATF2 cooperates with v-jun to promote growth factor-independent proliferation in vitro and tumor formation in vivo. Mol Cell Biol. 1998 Dec;18(12):7020-9.

30. Ikushima S, Inukai T, Inaba T, Nimer SD, Cleveland JL, Look AT. Pivotal role for the NFIL3/E4BP4 transcription factor in interleukin 3-mediated survival of pro-Bâ€‰lymphocytes. Proceedings of the National Academy of Sciences of the United States of America. 1997 March 18, 1997;94(6):2609-14.

31. Kataoka K, Shioda S, Yoshitomo-Nakagawa K, Handa H, Nishizawa M. Maf and jun nuclear oncoproteins share downstream target genes for inducing cell transformation. J Biol Chem. 2001 Sep 28;276(39):36849-56.

32. Kataoka K, Fujiwara KT, Noda M, Nishizawa M. MafB, a new maf family transcription activator that can associate with maf and fos but not with jun. Mol. Cell. Biol. 1994 November 1, 1994;14(11):7581-91.

33. Kataoka K, Nishizawa M, Kawai S. Structure-function analysis of the maf oncogene product, a member of the b-zip protein family. J Virol. 1993 Apr;67(4):2133-41.

34. Katsuoka F, Motohashi H, Ishii T, Aburatani H, Engel JD, Yamamoto M. Genetic evidence that small maf proteins are essential for the activation of antioxidant response element-dependent genes. Mol Cell Biol. 2005 Sep;25(18):8044-51.

35. Kerppola TK, Curran T. A conserved region adjacent to the basic domain is required for recognition of an extended DNA binding site by Maf/Nrl family proteins. Oncogene. 1994 Nov;9(11):3149-58.

36. Kuhlmann A, Villaudy J, Gazzolo L, Castellazzi M, Mesnard J, Duc Dodon M. HTLV-1 HBZ cooperates with JunD to enhance transcription of the human telomerase reverse transcriptase gene (hTERT). Retrovirology. 2007;4(1):92.

37. Kutok JL, Wang F. SPECTRUM OF EPSTEIN-BARR VIRUSâ€“ASSOCIATED DISEASES. Annual Review of Pathology: Mechanisms of Disease. 2006;1(1):375-404.

38. Kvansakul M, Yang H, Fairlie WD, Czabotar PE, Fischer SF, Perugini MA, Huang DC, Colman PM. Vaccinia virus anti-apoptotic F1L is a novel bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. Cell Death Differ. 2008 Oct;15(10):1564-71.

39. Lai C, Ting L. Transcriptional repression of human hepatitis B virus genes by a bZIP family member, E4BP4. J. Virol. 1999 April 1, 1999;73(4):3197-209.

40. Lefort S, Soucy-Faulkner A, Grandvaux N, Flamand L. Binding of kaposi's sarcoma-associated herpesvirus K-bZIP to interferon-responsive factor 3 elements modulates antiviral gene expression. J. Virol. 2007 October 15, 2007;81(20):10950-60.

41. Lemasson I, Lewis MR, Polakowski N, Hivin P, Cavanagh M, Thebault S, Barbeau B, Nyborg JK, Mesnard J. Human T-cell leukemia virus type 1 (HTLV-1) bZIP protein interacts with the cellular transcription factor CREB to inhibit HTLV-1 transcription. J. Virol. 2007 February 15, 2007;81(4):1543-53.

42. Levy AM, Gilad O, Xia L, Izumiya Y, Choi J, Tsalenko A, Yakhini Z, Witter R, Lee L, Cardona CJ, Kung H. Marek's disease virus meq transforms chicken cells via the v-jun transcriptional cascade: A converging transforming pathway for avian oncoviruses. PNAS. 2005 October 11, 2005;102(41):14831-6.

43. Levy AM, Izumiya Y, Brunovskis P, Xia L, Parcells MS, Reddy SM, Lee L, Chen H, Kung H. Characterization of the chromosomal binding sites and dimerization partners of the viral oncoprotein meq in marek's disease virus-transformed T cells. J. Virol. 2003 December 1, 2003;77(23):12841-51.

44. Lin S, Robinson DR, Miller G, Kung H. Kaposi's sarcoma-associated herpesvirus encodes a bZIP protein with homology to BZLF1 of epstein-barr virus. J. Virol. 1999 March 1, 1999;73(3):1909-17.

45. Lu R, Misra V. Zhangfei: A second cellular protein interacts with herpes simplex virus accessory factor HCF in a manner similar to luman and VP16. Nucleic Acids Res. 2000 Jun 15;28(12):2446-54.

46. Matsumoto J, Ohshima T, Isono O, Shimotohno K. HTLV-1 HBZ suppresses AP-1 activity by impairing both the DNA-binding ability and the stability of c-jun protein. Oncogene. 2005;24:1001-10.

47. Mesnard JM, Barbeau B, Devaux C. HBZ, a new important player in the mystery of adult-T-cell leukemia. Blood. 2006;108:3979-82.

48. Moitra J, Szilak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. Biochemistry. 1997 Oct 14;36(41):12567-73.

49. Nair V. Evolution of marek's disease - A paradigm for incessant race between the pathogen and the host. The Veterinary Journal. 2005;170(2):175-83.

50. Newman JR, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 Jun 27;300(5628):2097-101.

51. Oikarinen J, Hatamochi A, de Crombrugghe B. Separate binding sites for nuclear factor 1 and a CCAAT DNA binding factor in the mouse alpha 2(I) collagen promoter. J. Biol. Chem. 1987 August 15, 1987;262(23):11064-70.

52. Parkin SE, Baer M, Copeland TD, Schwartz RC, Johnson PF. Regulation of CCAAT/enhancer-binding protein (C/EBP) activator proteins by heterodimerization with C/EBPgamma (Ig/EBP). J Biol Chem. 2002 Jun 28;277(26):23563-72.

53. Petosa C, Morand P, Baudin F, Moulin M, Artero J, Muller CW. Structural basis of lytic cycle activation by the epstein-barr virus ZEBRA protein. Molecular Cell. 2006;21(4):565-72.

54. Pouponnot C, Sii-Felice K, Hmitou I, Rocques N, Lecoin L, Druillennec S, Felder-Schmittbuhl MP, Eychene A. Cell context reveals a dual role for maf in oncogenesis. Oncogene. 2005;25(9):1299-310.

55. Qian Z, Brunovskis P, Lee L, Vogt PK, Kung HJ. Novel DNA binding specificities of a putative herpesvirus bZIP oncoprotein. J. Virol. 1996 October 1, 1996;70(10):7161-70.

56. Qian Z, Brunovskis P, Rauscher F,3rd, Lee L, Kung HJ. Transactivation activity of meq, a marek's disease herpesvirus bZIP protein persistently expressed in latently infected transformed T cells. J. Virol. 1995 July 1, 1995;69(7):4037-44.

57. Rauscher Iii FJ, Sambucetti LC, Curran T, Distel RJ, Spiegelman BM. Common DNA binding site for fos protein complexesand transcription factor AP-1. Cell. 1988;52(3):471-80.

58. Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J Am Chem Soc. 2010 May 5;132(17):6025-31.

59. Rossetto C, Yamboliev I, Pari GS. Kaposi's sarcoma-associated Herpesvirus/Human herpesvirus 8 K-bZIP modulates LANA mediated suppression of lytic origin-dependent DNA synthesis. J. Virol. 2009 June 24, 2009:JVI.00922-09.

60. Ryseck RP, Bravo R. c-JUN, JUN B, and JUN D differ in their binding affinities to AP-1 and CRE consensus sequences: Effect of FOS proteins. Oncogene. 1991 Apr;6(4):533-42.

61. Satou Y, Yasunaga J, Yoshida M, Matsuoka M. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. Proc Natl Acad Sci U S A. 2006;103:720-5.

62. Schepers A, Pich D, Hammerschmidt W. Activation oforiLyt,the lytic origin of DNA replication of epstein-barr virus, by BZLF1. Virology. 1996;220(2):367-76.

63. Sinclair AJ. bZIP proteins of human gammaherpesviruses. J Gen Virol. 2003 August 1, 2003;84(8):1941-9.

64. Suchodolski PF, Izumiya Y, Lupiani B, Ajithdoss DK, Gilad O, Lee LF, Kung H, Reddy SM. Homodimerization of marek's disease virus-encoded meq protein is not sufficient for transformation of lymphocytes in chickens. J. Virol. 2009 January 15, 2009;83(2):859-69.

65. Thebault S, Basbous J, Hivin P, Devaux C, Mesnard JM. HBZ interacts with JunD and stimulates its transcriptional activity. FEBS Lett. 2004;562:165-70.

66. Thomas FS. The pleiotropic effects of kaposi's sarcoma herpesvirus. The Journal of Pathology. 2006;208(2):187-98.

67. Ubeda M, Wang XZ, Zinszner H, Wu I, Habener JF, Ron D. Stress-induced binding of the transcriptional factor CHOP to a novel DNA control element. Mol. Cell. Biol. 1996 April 1, 1996;16(4):1479-89.

68. van Straaten F, Muller R, Curran T, Van Beveren C, Verma IM. Complete nucleotide sequence of a human c-onc gene: Deduced amino acid sequence of the human c-fos protein. Proc Natl Acad Sci U S A. 1983a Jun;80(11):3183-7.

70. Vinson CR, Hai T, Boyd SM. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: Prediction and rational design. Genes Dev. 1993 Jun;7(6):1047-58.

71. Vogt PK. Jun, the oncoprotein. 1. 2001;20:2365-77.

72. Witt AE, Hines LM, Collins NL, Hu Y, Gunawardane RN, Moreira D, Raphael J, Jepson D, Koundinya M, Rolfs A, Taron B, Isakoff SJ, Brugge JS, LaBaer J. Functional proteomics

approach to investigate the biological activities of cDNAs implicated in breast cancer. J Proteome Res. 2006 Mar;5(3):599-610.

73. Wu FY, Wang SE, Chen H, Wang L, Hayward SD, Hayward GS. CCAAT/Enhancer binding protein {alpha} binds to the epstein-barr virus (EBV) ZTA protein through oligomeric interactions and contributes to cooperative transcriptional activation of the ZTA promoter through direct binding to the ZII and ZIIIB motifs during induction of the EBV lytic cycle. J. Virol. 2004 May 1, 2004;78(9):4847-65.

74. Wu FY, Wang SE, Tang Q, Fujimuro M, Chiou C, Zheng Q, Chen H, Hayward SD, Lane MD, Hayward GS. Cell cycle arrest by kaposi's sarcoma-associated herpesvirus replication-associated protein is mediated at both the transcriptional and posttranslational levels by binding to CCAAT/Enhancer-binding protein {alpha} and p21CIP-1. J. Virol. 2003 August 15, 2003;77(16):8893-914.

# CHAPTER 3

# Design of protein-interaction specificity gives selective bZIP-binding peptides

**Supporting information:**

This paper included supplemental tables and figures which are in appendix B.

**Reproduced with permission from:**

Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009 Apr 16;458(7240):859-64.

**Author Contributions** GG, AWR and AEK conceived the project. GG developed, implemented and applied the CLASSY formalism and carried out all computational analyses. AWR designed and performed all experiments. All authors analyzed data and guided the research plan. GG and AEK wrote the paper, in consultation with AWR.

**ABSTRACT**

Interaction specificity is a required feature of biological networks and a necessary characteristic of protein or small-molecule reagents and therapeutics. The ability to alter or inhibit protein interactions selectively would advance basic and applied molecular science. Assessing or modelling interaction specificity requires treating multiple competing complexes, which presents computational and experimental challenges. Here we present a computational framework for designing protein interaction specificity and use it to identify specific peptide partners for human bZIP transcription factors. Protein microarrays were used to characterize designed, synthetic ligands for all but one of 20 bZIP families. The bZIP proteins share strong sequence and structural similarities and thus are challenging targets to bind specifically. Yet many of the designs, including examples that bind the oncoproteins cJun, cFos and cMaf, were selective for their targets over all 19 other families. Collectively, the designs exhibit a wide range of novel interaction profiles, demonstrating that human bZIPs have only sparsely sampled the possible interaction space accessible to them. Our computational method provides a way to systematically analyze tradeoffs between stability and specificity and is suitable for use with many types of structure-scoring functions; thus it may prove broadly useful as a tool for protein design.

## INTRODUCTION

Designing peptides, proteins, or small molecules that bind to native protein targets is a promising route to new reagents and therapies. Yet dealing with the interaction specificity problem – i.e. achieving designs that are selective for their intended targets in preference to related alternatives – is difficult. Designing or assessing protein interaction specificity in a comprehensive manner is impeded by the challenges and costs inherent in modelling or measuring many competing complexes. Recent large-scale experiments that have characterized interaction specificity for a handful of protein families and/or domains represent significant progress in this area (Jones, et al. 2006, Stiffler, et al. 2007, Wiedemann, et al. 2004, Newman and Keating. 2003, Landgraf, et al. 2004, Skerker, et al. 2005). In particular, assays that provide a way to profile the interactions of a protein with many candidate partners offer an opportunity to explore how specificity can be introduced into proteins rationally, by design.

Computational design has led to remarkable advances in protein engineering over the past decade, including the design of protein-protein interactions (Havranek and Harbury. 2003, Kortemme, et al. 2004, Ali, et al. 2005, van der Sloot, et al. 2006, van der Sloot, et al. 2006, Reina, et al. 2002, Shifman and Mayo. 2003, Fu, et al. 2007, Bolon, et al. 2005). Introducing considerations of specificity into protein-design calculations raises interesting theoretical challenges that have been addressed in a few prior studies (Havranek and Harbury. 2003, Kangas and Tidor. 2000, Deutsch and Kurosky. 1996) and/or treated on a case-by-case basis in several applications (Havranek and Harbury. 2003, Kortemme, et al. 2004, Ali, et al. 2005, van der Sloot, et al. 2006, Bolon, et al. 2005). Most often, however, specificity is simply ignored in computational

protein design. Several proteins or peptides that were optimized solely for binding to a native target were shown *a posteriori* to be specific for their intended interaction partner over a few related alternatives (Reina, et al. 2002, Shifman and Mayo. 2003, Fu, et al. 2007, Yin, et al. 2007). However, focusing only on the stability of the desired complex led to a lack of specificity, both in computational design and experimental selections, in other examples (Bolon, et al. 2005, Deutsch and Kurosky. 1996, Mason, et al. 2006). Strategies that can simultaneously consider affinity and multi-state specificity in the design process are therefore highly desirable (Havranek and Harbury. 2003).

The basic-region leucine-zipper (bZIP) transcription factors provide an exciting but highly challenging opportunity to test strategies for interaction specificity design. The bZIPs homo- and/or heterodimerize by forming a parallel coiled coil (a "leucine zipper") and bind DNA using a region rich in basic amino acids (Vinson, et al. 2006). Approximately 53 human bZIP proteins that make up 20 families participate in a wide range of important biological processes and pose attractive targets for selective inhibition. Interest in inhibiting bZIPs dates to 1995, when Vinson and co-workers showed that heterodimers containing one bZIP subunit and one subunit with an acidic region replacing the basic region (A-ZIPs) are inactive. A-ZIPs have proven very useful for applications both *in vitro* and *in vivo* (Gerdes, et al. 2006, Krylov, et al. 1995). However, these inhibitors mimic the interaction preferences of the proteins from which they are derived and typically associate with multiple bZIP families. Extensive sequence similarity among the leucine-zipper domains hampers efforts to make specific peptides that could provide more selective A-ZIPs or other inhibitors. For example, strong undesirable off-target interactions were observed when experimentally selecting

synthetic partners for the cFos and cJun bZIP coiled coils out of peptide libraries (Mason, et al. 2006).

The bZIPs are also attractive design targets because experiments have probed sequence features that influence both structural and interaction specificity (Vinson, et al. 2006, Acharya, et al. 2006, Krylov, et al. 1998, Lupas and Gruber. 2005). Building upon these insights and taking advantage of large experimental data sets, computational models that provide useful predictions of bZIP interaction preferences have been developed (Newman and Keating. 2003, Mason, et al. 2006, Fong, et al. 2004, Grigoryan and Keating. 2006). These prior studies afford a relatively mature understanding of bZIP partnering and provide the potential for specificity design.

## RESULTS

### *Computational design of specificity*

We have developed a strategy for addressing specificity in protein-design calculations that rests on the trade-off between maximizing affinity and introducing specificity. The stability/specificity trade-off has been discussed previously(Havranek and Harbury. 2003, Bolon, et al. 2005, Kangas and Tidor. 2000, Deutsch and Kurosky. 1996), and has motivated the successful design of heterospecific coiled-coil pairs (Havranek and Harbury. 2003). For our work, we note that a protein designed to bind optimally to a native target may also bind strongly to one or more undesired competitors, indicating that the difference in energy between forming undesired complexes and the design•target complex is not sufficiently large. New designs can be sought that increase this gap and are thus more selective for the target, but these will necessarily have reduced target affinities relative to the design that is optimal for target

binding. The computational method presented here formalizes this trade-off by identifying sequences that minimize the stability sacrifice required to achieve increasing energy gaps from competing complexes. Such sequences posses the important property that they cannot simultaneously be improved both in predicted affinity and specificity.

Our framework, CLASSY (Cluster expansion and Linear programming-based Analysis of Specificity and Stability), makes use of two computational techniques to implement the above idea. The first is integer linear programming (ILP), an optimization method that has been applied to the energy-minimization problem in protein design (Kingsford, et al. 2005). The second is cluster expansion (CE), which we use to convert a structure-based interaction model into a sequence-based scoring function that is very fast to evaluate (Grigoryan, et al. 2006, Zhou, et al. 2005). Importantly, CE allows us to apply ILP at the sequence level, rather than at the structure level. This makes it possible to impose constraints on the energies of design•undesired partner interactions during optimization of the design•target energy, which is the keystone of the CLASSY approach. The power of CE and ILP mean that arbitrary numbers of desired and undesired states and relationships between them can be included in CLASSY designs. Thus, CLASSY can deal with problems beyond the scope of traditional design methods, making it an appropriate approach for designing specific anti-bZIP peptides.

As one example of how CLASSY can be used, we implemented a procedure called a *specificity sweep* to identify sequences of optimal stability that satisfy increasing requirements on specificity. For this purpose, the quantity $\Delta$ was defined as the energy gap between the lowest-energy undesired state and the desired target state

(Figure 3.1A). A specificity sweep begins by using ILP to find the sequence with the highest binding affinity for the target, ignoring specificity. An initial value for the quantity $\Delta$ is then computed by predicting the energies of all possible complexes involving this design. The ILP optimization is repeated, this time designing a protein that optimizes binding with the target subject to the constraint that all undesired states have energy gaps to the designed state that are larger than $\Delta$ plus a small increment. This is repeated, gradually increasing the value of $\Delta$, until it is no longer possible to find design sequences that satisfy the constraints. Although CLASSY can be run with any value assigned to $\Delta$, one advantage of the specificity sweep exploring a broad range of $\Delta$ values is that no assumption of how much stability or specificity is "enough" need be made prior to the calculation.

Candidate designs from a specificity sweep list may be selected for testing by a user, after considering predicted stability:specificity tradeoffs and the sequence changes that bring these about. Other considerations may be included, as CLASSY provides the ability to restrict arbitrary linear functions of sequence. In our application, a bias for the bZIP coiled-coil fold was imposed by constraining designs to be leucine-zipper like according to a position-specific scoring matrix (PSSM). Similar constraints could also be used, for example, to place requirements on predicted solubility. Such considerations, which are often included in designs in an ad hoc manner or by employing manual post-evaluation and filtering, can be naturally incorporated into the CLASSY procedure.

*Design of anti-bZIP peptides*

We applied CLASSY to design partners for nearly all human bZIPs and used our computational results to assess the difficulty of the bZIP interaction specificity design problem. We sought anti-bZIP designs predicted to bind their targets and yet interact minimally with themselves and with members of the 19 non-target bZIP families. Because of the extremely high sequence similarity within families, we did not require that the designs discriminate between siblings in the target family. The desired design•target heteromeric complex, as well as undesired design•design and design•off-target complexes, were modelled as coiled-coil dimers on a fixed-backbone template and evaluated using energy functions similar to that of reference (Grigoryan and Keating. 2006), which was shown previously to give good performance predicting native bZIP interaction preferences (Grigoryan and Keating. 2006) (also see appendix B).

Specificity sweeps were computed for the 46 bZIPs in reference (Newman and Keating. 2003). These calculations predicted that specificity will arise only rarely among bZIP partners optimized for stability alone. Such designs are almost all predicted to form strong homodimers, regardless of the family they are targeted against (Figure B.S2). Negative design is also required to disfavour complexes with undesired bZIP competitors. Approximately 65% of 46 designs optimized for affinity alone were judged to face significant competition from non-target families; this can be addressed in CLASSY by sacrificing stability, as shown in Figure B.S2. We carried out additional computational analyses to estimate how candidate bZIP partners are distributed in stability-specificity space (Figure B.S12). Even when the design•design homodimer is the only undesired state, the vast majority of sequence space is predicted to be non-

specific. Thus, addressing specificity is critical, but the drastic reduction this imposes on acceptable sequences makes the design problem challenging.

*Testing of anti-bZIP designs*

We next tested 48 peptides designed to bind representative targets from all 20 bZIP families, using a protein microarray assay that has been validated for measuring interaction preferences for bZIPs (Newman and Keating. 2003). Sequences to be tested were selected from the specificity sweeps by hand, considering the magnitude of $\Delta$, the amount of stability lost relative to the most stable design, and sequence features such as excessive loss of hydrophobic interactions in the core (see Figure 3.1C for the example of anti-SMAF; Table B.S1 provides detailed descriptions of the origin of each design). In a few of the cases where we designed more than one peptide against a given target, experimental results for initial designs were incorporated to guide the CLASSY design procedure. For example, anti-ZF was designed using a modified specificity sweep that up-weighted the influence of XBP-1 in determining $\Delta$, after this protein was experimentally determined to be a problematic competitor. The ability to easily incorporate information about known competitors is one advantage of CLASSY.

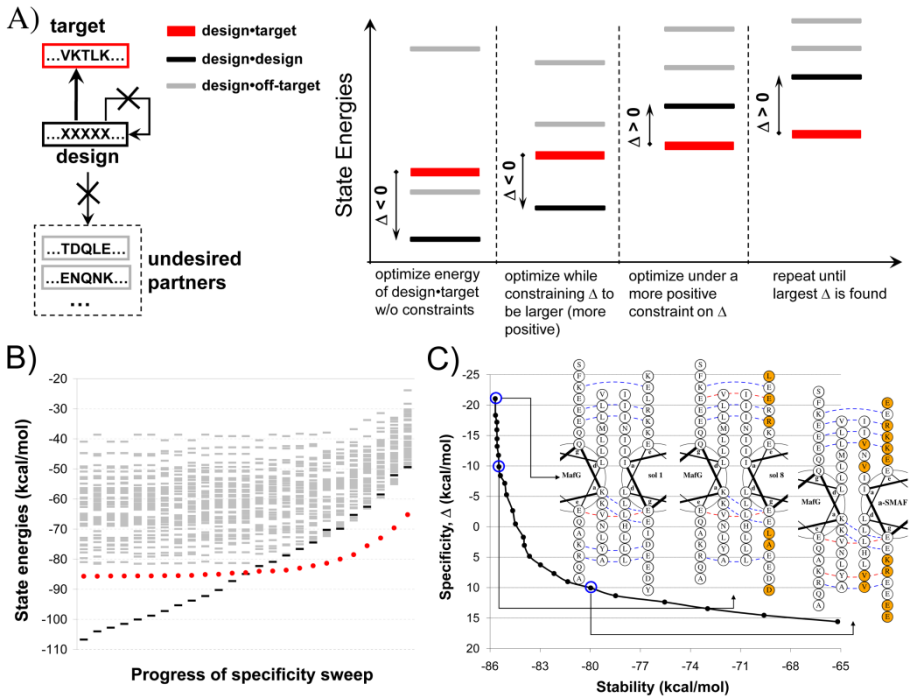



Figure 3.1. Designing specific peptides using CLASSY. A) Specificity sweep scheme. A sequence (black) is sought that binds a target (red) but not several undesired partners (gray) or itself. Panels from left to right illustrate iterations of the CLASSY procedure, during which the specificity gap Δ is increased. B) and C) A specificity sweep with MafG as the target and all other human bZIP coiled coils (except MafK, in the same family as MafG) and the design homodimer as undesired complexes. The plot in B corresponds to the cartoon in A. Red dots, black bars and gray bars represent energies of the design•target, design•design, and design•other bZIP complexes, respectively. C plots design•target complex stability vs. specificity (Δ). Portions of several designed complexes are shown using helical wheels (orange highlights amino-acid changes from the previously shown sequence). The rightmost solution is anti-SMAF.

In total, 48 peptides designed against 20 targets were tested for interaction with 33 representative human bZIP coiled coils and for self-association. Fluorescence intensities measured on bZIP arrays have previously been shown to reflect relative interaction strengths measured in solution (Newman and Keating. 2003). Each peptide in turn (both designed and native) was labelled with the fluorescent dye Cy-3 and used to probe aldehyde-derivatized slides printed with potential partners. Of the 48 designs tested, 40

bound to their intended target, as assessed by fluorescence signal above background (Figure B.S1). The probability of this occurring by chance, given the distribution of design-human interaction signals from the arrays, was ~$10^{-11}$. Self-association of the designs was also evaluated. Only 40% of the designs showed detectable self-interactions using the same criterion, and all but 6 interacted with a human bZIP more strongly than they interacted with themselves (Figure 3.2A and Figure B.S1).

To determine the interaction *specificity* of the designed molecules, we used Cy-3 labelled designed peptides and compared the array signal for interaction with the target to that for interaction with non-target competitors. Results for the most specific design identified for each of the 20 families are shown in Figure 3.2A. These designs are named using the target family name. For 10 designs, the strongest interaction observed was with the intended target. Strikingly, 8 of these designs bound their targets with array signals distinctly greater than for any other non-target-family partner (targets: ZF, cFos, MafG, ATF-2, cJun, cMaf, XBP-1, ATF-4, leftmost in the *Specificity* panel of Figure 3.2A). This indicates measurable interaction specificity on the arrays. For 2 more designs, fluorescence signal for interaction with the target was only marginally greater than that for interaction with 1-2 other proteins (targets: ATF-3, C/EBPγ). Nine other designs bound their targets, but less strongly than they bound to members of other families. For one target family (PAR), the designed peptide did not show detectable binding above background.

To assess the *stability* of each design•target interaction, we labelled each native bZIP target with Cy-3 and probed an array containing 33 representative human bZIP peptides as well as the anti-target design. This experiment assayed design•target

stability relative to interactions of the target with its native partner(s). The strongest signal was often from the design•target complex, indicating that many designs can be expected to out-compete native partners of the targets, using modest concentrations (summarized in Figure 3.2A, complete data in appendix B). Less stable designs can likely be improved through generic strategies such as the addition of acidic extensions, as for the A-ZIPs (Gerdes, et al. 2006).

To validate the array assay, 28 mixtures involving the 7 best designs were characterized in solution using thermal denaturation monitored by circular dichroism. Each designed peptide was tested for interaction with (1) its target, (2) its next-best interaction partner, as reported by the array, (3) a protein closely related by sequence to the target, and (4) itself. We monitored whether the mixtures showed an increase in the temperature of denaturation ($T_m$) compared to that expected from the average of the signals of the individual components (Figs 2B-E and Figure B.S3-8). In all cases, the $T_m$ studies supported binding of each design to its intended target. For the 21 undesired complexes tested, 18 either showed no evidence for interaction or a $T_m$ that was clearly lower than that of the design•target complex. For the remaining 3 undesired complexes, formation of mixtures complicated the analyses, although these are probably also weaker than the corresponding design•target complexes (Figure B.S4- 6). Solution data were also examined for consistency with the array measurements and supported the same relative ordering of stabilities for 35 of 41 comparable cases (see appendix B).

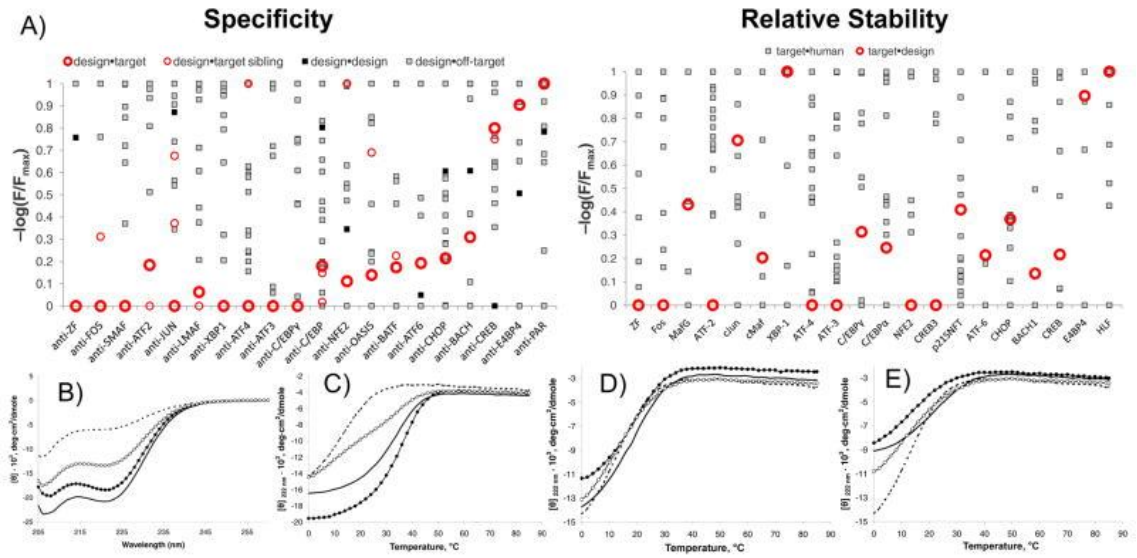



Figure 3.2. Experimental testing of anti-bZIP designs. A) Peptide array results for the most specific design identified for each human bZIP family. Columns show experiments using the indicated protein to probe an array. For the Specificity panel (left), designs in solution were used to probe human bZIPs and designs on the surface. In the Relative Stability panel (right), human bZIP targets were used to probe an array containing the cognate design of each target and 33 human bZIPs. Data are plotted as $-\log(F/F_{max})$, with F the fluorescence signal on the array, such that the strongest interaction has a value of zero. Values of $-\log(F/F_{max})$ above 1.0 were set to 1.0. Thick red circles – design•target; thin red circles – design interactions with siblings in the target family; grey squares – interactions with other human bZIPs; black squares – design•design. Designs are named using the family of their target. B) Solution testing of anti-SMAF complexes assayed using circular dichroism. In each panel, anti-SMAF alone is shown with dashed lines, the partner being tested with a solid line, the numerical average of these two signals with open circles (○) and the mixture of the two peptides with closed circles (●). (B, C) Anti-SMAF interacts with target MafG ($T_m$ ~ 38 °C). (D) Anti-SMAF interacts, at most, very weakly with cJun, the closest competitor according to microarray data. (E) There is no evidence for anti-SMAF interacting with MafB, a sequence closely related to the target. CD spectra in (B) were collected at 25 °C. Anti-SMAF unfolds with $T_m$ ~12 °C. Similar data for other complexes are included in Figures B.S3-8.

Three of our best designs target cJun, cFos, and ATF-2. These proteins are constituents of the AP-1 transcription factor complexes involved in cell proliferation and oncogenesis. The cJun•cJun, cJun•cFos, and cJun•ATF-2 dimers are involved in these important processes in ways that have not been fully elucidated. Complexes involving cJun have previously been targeted for disruption using a dominant-negative A-ZIP version of cFos (Gerdes, et al. 2006). But because cFos also binds ATF-2 and its family members (Newman and Keating. 2003), the A-ZIP strategy is not as specific as might be desired. The same is true for cJun and ATF-2: native partners of these targets also bind to additional families. Attempts to identify new partners for cFos and cJun using experimental selection strategies gave peptides that strongly self-associated and also bound bZIPs non-specifically (i.e. the intended anti-cFos and anti-Jun peptides bound both FOS and JUN family members tightly) (Mason, et al. 2006, Mason, et al. 2007). Our designed peptides provide a way to introduce specificity, e.g. to disrupt cJun•cFos but not cJun•cJun or cJun•ATF-2, using anti-FOS.

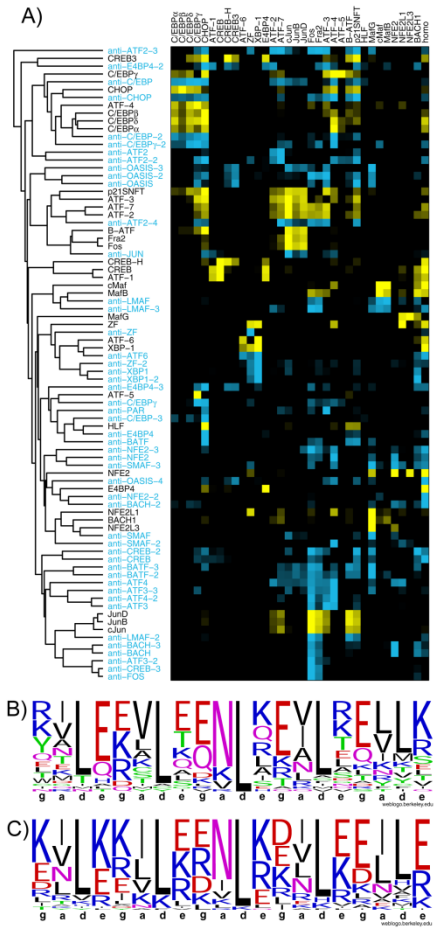



Figure 3.3. Properties of designed peptides compared to human bZIP leucine-zippers. A) Hierarchical clustering of interaction profiles for 33 human peptides and 48 designs; an interaction profile consists of the array signals for interactions with 33 surface-bound human peptides. Proteins on the surface are in columns and those in solution are in rows, with designed proteins and their interaction profiles in blue and human bZIP interaction profiles in yellow. B), C) Sequence logos for **a**, **d**, **e**, and **g** positions from the first 5 heptads of the 33 human bZIP leucine zippers in B) and the 48 designed peptides in C) (http://weblogo.berkeley.edu).

*Properties of the anti-bZIP designs*

Figure 3.3A shows the interaction profiles of native bZIP leucine zippers and the designed anti-bZIP peptides. The native proteins exhibit diverse interaction properties, despite their limited sequence variability (Figure 3.3B)[4]. The designed peptides are even more limited in sequence diversity, yet they encode many additional, novel specificity profiles, suggesting that bZIP-like coiled-coil interaction space is only sparsely sampled

by the human proteins (Figure 3.3C). Based on the frequency of success of our interaction prediction model, and results from CLASSY analysis, we conservatively estimate that >1,900 very distinct interaction profiles can be encoded using the restricted sequence space employed in our designs. This may prove useful for applications in synthetic biology (see appendix B).

CLASSY designs exhibited canonical bZIP specificity determinants, such as a preference for Asn residues at **a** positions to pair across helices, and charge complementarity at **g**-to-**e'** pairs (see Figure 3.1C for coiled-coil heptad positions; a prime indicates a residue on the opposite helix, see Figure B.S15) (Vinson, et al. 2006, Lupas and Gruber. 2005). Interestingly, **g**-to-**a**' pairs were predicted to make a comparable, if not larger, contribution to specificity than **g**-to-**e'** pairs. Other unanticipated specificity patterns also emerged, involving steric interactions between **a** and **d**′ sites (see appendix B for a fuller discussion). The significance of such interactions has not been broadly recognized in parallel coiled coils, although recent studies suggest their importance in anti-parallel dimers (Hadley, et al. 2008).

## DISCUSSION

CLASSY provides a way to analyze and optimize stability/specificity tradeoffs in protein design. The CE/ILP procedure imposes few formal requirements on the type of scoring function that can be used or the type of specificity problem that can be addressed. However, measuring and predicting interaction specificity for proteins generally remains challenging. Here, the bZIPs provided several advantages. The bZIP microarray assay benefits from reversible folding of short coiled coils, and data from prior array measurements of many bZIP transcription factor pairs were critical for

developing predictive models (Newman and Keating. 2003, Fong, et al. 2004, Grigoryan and Keating. 2006). Experimental helix propensities contributed to the quality of these models, and knowledge of particular specificity determinants (e.g. the special role of Asn pairs) improved predictions and also disfavoured the formation of higher-order oligomers (Vinson, et al. 2006). Finally, symmetric fixed-backbone models proved adequate for this application (Grigoryan and Keating. 2006). This facilitated both structural modelling and cluster-expansion training, although CE can also be used for asymmetric structures and with flexible backbones (Apgar, et al. 2009). Further details about features specific to bZIP modelling are in Methods and appendix B.

Determinants of protein interaction specificity are not yet as well understood for other complexes, but significant progress in this area is evident. Zinc-finger/DNA, SH2/peptide and PDZ/peptide complexes have been extensively studied, and both assays and interaction models have been developed that make these good candidates for design using CLASSY (see appendix B) (Jones, et al. 2006, Wiedemann, et al. 2004, Reina, et al. 2002, Sanchez, et al. 2008, Kaplan, et al. 2005). Large-scale interaction experiments are becoming more common, and general-purpose models to describe protein structures and energies are under development (Sanchez, et al. 2008, Boas and Harbury. 2007, Das and Baker. 2008, Zhou and Zhou. 2002). Advances in these areas will expand the problems that can be addressed using CLASSY. In the long term, we hope this approach will help address how interaction crosstalk can be controlled in both evolved and designed protein systems.

## METHODS SUMMARY

Structure-based modelling of coiled-coil interactions was done as previously described, with modifications detailed in the Methods and appendix B (Grigoryan and Keating. 2006). Using the technique of cluster expansion, structure-based models were converted to functions of sequence that included constant, single-residue and residue-pair terms. Training of the cluster expansion used 61,780 random bZIP-like sequences that were modelled structurally (Grigoryan, et al. 2006, Zhou, et al. 2005). A limited amino-acid alphabet was considered, which included the 10 residues most frequently found at each coiled-coil heptad position in native bZIPs. Constrained optimization employing integer linear programming (ILP) was used to design **a**, **d**, **e** and **g** sites. ILP optimization minimized the energy of design•target complexes, subject to constraints on the energy gap with respect to undesired complexes and the match of the design sequence to a position-specific scoring matrix derived from 432 native bZIP leucine zippers. Other positions in the coiled-coil repeat (**b**, **c** and **f** positions) were chosen to be consistent with the designed interface **a**, **d**, **e** and **g** residues, using a probabilistic framework. For each design target, the ILP optimization was repeated with increasing values of the specificity gap parameter $\Delta$, in a procedure termed a specificity sweep. Sequences for experimental testing were selected manually from candidates generated using the specificity sweeps.

For experimental testing, $His_6$-tagged peptides were expressed in RP3098 cells and purified by Ni-NTA followed by reverse-phase HPLC. Coiled-coil microarrays were printed, processed and probed as described previously (Newman and Keating. 2003). Fluorescence signals from the arrays were processed to remove background and normalized. Circular dichroism measurements were performed using standard techniques to measure spectra between 195 and 280 nm at 25 °C or thermal stability by

monitoring ellipticity at 222 nm. Data were fit to appropriate thermodynamic equations to obtain apparent $T_m$s. Detailed descriptions of all procedures are included in Methods and Appendix B.

## Methods

*Modelling bZIP leucine-zipper interactions*

Two variants of the previously described energy function HP/S/C were used to evaluate the relative stability of coiled-coil dimer structures (Grigoryan and Keating. 2006). Models were constructed using a single backbone, with rotameric sampling and continuous relaxation used to position side chains. HP/S/Ca is the model as published (Grigoryan and Keating. 2006), with scale factor $s = 0$ such that intra-chain interactions in the folded structure do not directly contribute to stability (though there are indirect contributions). HP/S/Ca replaces core **a-a'** and **d-d'** terms derived from structure-based calculations with weights from a machine-learning algorithm (Grigoryan and Keating. 2006). In the variant model HP/S/Cv, structure-based **a-a'** interactions were replaced with **a-a'** experimental coupling energies for 55 amino-acid combinations (Acharya, et al. 2006) and the **d-d'** interaction for Leu-Leu was replaced with the empirical value –2 kcal/mol. Following cluster expansion (see below), **a**-position point contributions were adjusted such that 100 folding free energies measured by (Acharya, et al. 2006) were predicted optimally (in the least squares sense, see Figure B.S10). The following 10 amino acids were allowed: V, L, N, I, K, A, R, T, S, and E for **a** positions; L, V, I, M, H, Y, T, A, K, and F for **d** positions; E, K, R, Q, L, S, T, A, V, and I for **e** positions; E, K, Q, R, L, Y, T, D, A, and I for **g** positions. These are the 10 amino acids most frequently encountered in the respective positions in bZIPs. Additionally, for the **a**

position, these are also the 10 amino acids for which Vinson and co-workers have measured coupling energies (Acharya, et al. 2006).

*Cluster expansion*

Cluster expansion (CE) provides a way to express the energy of a sequence adopting a particular backbone structure as an algebraic function of the sequence itself[28]. The formal basis of the technique is described in the appendix B. In this study, the desired and undesired structures had the same backbone, and thus one cluster expansion (for parallel, coiled-coil dimers) was sufficient. CE calculations for both HP/S/Ca and HP/S/Cv included single-residue and residue-pair terms. A training set was built by randomly generating 61,780 coiled-coil sequences with heptad position-specific amino-acid probabilities taken from a multi-species alignment of 432 bZIPs (personal communication with Mona Singh, Princeton University). Gly and Pro were not included. Pair contributions were included only for amino-acid pairs $\leq 7$ residues apart, resulting in 9,929 possible effective cluster interactions (ECI): 1 constant, 68 point and 9,860 pair terms. After the fitting procedure, 2,544 and 2,470 ECI survived the statistical significance test (e.g. lowered the cross-validated error (Grigoryan, et al. 2006)) for HP/S/Ca and HP/S/Cv, respectively. The performance of the resulting cluster expansions on a similarly generated test set of 10,000 sequences not used in training is shown in Figure B.S11.

*Multi-state design optimization*

Design sequences were optimized for interaction with the target using integer linear programming (ILP), imposing constraints on the design interaction energy with

competitors and a degree of match to a bZIP position-specific scoring matrix (PSSM). The ILP and PSSM are detailed in the appendix B. We performed two types of CLASSY calculations. The first, a *specificity sweep*, starts by using ILP to identify the design sequence that produces the provably lowest predicted binding energy to a target. Given this sequence, energy gaps between the design•target dimer and all design•competitor dimers, including design•design, are calculated as $gap = E_{design.competitor} - E_{design.target}$. The minimal gap (which may be negative) is defined as Δ. In the next iteration of the specificity sweep, the design•target energy is re-optimized, this time imposing constraints to require that all gaps be greater than Δ + 1 kcal/mol. In each round, Δ is updated and this procedure is repeated until no more solutions exist (Figure 3.1A). Designs to be tested are chosen from this list of optimized sequences, as discussed in the main text.

Anti-bZIP designs were tested in three rounds of microarray experiments. When we sought to improve upon a previously tested design, we sometimes used experimental results to formulate *biased specificity sweeps*. In these calculations, custom offsets were applied to enhance or diminish the significance of some gaps relative to others; the remainder of the procedure was identical to that for a standard specificity sweep. For example, a biased specificity sweep was used to design anti-ZF after the first design tested (named as anti-ZF-2) interacted with XBP-1 more strongly than with ZF, contrary to predictions of the model. This is illustrated and explained further in Figure B.S9. Table B.S1 contains a list of all designs and the procedures by which they were obtained, including the details of any biased specificity sweeps employed.

In all CLASSY procedures, except where noted in Table B.S1, 46 human bZIPs were considered (sequences take from (Fong, et al. 2004)), and the modelled states were

as follows: the design•target complex was the only desired state; design•off-target bZIP complexes for all bZIPs not in the family of the target bZIP were treated as undesired; the design•design homodimer was also an undesired state.

Further details on the theory behind CLASSY, as well as other computational analyses performed in this study, are in appendix B.

*Choosing 33 representative human bZIPs*

To avoid redundancy and conserve resources and time we used a representative set of 33 human bZIPs that covered all 20 families (see Figure B.S13). Representatives were chosen based on sequence similarity and reagent availability and described well the distinct interaction profiles reported by Newman and Keating (Newman and Keating. 2003). Computational design was nevertheless conducted with 46 human bZIPs taken from (Newman and Keating. 2003).

*Plasmid construction and peptide expression, purification and labelling*

Synthetic genes encoding all designs were constructed using DNAWorks (Hoover and Lubkowski. 2002) to design primers that contained flanking BamHI and XhoI restriction sites. A two-step PCR method was used to assemble the primers and the PCR products were digested with BamHI and XhoI and cloned into a modified pDEST17 vector (Newman and Keating. 2003). All synthetic genes were confirmed to be correct by sequencing. Plasmids encoding human leucine-zipper peptides have been previously published in (Newman and Keating. 2003) with the exceptions of modified Jun family constructs that are described in the appendix B.

Plasmids were transformed into RP3098 cells and 1 L cultures in LB were grown to 0.4-0.6 OD and induced at 37 °C for 3-4 hours with addition of 1mM IPTG. Peptides were purified under denaturing conditions (guanidine hydrochloride, GuHCl) by binding to Ni-NTA resin and eluted with 60% acetonitrile/1% TFA. Following reduction with 10 mM TCEP in 5% acetic acid for 3 minutes at 65 °C, peptides were further purified using reverse-phase HPLC. The molecular weights of all designed peptides were confirmed as correct to within 0.15% by mass spectrometry. To generate dye labelled-peptides, 10 molar excess of Cy3 NHS ester in 6 M GuHCl/100 mM phosphate (pH 7.5) was added to lyophilized aliquots of protein and incubated for 2 hours at room temperature. Free dye was removed using size-exclusion spin columns. Labelled peptides were stored at -80 °C.

*Preparation and probing of arrays*

Lyophilized aliquots of protein were resuspended to a concentration of 40 µM in 6 M GuHCl/100 mM phosphate (pH 7.5)/0.04% Tween-20/10 µM Alexa Fluor 633 hydrazide. Proteins were printed on aldehyde-presenting glass slides (Thermo Fisher Scientific) using a Microgrid TAS Arrayer. Twelve identical subarrays were printed on each slide. Each protein was spotted twice, in two different print orders, for a total of four spots for each protein per subarray. After printing, slides were divided into subarrays by drawing a hydrophobic boundary (PAP pen, Electron Microscopy Sciences). Slides were stored at -80 °C for up to 1 month.

Slides were prepared for probing by: (1) washing face up in -80 °C ethanol for 30 seconds; (2) transferring to 80% ethanol/10 mM NaOH and incubating with shaking for 15 minutes; (3) washing in $H_2O$ for 15 seconds; (4) incubating in PBS/0.1% Tween-

20 for 15 minutes with shaking; (5) drying by centrifugation. Slides were then immediately probed by diluting labelled peptide in 6 M GuHCl/100 mM phosphate (pH 7.5)/6 mM TCEP 6-fold into 1.2X Buffer (1.2% BSA, 1.2X PBS, 0.12% Tween-20). The resulting solution was mixed and 35 µl was immediately pipetted onto each subarray. Each sample was probed in duplicate on adjacent subarrays, for a total of 8 spots used to detect each interaction. Slides were covered with a box and incubated for 1 hour. Slides were washed in PBS/0.1% Tween-20 for 15 seconds and then $H_2O$ for 15 seconds and were then dried by centrifugation. Slides were scanned using a DNA Microarray Scanner (Agilent) at several photo-multiplier tube voltage levels. The concentration of probe was 160 nM unless otherwise indicated.

Additional details on experimental techniques and data analysis are provided in appendix B.

**ACKNOWLEDGEMENTS**

This work was supported by NIH award GM67681 and used computer equipment purchased under NSF award 0216437. We thank the MIT BioMicro center for arraying instrumentation and R.T. Sauer, M. Singh, B. Tidor, M. Laub, T.A. Baker, J.H. Davis, M.S. Kay, J.R.S. Newman, W. F. DeGrado and members of the Keating lab, especially O. Ashenberg and T.C.S. Chen, for thoughtful comments on the manuscript.

## REFERENCES

1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006;45(38):11324-32.

2. Ali MH, Taylor CM, Grigoryan G, Allen KN, Imperiali B, Keating AE. Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure. Structure. 2005 Feb;13(2):225-34.

3. Apgar JR, Hahn S, Grigoryan G, Keating AE. Cluster expansion models for flexible-backbone protein energetics. J Comput Chem. 2009 Nov 30;30(15):2402-13.

4. Boas FE, Harbury PB. Potential energy functions for protein design. Curr Opin Struct Biol. 2007 Apr;17(2):199-204.

5. Bolon DN, Grant RA, Baker TA, Sauer RT. Specificity versus stability in computational protein design. Proc Natl Acad Sci U S A. 2005 Sep 6;102(36):12724-9.

6. Das R, Baker D. Macromolecular modeling with rosetta. Annu Rev Biochem. 2008;77:363-82.

7. Deutsch JM, Kurosky T. New algorithm for protein design. Phys Rev Lett. 1996 Jan 8;76(2):323-6.

8. Fong J, Keating A, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biology. 2004;5(2):R11.

9. Fu X, Apgar JR, Keating AE. Modeling backbone flexibility to achieve sequence diversity: The design of novel alpha-helical ligands for bcl-xL. J Mol Biol. 2007 Aug 24;371(4):1099-117.

10. Gerdes MJ, Myakishev M, Frost NA, Rishi V, Moitra J, Acharya A, Levy MR, Park SW, Glick A, Yuspa SH, Vinson C. Activator protein-1 activity regulates epithelial tumor cell identity. Cancer Res. 2006 Aug 1;66(15):7578-88.

11. Grigoryan G, Zhou F, Lustig SR, Ceder G, Morgan D, Keating AE. Ultra-fast evaluation of protein energies directly from sequence. PLoS Comput Biol. 2006 Jun 16;2(6):e63.

12. Grigoryan G, Keating AE. Structure-based prediction of bZIP partnering specificity. Journal of Molecular Biology. 2006;355(5):1125-42.

13. Hadley EB, Testa OD, Woolfson DN, Gellman SH. Preferred side-chain constellations at antiparallel coiled-coil interfaces. Proc Natl Acad Sci U S A. 2008 Jan 15;105(2):530-5.

14. Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. Nat Struct Biol. 2003 Jan;10(1):45-52.

15. Hoover DM, Lubkowski J. DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. Nucl. Acids Res. 2002 May 15, 2002;30(10):e43.

16. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006 Jan 12;439(7073):168-74.

17. Kangas E, Tidor B. **Electrostatic specificity in molecular ligand design**. J Chem Phys. 2000;112(20):9210-11.

18. Kaplan T, Friedman N, Margalit H. Ab initio prediction of transcription factor targets using structural knowledge. PLoS Comput Biol. 2005 Jun;1(1):e1.

19. Kingsford CL, Chazelle B, Singh M. Solving and analyzing side-chain positioning problems using linear and integer programming. Bioinformatics. 2005 Apr 1;21(7):1028-36.

20. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol. 2004 Apr;11(4):371-9.

21. Krylov D, Barchi J, Vinson C. Inter-helical interactions in the leucine zipper coiled coil dimer: PH and salt dependence of coupling energy between charged amino acids. J Mol Biol. 1998 Jun 19;279(4):959-72.

22. Krylov D, Olive M, Vinson C. Extending dimerization interfaces: The bZIP basic region can form a coiled coil. EMBO J. 1995 Nov 1;14(21):5329-37.

23. Landgraf C, Panni S, Montecchi-Palazzi L, Castagnoli L, Schneider-Mergener J, Volkmer-Engert R, Cesareni G. Protein interaction networks by proteome peptide scanning. PLoS Biol. 2004 Jan;2(1):E14.

24. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. Adv Protein Chem. 2005;70:37-78.

25. Mason JM, Muller KM, Arndt KM. Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. Biochemistry. 2007 Apr 24;46(16):4804-14.

26. Mason JM, Schmitz MA, Müller KM, Arndt KM. Semirational design of jun-fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. Proceedings of the National Academy of Sciences. 2006 June 13, 2006;103(24):8989-94.

27. Newman JRS, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 June 27, 2003;300(5628):2097-101.

28. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C. Computer-aided design of a PDZ domain to recognize new target sequences. Nat Struct Biol. 2002 Aug;9(8):621-7.

29. Sanchez IE, Beltrao P, Stricher F, Schymkowitz J, Ferkinghoff-Borg J, Rousseau F, Serrano L. Genome-wide prediction of SH2 domain targets using structural information and the FoldX algorithm. PLoS Comput Biol. 2008 Apr 4;4(4):e1000052.

30. Shifman JM, Mayo SL. Exploring the origins of binding specificity through the computational redesign of calmodulin. Proc Natl Acad Sci U S A. 2003 Nov 11;100(23):13274-9.

31. Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: A system-level analysis. PLoS Biology. 2005 October 01, 2005;3(10):e334.

32. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. Science. 2007 July 20, 2007;317(5836):364-9.

33. van der Sloot AM, Tur V, Szegezdi E, Mullally MM, Cool RH, Samali A, Serrano L, Quax WJ. Designed tumor necrosis factor-related apoptosis-inducing ligand variants initiating apoptosis exclusively via the DR5 receptor. Proc Natl Acad Sci U S A. 2006 Jun 6;103(23):8634-9.

34. Vinson C, Acharya A, Taparowsky EJ. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2006;1759(1-2):4-12.

35. Wiedemann U, Boisguerin P, Leben R, Leitner D, Krause G, Moelling K, Volkmer-Engert R, Oschkinat H. Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. J Mol Biol. 2004 Oct 22;343(3):703-18.

36. Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, DeGrado WF. Computational design of peptides that target transmembrane helices. Science. 2007 March 30, 2007;315(5820):1817-22.

37. Zhou F, Grigoryan G, Lustig SR, Keating AE, Ceder G, Morgan D. Coarse-graining protein energetics in sequence variables. Phys Rev Lett. 2005 Sep 30;95(14):148103.

38. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 2002 Nov;11(11):2714-26.

# CHAPTER 4

# A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering

**Reproduced with permission from:**
Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J Am Chem Soc. 2010 May 5;132(17):6025-31.

**Supporting information:**
This paper included supplemental tables and figures which are in appendix C.

**Collaborator notes:**
Robert Grant helped in solving the two SYNZIP crystal structures.

**ABSTRACT**

The versatile coiled-coil protein motif is widely used to induce and control macromolecular interactions in biology and materials science. Yet the types of interaction patterns that can be constructed using known coiled coils are limited. Here we greatly expand the coiled-coil toolkit by measuring the complete pair-wise interactions of 48 synthetic coiled coils and 7 human bZIP coiled coils using peptide microarrays. The resulting 55-member protein 'interactome' includes 27 pairs of interacting peptides that preferentially hetero-associate. The 27 pairs can be used in combinations to assemble sets of 3 to 6 proteins that compose networks of varying topologies. Of special interest are heterospecific peptide pairs that participate in mutually orthogonal interactions. Such pairs provide the opportunity to dimerize two separate molecular systems without undesired crosstalk. Solution and structural characterization of two such sets of orthogonal heterodimers provide details of their interaction geometries. The orthogonal pair, along with the many other network motifs discovered in our screen, provide new capabilities for synthetic biology and other applications.

## INTRODUCTION

The coiled coil is a fundamental building block for molecular engineering. Its simple structure, which consists of two or more alpha helices twisted into a supercoiled rod-like bundle, is encoded in a seven amino-acid repeat designated [**abcdefg**]$_n$. Coiled coils have been used to induce and stabilize protein oligomers, to promote protein-protein interactions, to rewire cellular networks, to assemble functional scaffolds, to construct hydrogel materials, and to self-assemble nano-scale fibers and/or recruit ligands to nanopartices (Bashor, et al. 2008, Diehl, et al. 2006, Eckert, et al. 1999, Papapostolou, et al. 2007, Takagi, et al. 2001, Wolfe, et al. 2003, Petka, et al. 1998, McAllister, et al. 2008, Mapp, et al. 2000). Important early advances in coiled-coil engineering included demonstrating that leucine-zipper peptides, which are short coiled coils of ~40 amino acids, can fold to give stable structures composed of two to four helices, and that coiled coils can be modified using charge patterning to encode heterospecificity and helix orientation (Mason, et al. 2007). In particular, peptide "Velcro" is a designed heterospecific coiled-coil dimer with glutamates at all interfacial **e** and **g** positions on one helix and lysines at all **e** and **g** positions on the other; this heterodimer and variants of it have been widely employed in bio-molecular engineering. Further experiments have illustrated how residues at the hydrophobic interface, particularly those in **a** positions, can be mutated to modulate interaction affinity and introduce additional specificity (Acharya, et al. 2006). Prior studies not only generated reagents that have found many uses, but also elucidated structural principles that control interaction selectivity (Arndt, et al. 2002, Moll, et al. 2001, O'Shea, et al. 1993a).

Heterodimeric coiled-coil pairs have proven particularly useful for molecular engineering (Arndt, et al. 2002, Moll, et al. 2001, O'Shea, et al. 1993c, Lai, et al. 2004, Diss and Kennan. 2008, Bromley, et al. 2009, Mason, et al. 2006). Exciting recent applications have included using coiled-coil heterodimers to modulate MAP kinase signaling in yeast and inducing ordered structure via coiled coils in nano-scale fibers. Notably, while coiled-coil reagents for inducing homo-oligomerization or hetero-oligomerization of single complexes are widely used, the modern coiled-coil toolkit does not provide access to more complex interaction patterns.

Lacking is a large set of coiled coils that participate in specific and defined interactions with one another. Such reagents could be used to construct interaction networks containing multiple associations in a logical manner. For example, when engineering

cellular circuits it might be desirable to implement multiple parallel pathways, each using coiled coils to direct assembly of signaling complexes without crosstalk. Likewise, to engineer artificial transcription factors, heterodimers with specified cross-interactions could provide access to combinatorial control of binding to different DNA sites. For complex applications such as these, greater versatility is required than is currently provided by characterized coiled-coil peptides.

## RESULTS AND DISCUSSION

We recently reported the computational design of synthetic peptides that interact with the coiled-coil regions of human bZIP transcription factors. These designed peptides are 35-54 residues in length and share an amino-acid composition characteristic of bZIP leucine zippers (Figure C.S1, Table C.S1). Homodimerization of the designed peptides was disfavored by a variety of strategies, and experiments confirmed that most designs do not form strong self-associations (Grigoryan, et al. 2009b). Speculating that this set of heterospecific reagents might harbor interesting and useful interactions patterns, we systematically measured all pair-wise interactions involving 48 designed peptides and 7 additional coiled coils from human bZIPs that do not strongly self-associate.

To identify new heterospecific coiled-coil interactions in a high-throughput manner, we used a protein microarray assay. A complete 55 x 55 interaction matrix was generated by spotting small amounts of each peptide onto aldehyde-derivatized slides (Figure C.S2, Table C.S2). Each of the 55 proteins in turn was labeled with Cy3 dye and used in solution to probe subarrays printed on the slides. This assay is highly reproducible and shows good reciprocity with respect to which protein is immobilized (Figures C.S2 and C.S3). The relative ordering of fluorescence intensities on the arrays has also been shown to agree qualitatively with solution stability measurements (Grigoryan, et al. 2009b, Newman and Keating. 2003).

To discover new pairs of hetero-associating coiled coils, the interaction matrix was examined for peptides that: (1) did not show evidence of homo-association and (2) made strong, reciprocal interactions with a partner. Interacting and non-interacting pairs were chosen conservatively based on comparisons of prior array data with solution data. A total of 27 heterospecific pairs involving 23 synthetic peptides (named SYNZIPs 1-23) and 3 human bZIPs were selected for further analysis (Figure 4.1).

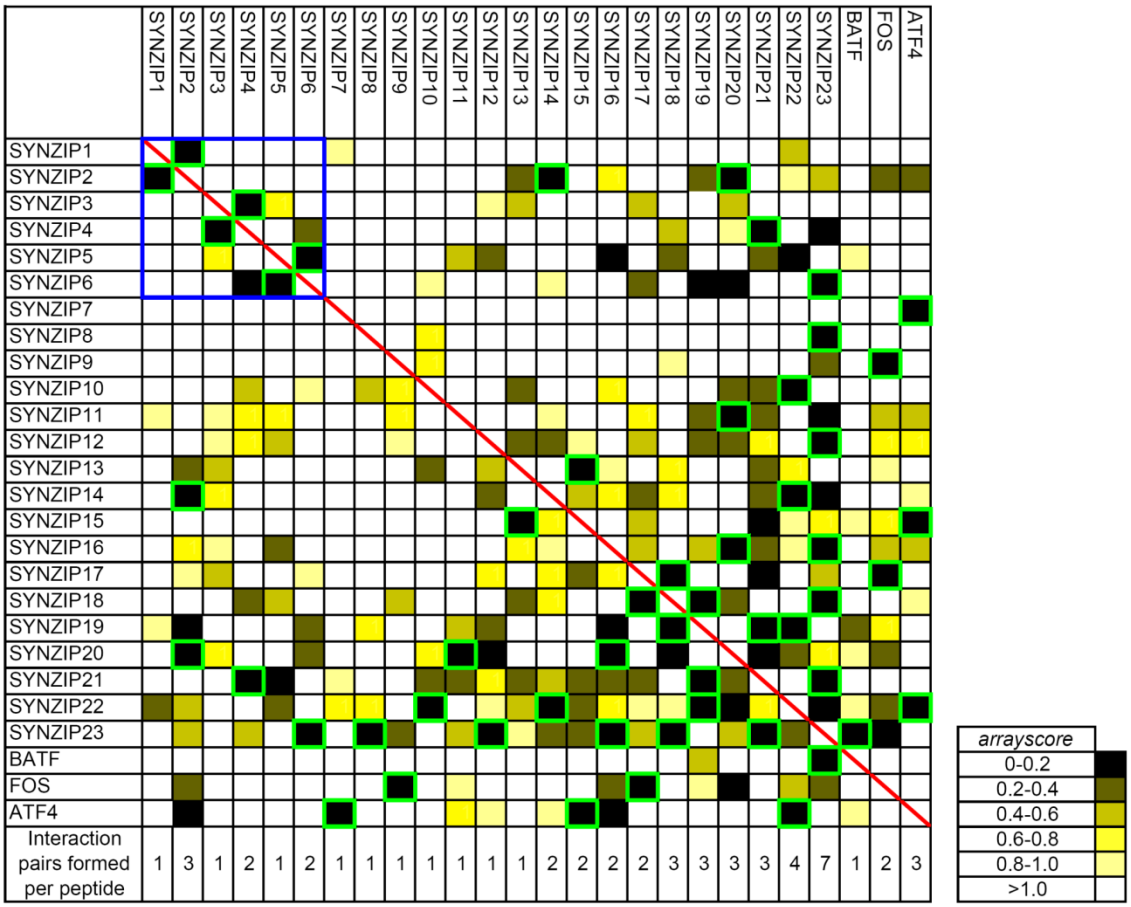



Figure 4.1. Array data describing the interactions of 26 peptides that form specific interaction pairs. Peptides printed on the surface are listed in rows, and fluorescently labeled peptides in solution are listed in columns. Color indicates the strength of the array fluorescence signal, given as *arrayscore* values (see Methods) according to the color bar at right with 0 (black) indicating the strongest signal and >1 (white) indicating the weakest. SYNZIP peptides 1-6, which are further described in Figures 4.2-4, are in the top left corner, boxed in blue. The red diagonal highlights the absence of homoassociation of peptides on the arrays. Interactions that showed arrayscore $\leq 0.2$ in both measurement directions are boxed in green. The number of strong, reciprocal interactions formed by each peptide is listed at bottom of each column.

Coiled coils can vary in their oligomerization state, helix orientation and axial helix alignment (Grigoryan and Keating. 2008). For the heterospecific pairs uncovered in this assay to be maximally useful, knowledge of their interaction geometry is important. The synthetic coil-coiled peptides were designed to interact with individual human bZIPs as parallel dimers, and we hypothesize that most of the design-design and design-human complexes detected on the arrays also form parallel dimers. Several lines of evidence support this. First is the special role of paired **a**-position asparagines in leucine zippers. Interaction of an asparagine residue with another asparagine on an opposing helix is

common in coiled-coil dimers and is much more favorable than an interaction with a hydrophobic residue (which we term an "Asn mismatch," unless the Asn occurs very close to the end of the coiled coil) (Mason, et al. 2007, Acharya, et al. 2006). Paired asparagines at **a** favor parallel dimer formation and are strongly conserved in the parallel, dimeric leucine-zipper transcription factors (Mason, et al. 2007, Moll, et al. 2001, Harbury, et al. 1993). Almost all (23 out of 26) peptides analyzed here contain at least one Asn residue at a coiled-coil **a** position, and of the 27 heterospecific pairs considered, 24 can be aligned such that two **a**-position Asn residues are paired. All heterospecific pairs can be aligned as parallel dimers without any Asn mismatches (Acharya, et al. 2006). In addition to the role of Asn residues, half of the 26 peptides also include a charged residue in one or two non-terminal **a** positions. Lysine in **a** positions has been reported to favor dimer formation over higher order oligomerization, presumably because **a** positions in dimers are less buried (Mason, et al. 2007, Campbell, et al. 2002); this likely applies for other charged side chains as well, as is supported by the lower frequencies of Lys, Arg and Glu residues in **a** positions of parallel trimers compared to parallel dimers (K. Gutwin and A. Keating, unpublished data). Additional indirect criteria support parallel dimer formation. For example, when considered as parallel dimers, all pairs can be aligned such that net **g-e'** electrostatic interactions are not unfavorable and destabilizing (Mason, et al. 2007, O'Shea, et al. 1993b). Finally, none of the heterospecific interactions encode a motif that has been reported to favor trimer formation (Kammerer, et al. 2005).

Given 27 heterospecific pairs among 26 peptides that likely form parallel coiled-coil dimers, we analyzed these to identify higher-order patterns of interaction and non-interaction. Each of the 26 peptides participates in 1-7 interactions, suggesting that subnetworks involving more than 2 peptides could be common in our data (Figure 4.1). We searched exhaustively for all subnetworks containing 3-6 proteins and found examples of the 10 topologies shown in Figure 4.2A (Table C.S3) (Wernicke and Rasche. 2006). In that figure, an edge indicates a high-confidence observation of an interaction on the array and the absence of an edge indicates that an interaction was not observed. Most networks are based on motifs we describe as "pair", "line", or "hub" structures. Many networks are composed of smaller networks, such as the 4 node "orthogonal pair" (2 pairs with no cross-interactions), "orthogonal triplet" (3 pairs with no cross-interactions) or the 5 node "pair + line" (similarly with no cross interactions). Interestingly, protein

nodes in the networks are sparsely connected. It may be that features engineered to diminish self-association also reduce interaction promiscuity more broadly.

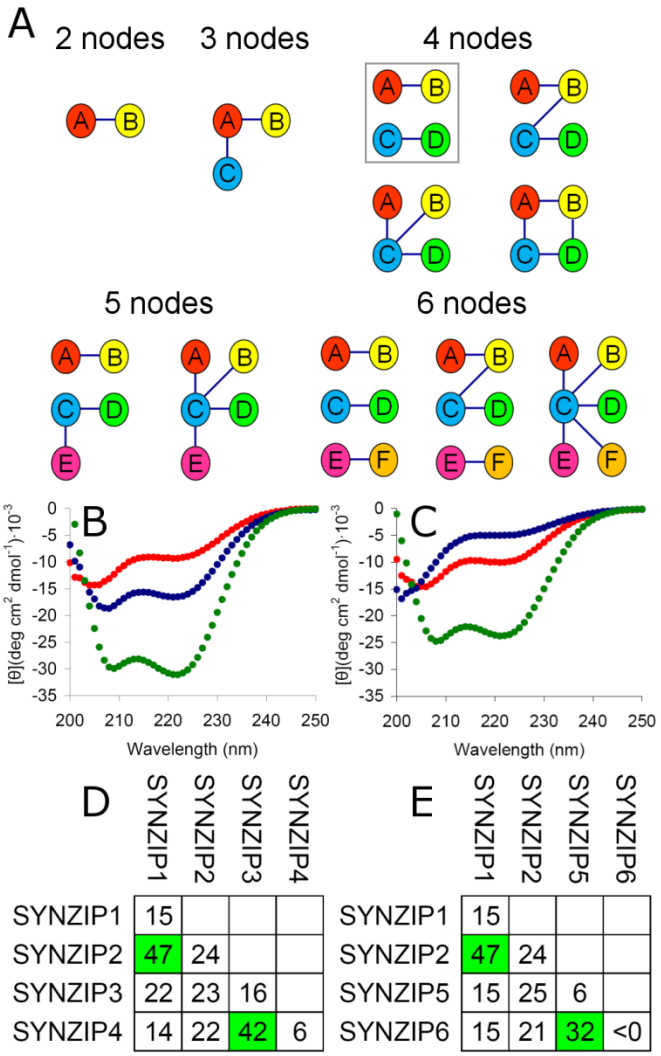



Figure 4.2. SYNZIP coiled coils form specific interaction subnetworks. (A) Graphical representation of subnetworks detected in the coiled-coil array data. Edges indicate an interaction and the absence of an edge between nodes indicates no interaction in the microarray screen. The orthogonal pair motif is boxed in grey. (B, C) CD spectra for two pairs of heterospecific coiled coils (4 μM of each protein and 8 μM total for mixtures, 25 °C). (B) SYNZIP2 (blue), SYNZIP1 (red), and SYNZIP2 + SYNZIP1 (green). (C) SYNZIP4 (blue), SYNZIP3 (red), and SYNZIP4 + SYNZIP3 (green). (D, E) Melting temperatures ($T_m$s) derived from fits to thermal melts of peptide mixtures. $T_m$ values for the interacting pair mixtures are highlighted in green.

Because of its immediate utility, e.g. for direct extension of existing applications, we chose the orthogonal-pair motif for further characterization (Bashor, et al. 2008, Diehl, et al. 2006, Wolfe, et al. 2003). Three coiled-coil pairs were selected that participate in two sets of orthogonal interactions. All three pairs were evaluated in solution using circular dichroism (Figure 4.2B and C, Figure C.S4). The six individual peptides gave only weak helical signal in isolation. But mixing each peptide with its appropriate partner gave a spectrum characteristic of a coiled coil, confirming heterospecific interaction. The orthogonal sets that can be constructed from these three pairs each consist of four peptides that participate in two interactions ('on' states) and eight non-interactions ('off' states). We measured the thermal stabilities of the ten possible interactions for each set (Figure 4.2D and E, Figure C.S5). The 'on' states had melting temperatures between 32 and 47 °C, at 8 μM total peptide concentration. For [SYNZIP6:SYNZIP5, SYNZIP1:SYNZIP2] the difference between the weakest 'on' state and the strongest 'off' state was ~8 °C. For [SYNZIP4:SYNZIP3, SYNZIP1:SYNZIP2] the difference was ~18 °C. (See C.S6 for characterization of an additional orthogonal set.) Previously published orthogonal coiled-coil pairs are either much less stable than this, have the property that at least one 'off' interactions is more stable than one 'on' interaction, or incorporate non-natural amino acids (Lai, et al. 2004, Diss and Kennan. 2008, Bromley, et al. 2009).

To confirm the interaction geometry of complexes composing the orthogonal pairs, we solved the structures of SYNZIP6:SYNZIP5 and SYNZIP1:SYNZIP2 to 2.5 and 1.8 Å, respectively (Figure C.S7, Table C.S4). Both complexes are parallel heterodimers, as anticipated (Figure 4.3A and B). We were unable to obtain crystals of SYNZIP4:SYNZIP3. While it is likely that this pair forms a parallel dimer (it includes **a**-position Asn and Lys residues and highly charged **e**- and **g**-position residues), SYNZIP3 is shorter than SYNZIP4, and the precise axial alignment of its two helices is uncertain. Either of two Asn residues in SYNZIP4 could be paired with the single **a**-position Asn in SYNZIP3, while maintaining a similar extent of coiled-coil dimer. To experimentally determine the alignment, two truncated versions of SYNZIP4 were generated. Each was mixed with SYNZIP3, and the thermal stabilities of the resulting complexes were measured by CD. The N-terminal SYNZIP4 truncation had very similar stability to the full-length peptide, while the C-terminal truncation was markedly destabilized (Figure 4.3C). Thus, the two most N-terminal heptads of SYNZIP4 are dispensable for the

interaction. Based on these experiments, helical wheel diagrams were generated for the three heterospecific pairs (Figure 4.3 D-F).

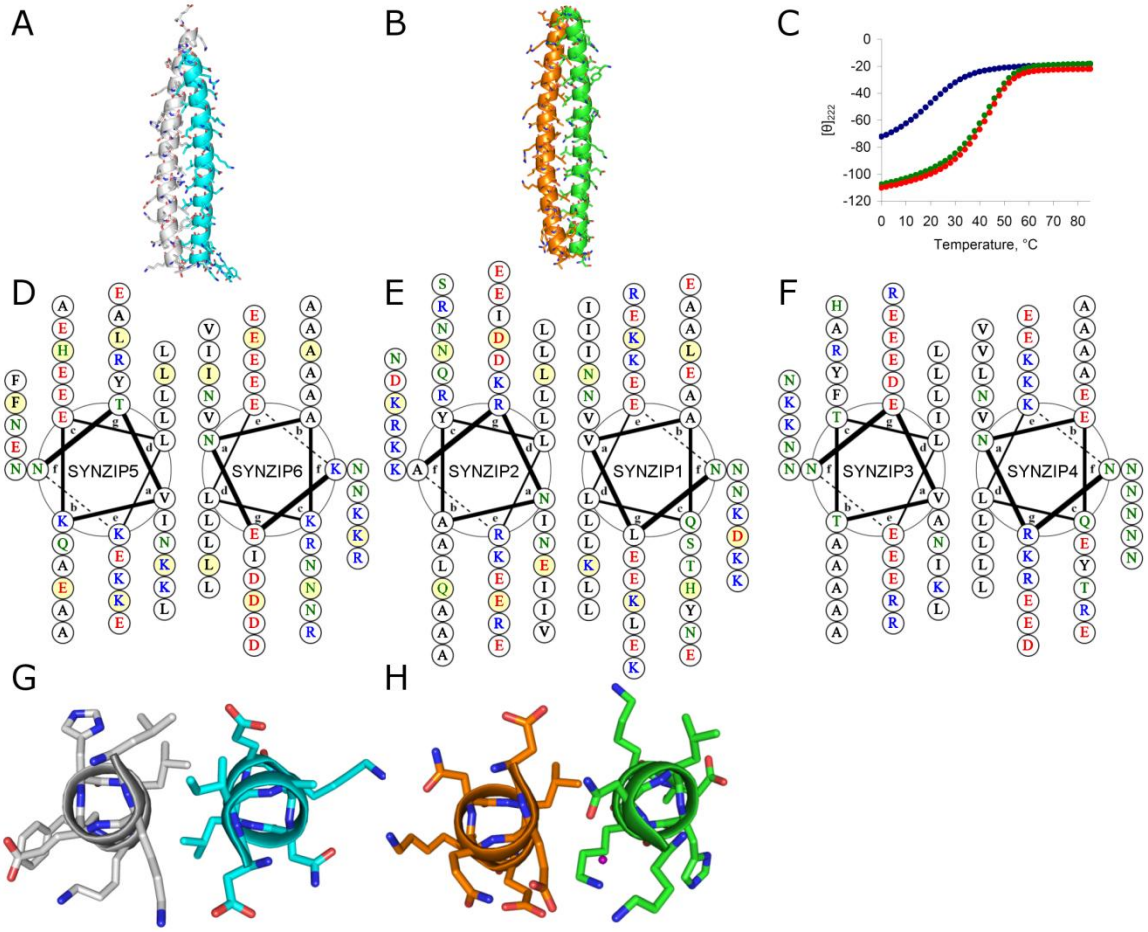



Figure 4.3. Interaction geometries for three heterospecific SYNZIP pairs. (A, B) Crystal structures of SYNZIP5:SYNZIP6 (A) (grey:teal) and SYNZIP2:SYNZIP1 (B) (orange:green) show that both complexes are parallel coiled-coil heterodimers. (C) Determination of the axial alignment of SYNZIP4:SYNZIP3 using CD thermal melts. $SYNZIP4_{1\text{-}54}$: SYNZIP3 (red), $SYNZIP4_{1\text{-}42}$: SYNZIP3, (blue), and $SYNZIP4_{15\text{-}54}$: SYNZIP3 (green). Each mixture was measured at 8 μM total peptide concentration, 4 μM of each peptide. (D-F) Helical wheel diagrams for SYNZIP5:SYNZIP6 (D), SYNZIP2:SYNZIP1 (E), and SYNZIP3:SYNZIP4 (F). Charged residues are colored red/blue, polar residues are in green, and hydrophobic residues are in black. Residues shaded yellow in (D) and (E) correspond to those shown in panels (G) and (H), respectively. (G) The fourth heptad of SYNZIP5 (residues 23-29):SYNZIP6 (residues 37-43), and (H) the fourth heptad of SYNZIP2 (residues 23-29):SYNZIP1 (residues 23-29) are shown in cross-section, as viewed from the N-terminus. A partially buried water molecule is represented in purple. Crystal structure figures generated using PyMOL (DeLano Scientific, Palo Alto, CA). Helical wheel diagrams created using DrawCoil 1.0. (http://www.gevorggrigoryan.com/drawcoil/)

These experiments suggested that portions of each complex were dispensable for the formation of orthogonal pairs. To demonstrate that shorter experimentally determined interaction regions interact specifically, truncated versions of SYNZIPs 1-6 (shown in Figure 4.3 D-F) were cloned with an N-terminal cysteine. Each protein was labeled with biotin. SYNZIPs 1 and 2 were also labeled with Alexa Fluor 546, and SYNZIPs 3, 4, 5, and 6 were labeled with Alexa Fluor 488. For each orthogonal set, each biotinylated protein was pre-mixed with the three other fluorescent proteins and then incubated with NeutrAvidin coated beads. These pull-down experiments showed that each biotinylated protein interacted specifically with its cognate partner (Figure 4.4A and B). Thus, the shorter peptides are sufficient to form specific interactions in four-component mixtures.

The crystal structures of SYNZIP6:SYNZIP5 (PDB ID 3HE4) and SYNZIP1:SYNZIP2 (PDB ID 3HE5) reveal interactions involving polar and charged residues that likely play a role in encoding specificity. Both structures include paired asparagines at **a-a'** positions that adopt conformations seen frequently in other parallel coiled-coil dimers. Neither structure contains any asparagine mismatches at non-terminal heptad positions, although both have mismatches at the extreme N-terminal heptad. At that position, asparagine is paired with valine but remains largely solvent exposed due to its location at the end of the helix. In the SYNZIP6:SYNZIP5 complex, in both the fourth and fifth heptads, Lys at **a** across from Ile interacts with an aspartate at the proceeding **g'** position (Figure 4.3G). In the SYNZIP1:SYNZIP2 complex, the fourth heptad contains a complex polar network involving a partly buried water molecule. The water is coordinated by SYNZIP1 residues Asn 24 at **a** and Lys 27 at **d**, as well as by SYNZIP2 residue Glu 24 at **a'**. In the 3 copies of the heterodimer in the asymmetric unit, Lys 23 at **g** on SYNZIP1, as well as Gln 25 at **b'** and Glu 28 at **e'** on SYNZIP2, are involved to varying degrees in this extended network (Figure 4.3H). These interactions suggest that charged residues in coiled-coil core positions can contribute specificity in parallel dimers, although such residues may be accommodated in ways that are difficult to anticipate, as illustrated here by incorporation of a water molecule.

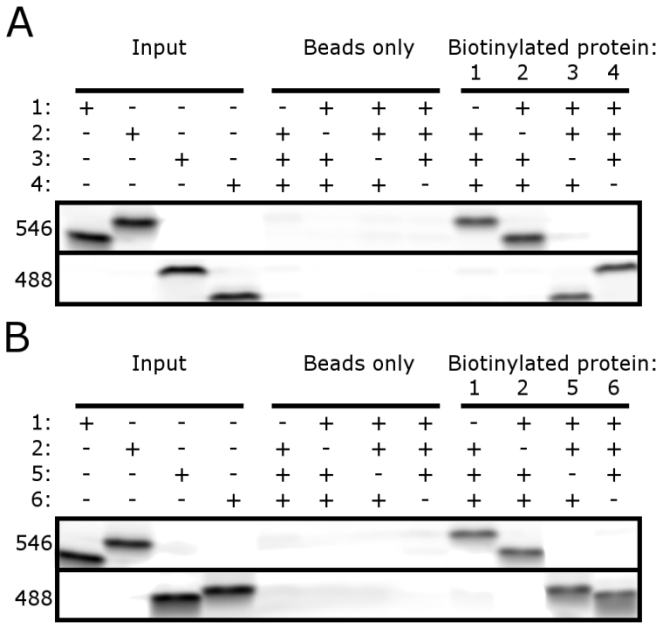



Figure 4.4. Biotin pull-down assay demonstrating specific interactions in each orthogonal set. (A, B) SYNZIPs 1 and 2 were labeled with Alexa Fluor 546 and SYNZIPS 3, 4, 5, and 6 were labeled with Alexa Fluor 488. Input lanes show each protein run individually. The beads-only lanes shows mixtures of the indicated fluorescent proteins incubated with NeutrAvidin beads. The biotinylated-protein lanes show mixtures of the 3 indicated fluorescent proteins (4 µM each) mixed with the indicated biotinylated protein at 4 µM, then incubated with NeutrAvidin beads. The two fluorescent channels 546 nm (top) and 488 nm (bottom) are indicated. (A) SYNZIP pairs 1-2 and 3-4. (B) SYNZIP pairs 1-2 and 5-6.

It is interesting to speculate about how specificity in the orthogonal sets is determined. The simple ACID-BASE charge repulsion strategy used in peptide "Velcro" is not sufficient to encode complex interaction patterns in coiled coils only ~40 amino acids long. How are so many different 'off' states disfavored? Using a simple model, 5 of the 14 'off' pairs among the two orthogonal pair sets have net repulsive electrostatic interactions at **g-e'** positions, when considered as parallel dimers. Unavoidable Asn mismatches appear in an additional 2 pairs. In the remainder, charged residues at **a** and **d** positions appear important, with **a**-position Lys and Glu residues disfavoring homodimerization, and repulsive charges at **g-a'** and **d-e'** pairs disfavoring both homo- and heterodimers (Acharya, et al. 2006). All of these interactions are implicated as useful and important negative design features. In terms of improving specificity, if this is required, we stress that the undesired complexes that form are weak and are not necessarily parallel dimers.

The orthogonal pairs introduced here dramatically increase the number of small, heterospecific protein-protein interaction partners that can be used as modular components for molecular engineering (Bromley, et al. 2008). The peptides can be over-expressed in *Escherichia coli*, contain aromatic amino acids for quantification using spectrometry and lack cysteines. While most of these peptides do partner with human bZIPs, they are likely to be effective for applications in yeast or bacteria, where human orthologs are absent, as well as *in vitro* and for materials applications. These reagents, or molecular parts, are also likely to be useful when paired with other types of synthetic or native interaction domains such as zinc fingers (Giesecke, et al. 2006). It is reasonable to consider using them to design novel transcription factors that do not cross-interact, or to elaborate molecular scaffolds (Bashor, et al. 2008, Wolfe, et al. 2003). Finally, the large number of interactions measured in the course of characterizing these peptides will be useful for testing computational models and further understanding the interaction specificity of "simple" coiled coils.

## METHODS AND MATERIALS

*Plasmid construction, protein expression and purification*

Proteins used in the array experiments were cloned, expressed and purified as published previously (Grigoryan, et al. 2009a). For solution studies and crystallography, genes were cloned into pSV282 (Vanderbilt University Medical Center, Center for Structural Biology) using BamHI and XhoI restriction enzymes (NEB). For the pull-down assays, synthetic genes for truncated peptides including an N-terminal cysteine and a short linker (GSCGS) were cloned based on experimentally determined alignments. SYNZIP6 was mutated at a **c**-position lysine to include a tyrosine for concentration determination. Each plasmid was transformed into RP3098 cells and 1 L cultures in LB were grown to 0.4-0.6 OD and induced at 37 °C for 3-4 hours with the addition of 1mM IPTG. MBP fusion proteins with a $His_6$ tag were purified under native conditions by binding to Ni-NTA resin (Qiagen) and eluting with 8 ml elution buffer (300 mM imidazole, 20 mM Tris, 500 mM NaCl, 1mM DTT, pH 7.9). Fusion proteins were then dialyzed overnight at 4 °C in TEV cleavage buffer (50 mM Tris, 50 mM NaCl, 1 mM DTT, 0.5 mM EDTA, pH 7.5). Peptides were cleaved from MBP by incubating with 100 µl TEV protease (1mg/ml) for 3 hours at room temp. After cleavage, the mixture was added to Ni-NTA resin and the flow through was collected. In the case of SYNZIP2, the

peptide bound the Ni-NTA resin after cleavage. SYNZIP2 was eluted from the resin with 6 M guanidine-HCl and the elute was then dialyzed into water. Peptides were additionally purified using reverse-phase HPLC and lyophilized. The molecular weights of the peptides were confirmed by mass spectrometry. Protein concentrations were determined using the Edelhoch method (Edelhoch. 1967) of UV absorbance at 280 in 6 M guanidine-HCl/100 mM sodium phosphate pH 7.4. Protein and DNA sequences are listed in Table C.S1.

*Coiled-coil array assay*

All array experiments were carried out as previously published (Grigoryan, et al. 2009b), with the exception that only two spots for each protein were printed per subarray, for a total of 8 measurements of each heteromeric interaction. Briefly, lyophilized proteins were resuspended to a concentration of 40 µM in 6 M guanidine-HCl/100 mM sodium phosphate pH 7.5/0.04% Triton X-100/10 µM Alexa Fluor 633 hydrazide. Proteins were printed on aldehyde-derivatized glass slides and 12 identical subarrays per slide were physically divided by drawing a hydrophobic boundary. Slides were blocked, and then each subarray was probed with Cy3-labeled proteins diluted six-fold from 6 M guanidine-HCl/100 mM sodium phosphate pH 7.5/6 mM TCEP to a concentration of ~160 nM in 1.2X buffer (1.2% BSA, 1.2X PBS, 0.12% Tween-20). Slides were then washed, dried, and scanned to obtain fluorescence values for each spot. Average background-corrected fluorescence values are listed in Table C.S2**.**

*Data analysis*

For each peptide pair, fluorescence intensities for the 4 replicate spots corresponding to the same surface/solution arrangement were corrected for background and then averaged. Averages were corrected further by subtracting the median signal for all proteins on the surface interacting with the same solution probe; this gave a value F. The quantity *arrayscore* was calculated by taking $-\log(F/F_{max})$ where $F_{max}$ was the maximum F value for a given solution probe. To identify heterospecific pairs, a strict criterion was employed by comparing *arrayscore* values to $T_m$ measurements of previously published data (Grigoryan, et al. 2009b). Non-interactions were required to have *arrayscore* > 1, which corresponds to an average $T_m$ of 14 °C (based on 13 comparisons). Interactions were required to have *arrayscore* < 0.2, which corresponds to

an average $T_m$ of 43 °C (based on 7 comparisons). These same criteria for interactions and non-interactions were employed to identify subnetworks when using Fanmod (Wernicke and Rasche. 2006) to search for all possible 3-6 node networks. Motifs are listed in Table C.S3**.**

*Circular dichroism*

Circular dichroism spectra were measured on an AVIV 400 spectrometer in 12.5 mM potassium phosphate (pH 7.4)/150 mM KCl. Individual measurements were made at 4 µM peptide or 4 µM of each peptide (8 µM total peptide) for mixtures. All measurements were made in a 1 cm cuvette. Mixtures of peptides were incubated for several hours at room temperature before measurement. Spectra were measured at 25 °C. Wavelength scans were monitored from 280 nm to 195 nm in 1 nm steps, averaging for 5 seconds at each wavelength. Three scans for each sample were averaged. Thermal unfolding curves were performed at 4 µM peptide for individual measurements or 4 µM of each peptide (8 µM total peptide) for mixtures and measured in a 1 cm cuvette with stirring. Melting curves were determined by monitoring ellipticity at 222 nm with an averaging time of 30 seconds, an equilibration time of 1.5 minutes, and a scan rate of 2 °C/min. All samples were measured from 0 to 85 °C. $T_m$ values were estimated as reported previously (Grigoryan, et al. 2009b). All thermal denaturations were reversible, with differences in $T_m$ values upon folding vs. unfolding of < 2°C for all but 2 weak complexes, and < 5 °C in all cases.

For a third orthogonal set of coiled-coil heterodimers, a slightly modified CD protocol was employed. The CD spectra in Figure 4.S6 were measured on an Aviv Model 202 spectrometer in 12.5 mM potassium phosphate (pH 7.4)/150 mM KCl. Individual measurements were made at 40 µM peptide and mixtures at 20 µM of each peptide, 40 µM total peptide. Mixtures of peptides were incubated for several hours at room temperature before measurement. Spectra were measured at 25 °C. Wavelength scans were performed in a 0.1 cm cuvette and were monitored from 260 nm to 195 nm in 1 nm steps averaging for 5 seconds at each wavelength.

*Crystallography*

Purified lyophilized protein was re-suspended in water to a concentration of 20 mg/ml and mixed to give 20 mg/ml of each complex. Crystals were grown by the hanging

drop method at room temperature by mixing 1 µl protein solution with 1 µl of reservoir solution. SYNZIP1:SYNZIP2 was grown in 45% MPD, 100 mM Tris pH 8.0, and 160 mM ammonium acetate. SYNZIP6:SYNZIP5 was grown in 100 mM Tris pH 8.2 and 20% MPD. Crystals were frozen in LN2 without addition of any cryoprotectant. Diffraction data were collected at 100K on a Rigaku MicroMax007-HF with VariMax-HR optics and a RAXIS-IV detector (SYNZIP1: SYNZIP2) or at the NE-CAT 24ID-E beam line of the Advanced Photon Source (SYNZIP6:SYNZIP5) and processed using HKL2000 (Otwinowski, et al. 1997). Both structures were solved by molecular replacement using PHASER (McCoy, et al. 2005). In each case the search model was derived from a single energy-minimized theoretical model selected from an ensemble of models spanning the space of parameters of native parallel dimeric coiled-coil structures. The ensemble was generated as previously described (Apgar, et al. 2008). The search models had no overhangs and the side chains at all non-interfacial positions (**b**, **c**, and **f**) were truncated to alanine. Model building was done using COOT (Emsley and Cowtan. 2004, Adams, et al. 2002) using twin law corrections for both structures (Table C.S4). Non-crystallographic symmetry (NCS) restraints between the four copies of the heterodimer in the asymmetric unit (ASU) of the SYNZIP6:SYNZIP5 crystals were used to aid in the refinement of that structure. Geometry was checked using MOLPROBITY (Davis, et al. 2007) and no outliers were identified (Table C.S4). Figures of structures were generated using PyMol (DeLano Scientific, Palo Alto, CA).

*Pull down assay*

Proteins containing a unique N-terminal cysteine were labeled by mixing 100 µM protein with 0.5 mM Alexa Fluor 488 or 546 maleimide (Invitrogen) or 2 mM maleimide-PEG11-biotin (Thermo Scientific) in 100 mM potassium phosphate pH 7.0/150 mM KCl/1 mM TCEP. Solutions were incubated for three hours at 18-22 °C. Free dye or biotin was removed using desalting spin columns (Thermo Scientific). Biotinylated proteins were concentrated using centrifugal filter units (Millipore). The concentration of unlabeled and biotinylated proteins was determined using the Edelhoch method. The concentration of dye labeled proteins was estimated by assuming a 50% recovery after desalting. Each dye labeled protein was mixed with the unlabeled version (at known concentration) in a 1:10 ratio. 400 pmoles of each protein indicated in Figure 4.4 were mixed in 75 µl binding buffer (12.5 mM potassium phosphate pH 7.4, 150 mM KCl, 1

mM DTT, 1% BSA, 0.1% Tween-20). Protein mixtures were incubated for 1 hour at 18-22 °C and then 50 µl of a 50% slurry of NeutrAvidin beads (Thermo Scientific) in binding buffer was added. Mixtures were incubated for 2 hours at 18-22 °C with rotation. Beads were then washed 3 times with 1 ml binding buffer at 4 °C and mixed with 100 µl of loading buffer (10 % glycerol, 2% SDS, 100 mM DTT, 0.01% bromophenol blue, 100 mM Tris pH 6.8). Following heating at 65 °C for 15 minutes, 10 µl of each sample was loaded onto an 18% Tris-glycine gel (Invitrogen). Gels were imaged on a Typhoon 9400 imager. Fluorsep software (Amersham Biosciences) was used to remove background fluorescent overlap.

*Sequence analysis*

Positions **a-g** in the coiled-coil heptad repeat were assigned manually, as designed previously (Grigoryan, et al. 2009b), based on conserved Leu residues and overall hydrophobic/polar patterning. Each peptide contains 5-7 full heptads. The following criteria were applied for sequence analysis. To predict the most probable alignments of coiled-coil dimers, all possible helix alignments that overlapped by at least 5 full heptads and did not contain an asparagine mismatch were considered. Asparagine mismatches were defined as an Asn residue at a non-terminal **a** position across from isoleucine, valine or leucine at a non-terminal **a** position. A terminal **a** position was defined as an **a** position ≤ 3 residues from the end of the coiled coil. For assessing **g-e'** electrostatics, the least repulsive alignment of ≥ 5 heptads that did not contain an asparagine mismatch was used. For this purpose, each attractive **g-e'** interaction was scored as + 0.5 and each repulsive **g-e'** interaction was scored as -0.5. Negatively charged glutamate and aspartate, and positively charged lysine and arginine were considered during scoring. Note that Glu, Lys, Arg and – to a lesser extent – Asp overwhelmingly predominate at **g** and **e** positions of the 26 peptides considered (Table C.S1).

## ACKNOWLEDGEMENTS

This work was supported by NIH award GM067681. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, supported by award RR-15301 from the National Center for Research Resources at the National Institute of Health. Use of the Advanced Photon Source is

supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02- We thank the MIT BioMicro center for arraying instrumentation, J.R. Apgar for generating models used for structure determination and G. Grigoryan for computational design of synthetic peptides not described elsewhere. We thank members of the Keating laboratory for comments on the manuscript.

## REFERENCES


1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006;45(38):11324-32.

2. Adams PD, Grosse-Kunstleve RW, Hung L, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: Building new software for automated crystallographic structure determination. Acta Crystallographica Section D. 2002;58(11):1948-54.

3. Apgar JR, Gutwin KN, Keating AE. Predicting helix orientation for coiled-coil dimers. Proteins: Structure, Function, and Bioinformatics. 2008;72(3):1048-65.

4. Arndt KM, Pelletier JN, Müller KM, Plückthun A, Alber T. Comparison of in vivo selection and rational design of heterodimeric coiled coils. Structure. 2002;10(9):1235-48.

5. Bashor CJ, Helman NC, Yan S, Lim WA. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. Science. 2008 March 14, 2008;319(5869):1539-43.

6. Bromley EHC, Sessions RB, Thomson AR, Woolfson DN. Designed alpha-helical tectons for constructing multicomponent synthetic biological systems. J Am Chem Soc. 2009;131(3):928-30.

7. Bromley EHC, Channon K, Moutevelis E, Woolfson DN. Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. ACS Chemical Biology. 2008;3(1):38-50.

8. Campbell KM, Sholders AJ, Lumb KJ. Contribution of buried lysine residues to the oligomerization specificity and stability of the fos coiled coil. Biochemistry. 2002 Apr 16;41(15):4866-71.

9. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, III AWB, Snoeyink J, Richardson JS, Richardson DC. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucl. Acids Res. 2007 July 13, 2007;35(suppl_2):W375-383.

10. Diehl MR, Zhang K, Lee HJ, Tirrell DA. Engineering cooperativity in biomotor-protein assemblies. Science. 2006 March 10, 2006;311(5766):1468-71.

11. Diss ML, Kennan AJ. Simultaneous directed assembly of three distinct heterodimeric coiled coils. Org Lett. 2008 Sep 4;10(17):3797-800.

12. Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS. Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. Cell. 1999;99(1):103-15.

13. Edelhoch H. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry. 1967 Jul;6(7):1948-54.

14. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallographica Section D. 2004;60(12 Part 1):2126-32.

15. Giesecke AV, Fang R, Joung JK. Synthetic protein-protein interaction domains created by shuffling Cys2His2 zinc-fingers. Mol Syst Biol. 2006;2:2006.0011.

16. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009a Apr 16;458(7240):859-64.

18. Grigoryan G, Keating AE. Structural specificity in coiled-coil interactions. Curr Opin Struct Biol. 2008 Aug;18(4):477-83.

19. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science. 1993 November 26, 1993;262(5138):1401-7.

20. Kammerer RA, Kostrewa D, Progias P, Honnappa S, Avila D, Lustig A, Winkler FK, Pieters J, Steinmetz MO. A conserved trimerization motif controls the topology of short coiled coils. Proceedings of the National Academy of Sciences of the United States of America. 2005 September 27, 2005;102(39):13891-6.

21. Lai JR, Fisk JD, Weisblum B, Gellman SH. Hydrophobic core repacking in a coiled-coil dimer via phage display: Insights into plasticity and specificity at a protein-protein interface. Journal of the American Chemical Society. 2004;126(34):10514-5.

22. Mapp AK, Ansari AZ, Ptashne M, Dervan PB. Activation of gene expression by small molecule transcription factors. Proceedings of the National Academy of Sciences of the United States of America. 2000 April 11, 2000;97(8):3930-5.

23. Mason JM, Muller KM, Arndt KM. Considerations in the design and optimization of coiled coil structures. Methods Mol Biol. 2007;352:35-70.

24. Mason JM, Schmitz MA, Muller KM, Arndt KM. Semirational design of jun-fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. Proc Natl Acad Sci U S A. 2006 Jun 13;103(24):8989-94.

25. McAllister KA, Zou H, Cochran FV, Bender GM, Senes A, Fry HC, Nanda V, Keenan PA, Lear JD, Saven JG, Therien MJ, Blasie JK, DeGrado WF. Using α-helical coiled-coils to design nanostructured metalloporphyrin arrays. Journal of the American Chemical Society. 2008;130(36):11921.

26. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. Acta Crystallographica Section D. 2005;61(4):458-64.

27. Moll JR, Ruvinov SB, Pastan I, Vinson C. Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10-15 M. Protein Sci. 2001 March 1, 2001;10(3):649-55.

28. Newman JR, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 Jun 27;300(5628):2097-101.

29. O'Shea EK, Lumb KJ, Kim PS. Peptide 'velcro': Design of a heterodimeric coiled coil. Curr Biol. 1993a Oct 1;3(10):658-67.

32. Otwinowski Z, Minor W, Carter CW,Jr. [20] Processing of X-ray diffraction data collected in oscillation mode In: Methods in Enzymology. Academic Press; 1997; p. 307-26.

33. Papapostolou D, Smith AM, Atkins EDT, Oliver SJ, Ryadnov MG, Serpell LC, Woolfson DN. Engineering nanoscale order into a designed protein fiber. Proceedings of the National Academy of Sciences. 2007 June 26, 2007;104(26):10853-8.

34. Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA. Reversible hydrogels from self-assembling artificial proteins. Science. 1998 July 17, 1998;281(5375):389-92.

35. Takagi J, Erickson HP, Springer TA. C-terminal opening mimics 'inside-out' activation of integrin alpha5beta1. Nat Struct Mol Biol. 2001;8(5):412-6.

36. Wernicke S, Rasche F. FANMOD: A tool for fast network motif detection. Bioinformatics. 2006 May 1, 2006;22(9):1152-3.

37. Wolfe SA, Grant RA, Pabo CO. Structure of a designed dimeric zinc finger protein bound to DNA. Biochemistry. 2003;42(46):13401-9.

# Chapter 5

# Conservation and rewiring of bZIP protein-protein interaction networks

A modified version of this chapter will be submitted for publication with Aaron W. Reinke, Judy Baek, Orr Ashenberg and Amy E. Keating as authors.

**Collaborator notes:**

Judy Baek cloned genes, purified proteins, and measured interactions. Orr Ashenberg developed ODE models for fitting binding curves to the data.

**ABSTRACT**

Molecular functions such as protein-protein interactions are often conserved throughout evolution, but it is unclear when and how frequently changes to interactions occur. Knowing how protein-protein interaction specificity evolves is important for understanding how changes in interactions can lead to changes in phenotype. To study the conservation of protein-protein interaction networks, bZIP transcription factor protein-protein interaction networks were measured for 5 metazoan and 2 single-cell species. The metazoan interaction networks displayed broadly similar interaction properties that were distinct from the single cell species. A core network of interactions was observed in the 5 metazoan species. This network was diversified in each species, both through rewiring of interactions between conserved proteins as well as the addition of new proteins and interactions. A cross-species interaction network including proteins from *C. intestinalis* and human revealed that several proteins have highly conserved specificity profiles, though for others, there are distinct changes in interactions. Minor sequence changes were identified that could exert major changes on interaction profiles. These results indicate that the bZIP interaction domain is flexible in its ability to evolve and rewire interactions.

## INTRODUCTION

Molecular changes drive phenotypic diversity throughout evolution. Differences in transcriptional regulation have been shown to be major contributors to developmental and cellular outcomes (Carroll. 2008). While much emphasis has been placed on changes to *cis* regulatory elements, it is unclear what impact mutations in transcription factors can have on gene regulation. Several recent studies suggest that changes to the molecular function of both protein-DNA and protein-protein interactions are more common than previously assumed. For example, the DNA-binding specificity of some transcription factors has been demonstrated to diverge extensively, coevolving with *cis* regulatory elements (Kuo, et al. 2010, Baker, et al. 2011). Different alleles of the human zinc finger protein PRDM9, which is involved in specifying hotspots in meiotic recombination, have different DNA binding specificities (Baudat, et al. 2009). An interaction between the transcription factors HoxA11 and Foxo1a evolved regulatory changes outside the interaction interface (Brayer, et al. 2011). Mutations to a phosphorylated regulatory site of the bZIP transcription factor CEBPB are responsible for changing the protein from a repressor to an activator upon phosphorylation (Lynch, et al. 2011). Some orthologs of human and *Caenorhabditis elegans* PDZ domains also show differences in binding specificity (Tonikian, et al. 2008). These studies suggest that biochemical functions of orthologous proteins are not necessarily conserved (Dickinson, et al. 2011).

The basic-leucine zipper (bZIP) proteins are a large class of transcription factors present in most eukaryotes. These proteins can form both homodimers and heterodimers, and the complex that forms influences the DNA sites that can be bound. The bZIP proteins provide an ideal system to study the evolution of interaction specificity. Fourteen bZIP families are conserved as far back as cnidarians, and bZIPs form a closed interaction network, allowing

potential partners to be identified by sequence (Amoutzias, et al. 2007). Interactions for the human network have been previously reported and models have been developed that can predict interactions with good, but limited accuracy (Grigoryan and Keating. 2006, Newman and Keating. 2003).

Though families of bZIP proteins are conserved throughout metazoan evolution, it is unclear if their interactions are. To address this question, we report quantitative *in vitro* measurements of bZIP protein-protein interaction networks from 7 species. These networks reveal that while a conserved set of interactions exists, extensive expansion and rewiring of the bZIP protein-protein interaction network has occurred, especially in humans compared to simpler metazoans.

**RESULTS**

*Measurement of bZIP protein-protein interactions*

We measured the bZIP protein-protein interaction networks of seven species. The species were selected based on evolutionary position and their status as established or emerging model organisms. They include 5 metazoan species: human, sea squirt (*Ciona intestinalis*), fruit fly (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), and sea anemone (*Nematostella vectensis*). Also included were two single-cell organisms, choanoflagellates (*Monosiga brevicollis*), the closest sister group of metazoans, and the yeast (*Saccharomyces cerevisiae*). There are 21 bZIP families in humans. 18 of these families are conserved in *C. intestinalis*. 14 of them occur in the last common ancestor of human and sea anemone. There are an additional 4 families that arose through duplication after the divergence of sea anemone. Both *M. brevicollis* and *S. cerevisiae* have only a few of the 14 metazoan ancestral families, indicating the majority of the metazoan ancestral families appeared at the emergence of metazoans. Each species also

has a number of novel families that are not conserved with any of the other species examined (Figure 5.1A, Figure 5.2, and Table 5.1).

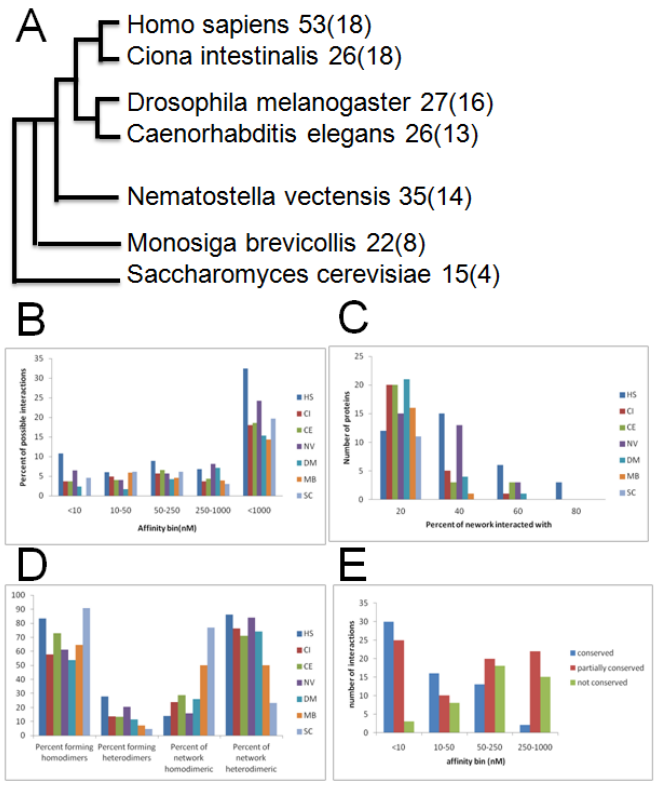



Figure 5.1. Characteristics of bZIP protein-protein interaction networks from 7 species.
**A,** Evolutionary tree of studied species. After each species name the number of bZIPs in that species is given, with the number of families in parentheses. **B**-**D**, Species abbreviations are as follows**:** HS, Human; CI, *C. intestinalis*; DM, *D. melanogaster;* CE, *C. elegans;* NV, N. *vectensis*; MB, *M. brevicollis***;** and SC, *S. cerevisiae*. **B**, Percentage of possible interactions observed with different affinities in each network. **C**, Histogram of the connectedness of each network. **D**, Frequency of heterodimeric vs**.** homodimeric interactions in each network. **E**, Relationship between conservation and interaction affinity.

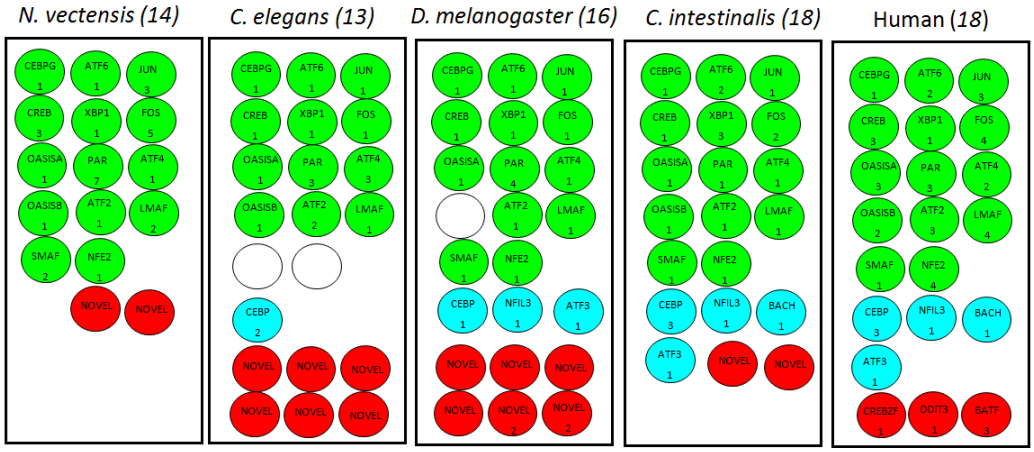



Figure 5.2. The bZIP family repertoire of each species.
The number of family in each species is indicated in parentheses. Circles represent bZIP families with the name of each family given along with the number of family members. Green circles, ancestral families. Blue circles, families conserved in at least 2 species. Red circles, novel families.

Interactions between bZIP proteins were quantified *in vitro* using a solution-based FRET assay. Each protein was expressed, purified and labeled with a small molecule fluorophore. Two versions of each bZIP were generated, either with an acceptor or donor fluorophore. Acceptor-labeled proteins were titrated at ~1 nM to 1 μM into 10 nM donor-labeled protein. Binding curves were measured at 21 °C and equilibrium dissociation constants were determined (see Methods). In humans there are 53 bZIPs, many of which are highly similar. 36 were selected that represent all families and cover most of the human bZIP sequence diversity. For the remaining species all possible pairwise interactions between bZIPs were measured. Each heteromeric interaction was measured twice, as each donor-labeled protein was measured against each acceptor-labeled protein; mostly similar affinities were observed for both measurements. The data was also highly reproducible **(**Figures 5.3), and the data for human proteins compared well to a previous array study (Newman and Keating. 2003).

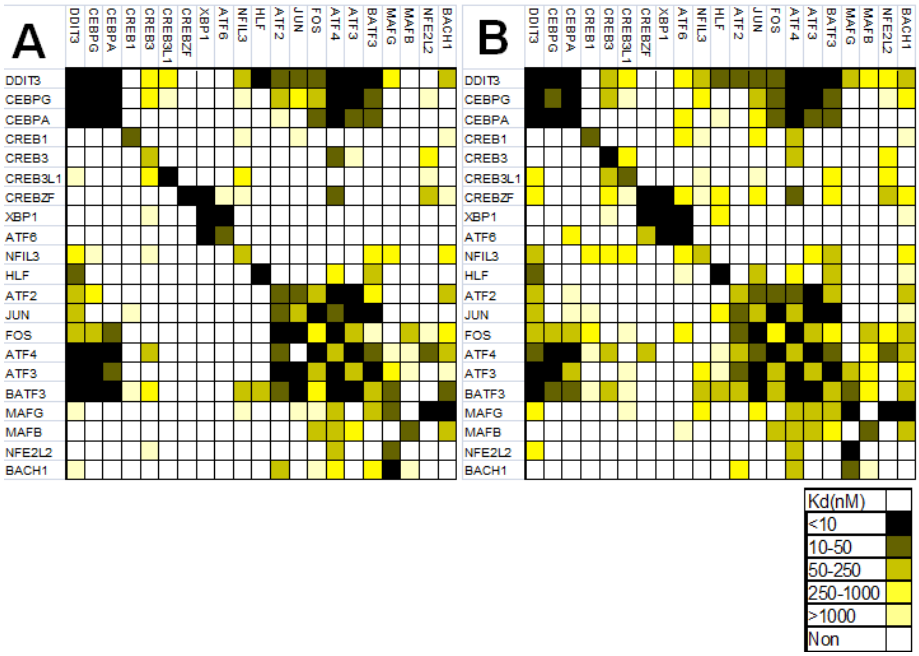



Figure 5.3. Reproducibility of measured bZIP interactions. Data are presented as a heat map with the strength of interaction indicated by the scale at the bottom. **A, B**, Two independent measurements of interactions among 21 human proteins.

*Properties of bZIP interaction networks.*

Interactions in each species were observed over a range of affinities, with the human network the densest (Figure 5.1B, Figures 5.4-10, and Table 5.2). The human network also had the most highly connected proteins, with choanoflagellates and yeast having the least connected networks (Figure 5.1C). The majority of proteins in each network were capable of forming homodimers and the majority of possible heterodimers were not observed. However, the number of possible heterodimers in each network is much greater than the number of possible homodimers. Thus, for the 5 metazoan species the interaction networks are composed of mostly heterodimers. Interestingly, yeast shows the opposite composition, with the majority of the network being composed of homodimers. The choanoflagellates network is composed of an approximately equal number of homodimers and heterodimers (Figure 5.1D).

For further analysis interactions were compared at the family level (see Methods). There are three categories of interactions; those that are always conserved in all species, those that interact in some organisms but not others, and those that occur in only one organism and thus are not conserved. Interestingly, the majority of conserved interactions were of high affinity and the majority of non-conserved interactions were of weaker affinity (Figure 5.1E). We compared how conserved each metazoan network was with each other network, and the overlap of interactions ranged from ~25% to ~75% (Figure 5.11).

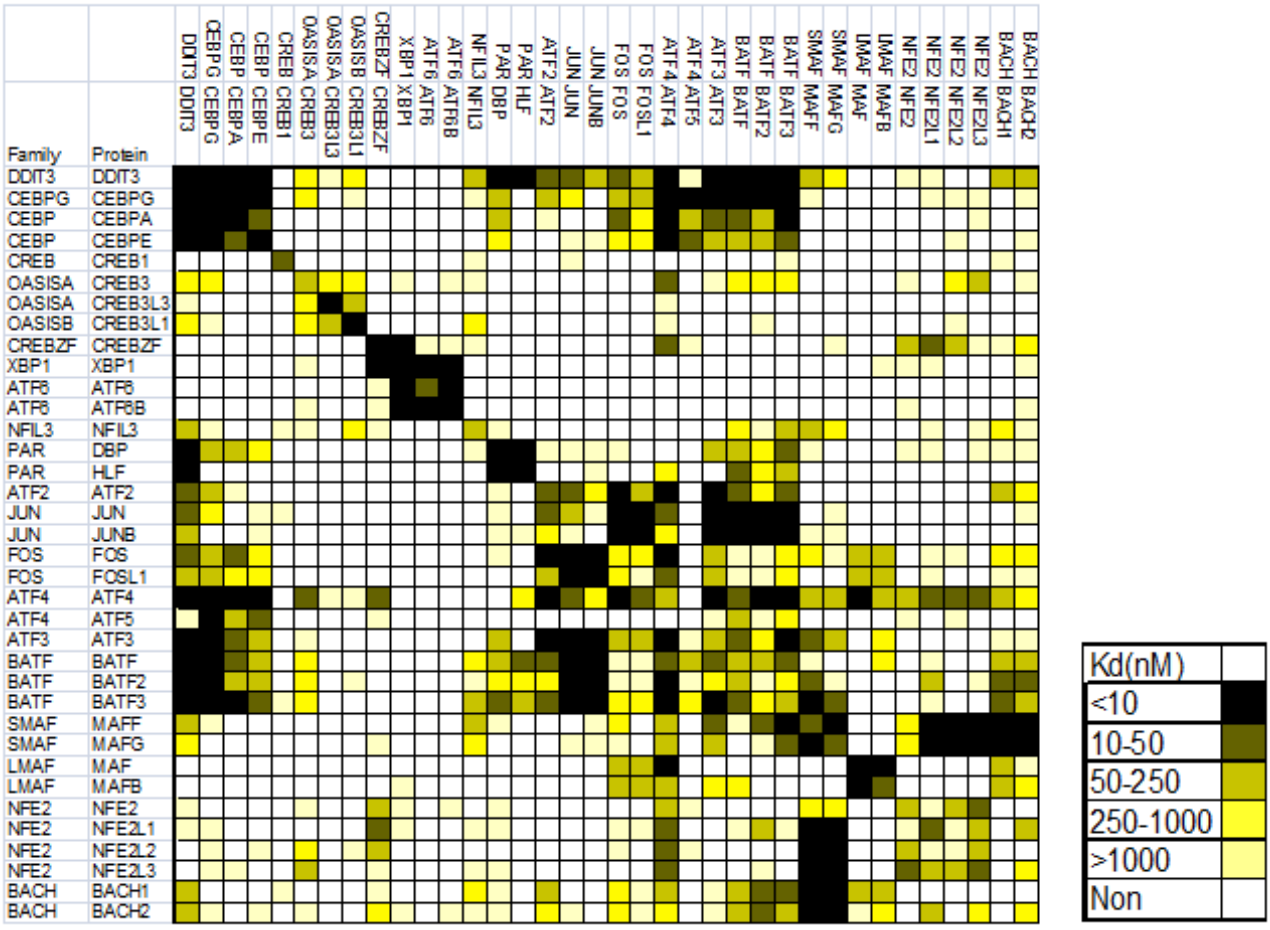



Figure 5.4. Human bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

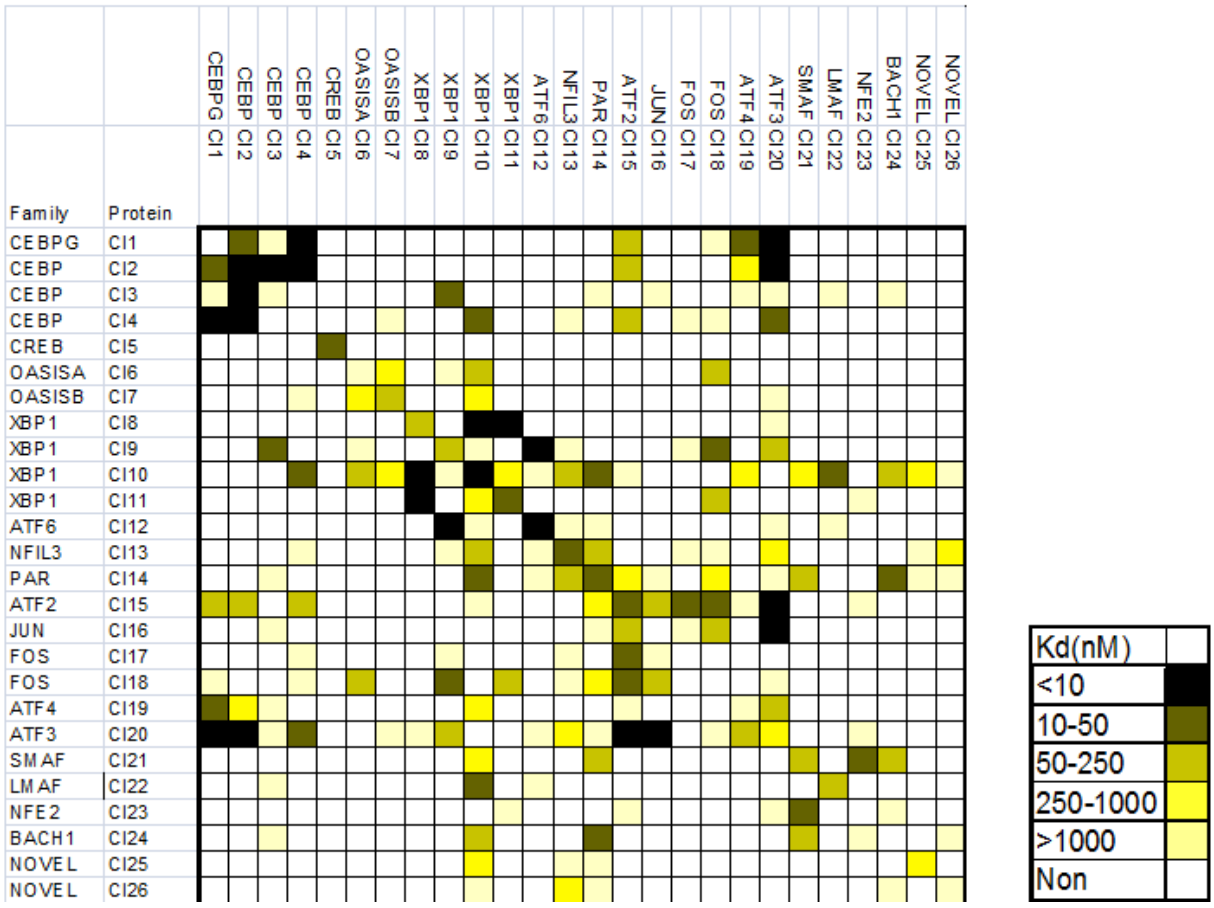



Figure 5.5. *C. intestinalis* bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

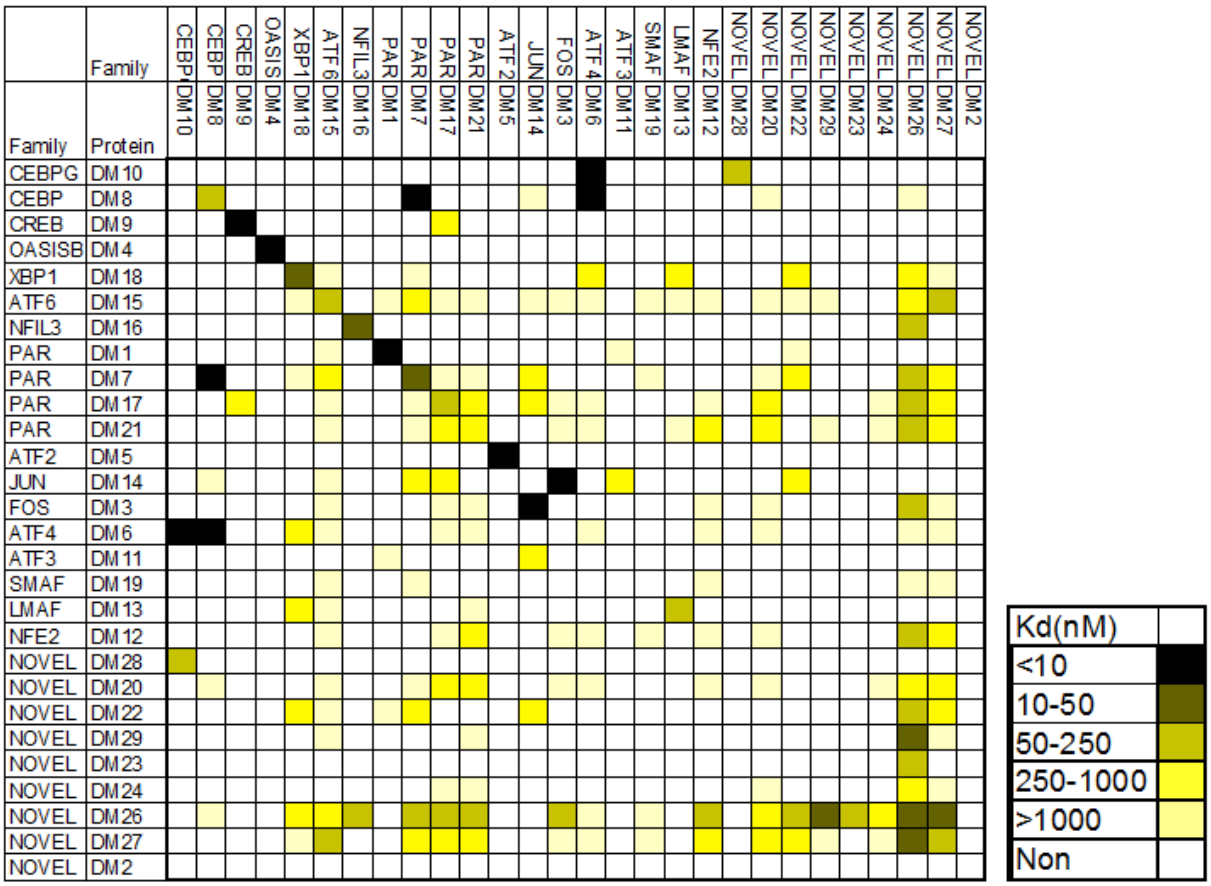



Figure 5.6. *D. melanogaster* bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

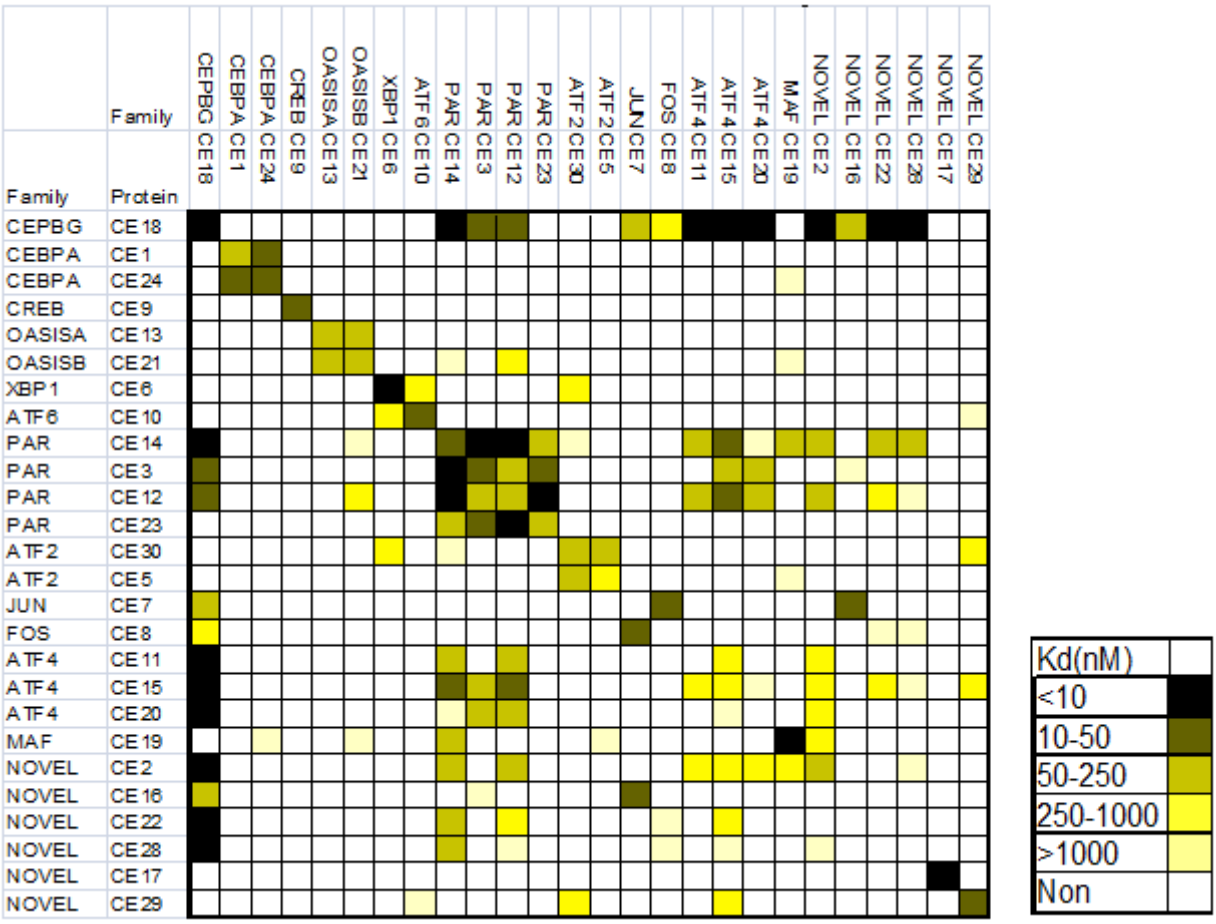



Figure 5.7. *C. elegans* bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

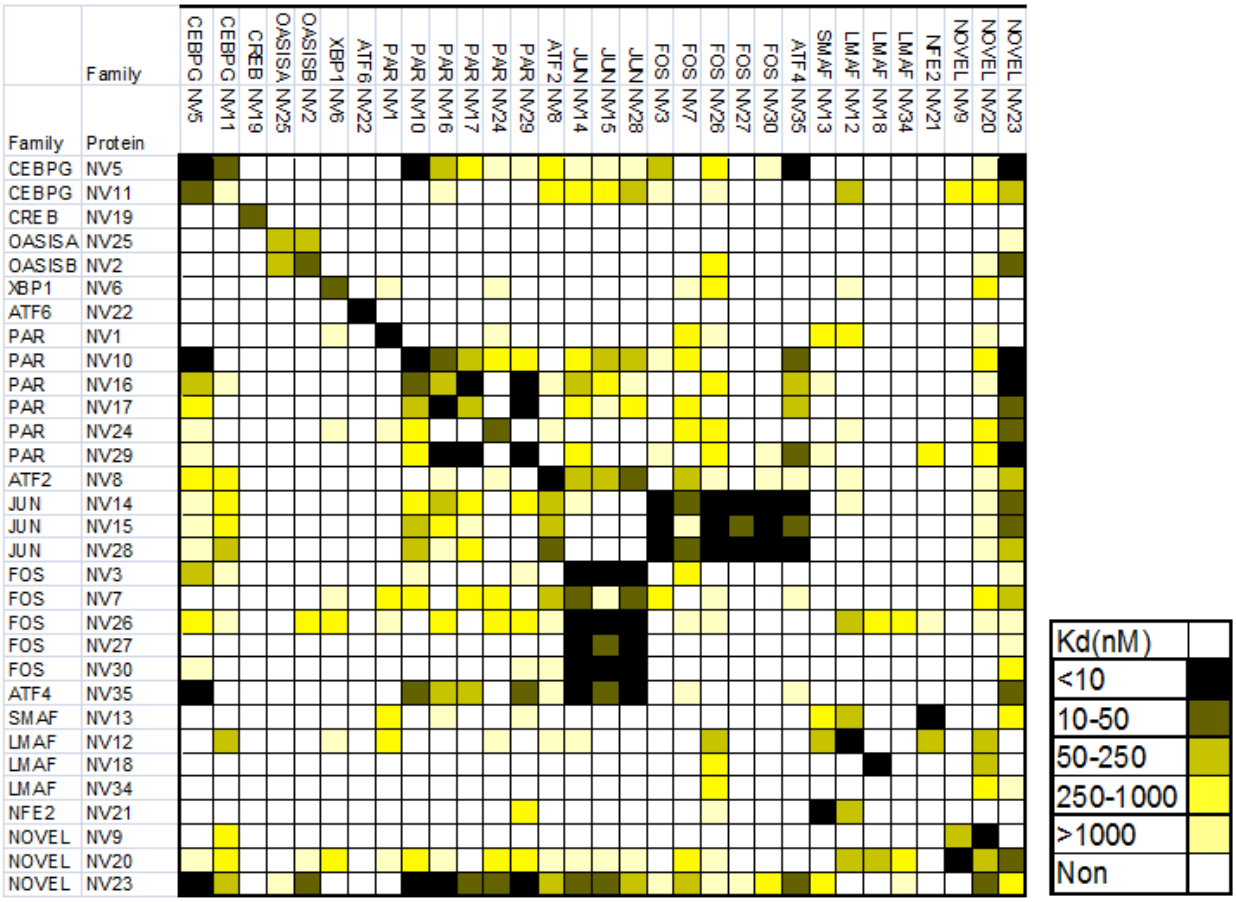



Figure 5.8. N. *vectensis* bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

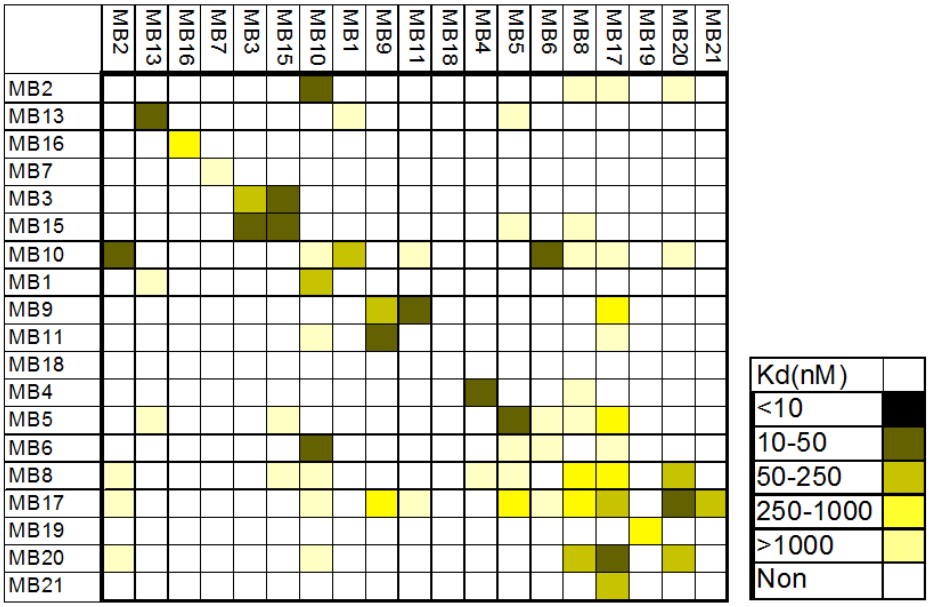



Figure 5.9. *Monosiga brevicollis* bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

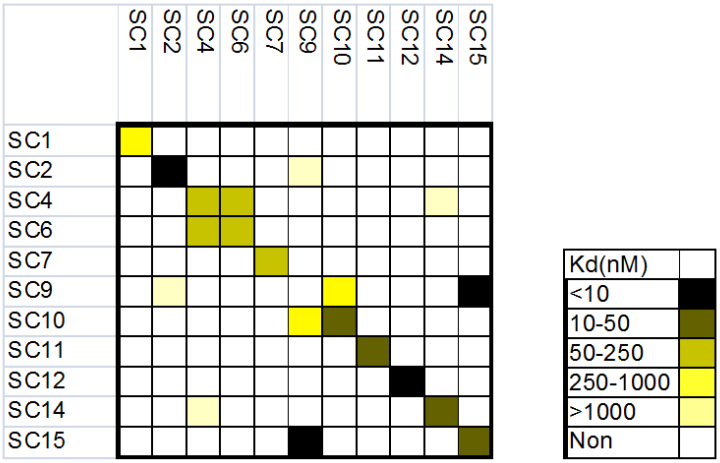



Figure 5.10. *S. cerevisiae* bZIP interaction network.
Data are presented as a heat map with the strength of interactions indicated by the scale at the right.

| | HS | CI | DM | CE |
|---|---|---|---|---|
| CI | **35** | | | |
| DM | **25** | **30** | | |
| CE | **42** | **34** | **53** | |
| NV | **53** | **42** | **43** | **72** |

Figure 5.11. Comparison of interaction networks between species.
The percentage overlap of interactions between each pair of species (see methods).

*Conservation and rewiring of bZIP interaction networks*

Our data indicate that the extant bZIP interactions networks are the result of both rewiring interactions among a set of ancestral families as well as the addition of new bZIP families. To compare how metazoan interactions changed over time, we used considerations of parsimony to infer a bZIP interaction network for the last common ancestor of metazoans (See Methods). This network is composed of the 14 ancestral metazoan families and contains 10 homodimeric and 9 heterodimeric interactions (Figure 5.12A). Compared to this ancestral network, several gains and losses occurred in N. *vectensis* and *C. elegans*, and a large number of interactions were lost in *D. melanogaster* (Figure 5.12B-D). In the higher species, human and *C. intestinalis*, a much larger number of changes occurred, and many new interactions were introduced. Many of the gains of interactions were observed with the four-member XBP1 family in *C. intestinalis* and the two-member ATF4 family in human (Figure 5.12E, F). The 4 families that arose from duplication of ancestral families (CEBPG-CEBP, PAR-NFIL3, FOS-ATF3, and NFE2-BACH) also led to diversification of the networks by adding additional partners and interactions. These duplicated families often maintained many of the same interactions, but also changed to add additional partners (Figure 5.13). Finally, novel families arose that interact with many of the conserved families (Figure 5.14). Taken together, rewiring of interactions among

ancestral proteins, the addition of conserved duplicated families, and the introduction of novel families has allowed each species to evolve a highly distinct bZIP interaction network (Figure 5.15).

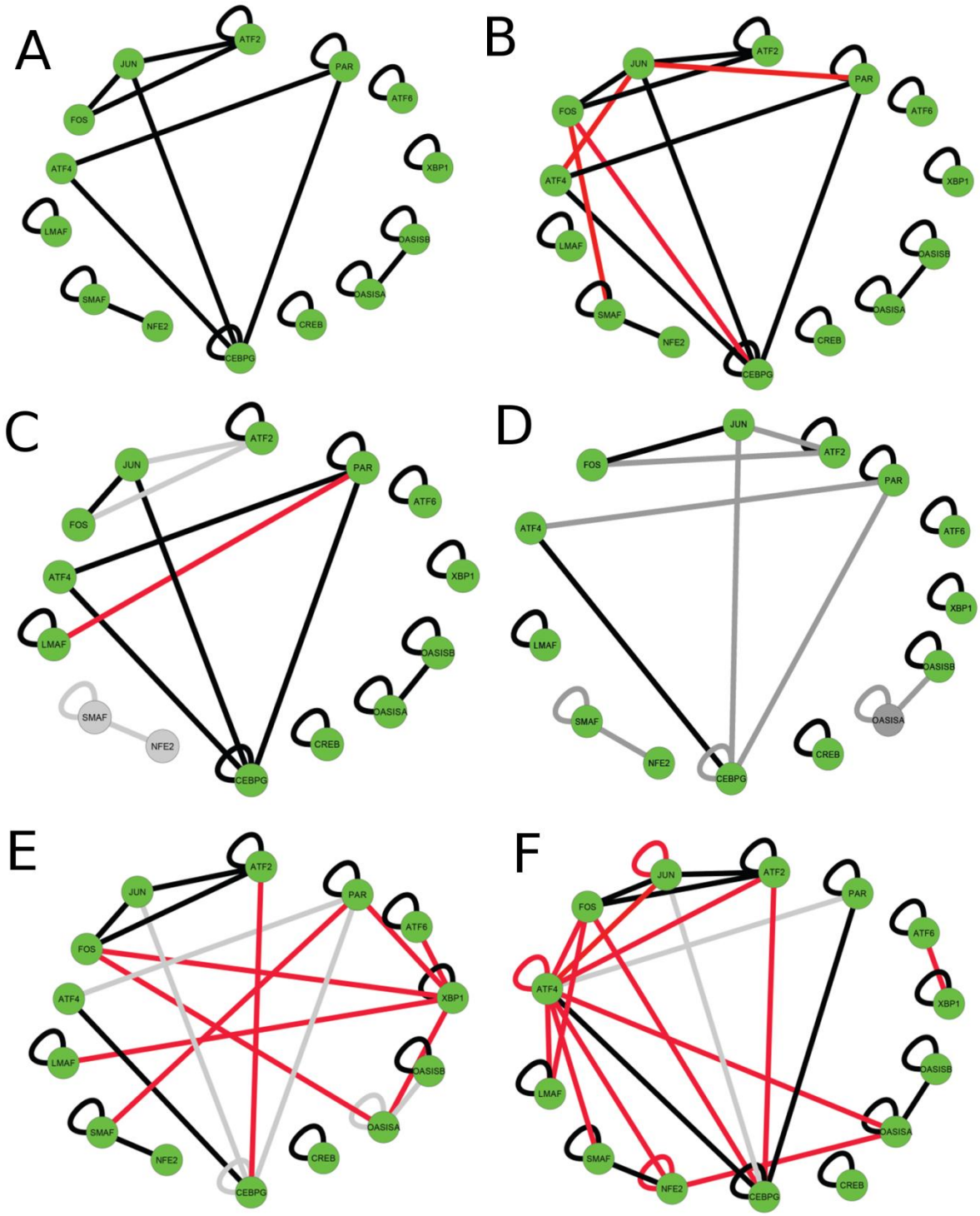



Figure 5.12. Rewiring of metazoan bZIP interactions networks.
**A-F,** Interactions involving proteins in the 14 ancestral families. Green circles, extant families. Grey circles, lost families. Black lines, inferred ancestral interactions. Red lines, gained interactions. Grey lines, lost interactions. **A,** Inferred ancestral network **B,** N. *vectensis*. **C,** *C. elegans*. **D,** *D. melanogaster*. **E,** *C. intestinalis*. **F,** Human. Graphs created using Cytoscape (http://www.cytoscape.org/).

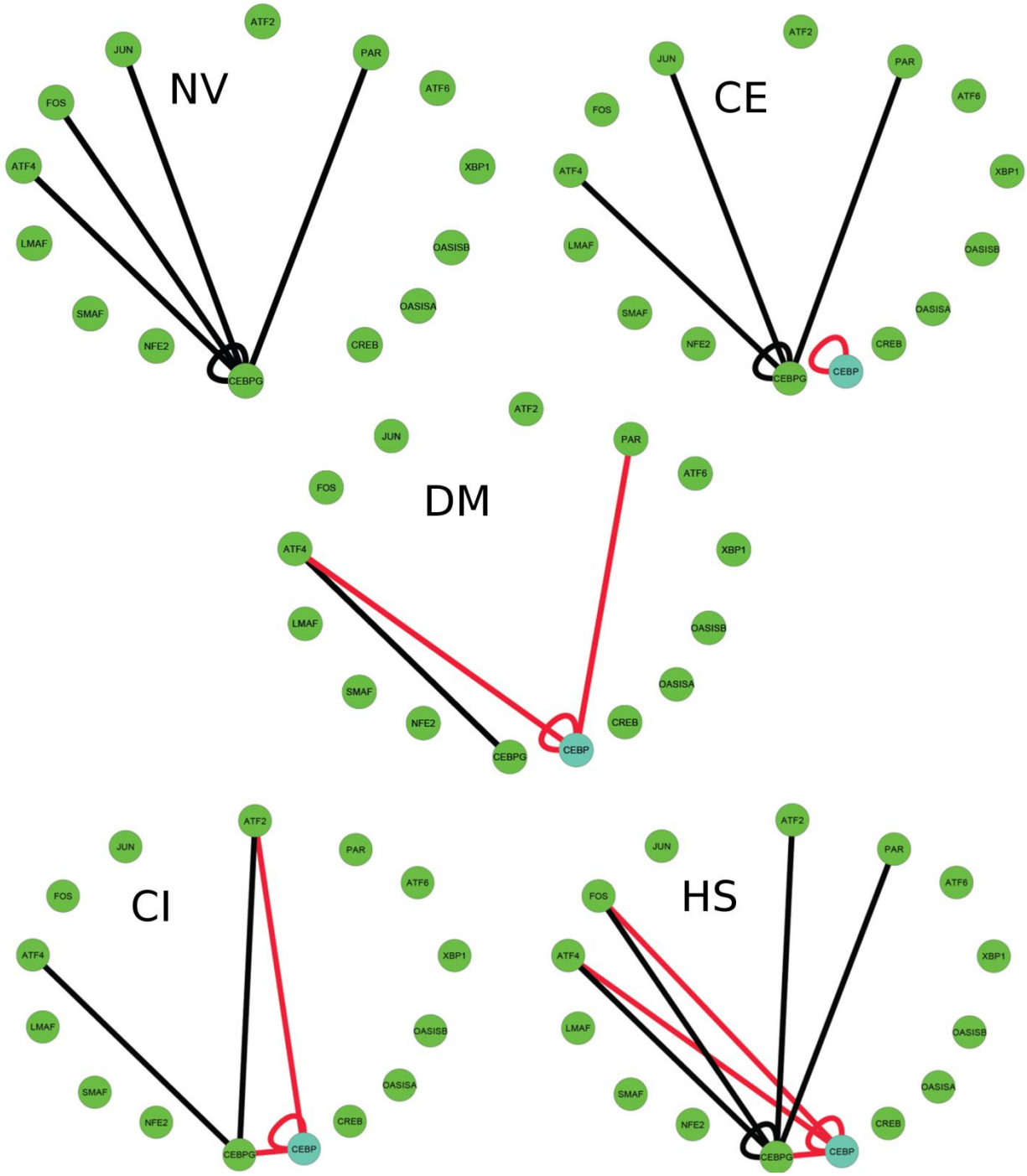



Figure 5.13. Interactions of CEBPG and CEBP families following the CEBPG-CEBP duplication.

Green circles represent ancestral families, blue circles show CEBP. Black lines are interactions with CEBPG and red lines are interactions with CEBP. Species abbreviations are the same as in Figure 5.1. Graphs created using Cytoscape (http://www.cytoscape.org/).

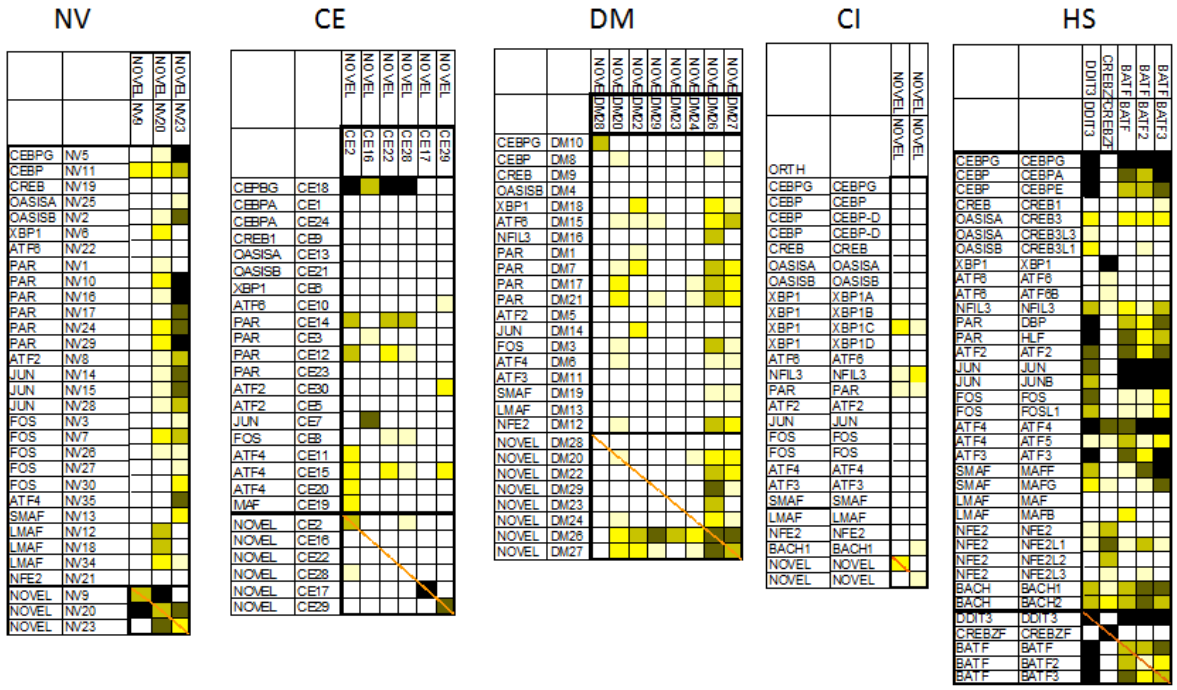



Figure 5.14. Interactions of novel bZIP families show extensive connections to conserved families.
Data are presented as in Figure 5.3.

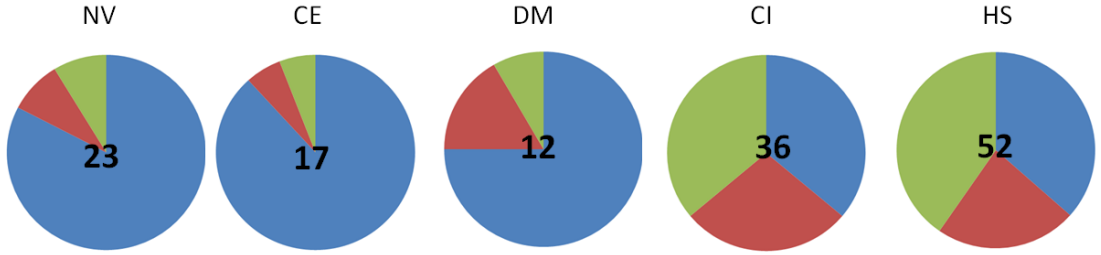



Figure 5.15. Origins of interactions in extant bZIP interaction networks.
Blue, ancestral interactions. Red other conserved or partially conserved interactions. Green, new interactions. The total number of interactions that occur in each species between conserved families is in the middle of each chart. Species abbreviations are the same as in Figure 5.1

*Evolution of bZIP interaction profiles*

When an interaction is gained or lost, it difficult to know which interaction partner was responsible for the change. To pinpoint the mechanism of how interactions change between orthologs, proteins must be profiled against a common set of partners. Towards this end, we measured 32 human bZIPs against 24 *C. intestinalis* bZIPs. The resulting interspecies interaction network revealed 5 families with highly similar interaction specificity profiles in each species

(Figure 5.16). The remaining families showed differences in specificity to varying degrees. These data allow identification of which partners change their specificity. For example, ATF4 from human interacts with ATF2 from both species, but *C. intestinalis* ATF4 doesn't interact with either ATF2. This indicates that *C. intestinalis* ATF2 is competent to interact with ATF4, but there are changes in the *C. intestinalis* version of ATF4 that prevent the interaction from occurring (Figure 5.17). For roughly half the cases where differences of interaction occur between the two species, there are changes in both partners, and for the rest there are changes in just one partner (Figure 5.16). This suggests that there is flexibility in bZIPs to evolve their interactions, by adding new partners while maintaining existing ones.

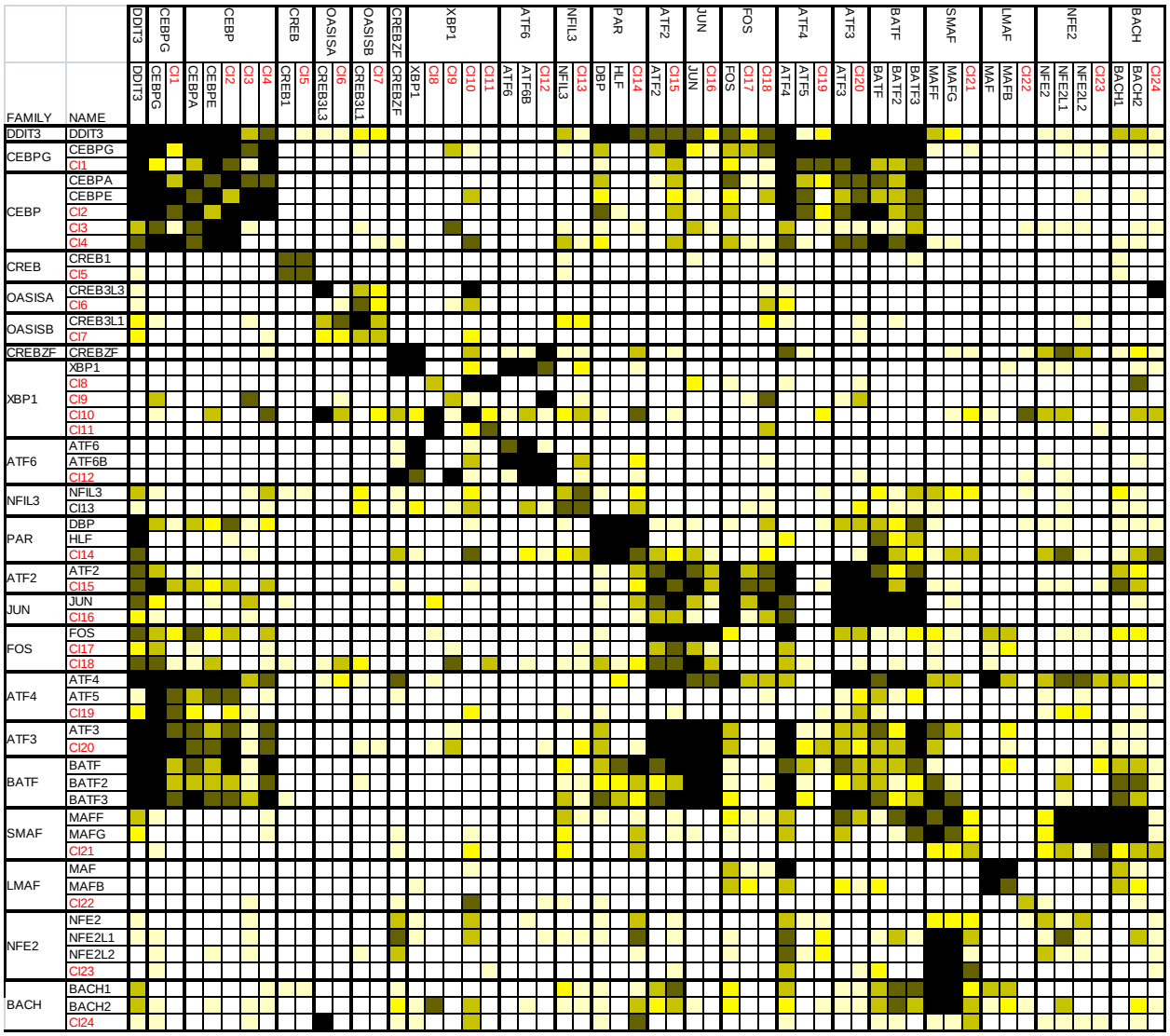


Figure 5.16. *C. intestinalis* and Human interspecies bZIP interaction network.
32 Human bZIPs measured against 24 *C. intestinalis* bZIPs. Data presented as in Figure 5. 3. Human proteins are in black and *C. intestinalis* are proteins in red.

Of particular interest is the ATF4 family, where a large number of interactions occur in human but not in *C. intestinalis*. In humans there are two ATF4 family proteins ATF4 and ATF5; ATF4 has a very promiscuous interaction profile while ATF5 is much more specific. The *C. intestinalis* ATF4 is similar in interaction specificity to ATF5 but not ATF4 (Figure 5.17). We also measured ATF4 from sea anemone and zebra fish (*Danio rerio*) against the human proteins.

Zebra fish ATF4 has similar specificity to human, indicating that the change happened before the last common ancestor of human and zebra fish. Sea anemone also is more similar in interaction specificity to human ATF4 than to *C. intestinalis*, though many of the strong interactions are with protein families that don't occur in N. *vectensis.* As a result N. *vectensis* ATF4 has more interactions with human and *C. intestinalis* proteins than it has in its own species (Figure 5.17).

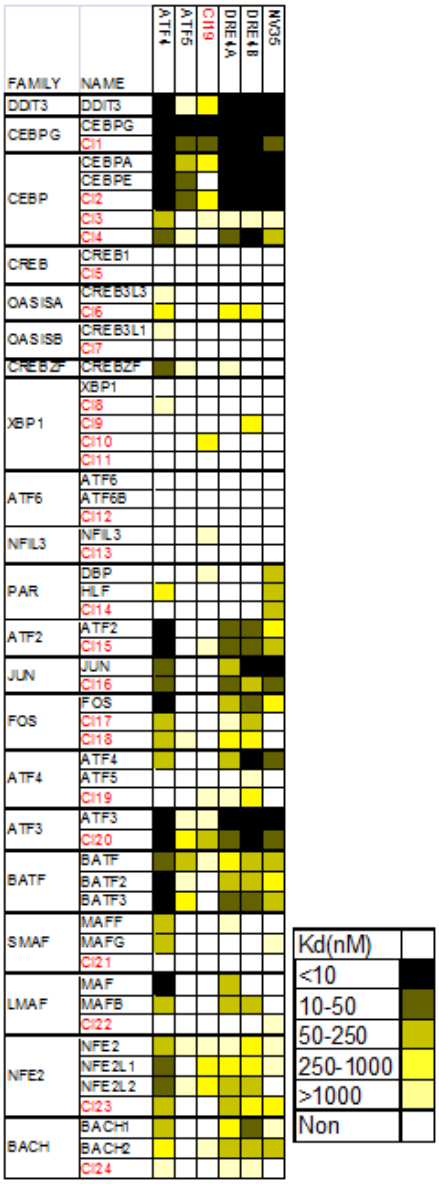



Figure 5.17 ATF4 family interaction specificity.
ATF4 proteins from human, *C. intestinalis, D. rerio, and N. vectensis* were measured against proteins from human and *C. intestinalis.* Data are presented as a heat map with the strength of interaction indicated by the scale at the right. Human proteins are in black and *C. intestinalis* are proteins in red.

Negative selection has been proposed by Lim and coworkers as an important means of ensuring specificity in interaction networks (Zarrinpar, et al. 2003). They observed for SH3 ligand Pbs2 that intraspecies interactions were more specific while interspecies interactions were more promiscuous. We do not observe the same trend in our data. Although there are interactions in the human-*C. intestinalis* interspecies network that do not occur in either of the intraspecies networks, the total number of interactions is less than in either intraspecies network (Figure 5.18). This could indicate that negative design is not a prominent force in shaping bZIP interaction networks. Alternatively, negative design that reduces promiscuity in intraspecies networks could work globally to reduce non-specific interactions.

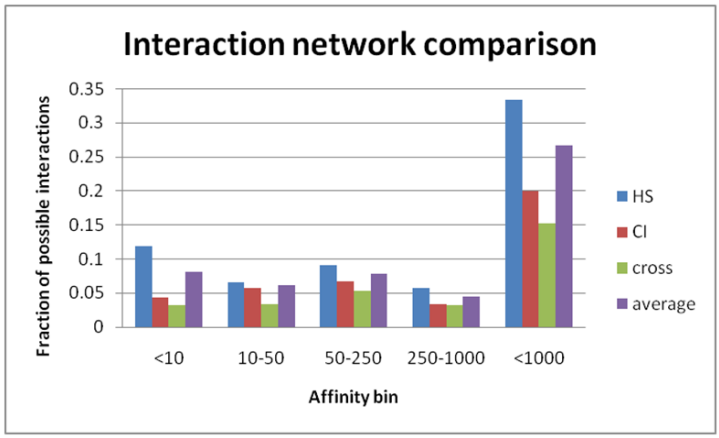



Figure 5.18. Characteristics of the Human, *C. intestinalis*, and interspecies interaction networks. Fraction of possible interactions of different affinities in each species' network. Blue, Human. Red, *C. intestinalis.* Green, Human- *C. intestinalis* cross-species network. Purple, average of Human and *C. intestinalis* interactions.

The relationship between sequence and interactions for bZIP paralogs is complex. There are instances where small numbers of sequence changes lead to large differences in interaction specificity, and conversely cases where large numbers of sequence changes do not significantly alter interaction specificity. There is at best only a very weak correlation between sequence identity and the conservation of an interaction (Figure 5.19). For orthologs, sequence conservation >80% did correlate with higher conservation, but any trend at lower sequence identity was very weak (Figure 5.20). Based on what is known about determinants of coiled-coil interaction specificity (Vinson, et al. 2006), we investigated the detailed origins of certain specificity changes.

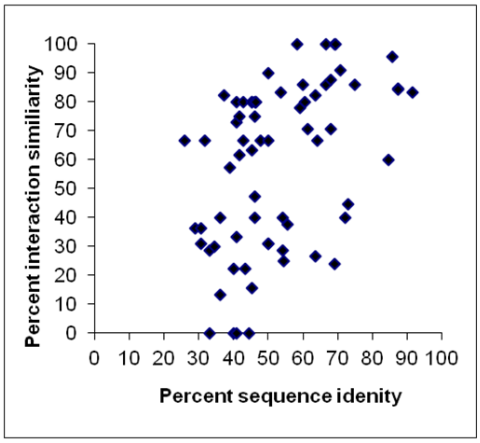



Figure 5.19. Sequence identity at the coiled-coil interface vs. interaction similarity of paralogs. Each point is a comparison between paralogs in the same species (see Methods).

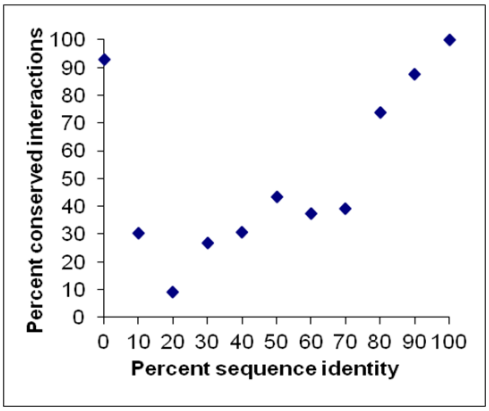



Figure 5.20. Sequence identity at the coiled-coil interface vs. interaction similarity of orthologs. Each point is the average percentage of conserved interactions in each sequence identity bin (see Methods).

The PAR family in *C. elegans* contains several family members with different interaction specificities (Figure 5.13). One member of the family, Y51H4A.4 contains an asparagine at an **a** position where the others do not (Figure 5.21A). Asparagines at **a** positions have been shown to be highly destabilizing when positioned across the interface from hydrophobic amino acids, but not when pairs with asparagines (Acharya, et al. 2006). We mutated the asparagine in Y51H4A.4 to alanine, which is the residue found at the same position in ces-2; we also made the reverse alanine-to-asparagine mutation in the ces-2 protein. These changes led to a switch in specificity (Figure 5.21B).

A similar result was observed for PAR family proteins in *D. melanogaster* (Figure 5.22). A second mechanism proposed to destabilize interactions is packing multiple amino acids branched at the beta carbon (e.g. valine or isoleucine) into the core of the coiled-coil interface (Grigoryan, et al. 2009). In humans, the ATF4 family member ATF5 has two consecutive **d** position valines, which are leucines in ATF4. Both paralogs also have an isoleucine at the following **a** position (Figure 5.21A). To test whether these differences contribute to ATF5 having a much more specific interaction profile than ATF4, we mutated the valines to leucines in ATF5, and made the reverse mutations in ATF4. This conferred an ATF-4 interaction profile on ATF5, and the ATF4 mutant also became very ATF5 like (Figure 5.21C). Mutations were also tested to switch specificity between orthologs of human and *C. intestinalis*. These either only subtly changed interaction specificity or led to changes of specificity in one of the orthologs, but not the other (Figure 5.23). Overall, these examples highlight the plasticity of the bZIP interactome, which can be rewired with only modest sequence changes.

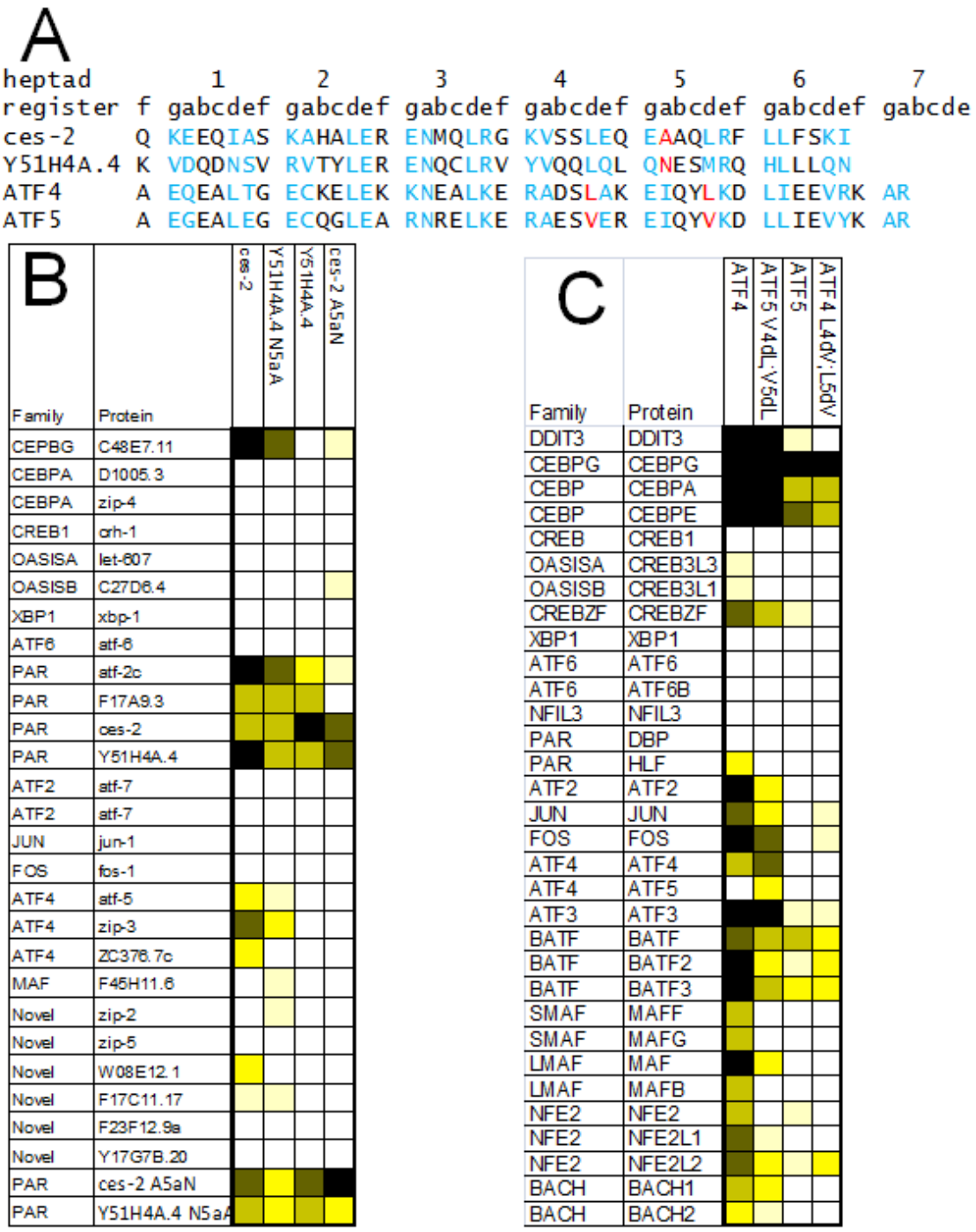



Figure 5.21. Switching interaction profiles between bZIP paralogs.

**A,** Sequences of PAR family proteins in *C. elegans* (top) and ATF4 family proteins in human (bottom). Interface positions are in blue and mutated residues are in red. **B, C,** Heat maps of interaction data, plotted in the same way as in Figure 5.3. Columns one and three are the wild-type proteins. Column two is the mutant version of column three, and column four is the mutant version of column one. Mutants are named by wild-type residue at the heptad number and position followed by the mutant residue. **B,** PAR family mutants in *C. elegans* **C,** ATF4 family mutants in human.

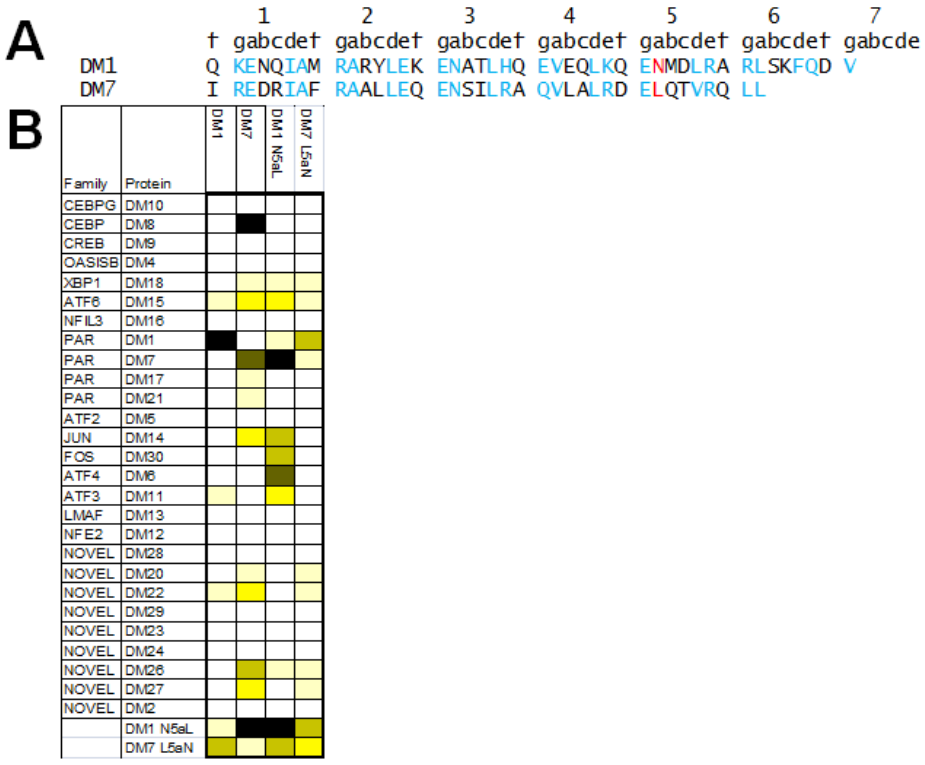



Figure 5.22. PAR family mutants in *D. melanogaster*.
Data presented as in Figure 5.21. **A,** Sequences of PAR family proteins in *D. melanogaster* **B,** PAR family mutants in *D. melanogaster*. Data are presented as a heat map.

**A**

| heptad | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| register | f | gabcdef | gabcdef | gabcdef | gabcdef | gabcdef | gabcdef | gabcde |
| CREB3L3 | E | YIDGLET | RMSACTA | QNQELQR | KVLHLEK | QNLSLLE | QLKKLQA | IVV |
| CI6 | D | YIGGLEA | RVTKCTN | LNQALSQ | RVKQLEQ | QNFTLLE | QLKQVHD | AVK |
| ATF6B | E | YLQGLEA | RLQAVLA | DNQQLRR | ENAALRR | RLEALLA | ENSELKL | |
| CI12 | E | YVSTLEQ | QMLECLD | DNNKLRS | MNQQLRD | KVMELEN | ENTRLR | |
| FOS | E | LTDTLQA | ETDQLED | EKSALQT | EIANLLK | EKEKLEF | ILAAHR | |
| CI17 | E | LTDRLQG | ETDHLED | HQSILHQ | EIMSLQQ | EKEHLEF | LLAAHS | |
| DBP | L | KENQISV | RAAFLEK | ENALLRQ | EVVAVRQ | ELSHYRA | VLSRYQA | QH |
| HLF | L | KENQIAI | RASFLEK | ENSALRQ | EVADLRK | ELGKCKN | ILAKYEA | RH |
| CI14 | I | KENQIAM | RANFLEK | ENESLKM | EVADLRS | ELKRVMN | TLRVYEK | EII |
| BACH1 | D | CIQNLES | EIEKLQS | EKESLLK | ERDHILS | TLGETKQ | NLTGLCQ | KV |
| CI24 | D | CIRSLQC | QLEQLRE | EHLNLMG | ERRTCQD | KSLKLAE | MFQKRYE | QVF |

B

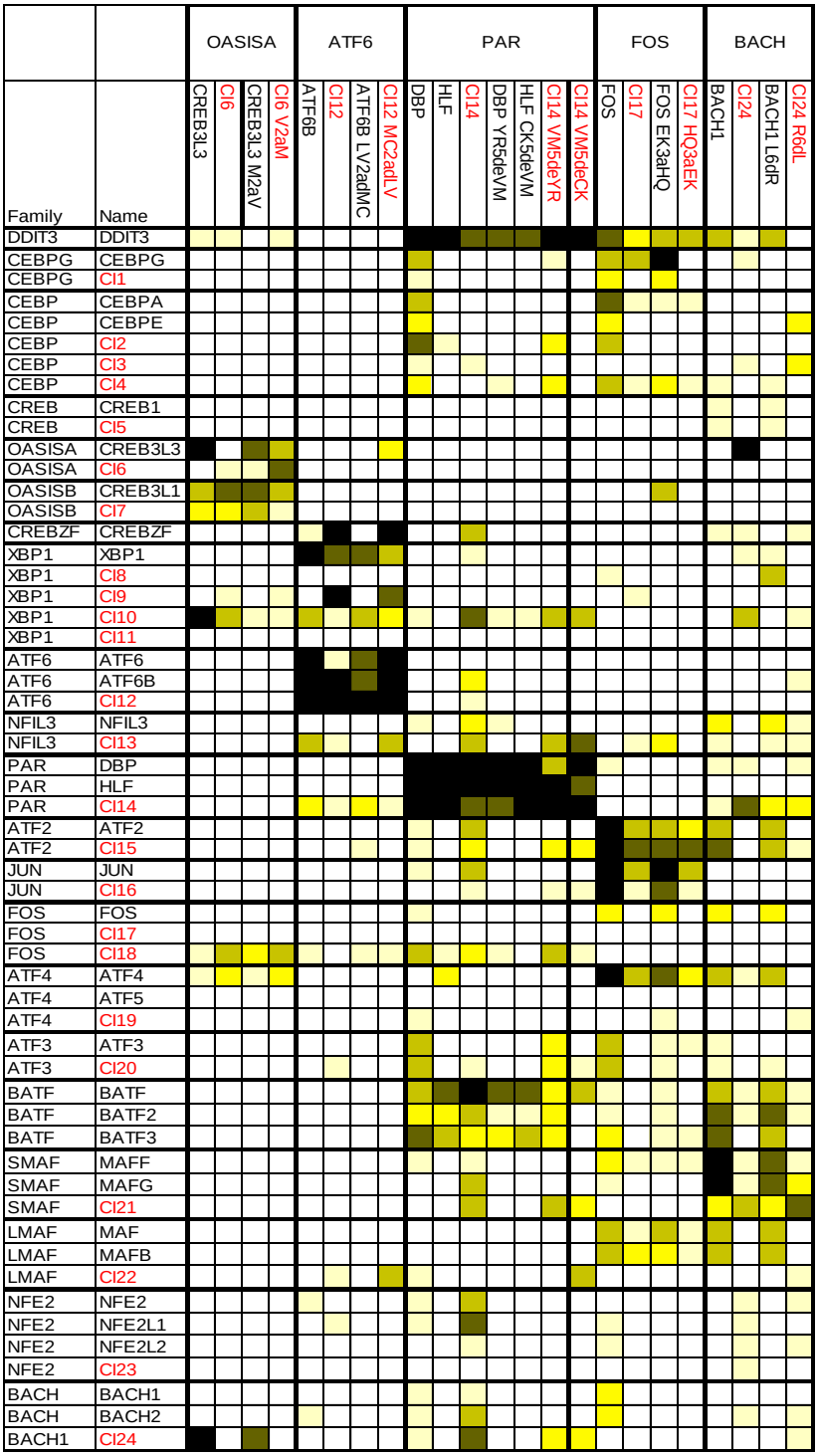



Figure 5.23. Mutants of Human and *C. intestinalis* orthologs.
Data presented as in Figure 5.21. Human proteins are in black and *C. intestinalis* proteins are in red. **A,** Sequences of Human and *C. intestinalis* orthologs. **B,** Mutants of Human and *C. intestinalis* orthologs. Data are presented as a heat map.

## DISCUSSION

The biochemical measurements in this study uncover interactions that do not necessarily occur *in vivo.* Further, interactions were assayed in the absence of DNA (preferred binding sites for most bZIP pairs are not yet known), and the presence of DNA could stabilize certain complexes not observed in our assay. Nonetheless, bZIP interactions that are conserved between species are likely to be functionally relevant. Indeed, we have discovered a core set of interactions that are conserved throughout metazoan evolution, and likely are involved in essential processes. Additionally, because those interactions that are conserved are of higher affinity, this suggests that the higher affinity interactions are also likely to be functionally important. The converse argument that lack of conservation probably implies functional irrelevance does not hold, however. One example is the interactions of JUN-FOS, JUN-ATF2, and JUN-JUN. The JUN-FOS interaction is always conserved. JUN-ATF2 is not observed to interact in *C. elegans* or *D. melanogaster* but does interact in the other three species. JUN-JUN only interacts in human. They have different extents of conservation yet have all been reported to be functionally important in humans (van Dam and Castellazzi. 2001).

A striking result is the number of interactions that change between networks. The bZIP interaction interface allows for both drastically changing interaction profiles with small number of changes, as well as being able to add or lose a small number of interactions while keeping many interactions constant. This has been observed previously in designed bZIP interactions (Grigoryan, et al. 2009), and some of the molecular mechanism that make this possible are understood. Here we were able to use these principles to rationally alter the interaction profiles of paralogous or orthologous proteins to make them more similar.

Our data can be used to inform research in several areas going forward. First, there is considerable interest in using interactions measured in one species to annotate other organisms

(Yu, et al. 2004). Our results show some relationship between sequence identities and interactions at high sequence identities as previously observed, but also suggest that annotation based on homology alone for bZIPs is a poor indicator of which interactions occur (Yu, et al. 2004). This suggests a cautious and conservative approach to in inter-species interaction transfer. Second, our data provide a very large number of quantitative interactions that can be used to test and improve models for predicting bZIP interactions from sequence. Such improved models could potentially be used to predict interactions in other species. Improved insight into more general methods for modeling protein interaction specificity could also arise from computational studies using these data. Third, this data can potentially be used predict which interactions occur in various cell types using comprehensive expression data (Ravasi, et al. 2010). Finally, this work provides a resource and starting to point to investigate the potential functional consequences of rewiring of the bZIP interaction networks.

## METHODS

### *bZIP identification.*

Proteins containing bZIP domains were identified by searching with custom made HMM models built using the program HMMER (Eddy. 1998). Initial models were constructed using 53 human bZIP domains. Additional versions of the model were generated using bZIPs identified from other species. Genomes of each species were searched using multiple HMM models and putative bZIP domains were manually inspected for the following features: highly conserved basic residues, spacing between basic region and leucine zipper, and predominantly hydrophobic coiled-coil core. Sequences were aligned using the previously described features. The N-terminal domain boundary of each bZIP was defined as 10 amino acids beyond the end of the minimal basic region (defined as the N-terminal end of the GCN4:DNA co-crystal structure) (Ellenberger,

et al. 1992). For three *N. vectensis* proteins where the bZIP domain is at the extreme N-terminus, the native N-terminal was used as the boundary instead. C-terminal boundaries were determined by manual inspection for polar amino acids in core positions, glycines or prolines, or the native C-terminal end of the protein. To determine ortholog assignments of each bZIP, phylogenetic trees of bZIP domains built using the neighbor joining method were generated as described previously (Grigoryan, et al. 2009) and reciprocal best hit databases were also used (Waterhouse, et al. 2011, Ostlund, et al. 2010, Powell, et al. 2012, Chen, et al. 2006). In the few cases where ortholog assignment was ambiguous, interaction profile similarities were used. bZIP family names are consistent with (Amoutzias, et al. 2007).

*Cloning, expression, purification, and labeling*

*C. elegans* bZIP domains were cloned from cDNA. *D. rerio* ATF4 genes were cloned by gene synthesis using the program DNAWorks to design primers, which were annealed using a two-step PCR method (Hoover and Lubkowski. 2002). These genes were cloned as intein-chitin binding domain fusions using a modified pTXB1 (NEB) plasmid. Genes were cloned into the plasmid using the SLIC method (Li and Elledge. 2007) or restriction digested with XhoI and NsiI. The remaining genes were ordered synthesized from GENEWIZ. All clones were sequence verified. Proteins were expressed in RP3098 cells by growing 1 L LB cultures at 37 °C to $OD_{600}$ 0.4-0.8 at which point expression was induced with 0.5 mM IPTG. Cultures were then incubated for 3-4 hours and cells pelleted. For poorly expressing proteins an alternative protocol of induction at 18 °C for 12-16 hours was used. Cells were resuspended in buffer (20 mM HEPES pH 8.0, 500 mM NaCl, 2 mM EDTA, 1 M guanidine-HCl, 0.2 mM PMSF, and 0.1% Trition X-100) and lysed using sonication. The lysate was then split and each half was pored over a column containing 1-ml chitin beads (NEB). The column was washed and then equilibrated with EPL

buffer (50 mM HEPES pH 8.0, 500 mM NaCl, 200 mM MESNA, 1 M guanidine-HCl). To cleave the intein and label the proteins with a fluorescent dye, the columns were incubated with EPL buffer containing 1mg/ml of cysteine-lysine-dye where dye is either fluorescein or TAMRA (CELTEK). Columns were capped and incubated for at least 16 hours. Cleaved and labeled proteins were then eluted and diluted 5-fold into denaturing buffer (6 M guanidine-HCL, 5 mM imidazole, 0.5 M NaCl, 20 mM TRIS ,1 mM DTT, pH 7.9). This solution was then flowed over columns containing 1ml Ni-NTA resin. After washing, proteins were eluted with 60% ACN/0.1% TFA. Labeled proteins were lyophilized, resuspended, and desalted using spin-columns (Bio-Rad). Proteins were stored in 10 mM potassium phosphate pH 4.5 at -80 °C. Peptide concentrations were measured in 6 M guanidine-HCl/100 mM sodium phosphate pH 7.4 using the absorbance of the dye with an extinction coefficient of 68,000 $M^{-1}$ $cm^{-1}$ at 499 nm for fluorescein and 86,000 $M^{-1}$ $cm^{-1}$ at 560 nm for TAMRA. To determine the accuracy of protein concentration determination using dye absorbance, amino acid analysis was performed (UC Davis proteomics core facility). Three fluorescein-labeled and three TAMRA-labeled proteins were analyzed and all were less than 15% from the correct concentration. Molecular weights of fluorescein-labeled *C. elegans* proteins were measured by mass spectrometry and were correct within 0.15% and no evidence of unlabeled proteins was observed. Care was taken during the purification process to protect the labeled proteins from photo damage.

*Interaction measurements.*

Fluorescein-labeled proteins were diluted to 80 nM in 1 mM TCEP in low protein binding tubes (Eppendorf). Dilutions of fluorescein-labeled proteins were then transferred to an entire row of a black 96-well non-binding surface plate (Corning). TAMRA-labeled proteins were diluted to 2.67 µM in 1 mM TCEP and 60 µl of each protein was transferred to a well in the

first two columns of a black 384-well non-binding surface plate (Corning). The remaining wells were filled with 30 µl of 1 mM TCEP. Each TAMRA labeled protein was serially diluted by aspirating 30 µl of protein and mixing it in a well containing 30 µl of 1 mM TCEP. Ten 2-fold dilutions of each protein were done, resulting in 11 concentrations of each TAMRA protein and a well containing no TAMRA-labeled protein. 10 µl of each donor was then transferred from the 96-well plate to the 384-well plate and mixed. 40 µl of 2X binding buffer (100 mM potassium phosphate pH 7.4, 300 mM KCl, 0.2% BSA, 0.2% Tween-20) was then added to each well and mixed. All binding reactions were set up using a Tecan Freedom EVO liquid handling robot, except for replicate experiments which were done using a multichannel pipette. Plates were then incubated for 60-120 minutes at 37 °C. Following incubation, plates were read using a fluorescence plate reader (Molecular Devices) with excitation at 480 nm and emission at 525 nm. Plates were then transferred to 21 °C and incubated for 60-90 minutes and measured again. Plates were then transferred to 4 °C and incubated for 60-90 minutes and measured again.

*Fitting equilibrium disassociation constants.*

Data were fit to the saturation binding equation $F_{obs}=F_{max}-(( F_{max} - F_{min})/(2*[donor]))*((Kd+[donor]+[acceptor])-((Kd+x+[donor])\text{^}2-4*[acceptor]*.[donor])\text{^}0.5)$ where $F_{obs}$ is the observed fluorescence of the donor and $F_{max}$ and $F_{min}$ are the maximum and minimum fluorescence intensity (Kohler and Schepartz. 2001). Using an ODE model that accounted for homodimerization of donor and acceptor proteins gives similar, but improved results (Ashenberg, et al. 2011). Curves were required to have a change of at least 20% between $F_{max}$ and $F_{min.}$ Reported values are between 1 and 5000 nM. Interactions <1 nM or greater than 5000 nM were identified but could not be accurately quantified. For each heterodimer two measurements were made and the lower value was used as the value for the interaction of the

heterodimer, because this was judged to be the value least affected by not accounting for competing homodimerization. $K_d$ values are reported in Table 5.2.

*Interaction data analysis*

To determine conservation, interactions were analyzed at the family level. Two families were considered to interact if at least one member of the family interacted with at least one member of the other family and this interaction was tighter than 1000 nM. Interactions that occurred in all species where both families were present were considered conserved, those that occurred in at least 2, but not all, were considered partially conserved, and those that only occurred in one species were considered not conserved.

Overlap of networks between species was determined by counting the number of interactions between protein families that are shared by each pair of species. The number of shared interactions for each pair of species was divided by the sum of all the interactions that occurred in that pair of species.

The ancestral bZIP interaction network was inferred using parsimony. To be included in the ancestral interaction network, an interaction had to occur in *N. vectensis* and in at least one of the lower metazoans (*C. elegans* and *D. melanogaster)* or both chordates (*C. intestinalis* and Human).

To determine the relationship between sequence identity and interaction properties for paralogs, all possible pairs of paralogs from each species were used. The percent sequence identity was calculated from the interface positions (**adeg).** The percent interaction similarity was calculated using (shared interactions*2)/(interaction differences + shared interactions*2).

To determine sequence identity vs. interaction similarity of orthologs, all possible interolog pairs were compared between species. The percent sequence identity was calculated from the interface positions (**adeg)** using the combined sequences of each pair of orthologs.

**AKNOWLEDGEMENTS**

We thank the MIT BioMicro center for use of TECAN liquid handling robot, J.J. Sims for advice and guidance with intein-protein labeling chemistry, L. Reich for analytical models for curve fitting, and A.M. Safer for the *C. elegans* used for making cDNA.

## REFERENCES


1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006;45(38):11324-32.

2. Amoutzias GD, Veron AS, Weiner J,3rd, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL. One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. Mol Biol Evol. 2007 Mar;24(3):827-35.

3. Ashenberg O, Rozen-Gagnon K, Laub MT, Keating AE. Determinants of homodimerization specificity in histidine kinases. J Mol Biol. 2011 Oct 14;413(1):222-35.

4. Baker CR, Tuch BB, Johnson AD. Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc Natl Acad Sci U S A. 2011 May 3;108(18):7493-8.

5. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2009 Feb 12;327(5967):836-40.

6. Brayer KJ, Lynch VJ, Wagner GP. Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proc Natl Acad Sci U S A. 2011 Aug 9;108(32):E414-20.

7. Carroll SB. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. Cell. 2008 Jul 11;134(1):25-36.

8. Chen F, Mackey AJ, Stoeckert CJ,Jr, Roos DS. OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D363-8.

9. Dickinson DJ, Weis WI, Nelson WJ. Protein evolution in cell and tissue development: Going beyond sequence and transcriptional analysis. Dev Cell. 2011 Jul 19;21(1):32-4.

10. Eddy SR. Profile hidden markov models. Bioinformatics. 1998;14(9):755-63.

11. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted [alpha] helices: Crystal structure of the protein-DNA complex. Cell. 1992;71(7):1223-37.

12. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009;458(7240):859-64.

13. Grigoryan G, Keating AE. Structure-based prediction of bZIP partnering specificity. Journal of Molecular Biology. 2006;355(5):1125-42.

14. Hoover DM, Lubkowski J. DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. Nucl. Acids Res. 2002 May 15, 2002;30(10):e43.

15. Kohler JJ, Schepartz A. Kinetic studies of fos.jun.DNA complex formation: DNA binding prior to dimerization. Biochemistry. 2001 Jan 9;40(1):130-42.

16. Kuo D, Licon K, Bandyopadhyay S, Chuang R, Luo C, Catalana J, Ravasi T, Tan K, Ideker T. Coevolution within a transcriptional network by compensatory *trans* and *cis* mutations. Genome Res. 2010 Dec;20(12):1672-8.

17. Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods. 2007 Mar;4(3):251-6.

18. Lynch VJ, May G, Wagner GP. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature. 2011 Nov 13;480(7377):383-6.

19. Newman JRS, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 June 27, 2003;300(5628):2097-101.

20. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer EL. InParanoid 7: New algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 2010 Jan;38(Database issue):D196-203.

21. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, Jensen LJ, von Mering C, Bork P. eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. 2012 Jan 1;40(D1):D284-9.

22. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, Hashimoto T, Hide W, Hofmann O, Kamburov A, Kaur M, Kawaji H, Kubosaki A, Lassmann T, van Nimwegen E, MacPherson CR, Ogawa C, Radovanovic A, Schwartz A, Teasdale RD, Tegner J, Lenhard B, Teichmann SA, Arakawa T, Ninomiya N, Murakami K, Tagami M, Fukuda S, Imamura K, Kai C, Ishihara R, Kitazume Y, Kawai J, Hume DA, Ideker T, Hayashizaki Y. An atlas of combinatorial transcriptional regulation in mouse and man. Cell. 2010 Mar 5;140(5):744-52.

23. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS. A specificity map for the PDZ domain family. PLoS Biol. 2008 Sep 30;6(9):e239.

24. van Dam H, Castellazzi M. Distinct roles of jun : Fos and jun : ATF dimers in oncogenesis. Oncogene. 2001 Apr 30;20(19):2453-64.

25. Vinson C, Acharya A, Taparowsky EJ. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2006;1759(1-2):4-12.

26. Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. OrthoDB: The hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 2011 Jan;39(Database issue):D283-8.

27. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M. Annotation transfer between genomes: Protein-protein interologs and protein-DNA regulogs. Genome Res. 2004 Jun;14(6):1107-18.

28. Zarrinpar A, Park S, Lim WA. Optimization of specificity in a cellular protein interaction network by negative selection. Nature. 2003;426(6967):676-80.

## TABLES

Table 5.1: List of bZIP sequences used in this study

| Species | Family | Name | shorthand | Protein sequence |
|---|---|---|---|---|
| Human | DDIT3 | DDIT3 | | LAQEEEEEDQGRTRKRKQSG HSPARAGKQRMKEKEQENER KVAQLAEENERLKQEIERLTR EVEATRRALIDRMVNLHQA |
| Human | CEBPG | CEBPG | | KKSSPMDRNSDEYRQRRERN NMAVKKSRLKSKQKAQDTL QRVNQLKEENERLEAKIKLLT KELSVLKDLFLEHAHNLA |
| Human | CEBP | CEBPA | | KAKKSVDKNSNEYRVRRERN NIAVRKSRDKAKQRNVETQQ KVLELTSDNDRLRKRVEQLS RELDTLRGIFRQL |
| Human | CEBP | CEBPE | | KGKKAVNKDSLEYRLRRERN NIAVRKSRDKAKRRILETQQK VLEYMAENERLRSRVEQLTQ ELDTLRNLFRQI |
| Human | CREB | CREB1 | | LPTQPAEEAARKREVRLMKN REAARECRRKKKEYVKCLEN RVAVLENQNKTLIEELKALK DLYCHKSD |
| Human | OASISA | CREB3 | | LPLTKTEEQILKRVRRKIRNK RSAQESRRKKKVYVGGLESR VLKYTAQNMELQNKVQLLE EQNLSLLDQLRKLQAMVIEIS |
| Human | OASISA | CREB3L3 | | LPLTKAEEKALKRVRRKIKN KISAQESRRKKKEYVECLEK KVETFTSENNELWKKVETLE NANRTLLQQLQKLQTLVT |
| Human | OASISB | CREB3L1 | | LPTQLPLTKYEERVLKKIRRK IRNKQSAQESRKKKKEYIDGL ETRMSACTAQNQELQRKVLH LEKQNLSLLEQLKKLQAIVV |

| | | | | |
|---|---|---|---|---|
| Human | CREBZF | CREBZF | | GGGSGNDNNQAATKSPRKAAAAARLNRLKKKEYVMGLESRVRGLAAENQELRAENRELGKRVQALQEESRYLRAVLA |
| Human | XBP1 | XBP1 | | RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLREKTHGLVVENQELRQRL |
| Human | ATF6 | ATF6 | | MRNVGSDIAVLRRQQRMIKNRESACQSRKKKKEYMLGLEARLKAALSENEQLKKENGTLKRQLDEVVSENQRLKV |
| Human | ATF6 | ATF6B | | SCPPEVDAKLLKRQQRMIKNRESACQSRRKKKEYLQGLEARLQAVLADNQQLRRENAALRRRLEALLAENSELKL |
| Human | NFIL3 | NFIL3 | | REFIPDEKKDAMYWEKRRKNNEAAKRSREKRRLNDLVLENKLIALGEENATLKAELLSLKLKFGLISSTAY |
| Human | PAR | DBP | | KIQVPEEQKDEKYWSRRYKNNEAAKRSRDARRLKENQISVRAAFLEKENALLRQEVVAVRQELSHYRAVLSRYQAQH |
| Human | PAR | HLF | | KVFIPDDLKDDKYWARRRKNNMAAKRSRDARRLKENQIAIRASFLEKENSALRQEVADLRKELGKCKNILAKYEARH |
| Human | ATF2 | ATF2 | | RRRAANEDPDEKRRKFLERNRAAASRCRQKRKVWVQSLEKKAEDLSSLNGQLQSEVTLLRNEVAQLKQLLLAHKD |
| Human | JUN | JUN | | SPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHV |
| Human | JUN | JUNB | | SPINMEDQERIKVERKRLRNRLAATKCRKRKLERIARLEDKVKTLKAENAGLSSTAGLLREQVAQLKQKVMTHV |

| | | | | |
|---|---|---|---|---|
| Human | FOS | FOS | | KVEQLSPEEEEKRRIRRERNK MAAAKCRNRRRELTDTLQAE TDQLEDEKSALQTEIANLLKE KEKLEFILAAHR |
| Human | FOS | FOSL1 | | PCEQISPEEEERRRVRRERNK LAAAKCRNRRKELTDFLQAE TDKLEDEKSGLQREIEELQKQ KERLELVLEAHR |
| Human | ATF4 | ATF4 | | VAAKVKGEKLDKKLKKMEQ NKTAATRYRQKKRAEQEALT GECKELEKKNEALKERADSL AKEIQYLKDLIEEVRKAR |
| Human | ATF4 | ATF5 | | PYPHPATTRGDRKQKKRDQN KSAALRYRQRKRAEGEALEG ECQGLEARNRELKERAESVE REIQYVKDLLIEVYKAR |
| Human | ATF3 | ATF3 | | TKAEVAPEEDERKKRRRERN KIAAAKCRNKKKEKTECLQK ESEKLESVNAELKAQIEELKN EKQHLIYMLNLHR |
| Human | BATF | BATF | | PPGKQDSSDDVRRVQRREKN RIAAQKSRQRQTQKADTLHL ESEDLEKQNAALRKEIKQLTE ELKYFTSVLNSHE |
| Human | BATF | BATF2 | | GLLTQTDPKEQQRQLKKQKN RAAAQRSRQKHTDKADALH QQHESLEKDNLALRKEIQSLQ AELAWWSRTLHVHERLCP |
| Human | BATF | BATF3 | | QPQQQSPEDDDRKVRRREKN RVAAQRSRKKQTQKADKLH EEYESLEQENTMLRREIGKLT EELKHLTEALKEHE |
| Human | SMAF | MAFF | | RGLSAEEVTRLKQRRRTLKN RGYAASCRVKRVCQKEELQK QKSELEREVDKLARENAAMR LELDALRGKCEALQGFARSV A |
| Human | SMAF | MAFG | | RGLSKEEIVQLKQRRRTLKNR GYAASCRVKRVTQKEELEKQ KAELQQEVEKLASENASMKL ELDALRSKYEALQTFARTVA |

| | | | | |
|---|---|---|---|---|
| Human | LMAF | MAF | | RGVSKEEVIRLKQKRRTLKN RGYAQSCRFKRVQQRHVLES EKNQLLQQVDHLKQEISRLV RERDAYKEKYEKLV |
| Human | LMAF | MAFB | | RGFTKDEVIRLKHKRRTLKN RGYAQSCRYKRVQQKHHLE NEKTQLIQQVEQLKQEVSRL ARERDAYKVKCEKLA |
| Human | NFE2 | NFE2 | | YPLTESQLALVRDIRRRGKNK VAAQNCRKRKLETIVQLERE LERLTNERERLLRARGEADR TLEVMRQQLTELYRDIFQHL |
| Human | NFE2 | NFE2L1 | | YQLSEAQLSLIRDIRRRGKNK MAAQNCRKRKLDTILNLERD VEDLQRDKARLLREKVEFLR SLRQMKQKVQSLYQEVFGRL |
| Human | NFE2 | NFE2L2 | | EQFNEAQLALIRDIRRRGKNK VAAQNCRKRKLENIVELEQD LDHLKDEKEKLLKEKGENDK SLHLLKKQLSTLYLEVFSML |
| Human | NFE2 | NFE2L3 | | YYLTDLQVSLIRDIRRRGKNK VAAQNCRKRKLDIILNLEDD VCNLQAKKETLKREQAQCN KAINIMKQKLHDLYHDIFSRL |
| Human | BACH | BACH1 | | HKLTPEQLDCIHDIRRRSKNRI AAQRCRKRKLDCIQNLESEIE KLQSEKESLLKERDHILSTLG ETKQNLTGLCQKV |
| Human | BACH | BACH2 | | HKLTSEQLEFIHDVRRRSKNR IAAQRCRKRKLDCIQNLECEI RKLVCEKEKLLSERNQLKAC MGELLDNFSCLSQEV |
| *D. rerio* | ATF4 | NP_001096662.1 | DRE4A | SASGSKVVVEKKKLKKMEQ NKTAATRYRQKKRAEQETLL SECAVLEERNQELAEKAESLT KEIQYLKELMEEVKRAR |
| *D. rerio* | ATF4 | NP_998398.1 | DRE4B | VKTSSGAPKVEKKLKKMEQN KTAATRYRQKKRVEQESLNS ECSELEKKNRELSEKADSLSR EIQYLRDLLEEMRTAK |

| | | | | |
|---|---|---|---|---|
| *C. intestinalis* | CEBPG | ENSCINP00000023749 | CI1 | IVKGELEGDQDDYVKRRQRN NIAVKKSREKSREKSQITSERI DQLKEENCVLENKVEVLNQE LKVLKQVFMDHA |
| *C. intestinalis* | CEBP | ENSCINP00000023748 | CI2 | KKTKILIKGSKEYVQKRERNN VAVRRSRDKAKRKAAETQV KVDQLQNENLKLHEKVAELT HELTTLKNLLKAL |
| *C. intestinalis* | CEBP | ENSCINP00000023074 | CI3 | DLTDAPSTSGVKVSRKRDRN NAACRESRKKKKMKLVEAE MEVVRLVEDNEVQRLKIARL EVEVKETKALLLSKM |
| *C. intestinalis* | CEBP | ENSCINP00000002651 | CI4 | NSLIPILHELDKIDRRRIRNNE ACKKSRMRRKQRKMDKERE AERLAAQNLHLKRKISMLQS ECDKIRRQVLQAR |
| *C. intestinalis* | CREB | ENSCINP00000023419 | CI5 | SPQQMAEEASRKRELRLMKN REAAKECRRRKKEYVKCLET RVAVLENQNKQLIDELKTLK ELYVHKQN |
| *C. intestinalis* | OASISA | ENSCINP00000025439 | CI6 | LPLTKYEERVLKKVRRKIRN KKSAMASRQKKKDYIGGLEA RVTKCTNLNQALSQRVKQLE QQNFTLLEQLKQVHDAVK |
| *C. intestinalis* | OASISB | ENSCINP00000011026 | CI7 | LPLTKSEEKSLKKVRRKIKNK ISAQESRRKKKEYVETLEKR MDVYNRENTELRHKLDSLES SNRSLLSQLKSLQVLVA |
| *C. intestinalis* | XBP1 | ENSCINP00000024434 | CI8 | QENIPLSAIEDKELRKKLRNR QSALAARERKKARMMELEK QVAELQETNRRMEDENQHLR ARLDNII |
| *C. intestinalis* | XBP1 | ENSCINP00000010446 | CI9 | RKRLTHLTTEEKVMRRKLKN RVAAQTARDRKKVRMECLE DNIQKVQQQAKELLDVNMQ LLERAEALERENRELRVRL |
| *C. intestinalis* | XBP1 | ENSCINP00000021189 | CI10 | EQDDGYADVDEKELRKKLR NRESAQRARDRQKARMQWL EHEVSMLQVRNLTLTRENNL LRNLLA |

| | | | | |
|---|---|---|---|---|
| *C. intestinalis* | XBP1 | ENSCINP00000015310 | CI11 | TPRKSFEHVTDKELRKKLKN RESAQAARDRKKAKMLSLER QISELLERNRIVETENQELRSR IQRME |
| *C. intestinalis* | ATF6 | ENSCINP00000028986 | CI12 | LTIKDLDGRAMKRQQRMIKN REAACQSRQRRKEYVSTLEQ QMLECLDDNNKLRSMNQQL RDKVMELENENTRLR |
| *C. intestinalis* | NFIL3 | ENSCINP00000016562 | CI13 | PNSMCEDSKNKDYWVRRRK NNEAARRSREKRRMNDLLLE RRVLQLSEENKQLRAQLVAL KIKYGDTE |
| *C. intestinalis* | PAR | ENSCINP00000004693 | CI14 | KVHVSSDSKDVKYWNRRNK NNVAAKRSREARRIKENQIA MRANFLEKENESLKMEVADL RSELKRVMNTLRVYEKEII |
| *C. intestinalis* | ATF2 | ENSCINP00000005786 | CI15 | GRQQQDVDPDIKRQRFLERN RAAASRCRSKKKNWVVGLE SKAKTLSQTNVMLQNEITQL KDEIASLKQLLLSHR |
| *C. intestinalis* | JUN | ENSCINP00000018871 | CI16 | SPINMDHQELIKSERKRLRNR VAASKCRKRKLERISRLEDK VNNLKNQNLELTSSANLLRQ QVAELKSKVMTHV |
| *C. intestinalis* | FOS | 100130316 | CI17 | QDHELSPAEATKRHIRRERNK IAAAKCRNRRRELTDRLQGE TDHLEDHQSILHQEIMSLQQE KEHLEFLLAAHS |
| *C. intestinalis* | FOS | ENSCINP00000007607 | CI18 | DLDDLSDDERERMRVKRERN RVAAAKCRNRRRELLERLEK EAEQLEREQELLRESVKRLQS QKRKLGVMLDEHE |
| *C. intestinalis* | ATF4 | ENSCINP00000022333 | CI19 | GRKSKVTTTVERKQRKRDQN KNAATRYRERKRLEFSKQES EQRVLEEKNKSLHDNVNRVT REIEYLKELMIEVYKIKGLIK |
| *C. intestinalis* | ATF3 | ENSCINP00000005786 | CI20 | QNDEISPETLLKRERRRERNK VAAAKCRFKKKILSEQLQEES EHLENLNAKLKREIEKLQEER QKLMYLLNGHK |

| | | | | |
|---|---|---|---|---|
| *C. intestinalis* | SMAF | ENSCINP00000002543 | CI21 | RSLSPEESRRLKQRRRTLKNR GYAASCRIKRLTQKDELDIER IQLQNEVDRVTQENQRMKLE LEAFQKKFHDLEQFAKSI |
| *C. intestinalis* | LMAF | ENSCINP00000002531 | CI22 | RGLSKEDVMALKQRRRTLKN RGYAQSCRTKRVMQRHILEK EKDALQIQLNQVRDHLAAMS KERDDYKTKFERLRKFFL |
| *C. intestinalis* | NFE2 | ENSCINP00000024999 | CI23 | TPLTTAQQTLIKDIRRRGKNK VAAQNCRKRKIETITTMEED VDVLRGRKNDLEMEQDELE ARKQNLKSQYNALYQQIF |
| *C. intestinalis* | BACH | ENSCINP00000026548 | CI24 | PSLSPQQITAIHEIRRRGKNRI AAQRCRKRKMDCIRSLQCQL EQLREEHLNLMGERRTCQDK SLKLAEMFQKRYEQVF |
| *C. intestinalis* | NOVEL | 100138308 | CI25 | SLVAQLQNKDLQKFGDKS RNAVLAKLNREKKKHKIA LLETEVHHLRGKNNRLEK MNQEFSTSILDLQHEVKYL RGVIA |
| *C. intestinalis* | NOVEL | 100135870 | CI26 | RHFVPNECKDEYYWRKRKK NNEAARKSREKRKTIDSVLE DKVLFLSQENLCLRNELYAL KVNFT |
| *D. melanogaster* | PAR | CG17888-PE | DM1 | KQFVPDELKDDKYWARRRK NNIAAKRSRDARRQKENQIA MRARYLEKENATLHQEVEQL KQENMDLRARLSKFQDV |
| *D. melanogaster* | NOVEL | CG15479-PA | DM2 | VNMVRKFPKKERSPKDQERR NKNTIACRMSRRKKKFDDLQ IEQQYKECSDEHLKIAEQSLR ARVYLNHLKQLVK |
| *D. melanogaster* | FOS | CG33956-PD | DM3 | RSTNMTPEEEQKRAVRRERN KQAAARCRKRRVDQTNELTE EVEQLEKRGESMRKEIEVLTN SKNQLEYLLATHRATCQKIRS DMLSVVTCNGLIA |
| *D. melanogaster* | OASISB | CG7450-PB | DM4 | LPLTKAEEKSLKKIRRKIKNKI SAQESRRKKKEYMDQLERRV EILVTENHDYKKRLEGLEETN ANLLSQLHKLQALVSKHN |

| | | | | |
|---|---|---|---|---|
| *D. melanogaster* | ATF2 | CG30420-PC | DM5 | PPKAAKAKDRSRDEDCMERR RAAASRYRNKMRNEHKDLIK QNAQLQQENQELHERISRLE KELQQHR |
| *D. melanogaster* | ATF4 | CG8669-PA | DM6 | RTRTYGRGVEDRKIRKKEQN KNAATRYRQKKKLEMENVL GEEHVLSKENEQLRRTLQER HNEMRYLRQLIREFYHERK |
| *D. melanogaster* | PAR | CG7786-PA | DM7 | KRPIPEAQKDAKYFERRKRN NEAAKRSRDARKIREDRIAFR AALLEQENSILRAQVLALRDE LQTVRQLL |
| *D. melanogaster* | CEBP | CG4354-PA | DM8 | HSNKHVDKGTDEYRRRRERN NIAVRKSREKAKVRSREVEE RVKSLLKEKDALIRQLGEMT NELQLHKQIYMQLM |
| *D. melanogaster* | CREB | CG6103-PF | DM9 | DNSGIAEDQTRKREIRLQKNR EAARECRRKKKEYIKCLENR VAVLENQNKALIEELKSLKEL YCQTKN |
| *D. melanogaster* | CEBPG | CG6272-PA | DM10 | DSPLSPHTDDPAYKEKRKKN NEAVQRTREKTKKSAEERKK RIDDLRKQNDALKVQIETSEK HISTLRDLII |
| *D. melanogaster* | ATF3 | CG11405-PA | DM11 | QPKGLTPEDEDRRRRRRERN KIAATKCRMKKRERTQNLIK ESEVLDTQNVELKNQVRQLE TERQKLVDMLKSH |
| *D. melanogaster* | NFE2 | CG17894-PC | DM12 | YDLSENQLSLIRDIRRRGKNK VAAQNCRKRKLDQILTLEDE VNAVVKRKTQLNQDRDHLE SERKRISNKFAMLHRHVFQY L |
| *D. melanogaster* | LMAF | CG10034-PA | DM13 | HGCPREEVVRLKQKRRTLKN RGYAQNCRSKRLHQRHELEK ANRVLNQDLHRLKLEYSRVC QERDALMQRLQ |
| *D. melanogaster* | JUN | CG2275-PA | DM14 | NPIDMEAQEKIKLERKRQRN RVAASKCRKRKLERISKLEDR VKVLKGENVDLASIVKNLKD HVAQLKQQVMEHI |

| | | | | |
|---|---|---|---|---|
| *D. melanogaster* | ATF6 | CG3136-PA | DM15 | TPSHTMDDKIYKKYQRMIKN RESASLSRKKRKEYVVSLETR INKLEKECDSLKAENITLRDQI FLLA |
| *D. melanogaster* | NFIL3 | CG14029-PA | DM16 | REFTPDNKKDESYWDRRRRN NEAAKRSREKRRYNDMVLE QRVIELTKENHVLKAQLDAIR DKFNISGENLVSVEKILASL |
| *D. melanogaster* | PAR | CG7952-PB | DM17 | GISSGSQVKDAAYYERRRKN NAAAKKSRDRRRIKEDEIAIR AAYLERQNIELLCQIDALKVQ LAAFTSAKV |
| *D. melanogaster* | XBP1 | CG9415-PB | DM18 | KRRLDHLTWEEKVQRKKLK NRVAAQTSRDRKKARMEEM DYEIKELTDRTEILQNKCDSL QAINESLLAKNHKLDSELELL RQELAELK |
| *D. melanogaster* | SMAF | CG9954-PA | DM19 | RGLNREEIVRMKQRRRTLKN RGYAASCRIKRIEQKDELETK KSYEWTELEQMHEDNEQVR REVSNWKNKYKALL |
| *D. melanogaster* | NOVEL | CG18619-PA | DM20 | KPGRKTSTEKLDMKAKLERS RQSARECRARKKLRYQYLEE LVADREKAVVALRTELERLI QWNNQLSESNT |
| *D. melanogaster* | PAR | CG4575-PA | DM21 | GISSGSHVKDTAYYERRRKN NAAAKKSRDRRRIKEDEIAIR AAYLERQNIELLCRIDALEVQ LAAITSAKV |
| *D. melanogaster* | NOVEL | CG33719 | DM22 | QKENERLQTEVQLMKQELDA AEKAAISRAKKQAQIGELMQ RIKELEEMQSSLEDEASELRE QNELLEFRILELEDDSDKME |
| *D. melanogaster* | NOVEL | CG13624 | DM23 | MTPVSELPFNVRPKSRKEKN KLASRACRLKKKAQHEANKI KLFGLEIEHSEFNVKAVEIS |
| *D. melanogaster* | NOVEL | CG13624b | DM24 | MTPVSELPFNVRPKSRKEKN KLASRACRLKKKAQHEANKI KLFGLEIEHKRLMNGIAELKQ ALV |

| | | | | |
|---|---|---|---|---|
| *D. melanogaster* | NOVEL | CG16813 | DM26 | RNHHKRRQRSPQEQLRRDRN TLASLRHRRSQQQQQQLIEQ QYLTSRIQHEANLQQQIRLSL YYVRFL |
| *D. melanogaster* | NOVEL | CG16815 | DM27 | RRSNTNRQRSPKEQMRRDRN TLACLLSRRAKQAQEEQVGQ QYEQYRSHHAAMLEQQVRL SLYYRHIL |
| *D. melanogaster* | NOVEL | CG17836 | DM28 | LMSSMKSEEERKAYQDRLKN NEASRVSRRKTKVREEEEKR AEDTLLAENLRLRARADEVA SRERKFKKYLM |
| *D. melanogaster* | NOVEL | CG1641 | DM29 | DVKDAQRQRAESCRKSRYN NKIKKAKLRFRHKFVSGQLK KSAVMLDTMRDVIAQAERQL L |
| *D. melanogaster* | FOS | CG33956-PD | DM30 | RSTNMTPEEEQKRAVRRERN KQAAARCRKRRVDQTNELTE EVEQLEKRGESMRKEIEVLTN SKNQLEYLLATHR |
| *C. elegans* | CEBPA | D1005.3 | CE1 | KLKADEEKAEPTYKLKRA RNNDAVRKSRKKAKELQ DKKEAEHDKMKRRIAELE GLLQSERDARRRDQDTLE QLLRNK |
| *C. elegans* | NOVEL | zip-2 | CE2 | KTSSVSSDSSSDYRHKRDK NNLASQKSRQKRQAKIRE SKEERERLEKRKVQLQAM VLTLETQVEDYKRLVMMF VKR |
| *C. elegans* | PAR | F17A9.3 | CE3 | LKRKKDQVKDVAYWERR RKNNDAAKRSRDQRRMK EDEMAHRATSLERENMLL RVELDQLRAETDKLRALIL |
| *C. elegans* | NFIL3 | atf-2n | CE4 | NSVNESVIKDEHYWERRR RNNDASRRSREKRRQNDL AMEEKIMLLSAENERLKS QL |
| *C. elegans* | ATF2 | atf-7 | CE5 | RSTTADMQPDERRNTILER NKAAAVRYRKRKKEEHD DMMGRVQAMEAEKNQLL AIQTQNQVLRRELERVTAL LTERESRCVCLK |

| | | | | |
|---|---|---|---|---|
| *C. elegans* | XBP1 | xbp-1 | CE6 | RERLNHLSQEEKMDRRKL KNRVAAQNARDKKKERS AKIEDVMRDLVEENRRLR AENERLRRQNKNLMNQQ N |
| *C. elegans* | JUN | jun-1 | CE7 | CGMALDDQEKKKLERKR ARNRQAATKCRQKKMDRI KELEEQVLHEKHRGQRLD AELLELNRALEHFRRTVEH HS |
| *C. elegans* | FOS | fos-1 | CE8 | EEDNMEDDDDDKRLKRR QRNKEAAARCRQRRIDLM KELQDQVNDFKNSNDKK MAECNNIRNKLNSLKNYL ETHD |
| *C. elegans* | CREB | crh-1 | CE9 | GPLHGEDESNRKRQVRLL KNREAAKECRRKKKEYV KCLENRVSVLENQNKALIE ELKTLKELYCRKEKD |
| *C. elegans* | ATF6 | atf-6 | CE10 | VDIKAEPQVFTSEQNRKIR NRMYAQASRMRKKEADE HMKMNLQELLQENEILRT ENAALKQRLAFFEHEE |
| *C. elegans* | ATF4 | atf-5 | CE11 | EKSYHPYKTPEKKERKKA QNRLAATRYREKKRREKE EAMTCIEGLSVTNGKLKD QVSELEREIRYFKKFMTEM |
| *C. elegans* | PAR | ces-2 | CE12 | SVPIPEEKKDSAYFERRRK NNDAAKRSRDARRQKEEQ IASKAHALERENMQLRGK VSSLEQEAAQLRFLLFSKI |
| *C. elegans* | OASISA | let-607 | CE13 | FPLTKAEERDLKRIRRKIR NKRSAQTSRKRKQDYIEQ LEDRVSESTKENQALKQQI ERLSSENQSVISQLKKLQA QL |
| *C. elegans* | PAR | atf-2c | CE14 | EDHSNYSNKSPQYVDRRR RNNEAAKRCRANRRAVFE YRSRRVQLLEGENEDLRT QIETLKAEIAHFKSVLAQR ASVVTALH |
| *C. elegans* | ATF4 | zip-3 | CE15 | EGREEEESPEEILRRKRIQN NLAAARYRKRQREARESA ESELGDLTRRNDELRDQV SRMEREIDRLKQAVL |

| | | | | |
|---|---|---|---|---|
| *C. elegans* | NOVEL | zip-5 | CE16 | TSCDEKLDLVSEDEKKRL RNTEAARRCREKIKRKTD DLETELTRLTARNEVMNQ HRIRLLSQVEEQMRMLENI KSRN |
| *C. elegans* | NOVEL | F23F12.9a | CE17 | DGSKIDPKRSPKYLEKRM KNNEAAKKSRASRKHREQ KNQTENELLKRKNAALEE ELKQAKCELAQMQITIRD MSIEREAYRRENEMLKMV NNKFADSKF |
| *C. elegans* | CEPBG | C48E7.11 | CE18 | NTSEPREDDEDDYSTKRK RNNEAVNRTRQKKRQEEN DTAEKVDELKKENETLER KVEQLQKELSFLKEMFMA YA |
| *C. elegans* | MAF | F45H11.6 | CE19 | MGQDRNVVMQWKQKRR TLKNRGYALNCRARRVNN QVQLEADNMMLRNQIKTL REALSEAQMRLHYYE |
| *C. elegans* | ATF4 | ZC376.7c | CE20 | RGVVLKPSVDEETDRRRM LNRIAAVRYREKKRAEKK GRKMEFQEVADRNRILLQ KERQLKREINSMKKELRK MGAIIQ |
| *C. elegans* | OASISB | C27D6.4 | CE21 | YPLTKSEEESLKIVRRKIKN KLSAQESRRKRKEYIDALE GRLHCFSEENKSLKKQVH QLEASNRDLQQKLHQYE |
| *C. elegans* | NOVEL | W08E12.1 | CE22 | KKEGSSNDETKLLSRKRQ QNKVAAARYRDKQKAKW QDLLDQLEAEEDRNQRLK LQAGHLEKEVAEMRQAFL AKL |
| *C. elegans* | PAR | Y51H4A.4 | CE23 | PVPVPENQKDEAYLDRRR RNNEAARKSRESRKKVDQ DNSVRVTYLERENQCLRV YVQQLQLQNESMRQHLLL QN |
| *C. elegans* | CEBPA | zip-4 | CE24 | NLKPDKEKVEPIYKLKRA RNNDAVRKSRNKAKELQL QKDEEYDEMKKRITQLEA ELQSEREGRERDQQLIKQL IREK |

| | | | | |
|---|---|---|---|---|
| *C. elegans* | NOVEL | R07H5.10 | CE25 | YEKVSEDQKDEKYSSKRE KNNLAVKRCREKKKNEEK YKKEAFENLIRSNLVKDQ KIEQLNNLVQSGKQRENA RIMEVQREKNILRQLKNEL TRI |
| *C. elegans* | NOVEL | zip-1 | CE26 | KAEMSRLTEKEKLERKKE QNRANAKNCVKNRNNSK EELKQTLEMLREKVQEAK RQNEMQENGLLAAYETNI |
| *C. elegans* | NOVEL | Y116F11B.6 | CE27 | SQAAQSNIPSGKAKTKRER NRIAAAKSRRLEKELMRK TQAIYETKKITHEQLCAYN NSNDSLFKTAVESVL |
| *C. elegans* | NOVEL | F17C11.17 | CE28 | PVSLVNLSDEEIAERKKQQ NRAAALRYRQKLRESRVM SVSVKETLTQRNAYLRDE AERLSKECEVIRRLIFDKL GKNA |
| *C. elegans* | NOVEL | Y17G7B.20 | CE29 | VKSPSSKRGRPSKVTSNSK MANYARNYREQKKNEMS TLQMHNSELEAELRLARE ENAKMKKALAKASDEITQ LKKVIDQDSQIARVV |
| *C. elegans* | ATF2 | atf-7 | CE30 | RSTTADMQPDERRNTILER NKAAAVRYRKRKKEEHD DMMGRVQAMEAEKNQLL TQNQVLRRELERVTALLT ERESRCVCLK |
| *C. elegans* | NOVEL | R07H5.10 | CE31 | YEKVSEDQKDEKYSSKREKN NLAVKRCREKKKNEEKYKK EAFENLIRSNLVKDQKIEQLN NLVQ |
| *N. vectensis* | PAR | 165267 | NV1 | TEKHFNNNKDNKYWEKRQR NNASAKRSRDARRVRELECQ IRAEFLEEENHKYKVENEML REENERLLKIIESFNNKQ |
| *N. vectensis* | OASISB | 28519 | NV2 | LPLTKVEERALKKVRRKIKN KISAQESRRKKKEYMETLEK RVETCSSENLELRKKLDSLEN TNRNLIGQLQKLQALIS |

| | | | | |
|---|---|---|---|---|
| *N. vectensis* | FOS | 233229 | NV3 | RKEQLTPEEEEKRRLRRERNKQAANRCRKRKRDKIEMLERTAQEIDDSNKALETDIANMRTELTELMSVLRSHDCVMRSRD |
| *N. vectensis* | NOVEL | 238168 | NV4 | LKPIHSLPLNARNKSRKEKNKLASRACRLKKKAQHEANKLKLHGLELEQQRLIHVIEKVRSEII |
| *N. vectensis* | CEBPG | 114346 | NV5 | DDEYIRKRERNNEAVRKSRKKAKQRIQETQQRVTELSKENEELRSKVTLLQKELSVLRSLFA |
| *N. vectensis* | XBP1 | 211292 | NV6 | RRRLDNLTVEERALRRKLKNRVAAQTARDRKKARMQDLEEAVESLERENKRLREENKRLNKSTESLAIENSELRVRL |
| *N. vectensis* | FOS | 246444 | NV7 | TPQPRPPEKPEVVEQRRRQNKFAAMKSRKKRTERINRLRQKTRKYEESIRKHGMVVKKLREEAEQLKQYLISHNCCKN |
| *N. vectensis* | ATF2 | 34679 | NV8 | RRSQEELDPDERRRKFLERNRAAATRCREKRKIWVQQLEKKADDLSNTNTQLQNEISLLRTEVAQLKSLLLAHK |
| *N. vectensis* | NOVEL | 248021 | NV9 | FNEPLTEEELKDIEDKNKKNAIAARENRAKKKKYMEDLEKTVQDLKKENQELQTGHSKLQKTVEALNDEVSYYKNVLA |
| *N. vectensis* | PAR | 241379 | NV10 | PSPSSSSGDSDKSEEKRKRNNQASKKFRQARKGKQQALFAKESELERENYSLKVQVEQLIRELNQLKAALH |
| *N. vectensis* | CEBPG | 104726 | NV11 | SKRNSMDKHSEEYRQKRERNNVAVRKSRFKSKQKFIETQSRVEELTEENERLHSRIDIITKELNALRSLFS |
| *N. vectensis* | LMAF | 118896 | NV12 | KGLSTEEQSRIKYRRRTLKNRGYAHNCRIKRISQKKSLEETNWELVQDLENLRKELEASKRERDMYKRKYENLYAMVM |

| | | | | |
|---|---|---|---|---|
| *N. vectensis* | SMAF | 99714 | NV13 | RGLPEDDVFKLKQRRRTLKN RGYAQNSRTKRVRQREDLEY ERQQLKDELFMVSKENEDLR RERDEAKRKYDSLQKLLT |
| *N. vectensis* | JUN | 238589 | NV14 | PPIDLDLQEAVKNERKKLRN RLAASKCRKRKLEKEAELED KVKVLKDKNTKLVSEAQELR RLVCELKEQVMNHV |
| *N. vectensis* | JUN | 95962 | NV15 | QPIDLEIQEVVKRERKKQRNR IASSKCRKRKLEREARLENRV KDLKERNIELNAVANALKQQ VCDLKQRVMDHV |
| *N. vectensis* | PAR | 80243 | NV16 | AYGRDKDQKYIEKRMKNNL AAKRSREAKRQREIEAMQKT LTLEKENSDLNKEVNKLKKM IARLENKLR |
| *N. vectensis* | PAR | 86952 | NV17 | RTSVPGEMKDQKYWERRLK NNVAAKRSRDLKRQKEMTV AKRAQNLEIENEKLRNEVTM LKKRLQTLNGKLD |
| *N. vectensis* | LMAF | 242787 | NV18 | RGLPSTEIDTIRKRRRSLKNR GYAMNCRTKREQENKELAK MNKKLARDVVSMKEELRKIK KERDAMKTKYDKMREVLNR LC |
| *N. vectensis* | CREB | 243410 | NV19 | SNQQIAEEATRKREMRLMKN REAAKECRRKKKEYVKCLEN RVAVLENQNKTLIEELKALK DLYCHKSE |
| *N. vectensis* | NOVEL | 127893 | NV20 | VSPTQLDMDRYVSDEGINRQ AIMAKINREKKKQYVQELEG SVEEYKSKNAVLQKDCEDM KGLVKDLQMEIAYLKGVLA |
| *N. vectensis* | NFE2 | 245260 | NV21 | EKLSDAQAKYVRDVRRRGK NKEAARICRKRKMDAIETLD DEITRLKQQRQSMFDERKDL QQETAELKRKISELESSLF |
| *N. vectensis* | ATF6 | 242270 | NV22 | QQPKVLDEKILRRQQRMIKN RESACLSRKKKKEYLQSLET QIKEVNLLNDKLSEENIKLKK RVQELENENNILKAKN |

| | | | | |
|---|---|---|---|---|
| *N. vectensis* | NOVEL | 197394 | NV23 | SRDTKASTSIDKATERRIKNNI ASRHTRAARRQREQELFEKE EYLKKNNEELKQQIVELTKET EILRKLVIQRLSSVN |
| *N. vectensis* | PAR | 243817 | NV24 | LSKAIADVKTEQYREKRRKN NASAKRSREARKMREIHAQT AAAYLQDENAKLRALVNVL KEENVYLREIML |
| *N. vectensis* | OASISA | 244559 | NV25 | VSLTKAEERVLKKVRRKIKN KQSAQESRKKKKDYVDGLE MRVKVCTEKNTSLQKKVDN LEKQNLTLMDQLKQLQAIVA |
| *N. vectensis* | FOS | 232694 | NV26 | FFCQLTPAEELKIIRRRQRNK QAASRCREKRRQRLEELQRE ATELEEQNAEVERDIATLRVE YNELEALLTEHACVL |
| *N. vectensis* | FOS | 126097 | NV27 | KVEELSPAELEKRRIRRERNK LAAFKCRQRRKEHIQELEIES EGIEDSNKELEREISELHEQRQ QLEEMLKTHSCKLS |
| *N. vectensis* | JUN | 150375 | NV28 | PPIDLELQEIVKRERKKQKNR VAASKCRRKKLEREAQLEVR VQQLKEKSIELNAVASALRQ QVGELKQRVLEHV |
| *N. vectensis* | PAR | 29743 | NV29 | RKFVPDQEKDDRYWARRVK NNVAARRSRDMRRQKEIEIS MKWKQLEKENARLREELQQ LKDRASELEKKLSEKQ |
| *N. vectensis* | FOS | 248713 | NV30 | LTPEEETRRKVRRQRNKVAA SKCRLKRREHVKNLLKASEE LESANSKLESDIACLNAEKEQ LERMLDAHK |
| *N. vectensis* | PAR | 242269 | NV31 | TPVKVSKMDLQREAEKRRKN NEASKRTREKRRNKEQELLK EKEIKEKENKALRTQVEDLE KQIKDIRSALDQRL |
| *N. vectensis* | NOVEL | 18367 | NV33 | GKKRGRKPSQIDLEAKLERSR QSARECRARKKLRYKCLEDT VTRKESEVSKLRQELDMYVR WCKAIDQGVY |

| | | | | |
|---|---|---|---|---|
| *N. vectensis* | LMAF | 39846 | NV34 | RGLENSEIVRLRKRRRSLKNR IYASVCKKKRVAEQKTYEVQ NRILVKERNTLKMELEKVKT ERDKIKEAYQTL |
| *N. vectensis* | ATF4 | sca_303 | NV35 | DTVSPKLKTPAQRQRKRVQN KDAATRYRVKKKDEQSRLFD EAEKLEKENNELKDEVGSLS KEIEYLKNLMLEVYQTKQ |
| *N. vectensis* | NOVEL | 242703 | NV36 | SKDEQDLHRKLQEIQATQGD MEEAQREIEKKKTEIEKIKAE LEELQQKTVTLNRKRKSLSSE CSQLQKKLHYCDSVLQVV |
| *M. brevicollis* | | 37668 | MB1 | RASSVDPPIDERRLKHLERNR AAATRCRERKKQWLQQLQQ KAATLTTSNRQMHEELKRLR DEVLNLKGNLV |
| *M. brevicollis* | | 36000 | MB2 | DSEEAFSLAWQEWRSVRQKN NAAVHKSRQKAKARRAVDR HAAREKERKAAQLAMEAEM LRKNVDVLIKAVR |
| *M. brevicollis* | | 11417 | MB3 | DECTKHMKGMTPAQKKRLR NKHASCVSRLKKKLYICNLV RELDRAKETAAAFQDDMDA LRARVTELEAENQHLR |
| *M. brevicollis* | | 38819 | MB4 | DIKPDTTATAKRPSNKRASNR ESARRFRQRRKEYIGQLEKK VSRLISENQRLRALLTAHL |
| *M. brevicollis* | | 32288 | MB5 | RRRRIADLAEADRARLRRLN REAARKHRERSKWRDESAA QDLQRLVLHHKQLASEAAAL RTEVSTLREVVRTLY |
| *M. brevicollis* | | 9939 | MB6 | DTEDEYQDAWQRWRAIRDH NNESVKRSRENARHRKHQHE AACRERERENSQLATEVDRL KDQVVLLTKVLK |
| *M. brevicollis* | | 30420 | MB7 | GLITQAQSRELKRMRRKVKN KLSAKDSRRRRKEYVTQLEE ENAQLRARLVTLHDQSMAR Q |

| | | | | |
|---|---|---|---|---|
| *M. brevicollis* | | 34232 | MB8 | PTSPASTVDSQLTDRTREFNRIAALRHRQRAKMRRLELDQRLLEASRHQQQLKMEMEELSKKHHSLLELCFTLY |
| *M. brevicollis* | | 32251 | MB9 | ELSGETTSKRAKTTDKKQLNKQAADRYRRKKRQQFEELQSQSSELADENKALSVKCERLENEVAYLKDLLM |
| *M. brevicollis* | | 31571 | MB10 | KVAVAEHLKDEAYLAYRELNNERARRCREKKREEKRQASRRLQTLDAENERLKDEMHRLQDALKDLVQAMQARV |
| *M. brevicollis* | | 31254 | MB11 | HSSDTHEDDDDHAGSTSNPNKSAADRYRKKKREEFERLQHDTEAMKAENLELKTRLSKLRNEAEFLANMLQ |
| *M. brevicollis* | | 38264 | MB12 | TKPSAGLSKAQLAEWRRDNNRTAAKDLRDRKRQFEEDVSHVVELAEAENAKLAARAQQLEHHHATMRARLGAFMHTFNQVT |
| *M. brevicollis* | | 31590 | MB13 | PANTPRGEDSNNYRIKRIRNNEAVRRCRIKKKQEMEEKAMRLELLEHKVSDLENCNRKLSELIVEQQKEIQRLRSERDTL |
| *M. brevicollis* | | 24481 | MB14 | ATDDEYNLAWIKWRQSRDSNNRSVKRSREKARERYQEIEIQKDHLVQHNTELLNQLRQAQ |
| *M. brevicollis* | | 22289 | MB15 | LSPVELLEIKEKKERRMLKNRESASLSRKRKKEYLETLEHQLHDAQQQLGRAQHQIQQLQNDNHVLREQLANYHGFVN |
| *M. brevicollis* | | 32766 | MB16 | RALTKAEEKELKKVRRKVKNKISAQDSRKRRKEYLSQLEDKVKSAMTNNRSLKTRVSSLERQNSNLMEQINELHARLA |
| *M. brevicollis* | | 33073 | MB17 | KRKLVADLNSSELEKMREVNRIAAQRHRLIEKAKRRERQDRFDSAIRLQKALQEEVVVLENELATLRRLVIELY |

| | | | | |
|---|---|---|---|---|
| *M. brevicollis* | | 6968 | MB18 | AGLSKSEVADVKAKRRRLKN RLSARLCSNKKREKCSELEDT NRDLLAKLRQVAQENKTLKS ETNRLKEANTALT |
| *M. brevicollis* | | 38380 | MB19 | EHETSEQQAELRKRRRRTQN RSAAKTSALRRKTNSLSTHA KLAKFEEENRSLQQQLSLAR QEKEDLLLANRILRAEIA |
| *M. brevicollis* | | 10973 | MB20 | KALEAINTAGTDAARRKTRN RLASAVSRARKKVFLHRLRS ELLQLAARYQVSTIESQQFRL QSLQAQRELWDLK |
| *M. brevicollis* | | 31046 | MB21 | GSEEEYQLAWTKWRESRDN NNESVKRSRMMAKKKREEQ ERVHEEREAQNRKLETVVSS MRDEVKFLNKVLK |
| *M. brevicollis* | | 10034 | MB22 | LVDQLEARLETMTQHATEQN KQLLRTTKKRQLEIDSELTSS SEAIAKQIAQVQADLATLRRN NKEIETKLR |
| *S. cerevisiae* | | YAP3 | SC1 | SVAHNENVPDDSKAKKKAQ NRAAQKAFRERKEARMKEL QDKLLESERNRQSLLKEIEEL RKANTEINAENRLLL |
| *S. cerevisiae* | | YAP5 | SC2 | HEDYETEENDEELQKKKRQN RDAQRAYRERKNNKLQVLEE TIESLSKVVKNYETKLNRLQN ELQAKESENHALKQKLETLT LKQASV |
| *S. cerevisiae* | | HAC1 | SC3 | KRAKTKEEKEQRRIERILRNR RAAHQSREKKRLHLQYLERK CSLLENLLNSVNLEKLADH |
| *S. cerevisiae* | | CST6 | SC4 | QGNPIPGTTAWKRARLLERN RIAASKCRQRKKVAQLQLQK EFNEIKDENRILLKKLNYYEK LISKFKKFSKIHLREHEKLN |
| *S. cerevisiae* | | SKO1 | SC5 | VTLDENEEQERKRKEFLERN RVAASKFRKRKKEYIKKIEND LQFYESEYDDLTQVIGKLCGII |

| | | | | |
|---|---|---|---|---|
| *S. cerevisiae* | | ACA1 | SC6 | TAGLKDGAKAWKRARLLER NRIAASKCRQRKKMSQLQLQ REFDQISKENTMMKKKIENY EKLVQKMKKISRLHM |
| *S. cerevisiae* | | CIN5 | SC7 | GQLIGKTGKPLRNTKRAAQN RSAQKAFRQRREKYIKNLEE KSKLFDGLMKENSELKKMIE SLKSKLKE |
| *S. cerevisiae* | | YAP1 | SC8 | KTSKKQDLDPETKQKRTAQN RAAQRAFRERKERKMKELEK KVQSLESIQQQNEVEATFLRD QLITLVNELKKYR |
| *S. cerevisiae* | | MET28 | SC9 | APVSTSNELDKIKQERRRKNT EASQRFRIRKKQKNFENMNK LQNLNTQINKLRDRIEQLNKE NEFWKAKLNDINEIKS |
| *S. cerevisiae* | | GCN4 | SC10 | PLSPIVPESSDPAALKRARNTE AARRSRARKLQRMKQLEDK VEELLSKNYHLENEVARLKK LV |
| *S. cerevisiae* | | CAD1 | SC11 | GRPGRKRIDSEAKSRRTAQN RAAQRAFRDRKEAKMKSLQ ERVELLEQKDAQNKTTTDFL LCSLKSLLSEITKYRAKNSDD ERILAFLD |
| *S. cerevisiae* | | YAP7 | SC12 | GNGSGDENGVDSVEKRRRQ NRDAQRAYRERRTTRIQVLE EKVEMLHNLVDDWQRKYKL LESEFSDTKENLQKSIALNNE LQKAL |
| *S. cerevisiae* | | YAP6 | SC13 | TQLISSSGKTLRNTRRAAQNR TAQKAFRQRKEKYIKNLEQK SKIFDDLLAENNNFKSLNDSL RNDNNILIAQHEAIRNAITML RSEYD |
| *S. cerevisiae* | | ARR1 | SC14 | RKGGRKPSLTPPKNKRAAQL RASQNAFRKRKLERLEELEK KEAQLTVTNDQIHILKKENEL LHFMLRSLLT |
| *S. cerevisiae* | | MET4 | SC15 | HGFEKKQLIKKELGDDDEDL LIQSKKSHQKKKLKEKELESS IHELTEIAASLQKRIHTLETEN KLLKNLVL |

Table 5.2: Equilibrium dissociation constants. $K_d$ values, in nM, for each interaction measured. NB, Non-binders.

Human bZIP interactions.

| | Family | DDIT3 | CEBPG | CEBP | CEBP | CREB | OASISA |
|---|---|---|---|---|---|---|---|
| Family | Protein | DDIT3 | CEBPG | CEBPA | CEBPE | CREB1 | CREB3 |
| DDIT3 | DDIT3 | 8.1 | <1 | <1 | <1 | NB | 315.1 |
| CEBPG | CEBPG | <1 | 2.0 | <1 | <1 | NB | 365.4 |
| CEBP | CEBPA | <1 | <1 | 7.9 | 19.0 | NB | NB |
| CEBP | CEBPE | <1 | <1 | 19.0 | <1 | NB | NB |
| CREB | CREB1 | NB | NB | NB | NB | 20.7 | NB |
| OASISA | CREB3 | 315.1 | 365.4 | NB | NB | NB | 78.0 |
| OASISA | CREB3L3 | >5000 | NB | NB | NB | NB | 527.0 |
| OASISB | CREB3L1 | 600.9 | 4476.0 | NB | NB | NB | 355.5 |
| CREBZF | CREBZF | NB | NB | NB | NB | NB | NB |
| XBP1 | XBP1 | NB | NB | NB | NB | NB | 3169.0 |
| ATF6 | ATF6 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6B | NB | NB | NB | NB | NB | 2693.4 |
| NFIL3 | NFIL3 | 219.0 | 4135.8 | NB | NB | 2869.6 | 2144.8 |
| PAR | DBP | 1.3 | 101.9 | 245.2 | 278.5 | NB | NB |
| PAR | HLF | 7.9 | NB | NB | NB | NB | NB |
| ATF2 | ATF2 | 14.8 | 80.8 | 1935.4 | NB | NB | NB |
| JUN | JUN | 16.8 | 425.9 | NB | 3859.1 | 2566.9 | NB |
| JUN | JUNB | 146.1 | NB | NB | 2222.3 | NB | NB |
| FOS | FOS | 18.9 | 61.4 | 28.5 | 290.3 | NB | NB |
| FOS | FOSL1 | 115.0 | 170.3 | 399.8 | 363.1 | NB | NB |
| ATF4 | ATF4 | <1 | <1 | <1 | <1 | NB | 48.2 |
| ATF4 | ATF5 | 2261.4 | <1 | 76.8 | 29.2 | NB | NB |
| ATF3 | ATF3 | <1 | <1 | 29.2 | 63.8 | NB | 1535.9 |
| BATF | BATF | <1 | <1 | 24.0 | 69.2 | NB | 288.6 |
| BATF | BATF2 | 5.6 | 1.5 | 77.4 | 115.7 | NB | 352.3 |
| BATF | BATF3 | <1 | 9.6 | 7.5 | 45.7 | 2453.2 | 542.1 |
| SMAF | MAFF | 168.3 | 1156.8 | NB | NB | NB | NB |
| SMAF | MAFG | 447.5 | NB | NB | NB | NB | NB |
| LMAF | MAF | NB | NB | NB | NB | NB | NB |
| LMAF | MAFB | NB | NB | NB | NB | NB | NB |
| NFE2 | NFE2 | 3271.5 | NB | NB | NB | NB | 2433.2 |
| NFE2 | NFE2L1 | 2898.0 | 1738.2 | NB | NB | NB | NB |
| NFE2 | NFE2L2 | NB | 1637.0 | NB | 2368.5 | NB | 439.5 |
| NFE2 | NFE2L3 | NB | 1124.0 | 4696.3 | NB | NB | 191.4 |
| BACH | BACH1 | 59.5 | NB | NB | NB | 3259.0 | NB |
| BACH | BACH2 | 79.2 | 2290.5 | NB | 2491.4 | NB | 2634.8 |

| | Family | OASISA | OASISB | CREBZF | XBP1 | ATF6 | ATF6 |
|---|---|---|---|---|---|---|---|
| Family | Protein | CREB3L3 | CREB3L1 | CREBZF | XBP1 | ATF6 | ATF6B |
| DDIT3 | DDIT3 | >5000 | 600.9 | NB | NB | NB | NB |
| CEBPG | CEBPG | NB | 4476.0 | NB | NB | NB | NB |
| CEBP | CEBPA | NB | NB | NB | NB | NB | NB |
| CEBP | CEBPE | NB | NB | NB | NB | NB | NB |
| CREB | CREB1 | NB | NB | NB | NB | NB | NB |
| OASISA | CREB3 | 527.0 | 355.5 | NB | 3169.0 | NB | 2693.4 |
| OASISA | CREB3L3 | 5.1 | 96.1 | NB | NB | NB | NB |
| OASISB | CREB3L1 | 96.1 | 8.9 | NB | NB | NB | NB |
| CREBZF | CREBZF | NB | NB | 1.9 | <1 | 3864.7 | 3517.1 |
| XBP1 | XBP1 | NB | NB | <1 | 6.2 | <1 | <1 |
| ATF6 | ATF6 | NB | NB | 3864.7 | <1 | 15.2 | <1 |
| ATF6 | ATF6B | NB | NB | 3517.1 | <1 | <1 | 1.1 |
| NFIL3 | NFIL3 | NB | 346.9 | 3666.6 | NB | NB | NB |
| PAR | DBP | NB | NB | NB | NB | NB | NB |
| PAR | HLF | NB | NB | NB | NB | NB | NB |
| ATF2 | ATF2 | NB | NB | NB | NB | NB | NB |
| JUN | JUN | NB | NB | NB | NB | NB | NB |
| JUN | JUNB | NB | NB | NB | NB | NB | NB |
| FOS | FOS | NB | NB | NB | NB | NB | NB |
| FOS | FOSL1 | NB | NB | NB | NB | NB | NB |
| ATF4 | ATF4 | 2883.8 | >5000 | 35.9 | NB | NB | NB |
| ATF4 | ATF5 | NB | NB | 2007.2 | NB | NB | NB |
| ATF3 | ATF3 | NB | NB | NB | NB | NB | NB |
| BATF | BATF | NB | NB | NB | NB | NB | NB |
| BATF | BATF2 | NB | >5000 | NB | NB | NB | NB |
| BATF | BATF3 | NB | NB | NB | NB | NB | NB |
| SMAF | MAFF | NB | NB | NB | NB | NB | NB |
| SMAF | MAFG | NB | NB | 4377.7 | NB | NB | NB |
| LMAF | MAF | NB | NB | NB | NB | NB | NB |
| LMAF | MAFB | NB | NB | NB | >5000 | NB | NB |
| NFE2 | NFE2 | NB | NB | 167.2 | 2792.3 | NB | 2826.2 |
| NFE2 | NFE2L1 | NB | NB | 33.8 | 1903.1 | NB | NB |
| NFE2 | NFE2L2 | NB | 1675.7 | 86.3 | NB | NB | NB |
| NFE2 | NFE2L3 | NB | NB | 1922.9 | NB | NB | NB |
| BACH | BACH1 | NB | NB | 2576.1 | NB | NB | NB |
| BACH | BACH2 | NB | NB | 458.1 | 1887.3 | NB | 2313.7 |

| | Family | NFIL3 | PAR | PAR | ATF2 | JUN | JUN |
|---|---|---|---|---|---|---|---|
| Family | Protein | NFIL3 | DBP | HLF | ATF2 | JUN | JUNB |
| DDIT3 | DDIT3 | 219.0 | 1.3 | 7.9 | 14.8 | 16.8 | 146.1 |
| CEBPG | CEBPG | 4135.8 | 101.9 | NB | 80.8 | 425.9 | NB |
| CEBP | CEBPA | NB | 245.2 | NB | 1935.4 | NB | NB |
| CEBP | CEBPE | NB | 278.5 | NB | NB | 3859.1 | 2222.3 |
| CREB | CREB1 | 2869.6 | NB | NB | NB | 2566.9 | NB |
| OASISA | CREB3 | 2144.8 | NB | NB | NB | NB | NB |
| OASISA | CREB3L3 | NB | NB | NB | NB | NB | NB |
| OASISB | CREB3L1 | 346.9 | NB | NB | NB | NB | NB |
| CREBZF | CREBZF | 3666.6 | NB | NB | NB | NB | NB |
| XBP1 | XBP1 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6B | NB | NB | NB | NB | NB | NB |
| NFIL3 | NFIL3 | 52.1 | 3964.6 | NB | NB | NB | NB |
| PAR | DBP | 3964.6 | 6.5 | <1 | >5000 | 3768.2 | 3842.1 |
| PAR | HLF | NB | <1 | 2.9 | NB | NB | 3616.8 |
| ATF2 | ATF2 | NB | >5000 | NB | 29.4 | 16.4 | 939.4 |
| JUN | JUN | NB | 3768.2 | NB | 16.4 | 185.9 | 2961.3 |
| JUN | JUNB | NB | 3842.1 | 3616.8 | 939.4 | 2961.3 | NB |
| FOS | FOS | NB | 2373.4 | NB | 6.7 | <1 | <1 |
| FOS | FOSL1 | NB | NB | NB | 236.7 | <1 | 2.4 |
| ATF4 | ATF4 | NB | NB | 721.2 | 7.7 | 25.2 | 258.8 |
| ATF4 | ATF5 | NB | NB | NB | NB | NB | NB |
| ATF3 | ATF3 | NB | 129.3 | NB | <1 | <1 | 5.6 |
| BATF | BATF | 262.9 | 117.3 | 28.8 | 18.2 | <1 | <1 |
| BATF | BATF2 | 1251.4 | 457.0 | 779.8 | 635.7 | <1 | 2.1 |
| BATF | BATF3 | 184.0 | 43.3 | 92.1 | 28.2 | <1 | <1 |
| SMAF | MAFF | 50.3 | 4403.3 | NB | NB | NB | >5000 |
| SMAF | MAFG | 739.4 | NB | NB | NB | 4149.9 | 3624.5 |
| LMAF | MAF | NB | NB | NB | NB | NB | NB |
| LMAF | MAFB | NB | NB | NB | NB | NB | NB |
| NFE2 | NFE2 | NB | >5000 | NB | NB | NB | NB |
| NFE2 | NFE2L1 | 1751.8 | 3953.2 | NB | NB | NB | NB |
| NFE2 | NFE2L2 | NB | NB | NB | NB | NB | NB |
| NFE2 | NFE2L3 | 1198.9 | 3278.5 | NB | NB | NB | NB |
| BACH | BACH1 | 586.7 | 4238.2 | NB | 60.4 | NB | NB |
| BACH | BACH2 | 2408.6 | 3457.4 | NB | 667.6 | 4461.0 | NB |

| | Family | FOS | FOS | ATF4 | ATF4 | ATF3 | BATF |
|---|---|---|---|---|---|---|---|
| Family | Protein | FOS | FOSL1 | ATF4 | ATF5 | ATF3 | BATF |
| DDIT3 | DDIT3 | 18.9 | 115.0 | <1 | 2261.4 | <1 | <1 |
| CEBPG | CEBPG | 61.4 | 170.3 | <1 | <1 | <1 | <1 |
| CEBP | CEBPA | 28.5 | 399.8 | <1 | 76.8 | 29.2 | 24.0 |
| CEBP | CEBPE | 290.3 | 363.1 | <1 | 29.2 | 63.8 | 69.2 |
| CREB | CREB1 | NB | NB | NB | NB | NB | NB |
| OASISA | CREB3 | NB | NB | 48.2 | NB | 1535.9 | 288.6 |
| OASISA | CREB3L3 | NB | NB | 2883.8 | NB | NB | NB |
| OASISB | CREB3L1 | NB | NB | >5000 | NB | NB | NB |
| CREBZF | CREBZF | NB | NB | 35.9 | 2007.2 | NB | NB |
| XBP1 | XBP1 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6B | NB | NB | NB | NB | NB | NB |
| NFIL3 | NFIL3 | NB | NB | NB | NB | NB | 262.9 |
| PAR | DBP | 2373.4 | NB | NB | NB | 129.3 | 117.3 |
| PAR | HLF | NB | NB | 721.2 | NB | NB | 28.8 |
| ATF2 | ATF2 | 6.7 | 236.7 | 7.7 | NB | <1 | 18.2 |
| JUN | JUN | <1 | <1 | 25.2 | NB | <1 | <1 |
| JUN | JUNB | <1 | 2.4 | 258.8 | NB | 5.6 | <1 |
| FOS | FOS | 386.3 | 604.5 | 1.5 | NB | 169.2 | 1001.3 |
| FOS | FOSL1 | 604.5 | 3739.7 | 42.6 | NB | 227.9 | 1996.2 |
| ATF4 | ATF4 | 1.5 | 42.6 | 186.0 | NB | <1 | 10.1 |
| ATF4 | ATF5 | NB | NB | NB | NB | 3641.2 | 183.4 |
| ATF3 | ATF3 | 169.2 | 227.9 | <1 | 3641.2 | 113.6 | 16.2 |
| BATF | BATF | 1001.3 | 1996.2 | 10.1 | 183.4 | 16.2 | 104.6 |
| BATF | BATF2 | 4824.7 | >5000 | 9.8 | 1601.0 | 568.2 | 164.5 |
| BATF | BATF3 | 382.0 | 381.9 | 1.3 | 383.3 | 2.0 | 42.2 |
| SMAF | MAFF | 257.3 | NB | 192.9 | NB | 22.6 | 2493.7 |
| SMAF | MAFG | 3608.1 | NB | 215.4 | NB | 120.5 | NB |
| LMAF | MAF | 185.1 | 202.5 | 1.7 | NB | NB | NB |
| LMAF | MAFB | 60.5 | 151.6 | 84.8 | NB | 485.1 | 611.8 |
| NFE2 | NFE2 | NB | NB | 65.6 | 2900.8 | NB | NB |
| NFE2 | NFE2L1 | 1917.5 | 2137.5 | 38.6 | NB | NB | 2118.9 |
| NFE2 | NFE2L2 | 3977.2 | NB | 12.2 | 1717.7 | NB | NB |
| NFE2 | NFE2L3 | NB | NB | 34.4 | NB | NB | NB |
| BACH | BACH1 | 468.9 | 1224.3 | 160.4 | NB | 1866.6 | 114.6 |
| BACH | BACH2 | 363.9 | 1745.7 | 452.6 | NB | 3371.9 | 199.6 |

| | Family | BATF | BATF | SMAF | SMAF | LMAF | LMAF |
|---|---|---|---|---|---|---|---|
| Family | Protein | BATF2 | BATF3 | MAFF | MAFG | MAF | MAFB |
| DDIT3 | DDIT3 | 5.6 | <1 | 168.3 | 447.5 | NB | NB |
| CEBPG | CEBPG | 1.5 | 9.6 | 1156.8 | NB | NB | NB |
| CEBP | CEBPA | 77.4 | 7.5 | NB | NB | NB | NB |
| CEBP | CEBPE | 115.7 | 45.7 | NB | NB | NB | NB |
| CREB | CREB1 | NB | 2453.2 | NB | NB | NB | NB |
| OASISA | CREB3 | 352.3 | 542.1 | NB | NB | NB | NB |
| OASISA | CREB3L3 | NB | NB | NB | NB | NB | NB |
| OASISB | CREB3L1 | >5000 | NB | NB | NB | NB | NB |
| CREBZF | CREBZF | NB | NB | NB | 4377.7 | NB | NB |
| XBP1 | XBP1 | NB | NB | NB | NB | NB | >5000 |
| ATF6 | ATF6 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6B | NB | NB | NB | NB | NB | NB |
| NFIL3 | NFIL3 | 1251.4 | 184.0 | 50.3 | 739.4 | NB | NB |
| PAR | DBP | 457.0 | 43.3 | 4403.3 | NB | NB | NB |
| PAR | HLF | 779.8 | 92.1 | NB | NB | NB | NB |
| ATF2 | ATF2 | 635.7 | 28.2 | NB | NB | NB | NB |
| JUN | JUN | <1 | <1 | NB | 4149.9 | NB | NB |
| JUN | JUNB | 2.1 | <1 | >5000 | 3624.5 | NB | NB |
| FOS | FOS | 4824.7 | 382.0 | 257.3 | 3608.1 | 185.1 | 60.5 |
| FOS | FOSL1 | >5000 | 381.9 | NB | NB | 202.5 | 151.6 |
| ATF4 | ATF4 | 9.8 | 1.3 | 192.9 | 215.4 | 1.7 | 84.8 |
| ATF4 | ATF5 | 1601.0 | 383.3 | NB | NB | NB | NB |
| ATF3 | ATF3 | 568.2 | 2.0 | 22.6 | 120.5 | NB | 485.1 |
| BATF | BATF | 164.5 | 42.2 | 2493.7 | NB | NB | 611.8 |
| BATF | BATF2 | 2359.5 | 303.7 | 43.3 | 1039.6 | NB | NB |
| BATF | BATF3 | 303.7 | 88.2 | 9.8 | 46.1 | NB | NB |
| SMAF | MAFF | 43.3 | 9.8 | 17.3 | 5.1 | NB | NB |
| SMAF | MAFG | 1039.6 | 46.1 | 5.1 | 13.7 | NB | NB |
| LMAF | MAF | NB | NB | NB | NB | <1 | 2.5 |
| LMAF | MAFB | NB | NB | NB | NB | 2.5 | 39.4 |
| NFE2 | NFE2 | NB | NB | 475.2 | 692.4 | NB | NB |
| NFE2 | NFE2L1 | 109.6 | 3264.6 | <1 | <1 | NB | NB |
| NFE2 | NFE2L2 | NB | NB | 9.8 | 1.5 | NB | NB |
| NFE2 | NFE2L3 | 4802.8 | NB | <1 | <1 | NB | NB |
| BACH | BACH1 | 26.9 | 37.9 | 1.3 | <1 | 187.7 | 114.9 |
| BACH | BACH2 | 32.7 | 65.7 | <1 | <1 | 1307.1 | 371.6 |

| | Family | NFE2 | NFE2 | NFE2 | NFE2 | BACH | BACH |
|---|---|---|---|---|---|---|---|
| Family | Protein | NFE2 | NFE2L1 | NFE2L2 | NFE2L3 | BACH1 | BACH2 |
| DDIT3 | DDIT3 | 3271.5 | 2898.0 | NB | NB | 59.5 | 79.2 |
| CEBPG | CEBPG | NB | 1738.2 | 1637.0 | 1124.0 | NB | 2290.5 |
| CEBP | CEBPA | NB | NB | NB | 4696.3 | NB | NB |
| CEBP | CEBPE | NB | NB | 2368.5 | NB | NB | 2491.4 |
| CREB | CREB1 | NB | NB | NB | NB | 3259.0 | NB |
| OASISA | CREB3 | 2433.2 | NB | 439.5 | 191.4 | NB | 2634.8 |
| OASISA | CREB3L3 | NB | NB | NB | NB | NB | NB |
| OASISB | CREB3L1 | NB | NB | 1675.7 | NB | NB | NB |
| CREBZF | CREBZF | 167.2 | 33.8 | 86.3 | 1922.9 | 2576.1 | 458.1 |
| XBP1 | XBP1 | 2792.3 | 1903.1 | NB | NB | NB | 1887.3 |
| ATF6 | ATF6 | NB | NB | NB | NB | NB | NB |
| ATF6 | ATF6B | 2826.2 | NB | NB | NB | NB | 2313.7 |
| NFIL3 | NFIL3 | NB | 1751.8 | NB | 1198.9 | 586.7 | 2408.6 |
| PAR | DBP | >5000 | 3953.2 | NB | 3278.5 | 4238.2 | 3457.4 |
| PAR | HLF | NB | NB | NB | NB | NB | NB |
| ATF2 | ATF2 | NB | NB | NB | NB | 60.4 | 667.6 |
| JUN | JUN | NB | NB | NB | NB | NB | 4461.0 |
| JUN | JUNB | NB | NB | NB | NB | NB | NB |
| FOS | FOS | NB | 1917.5 | 3977.2 | NB | 468.9 | 363.9 |
| FOS | FOSL1 | NB | 2137.5 | NB | NB | 1224.3 | 1745.7 |
| ATF4 | ATF4 | 65.6 | 38.6 | 12.2 | 34.4 | 160.4 | 452.6 |
| ATF4 | ATF5 | 2900.8 | NB | 1717.7 | NB | NB | NB |
| ATF3 | ATF3 | NB | NB | NB | NB | 1866.6 | 3371.9 |
| BATF | BATF | NB | 2118.9 | NB | NB | 114.6 | 199.6 |
| BATF | BATF2 | NB | 109.6 | NB | 4802.8 | 26.9 | 32.7 |
| BATF | BATF3 | NB | 3264.6 | NB | NB | 37.9 | 65.7 |
| SMAF | MAFF | 475.2 | <1 | 9.8 | <1 | 1.3 | <1 |
| SMAF | MAFG | 692.4 | <1 | 1.5 | <1 | <1 | <1 |
| LMAF | MAF | NB | NB | NB | NB | 187.7 | 1307.1 |
| LMAF | MAFB | NB | NB | NB | NB | 114.9 | 371.6 |
| NFE2 | NFE2 | 212.4 | 2505.4 | 240.9 | 17.9 | NB | NB |
| NFE2 | NFE2L1 | 2505.4 | 25.3 | 1466.9 | 64.7 | NB | 84.4 |
| NFE2 | NFE2L2 | 240.9 | 1466.9 | 2212.0 | 103.9 | NB | NB |
| NFE2 | NFE2L3 | 17.9 | 64.7 | 103.9 | 10.9 | NB | 458.0 |
| BACH | BACH1 | NB | NB | NB | NB | NB | NB |
| BACH | BACH2 | NB | 84.4 | NB | 458.0 | NB | 548.8 |

*C. intestinalis* bZIP interactions.

| | Family | CEBPG | CEBP | CEBP | CEBP | CREB | OASISA |
|---|---|---|---|---|---|---|---|
| Family | Protein | CI1 | CI2 | CI3 | CI4 | CI5 | CI6 |
| CEBPG | CI1 | NB | 37.5 | >5000 | 7.2 | NB | NB |
| CEBP | CI2 | 37.5 | <1 | <1 | <1 | NB | NB |
| CEBP | CI3 | >5000 | <1 | 2967.3 | NB | NB | NB |
| CEBP | CI4 | 7.2 | <1 | NB | NB | NB | NB |
| CREB | CI5 | NB | NB | NB | NB | 10.9 | NB |
| OASISA | CI6 | NB | NB | NB | NB | NB | 1476.3 |
| OASISB | CI7 | NB | NB | NB | >5000 | NB | 359.1 |
| XBP1 | CI8 | NB | NB | NB | NB | NB | NB |
| XBP1 | CI9 | NB | NB | 33.1 | NB | NB | 1243.0 |
| XBP1 | CI10 | NB | NB | NB | 19.4 | NB | 94.7 |
| XBP1 | CI11 | NB | NB | NB | NB | NB | NB |
| ATF6 | CI12 | NB | NB | NB | NB | NB | NB |
| NFIL3 | CI13 | NB | NB | NB | 3171.0 | NB | NB |
| PAR | CI14 | NB | NB | 1617.7 | NB | NB | NB |
| ATF2 | CI15 | 184.4 | 62.4 | NB | 97.9 | NB | NB |
| JUN | CI16 | NB | NB | 3381.8 | NB | NB | NB |
| FOS | CI17 | NB | NB | NB | 2816.3 | NB | NB |
| FOS | CI18 | 4964.7 | NB | NB | 1726.3 | NB | 145.5 |
| ATF4 | CI19 | 10.2 | 552.2 | 2283.9 | NB | NB | NB |
| ATF3 | CI20 | 6.6 | 9.1 | 1852.8 | 23.2 | NB | NB |
| SMAF | CI21 | NB | NB | NB | NB | NB | NB |
| LMAF | CI22 | NB | NB | 1620.4 | NB | NB | NB |
| NFE2 | CI23 | NB | NB | NB | NB | NB | NB |
| BACH1 | CI24 | NB | NB | 3706.2 | NB | NB | NB |
| NOVEL | CI25 | NB | NB | NB | NB | NB | NB |
| NOVEL | CI26 | NB | NB | NB | NB | NB | NB |

| | Family | OASISB | XBP1 | XBP1 | XBP1 | XBP1 | ATF6 |
|---|---|---|---|---|---|---|---|
| Family | Protein | CI7 | CI8 | CI9 | CI10 | CI11 | CI12 |
| CEBPG | CI1 | NB | NB | NB | NB | NB | NB |
| CEBP | CI2 | NB | NB | NB | NB | NB | NB |
| CEBP | CI3 | NB | NB | 33.1 | NB | NB | NB |
| CEBP | CI4 | >5000 | NB | NB | 19.4 | NB | NB |
| CREB | CI5 | NB | NB | NB | NB | NB | NB |
| OASISA | CI6 | 359.1 | NB | 1243.0 | 94.7 | NB | NB |
| OASISB | CI7 | 144.9 | NB | NB | 556.2 | NB | NB |
| XBP1 | CI8 | NB | 74.5 | NB | 4.7 | 9.2 | NB |
| XBP1 | CI9 | NB | NB | 93.9 | 1481.2 | NB | 9.5 |
| XBP1 | CI10 | 556.2 | 4.7 | 1481.2 | 4.9 | 812.9 | 3294.7 |
| XBP1 | CI11 | NB | 9.2 | NB | 812.9 | 16.4 | NB |
| ATF6 | CI12 | NB | NB | 9.5 | 3294.7 | NB | 7.9 |
| NFIL3 | CI13 | NB | NB | 3193.5 | 169.7 | NB | 3480.2 |
| PAR | CI14 | NB | NB | NB | 47.6 | NB | >5000 |
| ATF2 | CI15 | NB | NB | NB | 1030.7 | NB | NB |
| JUN | CI16 | NB | NB | NB | NB | NB | NB |
| FOS | CI17 | NB | NB | 4386.6 | NB | NB | NB |
| FOS | CI18 | NB | NB | 31.3 | NB | 171.8 | NB |
| ATF4 | CI19 | NB | NB | NB | 468.7 | NB | NB |
| ATF3 | CI20 | 2777.9 | 3247.4 | 71.7 | NB | NB | 4368.6 |
| SMAF | CI21 | NB | NB | NB | 266.2 | NB | NB |
| LMAF | CI22 | NB | NB | NB | 39.0 | NB | 4727.0 |
| NFE2 | CI23 | NB | NB | NB | NB | 4589.4 | NB |
| BACH1 | CI24 | NB | NB | NB | 97.3 | NB | NB |
| NOVEL | CI25 | NB | NB | NB | 968.3 | NB | NB |
| NOVEL | CI26 | NB | NB | NB | 1090.2 | NB | NB |

| | Family | NFIL3 | PAR | ATF2 | JUN | FOS | FOS |
|---|---|---|---|---|---|---|---|
| Family | Protein | CI13 | CI14 | CI15 | CI16 | CI17 | CI18 |
| CEBPG | CI1 | NB | NB | 184.4 | NB | NB | 4964.7 |
| CEBP | CI2 | NB | NB | 62.4 | NB | NB | NB |
| CEBP | CI3 | NB | 1617.7 | NB | 3381.8 | NB | NB |
| CEBP | CI4 | 3171.0 | NB | 97.9 | NB | 2816.3 | 1726.3 |
| CREB | CI5 | NB | NB | NB | NB | NB | NB |
| OASISA | CI6 | NB | NB | NB | NB | NB | 145.5 |
| OASISB | CI7 | NB | NB | NB | NB | NB | NB |
| XBP1 | CI8 | NB | NB | NB | NB | NB | NB |
| XBP1 | CI9 | 3193.5 | NB | NB | NB | 4386.6 | 31.3 |
| XBP1 | CI10 | 169.7 | 47.6 | 1030.7 | NB | NB | NB |
| XBP1 | CI11 | NB | NB | NB | NB | NB | 171.8 |
| ATF6 | CI12 | 3480.2 | >5000 | NB | NB | NB | NB |
| NFIL3 | CI13 | 29.7 | 110.9 | NB | NB | 1443.4 | 2182.8 |
| PAR | CI14 | 110.9 | 12.5 | 320.7 | 1120.6 | NB | 973.7 |
| ATF2 | CI15 | NB | 320.7 | 22.3 | 52.3 | 25.0 | 41.1 |
| JUN | CI16 | NB | 1120.6 | 52.3 | NB | 2301.4 | 76.8 |
| FOS | CI17 | 1443.4 | NB | 25.0 | 2301.4 | NB | NB |
| FOS | CI18 | 2182.8 | 973.7 | 41.1 | 76.8 | NB | NB |
| ATF4 | CI19 | NB | NB | 2342.7 | NB | NB | NB |
| ATF3 | CI20 | 910.6 | 2482.5 | <1 | <1 | NB | 2591.0 |
| SMAF | CI21 | NB | 102.6 | NB | NB | NB | NB |
| LMAF | CI22 | NB | NB | NB | NB | NB | NB |
| NFE2 | CI23 | NB | NB | 1482.6 | NB | NB | NB |
| BACH1 | CI24 | NB | 43.2 | NB | NB | NB | NB |
| NOVEL | CI25 | 3045.4 | 1977.1 | NB | NB | NB | NB |
| NOVEL | CI26 | 709.1 | 3330.4 | NB | NB | NB | NB |

| | Family | ATF4 | ATF3 | SMAF | LMAF | NFE2 | BACH1 |
|---|---|---|---|---|---|---|---|
| Family | Protein | CI19 | CI20 | CI21 | CI22 | CI23 | CI24 |
| CEBPG | CI1 | 10.2 | 6.6 | NB | NB | NB | NB |
| CEBP | CI2 | 552.2 | 9.1 | NB | NB | NB | NB |
| CEBP | CI3 | 2283.9 | 1852.8 | NB | 1620.4 | NB | 3706.2 |
| CEBP | CI4 | NB | 23.2 | NB | NB | NB | NB |
| CREB | CI5 | NB | NB | NB | NB | NB | NB |
| OASISA | CI6 | NB | NB | NB | NB | NB | NB |
| OASISB | CI7 | NB | 2777.9 | NB | NB | NB | NB |
| XBP1 | CI8 | NB | 3247.4 | NB | NB | NB | NB |
| XBP1 | CI9 | NB | 71.7 | NB | NB | NB | NB |
| XBP1 | CI10 | 468.7 | NB | 266.2 | 39.0 | NB | 97.3 |
| XBP1 | CI11 | NB | NB | NB | NB | 4589.4 | NB |
| ATF6 | CI12 | NB | 4368.6 | NB | 4727.0 | NB | NB |
| NFIL3 | CI13 | NB | 910.6 | NB | NB | NB | NB |
| PAR | CI14 | NB | 2482.5 | 102.6 | NB | NB | 43.2 |
| ATF2 | CI15 | 2342.7 | <1 | NB | NB | 1482.6 | NB |
| JUN | CI16 | NB | <1 | NB | NB | NB | NB |
| FOS | CI17 | NB | NB | NB | NB | NB | NB |
| FOS | CI18 | NB | 2591.0 | NB | NB | NB | NB |
| ATF4 | CI19 | 1487.2 | 238.1 | NB | NB | NB | NB |
| ATF3 | CI20 | 238.1 | 607.7 | NB | NB | 2478.5 | NB |
| SMAF | CI21 | NB | NB | 111.9 | NB | 36.3 | 52.4 |
| LMAF | CI22 | NB | NB | NB | 120.0 | NB | NB |
| NFE2 | CI23 | NB | 2478.5 | 36.3 | NB | NB | >5000 |
| BACH1 | CI24 | NB | NB | 52.4 | NB | >5000 | NB |
| NOVEL | CI25 | NB | NB | NB | NB | NB | NB |
| NOVEL | CI26 | NB | NB | NB | NB | NB | 1794.5 |

| | Family | NOVEL | NOVEL |
|---|---|---|---|
| Family | Protein | CI25 | CI26 |
| CEBPG | CI1 | NB | NB |
| CEBP | CI2 | NB | NB |
| CEBP | CI3 | NB | NB |
| CEBP | CI4 | NB | NB |
| CREB | CI5 | NB | NB |
| OASISA | CI6 | NB | NB |
| OASISB | CI7 | NB | NB |
| XBP1 | CI8 | NB | NB |
| XBP1 | CI9 | NB | NB |
| XBP1 | CI10 | 968.3 | 1090.2 |
| XBP1 | CI11 | NB | NB |
| ATF6 | CI12 | NB | NB |
| NFIL3 | CI13 | 3045.4 | 709.1 |
| PAR | CI14 | 1977.1 | 3330.4 |
| ATF2 | CI15 | NB | NB |
| JUN | CI16 | NB | NB |
| FOS | CI17 | NB | NB |
| FOS | CI18 | NB | NB |
| ATF4 | CI19 | NB | NB |
| ATF3 | CI20 | NB | NB |
| SMAF | CI21 | NB | NB |
| LMAF | CI22 | NB | NB |
| NFE2 | CI23 | NB | NB |
| BACH1 | CI24 | NB | 1794.5 |
| NOVEL | CI25 | 298.4 | NB |
| NOVEL | CI26 | NB | 2782.5 |

*D. melanogaster* bZIP interactions.

| | Family | CEBPG | CEBP | CREB | OASISB | XBP1 | ATF6 |
|---|---|---|---|---|---|---|---|
| Family | Protein | DM10 | DM8 | DM9 | DM4 | DM18 | DM15 |
| CEBPG | DM10 | NB | NB | NB | NB | NB | NB |
| CEBP | DM8 | NB | 221.3 | NB | NB | NB | NB |
| CREB | DM9 | NB | NB | 5.8 | NB | NB | NB |
| OASISB | DM4 | NB | NB | NB | <1 | NB | NB |
| XBP1 | DM18 | NB | NB | NB | NB | 13.0 | 2346.0 |
| ATF6 | DM15 | NB | NB | NB | NB | 2346.0 | 80.1 |
| NFIL3 | DM16 | NB | NB | NB | NB | NB | NB |
| PAR | DM1 | NB | NB | NB | NB | NB | 3559.6 |
| PAR | DM7 | NB | <1 | NB | NB | 2650.5 | 685.5 |
| PAR | DM17 | NB | NB | 281.7 | NB | NB | 2520.1 |
| PAR | DM21 | NB | NB | NB | NB | NB | 2157.1 |
| ATF2 | DM5 | NB | NB | NB | NB | NB | NB |
| JUN | DM14 | NB | 2517.3 | NB | NB | NB | 1474.7 |
| FOS | DM3 | NB | NB | NB | NB | NB | 2326.5 |
| ATF4 | DM6 | <1 | 1.2 | NB | NB | 274.4 | 2591.1 |
| ATF3 | DM11 | NB | NB | NB | NB | NB | NB |
| SMAF | DM19 | NB | NB | NB | NB | NB | 2860.4 |
| LMAF | DM13 | NB | NB | NB | NB | 594.8 | 4511.8 |
| NFE2 | DM12 | NB | NB | NB | NB | NB | 1634.7 |
| NOVEL | DM28 | 227.0 | NB | NB | NB | NB | NB |
| NOVEL | DM20 | NB | 3063.6 | NB | NB | NB | 1642.9 |
| NOVEL | DM22 | NB | NB | NB | NB | 254.7 | 1218.4 |
| NOVEL | DM29 | NB | NB | NB | NB | NB | >5000 |
| NOVEL | DM23 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM24 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM26 | NB | 2822.5 | NB | NB | 810.1 | 392.6 |
| NOVEL | DM27 | NB | NB | NB | NB | 3573.2 | 186.8 |
| NOVEL | DM2 | NB | NB | NB | NB | NB | NB |

| | Family | NFIL3 | PAR | PAR | PAR | PAR | ATF2 |
|---|---|---|---|---|---|---|---|
| Family | Protein | DM16 | DM1 | DM7 | DM17 | DM21 | DM5 |
| CEBPG | DM10 | NB | NB | NB | NB | NB | NB |
| CEBP | DM8 | NB | NB | <1 | NB | NB | NB |
| CREB | DM9 | NB | NB | NB | 281.7 | NB | NB |
| OASISB | DM4 | NB | NB | NB | NB | NB | NB |
| XBP1 | DM18 | NB | NB | 2650.5 | NB | NB | NB |
| ATF6 | DM15 | NB | 3559.6 | 685.5 | 2520.1 | 2157.1 | NB |
| NFIL3 | DM16 | 35.7 | NB | NB | NB | NB | NB |
| PAR | DM1 | NB | 8.9 | NB | NB | NB | NB |
| PAR | DM7 | NB | NB | 34.8 | 1686.0 | 2964.4 | NB |
| PAR | DM17 | NB | NB | 1686.0 | 59.5 | 528.0 | NB |
| PAR | DM21 | NB | NB | 2964.4 | 528.0 | 264.2 | NB |
| ATF2 | DM5 | NB | NB | NB | NB | NB | 8.9 |
| JUN | DM14 | NB | NB | 290.2 | 869.3 | NB | NB |
| FOS | DM3 | NB | NB | NB | 2133.6 | 1213.1 | NB |
| ATF4 | DM6 | NB | NB | NB | 2307.7 | 2740.5 | NB |
| ATF3 | DM11 | NB | 4020.8 | NB | NB | NB | NB |
| SMAF | DM19 | NB | NB | 2474.6 | NB | NB | NB |
| LMAF | DM13 | NB | NB | NB | NB | 4636.2 | NB |
| NFE2 | DM12 | NB | NB | NB | 2923.0 | 885.9 | NB |
| NOVEL | DM28 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM20 | NB | NB | 1814.4 | 706.2 | 851.1 | NB |
| NOVEL | DM22 | NB | 3290.4 | 526.8 | NB | NB | NB |
| NOVEL | DM29 | NB | NB | NB | NB | 2879.1 | NB |
| NOVEL | DM23 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM24 | NB | NB | NB | 1220.2 | 3044.2 | NB |
| NOVEL | DM26 | 105.5 | NB | 175.9 | 78.7 | 145.0 | NB |
| NOVEL | DM27 | NB | NB | 313.1 | 506.0 | 428.0 | NB |
| NOVEL | DM2 | NB | NB | NB | NB | NB | NB |

| | Family | JUN | FOS | ATF4 | ATF3 | SMAF | LMAF |
|---|---|---|---|---|---|---|---|
| Family | Protein | DM14 | DM3 | DM6 | DM11 | DM19 | DM13 |
| CEBPG | DM10 | NB | NB | <1 | NB | NB | NB |
| CEBP | DM8 | 2517.3 | NB | 1.2 | NB | NB | NB |
| CREB | DM9 | NB | NB | NB | NB | NB | NB |
| OASISB | DM4 | NB | NB | NB | NB | NB | NB |
| XBP1 | DM18 | NB | NB | 274.4 | NB | NB | 594.8 |
| ATF6 | DM15 | 1474.7 | 2326.5 | 2591.1 | NB | 2860.4 | 4511.8 |
| NFIL3 | DM16 | NB | NB | NB | NB | NB | NB |
| PAR | DM1 | NB | NB | NB | 4020.8 | NB | NB |
| PAR | DM7 | 290.2 | NB | NB | NB | 2474.6 | NB |
| PAR | DM17 | 869.3 | 2133.6 | 2307.7 | NB | NB | NB |
| PAR | DM21 | NB | 1213.1 | 2740.5 | NB | NB | 4636.2 |
| ATF2 | DM5 | NB | NB | NB | NB | NB | NB |
| JUN | DM14 | NB | 8.4 | NB | 449.2 | NB | NB |
| FOS | DM3 | 8.4 | NB | NB | NB | NB | NB |
| ATF4 | DM6 | NB | NB | 3301.4 | NB | NB | NB |
| ATF3 | DM11 | 449.2 | NB | NB | NB | NB | NB |
| SMAF | DM19 | NB | NB | NB | NB | NB | NB |
| LMAF | DM13 | NB | NB | NB | NB | NB | 54.8 |
| NFE2 | DM12 | NB | 3873.9 | 3059.8 | NB | 4159.7 | NB |
| NOVEL | DM28 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM20 | NB | 3575.1 | 2825.7 | NB | NB | NB |
| NOVEL | DM22 | 259.6 | NB | NB | NB | NB | NB |
| NOVEL | DM29 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM23 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM24 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM26 | NB | 152.8 | 2231.2 | NB | 1554.0 | NB |
| NOVEL | DM27 | NB | 1962.3 | 1871.4 | NB | 2547.5 | NB |
| NOVEL | DM2 | NB | NB | NB | NB | NB | NB |

| | Family | NFE2 | NOVEL | NOVEL | NOVEL | NOVEL | NOVEL |
|---|---|---|---|---|---|---|---|
| Family | Protein | DM12 | DM28 | DM20 | DM22 | DM29 | DM23 |
| CEBPG | DM10 | NB | 227.0 | NB | NB | NB | NB |
| CEBP | DM8 | NB | NB | 3063.6 | NB | NB | NB |
| CREB | DM9 | NB | NB | NB | NB | NB | NB |
| OASISB | DM4 | NB | NB | NB | NB | NB | NB |
| XBP1 | DM18 | NB | NB | NB | 254.7 | NB | NB |
| ATF6 | DM15 | 1634.7 | NB | 1642.9 | 1218.4 | >5000 | NB |
| NFIL3 | DM16 | NB | NB | NB | NB | NB | NB |
| PAR | DM1 | NB | NB | NB | 3290.4 | NB | NB |
| PAR | DM7 | NB | NB | 1814.4 | 526.8 | NB | NB |
| PAR | DM17 | 2923.0 | NB | 706.2 | NB | NB | NB |
| PAR | DM21 | 885.9 | NB | 851.1 | NB | 2879.1 | NB |
| ATF2 | DM5 | NB | NB | NB | NB | NB | NB |
| JUN | DM14 | NB | NB | NB | 259.6 | NB | NB |
| FOS | DM3 | 3873.9 | NB | 3575.1 | NB | NB | NB |
| ATF4 | DM6 | 3059.8 | NB | 2825.7 | NB | NB | NB |
| ATF3 | DM11 | NB | NB | NB | NB | NB | NB |
| SMAF | DM19 | 4159.7 | NB | NB | NB | NB | NB |
| LMAF | DM13 | NB | NB | NB | NB | NB | NB |
| NFE2 | DM12 | 2659.8 | NB | 1344.4 | NB | NB | NB |
| NOVEL | DM28 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM20 | 1344.4 | NB | 1165.2 | NB | NB | NB |
| NOVEL | DM22 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM29 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM23 | NB | NB | NB | NB | NB | NB |
| NOVEL | DM24 | NB | NB | 1354.5 | NB | NB | NB |
| NOVEL | DM26 | 93.6 | NB | 390.0 | 132.9 | 33.4 | 70.2 |
| NOVEL | DM27 | 791.5 | NB | 847.0 | 803.9 | 3051.8 | NB |
| NOVEL | DM2 | NB | NB | NB | NB | NB | NB |

| | Family | NOVEL | NOVEL | NOVEL | NOVEL |
|---|---|---|---|---|---|
| Family | Protein | DM24 | DM26 | DM27 | DM2 |
| CEBPG | DM10 | NB | NB | NB | NB |
| CEBP | DM8 | NB | 2822.5 | NB | NB |
| CREB | DM9 | NB | NB | NB | NB |
| OASISB | DM4 | NB | NB | NB | NB |
| XBP1 | DM18 | NB | 810.1 | 3573.2 | NB |
| ATF6 | DM15 | NB | 392.6 | 186.8 | NB |
| NFIL3 | DM16 | NB | 105.5 | NB | NB |
| PAR | DM1 | NB | NB | NB | NB |
| PAR | DM7 | NB | 175.9 | 313.1 | NB |
| PAR | DM17 | 1220.2 | 78.7 | 506.0 | NB |
| PAR | DM21 | 3044.2 | 145.0 | 428.0 | NB |
| ATF2 | DM5 | NB | NB | NB | NB |
| JUN | DM14 | NB | NB | NB | NB |
| FOS | DM3 | NB | 152.8 | 1962.3 | NB |
| ATF4 | DM6 | NB | 2231.2 | 1871.4 | NB |
| ATF3 | DM11 | NB | NB | NB | NB |
| SMAF | DM19 | NB | 1554.0 | 2547.5 | NB |
| LMAF | DM13 | NB | NB | NB | NB |
| NFE2 | DM12 | NB | 93.6 | 791.5 | NB |
| NOVEL | DM28 | NB | NB | NB | NB |
| NOVEL | DM20 | 1354.5 | 390.0 | 847.0 | NB |
| NOVEL | DM22 | NB | 132.9 | 803.9 | NB |
| NOVEL | DM29 | NB | 33.4 | 3051.8 | NB |
| NOVEL | DM23 | NB | 70.2 | NB | NB |
| NOVEL | DM24 | NB | 304.2 | 1861.9 | NB |
| NOVEL | DM26 | 304.2 | 14.4 | 25.9 | NB |
| NOVEL | DM27 | 1861.9 | 25.9 | 94.0 | NB |
| NOVEL | DM2 | NB | NB | NB | NB |

*C. elegans* bZIP interactions.

| | Family | CEPBG | CEBPA | CEBPA | CREB | OASISA | OASISB |
|---|---|---|---|---|---|---|---|
| Family | Protein | CE18 | CE1 | CE24 | CE9 | CE13 | CE21 |
| CEPBG | CE18 | 2.3 | NB | NB | NB | NB | NB |
| CEBPA | CE1 | NB | 65.6 | 13.3 | NB | NB | NB |
| CEBPA | CE24 | NB | 13.3 | 18.2 | NB | NB | NB |
| CREB | CE9 | NB | NB | NB | 21.5 | NB | NB |
| OASISA | CE13 | NB | NB | NB | NB | 82.5 | 136.0 |
| OASISB | CE21 | NB | NB | NB | NB | 136.0 | 62.1 |
| XBP1 | CE6 | NB | NB | NB | NB | NB | NB |
| ATF6 | CE10 | NB | NB | NB | NB | NB | NB |
| PAR | CE14 | <1 | NB | NB | NB | NB | 3307.1 |
| PAR | CE3 | 29.2 | NB | NB | NB | NB | NB |
| PAR | CE12 | 12.4 | NB | NB | NB | NB | 330.9 |
| PAR | CE23 | NB | NB | NB | NB | NB | NB |
| ATF2 | CE30 | NB | NB | NB | NB | NB | NB |
| ATF2 | CE5 | NB | NB | NB | NB | NB | NB |
| JUN | CE7 | 157.2 | NB | NB | NB | NB | NB |
| FOS | CE8 | 359.0 | NB | NB | NB | NB | NB |
| ATF4 | CE11 | <1 | NB | NB | NB | NB | NB |
| ATF4 | CE15 | <1 | NB | NB | NB | NB | NB |
| ATF4 | CE20 | <1 | NB | NB | NB | NB | NB |
| MAF | CE19 | NB | NB | 2624.0 | NB | NB | 3020.0 |
| NOVEL | CE2 | <1 | NB | NB | NB | NB | NB |
| NOVEL | CE16 | 73.5 | NB | NB | NB | NB | NB |
| NOVEL | CE22 | <1 | NB | NB | NB | NB | NB |
| NOVEL | CE28 | <1 | NB | NB | NB | NB | NB |
| NOVEL | CE17 | NB | NB | NB | NB | NB | NB |
| NOVEL | CE29 | NB | NB | NB | NB | NB | NB |

| | Family | XBP1 | ATF6 | PAR | PAR | PAR | PAR |
|---|---|---|---|---|---|---|---|
| Family | Protein | CE6 | CE10 | CE14 | CE3 | CE12 | CE23 |
| CEPBG | CE18 | NB | NB | <1 | 29.2 | 12.4 | NB |
| CEBPA | CE1 | NB | NB | NB | NB | NB | NB |
| CEBPA | CE24 | NB | NB | NB | NB | NB | NB |
| CREB | CE9 | NB | NB | NB | NB | NB | NB |
| OASISA | CE13 | NB | NB | NB | NB | NB | NB |
| OASISB | CE21 | NB | NB | 3307.1 | NB | 330.9 | NB |
| XBP1 | CE6 | 7.1 | 407.3 | NB | NB | NB | NB |
| ATF6 | CE10 | 407.3 | 18.2 | NB | NB | NB | NB |
| PAR | CE14 | NB | NB | 16.6 | <1 | 2.2 | 85.3 |
| PAR | CE3 | NB | NB | <1 | 13.4 | 69.0 | 41.5 |
| PAR | CE12 | NB | NB | 2.2 | 69.0 | 156.9 | 5.6 |
| PAR | CE23 | NB | NB | 85.3 | 41.5 | 5.6 | 160.5 |
| ATF2 | CE30 | 608.8 | NB | 1874.2 | NB | NB | NB |
| ATF2 | CE5 | NB | NB | NB | NB | NB | NB |
| JUN | CE7 | NB | NB | NB | NB | NB | NB |
| FOS | CE8 | NB | NB | NB | NB | NB | NB |
| ATF4 | CE11 | NB | NB | 171.5 | NB | 79.6 | NB |
| ATF4 | CE15 | NB | NB | 37.1 | 74.1 | 45.6 | NB |
| ATF4 | CE20 | NB | NB | 1285.8 | 241.2 | 208.6 | NB |
| MAF | CE19 | NB | NB | 79.1 | NB | NB | NB |
| NOVEL | CE2 | NB | NB | 75.7 | NB | 182.7 | NB |
| NOVEL | CE16 | NB | NB | NB | 3697.4 | NB | NB |
| NOVEL | CE22 | NB | NB | 187.6 | NB | 317.6 | NB |
| NOVEL | CE28 | NB | NB | 216.6 | NB | 1005.0 | NB |
| NOVEL | CE17 | NB | NB | NB | NB | NB | NB |
| NOVEL | CE29 | NB | >5000 | NB | NB | NB | NB |

| | Family | ATF2 | ATF2 | JUN | FOS | ATF4 | ATF4 |
|---|---|---|---|---|---|---|---|
| Family | Protein | CE30 | CE5 | CE7 | CE8 | CE11 | CE15 |
| CEPBG | CE18 | NB | NB | 157.2 | 359.0 | <1 | <1 |
| CEBPA | CE1 | NB | NB | NB | NB | NB | NB |
| CEBPA | CE24 | NB | NB | NB | NB | NB | NB |
| CREB | CE9 | NB | NB | NB | NB | NB | NB |
| OASISA | CE13 | NB | NB | NB | NB | NB | NB |
| OASISB | CE21 | NB | NB | NB | NB | NB | NB |
| XBP1 | CE6 | 608.8 | NB | NB | NB | NB | NB |
| ATF6 | CE10 | NB | NB | NB | NB | NB | NB |
| PAR | CE14 | 1874.2 | NB | NB | NB | 171.5 | 37.1 |
| PAR | CE3 | NB | NB | NB | NB | NB | 74.1 |
| PAR | CE12 | NB | NB | NB | NB | 79.6 | 45.6 |
| PAR | CE23 | NB | NB | NB | NB | NB | NB |
| ATF2 | CE30 | 63.4 | 131.7 | NB | NB | NB | NB |
| ATF2 | CE5 | 131.7 | 730.1 | NB | NB | NB | NB |
| JUN | CE7 | NB | NB | NB | 31.0 | NB | NB |
| FOS | CE8 | NB | NB | 31.0 | NB | NB | NB |
| ATF4 | CE11 | NB | NB | NB | NB | NB | 545.3 |
| ATF4 | CE15 | NB | NB | NB | NB | 545.3 | 601.1 |
| ATF4 | CE20 | NB | NB | NB | NB | NB | >5000 |
| MAF | CE19 | NB | 1773.3 | NB | NB | NB | NB |
| NOVEL | CE2 | NB | NB | NB | NB | 397.5 | 702.1 |
| NOVEL | CE16 | NB | NB | 24.4 | NB | NB | NB |
| NOVEL | CE22 | NB | NB | NB | 1659.6 | NB | 400.8 |
| NOVEL | CE28 | NB | NB | NB | 1504.0 | NB | 2878.3 |
| NOVEL | CE17 | NB | NB | NB | NB | NB | NB |
| NOVEL | CE29 | 517.1 | NB | NB | NB | NB | 262.7 |

| | Family | ATF4 | MAF | NOVEL | NOVEL | NOVEL | NOVEL |
|---|---|---|---|---|---|---|---|
| Family | Protein | CE20 | CE19 | CE2 | CE16 | CE22 | CE28 |
| CEPBG | CE18 | <1 | NB | <1 | 73.5 | <1 | <1 |
| CEBPA | CE1 | NB | NB | NB | NB | NB | NB |
| CEBPA | CE24 | NB | 2624.0 | NB | NB | NB | NB |
| CREB | CE9 | NB | NB | NB | NB | NB | NB |
| OASISA | CE13 | NB | NB | NB | NB | NB | NB |
| OASISB | CE21 | NB | 3020.0 | NB | NB | NB | NB |
| XBP1 | CE6 | NB | NB | NB | NB | NB | NB |
| ATF6 | CE10 | NB | NB | NB | NB | NB | NB |
| PAR | CE14 | 1285.8 | 79.1 | 75.7 | NB | 187.6 | 216.6 |
| PAR | CE3 | 241.2 | NB | NB | 3697.4 | NB | NB |
| PAR | CE12 | 208.6 | NB | 182.7 | NB | 317.6 | 1005.0 |
| PAR | CE23 | NB | NB | NB | NB | NB | NB |
| ATF2 | CE30 | NB | NB | NB | NB | NB | NB |
| ATF2 | CE5 | NB | 1773.3 | NB | NB | NB | NB |
| JUN | CE7 | NB | NB | NB | 24.4 | NB | NB |
| FOS | CE8 | NB | NB | NB | NB | 1659.6 | 1504.0 |
| ATF4 | CE11 | NB | NB | 397.5 | NB | NB | NB |
| ATF4 | CE15 | >5000 | NB | 702.1 | NB | 400.8 | 2878.3 |
| ATF4 | CE20 | NB | NB | 916.4 | NB | NB | NB |
| MAF | CE19 | NB | 4.5 | 257.1 | NB | NB | NB |
| NOVEL | CE2 | 916.4 | 257.1 | 52.5 | NB | NB | 1095.4 |
| NOVEL | CE16 | NB | NB | NB | NB | NB | NB |
| NOVEL | CE22 | NB | NB | NB | NB | NB | NB |
| NOVEL | CE28 | NB | NB | 1095.4 | NB | NB | NB |
| NOVEL | CE17 | NB | NB | NB | NB | NB | NB |
| NOVEL | CE29 | NB | NB | NB | NB | NB | NB |

| | Family | NOVEL | NOVEL |
|---|---|---|---|
| Family | Protein | CE17 | CE29 |
| CEPBG | CE18 | NB | NB |
| CEBPA | CE1 | NB | NB |
| CEBPA | CE24 | NB | NB |
| CREB | CE9 | NB | NB |
| OASISA | CE13 | NB | NB |
| OASISB | CE21 | NB | NB |
| XBP1 | CE6 | NB | NB |
| ATF6 | CE10 | NB | >5000 |
| PAR | CE14 | NB | NB |
| PAR | CE3 | NB | NB |
| PAR | CE12 | NB | NB |
| PAR | CE23 | NB | NB |
| ATF2 | CE30 | NB | 517.1 |
| ATF2 | CE5 | NB | NB |
| JUN | CE7 | NB | NB |
| FOS | CE8 | NB | NB |
| ATF4 | CE11 | NB | NB |
| ATF4 | CE15 | NB | 262.7 |
| ATF4 | CE20 | NB | NB |
| MAF | CE19 | NB | NB |
| NOVEL | CE2 | NB | NB |
| NOVEL | CE16 | NB | NB |
| NOVEL | CE22 | NB | NB |
| NOVEL | CE28 | NB | NB |
| NOVEL | CE17 | 6.0 | NB |
| NOVEL | CE29 | NB | 36.9 |

*N. vectensis* bZIP interactions.

| | Family | CEBPG | CEBPG | CREB | OASISA | OASISB | XBP1 |
|---|---|---|---|---|---|---|---|
| Family | Protein | NV5 | NV11 | NV19 | NV25 | NV2 | NV6 |
| CEBPG | NV5 | 2.8 | 18.2 | NB | NB | NB | NB |
| CEBPG | NV11 | 18.2 | 4452.7 | NB | NB | NB | NB |
| CREB | NV19 | NB | NB | 14.0 | NB | NB | NB |
| OASISA | NV25 | NB | NB | NB | 57.1 | 104.6 | NB |
| OASISB | NV2 | NB | NB | NB | 104.6 | 16.7 | NB |
| XBP1 | NV6 | NB | NB | NB | NB | NB | 31.4 |
| ATF6 | NV22 | NB | NB | NB | NB | NB | NB |
| PAR | NV1 | NB | NB | NB | NB | NB | 2543.4 |
| PAR | NV10 | <1 | NB | NB | NB | NB | NB |
| PAR | NV16 | 201.4 | 1902.8 | NB | NB | NB | NB |
| PAR | NV17 | 901.0 | NB | NB | NB | NB | NB |
| PAR | NV24 | 2095.9 | NB | NB | NB | NB | 2768.2 |
| PAR | NV29 | 1679.4 | NB | NB | NB | NB | NB |
| ATF2 | NV8 | 890.8 | 704.4 | NB | NB | NB | NB |
| JUN | NV14 | 1490.2 | 348.6 | NB | NB | NB | NB |
| JUN | NV15 | 1567.0 | 318.6 | NB | NB | NB | NB |
| JUN | NV28 | 4526.9 | 104.7 | NB | NB | NB | NB |
| FOS | NV3 | 59.9 | >5000 | NB | NB | NB | NB |
| FOS | NV7 | NB | NB | NB | NB | NB | 2635.1 |
| FOS | NV26 | 546.9 | 4727.7 | NB | NB | 419.9 | 571.4 |
| FOS | NV27 | NB | NB | NB | NB | NB | NB |
| FOS | NV30 | 1290.8 | NB | NB | NB | NB | NB |
| ATF4 | NV35 | <1 | NB | NB | NB | NB | NB |
| SMAF | NV13 | NB | NB | NB | NB | NB | NB |
| LMAF | NV12 | NB | 241.3 | NB | NB | NB | 3116.3 |
| LMAF | NV18 | NB | NB | NB | NB | NB | NB |
| LMAF | NV34 | NB | NB | NB | NB | NB | NB |
| NFE2 | NV21 | NB | NB | NB | NB | NB | NB |
| NOVEL | NV9 | NB | 797.8 | NB | NB | NB | NB |
| NOVEL | NV20 | 1934.5 | 803.7 | NB | NB | 2844.4 | 984.5 |
| NOVEL | NV23 | <1 | 136.2 | NB | 1724.2 | 16.6 | NB |

| | Family | ATF6 | PAR | PAR | PAR | PAR | PAR |
|---|---|---|---|---|---|---|---|
| Family | Protein | NV22 | NV1 | NV10 | NV16 | NV17 | NV24 |
| CEBPG | NV5 | NB | NB | <1 | 201.4 | 901.0 | 2095.9 |
| CEBPG | NV11 | NB | NB | NB | 1902.8 | NB | NB |
| CREB | NV19 | NB | NB | NB | NB | NB | NB |
| OASISA | NV25 | NB | NB | NB | NB | NB | NB |
| OASISB | NV2 | NB | NB | NB | NB | NB | NB |
| XBP1 | NV6 | NB | 2543.4 | NB | NB | NB | 2768.2 |
| ATF6 | NV22 | <1 | NB | NB | NB | NB | NB |
| PAR | NV1 | NB | 6.9 | NB | NB | NB | 2665.2 |
| PAR | NV10 | NB | NB | 9.9 | 23.7 | 187.2 | 485.5 |
| PAR | NV16 | NB | NB | 23.7 | 61.2 | <1 | NB |
| PAR | NV17 | NB | NB | 187.2 | <1 | 69.6 | NB |
| PAR | NV24 | NB | 2665.2 | 485.5 | NB | NB | 25.7 |
| PAR | NV29 | NB | NB | 877.6 | <1 | 6.1 | NB |
| ATF2 | NV8 | NB | NB | NB | 4166.7 | NB | 2099.1 |
| JUN | NV14 | NB | NB | 549.7 | 189.1 | 370.2 | NB |
| JUN | NV15 | NB | NB | 104.9 | 436.3 | 1805.0 | NB |
| JUN | NV28 | NB | NB | 112.0 | 1456.1 | 568.4 | NB |
| FOS | NV3 | NB | NB | 3431.1 | NB | NB | NB |
| FOS | NV7 | NB | 433.8 | 307.8 | NB | 661.6 | 613.5 |
| FOS | NV26 | NB | 2488.2 | NB | 699.4 | NB | 324.5 |
| FOS | NV27 | NB | NB | NB | NB | NB | NB |
| FOS | NV30 | NB | NB | NB | NB | NB | NB |
| ATF4 | NV35 | NB | NB | 14.9 | 198.9 | 98.2 | NB |
| SMAF | NV13 | NB | 375.5 | NB | >5000 | NB | NB |
| LMAF | NV12 | NB | 627.0 | NB | NB | NB | 1771.8 |
| LMAF | NV18 | NB | NB | NB | NB | NB | NB |
| LMAF | NV34 | NB | NB | NB | NB | NB | NB |
| NFE2 | NV21 | NB | NB | NB | NB | NB | NB |
| NOVEL | NV9 | NB | NB | NB | NB | NB | NB |
| NOVEL | NV20 | NB | 1166.5 | 866.3 | 4629.3 | NB | 695.4 |
| NOVEL | NV23 | NB | NB | <1 | 1.7 | 19.8 | 15.4 |

| | Family | PAR | ATF2 | JUN | JUN | JUN | FOS |
|---|---|---|---|---|---|---|---|
| Family | Protein | NV29 | NV8 | NV14 | NV15 | NV28 | NV3 |
| CEBPG | NV5 | 1679.4 | 890.8 | 1490.2 | 1567.0 | 4526.9 | 59.9 |
| CEBPG | NV11 | NB | 704.4 | 348.6 | 318.6 | 104.7 | >5000 |
| CREB | NV19 | NB | NB | NB | NB | NB | NB |
| OASISA | NV25 | NB | NB | NB | NB | NB | NB |
| OASISB | NV2 | NB | NB | NB | NB | NB | NB |
| XBP1 | NV6 | NB | NB | NB | NB | NB | NB |
| ATF6 | NV22 | NB | NB | NB | NB | NB | NB |
| PAR | NV1 | NB | NB | NB | NB | NB | NB |
| PAR | NV10 | 877.6 | NB | 549.7 | 104.9 | 112.0 | 3431.1 |
| PAR | NV16 | <1 | 4166.7 | 189.1 | 436.3 | 1456.1 | NB |
| PAR | NV17 | 6.1 | NB | 370.2 | 1805.0 | 568.4 | NB |
| PAR | NV24 | NB | 2099.1 | NB | NB | NB | NB |
| PAR | NV29 | 4.0 | NB | 842.4 | NB | NB | 1657.0 |
| ATF2 | NV8 | NB | 1.1 | 85.3 | 112.9 | 11.0 | NB |
| JUN | NV14 | 842.4 | 85.3 | 3745.7 | NB | NB | <1 |
| JUN | NV15 | NB | 112.9 | NB | NB | NB | <1 |
| JUN | NV28 | NB | 11.0 | NB | NB | NB | <1 |
| FOS | NV3 | 1657.0 | NB | <1 | <1 | <1 | NB |
| FOS | NV7 | NB | 73.2 | 17.0 | 1289.7 | 26.5 | 338.5 |
| FOS | NV26 | 286.6 | 1890.3 | <1 | 2.5 | <1 | NB |
| FOS | NV27 | NB | NB | 2.0 | 15.7 | 6.4 | NB |
| FOS | NV30 | >5000 | 2919.1 | 2.0 | <1 | 3.2 | NB |
| ATF4 | NV35 | 37.8 | 2673.2 | 8.7 | 10.1 | <1 | NB |
| SMAF | NV13 | 1130.1 | NB | NB | NB | NB | NB |
| LMAF | NV12 | NB | 2308.7 | 3786.3 | NB | NB | NB |
| LMAF | NV18 | NB | NB | NB | NB | NB | NB |
| LMAF | NV34 | NB | NB | NB | NB | NB | NB |
| NFE2 | NV21 | 293.3 | NB | NB | NB | NB | NB |
| NOVEL | NV9 | NB | NB | NB | NB | NB | NB |
| NOVEL | NV20 | 856.4 | 4474.3 | 2245.3 | 2983.5 | 1443.9 | NB |
| NOVEL | NV23 | 9.5 | 90.0 | 18.8 | 18.9 | 51.0 | 1731.6 |

| | Family | FOS | FOS | FOS | FOS | ATF4 | SMAF |
|---|---|---|---|---|---|---|---|
| Family | Protein | NV7 | NV26 | NV27 | NV30 | NV35 | NV13 |
| CEBPG | NV5 | NB | 546.9 | NB | 1290.8 | <1 | NB |
| CEBPG | NV11 | NB | 4727.7 | NB | NB | NB | NB |
| CREB | NV19 | NB | NB | NB | NB | NB | NB |
| OASISA | NV25 | NB | NB | NB | NB | NB | NB |
| OASISB | NV2 | NB | 419.9 | NB | NB | NB | NB |
| XBP1 | NV6 | 2635.1 | 571.4 | NB | NB | NB | NB |
| ATF6 | NV22 | NB | NB | NB | NB | NB | NB |
| PAR | NV1 | 433.8 | 2488.2 | NB | NB | NB | 375.5 |
| PAR | NV10 | 307.8 | NB | NB | NB | 14.9 | NB |
| PAR | NV16 | NB | 699.4 | NB | NB | 198.9 | >5000 |
| PAR | NV17 | 661.6 | NB | NB | NB | 98.2 | NB |
| PAR | NV24 | 613.5 | 324.5 | NB | NB | NB | NB |
| PAR | NV29 | NB | 286.6 | NB | >5000 | 37.8 | 1130.1 |
| ATF2 | NV8 | 73.2 | 1890.3 | NB | 2919.1 | 2673.2 | NB |
| JUN | NV14 | 17.0 | <1 | 2.0 | 2.0 | 8.7 | NB |
| JUN | NV15 | 1289.7 | 2.5 | 15.7 | <1 | 10.1 | NB |
| JUN | NV28 | 26.5 | <1 | 6.4 | 3.2 | <1 | NB |
| FOS | NV3 | 338.5 | NB | NB | NB | NB | NB |
| FOS | NV7 | NB | 1005.7 | NB | NB | 2173.5 | NB |
| FOS | NV26 | 1005.7 | 2771.3 | NB | NB | NB | NB |
| FOS | NV27 | NB | NB | NB | NB | NB | NB |
| FOS | NV30 | NB | NB | NB | NB | NB | NB |
| ATF4 | NV35 | 2173.5 | NB | NB | NB | 1994.5 | NB |
| SMAF | NV13 | NB | NB | NB | NB | NB | 279.0 |
| LMAF | NV12 | NB | 96.9 | NB | NB | NB | 69.9 |
| LMAF | NV18 | NB | 558.1 | NB | NB | NB | NB |
| LMAF | NV34 | NB | 808.0 | NB | NB | NB | NB |
| NFE2 | NV21 | NB | 1327.0 | NB | NB | NB | <1 |
| NOVEL | NV9 | NB | NB | NB | NB | NB | NB |
| NOVEL | NV20 | 503.6 | 3124.7 | NB | NB | NB | NB |
| NOVEL | NV23 | 228.8 | 2104.5 | 2492.2 | 499.7 | 15.2 | 292.0 |

| | Family | LMAF | LMAF | LMAF | NFE2 | NOVEL | NOVEL | NOVEL |
|---|---|---|---|---|---|---|---|---|
| Family | Protein | NV12 | NV18 | NV34 | NV21 | NV9 | NV20 | NV23 |
| CEBPG | NV5 | NB | NB | NB | NB | NB | 1934.5 | <1 |
| CEBPG | NV11 | 241.3 | NB | NB | NB | 797.8 | 803.7 | 136.2 |
| CREB | NV19 | NB | NB | NB | NB | NB | NB | NB |
| OASISA | NV25 | NB | NB | NB | NB | NB | NB | 1724.2 |
| OASISB | NV2 | NB | NB | NB | NB | NB | 2844.4 | 16.6 |
| XBP1 | NV6 | 3116.3 | NB | NB | NB | NB | 984.5 | NB |
| ATF6 | NV22 | NB | NB | NB | NB | NB | NB | NB |
| PAR | NV1 | 627.0 | NB | NB | NB | NB | 1166.5 | NB |
| PAR | NV10 | NB | NB | NB | NB | NB | 866.3 | <1 |
| PAR | NV16 | NB | NB | NB | NB | NB | 4629.3 | 1.7 |
| PAR | NV17 | NB | NB | NB | NB | NB | NB | 19.8 |
| PAR | NV24 | 1771.8 | NB | NB | NB | NB | 695.4 | 15.4 |
| PAR | NV29 | NB | NB | NB | 293.3 | NB | 856.4 | 9.5 |
| ATF2 | NV8 | 2308.7 | NB | NB | NB | NB | 4474.3 | 90.0 |
| JUN | NV14 | 3786.3 | NB | NB | NB | NB | 2245.3 | 18.8 |
| JUN | NV15 | NB | NB | NB | NB | NB | 2983.5 | 18.9 |
| JUN | NV28 | NB | NB | NB | NB | NB | 1443.9 | 51.0 |
| FOS | NV3 | NB | NB | NB | NB | NB | NB | 1731.6 |
| FOS | NV7 | NB | NB | NB | NB | NB | 503.6 | 228.8 |
| FOS | NV26 | 96.9 | 558.1 | 808.0 | 1327.0 | NB | 3124.7 | 2104.5 |
| FOS | NV27 | NB | NB | NB | NB | NB | NB | 2492.2 |
| FOS | NV30 | NB | NB | NB | NB | NB | NB | 499.7 |
| ATF4 | NV35 | NB | NB | NB | NB | NB | NB | 15.2 |
| SMAF | NV13 | 69.9 | NB | NB | <1 | NB | NB | 292.0 |
| LMAF | NV12 | 1.5 | NB | NB | 78.1 | NB | 126.3 | NB |
| LMAF | NV18 | NB | <1 | NB | NB | NB | 185.4 | NB |
| LMAF | NV34 | NB | NB | NB | NB | NB | 875.6 | 2675.0 |
| NFE2 | NV21 | 78.1 | NB | NB | NB | NB | NB | NB |
| NOVEL | NV9 | NB | NB | NB | NB | 204.9 | 3.6 | NB |
| NOVEL | NV20 | 126.3 | 185.4 | 875.6 | NB | 3.6 | 67.7 | 16.6 |
| NOVEL | NV23 | NB | NB | 2675.0 | NB | NB | 16.6 | 359.2 |

*M. brevicollis* bZIP interactions.

| Protein | MB2 | MB13 | MB16 | MB7 | MB3 | MB15 | MB10 | MB1 | MB9 | MB11 |
|---|---|---|---|---|---|---|---|---|---|---|
| MB2 | NB | NB | NB | NB | NB | NB | 12.6 | NB | NB | NB |
| MB13 | NB | 19.5 | NB | NB | NB | NB | NB | 1842.2 | NB | NB |
| MB16 | NB | NB | 411.0 | NB | NB | NB | NB | NB | NB | NB |
| MB7 | NB | NB | NB | 4072.3 | NB | NB | NB | NB | NB | NB |
| MB3 | NB | NB | NB | NB | 178.5 | 19.2 | NB | NB | NB | NB |
| MB15 | NB | NB | NB | NB | 19.2 | 20.3 | NB | NB | NB | NB |
| MB10 | 12.6 | NB | NB | NB | NB | NB | 2558.4 | 160.3 | NB | >5000 |
| MB1 | NB | 1842.2 | NB | NB | NB | NB | 160.3 | NB | NB | NB |
| MB9 | NB | NB | NB | NB | NB | NB | NB | NB | 240.2 | 42.4 |
| MB11 | NB | NB | NB | NB | NB | NB | >5000 | NB | 42.4 | NB |
| MB18 | NB | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB4 | NB | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB5 | NB | 2779.1 | NB | NB | NB | 2419.2 | NB | NB | NB | NB |
| MB6 | NB | NB | NB | NB | NB | NB | 33.5 | NB | NB | NB |
| MB8 | 1139.4 | NB | NB | NB | NB | 4607.4 | 2874.9 | NB | NB | NB |
| MB17 | 2797.3 | NB | NB | NB | NB | NB | 2114.4 | NB | 270.9 | >5000 |
| MB19 | NB | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB20 | 1080.5 | NB | NB | NB | NB | NB | 1855.4 | NB | NB | NB |
| MB21 | NB | NB | NB | NB | NB | NB | NB | NB | NB | NB |

| Protein | MB18 | MB4 | MB5 | MB6 | MB8 | MB17 | MB19 | MB20 | MB21 |
|---|---|---|---|---|---|---|---|---|---|
| MB2 | NB | NB | NB | NB | 1139.4 | 2797.3 | NB | 1080.5 | NB |
| MB13 | NB | NB | 2779.1 | NB | NB | NB | NB | NB | NB |
| MB16 | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB7 | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB3 | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB15 | NB | NB | 2419.2 | NB | 4607.4 | NB | NB | NB | NB |
| MB10 | NB | NB | NB | 33.5 | 2874.9 | 2114.4 | NB | 1855.4 | NB |
| MB1 | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB9 | NB | NB | NB | NB | NB | 270.9 | NB | NB | NB |
| MB11 | NB | NB | NB | NB | NB | >5000 | NB | NB | NB |
| MB18 | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| MB4 | NB | 21.6 | NB | NB | 1706.3 | NB | NB | NB | NB |
| MB5 | NB | NB | 32.3 | 3736.2 | 1402.2 | 428.6 | NB | NB | NB |
| MB6 | NB | NB | 3736.2 | 1432.9 | NB | 1086.0 | NB | NB | NB |
| MB8 | NB | 1706.3 | 1402.2 | NB | 454.1 | 553.1 | NB | 211.2 | NB |
| MB17 | NB | NB | 428.6 | 1086.0 | 553.1 | 57.8 | NB | 41.9 | 52.3 |
| MB19 | NB | NB | NB | NB | NB | NB | 416.5 | NB | NB |
| MB20 | NB | NB | NB | NB | 211.2 | 41.9 | NB | 70.2 | NB |
| MB21 | NB | NB | NB | NB | NB | 52.3 | NB | NB | NB |

*S. cerevisiae* bZIP interactions.

| Protein | SC1 | SC2 | SC4 | SC6 | SC7 | SC9 | SC10 | SC11 | SC12 | SC14 | SC15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SC1 | 515.7 | NB | NB | NB | NB | NB | NB | NB | NB | NB | NB |
| SC2 | NB | 2.8 | NB | NB | NB | 2531.3 | NB | NB | NB | NB | NB |
| SC4 | NB | NB | 71.4 | 127.8 | NB | NB | NB | NB | NB | 2001.3 | NB |
| SC6 | NB | NB | 127.8 | 158.8 | NB | NB | NB | NB | NB | NB | NB |
| SC7 | NB | NB | NB | NB | 109.9 | NB | NB | NB | NB | NB | NB |
| SC9 | NB | 2531.3 | NB | NB | NB | NB | 704.4 | NB | NB | NB | 0.0 |
| SC10 | NB | NB | NB | NB | NB | 704.4 | 18.7 | NB | NB | NB | NB |
| SC11 | NB | NB | NB | NB | NB | NB | NB | 14.4 | NB | NB | NB |
| SC12 | NB | NB | NB | NB | NB | NB | NB | NB | 2.7 | NB | NB |
| SC14 | NB | NB | 2001.3 | NB | NB | NB | NB | NB | NB | 41.0 | NB |
| SC15 | NB | NB | NB | NB | NB | 0.0 | NB | NB | NB | NB | 21.9 |

# Chapter 6

# Conclusions and future directions

*Comparison to previously generated data*

While much progress has been made in understanding bZIP interaction specificity, no model can, with very high accuracy, describe the relationship between protein sequence and the energy of interaction. A useful experimental approach to this problem is one that increases both the amount and the quality of experimental data available. When I started my thesis research, the only large bZIP interaction data set consisted of most of the human and *S. cerevisiae* bZIPs, measured using protein arrays. I have expanded upon this by using arrays to generate data for viral bZIPs and designed coiled coils measured against human bZIPs, and designed bZIPs measured against themselves. Additionally, using a quantitative FRET-based solution assay, I quantified the bZIP interaction networks of human, *S. cerevisiae*, 5 additional species, cross-species interactions between *C. intestinalis* and human, and a number of single and double point mutants. Thus, I have measured ~8,000 interactions and non-interactions, which is an increase in the amount of available data of over 4 fold. Besides the increase in the amount of measured interactions, the new data have a number of advantages. The data from additional species represent a more diverse sequence space than that of the human bZIPs. The designed coiled-coil data represent a more simplified interaction space, as the designed peptides are less diverse in sequence than the human bZIPs. The data for bZIP point mutants are useful for looking at what influence only one or two amino-acid changes can have on interaction profiles. Additionally, the quantitative data set makes it possible to test predictions of affinity, rather than just discriminating strong binders from non-binders.

*Comparison of assays used to measure bZIP interactions*

Two different techniques, arrays and FRET-based solution assays, were used to measure bZIP interactions. The bZIP array assay involves expressing and purifying bZIPs both by Ni-

NTA and then further by HPLC. These purified, reduced, and denatured peptides are then printed onto aldehyde-presenting slides. Twelve identical subarrays of 56 proteins containing 4 spots each can be printed on each slide. Each protein to be tested is fluorescently labeled with a CY3 NHS-ester, on one or more primary amine. The arrays are then blocked and incubated with the fluorescently-labeled proteins. After washing, the arrays are imaged, and the fluorescence intensity of each spot is determined. While the bZIP array assay can measure many interactions in parallel, it does not allow for the quantification of interactions because the arrays are only probed at a single concentration. There is also the potential for false negatives, due to semi-specific chemical labeling as well as the measurement of interactions on a surface.

To improve upon the array assay, I developed a high-throughput solution-based FRET assay. In this experiment, proteins are expressed as intein-chitin binding domain fusions and are uniquely labeled at the C-terminus with a fluorescent dye using native chemical ligation. Two versions of each protein are generated, one with an acceptor fluorophore and the other with a donor fluorophore. The proteins are purified first over chitin beads and then over Ni-NTA. Donor proteins are mixed with 12 different concentrations of acceptor labeled protein, and the fluorescence emission of the donor is monitored. These binding curves can then be fit to determine equilibrium disassociation constants. While the solution-based FRET assay is time consuming and costly, it is superior to the array assay in that it provides high quality quantitative data.

*Biological implications*

The *in vitro* interaction data that I have generated between native bZIPs represents the set of interactions that can occur free from cellular influence. These measurements use a standardized set of reagents and measurement techniques and as a result both the array and the

FRET assay have low assay false positive and negative rates (defined as the differences in the interactions that are observed to interact *in vitro* vs. interactions that occur *in vitro* measured by a different technique). In contrast, this data will inevitably contain both biological false positives and false negatives (defined as the differences in the interactions that occur *in vivo* vs. *in vitro*). Biological false positives can occur if the proteins are not co-expressed or co-localized. Additionally, a strong interaction partner for a bZIP might prevent interactions with weaker partners. Biological false negatives can occur if the proteins are brought together by DNA, other domains, or posttranslational modifications (Gaudray, et al. 2002, Lynch, et al. 2011). It is not known how common either biological false positives or negatives will be for the bZIP interaction data I have generated as there is not a comprehensive set of interactions detected *in vivo* to compare to. Furthermore, it is challenging to measure interactions *in vivo* and these measurements can also suffer from assay false positives and negatives. Also, it is difficult to measure interactions in all cell types and conditions to rule out two proteins interacting under any condition. Cellular complexity and difficulty in measuring interactions *in vivo* together make it hard to identify biological relevant interactions (Walhout. 2011).

The bZIP interaction data, though measured *in vitro*, provide a resource to help elucidate the functional significance of bZIP interactions when combined with other types of biological data. One source of biological data is that of gene expression. A requirement for proteins to interact is to be co-expressed, and bZIPS that interact *in vitro* could be compared to see if there is any condition where both partners are expressed. There is now expression data available from various tissues and developmental stages for humans, *D. melanogaster,* and *C. elegans* (Malovannaya, et al. 2011, Ravasi, et al. 2010, Chintapalli, et al. 2007, Graveley, et al. 2011, Spencer, et al. 2011). A drawback of this gene expression data is that levels of mRNA don't

always correlate with protein concentration and the data doesn't have single-cell resolution or contain information about subcellular localization. An alternative approach would be to fuse each bZIP to a fluorescent protein at a single copy under the endogenous promoter, an approach that is now possible in *C. elegans* (Frokjaer-Jensen, et al. 2008). This would allow for quantifying the protein level of each bZIP in different cell types and subcellular locations, although this is an enormous amount of work. Another type of biological data is phenotypic data from gene knockouts or knockdowns, which exist for many genes in several species. Unfortunately, this data is difficult to interpret as many bZIPs have more than one partner and their deletion could result in pleotropic phenotypes. An approach that circumvents these issues is generating mutants that only disrupt individual interactions instead of removing the function of the entire protein (Dreze, et al. 2009). These mutations, when combined with compensatory mutations in the partner that restored the interaction, would be useful for determining the biological significance of interactions (See *Applications of more accurate models).* A third source of biological data is that from CHIP experiments. In combination with information on bZIP DNA-binding specificity, the interaction data could be used to infer which DNA sites are bound by which bZIP complexes. Although the DNA-binding specificity of some bZIPs is known, for many bZIP this has not been determined, especially for species other than human (See *Measuring the DNA binding specificity of bZIPs).*

*Increasing the throughput of quantitative in vitro binding assays*

While both the work of others and the studies described in my thesis have generated a large amount of data, there is likely a need for yet more data to fully understand the relationship between sequence and binding energy in bZIPs. Thus, there is a need for further development of assays that can quickly generate large amounts of quantitative data. The array assay I described

could potentially be increased in throughput by performing it in a 96-well format (Jones, et al. 2006). This is unlikely to provide the high quality, quantitative data desired, however, due to the issues mentioned above. The throughput of the FRET assay could be improved in several ways. One possibility is to make bZIPs fused to fluorescent proteins and express the constructs using *in vitro* extracts. Protein concentrations could then be estimated by fluorescence and the proteins potentially used without purification. This would allow for a large increase in the number of proteins that could be assayed, but it would need to be determined if the bZIPs function properly fused to a much larger fluorescent protein, if bZIP fusion proteins are bright enough to measure tight interactions, and if the proteins could be used without purification. The throughput of the assay could be improved by using a two-stage approach, where initial measurements are made at a single concentration, and then positive interactions are further quantified by making measurements at multiple concentrations. The cutoff for interactions would have to be determined to minimize false negatives as well as false positives. This approach would be especially useful for measuring a sparse interaction space. These modifications together would allow for a much larger number of interactions to be quantified.

Selection-based approaches are attractive since an extremely large number of sequences can be measured simultaneously. Recently there has been excitement around using next generation sequencing as a way to sample all binders, not just those of the highest affinity (Jolma, et al. 2010, Hietpas, et al. 2011, Rockberg, et al. 2008, Ernst, et al. 2010, Fowler, et al. 2010). Combining selection-based approaches and deep sequencing with saturation binding curves provides a potential way to generate a large amount of quantitative interaction data. Using ribosome display, a large number of bZIP coiled-coil variants could be expressed. A biotin-labeled bZIP could be used to isolate proteins that bind by pulling down with streptavidin beads.

These pull downs could be done at multiple concentrations of the biotin-labeled bZIP. To isolate expressing non-binders, displayed proteins could be fused to an epitope tag, incubated with a saturating amount of biotin-labeled bZIP and streptavidin beads, and those proteins not interacting with the biotin-labeled bZIP pulled down using antibody- conjugated beads. The pools of binders at each concentration and non-binders could then be deep sequenced. By using positive controls to calibrate the data, binding curves could be fit to determine Kds that covered several magnitudes of affinity. Many conditions for this assay would have to be determined including expression levels, washing, DNA amplification, sequencing, and data interpretation. Nonetheless, the development of such an assay would allow the rapid quantification of an extremely large number of interactions.

*Additional interactions to measure*

Plant bZIP networks are larger and have distinct sequences from the metazoans, providing an interesting interaction space to measure. Plant bZIPs have been shown to be involved in a number of diverse processes such as seed development, flower maturation, and stress responses (Nijhawan, et al. 2008). 13 families of bZIPs that are conserved throughout plant evolution are present in flowering plants (Correa, et al. 2008). The plant bZIPs represent a separate origin, as only one bZIP family is shared with metazoa and fungi (Correa, et al. 2008). Plant networks are also larger than in the metazoa, with 92 bZIPs in rice and 77 in *Arabidopsis thaliana* (Nijhawan, et al. 2008, Correa, et al. 2008, Deppmann, et al. 2004). The bZIPs from these two species contain bZIPs that are longer than those in human and have a larger number of asparagines at **a** positions (Nijhawan, et al. 2008, Deppmann, et al. 2004). The plant bZIPs have been suggested to primarily form homodimers and intrafamily heterodimers (Deppmann, et al. 2004). Which interactions actually occur is unclear, given the sequence differences between

metazoan and plant bZIPs and little experimental data. A high-throughput FRET assay could potentially be used to measure these interaction networks. Measurement of plant bZIP networks would generate interaction data for a more diverse sequence space as well as provide a useful resource to the plant research community.

*Improving bZIP binding models*

The quantitative interaction data generated from different species present a large and diverse data set for improving bZIP-binding models. Several recent approaches that directly use interaction data to derive predictive models have been applied to similar problems (Chen, et al. 2008, Shao, et al. 2011, AlQuraishi and McAdams. 2011, AlQuraishi and McAdams. 2011). Such models can be tested in cross validation, by withholding protein families or portions of the interaction data set when the model is derived. The devolvement of more accurate models could guide the selection of additional experiments to perform, which would improve the models even further. The solution-based FRET assay could be used to measure coupling energies for pairs of interactions that the models do poorly on.

*Applications of more accurate models*

A useful application of improved binding models would be the design of proteins with specific interaction properties. We previously showed that the CLASSY algorithm can be used to design proteins that bind to one bZIP family but not others. A different design problem involves eliminating one interaction, while maintaining all other interactions and non-interactions at the same affinity. This problem is much harder, as it puts more stringent constraints on the designed sequence. These types of mutants have been coined "edgetic" alleles and would be useful for testing the function of individual interactions *in vivo* (Dreze, et al. 2009). A useful system to apply this approach would be the *C. elegans* bZIP interaction network. This network has fewer

proteins than the human network and also has a lower density of interactions. The development of new techniques for introducing mutant alleles into *C. elegans* provides a convenient way to test the phenotypic effect of the mutant alleles (Frokjaer-Jensen, et al. 2008). An additional use of an improved model would be designing expanded synthetic networks. The synthetic networks described in chapter 4 were not designed to include specific sub-networks and there are several ways the existing networks could be improved. This could include the design of sub-networks not previously observed, or improving existing sub-networks by increasing specificity and/or affinity. These additional sub-networks could be added either to the existing SYNZIP network, or created de novo.

Having a model that approaches experimental accuracy opens a number of interesting opportunities to predict interactions. Almost all eukaryotic genomes contain bZIP proteins, and the number of sequenced genomes is growing at an increasing rate. Prediction of interactions would allow examination of interaction networks on a much larger scale than is accessible experimentally. A much more detailed understanding of bZIP evolution could be achieved by looking at which lineages and on what time scales different interactions were gained and lost. Ancestral sequences could also be inferred, and the interaction properties of the resulting networks predicted (Pinney, et al. 2007). It would also be interesting to see how bZIP interaction network properties evolve by predicting interactions of non-metazoans and seeing whether all of these networks are homodimeric and less connected, as was observed for *S. cerevisiae* and *M. brevicollis*. Finally, the space of synthetic interactions could be interrogated, looking for example at how many pairs of orthogonal bZIP-like coiled coils can exist in the same network.

*Measuring the DNA binding specificity of bZIPs*

Knowing the DNA-binding specificity of transcription factors is important for understanding which genes they regulate. While there has been much effort to map bZIP DNA-binding specificity, there has not been a full accounting of which sites can be bound by each homodimer and heterodimer. Recently several studies have shown that the DNA-binding specificity of transcription factors can be measured rapidly using an approach known as SELEX-SEQ (Jolma, et al. 2010, Zykovich, et al. 2009, Zhao, et al. 2009, Wong, et al. 2011, Slattery, et al. 2011). We have applied this approach to map the DNA-binding specificity of the human bZIP proteins (work in progress). First, a randomized DNA library is incubated with a biotin-labeled bZIP. The bZIP proteins are then pulled down with streptavidin beads, and the bound DNA is amplified and used in successive rounds of binding. After several rounds of enrichment, selected DNA is barcoded with a unique DNA tag, combined with other selections, and subjected to deep sequencing. Using a biotin-labeled bZIP in combination with an unlabeled bZIP partner, DNA-binding specificity of heterodimers can be measured as well. In collaboration with the Ansari lab, we have attempted to measure the DNA-binding specificity of 36 bZIP homodimers as well as a number of heterodimers. Many questions can be addressed from this data, such as what is the effect of protein-protein interactions on DNA binding and what is the space of DNA sequences that are bound by heterodimers but not by homodimers. This specificity profiling approach will also be useful for measuring the DNA-binding specificities of bZIPs from other species; the clones described in chapter 5 can be biotin labeled using the same intein method used for fluorescence labeling. Many of these proteins from other species have more diverse basic regions and thus might have different DNA-binding specificities. Additionally, this data will be useful for comparing how evolution of DNA-binding specificity compares with the evolution of protein-protein interactions.

*Final conclusions*

bZIPs are a great system for understanding both protein-protein and protein-DNA interactions due to their structural simplicity and experimental tractability. Previous work as well as the experiments described in my thesis help make bZIPs one of the best understood models of molecular specificity. By taking advantage of the data already generated and developing new techniques to measure even more interactions, it should be possible to understand in exquisite molecular detail the binding specificity of bZIPs. This knowledge will then allow the prediction of bZIP interaction specificity as well as the design of bZIPs with any specified properties.

## REFERENCES

1. AlQuraishi M, McAdams HH. Direct inference of protein-DNA interactions using compressed sensing methods. Proc Natl Acad Sci U S A. 2011 Sep 6;108(36):14819-24.

2. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G. Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol. 2008 Sep;26(9):1041-5.

3. Chintapalli VR, Wang J, Dow JA. Using FlyAtlas to identify better drosophila melanogaster models of human disease. Nat Genet. 2007 Jun;39(6):715-20.

4. Correa LG, Riano-Pachon DM, Schrago CG, dos Santos RV, Mueller-Roeber B, Vincentz M. The role of bZIP transcription factors in green plant evolution: Adaptive features emerging from four founder genes. PLoS One. 2008 Aug 13;3(8):e2944.

5. Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, Taparowsky EJ, Vinson C. Dimerization specificity of all 67 B-ZIP motifs in arabidopsis thaliana: A comparison to homo sapiens B-ZIP motifs. Nucleic Acids Res. 2004 Jun 29;32(11):3435-45.

6. Dreze M, Charloteaux B, Milstein S, Vidalain PO, Yildirim MA, Zhong Q, Svrzikapa N, Romero V, Laloux G, Brasseur R, Vandenhaute J, Boxem M, Cusick ME, Hill DE, Vidal M. 'Edgetic' perturbation of a C. elegans BCL2 ortholog. Nat Methods. 2009 Nov;6(11):843-9.

7. Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, Sidhu SS. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol Biosyst. 2010 Oct;6(10):1782-90.

8. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010 Sep;7(9):741-6.

9. Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen SP, Grunnet M, Jorgensen EM. Single-copy insertion of transgenes in caenorhabditis elegans. Nat Genet. 2008 Nov;40(11):1375-83.

10. Gaudray G, Gachon F, Basbous J, Biard-Piechaczyk M, Devaux C, Mesnard J. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J. Virol. 2002 December 15, 2002;76(24):12813-22.

11. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, Langton L, Perrimon N, Sandler JE, Wan KH, Willingham A,

Zhang Y, Zou Y, Andrews J, Bickel PJ, Brenner SE, Brent MR, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Oliver B, Celniker SE. The developmental transcriptome of drosophila melanogaster. Nature. 2011 Mar 24;471(7339):473-9.

12. Hietpas RT, Jensen JD, Bolon DN. Experimental illumination of a fitness landscape. Proc Natl Acad Sci U S A. 2011 May 10;108(19):7896-901.

13. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpaa MJ, Bonke M, Palin K, Talukder S, Hughes TR, Luscombe NM, Ukkonen E, Taipale J. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. 2010 Jun;20(6):861-73.

14. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006 Jan 12;439(7073):168-74.

15. Lynch VJ, May G, Wagner GP. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature. 2011 Nov 13;480(7377):383-6.

16. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G, Kim BJ, Li C, Chen R, Li W, Wang Y, O'Malley BW, Qin J. Analysis of the human endogenous coregulator complexome. Cell. 2011 May 27;145(5):787-99.

17. Nijhawan A, Jain M, Tyagi AK, Khurana JP. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol. 2008 Feb;146(2):333-50.

18. Pinney JW, Amoutzias GD, Rattray M, Robertson DL. Reconstruction of ancestral protein interaction networks for the bZIP transcription factors. Proc Natl Acad Sci U S A. 2007 Dec 18;104(51):20449-53.

19. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, Hashimoto T, Hide W, Hofmann O, Kamburov A, Kaur M, Kawaji H, Kubosaki A, Lassmann T, van Nimwegen E, MacPherson CR, Ogawa C, Radovanovic A, Schwartz A, Teasdale RD, Tegner J, Lenhard B, Teichmann SA, Arakawa T, Ninomiya N, Murakami K, Tagami M, Fukuda S, Imamura K, Kai C, Ishihara R, Kitazume Y, Kawai J, Hume DA, Ideker T, Hayashizaki Y. An atlas of combinatorial transcriptional regulation in mouse and man. Cell. 2010 Mar 5;140(5):744-52.

20. Rockberg J, Lofblom J, Hjelm B, Uhlen M, Stahl S. Epitope mapping of antibodies using bacterial surface display. Nat Methods. 2008 Dec;5(12):1039-45.

21. Shao X, Tan CS, Voss C, Li SS, Deng N, Bader GD. A regression framework incorporating quantitative and negative interaction data improves quantitative prediction of PDZ domain-peptide interaction from primary sequence. Bioinformatics. 2011 Feb 1;27(3):383-90.

22. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS. Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. Cell. 2011 Dec 9;147(6):1270-82.

23. Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J, Reinke V, Petrella L, Strome S, Von Stetina SE, Katz M, Shaham S, Ratsch G, Miller DM,3rd. A spatial and temporal map of C. elegans gene expression. Genome Res. 2011 Feb;21(2):325-41.

24. Walhout AJ. What does biologically meaningful mean? A perspective on gene regulatory network validation. Genome Biol. 2011;12(4):109.

25. Wong D, Teixeira A, Oikonomopoulos S, Humburg P, Lone IN, Saliba D, Siggers T, Bulyk M, Angelov D, Dimitrov S, Udalova IA, Ragoussis J. Extensive characterization of NF-kappaB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. Genome Biol. 2011 Jul 29;12(7):R70.

26. Zhao Y, Granas D, Stormo GD. Inferring binding energies from selected binding sites. PLoS Comput Biol. 2009 Dec;5(12):e1000590.

27. Zykovich A, Korf I, Segal DJ. Bind-n-seq: High-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. Nucleic Acids Res. 2009 Dec;37(22):e151.

# APPENDIX A

# Supplementary Information for "Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays"

**Reproduced with permission from:**
Reinke AW, Grigoryan G, Keating AE. Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. Biochemistry. 2010 Mar 9;49(9):1985-97.

**Collaborator notes:**
Gevorg Grigoryan computationally designed the anti-MEQ peptide.

**SUPPLEMENTARY EXPERIMENTS**

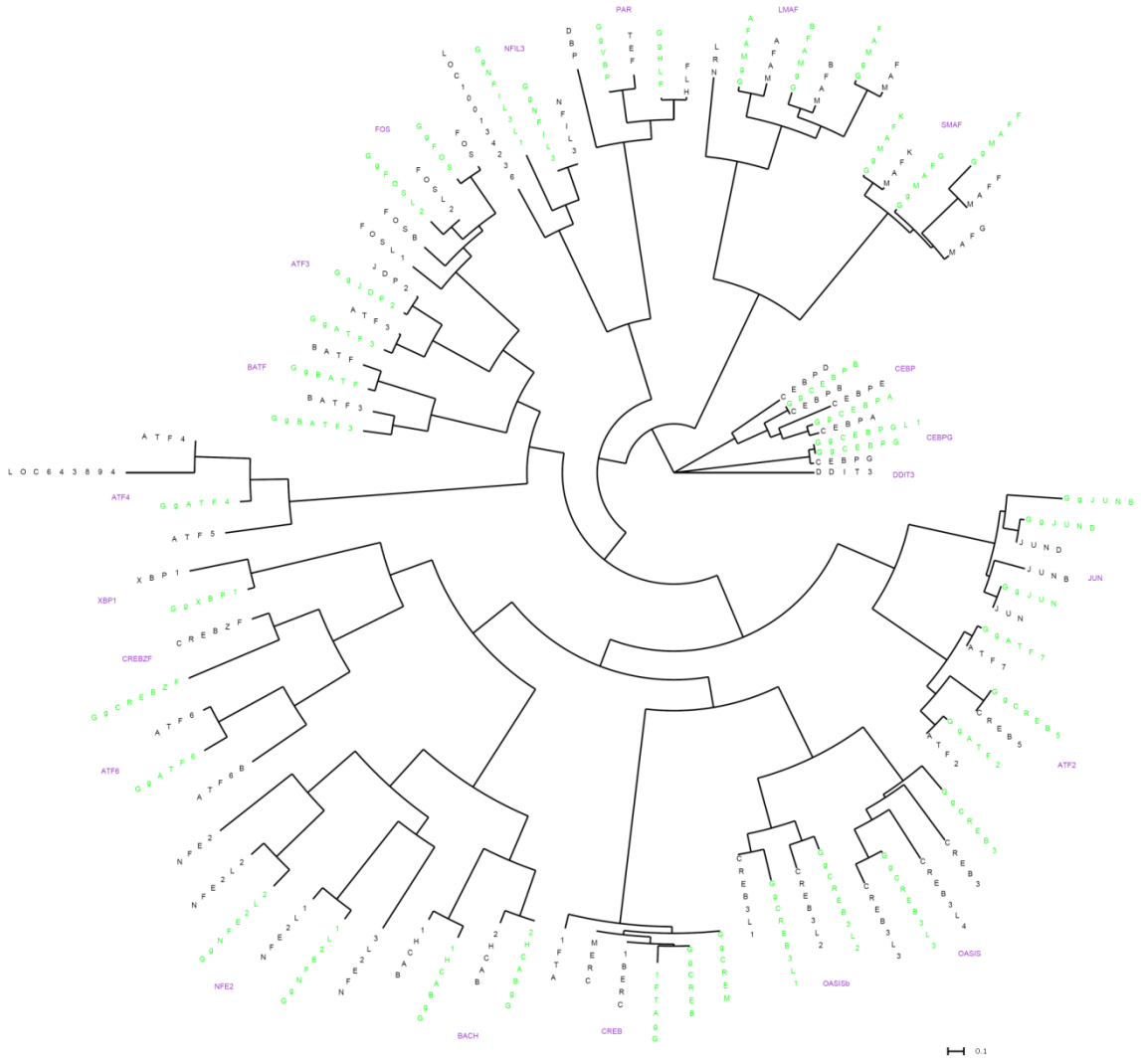



Figure A.S1. Comparison of Human and Chicken bZIPs. Tree is inferred by neighbor-joining using the leucine-zipper sequence of each human bZIP and each *G. galus* bZIP as described in the methods. Human sequences are in black and chicken sequences are in green. Family names are listed in purple. The scale bar refers to amino-acid changes per position. Overall, the chicken sequences are highly homologous to the human sequences, as judged by the short branch lengths between orthologs. All families are conserved between chicken and human, except for DDIT3, which is human specific.

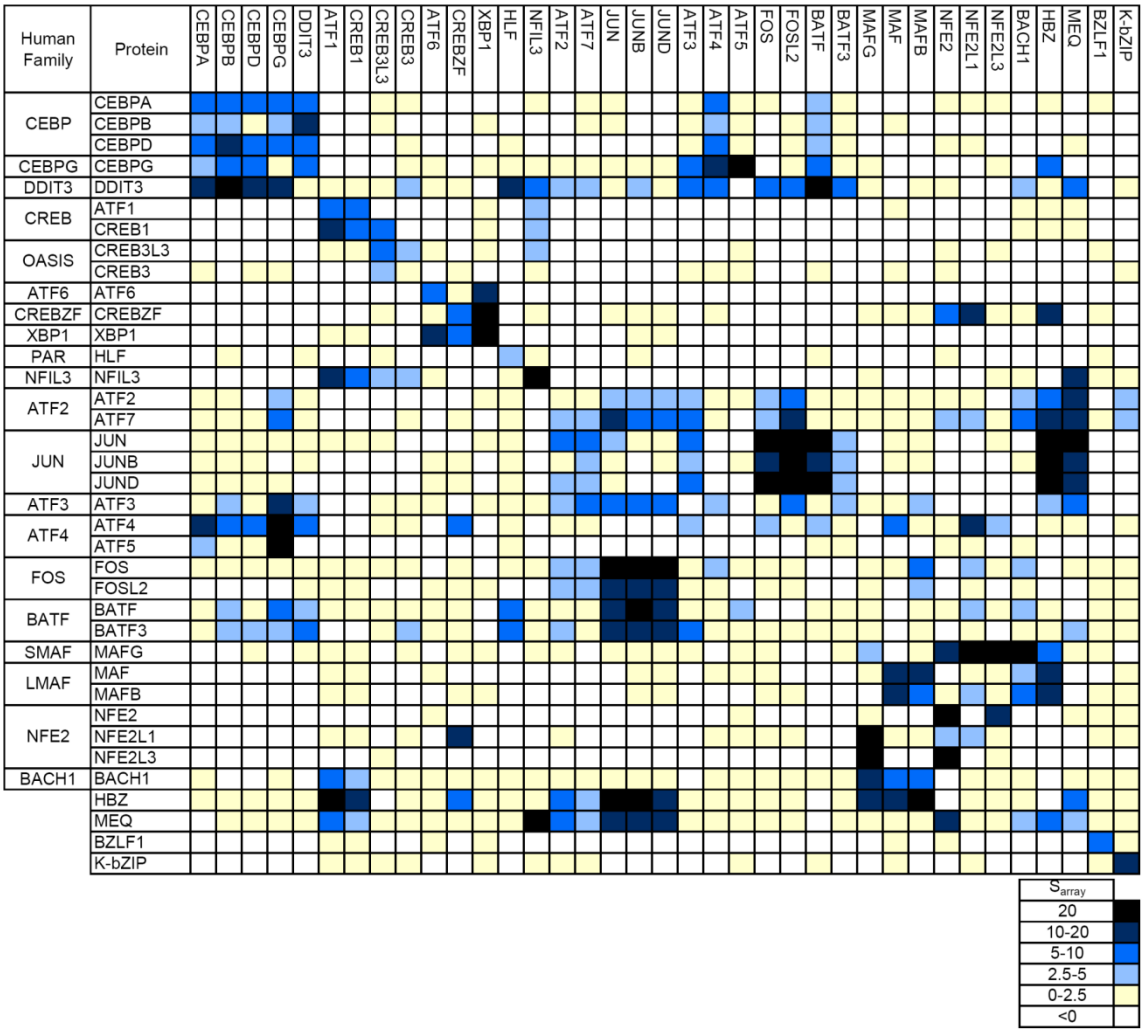



Figure A.S2. Complete interaction matrix of 33 human bZIPs and 4 viral bZIPs. Data are displayed as in Figure 2.2**.** Solution probe proteins are in columns, and proteins on the surface are in rows.

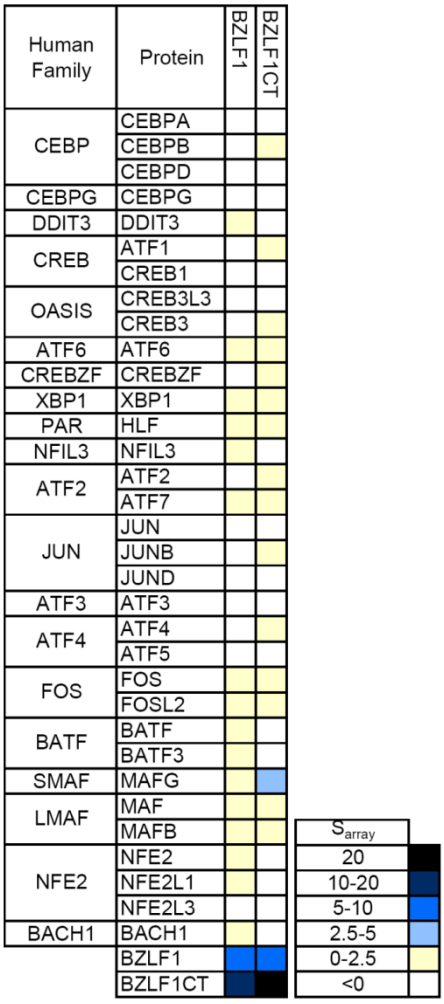



Figure A.S3. Neither the BZLF1 leucine zipper nor BZLF1 with additional C-terminal residues binds strongly to any human bZIP. Fluorescently labeled BZLF1 at 1280 nM and BZLF1 with the C-terminal region (BZLF1CT) at 160 nM in solution are listed in columns and potential partners on the surface are listed in rows. Data are displayed as in Figure 2.2.

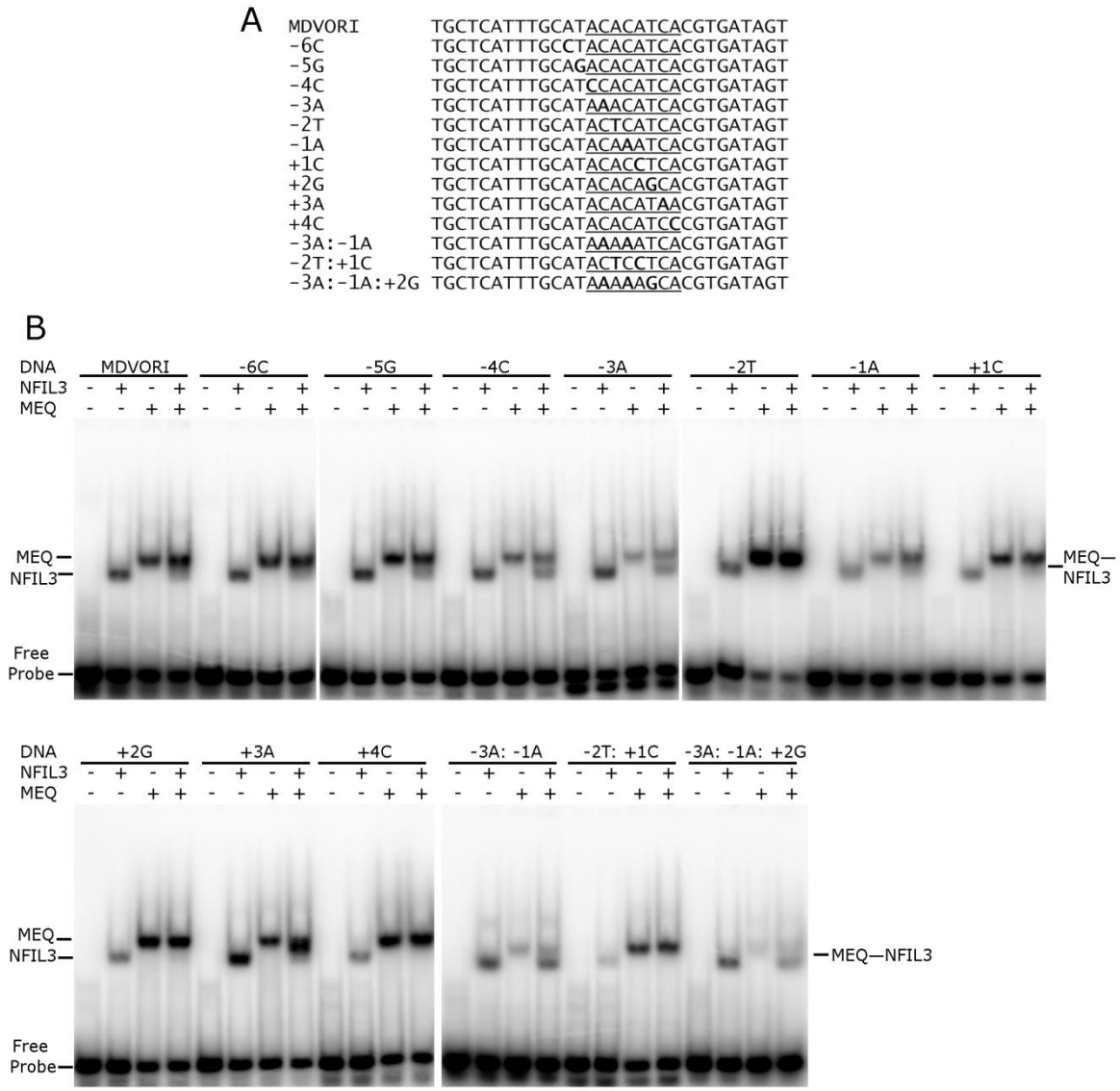



Figure A.S4. Gel shifts showing MEQ and NFIL3 directly binding to variants of the MDV DNA site. (A) DNA sequences used in gel-shift assays. The putative binding site is underlined. (B) Gel-shift experiments with MEQ and NFIL3 binding to different DNA sites. The concentration of MEQ and NFIL3 was 80 nM each or 160 nM total protein for mixtures. Each lane had 20 nM radiolabeled DNA. Each homodimer, heterodimer, and free probe is indicated at left. Strong heterodimer formation is observed on the +3A site.

Table A.S1. Protein sequences used in this study. Linker region of proteins is in bold.

| Proteins used in array studies | | |
|---|---|---|
| Name | Protein | Source |
| CEBPA | **SYYHHHHHHLESTSLYKKAGSGS**QRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRGIFRQL**LE** | Grigoryan, et al. 2009 |
| CEBPB | **SYYHHHHHHLESTSLYKKAGSEF**SDEYKIRRERNNIAVRKSRDKAKMRNLETQHKVLELTAENERLQKKVEQLSRELSTLRNLFKQLPEPLLASSGHC | Newman, et al. 2003 |
| CEBPD | **SYYHHHHHHLESTSLYKKAGSGS**RRNQEMQQKLVELSAENEKLHQRVEQLTRDLAGLRQFFKQL**LE** | Newman, et al. 2003 |
| CEBPG | **SYYHHHHHHLESTSLYKKAGSEFG**ERNNMAVKKSRLKSKQKAQDTLQRVNQLKEENERLEAKIKLLTKELSVLKDLFLEHAHNLADNVQSISTENTTADG**LE** | Newman, et al. 2003 |
| DDIT3 | **SYYHHHHHHLESTSLYKKAGSEF**RMKEKEQENERKVAQLAEENERLKQEIERLTREVEATRRALIDRMVNLHQA | Newman, et al. 2003 |
| ATF1 | **SYYHHHHHHLESTSLYKKAGSEF**DPQLKREIRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKTLKDLYSNKSV | Newman, et al. 2003 |
| CREB1 | **SYYHHHHHHLESTSLYKKAGSGSK**EAARKREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD | Newman, et al. 2003 |
| CREB3L3 | **SYYHHHHHHLESTSLYKKAGSGS**EYIDGLETRMSACTAQNQELQRKVLHLEKQNLSLLEQLKKLQAIVVQSTSLE | Newman, et al. 2003 |
| CREB3 | **SYYHHHHHHLESTSLYKKAGSEFW**RRKIRNKRSAQESRRKKKVYVGGLESRVLKYTAQNMELQNKVQLLEEQNLSLLDQLRKLQAMVIEISNKTSS**RLE** | Newman, et al. 2003 |
| ATF6 | **SYYHHHHHHLESTSLYKKAGSEF**ACQSRKKKKEYMLGLEARLKAALSENEQLKKENGTLKRQLDEVVSENQRLKVPSPKRRV**LE** | Newman, et al. 2003 |
| CREBZF | **SYYHHHHHHLESTSLYKKAGSEFC**RLNRLKKKEYVMGLESRVRGLAAENQELRAENRELGKRVQALQEESRYLRAVLANETGL**E** | Newman, et al. 2003 |
| XBP1 | **SYYHHHHHHLESTSLYKKAGSEF**QTARDRKKARMSELEQQVVDLEEENQKLLLENQLLREKTHGLVVENQELRQRLGMDALVAEEEAEAKGNEV**LE** | Newman, et al. 2003 |
| HLF | **SYYHHHHHHLESTSLYKKAGSG**RLKENQIAIRASFLEKENSALRQEVADLRKELGKCKNILAKYEARH**LE** | Newman, et al. 2003 |
| NFIL3 | **SYYHHHHHHLESTSLYKKAGSEF**EKRRLNDLVLENKLIALGEENATLKAELLSLKLKFGLISSTAYAQEIQKLSNSTAVYFQDYQTSKSNV**LE** | Newman, et al. 2003 |
| ATF2 | **SYYHHHHHHLESTSLYKKAGSEFD**LERNRAAASRCRQKRKVWVQSLEKKAEDLSSLNGQLQSEVTLLRNEVAQLKQLLLAHKDCPVTAMQKKSG**FLE** | Newman, et al. 2003 |
| ATF7 | **SYYHHHHHHLESTSLYKKAGSEF**KRKLWVSSLEKKAEELTSQNIQLSNEVTLLRNEVAQLKQLLLAHKDCPVTALQK**LE** | Newman, et al. 2003 |
| JUN | **SYYHHHHHHLESTSLYKKAGSGS**RKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNH**LE** | Grigoryan, et al. 2009 |
| JUNB | **SYYHHHHHHLESTSLYKKAGSGS**RKLERIARLEDKVKTLKAENAGLSSTAGLLREQVAQLKQKVMNH**LE** | Grigoryan, et al. 2009 |
| JUND | **SYYHHHHHHLESTSLYKKAGSGS**RKLERISRLEEKVKTLKSQNTELASTASLLREQVAQLKQKVMNH**LE** | Grigoryan, et al. 2009 |
| ATF3 | **SYYHHHHHHLESTSLYKKAGSGSC**PEEDERKKRRRERNKIAAAKCRNKKKEKTECLQKESEKLESVNAELKAQIEELKNEKQHLIYMLNLHRPTCIVRAQNGRTPED**LE** | Newman, et al. 2003 |

| | | |
|---|---|---|
| ATF4 | **SYYHHHHHHLESTSLYKKAGSEF**NKTAATRYRQKKRAEQEALTGE CKELEKKNEALKERADSLAKEIQYLKDLIEEVRKARGKKRVP | Newman, et al. 2003 |
| ATF5 | **SYYHHHHHHLESTSLYKKAGSEFN**RKQKKRDQNKSAALRYRQRKR AEGEALEGECQGLEARNRELKERAESVEREIQYVKDLLIEVYKARSQRT RSC | Newman, et al. 2003 |
| FOS | **SYYHHHHHHLESTSLYKKAGSEFF**RRERNKMAAAKCRNRRRELTD TLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDDLG FPEEMS**LE** | Newman, et al. 2003 |
| FOSL2 | **SYYHHHHHHLESTSLYKKAGSEF**ERNKLAAAKCRNRRRELTEKLQA ETEELEEEKSGLQKEIAELQKEKEKLEFMLVAHGPVCKISP**LE** | Newman, et al. 2003 |
| BATF | **SYYHHHHHHLESTSLYKKAGSGS**QKADTLHLESEDLEKQNAALRKE IKQLTEELKYFTSVLNSHE**LE** | Newman, et al. 2003 |
| BATF3 | **SYYHHHHHHLESTSLYKKAGSEF**RSRKKQTQKADKLHEEYESLEQE NTMLRREIGKLTEELKHLTEALKEHEKMCPLLLCPMNFV**HLE** | Newman, et al. 2003 |
| MAFG | **SYYHHHHHHLESTSLYKKAGSEFG**VTQKEELEKQKAELQQEVEKLA SENASMKLELDALRSKYEALQTFARTVARSPVAPARGPLAAGLGPLV PGKVAATSVITIVKSKTDA**LE** | Newman, et al. 2003 |
| MAF | **SYYHHHHHHLESTSLYKKAGSEFQ**RVQQRHVLESEKNQLLQQVDHL KQEISRLVRERDAYKEKYEKLVSSGFRENGSSSDNPSSPEFFM | Newman, et al. 2003 |
| MAFB | **SYYHHHHHHLESTSLYKKAGSEFQ**YKRVQQKHHLENEKTQLIQQVE QLKQEVSRLARERDAYKVKCEKLANSGFREAGSTSDSPSSPEFFL | Newman, et al. 2003 |
| NFE2 | **SYYHHHHHHLESTSLYKKAGSEFQ**RKLETIVQLERELERLTNERERLL RARGEADRTLEVMRQQLTELYRDIFQHLRDESGNS | Newman, et al. 2003 |
| NFE2L1 | **SYYHHHHHHLESTSLYKKAGSEF**KLDTILNLERDVEDLQRDKARLLR EKVEFLRSLRQMKQKVQSLYQEVFGRLRDENGRPY**YLEII** | Newman, et al. 2003 |
| NFE2L3 | **SYYHHHHHHLESTSLYKKAGSEFG**DIRRRGKNKVAAQNCRKRKLDII LNLEDDVCNLQAKKETLKREQAQCNKAINIMKQKLHDLYHDIFSRLR DDQGRPV**LE** | Newman, et al. 2003 |
| BACH1 | **SYYHHHHHHLESTSLYKKAGSEFG**CRKRKLDCIQNLESEIEKLQSEKE SLLKERDHILSTLGETKQNLTGLCQKVCKEAALSQEQN**LE** | Newman, et al. 2003 |
| HBZ | **SYYHHHHHHLESTSLYKKAGSGS**MQELGIDGYTRQLEGEVESLEAER RKLLQEKEDLMGEVNYWQGRLEAMWLQ | This study |
| MEQ | **SYYHHHHHHLESTSLYKKAGSGS**DYVDKLHEACEELQRANEHLRKE IRDLRTECTSLRVQLARHEP | This study |
| BZLF1 | **SYYHHHHHHLESTSLYKKAGSGS**AKFKQLLQHYREVAAAKSSEND RLRLLLKQM**GGRDYKDDDDK** | This study |
| K-bZIP | **SYYHHHHHHLESTSLYKKAGSGS**VSSKAYTRQLQQALEEKDAQLCF LAARLEAHKEQIIFLRDMLMRMCQQ**GGRDYKDDDDK** | This study |
| BZLF1CT | **SYYHHHHHHLESTSLYKKAGSGS**AKFKQLLQHYREVVAAKSSEND RLRLLLKQMCPSLDVDSIIPRTPDVLHEDLLNF**LE** | This study |
| anti-Meq | **SYYHHHHHHLESTSLYKKAGSGS**NLLATLRSTAAVLENENHVLEKE KEKLRKEKEQLLNKLEAYK | This study |

| Proteins used in circular dichroism and gel-shift studies | | |
|---|---|---|
| Name | Protein | Source |
| JUN | **SHHHHHHGESKEYKKGSGS** SPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNH | This study |
| CEBPG | **SHHHHHHGESKEYKKGSGS** KKSSPMDRNSDEYRQRRERNNMAVKKSRLKSKQKAQDTLQRVNQLKEENERLEAKIKLLTKELSVLKDLFLEHAHNLAD | This study |
| CREBZF | **SHHHHHHGESKEYKKGSGS** GGGSGNDNNQAATKSPRKAAAAAARLNRLKKKEYVMGLESRVRGLAAENQELRAENRELGKRVQALQEESRYLRAVLANETGL | This study |
| MAFB | **SHHHHHHGESKEYKKGSGS** SDDQLVSMSVRELNRHLRGFTKDEVIRLKQKRRTLKNRGYAQSCRYKRVQQKHHLENEKTQLIQQVEQLKQEVSRLARERDAYKVKCEKLANSG | This study |
| MAFG | **SHHHHHHGESKEYKKGSGS** TDEELVTMSVRELNQHLRGLSKEEIVQLKQRRRTLKNRGYAASCRVKRVTQKEELEKQKAELQQEVEKLASENASMKLELDALRSKYEALQTFARTVARS | This study |
| NFIL3 | **SHHHHHHGESKEYKKGSGS** REFIPDEKKDAMYWEKRRKNNEAAKRSREKRRLNDLVLENKLIALGEENATLKAELLSLKLKFGLIS | This study |
| ATF2 | **SHHHHHHGESKEYKKGSGS** RRRAANEDPDEKRRKFLERNRAAASRCRQKRKVWVQSLEKKAEDLSSLNGQLQSEVTLLRNEVAQLKQLLLAHKDC | This study |
| HBZ | **SHHHHHHGESKEYKKGSGS** KAADVARRKQEEQERRERKWRQGAEKAKQHSARKEKMQELGIDGYTRQLEGEVESLEAERRKLLQEKEDLMGEVNYWQGRLEAMWLQ | This study |
| MEQ | **SHHHHHHGESKEYKKGSGS** DGLSEEEKQKLERRRKRNRDAARRRRRKQTDYVDKLHEACEELQRANEHLRKEIRDLRTECTSLRVQLARHEP | This study |
| HBZLZ | **SHHHHHHGESKEYKKGSGS** MQELGIDGYTRQLEGEVESLEAERRKLLQEKEDLMGEVNYWQGRLEAMWLQ | This study |
| anti-Meq | **GS** NLLATLRSTAAVLENENHVLEKEKEKLRKEKEQLLNKLEAYK | This study |

Table A.S2**.** Average background-corrected fluorescence values from the array experiments. Peptides in solution are in columns and those on the surface are in rows. Concentrations are in nM.

| | CEBPA | CEBPB | CEBPD | CEBPG | DDIT3 | ATF1 | CREB1 | CREB3L3 | CREB3 | ATF6 | CREBZF | XBP1 | HLF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CEBPA | 16400.2 | 8465.3 | 27862.6 | 9709.6 | 15186.7 | -506.3 | -1345.1 | 39.0 | 854.9 | 447.3 | -2210.8 | 602.6 | 721.7 |
| CEBPB | 7461.5 | 4130.7 | 8591.9 | 7753.9 | 26228.3 | -908.5 | -1663.9 | -241.4 | -715.6 | 625.3 | -2029.1 | 1612.2 | 811.4 |
| CEBPD | 19228.3 | 10494.0 | 24816.0 | 13870.9 | 15487.3 | -704.8 | -1884.8 | -909.6 | 3206.9 | 141.8 | -3693.8 | -82.5 | 2449.6 |
| CEBPG | 8962.2 | 7403.8 | 27789.5 | 1671.9 | 13620.6 | -174.0 | 542.7 | -861.8 | 4113.4 | 1245.2 | -1881.1 | 1262.3 | 4043.1 |
| DDIT3 | 26600.9 | 20762.0 | 38960.3 | 32663.2 | 4609.1 | 602.1 | 1200.6 | 1171.1 | 9313.8 | 1412.0 | 1187.1 | 1250.8 | 39900.9 |
| ATF1 | -2325.3 | -752.7 | -3890.6 | -1944.6 | -1049.1 | 7323.9 | 10884.8 | -2272.6 | -2677.4 | 540.4 | -1896.8 | 2023.9 | -631.4 |
| CREB1 | -4412.3 | -1331.3 | -5543.4 | -3011.3 | -1302.8 | 12318.6 | 16312.6 | 8183.3 | 22.6 | -1502.1 | -5257.4 | 1707.9 | -2266.8 |
| CREB3L3 | -2019.1 | -547.3 | -3228.1 | -713.1 | -1378.3 | 1139.9 | 2580.6 | 6710.3 | 10436.3 | 1545.3 | -1101.1 | 1565.8 | -1001.6 |
| CREB3 | 40.0 | 75.4 | 387.8 | -597.7 | 295.3 | -503.5 | -114.5 | 2829.3 | 4008.9 | 718.2 | 32.5 | 1099.8 | 170.5 |
| ATF6 | -3547.4 | -593.4 | -3227.4 | -1915.9 | -1391.2 | -298.7 | -14.1 | -503.3 | -2670.7 | 9418.6 | 841.3 | 7659.4 | -1641.3 |
| CREBZF | -2999.2 | -615.9 | -5579.0 | -1615.1 | -1600.5 | -594.1 | -947.2 | -306.9 | -2843.9 | 1969.0 | 17003.4 | 13108.1 | -1889.5 |
| XBP1 | -6916.5 | -547.2 | -7262.9 | -2792.9 | -3011.3 | 754.8 | 2224.9 | -768.8 | -3298.7 | 13998.1 | 19643.8 | 20740.2 | -2661.9 |
| HLF | -1301.0 | 256.9 | -1111.6 | -1468.2 | 2365.1 | -154.1 | -418.7 | -161.4 | -702.9 | 242.4 | -75.2 | 1249.9 | 7040.4 |
| NFIL3 | -1599.6 | 74.3 | -1199.6 | -973.9 | 115.3 | 9223.2 | 13337.8 | 2318.7 | 8054.4 | 1508.5 | -344.8 | 1191.6 | 1051.4 |
| ATF2 | 1035.3 | 392.1 | -2091.9 | 5852.3 | 2062.8 | -529.5 | -146.8 | -1034.7 | -718.0 | 611.7 | -534.2 | 1981.5 | 868.6 |
| ATF7 | 3827.6 | 2387.9 | 1705.9 | 8493.0 | 5268.9 | -286.3 | 221.4 | -355.9 | 864.9 | 928.9 | 919.6 | 1260.1 | 1360.1 |
| JUN | -1138.9 | 1104.1 | 1273.4 | 1133.4 | 1112.1 | 837.5 | 2531.1 | -201.8 | -41.4 | 912.5 | -53.3 | 1307.4 | 4855.9 |
| JUNB | 329.0 | 328.1 | 333.1 | -1054.5 | 859.6 | -143.9 | 299.9 | -2793.0 | -2083.4 | 1165.6 | 210.2 | 884.7 | 2361.5 |
| JUND | -534.6 | 384.4 | 853.2 | 17.3 | -17.2 | 33.8 | 534.7 | -959.6 | -902.1 | 1144.2 | 19.0 | 1101.1 | 3347.6 |
| FOS | 3174.7 | 262.3 | 1517.2 | 2895.8 | 3169.9 | 729.7 | 1007.3 | 97.1 | 1038.8 | 718.4 | 1801.2 | 1088.4 | 996.4 |
| FOSL2 | -2308.9 | -751.7 | -4357.6 | -2292.1 | 777.8 | 441.7 | 1460.4 | -1100.6 | -1721.3 | 1350.8 | 158.9 | 1436.9 | 516.9 |
| ATF3 | 2806.1 | 4267.3 | 2664.2 | 22753.4 | 9553.3 | -106.5 | -425.6 | 232.3 | 2208.9 | 1007.6 | -972.4 | 2238.9 | 1037.1 |
| ATF4 | 42648.7 | 7702.9 | 31696.6 | 37828.5 | 16336.9 | -293.9 | -469.8 | 764.2 | 1426.9 | -389.4 | 11897.4 | 721.7 | 4694.9 |
| ATF5 | 7171.6 | 416.8 | 7273.9 | 37753.9 | -1748.5 | -1494.4 | -3841.6 | -1762.7 | -2506.3 | -907.9 | -3414.4 | 124.6 | 1124.4 |
| BATF | 3553.1 | 4548.5 | 5470.9 | 15419.5 | 11096.2 | 755.9 | 2427.5 | -81.2 | 5270.6 | 1826.2 | 722.6 | 1111.6 | 13375.9 |
| BATF3 | 2636.2 | 3614.8 | 9704.8 | 7478.2 | 15171.9 | -562.1 | 278.1 | 787.2 | 9465.5 | 1012.4 | 1747.7 | 1144.1 | 12205.6 |
| MAFG | -1183.9 | -63.0 | -311.9 | -622.7 | 1139.0 | 647.1 | 2002.4 | -1726.7 | 1331.4 | 670.8 | 343.2 | 1756.7 | 159.6 |
| MAF | -2329.4 | -339.8 | -4658.0 | -1457.8 | -1166.8 | 225.1 | 1056.9 | -785.8 | -1984.5 | 934.9 | -1061.8 | 767.5 | -1173.0 |
| MAFB | -5483.4 | -1013.9 | -6333.9 | -2140.1 | -2616.7 | 221.3 | 1218.9 | -920.0 | -3630.9 | 915.3 | 705.4 | 1588.4 | -1800.8 |
| NFE2 | -2821.6 | -614.0 | -4646.7 | -1510.3 | -920.4 | -208.2 | -396.0 | -840.0 | -545.9 | 1375.8 | -576.6 | 1034.0 | -2282.5 |
| NFE2L1 | -2720.7 | -1318.3 | -2439.9 | -4630.6 | -3339.6 | 1320.4 | 3225.5 | -1003.4 | 571.8 | 620.9 | 21919.8 | 790.6 | 428.9 |
| NFE2L3 | -2135.3 | -1394.6 | -2950.1 | -2078.8 | -1473.6 | -1587.1 | -1765.9 | -293.0 | -1720.3 | -930.3 | -2797.8 | 216.3 | -1946.4 |
| BACH1 | -533.6 | -139.6 | -2345.4 | 946.1 | 415.6 | 3900.5 | 6874.6 | -354.8 | 584.9 | 2954.9 | 3984.0 | 2655.3 | 2069.2 |
| HBZ | -972.4 | 1002.1 | 1611.6 | 3554.8 | 677.3 | 17760.8 | 25708.1 | -493.2 | 1685.6 | 3201.4 | 16017.4 | 1463.7 | 3430.4 |
| MEQ | -2124.4 | 949.3 | 2242.4 | 64.4 | 853.4 | 5993.6 | 10513.8 | 310.9 | 4790.3 | 1837.4 | 1849.3 | 1700.5 | 1016.8 |
| BZLF1 | -2770.5 | -769.1 | -4329.3 | -4146.8 | -1388.7 | 379.9 | 2380.1 | -1833.6 | -2351.9 | 1376.9 | -326.9 | 1404.2 | -618.6 |
| K-bZIP | -2506.1 | -261.5 | -1129.1 | -1520.8 | -1119.4 | 90.3 | 1383.8 | -83.6 | 659.2 | 693.1 | -100.8 | 1560.6 | -845.4 |

| | NFIL3 | ATF2 | ATF7 | JUN | JUNB | JUND | FOS | FOSL2 | ATF3 | ATF4 | ATF5 | BATF | BATF3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CEBPA | 311.0 | 177.7 | 2142.6 | 443.3 | 161.8 | 167.2 | 1028.4 | 123.3 | 1594.4 | 10297.5 | 1135.1 | 1670.0 | 7982.9 |
| CEBPB | -975.4 | -242.9 | 2616.4 | 131.1 | 411.6 | -31.4 | -336.8 | 176.7 | 2371.0 | 4304.3 | 215.2 | 3457.9 | 7326.8 |
| CEBPD | -42.6 | -1858.2 | -205.4 | -22.1 | -468.6 | -785.9 | 142.3 | 220.8 | 2765.6 | 13044.8 | 699.1 | 3465.8 | 14516.9 |
| CEBPG | 1323.0 | 3834.6 | 7042.0 | 109.3 | 585.5 | 478.8 | 337.8 | 958.9 | 12723.0 | 18182.6 | 15097.3 | 8235.8 | 13138.5 |
| DDIT3 | 9541.1 | 8325.1 | 10043.6 | 1175.8 | 2991.4 | 2757.3 | 9237.4 | 4533.2 | 15455.3 | 8893.6 | -417.9 | 28971.3 | 28594.8 |
| ATF1 | 3741.3 | -1573.1 | -700.9 | -126.0 | 190.1 | -350.2 | -999.0 | -169.2 | -1237.6 | -414.1 | -246.6 | -614.0 | -1975.0 |
| CREB1 | 3658.4 | -5291.3 | -1166.6 | -782.3 | -780.0 | -1985.9 | -2004.8 | -770.9 | -2966.8 | -1935.2 | -1794.8 | -827.1 | -5137.0 |
| CREB3L3 | 4204.5 | -261.5 | -371.4 | -138.6 | 388.7 | -376.4 | 28.5 | -55.8 | 90.6 | -360.2 | 340.1 | -731.8 | 1616.0 |
| CREB3 | 57.3 | -509.8 | -516.1 | 87.5 | -121.1 | 81.8 | -581.9 | -327.1 | 845.1 | 750.3 | 146.4 | -573.1 | 3517.6 |
| ATF6 | -554.3 | -2102.8 | -1167.4 | -720.4 | -154.4 | -564.4 | -1550.1 | -225.3 | -1524.8 | -427.6 | -256.4 | -860.6 | -2381.9 |
| CREBZF | 218.9 | -1429.3 | 632.1 | 468.1 | 1032.5 | 659.6 | -202.6 | 72.6 | -384.6 | 2317.1 | -353.2 | -950.6 | 1108.5 |
| XBP1 | -1162.3 | -2233.0 | -832.6 | -171.3 | 458.9 | 236.1 | -1021.7 | -120.2 | -1818.7 | -990.6 | -1059.9 | -1572.4 | -2693.0 |
| HLF | 710.3 | -2027.1 | -942.8 | -291.2 | 597.9 | -395.6 | -1201.5 | -259.3 | -1568.9 | 307.9 | -433.8 | 985.0 | 2552.6 |
| NFIL3 | 36187.8 | -2371.1 | -1067.3 | -13.9 | 221.2 | -460.8 | -1433.9 | -113.6 | 129.3 | -314.7 | -1169.5 | -1087.4 | 3450.6 |
| ATF2 | -2132.6 | 5073.2 | 5239.1 | 3518.8 | 3925.3 | 4649.3 | 5701.9 | 3343.9 | 7935.4 | 965.4 | 49.6 | 801.9 | 9572.6 |
| ATF7 | -536.4 | 6890.5 | 8203.3 | 8006.6 | 4837.2 | 7666.8 | 9036.9 | 6553.8 | 10295.4 | 2822.0 | -847.1 | 511.3 | 11128.9 |
| JUN | -169.6 | 19776.8 | 19617.1 | 2428.2 | 1385.8 | 1262.9 | 45551.8 | 28676.3 | 17605.2 | -405.3 | -571.8 | 26673.3 | 21878.4 |
| JUNB | -1109.3 | 5917.8 | 7557.8 | 393.0 | 270.0 | 356.9 | 29611.2 | 14296.9 | 8909.2 | 83.7 | 67.6 | 16984.4 | 19485.6 |
| JUND | -866.3 | 10644.5 | 13456.1 | 513.0 | 367.0 | 107.1 | 37736.6 | 19249.6 | 12655.8 | -394.1 | -294.9 | 22249.5 | 20503.4 |
| FOS | 522.0 | 8367.4 | 10322.2 | 20367.8 | 19903.6 | 22831.6 | 3277.1 | 1276.4 | 2225.3 | 3989.1 | 497.0 | -607.3 | 1681.9 |
| FOSL2 | -368.7 | 8843.4 | 8116.8 | 12360.6 | 13728.4 | 16881.4 | 3924.3 | 653.4 | 3644.9 | 715.6 | -421.1 | -2058.3 | -918.3 |
| ATF3 | 704.1 | 12564.9 | 14134.1 | 6003.6 | 8068.3 | 8131.3 | 3642.9 | 2656.0 | 454.6 | 5220.8 | -504.4 | 1485.6 | 22269.8 |
| ATF4 | -1019.1 | 778.9 | 2882.3 | -438.0 | -336.2 | -779.4 | 5731.6 | 588.9 | 5657.8 | -837.6 | -316.6 | 2929.7 | 11336.2 |
| ATF5 | -1773.8 | -3627.9 | -2386.1 | -1661.9 | -1547.1 | -2013.1 | -2763.5 | -372.4 | -2583.8 | -736.8 | -1240.7 | 223.8 | 5578.3 |
| BATF | 2610.4 | 4101.5 | 2571.9 | 11045.4 | 17121.3 | 18955.2 | 12.6 | -526.9 | 3272.2 | 2899.7 | 1919.1 | -225.9 | 5890.9 |
| BATF3 | 2602.0 | 6868.8 | 6773.6 | 9549.0 | 12115.1 | 11855.2 | 571.1 | 353.1 | 10329.4 | 3567.6 | 908.1 | 1052.8 | 9517.6 |
| MAFG | 1850.6 | 764.4 | 1274.1 | 949.8 | 1946.8 | 1579.7 | 1644.1 | -170.9 | 624.0 | 578.6 | 565.2 | -894.4 | 6731.1 |
| MAF | -805.8 | -294.3 | -868.0 | -13.8 | 1040.6 | 473.6 | 512.9 | 50.8 | -81.9 | 3239.0 | -101.0 | -1018.1 | -1526.8 |
| MAFB | -629.3 | -618.8 | -975.2 | -318.8 | 837.3 | 744.2 | 3415.9 | 1257.3 | -81.8 | -448.7 | -735.3 | -789.6 | -1361.7 |
| NFE2 | -783.9 | -1479.1 | -1312.3 | -778.5 | -422.6 | -833.8 | -1096.3 | -216.4 | -1347.9 | 188.1 | -74.3 | -1950.9 | -2877.1 |
| NFE2L1 | -2427.6 | 345.8 | 1426.8 | -1354.3 | -264.5 | -212.5 | 1023.4 | 282.4 | -1609.1 | 3815.4 | 257.0 | -1244.8 | 227.1 |
| NFE2L3 | -711.4 | -3174.6 | -1290.1 | -1013.0 | -980.9 | -1452.3 | -2584.0 | -534.7 | -2328.6 | -128.6 | -636.8 | -897.5 | -3180.9 |
| BACH1 | 1005.3 | 4546.8 | 3102.4 | 245.8 | 1229.9 | 1290.2 | 2196.3 | 574.4 | 285.2 | 805.4 | 1171.4 | 367.3 | 3569.8 |
| HBZ | 839.8 | 16118.8 | 12402.9 | 26402.8 | 29515.8 | 18528.8 | 4509.8 | 912.4 | 5012.8 | 1663.2 | 665.7 | -862.2 | 3634.7 |
| MEQ | 33263.3 | 17685.4 | 9372.8 | 13225.9 | 15470.0 | 14826.1 | 1461.8 | 569.2 | 4515.9 | 2899.0 | 257.3 | -2883.2 | 7867.2 |
| BZLF1 | -779.3 | -2752.1 | -1963.9 | -271.1 | -384.6 | -520.4 | -1860.7 | -202.9 | -961.0 | -785.6 | -280.1 | -2085.9 | -1992.5 |
| K-bZIP | 254.7 | 2668.3 | 1509.1 | -551.4 | -126.1 | -83.1 | 431.1 | 15.5 | -102.3 | -661.8 | 361.0 | -382.6 | -865.0 |

| | MAFG | MAF | MAFB | NFE2 | NFE2L1 | NFE2L3 | BACH1 | HBZ | MEQ | BZLF1 | K-bZIP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CEBPA | -575.3 | -934.9 | -293.4 | -638.4 | 1073.9 | -148.1 | -368.5 | 1309.2 | -708.9 | -783.2 | 562.8 |
| CEBPB | -133.1 | -727.9 | 105.6 | -1147.5 | -1948.4 | -577.6 | -910.1 | -477.9 | -298.4 | -1051.2 | 625.0 |
| CEBPD | -297.1 | -2206.0 | -2014.8 | -694.6 | -242.4 | -359.5 | -1313.5 | -1656.9 | 516.5 | -2122.3 | 109.3 |
| CEBPG | 173.1 | -852.9 | 231.3 | -1149.2 | -1961.5 | -111.7 | -702.4 | 12122.4 | -628.9 | -1955.2 | 644.0 |
| DDIT3 | 1536.8 | -1347.8 | 611.0 | -417.8 | -1159.9 | -529.6 | 3344.6 | 2995.1 | 11570.3 | -1620.2 | 1211.4 |
| ATF1 | -104.4 | -359.8 | 88.1 | -1059.5 | -1537.7 | -694.8 | 1387.8 | 1919.3 | 1229.4 | -952.4 | 470.0 |
| CREB1 | -2140.3 | -1584.5 | -1828.8 | -2132.1 | -2342.7 | -1817.6 | 1554.8 | 1390.1 | 308.1 | -5005.3 | -1437.0 |
| CREB3L3 | -118.4 | -1526.3 | -162.1 | -442.7 | -1033.6 | -145.8 | -1069.3 | -1271.8 | -1173.2 | -477.1 | 707.3 |
| CREB3 | -11.4 | -216.1 | 313.6 | -403.4 | -1329.1 | -388.5 | -680.3 | -2641.1 | -1215.8 | -1004.0 | 1465.9 |
| ATF6 | -362.3 | -3037.8 | -1734.8 | -1253.4 | -5055.6 | -1213.9 | -1130.1 | -3234.3 | -1562.9 | -1382.3 | -253.3 |
| CREBZF | 319.6 | -456.8 | 785.6 | 3389.6 | 28134.8 | -319.6 | 1476.9 | 22064.3 | -311.6 | -446.1 | 665.6 |
| XBP1 | -184.6 | -2038.6 | -1934.4 | -1843.4 | -3647.8 | -660.7 | -998.4 | -3391.2 | -2531.2 | -932.7 | -703.3 |
| HLF | -979.3 | -1756.4 | -2103.1 | -760.1 | -1636.1 | -497.4 | -143.3 | -1102.4 | -862.1 | -303.6 | 96.2 |
| NFIL3 | 541.3 | -1163.1 | 162.9 | -1204.9 | -2694.3 | -336.3 | 119.2 | -1272.2 | 23075.2 | 782.5 | 1249.1 |
| ATF2 | 296.9 | 403.2 | 1718.0 | -658.4 | 2436.4 | -564.0 | 3625.6 | 16538.5 | 23802.8 | -927.8 | 3657.6 |
| ATF7 | 587.5 | -452.3 | 734.8 | 1374.6 | 6706.9 | -1.9 | 5311.4 | 28123.2 | 19304.9 | -774.6 | 3157.4 |
| JUN | -343.4 | -127.4 | 422.3 | -1870.0 | -1487.1 | -336.9 | -122.0 | 48963.6 | 39542.5 | -1311.6 | 188.0 |
| JUNB | -801.1 | 15.2 | 344.4 | -973.4 | -2384.4 | -387.2 | 225.1 | 40558.8 | 31224.1 | -1748.8 | 21.7 |
| JUND | -66.3 | -1469.6 | 107.6 | -1095.0 | -1442.8 | -381.0 | -292.0 | 41619.3 | 35094.7 | -1585.0 | -153.9 |
| FOS | 614.1 | 1296.3 | 7607.3 | -800.9 | 5250.3 | 11.7 | 2734.4 | -779.8 | 556.6 | 147.2 | 2752.4 |
| FOSL2 | 308.5 | -915.1 | 5471.6 | -1209.8 | 581.1 | -562.3 | 797.1 | -1588.6 | 130.4 | 793.8 | 776.7 |
| ATF3 | 678.7 | 60.3 | 4383.3 | -979.4 | -1967.0 | -1035.5 | -74.9 | 5899.5 | 9460.0 | -2718.6 | -105.5 |
| ATF4 | -221.5 | 6439.6 | 992.2 | -332.4 | 19231.1 | 1636.9 | 104.1 | -315.4 | 1530.2 | -2728.5 | 1150.2 |
| ATF5 | -3506.4 | -1479.8 | -2058.9 | 151.6 | -9.1 | -1057.8 | 254.7 | -2894.6 | -2460.9 | -4593.2 | 284.1 |
| BATF | 279.3 | -3082.1 | 697.3 | -325.0 | 3975.8 | 277.4 | 2975.8 | 232.6 | -1556.4 | -379.3 | 1524.9 |
| BATF3 | 2334.0 | -995.0 | 1245.4 | -1131.8 | 14.5 | -516.3 | 1450.1 | 956.3 | 7311.6 | 218.6 | 724.1 |
| MAFG | 4335.0 | -1075.3 | 879.1 | 10390.6 | 40099.8 | 25157.4 | 23867.6 | 19230.9 | 2409.8 | 294.1 | 702.3 |
| MAF | 15.8 | 19582.3 | 15441.3 | -978.3 | 3799.9 | 46.8 | 3977.6 | 31679.4 | -967.6 | -947.9 | 1099.2 |
| MAFB | -558.1 | 18881.3 | 9776.6 | -676.4 | 5247.4 | -326.2 | 7328.3 | 24522.3 | -2220.7 | -235.6 | 1006.9 |
| NFE2 | 1434.5 | -1266.4 | 21.4 | 18844.8 | -1051.1 | 7592.4 | -732.0 | -2449.0 | 1464.1 | -453.8 | 1604.1 |
| NFE2L1 | 43009.4 | 453.8 | 2918.1 | 1124.9 | 4304.1 | -92.3 | -365.7 | -1764.1 | -1512.3 | -788.9 | 1046.8 |
| NFE2L3 | 52198.0 | -959.1 | -863.3 | 16419.4 | -2762.6 | 37.6 | -1337.8 | -3129.1 | -2098.4 | -2208.9 | -456.8 |
| BACH1 | 20474.3 | 6783.8 | 12128.5 | -886.4 | 323.8 | -273.4 | 2255.3 | -4121.1 | 1847.1 | 1551.6 | 2643.2 |
| HBZ | 11791.3 | 13803.1 | 30779.2 | -1041.8 | 1528.9 | -18.8 | 1256.8 | -1720.9 | 17075.6 | 162.7 | 2144.6 |
| MEQ | 861.3 | -279.1 | 1099.7 | 7238.4 | 2954.4 | -84.3 | 3116.5 | 10837.1 | 6219.6 | 994.6 | 1391.5 |
| BZLF1 | -235.0 | -705.2 | 122.9 | -446.6 | -2773.1 | -448.3 | -1377.4 | -2226.8 | -2414.6 | 9824.6 | 886.0 |
| K-bZIP | -391.5 | -602.8 | 285.6 | -1003.9 | -634.6 | -734.2 | -95.6 | -515.6 | -627.4 | 1041.9 | 12707.8 |

| | MEQ1800 | MEQ625 | MEQ125 | MEQ25 | MEQ5 | MEQ1 | HBZ1800 | HBZ625 | HBZ125 | HBZ25 | HBZ5 | HBZ1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CEBPA | -1956.4 | -1048.8 | -1137.8 | -41.9 | 6.6 | 11.7 | 2384.8 | 1231.9 | -637.9 | -104.5 | 3.9 | 9.2 |
| CEBPB | 2477.9 | 1657.6 | -257.4 | -7.1 | 23.5 | 25.1 | 849.8 | -179.6 | -826.4 | -106.7 | 15.1 | 26.2 |
| CEBPD | 303.1 | 130.4 | -549.8 | -36.6 | 8.0 | 11.9 | -3249.3 | -4236.0 | -1639.3 | -215.2 | -5.9 | 1.3 |
| CEBPG | 650.2 | -220.2 | -551.7 | -19.9 | 9.4 | 14.6 | 36408.3 | 27983.8 | 6476.6 | 180.3 | 12.9 | 19.0 |
| DDIT3 | 18879.9 | 14831.1 | 3871.3 | 148.0 | 55.4 | 30.9 | 7672.1 | 8868.3 | 1329.4 | -41.3 | -1.9 | 7.2 |
| ATF1 | 7006.1 | 2915.6 | -52.6 | -0.3 | 23.6 | 24.3 | 10475.4 | 4560.9 | 115.9 | -42.6 | 17.1 | 31.0 |
| CREB1 | 6904.3 | 2468.8 | -110.8 | -3.6 | 67.1 | 74.4 | 13052.6 | 3906.6 | -547.1 | -149.2 | 45.4 | 66.3 |
| CREB3L3 | -1350.9 | -1580.8 | -1121.0 | -37.2 | 4.8 | 6.9 | -2274.8 | -3930.3 | -1141.3 | -101.8 | -0.9 | 7.1 |
| CREB3 | -132.6 | -970.8 | -359.2 | -42.2 | 14.3 | 17.3 | -1840.8 | -1762.8 | -1190.9 | -71.9 | 6.6 | 16.1 |
| ATF6 | -2339.8 | -2589.6 | -1057.9 | -43.7 | 9.1 | 20.0 | -8566.8 | -4360.4 | -2222.2 | -227.2 | -29.6 | -1.6 |
| CREBZF | 606.3 | 629.6 | -69.9 | -19.4 | 21.9 | 23.9 | 44116.1 | 43809.8 | 18047.3 | 742.3 | 124.8 | 87.2 |
| XBP1 | -2754.9 | -2262.8 | -954.5 | -29.9 | 4.6 | 12.1 | -3006.0 | -5468.3 | -3432.9 | -202.6 | -3.3 | -0.8 |
| HLF | -1104.4 | -552.6 | -850.5 | -33.6 | 2.9 | 12.9 | -1963.8 | -1077.0 | -1917.5 | -149.6 | 11.9 | 0.4 |
| NFIL3 | 34569.3 | 25897.4 | 9320.6 | 500.5 | 132.3 | 53.6 | -787.4 | -3668.9 | -2195.8 | -241.9 | -18.3 | 5.9 |
| ATF2 | 43809.1 | 34865.6 | 11030.8 | 489.5 | 164.9 | 54.7 | 49945.8 | 37822.1 | 7240.8 | 267.3 | 39.1 | 39.6 |
| ATF7 | 33504.2 | 27277.8 | 10417.8 | 650.5 | 176.5 | 67.6 | 48362.3 | 49724.4 | 17226.0 | 1250.1 | 166.1 | 93.2 |
| JUN | 40956.9 | 35828.9 | 19746.9 | 1562.1 | 441.8 | 138.4 | 46284.6 | 48346.8 | 33404.8 | 4710.6 | 608.9 | 271.3 |
| JUNB | 38247.8 | 34017.6 | 14665.1 | 814.4 | 240.6 | 67.4 | 46897.5 | 49292.6 | 30957.0 | 2334.3 | 256.9 | 125.4 |
| JUND | 37200.4 | 35001.7 | 16522.2 | 1067.8 | 289.7 | 85.9 | 43240.8 | 45898.1 | 26671.4 | 2186.6 | 262.9 | 122.8 |
| FOS | 1533.3 | 1001.4 | -315.8 | -2.3 | 22.0 | 19.3 | -1246.9 | -806.4 | -476.1 | -65.4 | 10.9 | 16.6 |
| FOSL2 | 2431.9 | 462.6 | -4.6 | 7.8 | 17.2 | 21.2 | -1947.4 | -3642.8 | -1629.8 | -159.0 | 10.4 | 11.8 |
| ATF3 | 26154.6 | 16520.2 | 3833.8 | 174.4 | 75.2 | 51.3 | 28222.4 | 15280.8 | 1322.2 | 35.1 | 40.9 | 43.4 |
| ATF4 | 9900.7 | 4689.4 | 827.6 | 27.8 | 27.3 | 25.0 | 1286.8 | -1658.2 | -1008.5 | -108.5 | 15.8 | 22.3 |
| ATF5 | -3566.5 | -2411.4 | -1035.8 | -20.2 | 29.5 | 40.8 | -4428.8 | -1554.4 | -853.5 | -83.6 | 33.9 | 39.9 |
| BATF | -3492.7 | -3787.8 | -1855.7 | -114.3 | -13.1 | 2.6 | 882.6 | -3269.1 | -2850.0 | -261.4 | -22.4 | -18.2 |
| BATF3 | 19417.1 | 14058.3 | 4411.0 | 219.3 | 80.8 | 41.4 | 5455.3 | 65.9 | -454.7 | -120.3 | 3.9 | 26.9 |
| MAFG | 3525.3 | 984.3 | -9.8 | -21.8 | 15.9 | 16.0 | 43349.9 | 36152.3 | 11382.6 | 401.1 | 54.8 | 43.5 |
| MAF | -2530.1 | -1860.8 | -1105.5 | -20.2 | 5.8 | 12.3 | 32799.1 | 28090.8 | 10693.6 | 816.9 | 104.6 | 51.8 |
| MAFB | -2983.1 | -3317.3 | -1134.9 | -45.6 | 7.9 | 17.4 | 45207.9 | 33537.3 | 9535.5 | 349.5 | 42.8 | 53.8 |
| NFE2 | 6121.4 | 3238.4 | 272.9 | 23.5 | 19.1 | 16.3 | -2244.5 | -3426.1 | -1732.3 | -145.9 | -13.4 | 5.9 |
| NFE2L1 | -625.6 | -1323.4 | -740.9 | -30.4 | 6.6 | 13.6 | -4047.9 | -1621.0 | -1401.2 | -66.3 | 5.0 | 11.4 |
| NFE2L3 | -3538.4 | -3732.2 | -1775.9 | -93.2 | 4.6 | 19.9 | -4985.8 | -5444.1 | -2290.4 | -204.6 | -6.1 | 22.6 |
| BACH1 | 7620.1 | 4622.3 | 101.1 | 3.7 | 37.4 | 36.1 | -3649.9 | -2560.1 | -1027.5 | -132.0 | 8.4 | 27.5 |
| HBZ | 16059.6 | 14584.4 | 5662.5 | 382.9 | 120.4 | 51.4 | -657.0 | 74.8 | -461.1 | -42.8 | 9.1 | 12.3 |
| MEQ | 12345.1 | 7763.5 | 2379.4 | 83.4 | 38.2 | 24.3 | 25434.4 | 23555.3 | 7755.9 | 406.1 | 55.3 | 35.4 |

| | BZLF1 | BZLF1CT |
|---|---|---|
| CEBPA | -918.0 | -541.4 |
| CEBPB | -1035.4 | -215.1 |
| CEBPD | -1303.9 | -654.4 |
| CEBPG | -1356.1 | -554.1 |
| DDIT3 | -673.9 | -471.4 |
| ATF1 | -1020.0 | -127.6 |
| CREB1 | -1711.8 | -347.5 |
| CREB3L3 | -1068.1 | -575.7 |
| CREB3 | -905.4 | -114.4 |
| ATF6 | -826.5 | -62.1 |
| CREBZF | -884.4 | -232.1 |
| XBP1 | -434.5 | -29.5 |
| HLF | -553.8 | 225.8 |
| NFIL3 | -56.1 | -321.1 |
| ATF2 | -1065.1 | -220.6 |
| ATF7 | -656.3 | -103.8 |
| JUN | -1086.8 | -451.1 |
| JUNB | -1094.2 | -177.3 |
| JUND | -985.3 | -473.9 |
| FOS | -472.6 | 16.3 |
| FOSL2 | -352.9 | -25.4 |
| ATF3 | -1700.5 | -290.6 |
| ATF4 | -1323.3 | -101.4 |
| ATF5 | -1965.4 | -703.5 |
| BATF | -716.5 | -279.3 |
| BATF3 | -532.1 | -396.4 |
| MAFG | 104.1 | 521.1 |
| MAF | -355.6 | -199.3 |
| MAFB | -701.6 | -67.6 |
| NFE2 | -381.4 | -608.1 |
| NFE2L1 | -402.0 | -371.4 |
| NFE2L3 | -2073.9 | -552.9 |
| BACH1 | -545.3 | -451.0 |
| BZLF1 | 3400.1 | 1743.3 |
| BZLF1CT | 8831.0 | 7053.4 |

| | anti-MEQ | anti-MEQ2000 | anti-MEQ1000 | anti-MEQ500 | MEQ | ATF2 | JUN | BATF3 |
|---|---|---|---|---|---|---|---|---|
| CEBPA | -343.6 | -981.7 | -798.0 | -1089.6 | -1860.5 | -1209.5 | -220.3 | 2169.7 |
| CEBPB | -235.4 | -520.4 | -551.9 | -195.6 | -3720.1 | -2318.1 | -543.4 | 2156.1 |
| CEBPD | -502.5 | -1077.4 | -1297.4 | -771.9 | -2151.3 | -3235.8 | -540.0 | 2625.6 |
| CEBPG | 1982.6 | 1989.1 | 1100.7 | 509.0 | -658.6 | 5155.8 | 345.5 | 5172.2 |
| ATF1 | -211.4 | -924.7 | -499.9 | -85.8 | 609.1 | -1961.5 | -947.1 | -2072.6 |
| CREB1 | -3510.3 | -4096.5 | -1946.2 | -740.8 | 535.5 | -6256.7 | -1354.7 | -3906.8 |
| CREB3L3 | -48.9 | -808.8 | -836.8 | -345.1 | -2395.3 | -933.4 | -830.6 | -375.9 |
| CREB3 | -319.6 | -1159.1 | -1048.7 | -580.3 | -1330.6 | -820.6 | -279.0 | 438.1 |
| ATF6 | -751.9 | -2121.8 | -2701.0 | -771.7 | -1940.4 | -3780.2 | -1124.3 | -2020.6 |
| CREBZF | 698.6 | 1197.8 | 769.6 | 510.6 | 2290.8 | -789.8 | 296.4 | 1020.3 |
| XBP1 | -2030.7 | -2413.3 | -2652.4 | -1757.5 | -3871.8 | -2048.6 | -359.3 | -1979.3 |
| HLF | -199.3 | -605.1 | -873.8 | -71.1 | -3105.3 | -2699.3 | -188.0 | 121.6 |
| NFIL3 | -431.0 | -878.9 | -1388.9 | -372.2 | 16377.4 | -1942.6 | -513.1 | 305.7 |
| ATF2 | 15474.9 | 29845.1 | 18609.9 | 10084.3 | 29036.9 | 6994.6 | 4898.9 | 5202.8 |
| ATF7 | 18632.6 | 27526.6 | 17875.0 | 12471.2 | 22867.0 | 9594.1 | 8099.6 | 7512.3 |
| JUN | 4410.5 | 7446.3 | 5109.8 | 3051.3 | 33011.3 | 14760.4 | 1997.3 | 11768.7 |
| JUNB | 1590.3 | 3779.5 | 2474.8 | 1351.5 | 29327.3 | 4758.5 | 449.4 | 9728.8 |
| JUND | 1680.8 | 6417.1 | 3982.6 | 1944.3 | 34845.1 | 9823.1 | 788.6 | 11304.0 |
| FOS | 237.5 | -58.9 | 246.1 | 365.3 | 1014.5 | 13984.2 | 35014.5 | 703.8 |
| FOSL2 | -797.3 | -78.6 | -1304.8 | -2507.3 | -894.6 | 9788.4 | 19126.3 | 266.4 |
| ATF3 | 985.6 | 2340.3 | 1654.7 | 949.5 | 11393.0 | 19928.4 | 12049.6 | 13952.3 |
| ATF4 | 584.3 | 5432.9 | 1864.5 | 1157.2 | 3546.8 | 1728.8 | -1060.4 | 14458.4 |
| ATF5 | -1081.3 | -1536.0 | -2095.3 | -1318.5 | -3394.4 | -4831.8 | -1941.1 | 3954.8 |
| BATF | 1095.1 | 950.3 | 811.1 | -247.8 | -5102.5 | 2149.7 | 13378.4 | 2425.2 |
| BATF3 | 3505.8 | 7043.8 | 3208.4 | 1636.4 | 7014.1 | 9068.7 | 12668.4 | 5699.1 |
| MAFG | 434.1 | -71.2 | -514.9 | -44.6 | 476.1 | -253.8 | 268.3 | 2433.0 |
| MAF | -597.1 | -1622.8 | -869.6 | -473.8 | -2118.5 | -253.3 | 92.8 | -1080.9 |
| MAFB | 59.5 | -1223.2 | -483.7 | -469.0 | -2231.4 | -733.8 | -217.7 | -935.9 |
| NFE2 | -754.6 | -1774.4 | -1274.8 | -1921.6 | -327.3 | -2474.9 | -544.6 | -1826.8 |
| NFE2L1 | -784.6 | -2374.6 | -2141.1 | -1102.9 | -2314.9 | 98.8 | -284.6 | -769.6 |
| NFE2L3 | -486.8 | -1242.3 | -816.4 | -297.1 | -2088.4 | -4486.5 | -1628.7 | -3341.4 |
| BACH1 | 86.9 | -1658.8 | -247.4 | -25.9 | 3391.9 | 6185.9 | 456.6 | 895.5 |
| MEQ | 29778.6 | 28393.8 | 23575.5 | 19969.3 | 3605.4 | 18790.1 | 11986.9 | 4714.5 |
| anti-MEQ | -109.4 | -1402.1 | -1635.9 | -764.9 | 30928.2 | 23309.9 | 4148.3 | 2632.5 |

# APPENDIX B

# Supplementary Information for "Design of protein-interaction specificity gives selective bZIP-binding peptides"

**Reproduced with permission from:**

Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009 Apr 16;458(7240):859-64.

**Author Contributions** GG, AWR and AEK conceived the project. GG developed, implemented and applied the CLASSY formalism and carried out all computational analyses. AWR designed and performed all experiments. All authors analyzed data and guided the research plan. GG and AEK wrote the paper, in consultation with AWR.

## SUPPLEMENTARY METHODS

*Overview of Anti-bZIP Design Using CLASSY*

CLASSY is a computational design procedure for optimizing the stability of a particular structural state as a function of sequence, under an arbitrary number of constraints. It is compatible with many types of potential functions. Any linear analytical function of sequence variables can be constrained; examples include energy gaps towards other structures, or properties such as amino-acid composition or hydrophobicity.

CLASSY is based on two components: cluster expansion (CE) and integer linear programming (ILP) optimization. CE provides a way to express the energy of a sequence adopting a particular backbone structure as an algebraic function of the sequence itself (Grigoryan, et al. 2006). The formal basis of the technique is briefly described in the next section, but two properties of a cluster expansion are important for CLASSY: (1) it makes the evaluation of sequence energies many orders of magnitude faster than with direct structural methods, and (2) its simple functional form renders a new set of computational approaches applicable to protein design. We used CE in conjunction with ILP as a way to incorporate information about undesired states into design calculations.

*Theory of Cluster Expansion*

We have previously shown that the conformational energy of a protein sequence in a specified fold, defined numerically using structural calculations and optimization, can be expressed as a direct function of sequence using the method of cluster expansion (Grigoryan, et al. 2006, Zhou, et al. 2005). For completeness, we briefly describe this method here. Let $E_{\min}(\vec{\sigma})$ be the energy of sequence $\vec{\sigma}$ in a given backbone fold (subscript *min* stands for minimization

over side-chain degrees of freedom). Let $\vec{\sigma} = \{\sigma_1,...,\sigma_N\}$, where $\sigma_i$ is a discrete variable representing the amino acid at the *i*-th position of the sequence. For simplicity, and without loss of generality, assume that in our design problem there are *M* amino-acid possibilities at each position and $\sigma_i$ can take on values from 0 to *M*-1. We can then express $E_{\min}(\vec{\sigma})$ as a cluster expansion of the form:

$$E_{\min}(\vec{\sigma}) = J_o + \sum_{si=1}^{N}\sum_{i=1}^{M-1} J_{si}^{i} \cdot \varphi(si,i) + \sum_{si=1}^{N-1}\sum_{sj=si+1}^{N}\sum_{i=1}^{M-1}\sum_{j=1}^{M-1} J_{si,sj}^{i,j} \cdot \varphi(si,i)\cdot\varphi(sj,j) + \ldots,$$

where $\varphi(si,i)$ is a binary function that evaluates as 1 if site *si* is occupied with amino-acid *i* and zero otherwise. The summations are over sites and amino-acid identities. A collection of sites is referred to as a cluster, and a cluster populated by a given set of amino acids is a cluster function (CF). Terms *J* are the effective contributions of each cluster function to the overall energy (effective cluster interactions, or ECI). The three terms shown correspond to the constant, point and pair cluster-function contributions. If the expansion is written out in its entirety (i.e. up to the *N*-tuple cluster functions), then by virtue of having exactly the same number of ECI as possible sequences ($M^N$), it is exact. If the expansion is truncated at a given point, an approximation of $E_{\min}$ can be derived by fitting the ECI to minimize the error between CE-estimated energies and structure-derived energies for a training set of sequence-energy pairs. Once this procedure is carried out, the process of estimating the energy of a sequence adopting the specified structure is made many orders of magnitude more efficient (Grigoryan, et al. 2006).

*bZIP Models*

To model parallel dimeric coiled coils, we employed two variants of the energy function HP/S/C that was previously shown to perform well in predicting human bZIP interaction specificity

(Grigoryan and Keating. 2006). This function evaluates the relative stability of coiled-coil dimers primarily as a function of the amino acids at **a**, **d**, **e** and **g** positions, based on predicted structures of coiled-coil complexes. One of the key features of model HP/S/C is that core **a-a'** and **d-d'** terms derived from structure-based calculations are replaced with statistical weights from a machine-learning algorithm (Grigoryan and Keating. 2006). These terms can alternatively be replaced by experimentally determined thermodynamic coupling energies. However, these were only available for 15 amino-acid pairs at **a-a'** at the time of our earlier study (Acharya, et al. 2002), and using them gave inferior performance. Since then, Vinson and co-workers have measured coupling energies for 55 amino-acid pairs at **a-a'** (Acharya, et al. 2006a). Additionally, we recognized that almost all of the improvement upon replacing **d-d'** interactions with statistical weights can be attributed to Leu-Leu pairs, which are modelled as only slightly favourable in structure-based approaches, contrary to experimental data. As a result of these findings, we developed model HP/S/Cv. Structure-based **a-a'** interactions were replaced with **a-a'** coupling energies for 55 amino-acid combinations; the **d-d'** interaction for Leu-Leu was replaced with –2 kcal/mol (no experimental value is available), and the resulting model was expanded using CE. Because effective self contributions from our structural models and experimental coupling energies may be on different scales, point ECI values for the **a** position were adjusted such that 100 folding free energies measured by Acharya *et al*. were predicted optimally (in the least squares sense) by the overall CE model – see Figure B.S10.

As a way to account for pair-wise interactions in the reference state, both variant models used in this study ignored the energy of intra-chain side-chain interactions in the final predicted structure (see reference (Grigoryan and Keating. 2006)). Note, however, that because the process of placing side chains for structure prediction does take into account all side-chain interactions,

intra-chain interactions do make indirect contributions to the final energy, and corresponding ECI do emerge in cluster expansion.

*Integer Linear Programming*

Kingsford *et al.* have shown that the problem of finding the lowest-energy rotamer-based side-chain packing arrangement, in the context of protein design, can be expressed and solved as an ILP (Kingsford, et al. 2005). Given that CE provides the energies of the desired and undesired states as analytical functions of sequence, we introduced a similar approach for handling specificity in design. With notation as in Kingsford *et al.*, we represent the sequence space in our problem of designing a peptide of length *p* as an undirected *p*-partite graph with node set $V = V_1 \cup ... \cup V_p$. Set $V_i$ contains one node for each amino-acid possibility at position *i*. For each state *S*, each node $u \in V_i$ is assigned a weight $E_{uu}^S$ corresponding to its contribution to the energy of that state. If *S* is a heterodimer state (i.e. a state in which the design is complexed with a protein of fixed sequence), this individual contribution is simply the sum of the point ECI corresponding to *u* and pair ECI corresponding to pairs between *u* and all amino acids of the partner sequence. If *S* is the design•design homodimer state, then $E_{uu}^S$ is the sum of point ECI corresponding to *u* and pair ECI of *u* and its image on the opposite chain. The edges of the graph $D = \{(u,v): u \in V_i \text{ and } v \in V_j, i \neq j\}$ are assigned weights $E_{uv}^S$. If *S* is a heterodimer state, then $E_{uv}^S$ is the ECI of the corresponding intra-chain pair cluster function. If *S* is the design•design homodimer state, then additional contributions to $E_{uv}^S$ come from the ECI between *u* and the image of *v* as well as *v* and the image of *u*. Given these definitions, the energy of the design

sequence in any state *S* can be expressed as $\varepsilon^S = \sum_{u \in V} E^S_{uu} x_{uu} + \sum_{u,v \in D} E^S_{uv} x_{uv}$, where binary variables $x_{uu}$ and $x_{uv}$ determine which nodes and edges the sequence involves. Thus, the problem of optimizing the energy of state *S* can be expressed as an ILP seeking to minimize $\varepsilon^S$, under the constraint that the chosen nodes and edges correspond to one another. Further, because gaps between different states are also linear functions of decision variables $x_{uu}$ and $x_{uv}$, arbitrary gap constraints can also be incorporated. Finally, any additional function of these decision variables, such as a PSSM score, can also be incorporated. With *T* as the target state and $U_i$ representing undesired states, the ILP we used in this study is (where $V \backslash V_j$ stands for the set difference between *V* and $V_j$):

$$\text{Minimize}: \varepsilon^T = \sum_{u \in V} E^T_{uu} x_{uu} + \sum_{u,v \in D} E^T_{uv} x_{uv}$$

$$\text{subject to}:$$

$$\sum_{u \in V_j} x_{uu} = 1 \quad \text{for } j = 1,..., p$$

$$\sum_{u \in V_j} x_{uv} = x_{vv} \quad \text{for } j = 1,..., p \text{ and } v \in V \setminus V_j$$

$$\varepsilon^{U_1} - \varepsilon^T > gc_1, \text{where } \varepsilon^{U_1} = \sum_{u \in V} E^{U_1}_{uu} x_{uu} + \sum_{u,v \in D} E^{U_1}_{uv} x_{uv}$$

...

$$\varepsilon^{U_k} - \varepsilon^T > gc_k, \text{where } \varepsilon^{U_k} = \sum_{u \in V} E^{U_k}_{uu} x_{uu} + \sum_{u,v \in D} E^{U_k}_{uv} x_{uv}$$

$$\sum_{u \in V} W_u x_{uu} < pssmc$$

$$x_{uu}, x_{uv} \in \{0,1\}$$

Here *k* is the number of undesired states, $gc_i$ is the gap constraint for *i*-th state, *pssmc* is the PSSM constraint and $W_u$ is the PSSM weight corresponding to node *u*. We solved such ILPs with the *glpsol* tool from the GNU Linear Programming Kit (http://www.gnu.org/software/glpk/). Because of the simplicity of sequence-based expressions

obtained through CE, solutions to these ILPs with as many as 46 undesired states were generally obtained within 1-5 minutes on a single 2.7 GHz CPU.

Note that although everything was formulated in this instance for energy functions that are pair-wise decomposable at the sequence level, in principle this approach can be easily generalized for higher-order terms. Clearly, the CE methodology is already capable of taking higher-order interactions into account, should there be a need (Grigoryan, et al. 2006). The ILP formulation can be extended to handle higher-order terms by introducing additional decision variables. For example, $x_{uvw}$ would be 1 if there is a triplet interaction between nodes *u*, *v*, and *w*. Constraints for these new decision variables would also have to be imposed to ensure that higher-order interactions occur only between those nodes that are chosen (e.g. in this case $x_{uu}$, $x_{vv}$ and $x_{ww}$ are 1). Note that these higher-order decision variables would have to be introduced only for those clusters of sites that do, in fact, participate in higher-order interactions. This allows the complexity of the ILP problem to grow naturally with the size of the system (i.e. the number of variables and constraints grows linearly with the number of interactions in the system).

*PSSM Constraint*

To constrain CLASSY designs to favour a leucine-zipper fold, we derived heptad position-specific amino-acid frequencies from the multi-species alignment of 432 bZIP leucine zippers described above. These frequencies were then used to score all of the sequences in the alignment (taking into account only **a**, **d**, **e** and **g** positions), from which a length-normalized score distribution was derived. Based on this distribution, a cutoff value of 0.247 was imposed in CLASSY such that all of the designed sequences had a PSSM score of at least 0.247. Although this is a stringent cutoff, with 84% of native sequences scoring below it, the sequence space

remaining is still large. For example, for a six-heptad design sequence, where **a**, **d**, **e** and **g** positions are varied and 10 amino acids are allowed per position, the total sequence space is $10^{24}$, whereas after applying the PSSM cutoff of 0.247 it is still ~$10^{18}$ (calculated by convolving score distributions at individual positions to obtain the final distribution of scores and integrating it from 0.247 up).

*Choosing b, c and f Positions*

Positions **a**, **d**, **e** and **g** are assumed to encode most of the interaction specificity of the designed peptides (Vinson, et al. 2006, O'Shea, et al. 1992). Thus, we chose the identities of the **b**, **c** and **f** positions such that they were appropriate for the already selected **a**, **d**, **e**, and **g** positions, given what is observed in the multi-species dataset of 432 bZIP sequences referenced above. Thus, for each **b**, **c**, and **f** position $b_i$ we sought to optimize $P(b_i|a_1,...,a_n)$, where $a_1 \ldots a_n$ are the identities of the selected **a**, **d**, **e**, and **g** positions. We expressed this quantity in terms of probabilities we could measure from the dataset:

$$\begin{aligned} P(b_i|a_1,...,a_n) &= \frac{P(b_i,a_1,...,a_n)}{P(a_1,...,a_n)} = \frac{P(a_1|b_i,a_2,...,a_n)\cdot P(b_i,a_2,...,a_n)}{P(a_1,...,a_n)} \\ &= \frac{P(a_1|b_i,a_2,...,a_n)\cdot P(a_2|b_i,a_3,...,a_n)\cdot ...\cdot P(a_n|b_i)\cdot P(b_i)}{P(a_1,...,a_n)} \\ &\approx \frac{P(a_1|b_i)\cdot P(a_2|b_i)\cdot ...\cdot P(a_n|b_i)\cdot P(b_i)}{P(a_1,...,a_n)} \end{aligned}$$

The last step assumes that the pre-selected amino-acid decoration at positions **a**, **d**, **e**, and **g** represents well the natively observed decorations at these positions (i.e. probability $P(a_k|b_i)$ measured in the **adeg** context of the designed peptide and the probability averaged over all native contexts is the same). Quantity $P(a_1,...,a_n)$ is hard to estimate, but it is constant with respect to

**b**, **c** and **f** and is therefore not important. Conditional probabilities $P(a_k|b_i)$ can be easily measured from the native bZIP dataset, and for each **b**, **c** and **f** position the amino acid that optimizes the above probability can be found. Using this approach, we were able to obtain **b**, **c**, **f** decorations of natural content and distribution. However, we found that infrequently this procedure resulted in sequences with large charge and/or helix propensity (mostly due to the fact that the pre-selected **a**, **d**, **e**, and **g** amino acids already had high values of charge or helix propensity). Thus, we expressed the problem of finding the optimal **b**, **c** and **f** combination according to the above equation as an ILP (by taking the logarithm of the probability it can be decomposed into a sum of pre-computed probability logarithms) and incorporated constraints on total charge, charge content (number of charged residues) and helix propensity. For each property, the range of acceptable values was defined as $\mu \pm \sigma$, where $\mu$ and $\sigma$ are the mean and standard deviation of the corresponding property in the native bZIP dataset. In a few instances this resulted in no solutions (i.e. the selected **a**, **d**, **e** or **g** were already outside of the range for one of the properties) and for these cases more liberal intervals were allowed (either $\mu \pm 1.5\sigma$ or $\mu \pm 2\sigma$). Finally, because we wanted to rely on UV absorbance for determining concentration, we imposed the additional constraint that the **b**, **c**, **f** positions contain at least one Y or W residue (unless there was one already present at **a**, **d**, **e** or **g**).

*Uncovering Specificity-encoding Features*

We analyzed the 8 designs determined to be most specific using the arrays to identify specificity-encoding features. First, we compared each design•target complex with the corresponding design•undesired heterocomplexes. For each such comparison, we computed the contribution of each amino acid in the *i*-th position of the design sequence ($aa_i$) to the overall

stability and specificity. This was done by computing the interaction of $aa_i$ with the region of the target peptide from *i*-7 to *i*+7 (one heptad N- and C-terminal to $aa_i$) as well as the interaction of $aa_i$ with the same region of the undesired partner. The first value corresponded to the stability contribution of $aa_i$ and the difference between the two was the specificity contribution. To further isolate specificity determinants, this difference was decomposed into contributions from different positions on the target sequence and the corresponding positions on the undesired partner sequence.

We performed a similar analysis to elucidate features encoding specificity against the design•design homodimer, except the contribution of each amino acid $aa_i$ to specificity was considered as the difference between interaction of $aa_i$ with the residue opposing it in the target sequence and its interaction with itself in the design homodimer. The same analysis was repeated for pairs of amino acids at all position pairs (*i* and *j*) of the design sequence.

*Dividing Human bZIPs into 20 Families*

Human bZIPs were divided into 20 families based on the evolutionary analysis of (Amoutzias, et al. 2007) with the exception of including CHOP and ZF as individual families, and condensing OASIS and OASISb into a single family based on the similarity of their interaction profiles (Newman and Keating. 2003). The phylogenetic tree of human bZIPs shown in Figure B.S13 was made using only the leucine-zipper regions and was constructed with the PHYLIP (http://evolution.genetics.washington.edu/phylip.html) package using the Neighbour-Joining algorithm and the Jones-Taylor-Thornton (JTT) model of amino-acid replacements. TreeDyn (http://www.treedyn.org/) was used to visualize and annotate the tree.

*How Many Unique anti-bZIP Profiles Are There?*

Figure 3.3A shows that our CLASSY designs exhibit many novel interaction profiles when binding human bZIPs, while the sequence diversity used to generate these profiles is rather limited (Figure 3.3C). This suggests that there may be a very large number of different interaction profiles, of which our 48 designs have revealed only a very small portion. But how large is this number? To answer this question with high confidence we need either an extremely large number of designs and measurements or an extremely accurate model. At present, neither is available. However, if we have a good idea of a model's prediction accuracy and use this model to calculate the number of unique profiles that exist, we can then estimate a lower bound on the true number of profiles. Here, we used model HP/S/Cv for this purpose. Several steps were taken to ensure that our estimates were always below the true number of profiles.

The interaction profile of a peptide was defined as a binary vector indicating whether the peptide interacts (1) or does not interact (0) with each human bZIP. If two binary vectors are equal, the profiles are equivalent. In reality, there is a lot of space between such vectors, because interaction strength also plays a role in defining a profile. This is one way that we underestimated the total number of possible profiles. We also defined these vectors at the family level rather than the protein level – again, a significant underestimate of the real size of the profile space. We considered 19 out of the 20 families (due to difficulties assessing model performance on the ATF3 family), giving a total of 524,288 possible unique profiles. The following procedure was followed:

*Compute the total number of unique profiles predicted by HP/S/Cv.* For each human bZIP coiled coil $P_i$ we defined a computational energy cutoff $c_i$ to optimally discriminate interactions and non-interactions in the human bZIP interaction dataset (experimental interactions/non-

interactions taken from (Fong, et al. 2004)). To increase prediction confidence, we introduced a buffer parameter *b*, such that energy scores above $c_i+b$ were considered non-interactions, below $c_i - b$ were considered interactions, and scores between $c_i - b$ and $c_i+b$ were not considered as either (*b* was set to 3 kcal/mol by optimizing performance on the human bZIP interaction dataset). This parameter increases prediction confidence but reduces the number of peptides that can give rise to a profile, further reducing our final estimate. Next, we generated 1,000 random binary profile vectors and ran CLASSY to find the most stable sequence consistent with each profile (e.g. its interaction stability with each of the 40 bZIPs from the 19 considered families is either below $c_i - b$ or above $c_i+b$ in accordance to the profile). The bZIP PSSM constraint was applied. Out of these 1,000 cases, 5 produced a solution. Given that there are a total of 524,288 possible binary profiles, this translates into ~2,600 unique profiles that can be achieved in design.

*Estimate prediction rates*. The rates of true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*) predictions were estimated from anti-bZIP•bZIP interaction data. Performance is expected to be worse than for the human•human dataset for several reasons. First, the process of design tends to exacerbate errors in an energy function. Second, because designed sequences are different from native bZIPs in systematic ways, the ranges of HP/S/Cv scores for anti-bZIP•bZIP and bZIP•bZIP interactions will also be different, making cutoffs derived from the bZIP•bZIP dataset less applicable to anti-bZIP•bZIP interactions. Thus, although the prediction rates for the human•human interactions were *TP* = 0.84, *TN* = 0.91, *FP* = 0.16, *FN* = 0.09, they were worse for the anti-bZIP•bZIP interactions: *TP* = 0.39, *TN* = 0.94, *FP* = 0.61, *FN* = 0.06. The drastic difference between the two performance rates is a result of over-training optimal cutoffs to the case of human•human interactions, but since the most important goal here

is not to over-estimate the performance rate, this approach is still valid. The performance predicting relative stabilities of two complexes of anti-bZIP•bZIP is much better than this.

*Given two predicted distinct profiles, find the probability that they are in fact the same*. This probability, *ps*, is a product of the probabilities that each individual element of the profile (interaction or non-interaction with each human bZIP) is the same. Formally,

$$ps = (TN \cdot TN + FN \cdot FN)^{zz} \cdot (TN \cdot FP + FN \cdot TP)^{oz} \cdot (TP \cdot FN + FP \cdot TN)^{zo} \cdot (TP \cdot TP + FP \cdot FP)^{oo},$$

where *oo*, *oz*, *zo*, and *zz* are the number of corresponding profile elements that are both 1, 0 and 1, 1 and 0, or both 0, respectively. Probability *ps* was estimated to be $2.0 \cdot 10^{-4}$ by averaging over 1,000 pairs of randomly generated profiles.

*Calculate the probability distribution of the true number of profiles*. We predicted that there exist ~2,600 unique profiles. The first one we consider is certainly unique. The second one is predicted to be unique, but it is actually unique with probability 1 – *ps*. The third one is also predicted to be unique, but it is truly different from the first and the second with probability $(1 - ps)^2$. In general, if *P(k, n)* is the probability of having *k* unique profiles after considering *n* predicted unique profiles, then we can give the recursive definition $P(k,n) = P(k,n-1) \cdot k \cdot ps + P(k-1,n-1) \cdot (1-ps)^{k-1}$. Using this we generated the probability distribution of the true number of profiles after considering 2,600 profiles. This distribution had a sharp peak around 1,900 profiles and quickly fell to essentially zero before and after that (integral between 1,785 and 1950 is 0.9999). Based on this, there should exist at least ~1,900 unique peptide•human bZIP interaction profiles, and probably there are many more.

*A Picture of Multi-state Energy Phase Space*

Specificity-sweep calculations predict that designs selected solely for optimal binding to the target are often not specific, and are especially prone to homodimerization (see Figure B.S 2A). Many specificity problems can be eliminated by sacrificing relatively small amounts of stability (Figure B.S2C). However, it is not clear how severe the specificity constraint is and how much it restricts the choice of sequences. We investigated this in a simplified case where design•design homodimers are the only competing state. We constructed a 2D histogram of the entire design sequence space for several design problems, looking at the distribution of design•target energies versus design•design energies. In such a histogram, each 2D bin corresponds to energy ranges for the design•design and the design•target complexes and contains the number of sequences that satisfy these ranges.

If each amino acid at each site made an independent contribution to the total energy, this histogram could be built by convolving the 2D energy histograms of each individual site. However, amino acids at different sites interact with each other. To address this, we used the fact that amino acids more than a heptad apart do not interact in our CE energy expressions. As in the case for independent site contributions, sites were considered one-by-one and their histograms were convolved with the running total. However, at each step energy contributions from both single-residue and pair-wise interactions with residues in the preceding heptad were incorporated. In order to account for the pair-wise terms appropriately, individual histograms were maintained for each unique sequence combination in the preceding heptad. To limit memory usage, only 9 amino acids were considered at each site for this purpose. Note that because positions **b**, **c** and **f** were not explicitly considered in our models, there were a total of $9^4$ = 6,561 possible heptad sequences and 6,561 running total histograms needed to be kept at each stage. In the last step these 6,561 histograms were added to produce the final 2D histogram.

The results for ATF-2 and MafG are shown in Figure B.S12 (other bZIPs produced similar results). The dashed lines show where the design•design and design•target energies are equal. Clearly, most stable sequences are even more stable as homodimers (i.e. are below the line; note log scale), indicating that destabilization of the design homodimer is an extremely severe constraint that limits sequence space by many orders of magnitude.

*Jun family constructs*

The following peptides were used for the Jun family, which have more uniform length than those previously constructed by Newman & Keating.

cJun

MSYYHHHHHHLESTSLYKKAGSGSRKLERIARLEEKVKTLKAQNSELASTANMLREQV
AQLKQKVMNHLE,

JunB

MSYYHHHHHHLESTSLYKKAGSGSRKLERIARLEDKVKTLKAENAGLSSTAGLLREQV
AQLKQKVMNHLE,

JunD

MSYYHHHHHHLESTSLYKKAGSGSGSRKLERISRLEEKVKTLKSQNTELASTASLLREQ
VAQLKQKVMNHLE

*Data Analysis*

Scanned images of slides were analyzed using the program Digital Genome (Molecularware). For each probe the scan at the highest PMT voltage that did not show saturation was used for analysis. The signal in the red channel from the Alexa Fluor 633

hydrazide was used to identify the location of spots. The median signal and median background for each spot was determined, and signal less background for each spot was calculated. Missed spots and artifacts were manually flagged and removed from analysis; these represented less than 0.1% of all spots. For each pair of adjacent sub-arrays probed with the same labeled peptide, the average of 8 measurements for each protein on the surface was calculated and defined as *a*. These values are reported in Tables B.S3 – 5.

Two other quantities were used in analyses. Because a small number of probes showed high background, a corrected fluorescence signal was defined as F = *a* - $\tilde{a}$ , with $\tilde{a}$ the median of all signals measured using a common probe. The maximum of this quantity for a given probe was designated $F_{max}$. The quantity $-\log(F/F_{max})$ was used in Figure 3.2, Figure 3.3A, and Figure B.S1 and B.S14 to indicate relative array signal differences.

To distinguish signal from noise, and thus put an approximate lower bound on the signal required as evidence for an interaction, we defined the quantity $S_{array}$ as $S_{array}(a) = \frac{(a - \tilde{a})}{\sqrt{\sum_{i=1,\, a_i<\tilde{a}}^{N} (a_i - \tilde{a})^2 \Big/ N_{a<\tilde{a}}}}$, where $\tilde{a}$ is again the median of *a*, *N* is the number of unique printed proteins, and $N_{a<\tilde{a}}$ is the number of proteins producing *a* below the median. *N* and $N_{a<\tilde{a}}$ excluded other designed peptides on the surface when the solution probe was itself a designed peptide. $S_{array}$ is a Z-score-like quantity, where the distribution of signals below the median was assumed to be primarily noise-driven and thus was used to correct stronger signals. $S_{array}$ values are also provided in Tables B.S3 – 5.

For the purpose of estimating the number of designs that homodimerize, and how many designs interacted with their target, the following criterion was used: A and B were judged to

give signal above background, and thus to interact, if they produced an $S_{array}$ score above 2.5 either when A was on the surface and B was the probe or when B was on the surface and A was the probe. This cutoff was chosen based on reported homodimerization of bZIP families as well as our solution measurements of stability (Newman and Keating. 2003, Acharya, et al. 2006b, Vinson, et al. 2002).

*Interaction-Profile Clustering*

An interaction profile was defined using $-\log(F/F_{max})$ scores derived from microarrays, and profiles were clustered using Eucledian distance as the dissimilarity metric. Average linkage clustering was performed using the *linkage* command in Matlab 6.5.

*Circular Dichroism*

Circular dichroism (CD) spectra were measured on AVIV 400 and 202 spectrometers in 12.5 mM potassium phosphate (pH 7.4)/150 mM KCl/0.25 mM EDTA/1M GuHCl/1 mM DTT. All mixtures of peptides were incubated at room temperature for several hours before measurement. Wavelength scans were performed at 40 μM total peptide concentration and measured at 25 °C in a 1-mm cuvette. Scans were monitored from 280 nm to 195 nm in 1 nm steps averaging for 5 seconds at each wavelength. Three scans for each sample were averaged. Thermal unfolding curves were performed at 4 μM total peptide concentration and measured in a 1-cm cuvette. Melting curves were determined by monitoring ellipticity at 222 nm with an averaging time of 30 seconds, an equilibration time of 1.5 minutes, and a scan rate of 2 °C/min. All samples were measured from 0 °C to 85 °C unless otherwise noted. All thermal denaturations were reversible. $T_m$ values were estimated by fitting thermal denaturation data to a monomer-

dimer equilibrium, assuming no change in heat capacity upon folding. Specifically, we fit the derivative of the CD signal with respect to temperature to the equation:

$$\frac{d(signal)}{dT} = A \cdot \frac{\Delta H}{RT^2} \cdot \exp\left(-\frac{\Delta H}{R}\left[\frac{1}{T_m} - \frac{1}{T}\right]\right)\left[\frac{4\exp\left(\frac{\Delta H}{R}\left[\frac{1}{T_m} - \frac{1}{T}\right]\right)+1}{\sqrt{8 \cdot \exp\left(\frac{\Delta H}{R}\left[\frac{1}{T_m} - \frac{1}{T}\right]\right)+1}} - 1\right].$$

Here $A$, $\Delta H$, and $T_m$ were fitting parameters, with $\Delta H$ and $T_m$ corresponding to the change in enthalpy upon folding and the apparent melting temperature, respectively. We fit the derivative of the CD signal to reduce the reliance of the fit on pre- and post-transition baselines (John and Weeks. 2000). For two-species mixtures AB, the difference between the melting curve of the AB mix and the average of melting curves of A and B ($S_{AB\text{-}A\text{-}B}$) was calculated and treated as the signal for the purposes of fitting the above equation. No fitting was performed for mixtures where $S_{AB\text{-}A\text{-}B}$ was positive at any point during the unfolding transition (i.e. the signal from the average was stronger than the signal from the mixture), as it was not clear which species was being melted. Those mixtures with $S_{AB\text{-}A\text{-}B} > 0$ over the entire temperature range were assumed to show no evidence of interaction. Fitting was performed using the non-linear least squares method in Matlab 6.0. The 95% confidence intervals resulting from the fits are reported in Table B.S2.

*Comparing CD and Array-based Stability Ordering*

Relative stability orders established by CD and microarray were compared conservatively. The arrays were only used to judge relative stabilities when two interactions involved the same solution probe interacting with partners on the same array surface. CD ranks

were determined by visual inspection of thermal melts, with cases where the order was not clearly obvious being assigned the same rank. Array ranks for interactions sharing a common probe were established based on the $S_{array}$ measure, with ranks differing by only one unit in normalized $S_{array}$ considered the same. All possible pair-wise comparisons of CD and array ranks were made, a total of 41 comparisons, 35 of which gave the same order by CD and microarray.

*Array Results were Highly Reproducible*

The array measurements were highly reproducible over replicate experiments and a range of concentrations, as shown in Figure B.S14. The complete array data (averaged background-corrected signals as well as $S_{array}$ scores) are given in Tables B.S3-6. Proteins listed in columns were fluorescently labelled and used in solution as probes against proteins on the surface, which are listed in rows. All protein probes were at 160 nM unless otherwise noted. Duplicates are labelled. Tables B.S3-5 contain values from experiments in rounds 1, 2, and 3 respectively. Table B.S6 contains experimentally determined $S_{array}$ scores for 33 human proteins.

## SUPPLEMENTARY DISCUSSION

*Beyond bZIPs: Requirements for Applying CLASSY to Other Systems*

There are a variety of reasons that we selected bZIP transcription factors for this study. They comprise a biologically important class of proteins for which questions of interaction specificity are central to function. But also, interaction specificity is probably better understood for the bZIPs than for any other protein complex, and convenient properties of these proteins facilitate modelling and measurement. To what extent can CLASSY be applied to other problems in molecular recognition? To answer this it is important to distinguish between limitations that

arise from CLASSY itself – of which there are few – and limitations that arise from our understanding of specificity in other protein complexes. The systematic study of protein interaction specificity is a new, expanding research area. There are already several complexes amenable to study using CLASSY, and this number will increase with advances in experimental screening technologies and computational modelling.

Below we outline three requirements that must be met to apply CLASSY to a specificity design problem. For each, we comment on how the bZIPs satisfy the requirement and discuss prospects for other complexes.

1. Application of CLASSY requires that sets of desired and competing states be defined.

To address interaction specificity explicitly, one must define the universe of relevant complexes. For many problems, competing states of particular interest can be identified as those that share structural and evolutionary similarity with the target. In our bZIP application, the competitors were other bZIPs. These can be detected easily by sequence similarity. Many related interaction specificity problems can be posed. In the design of peptides to activate specific integrins, the competitors would be other integrins; in the design of specific PDZ domains the competitors would be undesired protein C-terminal peptides; in the design of BH3 peptides that bind specific Bcl-2 family members, the competitors would be other Bcl-2 proteins. Although criterion 2 (below) may not yet be satisfied for these examples, at least one prior example of a successful design calculation in each of these cases illustrates progress in modelling and highlights the types of applications where CLASSY may prove fruitful (Yin, et al. 2007, Reina, et al. 2002, Fu, et al. 2007). Similar examples can be constructed for any set of paralogous interaction domains; zinc-finger and homeodomain transcription factors as well as SH2, SH3 and

PDZ domains are discussed below.

2. A scoring function must provide information about the relative stabilities of the states under consideration.

Specificity can be designed using CLASSY only if a model captures information about the relative favourability of different states. CLASSY can use many types of scoring functions. Physical/structure-based models and empirical/statistical models are equally compatible with the requirements of the method. The only formal requirement is that the scoring function be expressed as a linear function of sequence variables (not necessarily limited to amino-acid pair terms). We have demonstrated that cluster expansion can accomplish this for complex structure-based energy functions and for several different protein folds (Grigoryan, et al. 2006, Zhou, et al. 2005, Apgar, et al. 2009). Cluster expansion can in theory also be applied directly to large experimental datasets, where available, to generate a predictive expression in the appropriate computational form.

In designing anti-bZIPs, we took advantage of experiments that elucidated some of the determinants of interaction specificity; we captured these in a hybrid structure-based/experiment-based model, which was tested using available peptide array data (Grigoryan and Keating. 2006, Newman and Keating. 2003). Specificity-scoring functions published for other protein domains can now be tested using CLASSY. For example, models based on fitting residue interactions to experimental data have been developed for PDZ domains and zinc fingers. Such scoring functions typically have the functional form required for CLASSY (Stiffler, et al. 2007, Wiedemann, et al. 2004, Kaplan, et al. 2005, Chen, et al. 2008). Scoring functions based on structural modelling have the greatest potential to be general. RosettaDesign has been used for

many applications, including the design of specific protein-protein interactions (Kortemme and Baker. 2004, Kortemme, et al. 2004). Other structure-based specificity models have been tested for PDZ (Reina, et al. 2002), SH2 (Sanchez, et al. 2008) and SH3 (Hou, et al. 2008, Hou, et al. 2006) domains. Structure-based models have also shown good performance for several transcription factor families (Jamal Rahi, et al. 2008, Siggers and Honig. 2007, Paillard, et al. 2004, Morozov, et al. 2005). Physical structure-based models face significant challenges, in particular capturing side-chain and backbone relaxation that can impact specificity. But as new methods for modelling structural relaxation are developed (and several groups report progress in this area (Das and Baker. 2008, Smith and Kortemme. 2008, Friedland, et al. 2008)), there are no obvious barriers to employing them in conjunction with CLASSY. In fact, we recently demonstrated that cluster expansion works well when applied to models that incorporate backbone flexibility (Apgar, et al. 2009). Finally, structural approaches that use atom-based or residue-based statistical potentials can give good predictions of binding energies and can capture some interaction specificity trends (Zhou and Zhou. 2002, Aloy and Russell. 2002, Apgar, et al. 2008); such models may prove especially useful for negative design.

How good do the scoring functions need to be? Our bZIP scoring functions, while capable of distinguishing strong interactions from non-interactions, do not provide quantitative predictions of relative stability (they do not correlate strongly with experimental $\Delta\Delta G$ estimates). Models can likely be effective for use in CLASSY if they (1) accurately capture some key specificity determinants and (2) are not under-defined. A model is under-defined if it has many missing or inappropriate weights; these can allow the design optimization calculations to proceed into non-sensible regions of sequence space. In our bZIP study, the experiments of Vinson and colleagues provided valuable data contributing to (1), though these experiments did not

comprehensively assess all possible specificity determinants (Acharya, et al. 2006a, Vinson, et al. 2006). To address (2), we used structural modelling to impose a physically realistic description of all amino-acid interactions that were not defined by experiments. A similar combined approach is likely to be appropriate for other domains. For example, for PDZ domains and zinc fingers, a small set of weights derived from experiments seem to predict much of the observed specificity (Wiedemann, et al. 2004, Chen, et al. 2008). But structural modelling may be required to provide reasonable (even if not highly accurate) estimates for the many amino-acid interactions that are not constrained by experiments. Also important for addressing (2) is the ability of CLASSY to incorporate sequence property constraints (e.g. the PSSM constraint used in this study), which can be used to ensure that only the sequence space that is reasonably well described by the underlying model is considered in design.

Finally, energy gaps in CLASSY can be chosen according to the estimated accuracy of the underlying energy function. Thus, if errors in predicted energies are known to be large, the user can choose to impose large energy gaps as constraints, ensuring that any designs returned are predicted to have a significant preference for the desired state over others (at the risk of finding either no solutions or only poorly stable solutions).

In summary, while we do not yet know if breakthroughs in predicting specificity will come primarily from improvements in modelling or from fitting to large experimental data sets, this likely does not matter in terms of applying CLASSY. Designing specific PDZ/SH2/SH3 domains or specific PDZ/SH2/SH3 ligands, or zinc-finger transcription factors with specialized binding profiles, are already good candidate applications for testing this method more broadly.

3. An experimental assay appropriate for testing the specificity of the proteins under study is

required.

It is impossible to know the quality of the scoring function, or the quality of CLASSY designs, without experiments that report on interaction specificity. Assessing specificity profiles generally involves testing many possible complexes. For the bZIPs, we took advantage of a previously validated peptide microarray assay (Newman and Keating. 2003). Similar large data sets exist for SH2, SH3, PTB, and PDZ domains, as well as for many transcription factors (Jones, et al. 2006, Matys, et al. 2003, Spaller. 2006, Tonikian, et al. 2008, Noyes, et al. 2008, Berger, et al. 2008). Exciting advances using SPOT arrays, protein microarrays, protein-binding DNA arrays, phage-display/phage ELISA, protein complementation assays and plate-based fluorescence assays expand the possibilities in this area, and suggest that many moderately sized binary complexes will be amenable to analysis (Newman and Keating. 2003, Stiffler, et al. 2007, Wiedemann, et al. 2004, Jones, et al. 2006, Tonikian, et al. 2008, Noyes, et al. 2008, Berger, et al. 2008, Landgraf, et al. 2004, Tarassov, et al. 2008, Remy and Michnick. 2006).

*CLASSY Introduces Negative Design Using Familiar bZIP Features*

CLASSY designs employed a range of strategies to achieve specificity, but some trends were evident. Designs optimized for stability alone often had **a** and **d** positions with medium-to-large hydrophobic residues (Acharya, et al. 2006a), and CLASSY initially improved specificity by maintaining these cores and modulating electrostatic **g-e'** interactions in early iterations of the specificity sweeps (see Figure 3.1C for definitions of coiled-coil heptad positions; a prime indicates a residue on the opposite helix). To achieve greater specificity (Δ), at a greater price in stability, CLASSY introduced core substitutions such as pairing of Ile with Ala (e.g. to destabilize homodimers using Ala-Ala pairs). The sequences selected for testing typically

included additional elements, such as charged amino acids in core positions. Such interactions imparted large amounts of specificity but were also predicted to be quite destabilizing. They were chosen for analysis because we judged specificity to be relatively more important; generic strategies such as ACID extensions could be used to improve stability if necessary (Ahn, et al. 1998).

Our 8 most specific designs exhibit canonical bZIP specificity determinants (Figure B.S 15A): there is a strong preference for Asn at an **a** position to be paired with Asn at the opposing **a′**, and electrostatic complementarity is exploited at **g-e′** positions (Grigoryan and Keating. 2006, Vinson, et al. 2006). Interestingly, a less recognized complementarity between **g-a′** positions is predicted to make a comparable, if not larger, contribution to specificity; this feature was extensively used in our designs (Figure B.S 15A) (McClain, et al. 2002). A strong preference for Leu-Leu over all other amino-acid pairs at **d-d′** positions was also exploited (Moitra, et al. 1997). Finally, our model predicts that interactions between **a** and **d′** can contribute significantly to specificity. In particular, a beta-branched residue at an **a** position strongly prefers a non-beta branched residue at the next **d** position of the opposing strand. Similar effects have been noted in anti-parallel coiled coils (Hadley, et al. 2008).

*Off-target Interactions May Form via Structures That Were Not Modelled*

In our computational modelling, we considered only parallel coiled-coil dimer structures with a unique axial alignment of helices. For the designs that bound to their targets, it is likely that the interaction occurred as modelled because the designs were restrained to have leucine zipper-like sequences, frequently retained buried Asn and Lys residues to favour dimers over other oligomers, and retained paired Asn residues at **a-a'** positions to favour particular parallel

alignments (Oakley and Kim. 1998, Gonzalez, et al. 1996). These features were selected automatically by CLASSY in most cases, and where they were not present in all candidate designs, we imposed a bias for such solutions when choosing examples for experimental testing. Further supporting the formation of dimers, interactions of designs with their targets were observed to occur irrespective of which peptide was printed on the array and which was labelled in solution, which is unlikely for some alternate stoichiometries.

When unexpected design•off-target interactions occurred, it is less clear what the structures of those complexes were. In several instances, we suspect that the complex formed was not one that was modelled as an undesired state. For example, the strong interaction between anti-SMAF-2 and ATF-4 (Figure B.S1) was predicted to be very unfavourable relative to anti-SMAF-2•MafG (Figure B.S16A-B). However, because the SMAF family has an Asn in a different heptad than most human bZIPs, the alignment used to model anti-SMAF-2 paired with ATF-4 left two asparagines at **a** positions unpaired (see Figure B.S16A). Asn residues have a strong preference to occur in pairs in coiled-coil dimers (Acharya, et al. 2006a), and it is unlikely that the anti-SMAF-2•ATF-4 interaction would occur in this way. More likely, the complex would adopt a shifted axial alignment (though this is also predicted to be unfavourable, Figure B.S16C), an anti-parallel helix orientation, or some other structure. Anti-BACH2-2, which showed strong homo-association on the array, illustrates another case where the complex formed may not be the one that was modelled as an undesired state. Anti-BACH-2 homodimer was predicted to be much less stable than anti-BACH-2•BACH1. However, although anti-BACH-2 has very strong anti-homodimerization features, they are heavily concentrated in the first two N-terminal heptads (see Figure B.S17). It is likely that this portion of the homodimer simply does not fold, and the rest of the sequence forms a stable association. Of course, if such problems can

be anticipated, additional constraints can be incorporated into CLASSY, where alternative alignments, coiled-coil lengths and orientations can be explicitly considered.

## SUPPLEMENTARY EXPERIMENTS

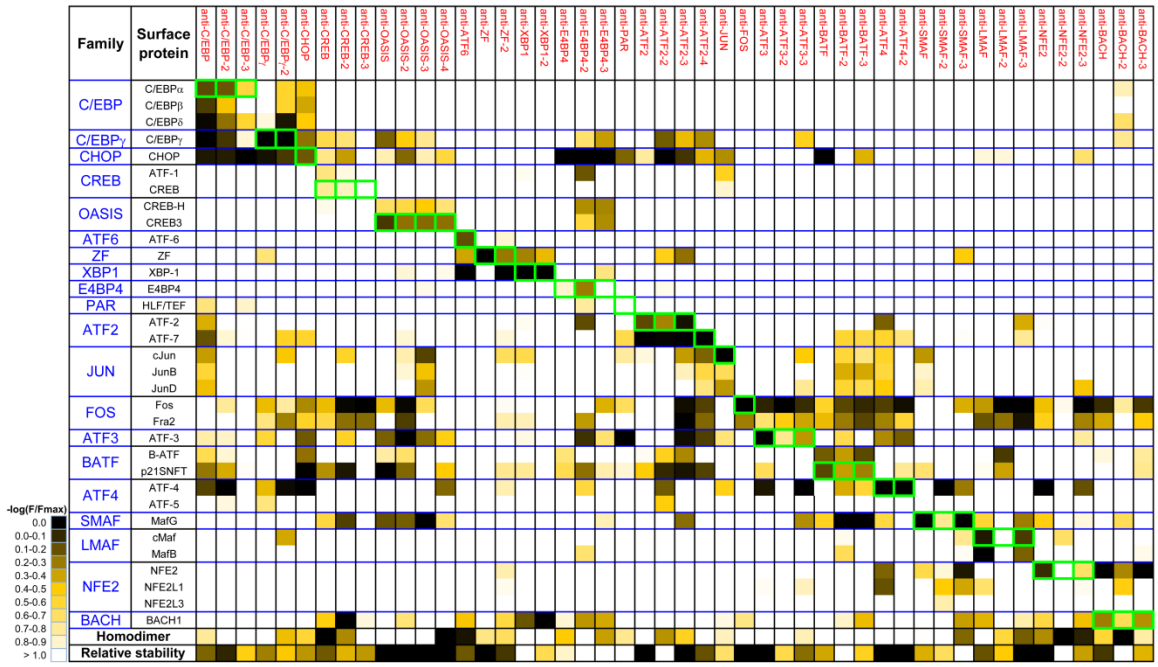
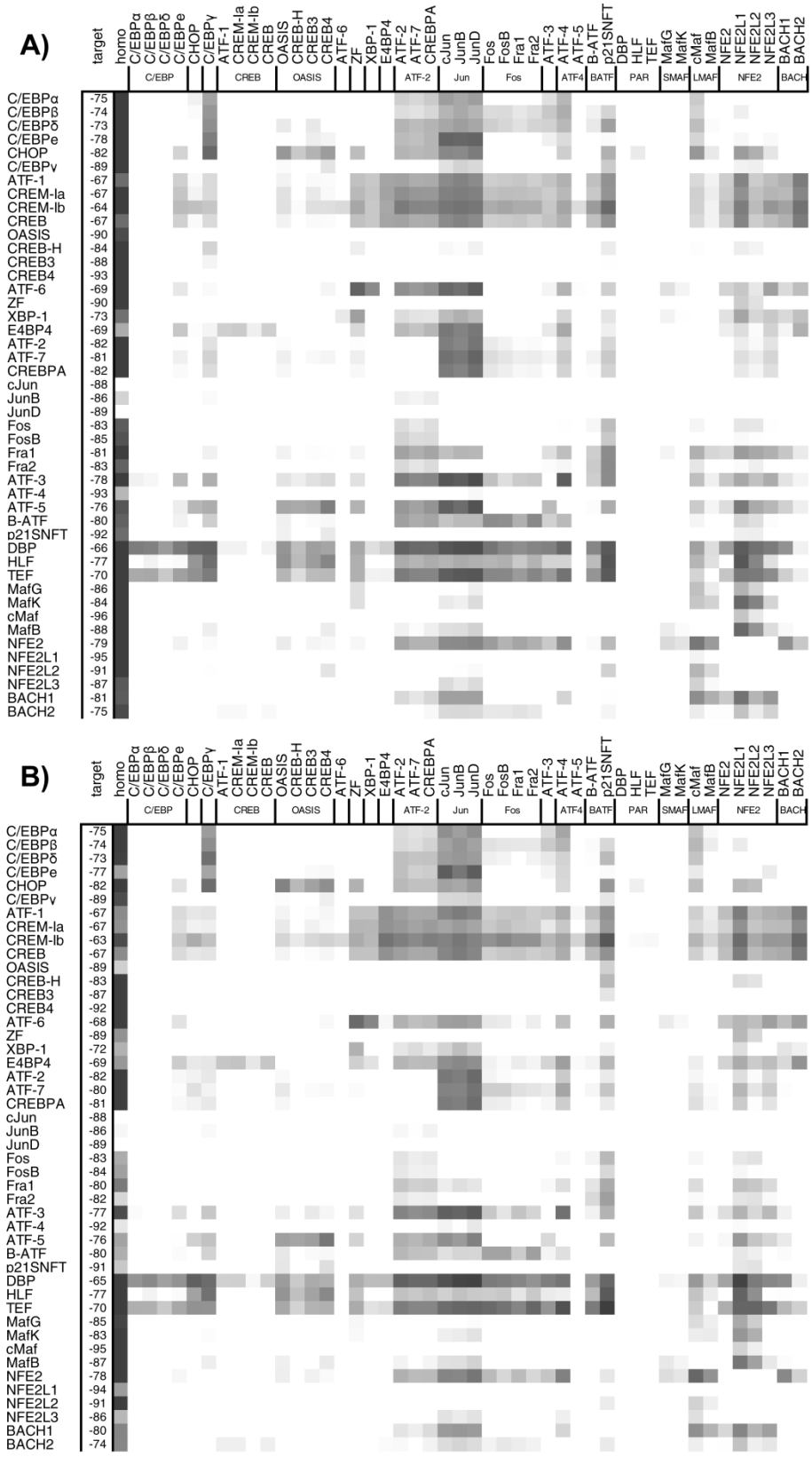
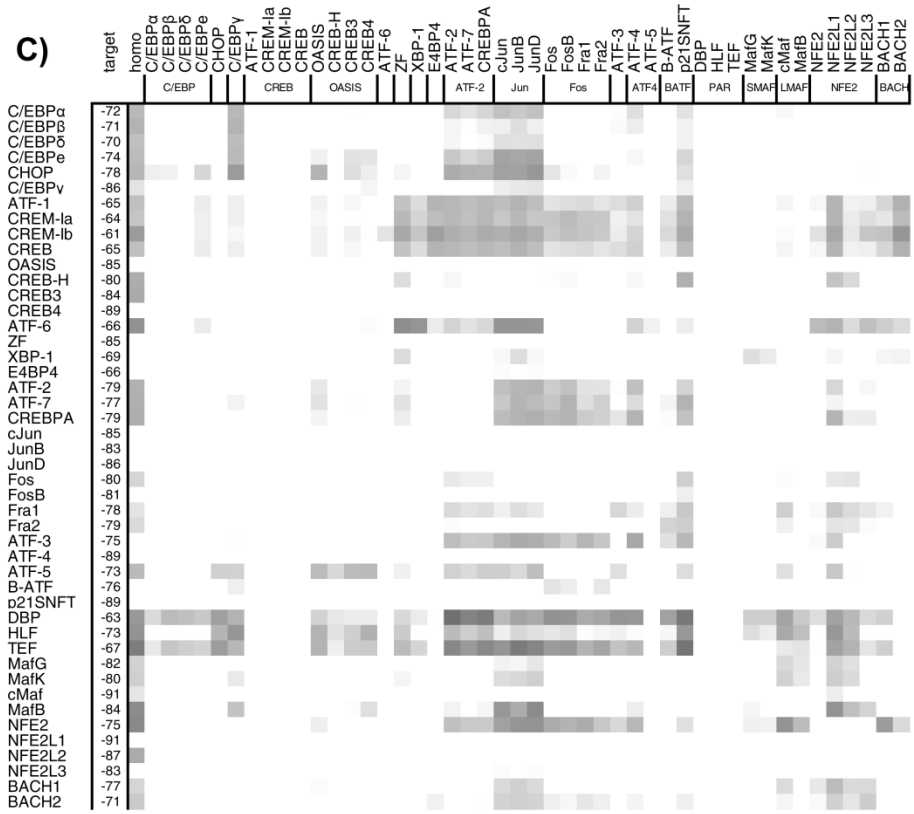
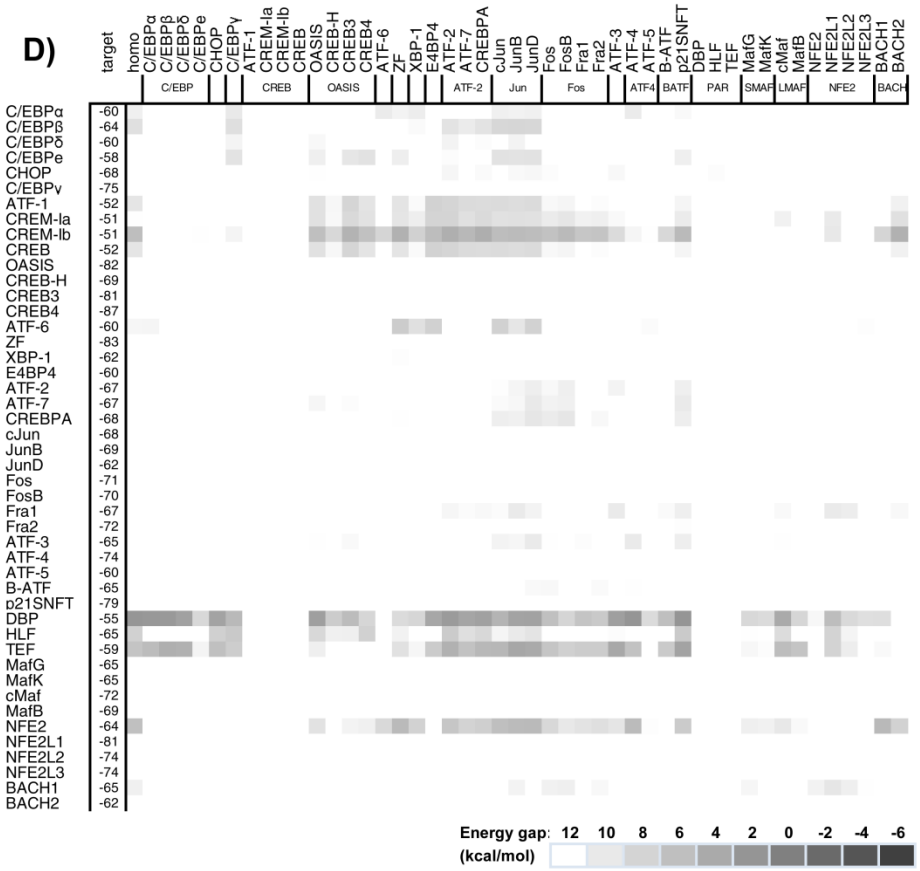
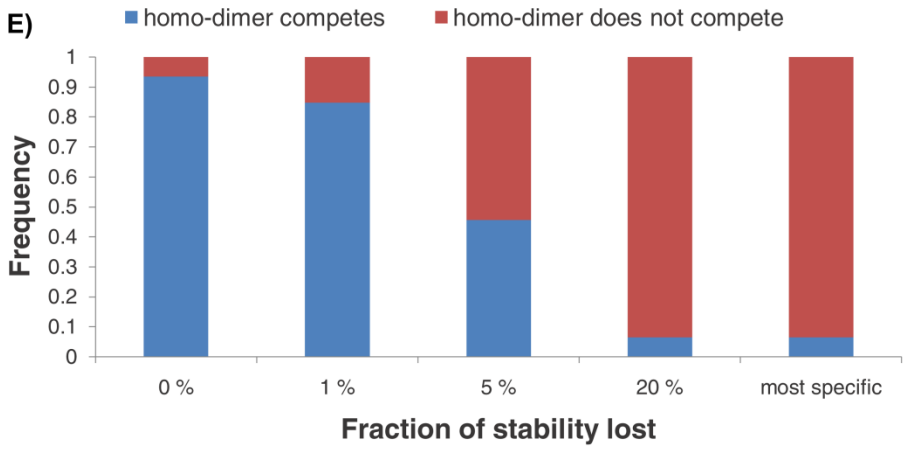



Figure B.S1. Array measurements characterizing all 48 designs. Designs are in columns. Human bZIPs on the arrays are in rows. Family names are in blue, with families separated by blue lines. Shown as a heat map are interaction $-\log(F/F_{max})$ scores (see section Data analysis), with lower scores (darker color) indicating stronger interactions. The "homodimer" row indicates the interaction of each design in solution with itself on the array, relative to the strongest interaction of that design with other partners on the array. The "relative stability" row indicates the interaction of each surface-attached design with its target in solution, relative to the target's strongest interaction (either the design or one of 33 human bZIPs on the same array). Green boxes indicate intended targets.


A)
target
homo
C/EBPα
C/EBPβ
C/EBPδ
C/EBPe
CHOP
C/EBPγ
ATF-1
CREM-Ia
CREM-Ib
CREB
OASIS
CREB-H
CREB3
CREB4
ATF-6
ZF
XBP-1
E4BP4
ATF-2
ATF-7
CREBPA
cJun
JunB
JunD
Fos
FosB
Fra1
Fra2
ATF-3
ATF-4
ATF-5
B-ATF
p21SNFT
DBP
HLF
TEF
MafG
MafK
cMaf
MafB
NFE2
NFE2L1
NFE2L2
NFE2L3
BACH1
BACH2
C/EBP
CREB
OASIS
ATF-2
Jun
Fos
ATF4
BATF
PAR
SMAF
LMAF
NFE2
BACH
B)


C)
target
homo
C/EBPα
C/EBPβ
C/EBPδ
C/EBPe
CHOP
C/EBPγ
ATF-1
CREM-la
CREM-lb
CREB
OASIS
CREB-H
CREB3
CREB4
ATF-6
ZF
XBP-1
E4BP4
ATF-2
ATF-7
CREBPA
cJun
JunB
JunD
Fos
FosB
Fra1
Fra2
ATF-3
ATF-4
ATF-5
B-ATF
p21SNFT
DBP
HLF
TEF
MafG
MafK
cMaf
MafB
NFE2
NFE2L1
NFE2L2
NFE2L3
BACH1
BACH2
C/EBP
CREB
OASIS
ATF-2
Jun
Fos
ATF4
BATF
PAR
SMAF
LMAF
NFE2
BACH
C/EBPα -72
C/EBPβ -71
C/EBPδ -70
C/EBPe -74
CHOP -78
C/EBPγ -86
ATF-1 -65
CREM-la -64
CREM-lb -61
CREB -65
OASIS -85
CREB-H -80
CREB3 -84
CREB4 -89
ATF-6 -66
ZF -85
XBP-1 -69
E4BP4 -66
ATF-2 -79
ATF-7 -77
CREBPA -79
cJun -85
JunB -83
JunD -86
Fos -80
FosB -81
Fra1 -78
Fra2 -79
ATF-3 -75
ATF-4 -89
ATF-5 -73
B-ATF -76
p21SNFT -89
DBP -63
HLF -73
TEF -67
MafG -82
MafK -80
cMaf -91
MafB -84
NFE2 -75
NFE2L1 -91
NFE2L2 -87
NFE2L3 -83
BACH1 -77
BACH2 -71


D)
target
homo
C/EBPα
C/EBPβ
C/EBPδ
C/EBPε
CHOP
C/EBPγ
ATF-1
CREM-Ia
CREM-Ib
CREB
OASIS
CREB-H
CREB3
CREB4
ATF-6
ZF
XBP-1
E4BP4
ATF-2
ATF-7
CREBPA
cJun
JunB
JunD
Fos
FosB
Fra1
Fra2
ATF-3
ATF-4
ATF-5
B-ATF
p21SNFT
DBP
HLF
TEF
MafG
MafK
cMaf
MafB
NFE2
NFE2L1
NFE2L2
NFE2L3
BACH1
BACH2
C/EBP
CREB
OASIS
ATF-2
Jun
Fos
ATF4
BATF
PAR
SMAF
LMAF
NFE2
BACH
Energy gap (kcal/mol): 12 10 8 6 4 2 0 -2 -4 -6


E)
homo-dimer competes
homo-dimer does not compete
Frequency
0 %
1 %
5 %
20 %
most specific
Fraction of stability lost

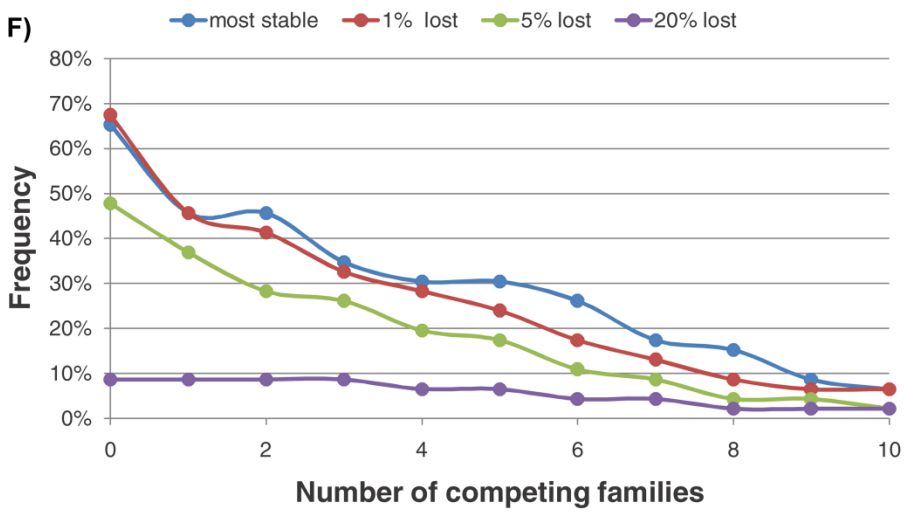



Figure B.S2. A global view of specificity sweeps with each human bZIP coiled coil as a target. In each row, the protein indicated at left is the target. The first column contains the score of the optimal design•target complex, whereas each subsequent column contains the energy gaps between the design•target complex and the corresponding design•competitor complex, including the design homodimer in the second column. A positive energy gap corresponds to design•target being more favorable than design•competitor. The color bar gives the energy scale. **(A)**, **(B)**, **(C)** and **(D)** correspond to designs from different stages of specificity sweeps. In **(A)** the design producing the most stable complex for each target was used to compute energies (first iteration). In **(B)** up to 1% of the stability score was sacrificed to gain specificity. In **(C)** up to 5% of stability was sacrificed and in **(D)** the most specific designs were considered. In **(E)** and **(F)** the specificity data are summarized as a function of decreasing stability. **(E)** shows the proportion of anti-human designs for which the design•design homodimer has a gap of less than 6 kcal/mol, and **(F)** shows the proportion of designs predicted to compete with a non-target-family human bZIP by the same criterion. Energies were computed using model HP/S/Cv.

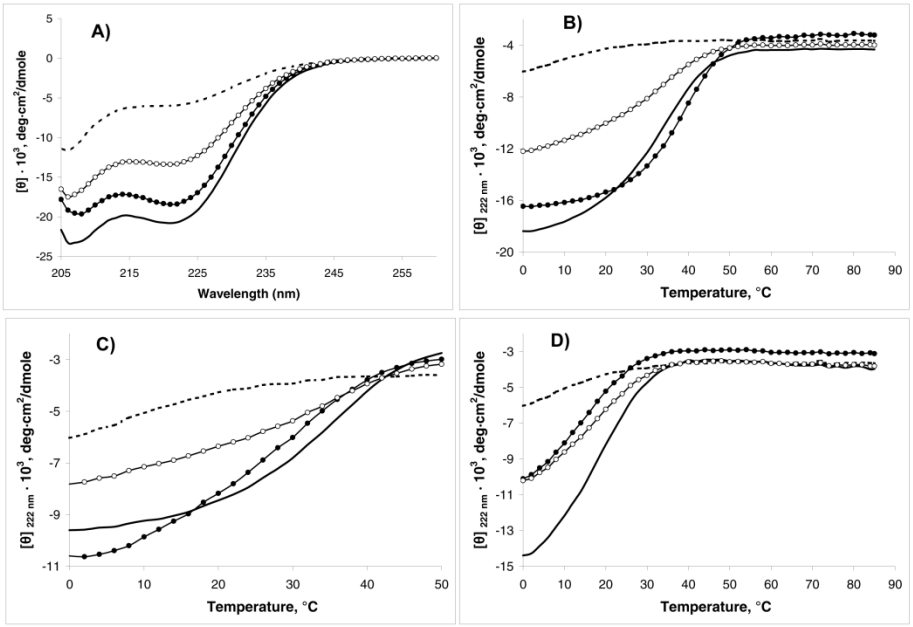



Figure B.S3. Solution characterization of anti-ATF2 by CD. Format and presentation is the same as in Fiigure 3.**2**B-E for anti-SMAF. The target protein is ATF-7 (which is in the same family as ATF-2) (in **A** and **B**), the closest off-target competitor is p21SNFT (in **C**), and the bZIP related to the target by sequence is cJun (in **D**). $T_m$ values are given in Table B.S2.

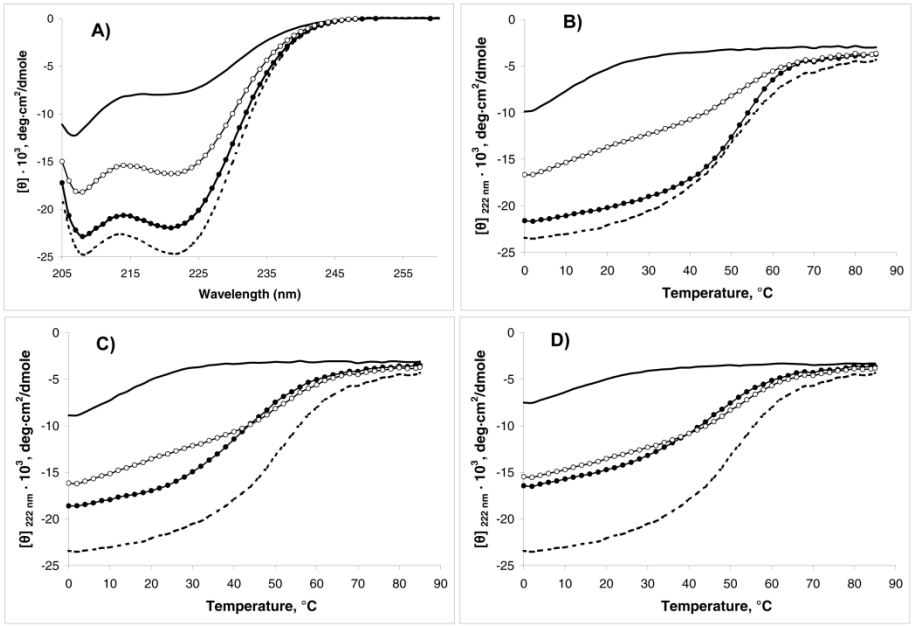



Figure B.S4. Solution characterization of anti-ATF4 by CD. Format and presentation is the same as in Figure 3.**2**B-E for anti-SMAF. The target protein is ATF-4 (in **A** and **B**), the closest off-target competitor is Fos (in **C**), and the bZIP related to the target by sequence is ATF-3 (in **D**). $T_m$ values are given in TableB.S2.

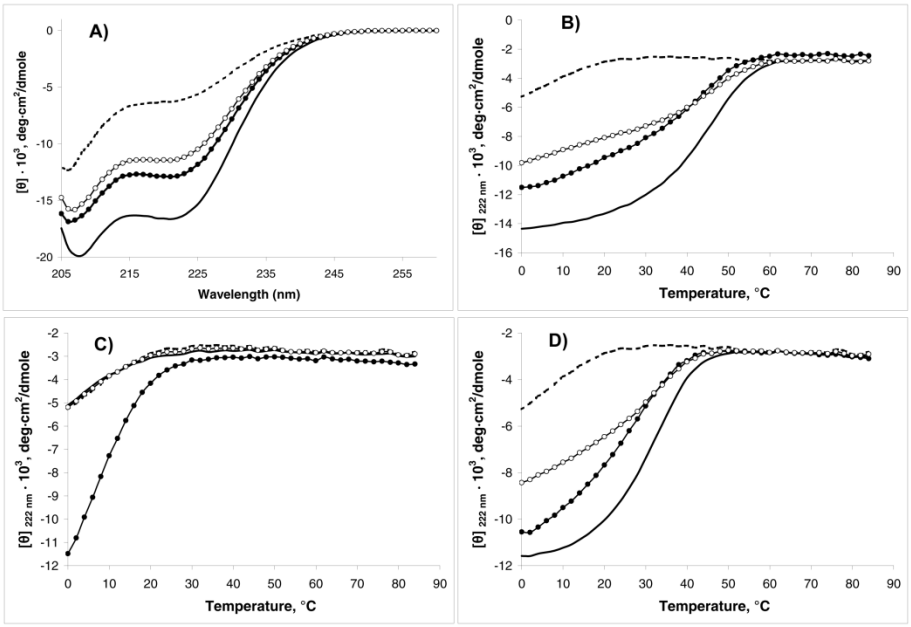



Figure B.S5. Solution characterization of anti-LMAF by CD. Format and presentation is the same as in Figure 3.**2**B-E for anti-SMAF. The target protein is cMaf (in **A** and **B**), the closest off-target competitor is Fra2 (in **C**), and the bZIP related to the target by sequence is MafG (in **D**). $T_m$ values are given in Table B.S2.

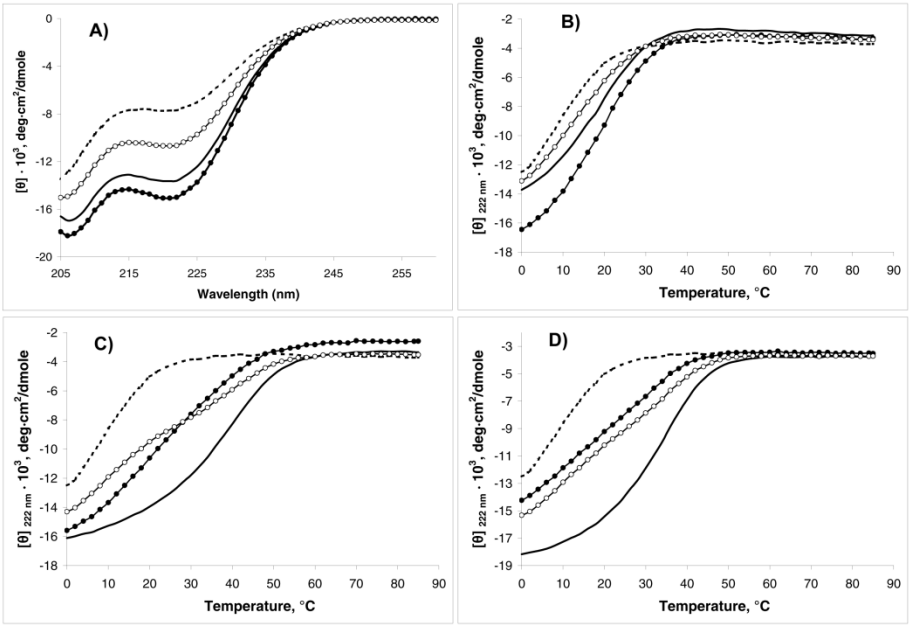



Figure B.S6. Solution characterization of anti-JUN by CD. Format and presentation is the same as in Figure 3.**2**B-E for anti-SMAF. The target protein is cJun (in **A** and **B**), the closest off-target competitor is CHOP (in **C**), and the bZIP related to the target by sequence is ATF-7 (in **D**). $T_m$ values are given in Table B.S2.

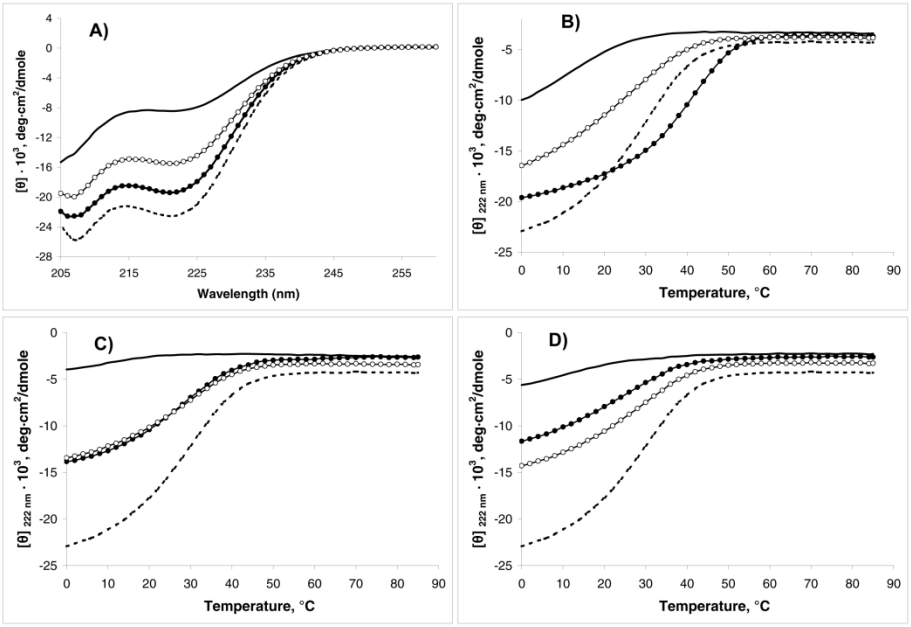



Figure B.S7. Solution characterization of anti-FOS by CD. Format and presentation is the same as in Figure 3.2B-E for anti-SMAF. The target protein is Fos (in A and B), closest off-target competitor is BACH1 (in C), and bZIP related to the target by sequence is ATF-3 (in D). $T_m$ values are given in Table B.S2.

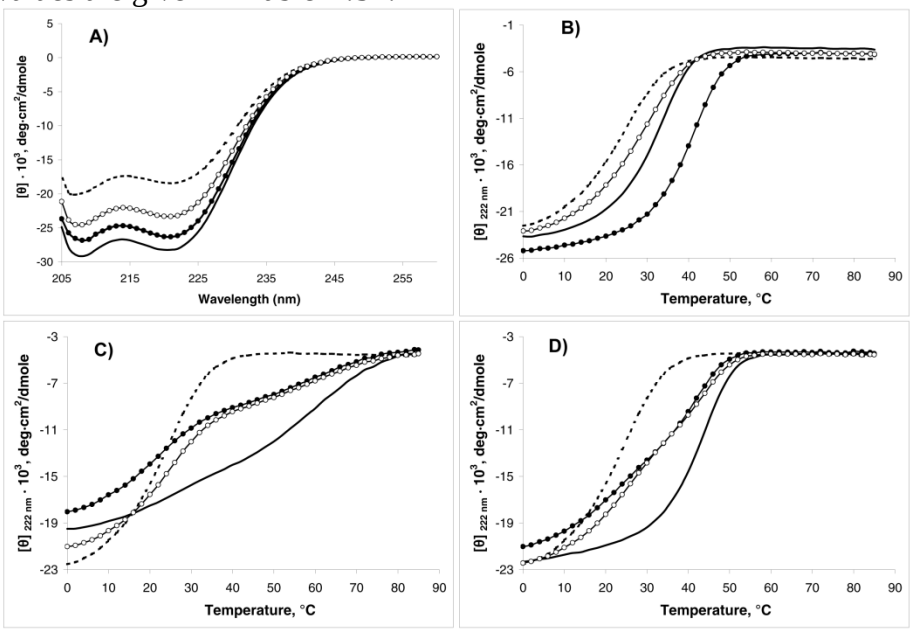



Figure B.S8. Solution characterization of anti-ZF by CD. Format and presentation is the same as in Figure 3.2B-E for anti-SMAF. The target protein is ZF (in A and B), closest off-target competitor is NFE2 (in C), and the bZIP related to the target by sequence is XBP-1 (in D). $T_m$ values are given in TableB.S2.

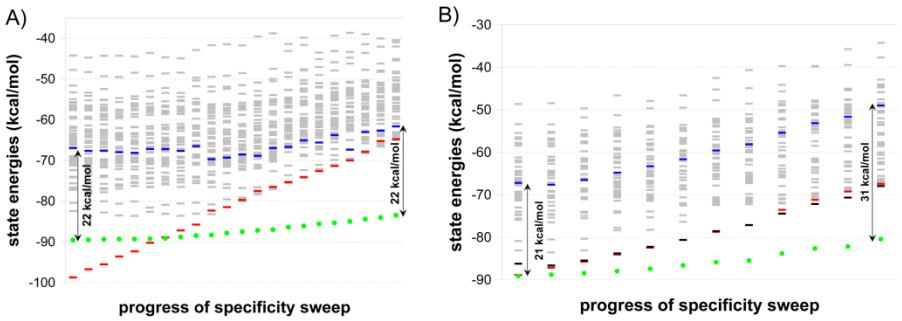



Figure B.S9. Specificity sweep (**A**) and biased specificity sweep (**B**) diagrams for the design of a peptide to bind the leucine-zipper region of ZF. Green dots correspond to the design•target complex and red bars to the design•design complex. Blue bars in **A)** correspond to the energy of the design•XBP-1 complex, which contrary to the prediction of the model showed evidence of strong interaction on the microarray. As a way of addressing this issue, a biased specificity sweep was conducted for ZF, where the gap between the energies of the design•ZF and design•XBP-1 complexes was shifted by 19 kcal/mol. This is shown in **(B)** with blue bars corresponding to the actual model-predicted design•XBP-1 energy, while the black bars are the energies used in the biased specificity sweep. Whereas in the regular specificity sweep there is no competition with the design•XBP-1 state, due to its incorrectly predicted high energy, in the biased specificity sweep this competition is imposed. This procedure generated a successful, highly specific design: anti-ZF.

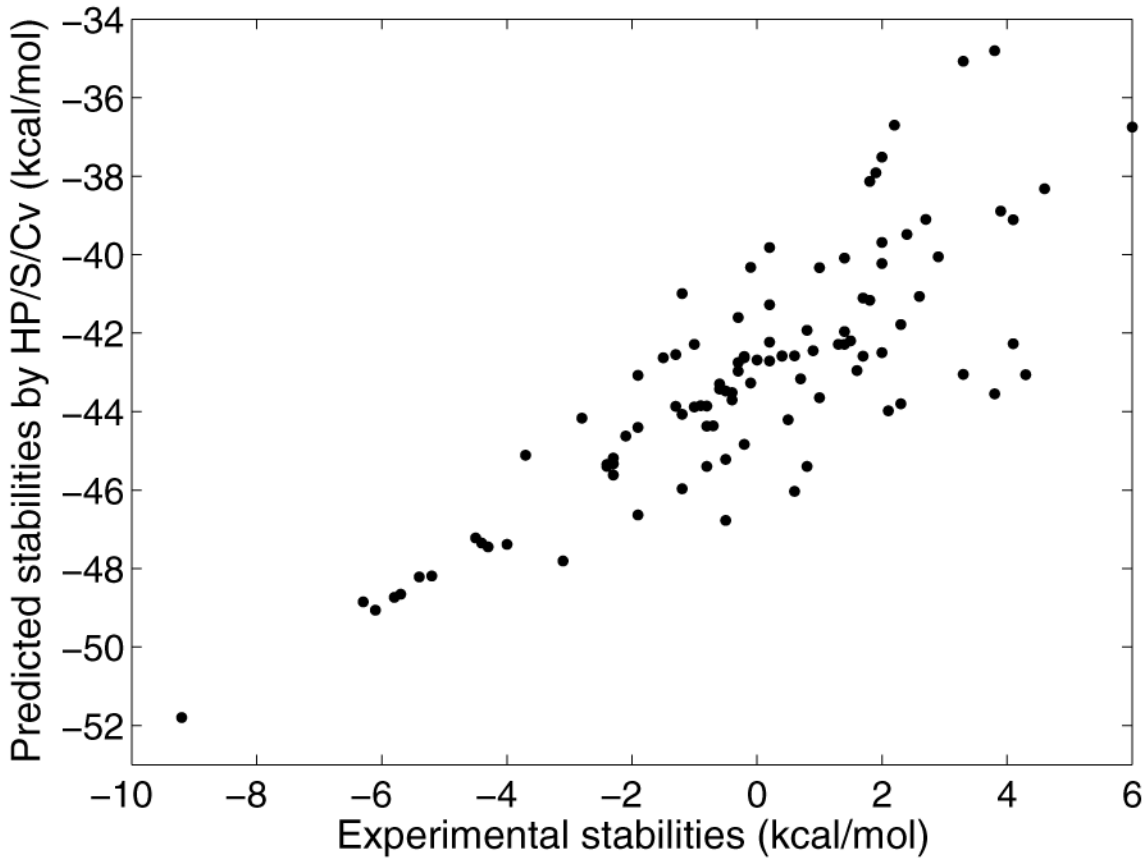



Figure B.S10. Adjusting the 9 **a**-position point ECI in model HP/S/Cv to optimally fit 100 stabilities experimentally measured by Vinson and co-workers[22]. R for the final fit is 0.83.

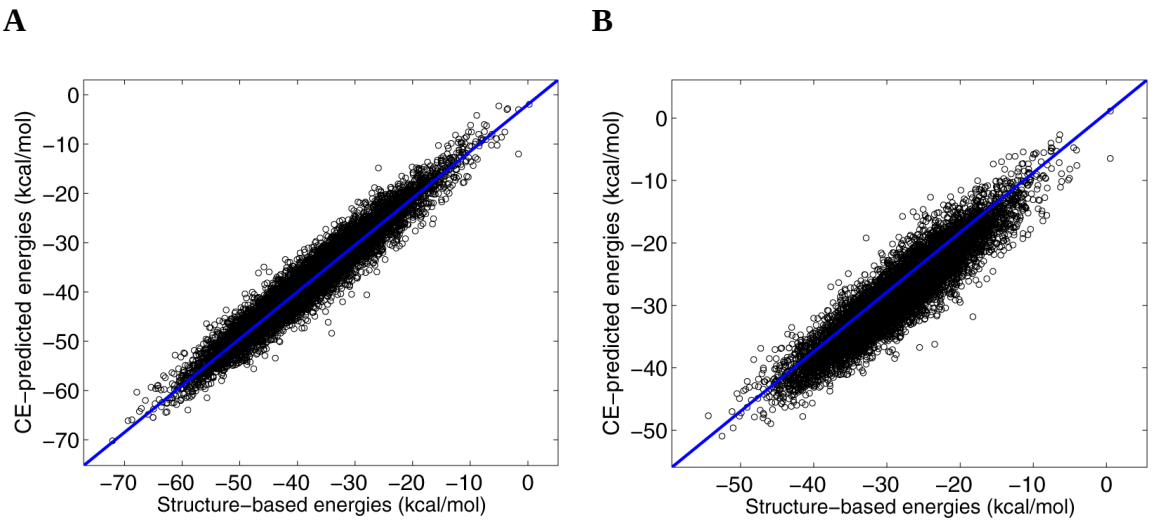
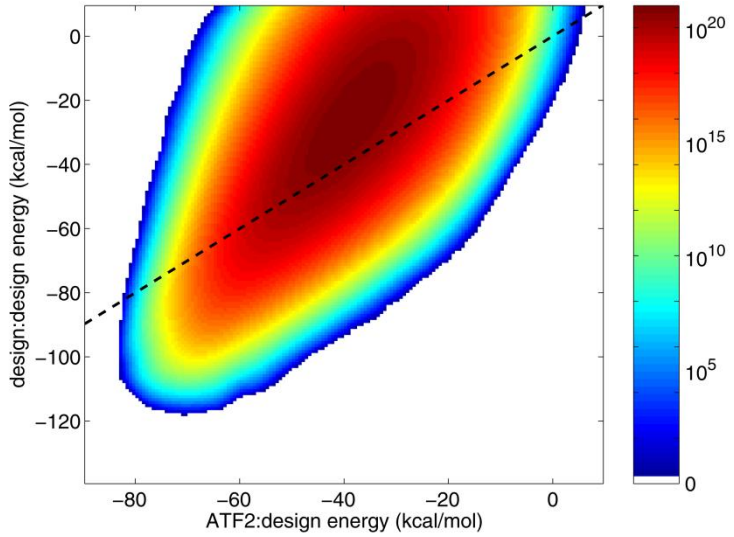



Figure B.S11. The performance of cluster-expanded versions of models HP/S/Ca and HP/S/Cv (panels **A** and **B**, respectively) on a randomly generated set of 10,000 sequences not present in the training set. Root mean square deviations between CE-predicted and structure-based energies are 2.4 and 2.6 kcal/mol for HP/S/Ca and HP/S/Cv, respectively. The cluster expansions contain 2,544 ECI for HP/S/Ca and 2,470 ECI for HP/S/Cv.

**A**




**B**

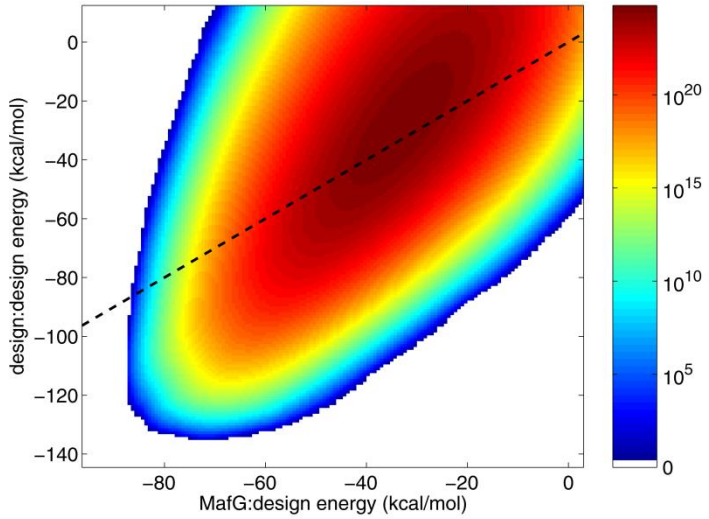



Figure B.S12. 2D energy histograms of two states – the design•target state and the design•design homodimer state. Color represents the total number of possible sequences in each bin (bin sizes are ~1 kcal/mol). The targets are ATF-2 and MafG in **(A)** and **(B),** respectively. The line where design•target and design•design scores are equal is shown. By optimizing only the design•target energy, sequences with high homodimerization propensity will be obtained in these examples. The specificity sweep procedure run with only one disfavoured state (design•design) locates the top boundary of this phase space.


Figure B.S13. Phylogentic tree constructed using the leucine-zipper regions of all human bZIP proteins. Protein names are in black and family names are in blue. Green dots indicate the 33 proteins used in the experiments in this study. The scale refers to amino-acid replacements per site.

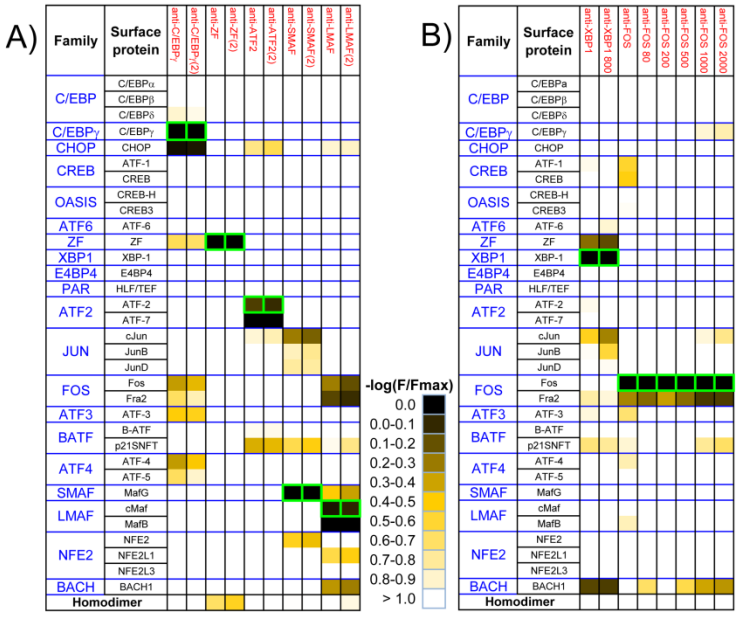



Figure B.S14. Reproducibility of protein-microarray measurements of design interactions probed in duplicate in **(A)** and at different concentrations in **(B)** (probe concentration in nM is shown as part of the probe name in the top row and is 160 nM where not indicated). Data are displayed in the same format as for Figure B.S1..

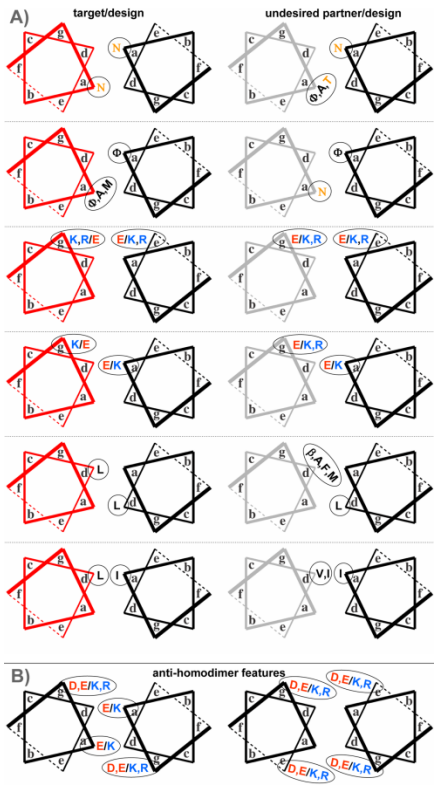



Figure B.S15. Common specificity mechanisms in successful designed peptides. **A)** Specificity features used for discriminating between design•target and design•off-target interactions. The design is in black, the target in red and the undesired partner in gray. Amino acids listed with single-letter codes are the residues comprising the specificity pattern. Slashes delineate subgroups of residues, with corresponding subgroups delineated similarly at the interacting position. Φ designates hydrophobic residues Ile, Val or Leu and β stands for beta-branched residues Ile or Val. In the last row, the **a-d'** interaction is between an **a** residue and the more C-terminal **d**' residue on the opposite helix. **B)** Specificity features commonly used in designed peptides to disfavor the design•design homodimer, using the same notation.

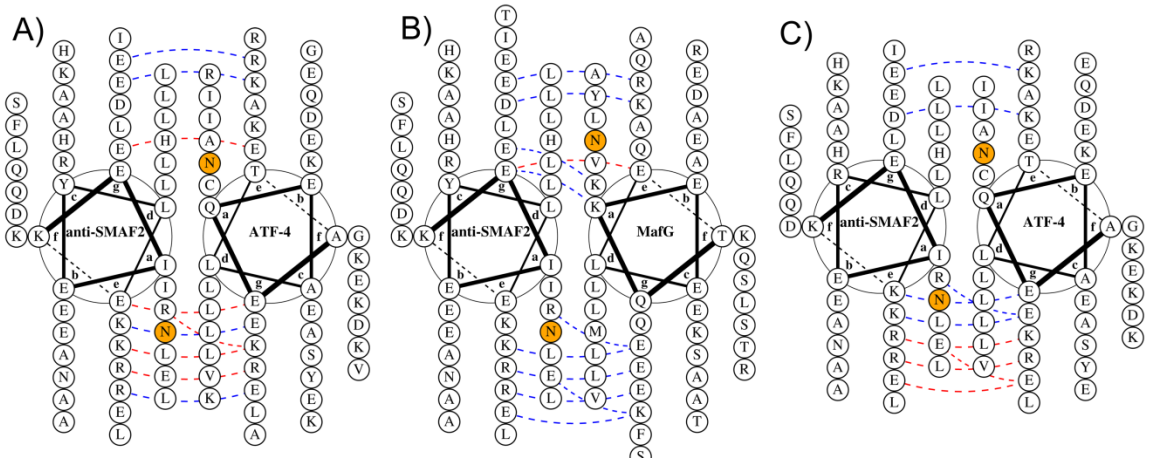



Figure B.S16. Helical-wheel diagrams for anti-SMAF-2 complexes with ATF-4 and MafG. **(A)** The anti-SMAF-2•ATF-4 complex is predicted to be much weaker than the anti-SMAF-2•MafG complex shown in **(B)**, in large part due to the misaligned asparagines at **a** positions in anti-SMAF-2•ATF-4. **(C)** A different alignment of anti-SMAF-2•ATF-4, where the asparagines match up, may be more favorable, although it is not predicted to be much stronger computationally. Diagrams made with DrawCoil 1.0 (http://www.gevorggrigoryan.com/drawcoil/).

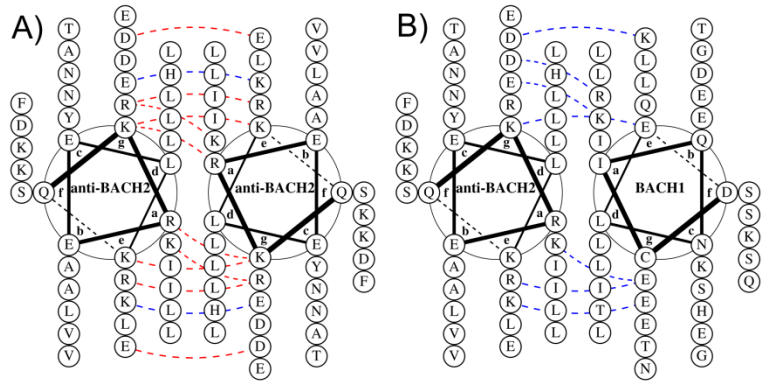



Figure B.S17. Helical-wheel diagrams of the anti-BACH-2 homodimer complex, shown in **(A)**, and the anti-BACH-2•BACH1 complex shown in **(B)**. The strong anti-homodimerization features of anti-BACH-2 are concentrated at the N-terminus of the sequence, leaving open the possibility that this portion simply does not fold, while the remainder of the coiled coil forms a stable complex. Diagrams made with DrawCoil 1.0 (http://www.gevorggrigoryan.com/drawcoil/).

Table B.S1. All designed sequences tested. For each design, listed in columns are: the name of the design, the name of the bZIP target for that design, the family of the target bZIP, the round of design/testing in which this sequence was produced, the count of attempts to design a partner for the given target, the energy function used and the designed sequence. Note that designs are named after the family of the target rather than the individual protein. There were three rounds of experiments. Attempts are different than rounds because not all targets were attempted in the first (or second) rounds. An attempt involved testing one or two designs (in one case, three) for each target considered in a set of experiments. When the first experimental attempt to identify a specific design was unsuccessful, alternative solutions from the specificity sweep were selected for testing in subsequent rounds (constituting further attempts). In a few cases, listed in the footnotes, these additional designs were created with a modified procedure aimed at addressing experimentally identified shortcomings of previous designs.

| **Design name** | **Target** | **Family** | **Round** | **Attempt** | **Method** | **Design sequence** |
|---|---|---|---|---|---|---|
| | | | | | | fgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabc |
| anti-C/EBP-2 | C/EBPα | C/EBP | 2 | 1 | HP/S/Ca | FENVTHEFILATLENENAKLRRLEAKLERELARLRNEVAWL |
| anti-C/EBP | C/EBPα | C/EBP | 3 | 2 | HP/S/Cv | AENQYVEDLIQYLEKENARLKKEVQRLVRELSYFRRRIAELA |
| anti-C/EBP-3 | C/EBPα | C/EBP | 3 | 2 | HP/S/Cv | AENQSVEDIIAKKEDENAHLKNEVKTLINELETLRKKIEYLA |
| anti-C/EBPγ | C/EBPγ | C/EBPγ | 2 | 1 | HP/S/Ca | NDLDAYEREAEKLEKKNEVLRNRLAALENELATLRQEVASMKQELQS |
| anti-C/EBPγ-2 | C/EBPγ | C/EBPγ | 2 | 1 | HP/S/Cv | RDLQNVEREIQSLEKKNESLKKKIASLENELATLKQEIAYFKRELAY |
| anti-CHOP | CHOP | CHOP | 3 | 1 | HP/S/Cv | DRLAVKENRVAVLKNENAKLRNIIANLKDRIAYFRRELAYLELEEEQLA |
| anti-CREB | CREB | CREB | 2 | 1 | HP/S/Cv | QLVAQLRSKVEQLVNRNQALKNKLEYLRQEIAETEQ |
| anti-CREB-2 | CREB | CREB | 3 | 2 | HP/S/Cv[1] | NKVEQLKNKVEQLKNRNAALKNDLARLEREIAYAEE |
| anti-CREB-3 | CREB | CREB | 3 | 2 | HP/S/Cv[1] | QKVESLKQKIEELKQRKAQLKNDIANLEKEIAYAET |
| anti-OASIS | CREB3 | OASIS | 2 | 1 | HP/S/Cv | QKVEQLKNKVEQKLKENESLENKVAELKNRNEYLKNKIENLINDITNLENDVAR |
| anti-OASIS-2 | CREB3 | OASIS | 2 | 1 | HP/S/Cv | QKVAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE |
| anti-OASIS-3 | CREB3 | OASIS | 3 | 2 | HP/S/Cv[5] | QKVAQLKNRVAYKLKENAKLENIVARLENDNANLEKDIANLEKDIANLERDVAR |
| anti-OASIS-4 | CREB3 | OASIS | 3 | 2 | HP/S/Cv[5] | QKVAQLKNIIAKKEDENAVLENLVAVLENENAYLEKELARLERDIARAERDVKV |

| | | | | | | |
|---|---|---|---|---|---|---|
| anti-ATF6 | ATF-6 | ATF6 | 3 | 1 | HP/S/Cv | EKIQELKRRLAYFRRENATLKNDNATLEN ELASVEAENEALRK |
| anti-ZF-2 | ZF | ZF | 2 | 1 | HP/S/Cv | QKIAYLRDRIAALKAENEALRAKNEALRS KIEELKKEKEELRDKIAQKKDR |
| anti-ZF | ZF | ZF | 3 | 2 | HP/S/Cv[6] | NLVAQLENEVASLENENETLKKKNLHKK DLIAYLEKEIANLRKKIEE |
| anti-XBP1-2 | XBP-1 | XBP1 | 2 | 1 | HP/S/Ca | SKYDALRNKLEALKNRNAQLRKENEQLR LEEAVLEVRNEVL |
| anti-XBP1 | XBP-1 | XBP1 | 2 | 1 | HP/S/Cv | QKIEYLKDKIAELKDRNAVKRSENAQLRQ AVATLEQKNEEL |
| anti-E4BP4-2 | E4BP4 | E4BP 4 | 2 | 1 | HP/S/Ca | QKRQELKQRLAVLENDNARLKNDLAQLE VEEAYIE |
| anti-E4BP4 | E4BP4 | E4BP 4 | 2 | 1 | HP/S/Cv | NKNNVKKNRLAVLENENATLRNELAWLR LELAAME |
| anti-E4BP4-3 | E4BP4 | E4BP 4 | 3 | 2 | HP/S/Cv[3] | EKNQELKNRLAVLENDNAALRNDLARLE REIAYME |
| anti-ATF2-2 | ATF-2 | ATF2 | 1 | 1 | HP/S/Ca | QKLQTLRDLLAVLENRNQELKQLRQHLK DLLKYLEDELATLEKE |
| anti-ATF2-3 | ATF-2 | ATF2 | 2 | 2 | HP/S/Cv | STVEELLRAIQELEKRNAELKNRKEELKN LVAHLRQELAAHKYE |
| anti-ATF2 | ATF-2 | ATF2 | 3 | 3 | HP/S/Cv | NTVKELKNYIQELEERNAELKNLKEHLKF AKAELEFELAAHKFE |
| anti-ATF2-4 | ATF-7 | ATF2 | 3 | 3 | HP/S/Cv | QKVEELKNKIAELENRNAVKKNRVAHLK QEIAYLKDELAAHEFE |
| anti-JUN | cJun | JUN | 1 | 1 | HP/S/Ca | SIAATLENDLARLENENARLEKDIANLERD LAKLEREEAYF |
| anti-FOS | Fos | FOS | 1 | 1 | HP/S/Ca | NEKEELKSKKAELRNRIEQLKQKREQLKQ KIANLRKEIEAYK |
| anti-ATF3 | ATF-3 | ATF3 | 1 | 1 | HP/S/Ca | ELTDELKNKKEALRKDNAALLNELASLEN EIANLEKEIAYFK |
| anti-ATF3-2 | ATF-3 | ATF3 | 1 | 1 | HP/S/Ca | NETEQLINKKEQLKNDNAALEKDAASLEK EIANLEKEIAYFK |
| anti-ATF3-3 | ATF-3 | ATF3 | 3 | 3 | HP/S/Cv[7] | NILASLENKKEELKKLNAHLLKEIENLEKE IANLEKEIAYFK |
| anti-ATF4 | ATF-4 | ATF4 | 2 | 1 | HP/S/Cv | KRIAYLRKKIAALKKDNANLEKDIANLEN EIERLIKEIKTLENEVASHEQ |
| anti-ATF4-2 | ATF-4 | ATF4 | 2 | 1 | HP/S/Cv | ARNAYLRKKIARLKKDNLQLERDEQNLE KIIANLRDEIARLENEVASHEQ |
| anti-BATF | p21SNF T | BATF | 2 | 1 | HP/S/Ca | NELESLENKKEELKNRNEELKQKREQLKQ KLAALRNKLDAYKNRL |
| anti-BATF-2 | p21SNF T | BATF | 3 | 2 | HP/S/Cv | NDIENLKDKIEELKQRKEELKQKIEYLKQK IEALRQKLAALKQRIA |
| anti-BATF-3 | p21SNF T | BATF | 3 | 2 | HP/S/Cv | EKIEELKDKIAELRSRNAALRNKIEALKQK LEALRQKIEYLKDRIA |
| anti-PAR | HLF | PAR | 3 | 1 | HP/S/Cv | NRLQELENKNEVLEKRKAELRNEVATLEQ ELAAHRYELAAIEKEIA |

| | | | | | | |
|---|---|---|---|---|---|---|
| anti-SMAF-2 | MafG | SMAF | 1 | 1 | HP/S/Ca | KEIEYLEKEIERLKDLREHLKQDNAAHRQELNALRLEEAKLEFILAHLLST |
| anti-SMAF-3 | MafG | SMAF | 1 | 1 | HP/S/Ca | KEIERLEKEIKTLINLLTTLRQDNAAHRKEAAALEKEEANLERDIQNLLRY |
| anti-SMAF | MafG | SMAF | 2 | 2 | HP/S/Cv | KEIANLEKEIASLEKKVAVLKQRNAAHKQEVAALRKEIAYVEDEIQYVEDE |
| anti-LMAF-2 | cMaf | LMAF | 3 | 1 | HP/S/Cv | NKNETLKNINARLRNDVARLKNRIARLKDDIENVEDEIQYLE |
| anti-LMAF-3 | cMaf | LMAF | 3 | 1 | HP/S/Cv | LENAQIKKEIAQLRKEVAQLKQKIEELKNDNARVEREIQYLE |
| anti-LMAF | cMaf | LMAF | 3 | 1 | HP/S/Ca | KDIANLKKEIAHLKNDLQRLESIRERLKFDILNHEQEEYALE |
| anti-NFE2 | NFE2 | NFE2 | 1 | 1 | HP/S/Ca | QKRQQLKQKLAALRRDIENLQDEIAYKEDEIANLKDKIEQLLS |
| anti-NFE2-2 | NFE2 | NFE2 | 3 | 2 | HP/S/Cv | QKIESLKDKLANKRDKIALLRSEVASFEKEIAYLEKEIANLEN |
| anti-NFE2-3 | NFE2 | NFE2 | 3 | 2 | HP/S/Cv[4] | EKIEYLKDKLAHKRNEVAQLRKEVTHKVDELTSLENEVAQLLK |
| anti-BACH-2 | BACH1 | BACH | 2 | 1 | HP/S/Ca | QKREELKSRKAYLRKEIANLKKDILNLLDDLVAHEFELVTL |
| anti-BACH | BACH1 | BACH | 2 | 1 | HP/S/Cv | QKIQYLKQRIAELRKKIANLRKDIANLEDDAAVKEDELVHL |
| anti-BACH-3 | BACH1 | BACH | 3 | 2 | HP/S/Cv[2] | EKIEYLKDRIAELRSKIAALRNDLTHLKNDKAHKENELAHLA |

[1] The only strong off-target interaction for design anti-CREB, produced in round 2, was the design•design homodimer. However, the specificity sweep produced no solutions that were significantly more specific against the homodimer. Thus, in the next round we sought to remove design homodimerization by considering only the homodimer as a competitor. In the resulting designs anti-CREB-2 and anti-CREB-3, homodimerization was indeed no longer a problem, but global specificity was reduced. This indicates that maintaining gaps to many states simultaneously can be important.

[2] The two strong off-target competitors for anti-BACH in round 2 were Fos and NFE2. The latter was deemed too close in sequence to effectively discriminate with our models. To improve specificity against Fos, a biased specificity sweep was used with a gap offset of -10 kcal/mol for Fos (making gaps with Fos more negative than they would be, which caused competition with Fos to be more stringent). However, anti-BACH-3 still interacted with Fos more strongly than with BACH-1.

[3] The initial two designs against E4BP4 were not very stable, and this was not predicted by the models. HP/S/Cv predicted that the most stable design against E4BP4 had a Lys at the N-terminal **d** position. To address this, we temporarily adjusted the ECI for Leu-Leu at **d-d'** in HP/S/Cv to be more favorable by 2 kcal/mol and reran the specificity sweep procedure. Anti-

E4BP4-3 was picked from this list. Although this resulted in a more hydrophobic core, there was no detectable increase in stability according to the microarray assay.
[4] The only strong off-target competitor for anti-NFE2 was ATF-4, so in this design we used a biased specificity sweep approach with a gap offset of -3 kcal/mol for ATF-4 (making gaps with ATF-4 more negative). However this design interacted with Fos, which had not previously been a strong competitor.
[5] To eliminate the only significant competitor of anti-OASIS, p21SNFT, a biased specificity sweep was run with a gap offset of -10 kcal/mol for p21SNFT. This did indeed eliminate p21SNFT as a competitor, but MafG emerged as a new strong competitor.
[6] Because the only significant competitor for the first design, anti-ZF-2, was XBP-1, we applied a biased specificity sweep approach with a gap offset of -10 kcal/mol for XBP-1. This successfully removed XBP-1 as a competitor and resulted in a very specific and stable design.
[7] Significant competitors for designs against ATF-3 were Fos and ATF-4, whereas the models considered JUN and ATF2 families more likely to interact. To bias the specificity sweep against the relevant competitors, gap offsets of +8 and +2 kcal/mol for JUN and ATF2 families respectively were imposed (making gaps with JUN and ATF2 family members less important in the optimization).

Table B.S2. Melting temperature ($T_m$) values estimated by fitting to CD-monitored melting curves. Corresponding 95% confidence intervals are given in brackets (see section Circular dichroism). Some measurements were made in duplicate to evaluate reproducibility; duplicate measurements are marked with a number two in parentheses.

| **bZIP•bZIP homodimers** | **Tm (° C)** | **95% CI** | | **design•design homodimers** | **Tm (° C)** | **95% CI** |
|---|---|---|---|---|---|---|
| CHOP | 36.4 | [35.8 36.9] | | anti-SMAF | 11.6 | [11.1 12.1] |
| BACH1 | 8.4 | [6.9 9.9] | | anti-ATF2 | 5.2 | [1.7 8.7] |
| XBP-1 | 42 | [41.7 42.3] | | anti-ATF4 | 48.6 | [48 49.3] |
| NFE2 | multiple transitions | | | anti-LMAF | 3 | [-3.4 9.3] |
| ZF | 31.6 | [31.3 31.8] | | anti-ZF | 22.1 | [21.7 22.4] |
| MafB | 19.8 | [19.1 20.6] | | anti-JUN | 7.3 | [6.6 8.1] |
| cMaf | 43.1 | [42.5 43.8] | | anti-FOS | 27.2 | [26.8 27.6] |
| Fra2 | <0 | [-13.4 5.7] | | | | |
| p21SNFT | 33 | [32.6 33.4] | | **design•bZIP heterodimers** | **Tm (° C)** | |
| ATF-4 | 7.9 | [6.1 9.7] | | anti-ATF4:ATF-4 | 52.1 | [51.4 52.8] |
| ATF-3 | 9.4 | [6.4 12.3] | | anti-ATF2:ATF-7 | 41 | [40.4 41.6] |
| ATF-3(2) | 6.6 | [4.3 9] | | anti-SMAF:MafG | 37.9 | [37 38.7] |
| Fos | 10.6 | [8.9 12.4] | | anti-JUN:cJun | 24.2 | [23.4 24.9] |
| Fos(2) | 9 | [8.1 9.9] | | anti-FOS:FOS | 43.6 | [42.7 44.4] |
| cJun | 16.6 | [16.0 17.3] | | anti-ZF:ZF | 43 | [42.7 43.4] |
| cJun(2) | 16.2 | [15.7 16.8] | | | | |
| ATF-7 | 31.4 | [31 31.8] | | | | |
| ATF-7(2) | 31.7 | [31.3 32.1] | | | | |
| MafG | 30.2 | [29.7 30.8] | | | | |
| MafG(2) | 31.8 | [31.5 32.2] | | | | |

Table B.S3. Average background-corrected fluorescence values and $S_{array}$ values from round 1 of array measurements. Peptides on the surface are in rows, those in solution in columns. Duplicate measurements are marked with a number two in parentheses. The anti-FOS peptide was tested at concentrations ranging from 80 nM to 2000 nM, as indicated in the probe names.

| protein | ATF-2 | cJun | Fos | Fra2 | ATF-3 | ATF-4 | p21SNFT | MafG |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -1209.5 | -220.3 | 290.7 | -789.9 | 4573.1 | 16459.4 | 2169.7 | -341.4 |
| C/EBPβ | -2318.1 | -543.4 | -1387.9 | -2406.9 | 7687.4 | 5320.2 | 2156.1 | 338.3 |
| C/EBPδ | -3235.8 | -540.0 | -1080.3 | -978.3 | 7598.0 | 15172.1 | 2625.6 | -244.6 |
| C/EBPγ | 5155.8 | 345.5 | 733.2 | 261.6 | 21941.3 | 34208.8 | 5172.2 | -7.9 |
| CHOP | 5110.1 | 996.5 | 5691.3 | 3996.1 | 24897.6 | 5879.3 | 18419.3 | 330.8 |
| ATF-1 | -1961.5 | -947.1 | -879.7 | -860.9 | -2213.3 | -2296.8 | -2072.6 | -318.6 |
| CREB | -6256.7 | -1354.7 | -1370.9 | -2093.1 | -4312.5 | -3194.2 | -3906.8 | -670.6 |
| CREB-H | -933.4 | -830.6 | -6.2 | -351.2 | -234.6 | -91.3 | -375.9 | -284.4 |
| CREB3 | -820.6 | -279.0 | -1116.9 | -1014.2 | 1396.7 | -1809.6 | 438.1 | -75.1 |
| ATF-6 | -3780.2 | -1124.3 | -980.4 | -1252.0 | -1825.8 | -1502.1 | -2020.6 | -867.1 |
| ZF | -789.8 | 296.4 | -359.9 | -385.4 | 3404.8 | 5528.6 | 1020.3 | 862.2 |
| XBP-1 | -2048.6 | -359.3 | -726.9 | -2537.3 | 445.3 | -815.4 | -1979.3 | -294.4 |
| E4BP4 | -1942.6 | -513.1 | -932.1 | -546.2 | 1080.1 | -1776.0 | 305.7 | 116.0 |
| ATF-2 | 6994.6 | 4898.9 | 5723.6 | 4635.6 | 20294.5 | 1121.1 | 5202.8 | -266.1 |
| ATF-7 | 9594.1 | 8099.6 | 6785.2 | 7509.6 | 22271.2 | 4419.6 | 7512.3 | -452.8 |
| cJun | 14760.4 | 1997.3 | 27052.4 | 24950.8 | 24562.3 | -319.9 | 11768.7 | -334.2 |
| JunB | 4758.5 | 449.4 | 16150.8 | 16856.8 | 18105.6 | -758.3 | 9728.8 | 17.1 |
| JunD | 9823.1 | 788.6 | 22888.9 | 22692.3 | 22719.3 | -1331.8 | 11304.0 | 98.3 |
| Fos | 13984.2 | 35014.5 | 2120.9 | 1451.1 | 6142.6 | 7326.7 | 703.8 | -854.8 |
| Fra2 | 9788.4 | 19126.3 | 3892.8 | 608.9 | 9627.2 | 1022.6 | 266.4 | -634.1 |
| ATF-3 | 19928.4 | 12049.6 | 2674.5 | 6099.9 | 194.9 | 11042.8 | 13952.3 | 180.7 |
| ATF-4 | 1728.8 | -1060.4 | 8750.9 | -223.4 | 21845.3 | -1508.4 | 14458.4 | -338.1 |
| ATF-5 | -4831.8 | -1941.1 | -2288.1 | -1926.1 | -2945.9 | -3321.8 | 3954.8 | -945.2 |
| B-ATF | 2149.7 | 13378.4 | 463.0 | -713.3 | 8148.4 | 4415.8 | 2425.2 | -675.4 |
| p21SNFT | 9068.7 | 12668.4 | 298.1 | -20.7 | 23988.6 | 8109.1 | 5699.1 | 1393.3 |
| HLF | -2699.3 | -188.0 | -1099.3 | -828.5 | -627.9 | -3384.4 | 121.6 | -474.0 |
| MafG | -253.8 | 268.3 | 767.9 | -532.1 | 3179.9 | -193.2 | 2433.0 | 1387.9 |
| cMaf | -253.3 | 92.8 | 211.8 | 127.9 | 2222.3 | 1815.9 | -1080.9 | -39.1 |
| MafB | -733.8 | -217.7 | 2278.7 | 1685.4 | 4347.4 | -1942.6 | -935.9 | -471.9 |
| NFE2 | -2474.9 | -544.6 | -443.6 | -1246.4 | -890.6 | -114.8 | -1826.8 | 345.8 |
| NFE2L1 | 98.8 | -284.6 | 771.0 | -13.0 | -1682.8 | 3979.2 | -769.6 | 38491.7 |
| NFE2L3 | -4486.5 | -1628.7 | -3240.7 | -2830.6 | -4049.0 | -415.3 | -3341.4 | 31185.9 |
| BACH1 | 6185.9 | 456.6 | 1932.9 | 2078.1 | -50.2 | 1253.1 | 895.5 | 18423.5 |
| anti-ATF2-2 | 735.1 | -410.0 | -874.0 | -1496.9 | -753.9 | -730.5 | 3897.1 | -170.9 |
| anti-ATF3 | 7223.3 | 1717.5 | 19632.5 | 15128.1 | 30646.4 | 16601.9 | 9409.2 | 3310.1 |
| anti-ATF3-2 | 579.0 | 562.3 | 13724.4 | 7512.8 | 8260.4 | 838.6 | 936.2 | 532.1 |
| anti-JUN | 2505.5 | 6966.4 | 2211.1 | 511.4 | 6446.6 | 7163.3 | 2591.8 | 925.5 |
| anti-FOS | -177.0 | 2685.3 | 39044.8 | 22696.1 | 4548.8 | -975.3 | 4052.0 | 1182.6 |
| anti-SMAF-2 | -2069.7 | 1877.4 | 1930.0 | -1085.1 | -447.6 | 21907.5 | -1261.8 | 7926.6 |
| anti-SMAF-3 | -811.6 | -280.1 | -80.3 | -743.6 | -1135.0 | -915.4 | -801.5 | -37.0 |
| anti-NFE2 | 67.3 | -304.7 | 6306.8 | 3092.9 | 479.4 | 10427.1 | 44.5 | 2958.3 |

| protein | NFE2 | NFE2L1 | BACH1 | anti-ATF2-2 | anti-ATF3 | anti-ATF3-2 | anti-JUN | anti-FOS |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -1513.3 | 91.9 | -206.9 | -818.8 | 1158.0 | 277.4 | -406.1 | -599.1 |
| C/EBPβ | -4574.9 | -3091.3 | -720.5 | -2531.9 | -310.7 | 367.1 | -282.7 | 357.2 |
| C/EBPδ | -1143.5 | -574.9 | -1533.1 | -565.9 | 955.0 | 350.7 | -118.3 | -616.6 |
| C/EBPγ | -861.3 | -716.8 | 82.1 | 8474.6 | 2587.9 | 390.4 | 615.8 | 2685.3 |
| CHOP | -979.9 | -830.3 | 3777.6 | 14294.7 | 3111.1 | 437.6 | 2289.4 | 1452.8 |
| ATF-1 | -3053.8 | -2446.8 | 1971.0 | -2297.3 | -687.1 | 410.1 | 1454.3 | -923.4 |
| CREB | -5052.8 | -2797.4 | 3077.3 | -2383.3 | -289.8 | 419.9 | 755.7 | -991.0 |
| CREB-H | -708.1 | -596.1 | -209.9 | 3.3 | -890.1 | 220.3 | 4.9 | 1466.1 |
| CREB3 | -816.0 | -1099.2 | -598.8 | 915.4 | -1117.3 | 356.9 | 443.9 | 1224.9 |
| ATF-6 | -2016.0 | -2267.0 | -254.0 | -991.8 | -1229.4 | 302.4 | -104.6 | -896.9 |
| ZF | 2730.1 | 21547.8 | 2352.9 | 4059.1 | 973.3 | 336.9 | -48.1 | -372.2 |
| XBP-1 | -1705.0 | -1383.3 | 696.1 | -1392.5 | -3462.4 | 334.6 | -686.9 | -903.9 |
| E4BP4 | -816.8 | -1684.9 | 101.3 | 1511.1 | -1434.5 | 306.0 | 531.5 | -113.3 |
| ATF-2 | -2235.5 | 3164.0 | 4399.2 | 6767.6 | 486.4 | 322.3 | -303.1 | -86.7 |
| ATF-7 | 724.4 | 7211.7 | 6091.9 | 13903.3 | 828.0 | 298.9 | 222.5 | 462.4 |
| cJun | -2589.8 | -1940.9 | 42.4 | 532.4 | 805.9 | 284.5 | 4862.2 | 3658.3 |
| JunB | -218.6 | -1815.1 | -299.3 | -457.1 | -226.1 | 332.1 | 1163.0 | 1083.6 |
| JunD | -712.3 | -1082.1 | 203.1 | -587.7 | 258.5 | 272.1 | 2164.3 | 1754.9 |
| Fos | -1298.1 | 5426.5 | 2637.5 | 200.9 | 24395.9 | 4055.4 | -196.8 | 35147.9 |
| Fra2 | -847.4 | 1584.0 | 2029.9 | -857.9 | 6044.1 | 1450.1 | -106.8 | 17465.9 |
| ATF-3 | -4021.9 | -2556.2 | 647.3 | -119.9 | 29694.3 | 968.5 | 1028.1 | 1567.6 |
| ATF-4 | -1788.3 | 27265.1 | 2516.6 | 8516.9 | 26002.3 | 594.9 | 1519.8 | -546.8 |
| ATF-5 | -704.1 | -996.6 | 1436.5 | 4261.2 | -128.5 | 374.2 | 125.9 | -1422.4 |
| B-ATF | -51.9 | 3341.1 | 4337.6 | 3925.6 | 1459.4 | 321.5 | -214.2 | 818.5 |
| p21SNFT | -1526.1 | -535.8 | 3288.4 | 11315.8 | 6627.7 | 489.4 | 249.2 | 2909.8 |
| HLF | -4778.1 | -1520.3 | 21.0 | -1036.3 | -370.9 | 254.1 | 22.9 | -633.5 |
| MafG | 11187.3 | 49004.5 | 25075.1 | 908.3 | 1654.0 | 567.8 | 594.9 | 1147.2 |
| cMaf | -899.0 | 3789.4 | 5396.6 | -1087.3 | -167.1 | 392.2 | 35.4 | 847.5 |
| MafB | -572.1 | 4157.4 | 11655.9 | -968.2 | -432.9 | 367.5 | 65.8 | 1440.7 |
| NFE2 | 15668.0 | -1367.0 | 759.1 | -1075.2 | 433.8 | 316.7 | -176.9 | 160.5 |
| NFE2L1 | -329.1 | 4830.4 | -125.9 | 65.6 | 3530.9 | 388.6 | 268.8 | -74.9 |
| NFE2L3 | 13034.3 | -3370.8 | -1015.4 | -1227.9 | -1316.4 | 189.1 | -1338.3 | -1048.0 |
| BACH1 | -618.1 | 331.9 | 2580.4 | -788.2 | 2616.3 | 373.6 | 705.9 | 6629.7 |
| anti-ATF2-2 | -2763.5 | 13087.5 | -1592.5 | 6187.1 | | | | |
| anti-ATF3 | 15014.4 | 44239.6 | 11233.2 | | -905.8 | | | |
| anti-ATF3-2 | 306.4 | 8928.4 | 294.6 | | | 259.1 | | |
| anti-JUN | 673.9 | 15897.4 | 6819.8 | | | | 804.3 | |
| anti-FOS | -682.9 | -6.0 | 17630.9 | | | | | 1828.7 |
| anti-SMAF-2 | 25607.3 | 42286.6 | 5227.0 | | | | | |
| anti-SMAF-3 | 1143.1 | 11088.9 | -95.0 | | | | | |
| anti-NFE2 | 32984.4 | 4472.4 | 7008.3 | | | | | |

| protein | anti-SMAF-2 | anti-SMAF-3 | anti-NFE2 | anti-FOS80 | anti-FOS200 | anti-FOS500 | anti-FOS1000 | anti-FOS2000 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -293.8 | 79.5 | 718.9 | 204.2 | 269.7 | 542.3 | 1006.1 | 765.1 |
| C/EBPβ | -400.9 | -1884.7 | -2306.9 | 326.9 | 292.4 | 394.1 | 1884.9 | 2025.0 |
| C/EBPδ | -131.1 | -273.5 | 153.0 | 215.8 | 119.1 | -173.7 | -262.5 | 46.9 |
| C/EBPγ | -112.2 | 1141.6 | 491.6 | 277.3 | 409.1 | 2203.7 | 6819.2 | 5537.9 |
| CHOP | -155.3 | -139.3 | 380.4 | 284.7 | 316.7 | 1676.3 | 5432.5 | 3920.5 |
| ATF-1 | 59.8 | -1207.3 | 832.1 | 407.0 | 388.4 | 263.1 | 912.7 | 597.1 |
| CREB | -315.5 | -2505.6 | -2082.8 | 420.4 | 373.6 | -534.3 | -2069.0 | -1908.0 |
| CREB-H | -112.5 | 217.1 | 50.6 | 197.1 | 217.1 | 1358.8 | 4158.9 | 3986.3 |
| CREB3 | 37.9 | 212.3 | 39.1 | 346.1 | 329.9 | 927.3 | 4478.6 | 3427.3 |
| ATF-6 | -368.9 | -995.1 | -1413.6 | 289.1 | 199.1 | -574.0 | -907.5 | -256.9 |
| ZF | 493.9 | 3841.1 | -862.1 | 292.4 | 245.5 | -124.1 | -415.3 | -415.5 |
| XBP-1 | -885.4 | -970.1 | -1292.3 | 300.0 | 296.3 | 464.1 | 1519.9 | 2019.6 |
| E4BP4 | -153.9 | -485.7 | -475.1 | 247.8 | 204.6 | 301.8 | 1727.7 | 1680.9 |
| ATF-2 | -332.3 | 433.0 | 5345.1 | 302.8 | 260.1 | 225.8 | 1281.9 | 1360.9 |
| ATF-7 | 58.1 | 1554.9 | 3605.4 | 241.4 | 222.6 | 556.1 | 2134.9 | 2013.0 |
| cJun | 1305.1 | 560.6 | 5047.0 | 267.1 | 418.5 | 2437.4 | 6727.9 | 5978.1 |
| JunB | 888.6 | 206.7 | 1224.9 | 289.0 | 272.8 | 907.7 | 3161.1 | 2901.4 |
| JunD | 1268.3 | 390.4 | 2837.4 | 250.9 | 273.8 | 1210.7 | 4386.9 | 3817.1 |
| Fos | 2324.4 | 4563.1 | 15804.4 | 842.0 | 3725.9 | 23273.9 | 38005.9 | 25564.4 |
| Fra2 | 239.9 | 609.2 | 13441.9 | 512.2 | 1444.0 | 11485.3 | 26424.6 | 17567.3 |
| ATF-3 | -197.5 | 1084.4 | -518.9 | 381.8 | 408.1 | 1836.2 | 5583.1 | 4559.3 |
| ATF-4 | 43709.7 | 5986.1 | 45654.4 | 362.3 | 249.4 | -547.3 | -1359.8 | -1113.3 |
| ATF-5 | 2216.7 | 1428.9 | -2845.4 | 311.5 | 221.1 | -957.7 | -3448.1 | -3212.2 |
| B-ATF | -117.8 | 371.0 | 359.6 | 231.8 | 253.8 | 1175.3 | 3755.0 | 3799.2 |
| p21SNFT | 26.2 | -42.3 | 765.4 | 353.7 | 489.8 | 2748.9 | 7858.3 | 6502.9 |
| HLF | -408.0 | -1247.6 | -2333.3 | 224.8 | 145.4 | -126.4 | -459.4 | -202.9 |
| MafG | 7202.8 | 11536.9 | 13977.4 | 274.2 | 234.8 | 1282.9 | 4000.7 | 3615.1 |
| cMaf | 283.4 | 2256.8 | 3369.4 | 334.1 | 336.7 | 1223.8 | 3534.1 | 2918.4 |
| MafB | 324.2 | 1583.5 | 1353.0 | 360.1 | 398.9 | 1641.6 | 5262.9 | 4428.6 |
| NFE2 | 5185.0 | 10083.2 | 35514.2 | 268.6 | 247.3 | 628.6 | 2240.4 | 1950.7 |
| NFE2L1 | 13690.0 | 4757.8 | 1653.6 | 270.4 | 226.1 | 105.8 | 1111.9 | 1050.1 |
| NFE2L3 | 6394.8 | -428.0 | -1980.8 | 283.1 | 217.7 | -588.6 | -1904.6 | -1727.6 |
| BACH1 | 3044.0 | 4953.7 | 11213.1 | 341.3 | 614.3 | 5078.1 | 14230.9 | 10864.3 |
| anti-ATF2-2 | | | | | | | | |
| anti-ATF3 | | | | | | | | |
| anti-ATF3-2 | | | | | | | | |
| anti-JUN | | | | | | | | |
| anti-FOS | | | | 269.6 | 299.3 | 941.1 | 3211.7 | 2952.8 |
| anti-SMAF-2 | -273.9 | | | | | | | |
| anti-SMAF-3 | | 6627.9 | | | | | | |
| anti-NFE2 | | | 21017.7 | | | | | |

$S_{array}$ values

| protein | ATF-2 | cJun | Fos | Fra2 | ATF-3 | ATF-4 | p21SNFT | MafG |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -0.4 | -0.3 | -0.2 | -0.4 | 0.3 | 6.3 | 0.5 | -0.6 |
| C/EBPβ | -0.8 | -0.7 | -1.2 | -1.7 | 0.9 | 1.8 | 0.5 | 0.8 |
| C/EBPδ | -1.2 | -0.7 | -1.0 | -0.6 | 0.9 | 5.7 | 0.7 | -0.4 |
| C/EBPγ | 2.0 | 0.3 | 0.0 | 0.4 | 4.1 | 13.4 | 1.8 | 0.1 |
| CHOP | 2.0 | 1.0 | 2.8 | 3.3 | 4.7 | 2.0 | 7.4 | 0.8 |
| ATF-1 | -0.7 | -1.1 | -0.9 | -0.5 | -1.2 | -1.3 | -1.3 | -0.6 |
| CREB | -2.3 | -1.5 | -1.2 | -1.5 | -1.7 | -1.6 | -2.1 | -1.3 |
| CREB-H | -0.3 | -1.0 | -0.4 | -0.1 | -0.8 | -0.4 | -0.6 | -0.5 |
| CREB3 | -0.2 | -0.4 | -1.0 | -0.6 | -0.4 | -1.1 | -0.2 | -0.1 |
| ATF-6 | -1.4 | -1.3 | -1.0 | -0.8 | -1.2 | -0.9 | -1.3 | -1.7 |
| ZF | -0.2 | 0.2 | -0.6 | -0.1 | 0.0 | 1.9 | 0.0 | 1.8 |
| XBP-1 | -0.7 | -0.5 | -0.8 | -1.8 | -0.7 | -0.7 | -1.3 | -0.5 |
| E4BP4 | -0.7 | -0.6 | -0.9 | -0.3 | -0.5 | -1.0 | -0.3 | 0.3 |
| ATF-2 | 2.7 | 5.1 | 2.8 | 3.8 | 3.7 | 0.1 | 1.8 | -0.5 |
| ATF-7 | 3.7 | 8.5 | 3.4 | 6.1 | 4.2 | 1.4 | 2.8 | -0.9 |
| cJun | 5.7 | 2.0 | 14.6 | 19.8 | 4.7 | -0.5 | 4.6 | -0.6 |
| JunB | 1.9 | 0.4 | 8.6 | 13.4 | 3.2 | -0.6 | 3.7 | 0.1 |
| JunD | 3.8 | 0.7 | 12.3 | 18.0 | 4.3 | -0.9 | 4.4 | 0.3 |
| Fos | 5.4 | 37.1 | 0.8 | 1.3 | 0.6 | 2.6 | -0.1 | -1.7 |
| Fra2 | 3.8 | 20.2 | 1.8 | 0.7 | 1.4 | 0.1 | -0.3 | -1.2 |
| ATF-3 | 7.6 | 12.7 | 1.1 | 5.0 | -0.7 | 4.1 | 5.5 | 0.4 |
| ATF-4 | 0.7 | -1.2 | 4.5 | 0.0 | 4.1 | -0.9 | 5.7 | -0.6 |
| ATF-5 | -1.8 | -2.2 | -1.7 | -1.3 | -1.4 | -1.7 | 1.3 | -1.9 |
| B-ATF | 0.9 | 14.1 | -0.2 | -0.4 | 1.0 | 1.4 | 0.6 | -1.3 |
| p21SNFT | 3.5 | 13.4 | -0.2 | 0.2 | 4.5 | 2.9 | 2.0 | 2.9 |
| HLF | -1.0 | -0.3 | -1.0 | -0.5 | -0.9 | -1.7 | -0.4 | -0.9 |
| MafG | 0.0 | 0.2 | 0.0 | -0.2 | -0.1 | -0.4 | 0.6 | 2.9 |
| cMaf | 0.0 | 0.0 | -0.3 | 0.3 | -0.3 | 0.4 | -0.9 | 0.0 |
| MafB | -0.2 | -0.3 | 0.9 | 1.5 | 0.2 | -1.1 | -0.8 | -0.9 |
| NFE2 | -0.9 | -0.7 | -0.7 | -0.8 | -0.9 | -0.4 | -1.2 | 0.8 |
| NFE2L1 | 0.1 | -0.4 | 0.0 | 0.2 | -1.1 | 1.3 | -0.8 | 79.2 |
| NFE2L3 | -1.6 | -1.8 | -2.2 | -2.0 | -1.6 | -0.5 | -1.9 | 64.2 |
| BACH1 | 2.4 | 0.4 | 0.7 | 1.8 | -0.8 | 0.2 | -0.1 | 38.0 |
| anti-ATF2-2 | 0.3 | -0.5 | -0.9 | -1.0 | -0.9 | -0.6 | 1.2 | -0.3 |
| anti-ATF3 | 2.8 | 1.7 | 10.5 | 12.1 | 6.0 | 6.3 | 3.6 | 6.9 |
| anti-ATF3-2 | 0.3 | 0.5 | 7.2 | 6.1 | 1.1 | 0.0 | 0.0 | 1.2 |
| anti-JUN | 1.0 | 7.3 | 0.8 | 0.6 | 0.7 | 2.5 | 0.7 | 2.0 |
| anti-FOS | 0.0 | 2.8 | 21.3 | 18.0 | 0.3 | -0.7 | 1.3 | 2.5 |
| anti-SMAF-2 | -0.7 | 1.9 | 0.7 | -0.7 | -0.8 | 8.4 | -1.0 | 16.4 |
| anti-SMAF-3 | -0.2 | -0.4 | -0.5 | -0.4 | -1.0 | -0.7 | -0.8 | 0.0 |
| anti-NFE2 | 0.1 | -0.4 | 3.1 | 2.6 | -0.6 | 3.8 | -0.4 | 6.2 |

| protein | NFE2 | NFE2L1 | BACH1 | anti-ATF2-2 | anti-ATF3 | anti-ATF3-2 | anti-JUN | anti-FOS |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -0.3 | 0.1 | -0.9 | -0.6 | 0.5 | -0.9 | -1.1 | -1.0 |
| C/EBPβ | -1.9 | -1.6 | -1.2 | -1.9 | -0.5 | 0.2 | -0.9 | -0.2 |
| C/EBPδ | -0.2 | -0.3 | -1.7 | -0.4 | 0.3 | 0.0 | -0.5 | -1.0 |
| C/EBPγ | 0.0 | -0.4 | -0.8 | 6.6 | 1.4 | 0.5 | 0.8 | 1.6 |
| CHOP | -0.1 | -0.4 | 1.3 | 11.1 | 1.8 | 1.0 | 3.9 | 0.6 |
| ATF-1 | -1.1 | -1.3 | 0.3 | -1.7 | -0.8 | 0.7 | 2.4 | -1.2 |
| CREB | -2.1 | -1.5 | 0.9 | -1.8 | -0.5 | 0.8 | 1.1 | -1.3 |
| CREB-H | 0.1 | -0.3 | -0.9 | 0.0 | -0.9 | -1.6 | -0.3 | 0.6 |
| CREB3 | 0.0 | -0.6 | -1.1 | 0.8 | -1.1 | 0.0 | 0.5 | 0.5 |
| ATF-6 | -0.6 | -1.2 | -0.9 | -0.7 | -1.1 | -0.6 | -0.5 | -1.2 |
| ZF | 1.8 | 11.4 | 0.5 | 3.2 | 0.3 | -0.2 | -0.4 | -0.8 |
| XBP-1 | -0.4 | -0.7 | -0.4 | -1.0 | -2.7 | -0.2 | -1.6 | -1.2 |
| E4BP4 | 0.0 | -0.9 | -0.7 | 1.2 | -1.3 | -0.6 | 0.7 | -0.6 |
| ATF-2 | -0.7 | 1.7 | 1.7 | 5.3 | 0.0 | -0.4 | -0.9 | -0.6 |
| ATF-7 | 0.8 | 3.8 | 2.6 | 10.8 | 0.3 | -0.7 | 0.1 | -0.1 |
| cJun | -0.9 | -1.0 | -0.8 | 0.5 | 0.2 | -0.9 | 8.8 | 2.4 |
| JunB | 0.3 | -1.0 | -1.0 | -0.3 | -0.5 | -0.3 | 1.8 | 0.3 |
| JunD | 0.1 | -0.6 | -0.7 | -0.4 | -0.1 | -1.0 | 3.7 | 0.9 |
| Fos | -0.2 | 2.9 | 0.7 | 0.2 | 16.3 | 45.6 | -0.7 | 27.0 |
| Fra2 | 0.0 | 0.8 | 0.3 | -0.6 | 3.8 | 13.5 | -0.5 | 13.2 |
| ATF-3 | -1.6 | -1.4 | -0.4 | 0.0 | 19.9 | 7.6 | 1.6 | 0.7 |
| ATF-4 | -0.5 | 14.5 | 0.6 | 6.6 | 17.4 | 3.0 | 2.5 | -0.9 |
| ATF-5 | 0.1 | -0.5 | 0.0 | 3.3 | -0.4 | 0.3 | -0.1 | -1.6 |
| B-ATF | 0.4 | 1.8 | 1.6 | 3.1 | 0.7 | -0.4 | -0.7 | 0.1 |
| p21SNFT | -0.4 | -0.3 | 1.0 | 8.8 | 4.2 | 1.7 | 0.1 | 1.8 |
| HLF | -2.0 | -0.8 | -0.8 | -0.8 | -0.6 | -1.2 | -0.3 | -1.0 |
| MafG | 6.0 | 26.0 | 13.2 | 0.7 | 0.8 | 2.6 | 0.8 | 0.4 |
| cMaf | 0.0 | 2.0 | 2.2 | -0.8 | -0.4 | 0.5 | -0.3 | 0.2 |
| MafB | 0.1 | 2.2 | 5.7 | -0.7 | -0.6 | 0.2 | -0.2 | 0.6 |
| NFE2 | 8.2 | -0.7 | -0.4 | -0.8 | 0.0 | -0.5 | -0.7 | -0.4 |
| NFE2L1 | 0.2 | 2.6 | -0.9 | 0.1 | 2.1 | 0.4 | 0.2 | -0.6 |
| NFE2L3 | 6.9 | -1.8 | -1.4 | -0.9 | -1.2 | -2.0 | -2.8 | -1.3 |
| BACH1 | 0.1 | 0.2 | 0.6 | -0.6 | 1.5 | 0.2 | 1.0 | 4.7 |
| anti-ATF2-2 | -1.0 | 6.9 | -1.7 | 4.8 | | | | |
| anti-ATF3 | 7.9 | 23.5 | 5.5 | | -0.9 | | | |
| anti-ATF3-2 | 0.6 | 4.7 | -0.6 | | | -1.2 | | |
| anti-JUN | 0.7 | 8.4 | 3.0 | | | | 1.2 | |
| anti-FOS | 0.1 | 0.0 | 9.0 | | | | | 0.9 |
| anti-SMAF-2 | 13.2 | 22.4 | 2.1 | | | | | |
| anti-SMAF-3 | 1.0 | 5.9 | -0.9 | | | | | |
| anti-NFE2 | 16.8 | 2.4 | 3.1 | | | | | |

| protein | anti-SMAF-2 | anti-SMAF-3 | anti-NFE2 | anti-FOS80 | anti-FOS200 | anti-FOS500 | anti-FOS1000 | anti-FOS2000 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -0.9 | -0.3 | 0.0 | -1.9 | -0.1 | -0.2 | -0.5 | -0.6 |
| C/EBPβ | -1.2 | -1.8 | -1.5 | 0.9 | 0.3 | -0.4 | -0.3 | -0.2 |
| C/EBPδ | -0.5 | -0.5 | -0.3 | -1.7 | -2.4 | -1.0 | -1.0 | -0.9 |
| C/EBPγ | -0.4 | 0.6 | -0.1 | -0.3 | 2.1 | 1.5 | 1.3 | 1.1 |
| CHOP | -0.6 | -0.4 | -0.2 | -0.1 | 0.7 | 1.0 | 0.9 | 0.5 |
| ATF-1 | 0.0 | -1.3 | 0.0 | 2.7 | 1.8 | -0.5 | -0.6 | -0.7 |
| CREB | -1.0 | -2.3 | -1.4 | 3.0 | 1.5 | -1.4 | -1.5 | -1.6 |
| CREB-H | -0.4 | -0.2 | -0.3 | -2.1 | -0.9 | 0.6 | 0.5 | 0.6 |
| CREB3 | 0.0 | -0.2 | -0.4 | 1.3 | 0.9 | 0.2 | 0.6 | 0.4 |
| ATF-6 | -1.1 | -1.1 | -1.1 | 0.0 | -1.1 | -1.4 | -1.2 | -1.0 |
| ZF | 1.2 | 2.7 | -0.8 | 0.1 | -0.4 | -1.0 | -1.0 | -1.0 |
| XBP-1 | -2.5 | -1.1 | -1.0 | 0.2 | 0.4 | -0.3 | -0.4 | -0.2 |
| E4BP4 | -0.6 | -0.7 | -0.6 | -0.9 | -1.1 | -0.5 | -0.3 | -0.3 |
| ATF-2 | -1.0 | 0.0 | 2.3 | 0.3 | -0.2 | -0.6 | -0.5 | -0.4 |
| ATF-7 | 0.0 | 0.9 | 1.4 | -1.1 | -0.8 | -0.2 | -0.2 | -0.2 |
| cJun | 3.4 | 0.1 | 2.2 | -0.5 | 2.2 | 1.8 | 1.3 | 1.3 |
| JunB | 2.3 | -0.2 | 0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.2 |
| JunD | 3.3 | 0.0 | 1.1 | -0.9 | 0.0 | 0.5 | 0.5 | 0.5 |
| Fos | 6.2 | 3.3 | 7.6 | 12.5 | 52.9 | 24.0 | 11.4 | 8.4 |
| Fra2 | 0.5 | 0.2 | 6.4 | 5.1 | 17.9 | 11.4 | 7.6 | 5.5 |
| ATF-3 | -0.7 | 0.5 | -0.6 | 2.1 | 2.1 | 1.1 | 0.9 | 0.8 |
| ATF-4 | 119.3 | 4.4 | 22.6 | 1.7 | -0.4 | -1.4 | -1.3 | -1.3 |
| ATF-5 | 5.9 | 0.8 | -1.8 | 0.5 | -0.8 | -1.8 | -2.0 | -2.1 |
| B-ATF | -0.5 | 0.0 | -0.2 | -1.3 | -0.3 | 0.4 | 0.3 | 0.5 |
| p21SNFT | -0.1 | -0.4 | 0.0 | 1.5 | 3.3 | 2.1 | 1.7 | 1.5 |
| HLF | -1.2 | -1.3 | -1.5 | -1.5 | -2.0 | -1.0 | -1.0 | -1.0 |
| MafG | 19.5 | 8.8 | 6.7 | -0.3 | -0.6 | 0.5 | 0.4 | 0.4 |
| cMaf | 0.6 | 1.5 | 1.3 | 1.0 | 1.0 | 0.5 | 0.3 | 0.2 |
| MafB | 0.8 | 0.9 | 0.3 | 1.6 | 1.9 | 0.9 | 0.8 | 0.7 |
| NFE2 | 14.0 | 7.6 | 17.5 | -0.5 | -0.4 | -0.1 | -0.1 | -0.2 |
| NFE2L1 | 37.3 | 3.4 | 0.5 | -0.4 | -0.7 | -0.7 | -0.5 | -0.5 |
| NFE2L3 | 17.3 | -0.7 | -1.4 | -0.1 | -0.9 | -1.4 | -1.5 | -1.5 |
| BACH1 | 8.2 | 3.6 | 5.3 | 1.2 | 5.2 | 4.6 | 3.7 | 3.1 |
| anti-ATF2-2 | | | | | | | | |
| anti-ATF3 | | | | | | | | |
| anti-ATF3-2 | | | | | | | | |
| anti-JUN | | | | | | | | |
| anti-FOS | | | | -0.4 | 0.4 | 0.2 | 0.2 | 0.2 |
| anti-SMAF-2 | -0.9 | | | | | | | |
| anti-SMAF-3 | | 4.9 | | | | | | |
| anti-NFE2 | | | 10.2 | | | | | |

Table B.S4. Average background-corrected fluorescence values and $S_{array}$ values) from round 2 of array measurements. Peptides on the surface are in rows, those in solution in columns. Duplicate measurements are marked with a number two in parentheses. The anti-XBP1 peptide was also tested at a concentration of 800 nM, as indicated in the probe name.

| protein | C/EBPα | C/EBPβ | C/EBPδ | C/EBPγ | CHOP | CREB | CREB3 | ATF-6 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 14378.8 | 10132.1 | 19625.2 | 3902.0 | 20179.0 | -633.1 | -452.9 | -540.1 |
| C/EBPβ | 12039.9 | 3042.9 | 13983.8 | 3908.8 | 42048.5 | -840.8 | -1298.5 | -1070.9 |
| C/EBPδ | 15951.5 | 10389.3 | 15986.0 | 7234.7 | 16729.9 | -239.9 | 255.3 | -496.9 |
| C/EBPγ | 13809.7 | 8788.3 | 24250.3 | 1623.4 | 22187.2 | 2080.9 | 1872.9 | 1414.9 |
| CHOP | 19799.0 | 31077.8 | 23783.9 | 14532.8 | 5784.9 | 2708.6 | 4397.6 | 1670.5 |
| ATF-1 | -840.7 | -1925.9 | -3695.5 | -1632.7 | -819.4 | 15719.9 | -1393.8 | -630.9 |
| CREB | -1528.8 | -3224.4 | -3356.7 | -1296.3 | -640.9 | 17969.9 | -319.3 | -4368.5 |
| CREB-H | 310.2 | -347.4 | -101.5 | 216.0 | 25.6 | 2800.9 | 3173.6 | 2037.7 |
| CREB3 | 399.4 | -1604.3 | -821.1 | 138.4 | 102.9 | 2102.9 | 2519.8 | 269.0 |
| ATF-6 | -217.0 | -1167.4 | -1166.7 | -669.8 | -613.3 | 569.8 | -208.3 | 14541.8 |
| ZF | 132.5 | -958.8 | -1345.8 | -420.2 | 682.1 | -730.3 | -318.0 | 3118.7 |
| XBP-1 | 668.6 | 48.5 | -852.8 | -180.8 | -632.3 | 3970.9 | 271.3 | 26091.6 |
| E4BP4 | 244.1 | -1225.6 | 1344.1 | 737.1 | 835.3 | 15648.5 | 2934.3 | 2081.4 |
| ATF-2 | 242.1 | -447.0 | -1214.4 | 2156.6 | 1323.1 | 1119.5 | 58.5 | 387.8 |
| ATF-7 | 4336.8 | 3877.6 | 3626.4 | 5025.4 | 6198.5 | 2764.8 | 897.5 | 1302.3 |
| cJun | 1976.4 | 1017.8 | 2582.0 | 797.0 | 1445.8 | 3880.9 | 20.9 | 1615.4 |
| JunB | 641.5 | -776.1 | -285.5 | -589.3 | -139.1 | 2234.1 | -575.7 | 1830.8 |
| JunD | 690.4 | -818.8 | 1029.9 | 101.4 | 779.2 | 3247.6 | -277.2 | 1564.9 |
| Fos | 4014.0 | 415.9 | 2099.0 | 885.0 | 3244.0 | 1392.1 | -263.8 | 366.6 |
| Fra2 | 1437.7 | -28.4 | 627.0 | 436.3 | 1301.9 | 2829.6 | -426.0 | 1812.5 |
| ATF-3 | 3521.3 | 4105.6 | 6042.5 | 13256.1 | 9384.3 | -557.1 | 311.8 | -1090.1 |
| ATF-4 | 41351.2 | 14860.3 | 45224.3 | 45710.6 | 29556.1 | 162.3 | 835.8 | -1395.9 |
| ATF-5 | 11769.3 | 1232.0 | 13935.9 | 43585.1 | -930.4 | -2876.8 | -844.9 | -3743.6 |
| B-ATF | 4622.4 | 5606.3 | 8968.8 | 7972.0 | 16112.8 | 4414.1 | 1948.8 | 2878.9 |
| p21SNFT | 6622.4 | 4153.6 | 13175.2 | 4778.7 | 25657.1 | 2509.4 | 4765.3 | 1481.5 |
| TEF | -581.9 | -2005.1 | -2087.2 | -845.8 | -327.3 | -1793.1 | -843.6 | -3237.4 |
| MafG | 30.3 | -714.3 | 307.3 | -113.3 | 929.2 | 4842.8 | -369.2 | 2885.7 |
| cMaf | 459.7 | 221.8 | -674.3 | -200.2 | -729.1 | 2294.7 | -119.3 | 1454.9 |
| MafB | 417.9 | -1434.6 | -2046.9 | -944.7 | -1315.3 | 2019.0 | -1269.5 | 1752.3 |
| NFE2 | -606.8 | -1494.0 | -1943.2 | -387.3 | -570.9 | 736.8 | -0.4 | 1107.4 |
| NFE2L1 | 540.0 | -611.9 | 807.6 | -764.6 | -1143.4 | 6114.6 | -197.1 | 1019.3 |
| NFE2L3 | -535.8 | -1643.8 | -1520.6 | -1795.9 | -1347.8 | -2402.1 | -1392.2 | -3045.3 |
| BACH1 | 484.6 | -190.1 | -961.0 | 159.6 | 456.4 | 7972.2 | 377.3 | 3421.4 |
| anti-XBP1-2 | 576.8 | -482.4 | -723.1 | 82.7 | -277.3 | 3063.0 | -157.6 | 6119.6 |
| anti-XBP1 | 931.6 | -701.3 | -865.4 | -532.0 | -1009.9 | 3870.8 | -203.9 | 6672.2 |
| anti-BATF | 1552.4 | 3090.8 | 1613.1 | 6935.3 | 25631.3 | 2435.3 | 1450.1 | 62.3 |
| anti-SMAF | 14.3 | -448.7 | -440.9 | 319.3 | -406.3 | 5856.8 | -150.9 | 1891.0 |
| anti-E4BP4-2 | 752.3 | -323.8 | -141.9 | 73.1 | -285.4 | 4925.5 | -52.9 | 3899.9 |
| anti-E4BP4 | 994.4 | -111.4 | -706.5 | 167.3 | 2327.4 | 4962.2 | 470.4 | 3062.4 |
| anti-C/EBPγ | 3702.3 | 7158.3 | 16172.7 | 22417.9 | 16943.1 | 3989.9 | 2729.5 | 3093.9 |
| anti-C/EBPγ-2 | 15069.1 | 12318.4 | 40112.9 | 14940.6 | 13519.5 | 4375.3 | 5279.5 | 1591.3 |
| anti-ATF4 | 10508.0 | 2319.8 | 22113.4 | 2202.3 | 3284.8 | 6850.2 | 16656.5 | 1201.3 |
| anti-ATF4-2 | 12446.6 | 897.1 | 16506.4 | 1646.1 | 4607.8 | 2106.3 | 11698.6 | 809.5 |
| anti-BACH-2 | 1721.8 | -1163.9 | -288.2 | -742.6 | -535.4 | 143.0 | -758.2 | -72.4 |
| anti-BACH | 4252.5 | 807.8 | 770.4 | 306.1 | -525.6 | 6946.3 | 1033.5 | 1306.5 |
| anti-ATF2-3 | 396.9 | -284.6 | 2357.2 | 3461.2 | 4946.7 | 11952.9 | 3663.7 | 5248.6 |
| anti-ZF-2 | -610.4 | -366.9 | 283.9 | 112.8 | 249.1 | 6619.8 | 2013.3 | 18946.3 |
| anti-CREB | -932.1 | -991.4 | -1724.9 | 1184.6 | 41.0 | 12012.5 | 1675.7 | 2727.9 |
| anti-C/EBP-2 | 33587.8 | 21296.4 | 36955.0 | 27319.6 | 25905.8 | 2010.7 | 1682.1 | 2094.3 |
| anti-OASIS | 2627.1 | 6574.8 | 10764.7 | 6089.2 | -325.4 | 5051.8 | 27429.9 | 1417.1 |
| Anti-OASIS-2 | 1857.7 | 1113.3 | 1578.4 | 5850.9 | 1885.6 | 3453.4 | 15075.7 | 1725.9 |
| | | | | | | | | |

| protein | ZF | XBP-1 | E4BP4 | ATF-2 | cJun | Fos | ATF-3 | ATF-4 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -2414.9 | 660.8 | -751.3 | 445.3 | -31.9 | 2536.2 | 3711.4 | 34359.1 |
| C/EBPβ | -1722.3 | 2157.9 | -1934.4 | -596.2 | -559.8 | -1107.8 | 4086.3 | 6426.5 |
| C/EBPδ | -4880.8 | 361.1 | -415.8 | -878.8 | -443.6 | -1296.7 | 7159.6 | 34679.4 |
| C/EBPγ | -214.3 | 2398.1 | 3749.6 | 6527.0 | 355.6 | 640.1 | 22896.1 | 24108.8 |
| CHOP | 9281.3 | 2843.9 | 8722.0 | 7308.7 | 1574.7 | 3996.6 | 23808.6 | 11390.3 |
| ATF-1 | -2319.9 | 3468.7 | 5413.0 | -1314.6 | -319.8 | -531.9 | 112.0 | -4293.8 |
| CREB | -8696.3 | 1457.4 | 4003.8 | -2087.8 | -1269.1 | -1561.0 | -1440.7 | -4408.7 |
| CREB-H | 1726.6 | 2271.0 | 3409.1 | 65.5 | 73.0 | -334.1 | 1992.5 | 840.3 |
| CREB3 | 1167.4 | 2201.3 | 876.5 | -452.3 | -583.4 | -943.0 | 1259.4 | -1027.1 |
| ATF-6 | 1511.0 | 9836.3 | -147.1 | -886.0 | -423.4 | -479.1 | 354.5 | -717.3 |
| ZF | 32236.8 | 22213.9 | -437.1 | 704.9 | 261.0 | -238.3 | 4145.3 | 8785.6 |
| XBP-1 | 27480.3 | 31541.8 | -161.8 | 221.9 | 66.7 | -1327.7 | 3023.8 | 1618.1 |
| E4BP4 | 1962.8 | 2244.8 | 40608.3 | -361.0 | -543.9 | -811.4 | 4647.6 | -3189.4 |
| ATF-2 | -555.1 | 3930.3 | -561.5 | 4995.7 | 4631.9 | 3683.7 | 12381.9 | 1308.3 |
| ATF-7 | 4956.8 | 2492.6 | -37.1 | 7407.3 | 8615.6 | 6191.6 | 19157.9 | 9845.1 |
| cJun | 6370.2 | 2653.2 | -744.8 | 13287.4 | 2943.1 | 28143.6 | 23634.2 | 703.5 |
| JunB | 4287.3 | 2020.8 | -1211.3 | 4806.7 | 715.9 | 18676.8 | 19483.0 | -3499.4 |
| JunD | 5962.6 | 1892.7 | -1449.3 | 8136.4 | 828.8 | 22002.7 | 21914.4 | -18.3 |
| Fos | 8015.2 | 996.9 | -1.2 | 8113.9 | 29520.1 | 1778.4 | 3488.7 | 10138.6 |
| Fra2 | 8316.2 | 3028.1 | -1182.9 | 7411.4 | 17576.8 | 2955.3 | 5901.2 | 4285.1 |
| ATF-3 | 6441.8 | 3085.6 | 888.3 | 10483.1 | 9526.3 | 1960.8 | 675.6 | 15171.7 |
| ATF-4 | 27125.9 | 900.9 | -851.6 | 1706.2 | -483.2 | 5716.6 | 11746.7 | -926.9 |
| ATF-5 | -1674.8 | -1932.1 | -835.9 | -2308.8 | -1581.6 | -2499.8 | -2249.5 | -4026.6 |
| B-ATF | 7504.8 | 1769.3 | 2782.0 | 4070.7 | 16188.1 | -582.0 | 8664.3 | 13362.1 |
| p21SNFT | 10062.6 | 2309.4 | 2341.2 | 7257.2 | 12525.1 | 374.5 | 19961.8 | 14456.7 |
| TEF | -2597.3 | 3076.5 | -751.1 | -1283.2 | -934.8 | -1822.0 | 123.3 | -3493.4 |
| MafG | 6196.1 | 4109.8 | 113.4 | 588.8 | 1459.4 | 163.1 | 1756.7 | 1110.1 |
| cMaf | 2434.0 | 2664.9 | 11.3 | 1171.9 | 245.6 | 656.3 | 3638.8 | 5309.9 |
| MafB | 4784.0 | 3499.1 | -522.7 | 833.6 | 22.9 | 1953.0 | 5563.9 | -1037.3 |
| NFE2 | 3751.2 | 914.6 | -1195.4 | 316.3 | -410.9 | -1275.6 | -692.5 | -718.5 |
| NFE2L1 | 27090.3 | 1167.9 | -298.1 | 1388.1 | -109.6 | 1058.5 | 363.6 | 6786.6 |
| NFE2L3 | -2936.2 | -866.4 | -787.3 | -2125.3 | -1338.3 | -1975.7 | -2200.8 | -1621.5 |
| BACH1 | 15754.4 | 4292.8 | 1411.6 | 5134.1 | 523.2 | 1516.8 | 1600.4 | 2707.9 |
| anti-XBP1-2 | 15814.3 | 8919.3 | -309.7 | 27.9 | 554.3 | -193.4 | 2804.9 | 730.9 |
| anti-XBP1 | 11306.4 | 5129.6 | -1151.9 | -652.3 | 990.0 | -325.0 | 1342.6 | 1316.3 |
| anti-BATF | 28913.5 | 3469.6 | -811.8 | 1959.2 | -425.8 | 12467.3 | 5325.4 | 4115.8 |
| anti-SMAF | 4048.3 | 2717.3 | -290.9 | 886.8 | 4516.6 | 1786.1 | 3155.2 | 1872.6 |
| anti-E4BP4-2 | 2103.6 | 2188.9 | 1218.9 | 394.9 | 506.8 | -648.4 | 3556.6 | -672.4 |
| anti-E4BP4 | 2764.3 | 2628.6 | 5120.1 | 54.1 | 450.8 | 415.0 | 2805.8 | 218.1 |
| anti-C/EBPγ | 27028.4 | 2991.2 | 737.5 | 1480.3 | 2045.8 | 5762.1 | 12269.3 | 11299.2 |
| anti-C/EBPγ-2 | 12425.8 | 3463.8 | 1396.4 | 9184.3 | 5304.9 | 4767.3 | 6109.3 | 32502.9 |
| anti-ATF4 | 24671.5 | 4353.9 | 3071.7 | 20937.1 | 9518.1 | 16634.8 | 20245.3 | 38930.5 |
| anti-ATF4-2 | 7739.3 | 1445.9 | -528.2 | 6264.4 | 1866.8 | 19898.6 | 17574.6 | 37258.2 |
| anti-BACH-2 | -1314.9 | 598.4 | -1776.3 | -838.9 | -298.7 | -244.9 | 1300.1 | -3506.2 |
| anti-BACH | 863.4 | 2160.3 | 465.4 | 1621.6 | 19.2 | 14561.6 | 4288.1 | 930.2 |
| anti-ATF2-3 | 28725.3 | 2959.3 | 4361.7 | 32284.5 | 2933.3 | 19785.3 | 16925.0 | 5705.9 |
| anti-ZF-2 | 24889.6 | 23267.8 | 2000.4 | 2820.9 | 2112.4 | 2913.6 | 3294.3 | 8336.9 |
| anti-CREB | 479.3 | 2586.8 | 1745.8 | -1018.0 | 328.2 | 1437.5 | 814.8 | -1579.0 |
| anti-C/EBP-2 | 18057.1 | 1833.5 | -841.7 | 3313.6 | 546.9 | 9064.3 | 17207.1 | 44465.6 |
| anti-OASIS | 1277.9 | 2128.5 | 8390.2 | -238.1 | 958.5 | 3741.5 | 13676.4 | 2058.8 |
| anti-OASIS-2 | 4944.3 | 2320.0 | 3345.6 | 2668.4 | 1646.3 | 3004.5 | 16412.8 | -527.3 |

| protein | p21SNFT | TEF | MafG | cMaf | NFE2 | BACH1 | anti-XBP1-2 | anti-XBP1 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 13324.3 | 445.9 | -192.1 | -290.8 | -1825.0 | -1600.7 | 132.8 | 238.3 |
| C/EBPβ | 6601.3 | -1122.6 | -675.8 | -128.6 | -5212.5 | -1075.1 | 384.8 | 460.9 |
| C/EBPδ | 24971.1 | 244.4 | -72.1 | -3718.6 | -1078.4 | -2389.9 | 257.3 | 503.1 |
| C/EBPγ | 18509.8 | 1652.1 | -72.7 | -13.1 | -1967.1 | 56.1 | 421.9 | 520.3 |
| CHOP | 26243.4 | 4398.8 | 19.8 | 1568.9 | -1860.4 | 6762.5 | 281.6 | 494.8 |
| ATF-1 | -11330.2 | -928.5 | -111.7 | -793.5 | -4380.5 | 2919.4 | 684.4 | 736.1 |
| CREB | -10522.6 | -4101.9 | -577.2 | -1270.0 | -6533.5 | 3999.6 | 259.1 | 482.2 |
| CREB-H | 3896.3 | 3423.8 | -123.0 | -351.1 | -851.3 | -426.2 | 466.6 | 536.6 |
| CREB3 | 4114.5 | -164.6 | -184.4 | 83.8 | -1348.8 | -514.8 | 376.8 | 550.4 |
| ATF-6 | -5640.9 | -60.1 | -330.7 | 191.1 | -3878.0 | -90.1 | 696.0 | 609.6 |
| ZF | 1039.9 | 700.9 | 31.4 | -209.6 | 1887.8 | 3316.3 | 3912.9 | 1391.6 |
| XBP-1 | -2073.5 | 2784.7 | -173.4 | -971.7 | -3023.4 | 476.1 | 10564.9 | 2314.0 |
| E4BP4 | 10000.5 | 2939.3 | -319.8 | -1818.8 | -979.3 | -74.4 | 477.6 | 510.0 |
| ATF-2 | 24722.8 | 958.6 | 65.1 | 660.2 | -2828.9 | 6905.6 | 69.3 | 735.9 |
| ATF-7 | 29215.8 | 2030.9 | -231.6 | 694.2 | 1295.8 | 9041.8 | 315.6 | 383.3 |
| cJun | 27603.6 | 2241.5 | 1.0 | 1493.8 | -1682.4 | 1392.6 | 248.8 | 945.9 |
| JunB | 22546.1 | 1888.7 | -96.3 | -116.7 | -1276.8 | -44.5 | 274.8 | 729.2 |
| JunD | 29892.1 | 1232.4 | -245.3 | 1303.6 | -857.7 | -214.9 | 481.0 | 515.9 |
| Fos | -404.2 | 2895.0 | -281.6 | 2130.1 | -2478.0 | 5113.0 | 177.6 | 690.6 |
| Fra2 | 2168.2 | 2199.8 | -359.5 | 1268.9 | -1696.9 | 2183.1 | 557.7 | 791.6 |
| ATF-3 | 31526.1 | 797.2 | 150.8 | 481.5 | -6966.8 | -493.4 | 616.3 | 746.8 |
| ATF-4 | 27759.2 | -3262.1 | 54.0 | 10884.5 | -1981.8 | 3517.3 | -28.6 | 595.3 |
| ATF-5 | 15657.6 | -4280.0 | -89.4 | -894.7 | -556.2 | 775.4 | 110.4 | 355.3 |
| B-ATF | 19458.4 | 8601.8 | -248.9 | 257.0 | 424.3 | 6754.7 | 307.3 | 468.6 |
| p21SNFT | 26991.1 | 4459.4 | 1263.4 | 855.8 | -3043.9 | 4764.4 | 724.9 | 830.8 |
| TEF | -5321.8 | 3396.4 | -419.1 | -849.2 | -5937.8 | -398.2 | 60.4 | 514.6 |
| MafG | 16893.3 | 2031.6 | 1969.6 | 69.3 | 17719.7 | 44305.4 | 560.3 | 573.1 |
| cMaf | -191.3 | 1700.3 | 52.5 | 24096.8 | -1070.4 | 9750.6 | 310.8 | 543.8 |
| MafB | -4965.8 | 2167.0 | -153.5 | 27520.9 | -1783.3 | 15627.7 | 606.1 | 657.8 |
| NFE2 | -5787.9 | 1823.6 | 380.2 | -557.8 | 19093.1 | 448.7 | 572.1 | 656.1 |
| NFE2L1 | -1869.8 | 1249.7 | 38273.3 | 1542.1 | 708.1 | 21.9 | 339.8 | 395.8 |
| NFE2L3 | -15234.8 | -3185.4 | 27391.5 | -1131.9 | 12016.6 | -1945.3 | 190.4 | 422.3 |
| BACH1 | 5125.7 | 3448.2 | 13684.1 | 8474.9 | -2608.5 | 3085.6 | 10529.3 | 1646.6 |
| anti-XBP1-2 | -1079.3 | 3235.8 | -59.6 | -423.9 | -352.8 | 9744.2 | 276.7 | |
| anti-XBP1 | -3135.4 | 3086.4 | -221.3 | -2219.3 | -485.8 | 2714.4 | | 406.4 |
| anti-BATF | 20417.3 | 9229.8 | 10298.6 | -84.5 | -1484.1 | 858.2 | | |
| anti-SMAF | 25246.4 | 2110.9 | 14168.2 | 2406.8 | 16335.1 | 5104.8 | | |
| anti-E4BP4-2 | -3405.6 | 4597.0 | -701.6 | -5.6 | -1181.0 | -733.8 | | |
| anti-E4BP4 | 7575.1 | 8804.1 | -23.4 | 43.4 | -3.6 | 1532.9 | | |
| anti-C/EBPγ | 24111.7 | 2821.7 | 305.2 | -1611.9 | -456.6 | 3704.6 | | |
| anti-C/EBPγ-2 | 24454.4 | 9770.1 | 413.5 | 32749.3 | 5732.5 | 5480.3 | | |
| anti-ATF4 | 27373.3 | 2251.5 | 6694.0 | 5583.4 | 30084.7 | 24527.0 | | |
| anti-ATF4-2 | 27211.6 | 542.3 | -173.4 | 1428.1 | 10552.1 | 5146.4 | | |
| anti-BACH-2 | -11212.4 | 3.4 | 3.5 | -551.8 | -628.9 | 2207.5 | | |
| anti-BACH | 7472.1 | 2119.3 | -200.9 | 6903.6 | 31129.8 | 33014.3 | | |
| anti-ATF2-3 | 24496.7 | 1652.6 | 2296.5 | 52.4 | -786.8 | 436.9 | | |
| anti-ZF-2 | 10086.7 | 525.9 | 1476.9 | 886.3 | 3044.9 | 11656.3 | | |
| anti-CREB | 15076.1 | 3523.3 | 611.5 | -219.1 | -1507.3 | 3384.6 | | |
| anti-C/EBP-2 | 28553.1 | 4660.8 | -149.3 | 2873.7 | 2558.2 | -156.1 | | |
| anti-OASIS | 21807.4 | 1829.9 | 3717.1 | -494.7 | -205.0 | -479.9 | | |
| anti-OASIS-2 | 20343.0 | 1975.1 | 2120.3 | 896.4 | -2148.8 | 3221.1 | | |

| protein | anti-BATF | anti-SMAF | anti-E4BP4-2 | anti-E4BP4 | anti-C/EBPγ | anti-C/EBPγ-2 | anti-ATF4 | anti-ATF4-2 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -621.8 | -378.9 | 351.2 | 118.4 | 1011.2 | 8545.8 | 2675.1 | 4290.0 |
| C/EBPβ | -971.1 | -1003.6 | 558.3 | 192.4 | 1258.9 | 8497.4 | -2170.3 | -1821.8 |
| C/EBPδ | -1583.3 | -588.2 | 556.9 | -292.6 | 3168.4 | 21820.2 | 4396.6 | 2623.7 |
| C/EBPγ | 1369.4 | 163.8 | 730.4 | 2936.7 | 23323.2 | 24328.1 | 1314.9 | -388.9 |
| CHOP | 24028.7 | 278.6 | 1064.4 | 37604.3 | 21183.6 | 18430.5 | 1951.2 | 3609.8 |
| ATF-1 | -866.1 | 282.9 | 905.1 | -118.7 | -880.3 | -1519.7 | -695.3 | -1609.2 |
| CREB | -3340.9 | -339.4 | 581.1 | -1504.4 | -523.8 | -1899.4 | -1782.3 | -1785.9 |
| CREB-H | 223.1 | 100.8 | 836.6 | 1076.1 | -450.6 | 222.5 | 357.9 | -83.7 |
| CREB3 | -785.2 | -108.3 | 748.8 | 116.5 | -590.8 | 702.8 | 8138.3 | 4481.9 |
| ATF-6 | -1388.6 | -212.5 | 653.0 | 526.0 | -1008.1 | -696.6 | -898.8 | -984.5 |
| ZF | 1145.9 | 319.2 | 557.6 | 571.6 | 4653.5 | 1968.3 | 5086.2 | 1821.2 |
| XBP-1 | 540.6 | -227.3 | 646.1 | 360.1 | -667.6 | -1705.1 | -1484.0 | -429.2 |
| E4BP4 | -804.8 | -618.3 | 855.1 | 5262.3 | 413.3 | -352.1 | -483.5 | -1838.6 |
| ATF-2 | -829.7 | -60.4 | 919.1 | 834.5 | -229.2 | 4940.3 | 25340.7 | 4014.0 |
| ATF-7 | -191.9 | 1639.7 | 486.4 | 1233.8 | 560.5 | 8720.0 | 14443.4 | 7965.4 |
| cJun | -1152.4 | 18511.9 | 496.9 | 1578.9 | 1205.7 | 9882.3 | 14654.0 | 3451.2 |
| JunB | -1038.8 | 5859.3 | 591.0 | 496.4 | -127.6 | 4819.8 | 13823.4 | -658.9 |
| JunD | -1265.7 | 6489.1 | 453.8 | 452.6 | 139.0 | 4999.3 | 16156.4 | 1327.0 |
| Fos | 5770.8 | 3978.6 | 661.3 | 1347.4 | 8366.1 | 6282.0 | 29530.1 | 34437.4 |
| Fra2 | 7879.5 | 1525.3 | 824.3 | 143.5 | 4514.6 | 13295.1 | 22081.8 | 10492.6 |
| ATF-3 | -387.4 | 1339.0 | 962.2 | 3528.4 | 6073.3 | 526.2 | 21316.4 | 22704.9 |
| ATF-4 | -587.6 | 746.6 | 748.1 | 282.1 | 8454.4 | 23207.6 | 39871.1 | 37368.8 |
| ATF-5 | -2197.9 | -632.6 | 522.4 | -1702.8 | 4543.3 | 1856.8 | 5264.9 | 1629.6 |
| B-ATF | 14011.9 | 650.6 | 645.8 | 7437.2 | 503.0 | -468.7 | 6056.6 | 1613.9 |
| p21SNFT | 15853.1 | 8520.8 | 869.1 | 8896.9 | 2080.4 | 4893.9 | 7767.7 | 7312.8 |
| TEF | -2356.1 | -457.4 | 714.1 | 2917.4 | -401.9 | -646.3 | -1845.6 | -2057.8 |
| MafG | 6080.5 | 42792.9 | 672.9 | 2946.1 | -602.6 | 1030.6 | 13716.0 | 584.3 |
| cMaf | 3.3 | 1748.2 | 618.8 | 1370.9 | -480.4 | 11262.8 | 699.6 | 565.5 |
| MafB | -568.6 | 847.7 | 729.1 | 754.2 | -1544.9 | 4030.3 | 381.8 | 128.5 |
| NFE2 | -765.6 | 10067.1 | 575.1 | 114.3 | -378.5 | -313.4 | 25018.7 | 3268.6 |
| NFE2L1 | -1128.8 | 938.8 | 448.0 | 206.2 | -894.1 | -681.3 | 27209.3 | 1491.0 |
| NFE2L3 | -1978.8 | -1158.5 | 476.7 | -1657.2 | -1318.4 | -1392.6 | 7552.6 | -1925.4 |
| BACH1 | 11.8 | 3091.8 | 827.4 | 5133.9 | 1349.4 | 1077.1 | 7647.7 | 1427.4 |
| anti-XBP1-2 | | | | | | | | |
| anti-XBP1 | | | | | | | | |
| anti-BATF | -1485.6 | | | | | | | |
| anti-SMAF | | 4741.2 | | | | | | |
| anti-E4BP4-2 | | | 593.8 | | | | | |
| anti-E4BP4 | | | | 12171.8 | | | | |
| anti-C/EBPγ | | | | | 1559.8 | | | |
| anti-C/EBPγ-2 | | | | | | 10249.5 | | |
| anti-ATF4 | | | | | | | 3540.6 | |
| anti-ATF4-2 | | | | | | | | 4570.4 |
| anti-BACH-2 | | | | | | | | |
| anti-BACH | | | | | | | | |
| anti-ATF2-3 | | | | | | | | |
| anti-ZF-2 | | | | | | | | |
| anti-CREB | | | | | | | | |
| anti-C/EBP-2 | | | | | | | | |
| anti-OASIS | | | | | | | | |
| anti-OASIS-2 | | | | | | | | |

| protein | anti-BACH-2 | anti-BACH | anti-ATF2-3 | anti-ZF-2 | anti-CREB | anti-C/EBP-2 |
|---|---|---|---|---|---|---|
| C/EBPα | 3026.8 | 819.7 | -2789.2 | -895.1 | -100.6 | 22362.9 |
| C/EBPβ | -2525.7 | -596.9 | -1430.0 | -1808.9 | -449.8 | 11580.0 |
| C/EBPδ | 4396.8 | 175.6 | -293.8 | -1413.6 | -873.8 | 18934.1 |
| C/EBPγ | 3661.0 | 806.5 | 22134.4 | -111.1 | 3934.9 | 26889.6 |
| CHOP | -974.1 | 1041.7 | 41196.5 | 1515.1 | 3048.2 | 30991.9 |
| ATF-1 | -3360.4 | 642.3 | 946.4 | -306.4 | 3148.6 | -1469.4 |
| CREB | -5286.4 | -261.1 | 6395.9 | -3750.5 | 2830.2 | -1195.2 |
| CREB-H | 457.4 | 1077.0 | 1467.7 | 1126.8 | 2010.1 | -811.4 |
| CREB3 | -661.9 | -535.3 | 4802.4 | 528.2 | 1683.2 | -1491.6 |
| ATF-6 | -1529.3 | -679.5 | -533.5 | 3824.5 | -637.5 | -934.2 |
| ZF | -2627.5 | -1016.6 | 30723.6 | 9987.6 | -478.3 | 1161.9 |
| XBP-1 | -893.6 | -1457.1 | 486.6 | 18013.8 | 20.0 | -1757.4 |
| E4BP4 | -55.8 | 143.1 | 7336.0 | 1712.3 | 230.4 | -2181.6 |
| ATF-2 | -941.4 | 2025.3 | 48886.9 | 1806.7 | -378.4 | -615.1 |
| ATF-7 | 243.8 | 433.4 | 53617.9 | 3740.3 | 580.1 | 3845.8 |
| cJun | 575.9 | -427.8 | 25474.4 | 5198.7 | 1563.1 | -416.6 |
| JunB | -120.2 | -583.4 | 12123.1 | 2123.8 | 168.8 | -1490.4 |
| JunD | 586.1 | -377.4 | 8849.9 | 2717.9 | -664.9 | -1019.7 |
| Fos | 3452.8 | 24319.0 | 49898.1 | 3532.8 | 5916.3 | 5173.0 |
| Fra2 | 330.4 | 12409.4 | 53166.4 | 4148.0 | 4027.0 | -1101.7 |
| ATF-3 | -4280.5 | 2880.0 | 50237.5 | -324.6 | 605.3 | 4072.4 |
| ATF-4 | 622.1 | 3715.4 | 917.6 | 4098.6 | -320.8 | 36727.6 |
| ATF-5 | -1828.6 | 137.1 | -4756.6 | -5862.7 | -359.6 | 3612.8 |
| B-ATF | 1873.9 | 4312.1 | 27785.4 | 1559.0 | 1262.4 | 3824.6 |
| p21SNFT | -1040.0 | 3808.8 | 48355.0 | 4508.4 | 7521.9 | 13473.9 |
| TEF | -3864.9 | -1713.0 | -2791.5 | -2796.9 | -1150.4 | -1515.3 |
| MafG | 4747.2 | 755.0 | 31517.3 | 2968.3 | 4127.9 | -854.8 |
| cMaf | -141.9 | 3569.3 | 1127.3 | 1845.4 | -397.1 | -130.3 |
| MafB | -618.8 | 3600.8 | -1693.4 | 1369.1 | -1146.1 | -2332.8 |
| NFE2 | 10477.2 | 30996.3 | -3207.4 | 4143.6 | -521.2 | -1285.9 |
| NFE2L1 | 6634.3 | 2780.9 | -1364.1 | 3576.6 | -127.1 | 1463.1 |
| NFE2L3 | -1925.3 | 201.1 | -5328.6 | -2656.3 | -1138.1 | -1718.4 |
| BACH1 | 4769.8 | 15576.9 | -693.9 | 5640.9 | 5151.3 | -1300.2 |
| anti-XBP1-2 | | | | | | |
| anti-XBP1 | | | | | | |
| anti-BATF | | | | | | |
| anti-SMAF | | | | | | |
| anti-E4BP4-2 | | | | | | |
| anti-E4BP4 | | | | | | |
| anti-C/EBPγ | | | | | | |
| anti-C/EBPγ-2 | | | | | | |
| anti-ATF4 | | | | | | |
| anti-ATF4-2 | | | | | | |
| anti-BACH-2 | 22601.3 | | | | | |
| anti-BACH | | 8223.3 | | | | |
| anti-ATF2-3 | | | 18632.6 | | | |
| anti-ZF-2 | | | | 5162.6 | | |
| anti-CREB | | | | | 16779.5 | |
| anti-C/EBP-2 | | | | | | -1467.1 |
| anti-OASIS | | | | | | |
| anti-OASIS-2 | | | | | | |

| protein | anti-OASIS | anti-OASIS-2 | anti-C/EBPγ(2) | anti-SMAF(2) | anti-XBP1800 |
|---|---|---|---|---|---|
| C/EBPα | -1009.8 | -748.7 | 905.3 | 478.0 | 237.5 |
| C/EBPβ | 951.6 | -1294.3 | 308.6 | 55.1 | 319.2 |
| C/EBPδ | -4.6 | -1457.8 | 2858.0 | -5.4 | 250.5 |
| C/EBPγ | 6182.5 | 6129.1 | 24063.3 | 2043.1 | 353.4 |
| CHOP | 1782.0 | 10525.9 | 20659.7 | 910.5 | 361.3 |
| ATF-1 | -1943.9 | 524.1 | -1120.9 | 1019.6 | 702.9 |
| CREB | -319.8 | -918.4 | -956.4 | 419.1 | 649.9 |
| CREB-H | 2421.3 | 4961.9 | -784.6 | 1199.8 | 407.9 |
| CREB3 | 7748.9 | 10026.6 | -1115.0 | -113.9 | 398.7 |
| ATF-6 | -260.6 | -374.9 | -893.6 | 235.9 | 1157.8 |
| ZF | -579.4 | 820.8 | 4418.3 | 1278.1 | 3886.3 |
| XBP-1 | 1126.6 | 3026.8 | -574.4 | 814.8 | 6189.5 |
| E4BP4 | 1063.3 | 497.4 | -1688.9 | 226.6 | 409.8 |
| ATF-2 | -450.9 | 262.1 | -539.6 | 1473.2 | 640.3 |
| ATF-7 | -956.2 | 3037.1 | 538.3 | 2280.6 | 373.7 |
| cJun | 1154.2 | 3427.1 | 513.6 | 22955.9 | 2842.2 |
| JunB | 67.4 | 3045.3 | -814.7 | 7609.4 | 1651.4 |
| JunD | 191.6 | 1698.8 | 257.9 | 7231.5 | 1081.8 |
| Fos | 3890.1 | 17550.4 | 7112.3 | 4832.6 | 915.1 |
| Fra2 | 801.3 | 13051.9 | 3398.8 | 1443.0 | 1139.6 |
| ATF-3 | 6281.5 | 18198.6 | 5734.2 | 1915.7 | 521.9 |
| ATF-4 | 943.7 | 796.4 | 6383.4 | 1328.9 | 294.3 |
| ATF-5 | -1073.9 | -1726.5 | 3136.0 | 1060.6 | 71.9 |
| B-ATF | 1871.4 | 4359.3 | 562.5 | 1649.7 | 448.6 |
| p21SNFT | 10563.2 | 10585.2 | 1450.6 | 11300.8 | 1313.9 |
| TEF | -1023.8 | -2560.5 | -925.2 | 968.2 | 340.4 |
| MafG | 6805.5 | 11081.9 | -324.3 | 42998.6 | 677.4 |
| cMaf | -213.1 | 802.3 | -489.3 | 2598.1 | 425.9 |
| MafB | -2845.6 | 1042.4 | -2180.1 | 1888.4 | 631.4 |
| NFE2 | 467.0 | -112.2 | -538.4 | 12839.1 | 1034.5 |
| NFE2L1 | 7.9 | 858.4 | -1276.0 | 1119.5 | 389.6 |
| NFE2L3 | -1970.3 | -2436.6 | -1654.3 | 569.1 | 188.9 |
| BACH1 | 394.6 | 4116.5 | 952.4 | 4105.9 | 4251.6 |
| anti-XBP1-2 | | | | | |
| anti-XBP1 | | | | | 455.1 |
| anti-BATF | | | | | |
| anti-SMAF | | | | 5382.6 | |
| anti-E4BP4-2 | | | | | |
| anti-E4BP4 | | | | | |
| anti-C/EBPγ | | | 1752.0 | | |
| anti-C/EBPγ-2 | | | | | |
| anti-ATF4 | | | | | |
| anti-ATF4-2 | | | | | |
| anti-BACH-2 | | | | | |
| anti-BACH | | | | | |
| anti-ATF2-3 | | | | | |
| anti-ZF-2 | | | | | |
| anti-CREB | | | | | |
| anti-C/EBP-2 | | | | | |
| anti-OASIS | 269.4 | | | | |
| anti-OASIS-2 | | -101.6 | | | |

$S_{array}$ values.

| protein | C/EBPα | C/EBPβ | C/EBPδ | C/EBPγ | CHOP | CREB | CREB3 | ATF-6 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 12.3 | 8.8 | 8.9 | 3.4 | 14.4 | -1.3 | -0.8 | -0.8 |
| C/EBPβ | 10.2 | 2.7 | 6.3 | 3.4 | 30.6 | -1.4 | -1.8 | -1.0 |
| C/EBPδ | 13.8 | 9.1 | 7.2 | 6.6 | 11.8 | -1.1 | 0.0 | -0.8 |
| C/EBPγ | 11.8 | 7.7 | 11.1 | 1.2 | 15.9 | -0.3 | 1.9 | -0.1 |
| CHOP | 17.3 | 26.9 | 10.9 | 13.8 | 3.7 | 0.0 | 4.8 | 0.0 |
| ATF-1 | -1.6 | -1.6 | -2.1 | -2.0 | -1.2 | 4.9 | -1.9 | -0.9 |
| CREB | -2.3 | -2.7 | -2.0 | -1.7 | -1.1 | 5.7 | -0.7 | -2.3 |
| CREB-H | -0.6 | -0.2 | -0.4 | -0.2 | -0.6 | 0.0 | 3.3 | 0.2 |
| CREB3 | -0.5 | -1.3 | -0.8 | -0.3 | -0.5 | -0.3 | 2.6 | -0.5 |
| ATF-6 | -1.1 | -0.9 | -0.9 | -1.1 | -1.0 | -0.8 | -0.5 | 5.1 |
| ZF | -0.7 | -0.7 | -1.0 | -0.8 | -0.1 | -1.3 | -0.7 | 0.6 |
| XBP-1 | -0.2 | 0.1 | -0.8 | -0.6 | -1.0 | 0.4 | 0.0 | 9.7 |
| E4BP4 | -0.6 | -1.0 | 0.3 | 0.3 | 0.0 | 4.8 | 3.1 | 0.2 |
| ATF-2 | -0.6 | -0.3 | -0.9 | 1.7 | 0.4 | -0.6 | -0.2 | -0.5 |
| ATF-7 | 3.1 | 3.4 | 1.4 | 4.5 | 4.0 | 0.0 | 0.7 | -0.1 |
| cJun | 1.0 | 1.0 | 0.9 | 0.4 | 0.5 | 0.4 | -0.3 | 0.0 |
| JunB | -0.3 | -0.6 | -0.5 | -1.0 | -0.7 | -0.2 | -1.0 | 0.1 |
| JunD | -0.2 | -0.6 | 0.1 | -0.3 | 0.0 | 0.2 | -0.6 | 0.0 |
| Fos | 2.8 | 0.5 | 0.6 | 0.4 | 1.8 | -0.5 | -0.6 | -0.5 |
| Fra2 | 0.5 | 0.1 | -0.1 | 0.0 | 0.4 | 0.0 | -0.8 | 0.1 |
| ATF-3 | 2.4 | 3.6 | 2.5 | 12.5 | 6.4 | -1.3 | 0.1 | -1.0 |
| ATF-4 | 37.0 | 12.9 | 21.0 | 44.2 | 21.3 | -1.0 | 0.7 | -1.2 |
| ATF-5 | 9.9 | 1.2 | 6.2 | 42.2 | -1.3 | -2.1 | -1.3 | -2.1 |
| B-ATF | 3.4 | 4.9 | 3.9 | 7.4 | 11.4 | 0.6 | 1.9 | 0.5 |
| p21SNFT | 5.2 | 3.7 | 5.9 | 4.2 | 18.4 | -0.1 | 5.2 | 0.0 |
| TEF | -1.4 | -1.6 | -1.4 | -1.3 | -0.8 | -1.7 | -1.3 | -1.9 |
| MafG | -0.8 | -0.5 | -0.2 | -0.5 | 0.1 | 0.8 | -0.7 | 0.5 |
| cMaf | -0.4 | 0.3 | -0.7 | -0.6 | -1.1 | -0.2 | -0.4 | 0.0 |
| MafB | -0.5 | -1.1 | -1.3 | -1.3 | -1.6 | -0.3 | -1.7 | 0.1 |
| NFE2 | -1.4 | -1.2 | -1.3 | -0.8 | -1.0 | -0.8 | -0.3 | -0.2 |
| NFE2L1 | -0.4 | -0.4 | 0.0 | -1.2 | -1.4 | 1.2 | -0.5 | -0.2 |
| NFE2L3 | -1.3 | -1.3 | -1.1 | -2.2 | -1.6 | -2.0 | -1.9 | -1.8 |
| BACH1 | -0.4 | -0.1 | -0.8 | -0.3 | -0.2 | 1.9 | 0.1 | 0.7 |
| anti-XBP1-2 | -0.3 | -0.3 | -0.7 | -0.3 | -0.8 | 0.1 | -0.5 | 1.8 |
| anti-XBP1 | 0.0 | -0.5 | -0.8 | -0.9 | -1.3 | 0.4 | -0.5 | 2.0 |
| anti-BATF | 0.6 | 2.8 | 0.4 | 6.3 | 18.4 | -0.1 | 1.4 | -0.6 |
| anti-SMAF | -0.8 | -0.3 | -0.6 | -0.1 | -0.9 | 1.1 | -0.5 | 0.1 |
| anti-E4BP4-2 | -0.2 | -0.2 | -0.4 | -0.4 | -0.8 | 0.8 | -0.4 | 0.9 |
| anti-E4BP4 | 0.1 | 0.0 | -0.7 | -0.3 | 1.1 | 0.8 | 0.2 | 0.6 |
| anti-C/EBPγ | 2.5 | 6.3 | 7.3 | 21.5 | 12.0 | 0.4 | 2.8 | 0.6 |
| anti-C/EBPγ-2 | 13.0 | 10.7 | 18.6 | 14.2 | 9.4 | 0.6 | 5.8 | 0.0 |
| anti-ATF4 | 8.8 | 2.1 | 10.1 | 1.7 | 1.9 | 1.5 | 18.8 | -0.1 |
| anti-ATF4-2 | 10.6 | 0.9 | 7.4 | 1.2 | 2.8 | -0.3 | 13.1 | -0.3 |
| anti-BACH-2 | 0.7 | -0.9 | -0.5 | -1.2 | -1.0 | -1.0 | -1.2 | -0.6 |
| anti-BACH | 3.0 | 0.8 | 0.0 | -0.1 | -1.0 | 1.6 | 0.9 | -0.1 |
| anti-ATF2-3 | -0.5 | -0.1 | 0.8 | 3.0 | 3.1 | 3.4 | 3.9 | 1.5 |
| anti-ZF-2 | -1.4 | -0.2 | -0.2 | -0.3 | -0.4 | 1.4 | 2.0 | 6.8 |
| anti-CREB | -1.7 | -0.8 | -1.2 | 0.7 | -0.5 | 3.5 | 1.6 | 0.5 |
| anti-C/EBP-2 | 29.9 | 18.5 | 17.1 | 26.3 | 18.6 | -0.3 | 1.6 | 0.2 |
| anti-OASIS | 1.6 | 5.8 | 4.7 | 5.5 | -0.8 | 0.8 | 31.2 | -0.1 |
| anti-OASIS-2 | 0.8 | 1.1 | 0.4 | 5.3 | 0.8 | 0.2 | 17.0 | 0.1 |

| protein | ZF | XBP-1 | E4BP4 | ATF-2 | cJun | Fos | ATF-3 | ATF-4 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -1.3 | -1.2 | -0.8 | -0.4 | -0.5 | 0.5 | -0.1 | 9.7 |
| C/EBPβ | -1.1 | -0.2 | -2.1 | -0.9 | -1.1 | -1.1 | 0.0 | 1.4 |
| C/EBPδ | -1.7 | -1.4 | -0.4 | -1.1 | -1.0 | -1.2 | 0.9 | 9.8 |
| C/EBPγ | -0.9 | -0.1 | 4.2 | 2.9 | -0.2 | -0.4 | 5.7 | 6.6 |
| CHOP | 0.7 | 0.2 | 9.6 | 3.3 | 1.1 | 1.1 | 6.0 | 2.9 |
| ATF-1 | -1.3 | 0.6 | 6.0 | -1.3 | -0.8 | -0.9 | -1.2 | -1.7 |
| CREB | -2.3 | -0.7 | 4.4 | -1.7 | -1.8 | -1.3 | -1.7 | -1.8 |
| CREB-H | -0.6 | -0.1 | 3.8 | -0.6 | -0.4 | -0.8 | -0.6 | -0.2 |
| CREB3 | -0.7 | -0.2 | 1.0 | -0.9 | -1.1 | -1.1 | -0.9 | -0.8 |
| ATF-6 | -0.6 | 4.8 | -0.1 | -1.1 | -0.9 | -0.9 | -1.1 | -0.7 |
| ZF | 4.7 | 12.9 | -0.4 | -0.3 | -0.2 | -0.7 | 0.0 | 2.1 |
| XBP-1 | 3.9 | 19.1 | -0.1 | -0.5 | -0.4 | -1.2 | -0.3 | 0.0 |
| E4BP4 | -0.5 | -0.2 | 44.7 | -0.8 | -1.1 | -1.0 | 0.2 | -1.4 |
| ATF-2 | -0.9 | 0.9 | -0.6 | 2.0 | 4.1 | 1.0 | 2.5 | -0.1 |
| ATF-7 | 0.0 | 0.0 | 0.0 | 3.3 | 8.1 | 2.1 | 4.6 | 2.4 |
| cJun | 0.2 | 0.1 | -0.8 | 6.5 | 2.4 | 11.8 | 5.9 | -0.3 |
| JunB | -0.1 | -0.3 | -1.3 | 1.9 | 0.2 | 7.6 | 4.7 | -1.5 |
| JunD | 0.2 | -0.4 | -1.6 | 3.7 | 0.3 | 9.1 | 5.4 | -0.5 |
| Fos | 0.5 | -1.0 | 0.0 | 3.7 | 29.1 | 0.2 | -0.2 | 2.5 |
| Fra2 | 0.6 | 0.4 | -1.3 | 3.3 | 17.1 | 0.7 | 0.6 | 0.8 |
| ATF-3 | 0.3 | 0.4 | 1.0 | 5.0 | 9.0 | 0.2 | -1.0 | 4.0 |
| ATF-4 | 3.8 | -1.0 | -0.9 | 0.3 | -1.0 | 1.9 | 2.3 | -0.8 |
| ATF-5 | -1.1 | -2.9 | -0.9 | -1.9 | -2.1 | -1.7 | -1.9 | -1.7 |
| B-ATF | 0.4 | -0.5 | 3.1 | 1.6 | 15.7 | -0.9 | 1.4 | 3.5 |
| p21SNFT | 0.9 | -0.1 | 2.6 | 3.3 | 12.0 | -0.5 | 4.8 | 3.8 |
| TEF | -1.3 | 0.4 | -0.8 | -1.3 | -1.4 | -1.4 | -1.2 | -1.5 |
| MafG | 0.2 | 1.1 | 0.2 | -0.3 | 1.0 | -0.6 | -0.7 | -0.2 |
| cMaf | -0.4 | 0.1 | 0.1 | 0.0 | -0.3 | -0.3 | -0.1 | 1.1 |
| MafB | 0.0 | 0.7 | -0.5 | -0.2 | -0.5 | 0.2 | 0.4 | -0.8 |
| NFE2 | -0.2 | -1.0 | -1.3 | -0.5 | -0.9 | -1.2 | -1.5 | -0.7 |
| NFE2L1 | 3.8 | -0.9 | -0.3 | 0.1 | -0.6 | -0.2 | -1.1 | 1.5 |
| NFE2L3 | -1.4 | -2.2 | -0.8 | -1.8 | -1.8 | -1.5 | -1.9 | -1.0 |
| BACH1 | 1.9 | 1.2 | 1.6 | 2.1 | 0.0 | 0.0 | -0.8 | 0.3 |
| anti-XBP1-2 | 1.9 | 4.2 | -0.3 | -0.6 | 0.0 | -0.7 | -0.4 | -0.3 |
| anti-XBP1 | 1.1 | 1.7 | -1.2 | -1.0 | 0.5 | -0.8 | -0.8 | -0.1 |
| anti-BATF | 4.1 | 0.6 | -0.9 | 0.4 | -0.9 | 4.9 | 0.4 | 0.7 |
| anti-SMAF | -0.2 | 0.1 | -0.3 | -0.2 | 4.0 | 0.2 | -0.3 | 0.1 |
| anti-E4BP4-2 | -0.5 | -0.2 | 1.4 | -0.4 | 0.0 | -0.9 | -0.2 | -0.7 |
| anti-E4BP4 | -0.4 | 0.1 | 5.7 | -0.6 | -0.1 | -0.5 | -0.4 | -0.4 |
| anti-C/EBPγ | 3.8 | 0.3 | 0.9 | 0.2 | 1.5 | 1.9 | 2.5 | 2.9 |
| anti-C/EBPγ-2 | 1.3 | 0.6 | 1.6 | 4.3 | 4.8 | 1.5 | 0.6 | 9.1 |
| anti-ATF4 | 3.4 | 1.2 | 3.4 | 10.6 | 9.0 | 6.7 | 4.9 | 11.0 |
| anti-ATF4-2 | 0.5 | -0.7 | -0.5 | 2.7 | 1.4 | 8.2 | 4.1 | 10.5 |
| anti-BACH-2 | -1.1 | -1.2 | -1.9 | -1.1 | -0.8 | -0.7 | -0.8 | -1.5 |
| anti-BACH | -0.7 | -0.2 | 0.6 | 0.2 | -0.5 | 5.8 | 0.1 | -0.2 |
| anti-ATF2-3 | 4.1 | 0.3 | 4.8 | 16.6 | 2.4 | 8.1 | 3.9 | 1.2 |
| anti-ZF-2 | 3.4 | 13.6 | 2.2 | 0.9 | 1.6 | 0.7 | -0.2 | 2.0 |
| anti-CREB | -0.8 | 0.1 | 2.0 | -1.2 | -0.2 | 0.0 | -1.0 | -0.9 |
| anti-C/EBP-2 | 2.3 | -0.4 | -0.9 | 1.1 | 0.0 | 3.4 | 4.0 | 12.7 |
| anti-OASIS | -0.6 | -0.2 | 9.3 | -0.8 | 0.5 | 1.0 | 2.9 | 0.1 |
| anti-OASIS-2 | 0.0 | -0.1 | 3.7 | 0.8 | 1.1 | 0.7 | 3.7 | -0.6 |

| protein | p21SNFT | TEF | MafG | cMaf | NFE2 | BACH1 | anti-XBP1-2 | anti-XBP1 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 0.0 | -0.6 | -0.5 | -0.3 | -0.3 | -1.5 | -1.6 | -2.3 |
| C/EBPβ | -0.4 | -1.1 | -2.3 | -0.2 | -1.7 | -1.3 | 0.2 | -0.7 |
| C/EBPδ | 0.7 | -0.7 | 0.0 | -3.1 | 0.0 | -1.8 | -0.7 | -0.3 |
| C/EBPγ | 0.3 | -0.1 | 0.0 | -0.1 | -0.4 | -0.8 | 0.4 | -0.2 |
| CHOP | 0.8 | 0.9 | 0.3 | 1.2 | -0.3 | 1.8 | -0.5 | -0.4 |
| ATF-1 | -1.5 | -1.1 | -0.2 | -0.7 | -1.3 | 0.3 | 2.3 | 1.4 |
| CREB | -1.5 | -2.2 | -1.9 | -1.1 | -2.2 | 0.7 | -0.7 | -0.5 |
| CREB-H | -0.6 | 0.5 | -0.2 | -0.3 | 0.1 | -1.0 | 0.8 | -0.1 |
| CREB3 | -0.6 | -0.8 | -0.5 | 0.0 | -0.1 | -1.1 | 0.1 | 0.0 |
| ATF-6 | -1.2 | -0.8 | -1.0 | 0.1 | -1.1 | -0.9 | 2.4 | 0.5 |
| ZF | -0.8 | -0.5 | 0.3 | -0.2 | 1.2 | 0.4 | 25.0 | 6.4 |
| XBP-1 | -1.0 | 0.3 | -0.4 | -0.9 | -0.8 | -0.7 | 71.8 | 13.4 |
| E4BP4 | -0.2 | 0.3 | -1.0 | -1.6 | 0.0 | -0.9 | 0.8 | -0.3 |
| ATF-2 | 0.7 | -0.4 | 0.5 | 0.5 | -0.7 | 1.8 | -2.0 | 1.4 |
| ATF-7 | 1.0 | 0.0 | -0.6 | 0.5 | 1.0 | 2.7 | -0.3 | -1.2 |
| cJun | 0.9 | 0.1 | 0.2 | 1.2 | -0.2 | -0.3 | -0.8 | 3.0 |
| JunB | 0.6 | -0.1 | -0.1 | -0.2 | -0.1 | -0.9 | -0.6 | 1.4 |
| JunD | 1.0 | -0.3 | -0.7 | 1.0 | 0.1 | -0.9 | 0.9 | -0.2 |
| Fos | -0.9 | 0.3 | -0.8 | 1.7 | -0.6 | 1.1 | -1.3 | 1.1 |
| Fra2 | -0.7 | 0.1 | -1.1 | 1.0 | -0.3 | 0.0 | 1.4 | 1.8 |
| ATF-3 | 1.1 | -0.4 | 0.8 | 0.3 | -2.4 | -1.1 | 1.8 | 1.5 |
| ATF-4 | 0.9 | -1.9 | 0.4 | 9.0 | -0.4 | 0.5 | -2.7 | 0.4 |
| ATF-5 | 0.1 | -2.3 | -0.1 | -0.8 | 0.2 | -0.6 | -1.7 | -1.4 |
| B-ATF | 0.4 | 2.4 | -0.7 | 0.2 | 0.6 | 1.8 | -0.4 | -0.6 |
| p21SNFT | 0.8 | 0.9 | 5.0 | 0.7 | -0.8 | 1.0 | 2.6 | 2.1 |
| TEF | -1.2 | 0.5 | -1.4 | -0.8 | -2.0 | -1.0 | -2.1 | -0.2 |
| MafG | 0.2 | 0.0 | 7.6 | 0.0 | 7.6 | 16.4 | 1.4 | 0.2 |
| cMaf | -0.8 | -0.1 | 0.4 | 20.0 | 0.0 | 2.9 | -0.3 | 0.0 |
| MafB | -1.1 | 0.1 | -0.4 | 22.8 | -0.3 | 5.2 | 1.7 | 0.8 |
| NFE2 | -1.2 | -0.1 | 1.7 | -0.5 | 8.2 | -0.7 | 1.5 | 0.8 |
| NFE2L1 | -0.9 | -0.3 | 144.0 | 1.2 | 0.7 | -0.9 | -0.1 | -1.1 |
| NFE2L3 | -1.8 | -1.9 | 103.1 | -1.0 | 5.3 | -1.6 | -1.2 | -0.9 |
| BACH1 | -0.5 | 0.5 | 51.6 | 7.0 | -0.6 | 0.3 | 71.5 | 8.3 |
| anti-XBP1-2 | -0.9 | 0.4 | 0.0 | -0.4 | 0.3 | 2.9 | -0.6 | |
| anti-XBP1 | -1.0 | 0.4 | -0.6 | -1.9 | 0.2 | 0.2 | | -1.1 |
| anti-BATF | 0.4 | 2.6 | 38.9 | -0.1 | -0.2 | -0.5 | | |
| anti-SMAF | 0.7 | 0.0 | 53.4 | 1.9 | 7.1 | 1.1 | | |
| anti-E4BP4-2 | -1.0 | 0.9 | -2.4 | -0.1 | 0.0 | -1.1 | | |
| anti-E4BP4 | -0.4 | 2.5 | 0.1 | 0.0 | 0.4 | -0.3 | | |
| anti-C/EBPγ | 0.7 | 0.3 | 1.4 | -1.4 | 0.3 | 0.6 | | |
| anti-C/EBPγ-2 | 0.7 | 2.8 | 1.8 | 27.2 | 2.8 | 1.3 | | |
| anti-ATF4 | 0.9 | 0.1 | 25.4 | 4.6 | 12.7 | 8.7 | | |
| anti-ATF4-2 | 0.9 | -0.5 | -0.4 | 1.1 | 4.7 | 1.1 | | |
| anti-BACH-2 | -1.5 | -0.7 | 0.2 | -0.5 | 0.2 | 0.0 | | |
| anti-BACH | -0.4 | 0.0 | -0.5 | 5.7 | 13.1 | 12.0 | | |
| anti-ATF2-3 | 0.7 | -0.1 | 8.9 | 0.0 | 0.1 | -0.7 | | |
| anti-ZF-2 | -0.2 | -0.5 | 5.8 | 0.7 | 1.7 | 3.7 | | |
| anti-CREB | 0.1 | 0.5 | 2.5 | -0.2 | -0.2 | 0.5 | | |
| anti-C/EBP-2 | 0.9 | 1.0 | -0.3 | 2.3 | 1.5 | -0.9 | | |
| anti-OASIS | 0.5 | -0.1 | 14.2 | -0.5 | 0.4 | -1.0 | | |
| anti-OASIS-2 | 0.4 | 0.0 | 8.2 | 0.7 | -0.4 | 0.4 | | |

| protein | anti-BATF | anti-SMAF | anti-E4BP4-2 | anti-E4BP4 | anti-C/EBPγ | anti-C/EBPγ-2 | anti-ATF4 | anti-ATF4-2 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 0.1 | -1.0 | -2.7 | -0.5 | 0.5 | 1.8 | -0.7 | 1.4 |
| C/EBPβ | -0.3 | -1.7 | -0.8 | -0.4 | 0.7 | 1.8 | -1.7 | -1.7 |
| C/EBPδ | -1.0 | -1.2 | -0.8 | -0.9 | 2.3 | 6.1 | -0.3 | 0.6 |
| C/EBPγ | 2.3 | -0.4 | 0.7 | 2.1 | 19.7 | 7.0 | -1.0 | -1.0 |
| CHOP | 27.0 | -0.2 | 3.8 | 33.8 | 17.9 | 5.0 | -0.8 | 1.1 |
| ATF-1 | -0.2 | -0.2 | 2.3 | -0.7 | -1.2 | -1.5 | -1.4 | -1.6 |
| CREB | -2.9 | -0.9 | -0.6 | -2.0 | -0.8 | -1.6 | -1.6 | -1.7 |
| CREB-H | 1.0 | -0.4 | 1.7 | 0.4 | -0.8 | -0.9 | -1.2 | -0.8 |
| CREB3 | -0.1 | -0.7 | 0.9 | -0.5 | -0.9 | -0.8 | 0.5 | 1.5 |
| ATF-6 | -0.8 | -0.8 | 0.0 | -0.1 | -1.3 | -1.2 | -1.4 | -1.3 |
| ZF | 2.0 | -0.2 | -0.8 | -0.1 | 3.6 | -0.3 | -0.1 | 0.1 |
| XBP-1 | 1.3 | -0.8 | 0.0 | -0.3 | -1.0 | -1.5 | -1.6 | -1.0 |
| E4BP4 | -0.1 | -1.2 | 1.9 | 4.2 | 0.0 | -1.1 | -1.4 | -1.7 |
| ATF-2 | -0.1 | -0.6 | 2.5 | 0.2 | -0.6 | 0.6 | 4.3 | 1.3 |
| ATF-7 | 0.5 | 1.3 | -1.5 | 0.5 | 0.1 | 1.9 | 1.9 | 3.3 |
| cJun | -0.5 | 20.1 | -1.4 | 0.8 | 0.6 | 2.2 | 2.0 | 1.0 |
| JunB | -0.4 | 6.0 | -0.5 | -0.2 | -0.5 | 0.6 | 1.8 | -1.1 |
| JunD | -0.6 | 6.7 | -1.8 | -0.2 | -0.3 | 0.7 | 2.3 | -0.1 |
| Fos | 7.1 | 3.9 | 0.1 | 0.6 | 6.8 | 1.1 | 5.3 | 16.9 |
| Fra2 | 9.4 | 1.2 | 1.6 | -0.5 | 3.5 | 3.4 | 3.6 | 4.6 |
| ATF-3 | 0.3 | 1.0 | 2.9 | 2.6 | 4.8 | -0.8 | 3.5 | 10.9 |
| ATF-4 | 0.1 | 0.3 | 0.9 | -0.3 | 6.9 | 6.6 | 7.5 | 18.4 |
| ATF-5 | -1.6 | -1.2 | -1.2 | -2.2 | 3.5 | -0.4 | -0.1 | 0.0 |
| B-ATF | 16.1 | 0.2 | 0.0 | 6.2 | 0.0 | -1.1 | 0.1 | 0.0 |
| p21SNFT | 18.1 | 9.0 | 2.0 | 7.5 | 1.4 | 0.6 | 0.5 | 3.0 |
| TEF | -1.8 | -1.1 | 0.6 | 2.1 | -0.7 | -1.2 | -1.7 | -1.9 |
| MafG | 7.4 | 47.2 | 0.2 | 2.1 | -0.9 | -0.6 | 1.8 | -0.5 |
| cMaf | 0.8 | 1.4 | -0.3 | 0.6 | -0.8 | 2.7 | -1.1 | -0.5 |
| MafB | 0.1 | 0.4 | 0.7 | 0.1 | -1.7 | 0.3 | -1.2 | -0.7 |
| NFE2 | -0.1 | 10.7 | -0.7 | -0.5 | -0.7 | -1.1 | 4.3 | 0.9 |
| NFE2L1 | -0.5 | 0.5 | -1.8 | -0.4 | -1.2 | -1.2 | 4.8 | 0.0 |
| NFE2L3 | -1.4 | -1.8 | -1.6 | -2.1 | -1.5 | -1.4 | 0.4 | -1.8 |
| BACH1 | 0.8 | 2.9 | 1.6 | 4.1 | 0.8 | -0.6 | 0.4 | -0.1 |
| anti-XBP1-2 | | | | | | | | |
| anti-XBP1 | | | | | | | | |
| anti-BATF | -0.9 | | | | | | | |
| anti-SMAF | | 4.7 | | | | | | |
| anti-E4BP4-2 | | | -0.5 | | | | | |
| anti-E4BP4 | | | | 10.5 | | | | |
| anti-C/EBPγ | | | | | 1.0 | | | |
| anti-C/EBPγ-2 | | | | | | 2.4 | | |
| anti-ATF4 | | | | | | | -0.5 | |
| anti-ATF4-2 | | | | | | | | 1.5 |
| anti-BACH-2 | | | | | | | | |
| anti-BACH | | | | | | | | |
| anti-ATF2-3 | | | | | | | | |
| anti-ZF-2 | | | | | | | | |
| anti-CREB | | | | | | | | |
| anti-C/EBP-2 | | | | | | | | |
| anti-OASIS | | | | | | | | |

| protein | anti-BACH-2 | anti-BACH | anti-ATF2-3 | anti-ZF-2 | anti-CREB | anti-C/EBP-2 |
|---|---|---|---|---|---|---|
| C/EBPα | 1.3 | 0.0 | -1.2 | -0.8 | -0.4 | 31.8 |
| C/EBPβ | -1.0 | -1.0 | -1.0 | -1.1 | -0.8 | 16.9 |
| C/EBPδ | 1.8 | -0.5 | -0.9 | -1.0 | -1.3 | 27.1 |
| C/EBPγ | 1.5 | 0.0 | 2.4 | -0.6 | 4.5 | 38.0 |
| CHOP | -0.4 | 0.2 | 5.2 | -0.1 | 3.4 | 43.7 |
| ATF-1 | -1.3 | -0.1 | -0.7 | -0.6 | 3.6 | -1.0 |
| CREB | -2.1 | -0.8 | 0.1 | -1.7 | 3.2 | -0.7 |
| CREB-H | 0.2 | 0.2 | -0.6 | -0.2 | 2.2 | -0.1 |
| CREB3 | -0.2 | -1.0 | -0.1 | -0.4 | 1.8 | -1.1 |
| ATF-6 | -0.6 | -1.1 | -0.9 | 0.6 | -1.0 | -0.3 |
| ZF | -1.0 | -1.4 | 3.7 | 2.4 | -0.8 | 2.6 |
| XBP-1 | -0.3 | -1.7 | -0.8 | 4.8 | -0.2 | -1.4 |
| E4BP4 | 0.0 | -0.5 | 0.3 | 0.0 | 0.0 | -2.0 |
| ATF-2 | -0.4 | 0.9 | 6.4 | 0.0 | -0.7 | 0.1 |
| ATF-7 | 0.1 | -0.3 | 7.1 | 0.6 | 0.5 | 6.3 |
| cJun | 0.3 | -0.9 | 2.9 | 1.0 | 1.7 | 0.4 |
| JunB | 0.0 | -1.0 | 1.0 | 0.1 | 0.0 | -1.1 |
| JunD | 0.3 | -0.9 | 0.5 | 0.3 | -1.0 | -0.4 |
| Fos | 1.5 | 17.9 | 6.5 | 0.5 | 6.9 | 8.1 |
| Fra2 | 0.2 | 8.8 | 7.0 | 0.7 | 4.6 | -0.5 |
| ATF-3 | -1.7 | 1.6 | 6.6 | -0.6 | 0.5 | 6.6 |
| ATF-4 | 0.3 | 2.2 | -0.7 | 0.7 | -0.6 | 51.6 |
| ATF-5 | -0.7 | -0.5 | -1.5 | -2.3 | -0.7 | 6.0 |
| B-ATF | 0.8 | 2.7 | 3.3 | -0.1 | 1.3 | 6.2 |
| p21SNFT | -0.4 | 2.3 | 6.3 | 0.8 | 8.9 | 19.5 |
| TEF | -1.6 | -1.9 | -1.2 | -1.4 | -1.6 | -1.1 |
| MafG | 2.0 | 0.0 | 3.8 | 0.3 | 4.8 | -0.2 |
| cMaf | 0.0 | 2.1 | -0.7 | 0.0 | -0.7 | 0.8 |
| MafB | -0.2 | 2.1 | -1.1 | -0.1 | -1.6 | -2.2 |
| NFE2 | 4.4 | 23.0 | -1.3 | 0.7 | -0.9 | -0.8 |
| NFE2L1 | 2.8 | 1.5 | -1.0 | 0.5 | -0.4 | 3.0 |
| NFE2L3 | -0.8 | -0.4 | -1.6 | -1.3 | -1.6 | -1.4 |
| BACH1 | 2.0 | 11.3 | -0.9 | 1.1 | 6.0 | -0.8 |
| anti-XBP1-2 | | | | | | |
| anti-XBP1 | | | | | | |
| anti-BATF | | | | | | |
| anti-SMAF | | | | | | |
| anti-E4BP4-2 | | | | | | |
| anti-E4BP4 | | | | | | |
| anti-C/EBPγ | | | | | | |
| anti-C/EBPγ-2 | | | | | | |
| anti-ATF4 | | | | | | |
| anti-ATF4-2 | | | | | | |
| anti-BACH-2 | 9.4 | | | | | |
| anti-BACH | | 5.7 | | | | |
| anti-ATF2-3 | | | 1.9 | | | |
| anti-ZF-2 | | | | 1.0 | | |
| anti-CREB | | | | | 20.1 | |
| anti-C/EBP-2 | | | | | | -1.0 |
| anti-OASIS | | | | | | |

| protein | anti-OASIS | anti-OASIS-2 | anti-C/EBPγ(2) | anti-SMAF(2) | anti-XBP1800 |
|---|---|---|---|---|---|
| C/EBPα | -1.0 | -0.9 | 0.5 | -1.0 | -1.9 |
| C/EBPβ | 0.5 | -1.2 | 0.0 | -1.5 | -1.3 |
| C/EBPδ | -0.3 | -1.3 | 1.9 | -1.6 | -1.8 |
| C/EBPγ | 4.4 | 2.8 | 17.4 | 0.8 | -1.0 |
| CHOP | 1.1 | 5.3 | 14.9 | -0.5 | -1.0 |
| ATF-1 | -1.7 | -0.2 | -1.0 | -0.4 | 1.6 |
| CREB | -0.5 | -1.0 | -0.9 | -1.1 | 1.2 |
| CREB-H | 1.6 | 2.2 | -0.8 | -0.2 | -0.6 |
| CREB3 | 5.5 | 5.0 | -1.0 | -1.7 | -0.7 |
| ATF-6 | -0.4 | -0.7 | -0.9 | -1.3 | 5.0 |
| ZF | -0.7 | -0.1 | 3.0 | -0.1 | 25.4 |
| XBP-1 | 0.6 | 1.1 | -0.6 | -0.7 | 42.6 |
| E4BP4 | 0.5 | -0.2 | -1.4 | -1.3 | -0.6 |
| ATF-2 | -0.6 | -0.4 | -0.6 | 0.1 | 1.1 |
| ATF-7 | -1.0 | 1.1 | 0.2 | 1.0 | -0.9 |
| cJun | 0.6 | 1.4 | 0.2 | 24.8 | 17.6 |
| JunB | -0.2 | 1.2 | -0.8 | 7.1 | 8.7 |
| JunD | -0.1 | 0.4 | 0.0 | 6.7 | 4.4 |
| Fos | 2.6 | 9.1 | 5.0 | 4.0 | 3.2 |
| Fra2 | 0.3 | 6.6 | 2.3 | 0.1 | 4.9 |
| ATF-3 | 4.4 | 9.5 | 4.0 | 0.6 | 0.2 |
| ATF-4 | 0.5 | -0.1 | 4.5 | -0.1 | -1.5 |
| ATF-5 | -1.0 | -1.5 | 2.1 | -0.4 | -3.1 |
| B-ATF | 1.1 | 1.9 | 0.2 | 0.3 | -0.3 |
| p21SNFT | 7.6 | 5.3 | 0.9 | 11.4 | 6.2 |
| TEF | -1.0 | -1.9 | -0.9 | -0.5 | -1.1 |
| MafG | 4.8 | 5.6 | -0.4 | 47.8 | 1.4 |
| cMaf | -0.4 | -0.1 | -0.6 | 1.4 | -0.5 |
| MafB | -2.4 | 0.1 | -1.8 | 0.6 | 1.1 |
| NFE2 | 0.1 | -0.6 | -0.6 | 13.2 | 4.1 |
| NFE2L1 | -0.2 | -0.1 | -1.1 | -0.3 | -0.7 |
| NFE2L3 | -1.7 | -1.9 | -1.4 | -0.9 | -2.2 |
| BACH1 | 0.0 | 1.7 | 0.5 | 3.1 | 28.1 |
| anti-XBP1-2 | | | | | |
| anti-XBP1 | | | | | -0.2 |
| anti-BATF | | | | | |
| anti-SMAF | | | | 4.6 | |
| anti-E4BP4-2 | | | | | |
| anti-E4BP4 | | | | | |
| anti-C/EBPγ | | | 1.1 | | |
| anti-C/EBPγ-2 | | | | | |
| anti-ATF4 | | | | | |
| anti-ATF4-2 | | | | | |
| anti-BACH-2 | | | | | |
| anti-BACH | | | | | |
| anti-ATF2-3 | | | | | |
| anti-ZF-2 | | | | | |
| anti-CREB | | | | | |
| anti-C/EBP-2 | | | | | |
| anti-OASIS | 0.0 | | | | |
| anti-OASIS-2 | | -0.6 | | | |

Table B.S5. Average background-corrected fluorescence values and $S_{array}$ values from round 3 of array measurements. Peptides on the surface are in rows, those in solution in columns. Duplicate measurements are marked with a number two in parentheses.

| protein | C/EBPα | C/EBPδ | C/EBPγ | CHOP | CREB | CREB3 | ATF-6 | ZF |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 13209.9 | 14793.7 | 5578.6 | 20538.1 | -3572.9 | -348.5 | -1109.8 | -3046.3 |
| C/EBPβ | 13655.8 | 12534.1 | 8038.8 | 43556.6 | -5424.4 | -835.6 | -2331.6 | -934.4 |
| C/EBPδ | 16058.1 | 12662.8 | 11335.4 | 19225.8 | -3982.1 | 127.2 | -1355.0 | -3096.3 |
| C/EBPγ | 12621.5 | 18023.1 | 3251.3 | 18315.5 | -547.1 | 3900.8 | -104.8 | -729.8 |
| CHOP | 18136.8 | 17304.7 | 12237.4 | 6192.9 | 998.0 | 5479.4 | 194.9 | 2076.3 |
| ATF-1 | -785.0 | -2284.6 | -380.3 | -1290.3 | 19529.9 | -1428.0 | -1519.2 | -1797.7 |
| CREB | -2492.4 | -3500.0 | -1550.3 | -1430.9 | 24165.9 | 283.8 | -4971.3 | -5890.9 |
| CREB-H | -428.6 | -70.8 | 622.0 | -337.6 | 1519.8 | 4298.0 | 632.3 | 189.1 |
| CREB3 | -49.9 | 421.8 | 358.1 | -1656.3 | -774.7 | 4662.0 | -1059.0 | 656.1 |
| ATF-6 | -522.0 | -1049.5 | -572.3 | -907.7 | -1652.0 | -889.6 | 15606.4 | 632.4 |
| ZF | -1218.1 | -1837.4 | -235.3 | 163.5 | -3134.9 | -928.9 | 1320.9 | 6024.5 |
| XBP-1 | 417.8 | -781.4 | 221.0 | -727.7 | 1185.1 | -101.2 | 23319.6 | 13396.5 |
| E4BP4 | -36.6 | 1346.3 | 1344.6 | -108.0 | 16860.4 | 3140.1 | 669.8 | 1191.6 |
| ATF-2 | 382.4 | -603.7 | 4215.1 | 1217.9 | -1842.3 | -465.7 | -477.2 | 103.3 |
| ATF-7 | 4210.5 | 3216.8 | 7322.9 | 7111.1 | -72.6 | 829.5 | 4.6 | 2472.9 |
| cJun | 1272.0 | 1988.3 | 1500.9 | 1819.6 | 1703.1 | 105.6 | 178.6 | 750.7 |
| JunB | -30.3 | 97.2 | 155.3 | -296.2 | -207.0 | -676.0 | 459.6 | 1210.4 |
| JunD | 696.9 | 1078.4 | 714.2 | 388.4 | 1006.4 | -587.1 | 163.5 | 1325.3 |
| Fos | 3548.1 | 1704.4 | 2049.2 | 4124.7 | -819.8 | -32.9 | 282.4 | 1736.5 |
| Fra2 | 1460.7 | 615.3 | 1240.8 | 1409.2 | 401.1 | -892.9 | 622.8 | 2158.4 |
| ATF-3 | 4066.0 | 4646.3 | 15836.8 | 8668.4 | -3591.8 | 644.9 | -1484.1 | -274.0 |
| ATF-4 | 35862.3 | 33317.2 | 47227.6 | 34394.5 | -3495.4 | 1130.4 | -2093.1 | 17038.3 |
| ATF-5 | 12591.1 | 10962.6 | 46411.8 | -1321.4 | -6165.7 | -2218.4 | -4080.1 | -3706.3 |
| B-ATF | 4399.9 | 7027.9 | 10870.6 | 18098.6 | 2221.4 | 2149.7 | 1758.4 | 2508.6 |
| p21SNFT | 5924.9 | 10041.2 | 6911.2 | 24887.8 | -188.7 | 5459.4 | 868.9 | 2257.5 |
| HLF | 74.8 | 595.2 | 638.8 | 4667.8 | -171.4 | -455.1 | -495.7 | 630.0 |
| MafG | 476.8 | 1097.2 | 1218.4 | 968.3 | 2225.0 | 536.3 | 1270.0 | 3135.9 |
| cMaf | -296.7 | -885.1 | -0.5 | -791.0 | -1619.1 | -863.1 | 341.9 | 595.3 |
| MafB | -161.8 | -1295.9 | -127.6 | -1323.4 | -765.9 | -1335.6 | 513.1 | 2527.3 |
| NFE2 | -385.6 | -1858.9 | -141.2 | -1908.1 | -2632.1 | -499.7 | -371.4 | -1048.1 |
| NFE2L1 | -132.1 | 201.7 | -588.4 | -1192.3 | 2935.9 | -20.9 | -967.8 | 8911.4 |
| NFE2L3 | -1097.4 | -1365.4 | -1335.9 | -1704.3 | -6599.8 | -2039.2 | -3744.5 | -2930.1 |
| BACH1 | -342.9 | -978.6 | 528.6 | 601.6 | 7603.6 | 486.9 | 2064.3 | 3666.3 |
| anti-CREB-2 | 3426.3 | 565.2 | 1771.2 | -181.0 | 9291.7 | 2176.9 | 349.3 | 166.9 |
| anti-CREB-3 | -23.7 | 5.9 | 713.6 | -1002.7 | 10743.3 | 3594.4 | 1934.2 | 17.1 |
| anti-BACH-3 | 655.3 | -18.3 | 673.1 | -959.1 | 3107.1 | 508.0 | 325.8 | 487.3 |
| anti-E4BP4-3 | -126.6 | 407.3 | 1406.9 | -109.4 | 3974.3 | 4394.6 | 2146.3 | 82.7 |
| anti-C/EBP | 20584.4 | 21468.8 | 21568.3 | 20689.9 | -1976.8 | 2616.1 | -301.9 | 142.7 |
| anti-C/EBP-3 | 10762.1 | 10968.4 | 4878.9 | 15697.3 | 658.4 | -316.8 | 919.1 | 192.8 |
| anti-NFE2-2 | -858.4 | -1002.6 | 314.7 | -1011.6 | 586.3 | -1253.9 | 1557.9 | -379.9 |
| anti-NFE2-3 | -379.0 | -1652.0 | -1036.6 | -2910.3 | -666.1 | -1247.1 | -131.9 | -561.9 |
| anti-OASIS-3 | 5571.9 | 12381.3 | 5384.7 | -373.9 | 3533.3 | 19185.3 | 2236.6 | 1671.3 |
| anti-OASIS-4 | 2210.6 | 4412.1 | 4143.6 | 2217.6 | 5246.9 | 15999.0 | 3511.4 | 1592.1 |
| anti-ZF | 364.5 | -1154.4 | 432.4 | -2419.6 | 47.3 | -785.5 | 2591.8 | 20271.9 |
| anti-ATF3-3 | 7028.8 | 6925.1 | 11063.5 | 4067.0 | 2507.4 | 2810.0 | 3335.0 | 8056.2 |
| anti-ATF2 | 546.8 | 1840.7 | 8594.2 | 9824.8 | 18551.5 | 2679.1 | 2197.3 | 4015.5 |
| anti-ATF2-4 | -199.1 | 20.0 | 8047.6 | 855.6 | 1292.8 | -566.8 | 494.0 | 326.2 |
| anti-CHOP | 13367.8 | 10774.3 | 18768.4 | 18869.4 | 911.3 | 1352.7 | 278.9 | 1138.9 |
| anti-ATF6 | 334.4 | -292.1 | 384.5 | -406.9 | 1419.0 | -530.4 | 14359.4 | 2530.7 |
| anti-LMAF-2 | 5432.4 | 1871.6 | 2815.4 | 2752.5 | 8000.9 | 2074.6 | -306.9 | 1250.2 |
| anti-LMAF-3 | -54.1 | -561.8 | 307.2 | -773.5 | 3296.1 | 138.3 | 1234.8 | 355.9 |
| anti-LMAF | 140.0 | 266.9 | 70.1 | -413.1 | 2122.3 | -70.0 | 1108.4 | 398.9 |
| anti-PAR | -311.4 | -792.9 | 646.6 | -212.9 | 657.4 | -686.9 | 906.9 | 1589.5 |
| anti-BATF-2 | 3976.1 | 3049.8 | 3833.1 | 1992.4 | 1257.1 | 8034.4 | -925.1 | -1142.6 |
| anti-BATF-3 | 5996.0 | 6996.4 | 5492.1 | 5901.9 | 9897.6 | 14557.3 | -249.8 | 1767.6 |

| protein | XBP-1 | E4BP4 | ATF-2 | ATF-7 | cJun | Fos | ATF-3 | ATF-4 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 129.1 | -162.4 | -1032.3 | 1516.3 | 111.4 | 377.0 | 1029.7 | 14649.7 |
| C/EBPβ | 1921.3 | -283.2 | -1335.3 | 2728.3 | -258.2 | -499.3 | 2757.1 | 6848.3 |
| C/EBPδ | -60.3 | -576.0 | -1908.1 | -899.8 | -47.8 | -583.3 | 3976.4 | 15766.3 |
| C/EBPγ | 1940.2 | 3345.4 | 5636.9 | 9823.9 | 1301.3 | 2071.9 | 17972.9 | 32061.8 |
| CHOP | 2038.8 | 8968.4 | 6154.9 | 13037.4 | 3220.8 | 7857.4 | 21974.4 | 6802.5 |
| ATF-1 | 3430.6 | 5915.8 | -1325.4 | -1590.2 | -275.8 | -786.3 | -4065.2 | -987.7 |
| CREB | 934.3 | 5504.4 | -3179.0 | -1697.4 | -1443.7 | -289.8 | -2972.5 | -2644.9 |
| CREB-H | 2412.2 | 3255.1 | -525.4 | -411.6 | 99.6 | -398.8 | -791.0 | 1050.1 |
| CREB3 | 2225.7 | 1811.0 | -1821.5 | -1421.9 | -351.6 | -989.3 | -1796.9 | 52.2 |
| ATF-6 | 10367.0 | 147.6 | -1838.1 | -923.7 | -803.8 | -1048.4 | -2662.3 | -727.4 |
| ZF | 13890.1 | -316.2 | -286.3 | 396.6 | 1033.8 | 307.9 | -50.1 | 6677.3 |
| XBP-1 | 23973.1 | 768.0 | -1654.5 | -695.7 | 637.0 | -543.3 | -725.4 | 296.8 |
| E4BP4 | 2769.9 | 35244.4 | -1527.5 | -699.9 | -719.6 | -555.3 | -1720.9 | -375.8 |
| ATF-2 | 3979.3 | -407.3 | 4629.9 | 9622.1 | 6359.8 | 6199.7 | 10201.1 | 1436.1 |
| ATF-7 | 2308.2 | 156.3 | 9273.3 | 13070.9 | 12902.2 | 10117.0 | 16814.5 | 5917.4 |
| cJun | 2520.0 | 38.9 | 14318.4 | 23433.9 | 4173.0 | 35885.9 | 24256.0 | 548.1 |
| JunB | 2247.8 | -605.8 | 5201.8 | 11126.3 | 996.7 | 20803.4 | 16463.1 | 106.5 |
| JunD | 1735.1 | -276.8 | 7661.6 | 17368.8 | 1622.3 | 27824.8 | 21844.3 | 333.4 |
| Fos | 586.4 | 59.1 | 7564.2 | 12781.3 | 33395.6 | 3086.9 | 1165.1 | 7057.3 |
| Fra2 | 2269.0 | -201.4 | 6752.6 | 11611.6 | 20026.5 | 4411.1 | 5184.6 | 3243.9 |
| ATF-3 | 911.3 | 332.7 | 14559.4 | 18778.8 | 12797.8 | 3479.4 | -1830.5 | 9263.3 |
| ATF-4 | 676.3 | -217.9 | 2658.9 | 10813.5 | 111.3 | 11242.1 | 9813.9 | -762.4 |
| ATF-5 | -2361.8 | -612.6 | -2975.8 | -1683.4 | -1713.1 | -2562.1 | -3205.2 | -2369.6 |
| B-ATF | 2099.9 | 2208.6 | 4180.8 | 4389.8 | 19465.9 | 543.7 | 6528.3 | 6194.4 |
| p21SNFT | 1949.1 | 2566.6 | 7246.0 | 9490.8 | 14414.6 | 708.8 | 15836.9 | 7807.6 |
| HLF | 3818.4 | 1646.7 | -1209.8 | 47.9 | 876.7 | 11.8 | -797.8 | 1519.9 |
| MafG | 4014.9 | 1596.3 | -182.9 | 2280.9 | 2626.5 | 555.7 | 466.8 | 1875.8 |
| cMaf | 2571.5 | -312.8 | -563.6 | -658.4 | 497.4 | 738.3 | -332.1 | 3776.1 |
| MafB | 2993.8 | -202.3 | -122.6 | -1806.8 | 580.4 | 3470.3 | 99.9 | -87.4 |
| NFE2 | 759.2 | -358.1 | -1662.3 | -1152.5 | -420.2 | -959.6 | -2428.0 | 654.8 |
| NFE2L1 | 990.2 | -607.9 | 158.0 | 796.4 | -328.9 | 1389.7 | -1961.3 | 4173.1 |
| NFE2L3 | -1561.1 | -868.3 | -3042.9 | -2222.9 | -1553.4 | -2322.6 | -4125.4 | 110.8 |
| BACH1 | 3163.0 | 1691.3 | 4289.5 | 5348.3 | 780.0 | 2171.4 | -1058.2 | 2162.1 |
| anti-CREB-2 | 2697.4 | 1096.6 | -424.2 | 369.3 | 1298.6 | 3561.6 | 1454.3 | 456.3 |
| anti-CREB-3 | 2107.1 | 890.6 | -411.8 | -338.5 | 4204.9 | 39553.5 | 5507.6 | 1247.6 |
| anti-BACH-3 | 3011.6 | 345.1 | 135.9 | -574.3 | -152.1 | 12210.6 | 1437.3 | 1599.3 |
| anti-E4BP4-3 | 2913.9 | 2975.4 | -821.6 | -882.1 | 667.8 | 157.0 | 785.3 | 881.2 |
| anti-C/EBP | 2680.1 | 2505.1 | 3523.8 | 9706.1 | 2582.1 | -167.0 | 890.1 | 4579.3 |
| anti-C/EBP-3 | 1939.4 | 106.0 | 498.1 | 5439.1 | 2067.0 | 5125.0 | 6282.2 | 4427.3 |
| anti-NFE2-2 | 3691.9 | -800.8 | -873.4 | -854.8 | 6.5 | 2470.0 | -548.5 | 895.7 |
| anti-NFE2-3 | 1999.2 | -319.8 | -2642.0 | -2155.8 | -321.6 | 1304.5 | -3494.1 | 2187.1 |
| anti-OASIS-3 | 1480.3 | 9622.2 | 2064.1 | 4018.5 | 7855.2 | 7248.3 | 8315.4 | 3467.0 |
| anti-OASIS-4 | 2889.8 | 5416.1 | -631.7 | 787.9 | 1174.3 | 2349.5 | 7525.4 | 6100.8 |
| anti-ZF | 2833.9 | -148.1 | -1599.1 | -1240.6 | 2008.1 | 779.7 | -2762.3 | 1801.9 |
| anti-ATF3-3 | 3412.1 | 486.1 | 3243.2 | 4351.4 | 9893.5 | 22313.4 | 13766.9 | 16658.9 |
| anti-ATF2 | 1880.0 | 186.9 | 35112.0 | 32055.9 | 6294.3 | 3430.3 | 4126.7 | 4296.3 |
| anti-ATF2-4 | 2497.2 | -639.6 | 3705.6 | 15048.9 | 3613.6 | 10982.3 | 10693.9 | 1821.9 |
| anti-CHOP | 3383.2 | 4315.3 | 6226.3 | 9176.3 | 3639.3 | 14943.8 | 16488.3 | 12488.0 |
| anti-ATF6 | 4569.6 | 97.4 | -870.7 | -627.5 | 411.1 | -431.6 | -265.1 | 809.5 |
| anti-LMAF-2 | 2022.6 | 224.6 | 118.0 | 257.3 | 382.1 | 29369.4 | 2351.1 | 2312.7 |
| anti-LMAF-3 | 2795.6 | 1805.9 | 1868.6 | 976.4 | 858.2 | 18032.8 | 1970.9 | 1227.8 |
| anti-LMAF | 1864.8 | -427.8 | 111.4 | -590.3 | -90.1 | 7380.7 | 103.3 | 1686.0 |
| anti-PAR | 2415.1 | -292.3 | -1249.0 | 761.1 | 2200.0 | 2456.1 | 6404.6 | 2066.4 |
| anti-BATF-2 | 5106.8 | 309.9 | 2781.6 | 4860.6 | 4138.9 | 12061.0 | 2702.1 | 5525.0 |
| anti-BATF-3 | 7117.7 | 8465.3 | 8727.6 | 7152.7 | 5984.1 | 15295.8 | 5413.6 | 8966.4 |

| protein | p21SNFT | HLF | MafG | cMaf | NFE2 | BACH1 | anti-CREB-2 | anti-CREB-3 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 5212.8 | -256.6 | -375.6 | -1249.1 | -1939.3 | -1975.6 | -100.6 | -639.7 |
| C/EBPβ | 6411.1 | 434.1 | -871.7 | -1005.7 | -6727.1 | -2369.6 | 47.6 | -197.4 |
| C/EBPδ | 10373.5 | 1443.1 | -667.3 | -4241.4 | -1143.3 | -2251.8 | -323.1 | -316.8 |
| C/EBPγ | 11231.8 | 3209.3 | 247.4 | -673.4 | -262.3 | -350.6 | 698.2 | 34.8 |
| CHOP | 31993.1 | 36937.2 | 592.2 | -237.7 | 725.1 | 3533.2 | 1177.2 | 127.9 |
| ATF-1 | -3789.8 | -1021.8 | -267.1 | -1377.4 | -4343.3 | 1982.5 | 544.4 | 226.9 |
| CREB | -4636.8 | -4761.1 | -655.8 | -1624.0 | -11555.7 | 1839.4 | 584.4 | 121.3 |
| CREB-H | 1012.3 | 420.5 | -275.0 | -422.8 | 118.9 | -818.6 | 355.6 | 263.2 |
| CREB3 | -252.7 | -979.5 | -140.0 | -698.0 | 230.1 | -1547.2 | 337.7 | -18.7 |
| ATF-6 | -2853.1 | -370.9 | -258.3 | -673.0 | -2123.3 | -913.5 | 15.4 | -358.4 |
| ZF | 2208.1 | -995.4 | 240.8 | -372.1 | 2886.3 | 780.3 | 139.8 | -231.8 |
| XBP-1 | -2151.3 | 60.1 | -228.0 | -1510.3 | -1632.3 | -163.9 | 353.7 | -570.3 |
| E4BP4 | 2472.8 | 967.2 | -362.7 | -1237.8 | 222.3 | -528.4 | -167.1 | -321.1 |
| ATF-2 | 7411.9 | 737.2 | -593.6 | -514.6 | -1567.8 | 3863.3 | -146.4 | -349.9 |
| ATF-7 | 11721.5 | 1508.9 | 563.6 | -38.8 | 3279.4 | 5888.4 | 270.3 | -161.8 |
| cJun | 24109.8 | 5642.1 | 30.5 | -95.1 | -1656.6 | -733.8 | 863.9 | 1241.2 |
| JunB | 25629.1 | 2137.8 | -247.3 | -1268.8 | -1216.0 | -1396.6 | 228.3 | 623.3 |
| JunD | 23364.3 | 2556.4 | -169.9 | 644.7 | 189.4 | -444.9 | 253.8 | 404.9 |
| Fos | -60.8 | 635.2 | -350.8 | 1058.9 | -1141.3 | 2665.6 | 2360.6 | 34629.8 |
| Fra2 | -1136.4 | 1150.3 | -221.8 | 984.3 | -382.9 | 898.6 | 1440.0 | 18814.9 |
| ATF-3 | 22942.9 | 1429.4 | -44.7 | -56.6 | -5603.0 | -660.1 | 867.3 | 607.8 |
| ATF-4 | 23444.8 | 8053.4 | -304.1 | 8585.9 | -75.6 | 660.1 | 38.1 | -327.7 |
| ATF-5 | 5509.9 | 603.0 | -1266.3 | -1964.4 | 1465.5 | -877.6 | -65.1 | -552.5 |
| B-ATF | 6636.4 | 14240.4 | -234.8 | -487.6 | 1713.6 | 4512.3 | 330.9 | -139.5 |
| p21SNFT | 8758.2 | 14211.5 | 1927.7 | -407.8 | -1488.6 | 3238.3 | 2173.4 | 279.4 |
| HLF | 3625.3 | 11523.4 | -213.9 | -876.1 | -1612.4 | -144.8 | 134.3 | -419.3 |
| MafG | 6846.1 | 589.3 | 5846.6 | -51.7 | 21970.8 | 29119.6 | 1777.0 | 782.9 |
| cMaf | -1601.7 | -130.6 | -321.6 | 16013.9 | -196.1 | 4326.8 | 92.4 | 1256.2 |
| MafB | -2949.5 | -622.6 | -269.8 | 21421.3 | 1380.3 | 11118.8 | 78.8 | 1530.8 |
| NFE2 | -3446.4 | -1088.9 | 1234.1 | -1039.0 | 19406.9 | -1453.3 | 414.6 | 586.6 |
| NFE2L1 | -2215.3 | -711.1 | 38564.2 | 811.8 | 2517.2 | -909.4 | 199.8 | 543.4 |
| NFE2L3 | -5146.1 | -3135.1 | 43490.9 | -1371.6 | 22893.8 | -3604.4 | -190.6 | -139.3 |
| BACH1 | 2420.9 | 2250.1 | 16170.7 | 3892.6 | -2251.5 | 2300.1 | 2436.3 | 4184.1 |
| anti-CREB-2 | 6702.6 | 53.0 | 1375.8 | -375.8 | 3324.8 | 4956.1 | 1282.3 | |
| anti-CREB-3 | 1344.0 | -218.8 | 1092.4 | 9720.4 | 8723.1 | 24989.0 | | 851.9 |
| anti-BACH-3 | 2733.1 | -486.4 | -412.9 | 1190.9 | 27545.9 | 10812.2 | | |
| anti-E4BP4-3 | -1046.2 | 1802.5 | -627.3 | -881.8 | 1405.9 | 452.1 | | |
| anti-C/EBP | 5508.0 | 13048.5 | -467.4 | -601.2 | 431.2 | -1633.0 | | |
| anti-C/EBP-3 | 5786.6 | 11024.0 | -593.9 | -1586.3 | 2412.1 | -238.1 | | |
| anti-NFE2-2 | -2332.3 | -63.0 | -604.1 | 2531.4 | 1582.0 | 597.3 | | |
| anti-NFE2-3 | -3802.4 | -1928.9 | -510.9 | -1200.9 | 10107.6 | -234.4 | | |
| anti-OASIS-3 | 2813.4 | 434.5 | 11150.1 | 169.4 | 6655.1 | 4290.9 | | |
| anti-OASIS-4 | 7236.8 | 632.5 | 4753.7 | -560.3 | 5339.3 | -376.8 | | |
| anti-ZF | -2180.6 | -1486.9 | -89.6 | -1228.1 | 29940.8 | 766.8 | | |
| anti-ATF3-3 | 7977.9 | 1679.0 | 17906.1 | 1766.5 | 24377.7 | 8858.4 | | |
| anti-ATF2 | 27165.6 | 7673.7 | -52.3 | 508.0 | 965.2 | 172.4 | | |
| anti-ATF2-4 | 13542.9 | -483.7 | -113.1 | -362.5 | 1589.6 | -873.4 | | |
| anti-CHOP | 34189.9 | 19272.0 | 26.8 | 385.8 | 4325.9 | -262.6 | | |
| anti-ATF6 | -1630.1 | -306.2 | -172.7 | -373.0 | 1762.9 | 174.2 | | |
| anti-LMAF-2 | 20609.8 | 7392.1 | -355.6 | 6390.9 | 7589.3 | 5216.8 | | |
| anti-LMAF-3 | 324.8 | 1829.8 | 9261.1 | 29157.9 | 7536.3 | 6535.4 | | |
| anti-LMAF | 624.9 | -373.6 | 2182.6 | 13285.5 | 4396.9 | 4952.5 | | |
| anti-PAR | -2122.5 | 364.9 | -907.8 | -1596.4 | -315.1 | -82.4 | | |
| anti-BATF-2 | 5226.4 | 415.4 | 8680.0 | 3679.1 | 1333.0 | 6845.6 | | |
| anti-BATF-3 | 11336.1 | 5908.9 | 12477.3 | -1994.9 | 11289.7 | 10738.5 | | |

| protein | anti-BACH-3 | anti-E4BP4-3 | anti-C/EBP | anti-C/EBP-3 | anti-NFE2-2 | anti-NFE2-3 | anti-OASIS-3 | anti-OASIS-4 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 2143.0 | 174.8 | 19483.5 | 4422.2 | -355.2 | 154.6 | 173.8 | 241.5 |
| C/EBPβ | 245.4 | 304.5 | 20832.1 | 1479.4 | -402.3 | 125.0 | 63.8 | -2135.4 |
| C/EBPδ | 139.9 | 84.4 | 27624.8 | 4766.2 | -865.6 | 163.5 | 172.4 | -697.9 |
| C/EBPγ | 1274.4 | 1639.6 | 28736.4 | 2177.7 | 180.6 | 248.3 | 517.2 | 1409.9 |
| CHOP | 2534.4 | 3108.3 | 25218.6 | 19179.5 | 1.9 | 417.3 | 455.3 | 6441.0 |
| ATF-1 | 763.9 | 232.3 | -1572.1 | -221.4 | -332.4 | 229.2 | -98.1 | -1156.4 |
| CREB | 106.3 | -21.8 | -2865.9 | -568.4 | -101.0 | 199.1 | -115.4 | -256.2 |
| CREB-H | 508.1 | 1785.8 | -599.1 | -326.8 | -792.2 | 155.1 | 888.4 | 5117.4 |
| CREB3 | -20.6 | 1744.7 | -360.0 | -420.2 | -511.9 | 224.5 | 1467.8 | 14325.4 |
| ATF-6 | -1299.8 | 498.3 | -555.5 | -316.5 | -401.5 | 184.8 | -95.3 | 4.4 |
| ZF | -170.4 | 786.3 | 59.3 | -330.2 | -26.6 | 272.4 | 192.4 | -935.0 |
| XBP-1 | -1104.3 | 1056.8 | -653.0 | -321.8 | -744.3 | 227.2 | -14.8 | 3068.1 |
| E4BP4 | 156.8 | 807.4 | 1630.3 | -259.8 | -955.6 | 139.9 | 375.6 | 2518.8 |
| ATF-2 | 912.0 | 201.9 | 11515.4 | 104.6 | -20.0 | 256.7 | -43.1 | -981.9 |
| ATF-7 | 850.2 | 572.3 | 18691.0 | 1781.6 | 90.1 | 292.5 | 146.1 | 292.7 |
| cJun | -340.8 | 556.6 | 12588.9 | 297.3 | -588.3 | 134.4 | 1798.6 | 1136.4 |
| JunB | -427.9 | 703.1 | 8170.8 | -399.3 | -537.7 | 305.0 | 932.6 | -2125.8 |
| JunD | -224.6 | 635.7 | 10636.1 | -221.4 | -215.6 | 522.9 | 1178.8 | -1050.4 |
| Fos | 18745.0 | 1533.2 | 415.0 | 1192.4 | 2169.0 | 1059.9 | 637.1 | 2170.4 |
| Fra2 | 8467.4 | 834.7 | -122.7 | -6.9 | 1296.9 | 643.9 | -251.7 | -1659.3 |
| ATF-3 | 1744.6 | 1092.6 | 5388.7 | 1015.4 | -103.4 | 386.7 | 1449.0 | 7954.6 |
| ATF-4 | 1385.2 | 596.6 | 19941.3 | 1410.6 | 50.6 | 765.7 | 84.5 | 15012.3 |
| ATF-5 | -354.7 | 232.8 | 1098.2 | -226.3 | -1003.9 | 211.5 | -475.1 | -802.1 |
| B-ATF | 4697.1 | 713.5 | 7386.6 | 847.1 | -426.7 | 214.4 | -223.8 | -71.4 |
| p21SNFT | 4964.8 | 976.1 | 15339.8 | 1843.2 | -369.8 | 327.7 | 40.9 | 8407.6 |
| HLF | 141.9 | 589.1 | 6304.3 | 2294.4 | -194.3 | 270.6 | -187.0 | -264.1 |
| MafG | 397.4 | 963.6 | 146.1 | -124.4 | -405.6 | 355.4 | 2598.4 | 5215.3 |
| cMaf | 3008.6 | 184.3 | -453.6 | -357.7 | 4059.5 | 169.9 | -127.3 | -591.6 |
| MafB | 3523.8 | 339.9 | -2087.1 | -462.3 | 194.8 | 330.0 | -100.8 | -1851.6 |
| NFE2 | 25614.5 | 280.7 | -386.0 | -263.9 | 107.9 | 419.8 | 119.8 | -520.1 |
| NFE2L1 | 1860.9 | 268.2 | -793.3 | -647.3 | 309.1 | 349.0 | 187.6 | 546.3 |
| NFE2L3 | -220.4 | -305.9 | -2198.9 | -406.8 | -599.0 | 1.2 | -388.6 | -2437.3 |
| BACH1 | 11790.9 | 1573.3 | 52.4 | -54.1 | 1463.1 | 531.9 | 67.6 | 402.4 |
| anti-CREB-2 | | | | | | | | |
| anti-CREB-3 | | | | | | | | |
| anti-BACH-3 | 4499.4 | | | | | | | |
| anti-E4BP4-3 | | 1629.3 | | | | | | |
| anti-C/EBP | | | 5673.6 | | | | | |
| anti-C/EBP-3 | | | | -398.7 | | | | |
| anti-NFE2-2 | | | | | 21301.4 | | | |
| anti-NFE2-3 | | | | | | 923.5 | | |
| anti-OASIS-3 | | | | | | | -964.1 | |
| anti-OASIS-4 | | | | | | | | 26870.5 |
| anti-ZF | | | | | | | | |
| anti-ATF3-3 | | | | | | | | |
| anti-ATF2 | | | | | | | | |
| anti-ATF2-4 | | | | | | | | |
| anti-CHOP | | | | | | | | |
| anti-ATF6 | | | | | | | | |
| anti-LMAF-2 | | | | | | | | |
| anti-LMAF-3 | | | | | | | | |
| anti-LMAF | | | | | | | | |
| anti-PAR | | | | | | | | |
| anti-BATF-2 | | | | | | | | |
| anti-BATF-3 | | | | | | | | |

| protein | anti-ZF | anti-ATF3-3 | anti-ATF2 | anti-ATF2-4 | anti-CHOP | anti-ATF6 | anti-LMAF-2 | anti-LMAF-3 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -25.4 | 5394.8 | -1254.0 | -339.6 | 11954.4 | 104.4 | -48.1 | -786.5 |
| C/EBPβ | -467.5 | 5149.2 | -627.7 | -46.3 | 13913.9 | 162.6 | -82.8 | -1109.7 |
| C/EBPδ | -22.4 | 5497.4 | -926.5 | -470.6 | 11357.5 | 124.1 | -106.1 | -3455.9 |
| C/EBPγ | 585.6 | 16082.0 | 1131.6 | 3701.4 | 18072.4 | 139.3 | 443.0 | -448.8 |
| CHOP | 71.6 | 7075.2 | 3529.8 | 2911.8 | 20927.1 | 165.3 | 1998.6 | 329.6 |
| ATF-1 | -144.6 | -798.6 | 387.5 | -120.4 | -817.9 | 165.4 | -186.4 | -436.8 |
| CREB | -53.6 | -1185.9 | 562.3 | -406.6 | -1352.9 | 156.6 | -458.7 | -397.1 |
| CREB-H | 67.8 | -1645.5 | -268.6 | -89.8 | -764.1 | 159.9 | 53.9 | -831.8 |
| CREB3 | -11.3 | -1139.3 | -705.1 | -486.4 | -1661.3 | 158.6 | -212.1 | -1038.1 |
| ATF-6 | 517.4 | -1231.6 | -906.9 | -398.4 | -1868.9 | 1466.5 | -345.4 | -429.8 |
| ZF | 44044.8 | 9265.1 | 248.5 | 255.5 | 626.8 | 964.9 | 585.7 | -285.3 |
| XBP-1 | 1180.7 | -211.6 | -728.6 | -493.7 | -2719.2 | 2181.1 | -755.8 | -2185.7 |
| E4BP4 | -45.6 | -1317.8 | -852.9 | -796.5 | 117.3 | 148.9 | -220.6 | 278.8 |
| ATF-2 | -94.0 | 2287.5 | 15892.6 | 565.5 | 3454.3 | 218.6 | -24.6 | 13479.1 |
| ATF-7 | 157.8 | 3166.9 | 24424.9 | 8310.8 | 9652.8 | 171.7 | 383.5 | 5741.6 |
| cJun | 2224.6 | 17397.1 | 2563.4 | 4664.2 | 4618.9 | 239.3 | -291.6 | -52.5 |
| JunB | 753.6 | 8353.2 | 831.0 | 1838.9 | 1212.8 | 254.3 | 24.0 | -872.8 |
| JunD | 902.7 | 11596.1 | 301.8 | 1392.8 | 1713.2 | 228.3 | 23.6 | -119.5 |
| Fos | 569.6 | 35039.6 | -535.3 | 5698.4 | 16116.5 | 183.7 | 16476.2 | 35641.2 |
| Fra2 | 190.6 | 20464.5 | -1010.9 | 4583.8 | 9976.9 | 204.6 | 8672.6 | 31956.7 |
| ATF-3 | 91.7 | 21730.2 | 535.7 | 5781.0 | 20725.1 | 152.9 | 934.6 | 2967.8 |
| ATF-4 | 832.9 | 43681.9 | 341.0 | -152.4 | 33308.7 | 198.9 | -123.6 | -19.4 |
| ATF-5 | 19.8 | -504.0 | -479.8 | -477.9 | -2248.8 | 96.5 | -588.4 | -1210.1 |
| B-ATF | 87.6 | 1219.1 | 2301.8 | 13.3 | 18218.4 | 190.6 | 11858.0 | 412.4 |
| p21SNFT | 719.1 | 7028.4 | 7298.4 | 5691.6 | 33126.8 | 222.6 | 6614.2 | 1208.6 |
| HLF | 86.3 | 29.3 | -471.7 | -252.3 | 3446.3 | 240.3 | 615.4 | -72.4 |
| MafG | 364.8 | 22641.6 | -119.4 | 266.3 | -475.6 | 211.1 | 660.3 | 17393.2 |
| cMaf | -62.5 | 1309.9 | -490.1 | 2.0 | -980.2 | 161.7 | 1045.0 | 25571.0 |
| MafB | 71.3 | -726.8 | -7.1 | -317.6 | 32.0 | 246.3 | 786.3 | 20431.5 |
| NFE2 | 4000.7 | 8994.6 | -1355.6 | -476.6 | -2691.4 | 177.6 | 495.0 | 2026.1 |
| NFE2L1 | 415.9 | 10572.8 | -806.2 | -609.9 | -2601.3 | 135.9 | 33.4 | 638.0 |
| NFE2L3 | 322.2 | -2263.1 | -347.4 | -890.5 | -1742.6 | 101.5 | -558.1 | -2065.5 |
| BACH1 | 853.4 | 6761.4 | -802.4 | -193.9 | -1099.7 | 833.6 | 445.3 | 5196.8 |
| anti-CREB-2 | | | | | | | | |
| anti-CREB-3 | | | | | | | | |
| anti-BACH-3 | | | | | | | | |
| anti-E4BP4-3 | | | | | | | | |
| anti-C/EBP | | | | | | | | |
| anti-C/EBP-3 | | | | | | | | |
| anti-NFE2-2 | | | | | | | | |
| anti-NFE2-3 | | | | | | | | |
| anti-OASIS-3 | | | | | | | | |
| anti-OASIS-4 | | | | | | | | |
| anti-ZF | 7836.2 | | | | | | | |
| anti-ATF3-3 | | 10722.9 | | | | | | |
| anti-ATF2 | | | -794.4 | | | | | |
| anti-ATF2-4 | | | | -243.1 | | | | |
| anti-CHOP | | | | | 9340.5 | | | |
| anti-ATF6 | | | | | | 1966.1 | | |
| anti-LMAF-2 | | | | | | | 3977.4 | |
| anti-LMAF-3 | | | | | | | | 20934.9 |
| anti-LMAF | | | | | | | | |
| anti-PAR | | | | | | | | |
| anti-BATF-2 | | | | | | | | |
| anti-BATF-3 | | | | | | | | |

| protein | anti-LMAF | anti-PAR | anti-BATF-2 | anti-BATF-3 | anti-ATF2(2) | anti-LMAF(2) | anti-ZF(2) |
|---|---|---|---|---|---|---|---|
| C/EBPα | -871.9 | -18.8 | -1112.1 | -2431.1 | -843.2 | -306.1 | -94.1 |
| C/EBPβ | -977.6 | 122.6 | -1571.9 | -2797.9 | -388.2 | -644.4 | -206.4 |
| C/EBPδ | -1595.4 | -70.9 | -1296.8 | -254.2 | -774.6 | -1321.4 | -49.8 |
| C/EBPγ | 359.0 | 350.3 | 2470.0 | 142.9 | 1287.6 | 622.3 | 602.6 |
| CHOP | 3041.4 | 1427.4 | 1917.8 | 5555.9 | 4240.5 | 2871.2 | 35.2 |
| ATF-1 | -458.1 | 154.3 | -1142.9 | -1438.6 | 297.2 | -182.7 | -22.4 |
| CREB | -237.5 | 97.6 | -3065.3 | -6040.4 | 668.7 | -656.1 | 108.4 |
| CREB-H | -748.9 | -10.4 | 17.9 | 1350.9 | -418.4 | -441.4 | 87.7 |
| CREB3 | -508.4 | 46.7 | 1320.1 | 1984.7 | -310.1 | -215.8 | 21.0 |
| ATF-6 | -838.7 | 70.3 | -3596.8 | -3331.4 | -429.6 | -904.7 | 388.8 |
| ZF | -705.8 | 319.1 | -403.5 | 1838.1 | -94.3 | -44.3 | 33820.1 |
| XBP-1 | -2027.8 | 71.1 | -2021.0 | -525.6 | -264.2 | -1098.0 | 895.4 |
| E4BP4 | -694.5 | -50.6 | -2017.1 | -654.3 | -1404.6 | -696.1 | -11.6 |
| ATF-2 | 1278.4 | 173.2 | 1509.4 | 2444.4 | 16847.8 | 2267.1 | 10.4 |
| ATF-7 | 579.2 | 674.0 | 4919.4 | 4116.1 | 22662.7 | 1375.7 | 176.6 |
| cJun | -1184.4 | 517.8 | 4791.6 | 3539.0 | 2955.0 | 155.3 | 1858.1 |
| JunB | -558.5 | 261.1 | 6590.2 | 5778.6 | 828.8 | -117.3 | 786.9 |
| JunD | 287.6 | 162.3 | 4922.3 | 4002.1 | 947.8 | 439.6 | 755.3 |
| Fos | 10262.9 | 633.7 | 10661.9 | 10178.0 | 198.3 | 11654.8 | 431.8 |
| Fra2 | 15005.8 | 397.6 | 8319.8 | 8848.3 | 207.3 | 14514.8 | 221.7 |
| ATF-3 | 1006.2 | 2399.6 | 4512.6 | 2635.9 | 434.3 | 1915.1 | 78.3 |
| ATF-4 | 1262.4 | 356.3 | 6632.5 | 4305.6 | 613.6 | 1880.6 | 801.8 |
| ATF-5 | 183.6 | -15.1 | -4658.7 | -954.3 | -268.9 | 296.9 | -39.9 |
| B-ATF | 0.6 | 324.3 | 7148.6 | 8748.6 | 2440.4 | 1211.7 | 115.7 |
| p21SNFT | 2773.7 | 439.8 | 7081.5 | 6828.0 | 6479.3 | 3476.6 | 605.9 |
| HLF | -497.1 | 280.5 | -2534.9 | -122.2 | 71.4 | -466.9 | 48.6 |
| MafG | 6120.6 | -16.5 | 15293.9 | 12396.9 | 70.4 | 6918.6 | 359.3 |
| cMaf | 20835.4 | 196.3 | 2919.0 | -1283.5 | -317.1 | 15396.6 | 11.7 |
| MafB | 24035.4 | 169.3 | 4354.2 | -568.1 | -463.9 | 19611.3 | 61.8 |
| NFE2 | 662.8 | 170.0 | -123.0 | 1418.1 | -774.6 | 1078.1 | 3372.8 |
| NFE2L1 | 4858.1 | 139.3 | -1771.3 | 1862.2 | -432.9 | 5059.7 | 432.1 |
| NFE2L3 | -1532.1 | 43.9 | -1858.4 | -3154.9 | -661.6 | -946.6 | 158.2 |
| BACH1 | 8826.7 | 232.6 | 3181.9 | 5344.6 | -213.9 | 8569.6 | 640.4 |
| anti-CREB-2 | | | | | | | |
| anti-CREB-3 | | | | | | | |
| anti-BACH-3 | | | | | | | |
| anti-E4BP4-3 | | | | | | | |
| anti-C/EBP | | | | | | | |
| anti-C/EBP-3 | | | | | | | |
| anti-NFE2-2 | | | | | | | |
| anti-NFE2-3 | | | | | | | |
| anti-OASIS-3 | | | | | | | |
| anti-OASIS-4 | | | | | | | |
| anti-ZF | | | | | | | 7694.6 |
| anti-ATF3-3 | | | | | | | |
| anti-ATF2 | | | | | -599.8 | | |
| anti-ATF2-4 | | | | | | | |
| anti-CHOP | | | | | | | |
| anti-ATF6 | | | | | | | |
| anti-LMAF-2 | | | | | | | |
| anti-LMAF-3 | | | | | | | |
| anti-LMAF | 2455.9 | | | | | 2647.3 | |
| anti-PAR | | 538.2 | | | | | |
| anti-BATF-2 | | | -2385.7 | | | | |
| anti-BATF-3 | | | | -823.6 | | | |

$S_{array}$ values

| protein | C/EBPα | C/EBPδ | C/EBPγ | CHOP | CREB | CREB3 | ATF-6 | ZF |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 12.7 | 8.5 | 3.0 | 13.0 | -1.3 | -0.5 | -0.7 | -1.7 |
| C/EBPβ | 13.2 | 7.2 | 4.7 | 27.8 | -1.8 | -0.9 | -1.4 | -0.7 |
| C/EBPδ | 15.6 | 7.3 | 7.0 | 12.1 | -1.4 | 0.0 | -0.9 | -1.7 |
| C/EBPγ | 12.2 | 10.5 | 1.3 | 11.6 | -0.4 | 3.6 | -0.2 | -0.6 |
| CHOP | 17.7 | 10.0 | 7.6 | 3.7 | 0.0 | 5.1 | -0.1 | 0.6 |
| ATF-1 | -1.3 | -1.7 | -1.2 | -1.1 | 5.4 | -1.5 | -1.0 | -1.1 |
| CREB | -3.0 | -2.5 | -2.0 | -1.2 | 6.8 | 0.1 | -2.7 | -3.0 |
| CREB-H | -0.9 | -0.4 | -0.5 | -0.5 | 0.2 | 4.0 | 0.1 | -0.2 |
| CREB3 | -0.5 | -0.1 | -0.7 | -1.3 | -0.5 | 4.3 | -0.7 | 0.0 |
| ATF-6 | -1.0 | -1.0 | -1.3 | -0.8 | -0.7 | -1.0 | 7.9 | 0.0 |
| ZF | -1.7 | -1.5 | -1.1 | -0.1 | -1.2 | -1.0 | 0.5 | 2.5 |
| XBP-1 | -0.1 | -0.8 | -0.8 | -0.7 | 0.1 | -0.2 | 11.8 | 5.8 |
| E4BP4 | -0.5 | 0.5 | 0.0 | -0.3 | 4.6 | 2.9 | 0.2 | 0.2 |
| ATF-2 | -0.1 | -0.7 | 2.0 | 0.5 | -0.8 | -0.6 | -0.4 | -0.3 |
| ATF-7 | 3.7 | 1.6 | 4.2 | 4.3 | -0.3 | 0.7 | -0.2 | 0.8 |
| cJun | 0.8 | 0.8 | 0.1 | 0.9 | 0.2 | 0.0 | -0.1 | 0.0 |
| JunB | -0.5 | -0.3 | -0.8 | -0.4 | -0.3 | -0.8 | 0.1 | 0.3 |
| JunD | 0.2 | 0.3 | -0.4 | 0.0 | 0.0 | -0.7 | -0.1 | 0.3 |
| Fos | 3.1 | 0.7 | 0.5 | 2.4 | -0.5 | -0.2 | 0.0 | 0.5 |
| Fra2 | 1.0 | 0.0 | -0.1 | 0.7 | -0.1 | -1.0 | 0.1 | 0.7 |
| ATF-3 | 3.6 | 2.4 | 10.2 | 5.3 | -1.3 | 0.5 | -0.9 | -0.4 |
| ATF-4 | 35.4 | 19.7 | 32.2 | 21.9 | -1.3 | 1.0 | -1.3 | 7.5 |
| ATF-5 | 12.1 | 6.2 | 31.6 | -1.1 | -2.1 | -2.2 | -2.3 | -2.0 |
| B-ATF | 3.9 | 3.9 | 6.7 | 11.4 | 0.4 | 1.9 | 0.7 | 0.8 |
| p21SNFT | 5.5 | 5.7 | 3.9 | 15.8 | -0.3 | 5.1 | 0.3 | 0.7 |
| HLF | -0.4 | 0.0 | -0.5 | 2.8 | -0.3 | -0.6 | -0.4 | 0.0 |
| MafG | 0.0 | 0.3 | -0.1 | 0.4 | 0.4 | 0.4 | 0.5 | 1.1 |
| cMaf | -0.8 | -0.9 | -0.9 | -0.8 | -0.7 | -0.9 | 0.0 | 0.0 |
| MafB | -0.6 | -1.1 | -1.0 | -1.1 | -0.5 | -1.4 | 0.1 | 0.9 |
| NFE2 | -0.9 | -1.5 | -1.0 | -1.5 | -1.0 | -0.6 | -0.4 | -0.8 |
| NFE2L1 | -0.6 | -0.2 | -1.4 | -1.0 | 0.6 | -0.1 | -0.7 | 3.8 |
| NFE2L3 | -1.6 | -1.2 | -1.9 | -1.3 | -2.2 | -2.1 | -2.1 | -1.6 |
| BACH1 | -0.8 | -0.9 | -0.6 | 0.1 | 1.9 | 0.3 | 0.9 | 1.4 |
| anti-CREB-2 | 3.0 | 0.0 | 0.3 | -0.4 | 2.4 | 1.9 | 0.0 | -0.2 |
| anti-CREB-3 | -0.5 | -0.4 | -0.4 | -0.9 | 2.9 | 3.3 | 0.8 | -0.3 |
| anti-BACH-3 | 0.2 | -0.4 | -0.5 | -0.9 | 0.6 | 0.4 | 0.0 | -0.1 |
| anti-E4BP4-3 | -0.6 | -0.1 | 0.0 | -0.3 | 0.9 | 4.1 | 0.9 | -0.3 |
| anti-C/EBP | 20.1 | 12.6 | 14.2 | 13.1 | -0.8 | 2.4 | -0.3 | -0.2 |
| anti-C/EBP-3 | 10.3 | 6.2 | 2.5 | 9.9 | -0.1 | -0.4 | 0.3 | -0.2 |
| anti-NFE2-2 | -1.3 | -1.0 | -0.7 | -0.9 | -0.1 | -1.3 | 0.6 | -0.5 |
| anti-NFE2-3 | -0.9 | -1.4 | -1.7 | -2.1 | -0.5 | -1.3 | -0.2 | -0.6 |
| anti-OASIS-3 | 5.1 | 7.1 | 2.8 | -0.5 | 0.8 | 18.1 | 1.0 | 0.5 |
| anti-OASIS-4 | 1.7 | 2.3 | 2.0 | 1.2 | 1.3 | 15.1 | 1.6 | 0.4 |
| anti-ZF | -0.1 | -1.1 | -0.6 | -1.8 | -0.3 | -0.9 | 1.2 | 9.0 |
| anti-ATF3-3 | 6.6 | 3.8 | 6.8 | 2.4 | 0.5 | 2.5 | 1.5 | 3.4 |
| anti-ATF2 | 0.1 | 0.7 | 5.1 | 6.1 | 5.1 | 2.4 | 1.0 | 1.5 |
| anti-ATF2-4 | -0.7 | -0.3 | 4.7 | 0.3 | 0.1 | -0.7 | 0.1 | -0.2 |
| anti-CHOP | 12.9 | 6.1 | 12.2 | 11.9 | 0.0 | 1.2 | 0.0 | 0.2 |
| anti-ATF6 | -0.1 | -0.5 | -0.7 | -0.5 | 0.1 | -0.6 | 7.2 | 0.9 |
| anti-LMAF-2 | 5.0 | 0.8 | 1.0 | 1.5 | 2.1 | 1.8 | -0.3 | 0.3 |
| anti-LMAF-3 | -0.5 | -0.7 | -0.7 | -0.7 | 0.7 | 0.0 | 0.5 | -0.1 |
| anti-LMAF | -0.3 | -0.2 | -0.9 | -0.5 | 0.4 | -0.2 | 0.4 | -0.1 |
| anti-PAR | -0.8 | -0.8 | -0.5 | -0.4 | -0.1 | -0.8 | 0.3 | 0.4 |
| anti-BATF-2 | 3.5 | 1.5 | 1.7 | 1.0 | 0.1 | 7.5 | -0.7 | -0.8 |
| anti-BATF-3 | 5.5 | 3.8 | 2.9 | 3.6 | 2.6 | 13.7 | -0.3 | 0.5 |

| protein | XBP-1 | E4BP4 | ATF-2 | ATF-7 | cJun | Fos | ATF-3 | ATF-4 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -1.4 | -0.6 | -0.7 | 0.4 | -0.7 | -0.7 | -0.1 | 6.7 |
| C/EBPβ | -0.3 | -0.8 | -0.8 | 1.1 | -1.0 | -1.1 | 0.4 | 2.6 |
| C/EBPδ | -1.5 | -1.3 | -1.2 | -0.9 | -0.8 | -1.1 | 0.8 | 7.3 |
| C/EBPγ | -0.3 | 5.5 | 3.2 | 5.0 | 0.2 | -0.1 | 5.3 | 15.9 |
| CHOP | -0.2 | 15.4 | 3.5 | 6.8 | 1.7 | 2.1 | 6.5 | 2.6 |
| ATF-1 | 0.6 | 10.0 | -0.8 | -1.3 | -1.0 | -1.2 | -1.8 | -1.5 |
| CREB | -0.9 | 9.3 | -1.9 | -1.4 | -1.9 | -1.0 | -1.4 | -2.4 |
| CREB-H | 0.0 | 5.4 | -0.4 | -0.7 | -0.7 | -1.0 | -0.7 | -0.4 |
| CREB3 | -0.1 | 2.8 | -1.1 | -1.2 | -1.0 | -1.2 | -1.0 | -1.0 |
| ATF-6 | 4.9 | -0.1 | -1.1 | -1.0 | -1.4 | -1.3 | -1.3 | -1.4 |
| ZF | 7.0 | -0.9 | -0.2 | -0.2 | 0.0 | -0.8 | -0.5 | 2.5 |
| XBP-1 | 13.2 | 1.0 | -1.0 | -0.8 | -0.3 | -1.1 | -0.7 | -0.8 |
| E4BP4 | 0.2 | 61.4 | -1.0 | -0.8 | -1.3 | -1.1 | -1.0 | -1.2 |
| ATF-2 | 1.0 | -1.0 | 2.7 | 4.9 | 4.1 | 1.4 | 2.8 | -0.2 |
| ATF-7 | -0.1 | -0.1 | 5.4 | 6.9 | 9.1 | 2.9 | 4.9 | 2.1 |
| cJun | 0.1 | -0.3 | 8.3 | 12.6 | 2.4 | 12.5 | 7.3 | -0.7 |
| JunB | -0.1 | -1.4 | 3.0 | 5.8 | 0.0 | 6.9 | 4.8 | -0.9 |
| JunD | -0.4 | -0.8 | 4.4 | 9.3 | 0.5 | 9.5 | 6.5 | -0.8 |
| Fos | -1.1 | -0.2 | 4.4 | 6.7 | 24.7 | 0.3 | -0.1 | 2.7 |
| Fra2 | -0.1 | -0.7 | 3.9 | 6.0 | 14.5 | 0.8 | 1.2 | 0.7 |
| ATF-3 | -0.9 | 0.3 | 8.5 | 10.0 | 9.0 | 0.4 | -1.0 | 3.9 |
| ATF-4 | -1.1 | -0.7 | 1.5 | 5.6 | -0.7 | 3.3 | 2.7 | -1.4 |
| ATF-5 | -2.9 | -1.4 | -1.8 | -1.4 | -2.1 | -1.8 | -1.5 | -2.2 |
| B-ATF | -0.2 | 3.5 | 2.4 | 2.0 | 14.1 | -0.7 | 1.6 | 2.3 |
| p21SNFT | -0.3 | 4.2 | 4.2 | 4.9 | 10.2 | -0.6 | 4.6 | 3.1 |
| HLF | 0.9 | 2.6 | -0.8 | -0.4 | -0.1 | -0.9 | -0.7 | -0.2 |
| MafG | 1.0 | 2.5 | -0.2 | 0.8 | 1.2 | -0.7 | -0.3 | 0.0 |
| cMaf | 0.1 | -0.9 | -0.4 | -0.8 | -0.4 | -0.6 | -0.6 | 1.0 |
| MafB | 0.4 | -0.7 | -0.1 | -1.5 | -0.3 | 0.4 | -0.4 | -1.0 |
| NFE2 | -1.0 | -1.0 | -1.0 | -1.1 | -1.1 | -1.2 | -1.2 | -0.6 |
| NFE2L1 | -0.9 | -1.4 | 0.0 | 0.0 | -1.0 | -0.4 | -1.1 | 1.2 |
| NFE2L3 | -2.4 | -1.9 | -1.9 | -1.7 | -1.9 | -1.7 | -1.8 | -0.9 |
| BACH1 | 0.5 | 2.6 | 2.5 | 2.5 | -0.2 | -0.1 | -0.8 | 0.2 |
| anti-CREB-2 | 0.2 | 1.6 | -0.3 | -0.2 | 0.2 | 0.5 | 0.0 | -0.7 |
| anti-CREB-3 | -0.2 | 1.2 | -0.3 | -0.6 | 2.4 | 13.9 | 1.3 | -0.3 |
| anti-BACH-3 | 0.4 | 0.3 | 0.0 | -0.8 | -0.9 | 3.7 | 0.0 | -0.1 |
| anti-E4BP4-3 | 0.3 | 4.9 | -0.5 | -0.9 | -0.3 | -0.8 | -0.2 | -0.5 |
| anti-C/EBP | 0.2 | 4.1 | 2.0 | 5.0 | 1.2 | -0.9 | -0.2 | 1.4 |
| anti-C/EBP-3 | -0.3 | -0.1 | 0.2 | 2.6 | 0.8 | 1.0 | 1.5 | 1.3 |
| anti-NFE2-2 | 0.8 | -1.7 | -0.6 | -0.9 | -0.8 | 0.0 | -0.6 | -0.5 |
| anti-NFE2-3 | -0.3 | -0.9 | -1.6 | -1.7 | -1.0 | -0.4 | -1.6 | 0.2 |
| anti-OASIS-3 | -0.6 | 16.5 | 1.1 | 1.8 | 5.2 | 1.8 | 2.2 | 0.8 |
| anti-OASIS-4 | 0.3 | 9.2 | -0.4 | 0.0 | 0.1 | 0.0 | 1.9 | 2.2 |
| anti-ZF | 0.3 | -0.6 | -1.0 | -1.1 | 0.8 | -0.6 | -1.3 | 0.0 |
| anti-ATF3-3 | 0.6 | 0.5 | 1.8 | 2.0 | 6.8 | 7.5 | 3.9 | 7.8 |
| anti-ATF2 | -0.3 | 0.0 | 20.5 | 17.5 | 4.0 | 0.4 | 0.9 | 1.3 |
| anti-ATF2-4 | 0.1 | -1.4 | 2.1 | 8.0 | 2.0 | 3.2 | 2.9 | 0.0 |
| anti-CHOP | 0.6 | 7.2 | 3.6 | 4.7 | 2.0 | 4.7 | 4.8 | 5.6 |
| anti-ATF6 | 1.3 | -0.2 | -0.6 | -0.8 | -0.4 | -1.0 | -0.5 | -0.6 |
| anti-LMAF-2 | -0.2 | 0.1 | 0.0 | -0.3 | -0.5 | 10.1 | 0.3 | 0.2 |
| anti-LMAF-3 | 0.2 | 2.8 | 1.0 | 0.1 | -0.1 | 5.9 | 0.2 | -0.3 |
| anti-LMAF | -0.3 | -1.1 | 0.0 | -0.8 | -0.8 | 1.9 | -0.4 | -0.1 |
| anti-PAR | 0.0 | -0.8 | -0.8 | 0.0 | 0.9 | 0.0 | 1.6 | 0.1 |
| anti-BATF-2 | 1.7 | 0.2 | 1.6 | 2.3 | 2.4 | 3.6 | 0.4 | 1.9 |
| anti-BATF-3 | 2.9 | 14.5 | 5.1 | 3.6 | 3.8 | 4.8 | 1.3 | 3.7 |

| protein | p21SNFT | HLF | MafG | cMaf | NFE2 | BACH1 | anti-CREB-2 | anti-CREB-3 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 0.3 | -0.5 | -0.5 | -0.8 | -0.8 | -1.5 | -1.2 | -1.7 |
| C/EBPβ | 0.5 | -0.1 | -1.8 | -0.6 | -2.0 | -1.7 | -0.8 | -0.7 |
| C/EBPδ | 1.3 | 0.5 | -1.3 | -3.4 | -0.6 | -1.6 | -1.9 | -1.0 |
| C/EBPγ | 1.4 | 1.5 | 1.1 | -0.3 | -0.4 | -0.5 | 1.2 | -0.2 |
| CHOP | 5.4 | 21.3 | 2.0 | 0.1 | -0.2 | 1.9 | 2.7 | 0.0 |
| ATF-1 | -1.4 | -1.0 | -0.2 | -0.9 | -1.4 | 0.9 | 0.7 | 0.2 |
| CREB | -1.6 | -3.1 | -1.2 | -1.1 | -3.2 | 0.8 | 0.9 | 0.0 |
| CREB-H | -0.5 | -0.1 | -0.3 | 0.0 | -0.3 | -0.8 | 0.2 | 0.3 |
| CREB3 | -0.7 | -0.9 | 0.1 | -0.3 | -0.3 | -1.2 | 0.1 | -0.3 |
| ATF-6 | -1.2 | -0.6 | -0.2 | -0.3 | -0.9 | -0.8 | -0.9 | -1.1 |
| ZF | -0.3 | -0.9 | 1.1 | 0.0 | 0.4 | 0.2 | -0.5 | -0.8 |
| XBP-1 | -1.1 | -0.3 | -0.1 | -1.0 | -0.7 | -0.4 | 0.2 | -1.5 |
| E4BP4 | -0.2 | 0.2 | -0.5 | -0.8 | -0.3 | -0.6 | -1.4 | -1.0 |
| ATF-2 | 0.7 | 0.1 | -1.1 | -0.1 | -0.7 | 2.1 | -1.4 | -1.0 |
| ATF-7 | 1.5 | 0.5 | 1.9 | 0.3 | 0.5 | 3.3 | -0.1 | -0.6 |
| cJun | 3.9 | 3.0 | 0.5 | 0.2 | -0.7 | -0.7 | 1.7 | 2.4 |
| JunB | 4.2 | 0.9 | -0.2 | -0.8 | -0.6 | -1.1 | -0.2 | 1.1 |
| JunD | 3.7 | 1.1 | 0.0 | 0.9 | -0.3 | -0.5 | -0.1 | 0.6 |
| Fos | -0.7 | 0.0 | -0.5 | 1.3 | -0.6 | 1.3 | 6.3 | 75.3 |
| Fra2 | -0.9 | 0.3 | -0.1 | 1.2 | -0.4 | 0.3 | 3.5 | 40.8 |
| ATF-3 | 3.7 | 0.5 | 0.3 | 0.3 | -1.7 | -0.7 | 1.7 | 1.1 |
| ATF-4 | 3.7 | 4.4 | -0.3 | 7.9 | -0.3 | 0.1 | -0.8 | -1.0 |
| ATF-5 | 0.4 | 0.0 | -2.8 | -1.4 | 0.0 | -0.8 | -1.1 | -1.5 |
| B-ATF | 0.6 | 8.0 | -0.2 | -0.1 | 0.1 | 2.5 | 0.1 | -0.6 |
| p21SNFT | 1.0 | 8.0 | 5.4 | 0.0 | -0.7 | 1.7 | 5.7 | 0.3 |
| HLF | 0.0 | 6.4 | -0.1 | -0.4 | -0.7 | -0.4 | -0.5 | -1.2 |
| MafG | 0.6 | 0.0 | 15.5 | 0.3 | 5.1 | 17.4 | 4.5 | 1.4 |
| cMaf | -1.0 | -0.4 | -0.4 | 14.5 | -0.4 | 2.4 | -0.6 | 2.5 |
| MafB | -1.2 | -0.7 | -0.2 | 19.2 | 0.0 | 6.5 | -0.7 | 3.1 |
| NFE2 | -1.3 | -1.0 | 3.6 | -0.6 | 4.5 | -1.2 | 0.3 | 1.0 |
| NFE2L1 | -1.1 | -0.8 | 99.5 | 1.0 | 0.3 | -0.8 | -0.3 | 0.9 |
| NFE2L3 | -1.7 | -2.2 | 112.1 | -0.9 | 5.3 | -2.5 | -1.5 | -0.6 |
| BACH1 | -0.2 | 1.0 | 42.0 | 3.8 | -0.9 | 1.1 | 6.5 | 8.9 |
| anti-CREB-2 | 0.6 | -0.3 | 4.0 | 0.0 | 0.5 | 2.7 | 3.0 | |
| anti-CREB-3 | -0.4 | -0.5 | 3.2 | 8.9 | 1.8 | 14.9 | | 1.6 |
| anti-BACH-3 | -0.2 | -0.6 | -0.6 | 1.4 | 6.5 | 6.3 | | |
| anti-E4BP4-3 | -0.9 | 0.7 | -1.2 | -0.4 | 0.0 | 0.0 | | |
| anti-C/EBP | 0.4 | 7.3 | -0.8 | -0.2 | -0.2 | -1.3 | | |
| anti-C/EBP-3 | 0.4 | 6.1 | -1.1 | -1.1 | 0.3 | -0.4 | | |
| anti-NFE2-2 | -1.1 | -0.4 | -1.1 | 2.6 | 0.1 | 0.1 | | |
| anti-NFE2-3 | -1.4 | -1.5 | -0.9 | -0.7 | 2.2 | -0.4 | | |
| anti-OASIS-3 | -0.2 | -0.1 | 29.1 | 0.5 | 1.3 | 2.3 | | |
| anti-OASIS-4 | 0.7 | 0.0 | 12.6 | -0.2 | 1.0 | -0.5 | | |
| anti-ZF | -1.1 | -1.2 | 0.2 | -0.8 | 7.1 | 0.2 | | |
| anti-ATF3-3 | 0.8 | 0.6 | 46.4 | 1.9 | 5.7 | 5.1 | | |
| anti-ATF2 | 4.5 | 4.1 | 0.3 | 0.8 | -0.1 | -0.2 | | |
| anti-ATF2-4 | 1.9 | -0.6 | 0.2 | 0.0 | 0.1 | -0.8 | | |
| anti-CHOP | 5.8 | 10.9 | 0.5 | 0.7 | 0.7 | -0.4 | | |
| anti-ATF6 | -1.0 | -0.5 | 0.0 | 0.0 | 0.1 | -0.2 | | |
| anti-LMAF-2 | 3.2 | 4.0 | -0.5 | 6.0 | 1.5 | 2.9 | | |
| anti-LMAF-3 | -0.6 | 0.7 | 24.2 | 26.1 | 1.5 | 3.7 | | |
| anti-LMAF | -0.6 | -0.6 | 6.0 | 12.1 | 0.8 | 2.7 | | |
| anti-PAR | -1.1 | -0.1 | -1.9 | -1.1 | -0.4 | -0.3 | | |
| anti-BATF-2 | 0.3 | -0.1 | 22.7 | 3.6 | 0.0 | 3.9 | | |
| anti-BATF-3 | 1.5 | 3.1 | 32.5 | -1.4 | 2.5 | 6.3 | | |

| protein | anti-BACH-3 | anti-E4BP4-3 | anti-C/EBP | anti-C/EBP-3 | anti-NFE2-2 | anti-NFE2-3 | anti-OASIS-3 | anti-OASIS-4 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | 1.2 | -1.1 | 8.5 | 20.3 | -0.3 | -1.0 | 0.1 | 0.1 |
| C/EBPβ | -0.5 | -0.7 | 9.1 | 7.3 | -0.5 | -1.3 | -0.1 | -1.6 |
| C/EBPδ | -0.6 | -1.3 | 12.3 | 21.8 | -1.5 | -0.9 | 0.1 | -0.6 |
| C/EBPγ | 0.4 | 2.4 | 12.8 | 10.4 | 0.9 | -0.1 | 0.6 | 0.9 |
| CHOP | 1.6 | 5.9 | 11.1 | 85.3 | 0.5 | 1.4 | 0.5 | 4.5 |
| ATF-1 | 0.0 | -0.9 | -1.4 | -0.2 | -0.3 | -0.3 | -0.3 | -0.9 |
| CREB | -0.7 | -1.5 | -2.0 | -1.7 | 0.2 | -0.6 | -0.3 | -0.3 |
| CREB-H | -0.3 | 2.8 | -0.9 | -0.7 | -1.4 | -1.0 | 1.2 | 3.5 |
| CREB3 | -0.8 | 2.7 | -0.8 | -1.1 | -0.7 | -0.4 | 2.0 | 10.1 |
| ATF-6 | -2.0 | -0.3 | -0.9 | -0.6 | -0.5 | -0.7 | -0.3 | -0.1 |
| ZF | -0.9 | 0.4 | -0.6 | -0.7 | 0.4 | 0.1 | 0.1 | -0.7 |
| XBP-1 | -1.8 | 1.1 | -0.9 | -0.7 | -1.2 | -0.3 | -0.2 | 2.1 |
| E4BP4 | -0.6 | 0.5 | 0.1 | -0.4 | -1.7 | -1.2 | 0.4 | 1.7 |
| ATF-2 | 0.1 | -1.0 | 4.7 | 1.2 | 0.4 | -0.1 | -0.2 | -0.8 |
| ATF-7 | 0.0 | -0.1 | 8.1 | 8.6 | 0.7 | 0.3 | 0.1 | 0.1 |
| cJun | -1.1 | -0.1 | 5.2 | 2.1 | -0.9 | -1.2 | 2.5 | 0.7 |
| JunB | -1.2 | 0.2 | 3.2 | -1.0 | -0.8 | 0.4 | 1.2 | -1.6 |
| JunD | -1.0 | 0.0 | 4.3 | -0.2 | 0.0 | 2.4 | 1.6 | -0.8 |
| Fos | 16.7 | 2.2 | -0.4 | 6.0 | 5.5 | 7.5 | 0.8 | 1.4 |
| Fra2 | 7.1 | 0.5 | -0.7 | 0.7 | 3.5 | 3.6 | -0.5 | -1.3 |
| ATF-3 | 0.9 | 1.1 | 1.9 | 5.2 | 0.2 | 1.2 | 2.0 | 5.5 |
| ATF-4 | 0.5 | 0.0 | 8.7 | 7.0 | 0.6 | 4.7 | 0.0 | 10.5 |
| ATF-5 | -1.1 | -0.9 | -0.1 | -0.2 | -1.8 | -0.5 | -0.9 | -0.7 |
| B-ATF | 3.6 | 0.2 | 2.8 | 4.5 | -0.5 | -0.5 | -0.5 | -0.1 |
| p21SNFT | 3.9 | 0.9 | 6.5 | 8.9 | -0.4 | 0.6 | -0.1 | 5.9 |
| HLF | -0.6 | -0.1 | 2.3 | 10.9 | 0.0 | 0.1 | -0.4 | -0.3 |
| MafG | -0.4 | 0.8 | -0.6 | 0.2 | -0.5 | 0.9 | 3.7 | 3.6 |
| cMaf | 2.1 | -1.0 | -0.8 | -0.8 | 9.8 | -0.9 | -0.3 | -0.5 |
| MafB | 2.5 | -0.7 | -1.6 | -1.3 | 0.9 | 0.6 | -0.3 | -1.4 |
| NFE2 | 23.1 | -0.8 | -0.8 | -0.4 | 0.7 | 1.5 | 0.0 | -0.5 |
| NFE2L1 | 1.0 | -0.8 | -1.0 | -2.1 | 1.2 | 0.8 | 0.1 | 0.3 |
| NFE2L3 | -1.0 | -2.2 | -1.7 | -1.0 | -0.9 | -2.5 | -0.7 | -1.8 |
| BACH1 | 10.2 | 2.3 | -0.6 | 0.5 | 3.8 | 2.5 | -0.1 | 0.2 |
| anti-CREB-2 | | | | | | | | |
| anti-CREB-3 | | | | | | | | |
| anti-BACH-3 | 3.4 | | | | | | | |
| anti-E4BP4-3 | | 2.4 | | | | | | |
| anti-C/EBP | | | 2.0 | | | | | |
| anti-C/EBP-3 | | | | -1.0 | | | | |
| anti-NFE2-2 | | | | | 49.6 | | | |
| anti-NFE2-3 | | | | | | 6.2 | | |
| anti-OASIS-3 | | | | | | | -1.6 | |
| anti-OASIS-4 | | | | | | | | 18.9 |
| anti-ZF | | | | | | | | |
| anti-ATF3-3 | | | | | | | | |
| anti-ATF2 | | | | | | | | |
| anti-ATF2-4 | | | | | | | | |
| anti-CHOP | | | | | | | | |
| anti-ATF6 | | | | | | | | |
| anti-LMAF-2 | | | | | | | | |
| anti-LMAF-3 | | | | | | | | |
| anti-LMAF | | | | | | | | |
| anti-PAR | | | | | | | | |
| anti-BATF-2 | | | | | | | | |
| anti-BATF-3 | | | | | | | | |

| protein | anti-ZF | anti-ATF3-3 | anti-ATF2 | anti-ATF2-4 | anti-CHOP | anti-ATF6 | anti-LMAF-2 | anti-LMAF-3 |
|---|---|---|---|---|---|---|---|---|
| C/EBPα | -0.8 | 0.0 | -1.8 | -0.9 | 3.8 | -1.7 | -0.2 | -0.6 |
| C/EBPβ | -2.7 | -0.1 | -0.6 | 0.2 | 4.5 | -0.4 | -0.3 | -0.8 |
| C/EBPδ | -0.8 | 0.0 | -1.2 | -1.4 | 3.6 | -1.3 | -0.4 | -2.7 |
| C/EBPγ | 1.7 | 1.9 | 2.8 | 14.3 | 6.0 | -0.9 | 1.1 | -0.3 |
| CHOP | -0.4 | 0.3 | 7.4 | 11.3 | 7.1 | -0.4 | 5.3 | 0.3 |
| ATF-1 | -1.3 | -1.1 | 1.3 | -0.1 | -0.8 | -0.3 | -0.6 | -0.3 |
| CREB | -0.9 | -1.2 | 1.7 | -1.1 | -1.0 | -0.6 | -1.4 | -0.3 |
| CREB-H | -0.4 | -1.3 | 0.1 | 0.1 | -0.8 | -0.5 | 0.0 | -0.6 |
| CREB3 | -0.8 | -1.2 | -0.8 | -1.4 | -1.1 | -0.5 | -0.7 | -0.8 |
| ATF-6 | 1.4 | -1.2 | -1.2 | -1.1 | -1.2 | 29.4 | -1.1 | -0.3 |
| ZF | 181.7 | 0.7 | 1.1 | 1.4 | -0.3 | 17.9 | 1.5 | -0.2 |
| XBP-1 | 4.2 | -1.0 | -0.8 | -1.5 | -1.5 | 45.7 | -2.2 | -1.7 |
| E4BP4 | -0.9 | -1.2 | -1.1 | -2.6 | -0.5 | -0.7 | -0.7 | 0.2 |
| ATF-2 | -1.1 | -0.6 | 31.3 | 2.5 | 0.7 | 0.9 | -0.2 | 10.5 |
| ATF-7 | -0.1 | -0.4 | 47.8 | 31.5 | 3.0 | -0.2 | 0.9 | 4.5 |
| cJun | 8.5 | 2.1 | 5.6 | 17.9 | 1.1 | 1.3 | -0.9 | 0.0 |
| JunB | 2.4 | 0.5 | 2.2 | 7.3 | -0.1 | 1.7 | -0.1 | -0.7 |
| JunD | 3.0 | 1.1 | 1.2 | 5.6 | 0.1 | 1.1 | -0.1 | -0.1 |
| Fos | 1.6 | 5.2 | -0.4 | 21.7 | 5.3 | 0.1 | 44.5 | 27.7 |
| Fra2 | 0.1 | 2.7 | -1.4 | 17.6 | 3.1 | 0.5 | 23.3 | 24.8 |
| ATF-3 | -0.3 | 2.9 | 1.6 | 22.0 | 7.0 | -0.6 | 2.4 | 2.3 |
| ATF-4 | 2.7 | 6.8 | 1.3 | -0.2 | 11.6 | 0.4 | -0.5 | 0.0 |
| ATF-5 | -0.6 | -1.1 | -0.3 | -1.4 | -1.3 | -1.9 | -1.7 | -0.9 |
| B-ATF | -0.4 | -0.7 | 5.0 | 0.4 | 6.1 | 0.2 | 32.0 | 0.3 |
| p21SNFT | 2.3 | 0.3 | 14.7 | 21.7 | 11.5 | 1.0 | 17.8 | 1.0 |
| HLF | -0.4 | -1.0 | -0.3 | -0.6 | 0.7 | 1.4 | 1.5 | 0.0 |
| MafG | 0.8 | 3.0 | 0.4 | 1.4 | -0.7 | 0.7 | 1.7 | 13.5 |
| cMaf | -1.0 | -0.7 | -0.4 | 0.4 | -0.9 | -0.4 | 2.7 | 19.9 |
| MafB | -0.4 | -1.1 | 0.6 | -0.8 | -0.5 | 1.5 | 2.0 | 15.9 |
| NFE2 | 15.8 | 0.6 | -2.0 | -1.4 | -1.5 | -0.1 | 1.2 | 1.6 |
| NFE2L1 | 1.0 | 0.9 | -1.0 | -1.9 | -1.5 | -1.0 | 0.0 | 0.5 |
| NFE2L3 | 0.6 | -1.4 | -0.1 | -2.9 | -1.2 | -1.8 | -1.6 | -1.6 |
| BACH1 | 2.8 | 0.2 | -1.0 | -0.3 | -0.9 | 14.9 | 1.1 | 4.1 |
| anti-CREB-2 | | | | | | | | |
| anti-CREB-3 | | | | | | | | |
| anti-BACH-3 | | | | | | | | |
| anti-E4BP4-3 | | | | | | | | |
| anti-C/EBP | | | | | | | | |
| anti-C/EBP-3 | | | | | | | | |
| anti-NFE2-2 | | | | | | | | |
| anti-NFE2-3 | | | | | | | | |
| anti-OASIS-3 | | | | | | | | |
| anti-OASIS-4 | | | | | | | | |
| anti-ZF | 31.7 | | | | | | | |
| anti-ATF3-3 | | 0.9 | | | | | | |
| anti-ATF2 | | | -0.9 | | | | | |
| anti-ATF2-4 | | | | -0.5 | | | | |
| anti-CHOP | | | | | 2.9 | | | |
| anti-ATF6 | | | | | | 40.8 | | |
| anti-LMAF-2 | | | | | | | 10.6 | |
| anti-LMAF-3 | | | | | | | | 16.3 |
| anti-LMAF | | | | | | | | |
| anti-PAR | | | | | | | | |
| anti-BATF-2 | | | | | | | | |
| anti-BATF-3 | | | | | | | | |

| protein | anti-LMAF | anti-PAR | anti-BATF-2 | anti-BATF-3 | anti-ATF2(2) | anti-LMAF(2) | anti-ZF(2) |
|---|---|---|---|---|---|---|---|
| C/EBPα | -0.9 | -1.4 | -0.7 | -1.2 | -1.7 | -0.8 | -1.5 |
| C/EBPβ | -1.0 | -0.4 | -0.9 | -1.3 | -0.7 | -1.1 | -2.1 |
| C/EBPδ | -1.5 | -1.7 | -0.8 | -0.5 | -1.5 | -1.7 | -1.2 |
| C/EBPγ | 0.1 | 1.3 | 0.3 | -0.4 | 2.6 | 0.1 | 2.5 |
| CHOP | 2.4 | 9.0 | 0.1 | 1.1 | 8.4 | 2.1 | -0.7 |
| ATF-1 | -0.6 | -0.1 | -0.7 | -0.9 | 0.6 | -0.7 | -1.1 |
| CREB | -0.4 | -0.5 | -1.3 | -2.2 | 1.4 | -1.1 | -0.3 |
| CREB-H | -0.8 | -1.3 | -0.4 | -0.1 | -0.8 | -0.9 | -0.5 |
| CREB3 | -0.6 | -0.9 | 0.0 | 0.1 | -0.6 | -0.7 | -0.8 |
| ATF-6 | -0.9 | -0.7 | -1.4 | -1.4 | -0.8 | -1.3 | 1.3 |
| ZF | -0.8 | 1.1 | -0.5 | 0.1 | -0.2 | -0.5 | 190.4 |
| XBP-1 | -1.9 | -0.7 | -1.0 | -0.6 | -0.5 | -1.5 | 4.1 |
| E4BP4 | -0.8 | -1.6 | -1.0 | -0.7 | -2.8 | -1.1 | -1.0 |
| ATF-2 | 0.9 | 0.0 | 0.0 | 0.2 | 33.5 | 1.6 | -0.9 |
| ATF-7 | 0.3 | 3.6 | 1.0 | 0.7 | 45.0 | 0.8 | 0.1 |
| cJun | -1.2 | 2.5 | 1.0 | 0.5 | 5.9 | -0.3 | 9.6 |
| JunB | -0.7 | 0.6 | 1.5 | 1.2 | 1.7 | -0.6 | 3.5 |
| JunD | 0.0 | -0.1 | 1.0 | 0.7 | 1.9 | -0.1 | 3.3 |
| Fos | 8.4 | 3.3 | 2.6 | 2.4 | 0.4 | 10.1 | 1.5 |
| Fra2 | 12.4 | 1.6 | 2.0 | 2.1 | 0.4 | 12.7 | 0.3 |
| ATF-3 | 0.6 | 16.0 | 0.9 | 0.3 | 0.9 | 1.3 | -0.5 |
| ATF-4 | 0.9 | 1.3 | 1.5 | 0.8 | 1.2 | 1.2 | 3.6 |
| ATF-5 | 0.0 | -1.3 | -1.7 | -0.7 | -0.5 | -0.2 | -1.2 |
| B-ATF | -0.2 | 1.1 | 1.6 | 2.0 | 4.9 | 0.6 | -0.3 |
| p21SNFT | 2.1 | 1.9 | 1.6 | 1.5 | 12.9 | 2.7 | 2.5 |
| HLF | -0.6 | 0.8 | -1.1 | -0.5 | 0.2 | -0.9 | -0.7 |
| MafG | 4.9 | -1.4 | 4.0 | 3.1 | 0.2 | 5.8 | 1.1 |
| cMaf | 17.3 | 0.2 | 0.4 | -0.8 | -0.6 | 13.5 | -0.9 |
| MafB | 20.0 | 0.0 | 0.8 | -0.6 | -0.9 | 17.4 | -0.6 |
| NFE2 | 0.4 | 0.0 | -0.4 | -0.1 | -1.5 | 0.5 | 18.1 |
| NFE2L1 | 3.9 | -0.2 | -0.9 | 0.1 | -0.8 | 4.1 | 1.5 |
| NFE2L3 | -1.5 | -0.9 | -0.9 | -1.4 | -1.3 | -1.3 | -0.1 |
| BACH1 | 7.2 | 0.4 | 0.5 | 1.1 | -0.4 | 7.3 | 2.7 |
| anti-CREB-2 | | | | | | | |
| anti-CREB-3 | | | | | | | |
| anti-BACH-3 | | | | | | | |
| anti-E4BP4-3 | | | | | | | |
| anti-C/EBP | | | | | | | |
| anti-C/EBP-3 | | | | | | | |
| anti-NFE2-2 | | | | | | | |
| anti-NFE2-3 | | | | | | | |
| anti-OASIS-3 | | | | | | | |
| anti-OASIS-4 | | | | | | | |
| anti-ZF | | | | | | | 42.6 |
| anti-ATF3-3 | | | | | | | |
| anti-ATF2 | | | | | -1.2 | | |
| anti-ATF2-4 | | | | | | | |
| anti-CHOP | | | | | | | |
| anti-ATF6 | | | | | | | |
| anti-LMAF-2 | | | | | | | |
| anti-LMAF-3 | | | | | | | |
| anti-LMAF | 1.9 | | | | | 1.9 | |
| anti-PAR | | 2.6 | | | | | |
| anti-BATF-2 | | | -1.1 | | | | |
| anti-BATF-3 | | | | -0.7 | | | |

Table B.S6. Calculated $S_{array}$ scores for the complete set of 33 human bZIP measurements. Peptides on the surface are in rows, those in solution are in columns.

| Protein | C/EBPa | C/EBPb | C/EBPd | C/EBPg | CHOP | ATF-1 | CREB | CREB-H | CREB3 |
|---|---|---|---|---|---|---|---|---|---|
| C/EBPa | 9.2 | 13.9 | 11.6 | 5.9 | 9.7 | -1.1 | -1.8 | 0.4 | 0.7 |
| C/EBPb | 4.3 | 6.6 | 3.6 | 4.7 | 17.1 | -1.8 | -2.1 | 0.1 | -0.5 |
| C/EBPd | 10.7 | 17.3 | 10.3 | 8.5 | 9.9 | -1.4 | -2.3 | -0.7 | 2.4 |
| C/EBPg | 5.1 | 12.1 | 11.5 | 1.0 | 8.6 | -0.5 | -0.1 | -0.6 | 3.0 |
| CHOP | 14.7 | 34.5 | 16.1 | 20.1 | 2.5 | 0.8 | 0.5 | 1.7 | 6.8 |
| ATF-1 | -1.0 | -1.6 | -1.5 | -1.2 | -1.3 | 12.3 | 9.2 | -2.2 | -1.9 |
| CREB | -2.1 | -2.6 | -2.2 | -1.9 | -1.5 | 20.9 | 14.0 | 9.6 | 0.1 |
| CREB-H | -0.8 | -1.3 | -1.3 | -0.5 | -1.5 | 1.7 | 1.7 | 7.9 | 7.6 |
| CREB3 | 0.3 | -0.2 | 0.2 | -0.4 | -0.4 | -1.1 | -0.7 | 3.5 | 2.9 |
| ATF-6 | -1.6 | -1.4 | -1.3 | -1.2 | -1.5 | -0.7 | -0.6 | -0.2 | -1.9 |
| ZF | -1.3 | -1.4 | -2.2 | -1.0 | -1.7 | -1.3 | -1.4 | 0.0 | -2.0 |
| XBP-1 | -3.4 | -1.3 | -2.9 | -1.8 | -2.6 | 1.1 | 1.4 | -0.5 | -2.3 |
| E4BP4 | -0.6 | -0.2 | -0.4 | -0.6 | -0.5 | 15.6 | 11.4 | 3.0 | 5.9 |
| ATF-2 | 0.9 | 0.3 | -0.8 | 3.6 | 0.8 | -1.1 | -0.7 | -0.8 | -0.5 |
| ATF-7 | 2.4 | 3.7 | 0.8 | 5.2 | 3.0 | -0.7 | -0.4 | -0.1 | 0.7 |
| cJun | -0.3 | 1.5 | 0.6 | 0.7 | 0.2 | 1.2 | 1.7 | 0.1 | 0.0 |
| JunB | 0.5 | 0.2 | 0.2 | -0.7 | 0.0 | -0.5 | -0.3 | -2.8 | -1.4 |
| JunD | 0.0 | 0.3 | 0.4 | 0.0 | -0.6 | -0.2 | -0.1 | -0.7 | -0.6 |
| Fos | 2.0 | 0.1 | 0.7 | 1.7 | 1.6 | 1.0 | 0.3 | 0.5 | 0.8 |
| Fra2 | -1.0 | -1.6 | -1.7 | -1.4 | 0.0 | 0.5 | 0.7 | -0.9 | -1.2 |
| ATF-3 | 1.8 | 6.8 | 1.2 | 14.0 | 5.9 | -0.4 | -1.0 | 0.6 | 1.7 |
| ATF-4 | 23.4 | 12.6 | 13.1 | 23.2 | 10.4 | -0.7 | -1.0 | 1.2 | 1.1 |
| ATF-5 | 4.2 | 0.3 | 3.1 | 23.2 | -1.8 | -2.8 | -4.0 | -1.6 | -1.7 |
| B-ATF | 2.2 | 7.3 | 2.3 | 9.5 | 6.9 | 1.1 | 1.6 | 0.3 | 3.9 |
| p21SNFT | 1.7 | 5.7 | 4.1 | 4.6 | 9.6 | -1.2 | -0.3 | 1.2 | 6.9 |
| HLF | -0.4 | 0.1 | -0.4 | -0.9 | 1.0 | -0.5 | -0.9 | 0.2 | -0.4 |
| MafG | -0.3 | -0.5 | -0.1 | -0.4 | 0.2 | 0.9 | 1.2 | -1.6 | 1.0 |
| cMaf | -1.0 | -0.9 | -1.9 | -0.9 | -1.4 | 0.2 | 0.4 | -0.5 | -1.4 |
| MafB | -2.7 | -2.1 | -2.5 | -1.4 | -2.3 | 0.1 | 0.5 | -0.7 | -2.5 |
| NFE2 | -1.2 | -1.4 | -1.8 | -1.0 | -1.2 | -0.6 | -0.9 | -0.6 | -0.3 |
| NFE2L1 | -1.2 | -2.6 | -0.9 | -2.9 | -2.8 | 2.0 | 2.3 | -0.8 | 0.5 |
| NFE2L3 | -0.9 | -2.7 | -1.1 | -1.3 | -1.6 | -3.0 | -2.2 | 0.0 | -1.2 |
| BACH1 | 0.0 | -0.6 | -0.9 | 0.5 | -0.3 | 6.4 | 5.6 | -0.1 | 0.5 |

| Protein | ATF-6 | ZF | XBP-1 | E4BP4 | ATF-2 | ATF-7 | cJun | JunB |
|---|---|---|---|---|---|---|---|---|
| C/EBPa | -0.8 | -1.8 | -1.3 | 0.5 | -0.1 | 1.2 | 0.5 | -0.6 |
| C/EBPb | -0.5 | -1.6 | 0.8 | -0.9 | -0.4 | 1.6 | -0.1 | -0.2 |
| C/EBPd | -1.3 | -3.1 | -2.7 | 0.1 | -1.4 | -1.0 | -0.4 | -1.7 |
| C/EBPg | 0.4 | -1.5 | 0.0 | 1.6 | 2.2 | 5.8 | -0.1 | 0.1 |
| CHOP | 0.6 | 1.2 | 0.0 | 10.6 | 5.1 | 8.5 | 1.8 | 4.0 |
| ATF-1 | -0.7 | -1.5 | 1.6 | 4.2 | -1.3 | -1.4 | -0.6 | -0.6 |
| CREB | -3.7 | -4.5 | 1.0 | 4.1 | -3.6 | -1.9 | -1.8 | -2.2 |
| CREB-H | 0.8 | -0.8 | 0.7 | 4.7 | -0.4 | -1.1 | -0.6 | -0.3 |
| CREB3 | -0.4 | 0.2 | -0.3 | 0.2 | -0.6 | -1.3 | -0.2 | -1.1 |
| ATF-6 | 12.3 | 0.9 | 13.0 | -0.5 | -1.6 | -1.9 | -1.7 | -1.1 |
| ZF | 1.4 | 15.4 | 24.1 | 0.4 | -1.2 | -0.2 | 0.5 | 0.8 |
| XBP-1 | 19.0 | 17.8 | 39.6 | -1.1 | -1.7 | -1.6 | -0.7 | -0.2 |
| E4BP4 | 0.7 | -0.1 | -0.1 | 39.8 | -1.8 | -1.8 | -0.4 | -0.5 |
| ATF-2 | -0.6 | -0.3 | 1.5 | -2.2 | 3.0 | 4.1 | 6.2 | 5.5 |
| ATF-7 | -0.1 | 1.0 | 0.0 | -0.5 | 4.2 | 6.8 | 14.5 | 6.9 |
| cJun | -0.1 | 0.1 | 0.1 | -0.1 | 12.4 | 17.4 | 4.2 | 1.4 |
| JunB | 0.2 | 0.4 | -0.7 | -1.1 | 3.5 | 6.2 | 0.4 | -0.5 |
| JunD | 0.2 | 0.2 | -0.3 | -0.8 | 6.6 | 11.7 | 0.6 | -0.3 |
| Fos | -0.4 | 1.8 | -0.3 | 0.7 | 5.1 | 8.8 | 37.4 | 31.4 |
| Fra2 | 0.5 | 0.3 | 0.4 | -0.3 | 5.4 | 6.8 | 22.6 | 21.3 |
| ATF-3 | 0.0 | -0.7 | 2.0 | 0.9 | 7.8 | 12.3 | 10.8 | 12.2 |
| ATF-4 | -2.0 | 10.9 | -1.0 | -1.0 | 0.3 | 1.9 | -1.1 | -1.4 |
| ATF-5 | -2.8 | -2.9 | -2.3 | -1.8 | -2.6 | -3.0 | -3.4 | -3.4 |
| B-ATF | 1.2 | 0.8 | -0.3 | 3.0 | 2.4 | 1.6 | 20.2 | 26.8 |
| p21SNFT | 0.0 | 1.7 | -0.2 | 3.0 | 4.1 | 5.5 | 17.4 | 18.7 |
| HLF | -1.1 | 0.1 | 0.0 | 0.9 | -1.5 | -1.7 | -0.9 | 0.1 |
| MafG | -0.5 | 0.5 | 1.1 | 2.2 | 0.2 | 0.4 | 1.4 | 2.3 |
| cMaf | -0.1 | -0.8 | -1.0 | -0.7 | -0.4 | -1.6 | -0.4 | 0.8 |
| MafB | -0.1 | 0.8 | 0.7 | -0.6 | -0.6 | -1.7 | -0.9 | 0.5 |
| NFE2 | 0.6 | -0.3 | -0.4 | -0.7 | -1.2 | -2.0 | -1.8 | -1.6 |
| NFE2L1 | -0.6 | 19.8 | -0.9 | -2.5 | 0.0 | 0.5 | -2.8 | -1.3 |
| NFE2L3 | -2.8 | -2.3 | -2.1 | -0.6 | -2.3 | -2.0 | -2.2 | -2.5 |
| BACH1 | 2.9 | 3.8 | 2.9 | 1.2 | 2.7 | 2.1 | 0.1 | 1.1 |

| Protein | JunD | Fos | Fra2 | ATF-3 | ATF-4 | ATF-5 | B-ATF | p21SNFT |
|---|---|---|---|---|---|---|---|---|
| C/EBPa | -0.2 | 0.7 | -0.2 | 0.7 | 12.8 | 1.2 | 1.7 | 1.6 |
| C/EBPb | -0.5 | -0.7 | -0.1 | 1.3 | 4.9 | 0.1 | 3.4 | 1.3 |
| C/EBPd | -1.6 | -0.2 | 0.0 | 1.6 | 16.5 | 0.7 | 3.4 | 4.1 |
| C/EBPg | 0.2 | 0.0 | 2.0 | 9.3 | 23.3 | 18.1 | 8.0 | 3.6 |
| CHOP | 3.5 | 8.8 | 11.7 | 11.5 | 11.0 | -0.6 | 28.1 | 9.5 |
| ATF-1 | -1.0 | -1.3 | -1.0 | -1.5 | -1.4 | -0.4 | -0.6 | -2.2 |
| CREB | -3.3 | -2.3 | -2.6 | -2.8 | -3.4 | -2.3 | -0.8 | -3.4 |
| CREB-H | -1.0 | -0.3 | -0.7 | -0.4 | -1.3 | 0.3 | -0.7 | -0.8 |
| CREB3 | -0.4 | -0.9 | -1.4 | 0.2 | 0.2 | 0.0 | -0.5 | -0.1 |
| ATF-6 | -1.3 | -1.9 | -1.2 | -1.7 | -1.4 | -0.4 | -0.8 | -2.4 |
| ZF | 0.5 | -0.5 | -0.4 | -0.8 | 2.2 | -0.6 | -0.9 | -1.0 |
| XBP-1 | -0.1 | -1.3 | -0.9 | -1.9 | -2.1 | -1.4 | -1.5 | -2.5 |
| E4BP4 | -1.1 | -1.8 | -0.9 | -0.4 | -1.3 | -1.5 | -1.0 | -0.1 |
| ATF-2 | 6.2 | 5.3 | 8.5 | 5.6 | 0.4 | -0.1 | 0.8 | 2.2 |
| ATF-7 | 10.5 | 8.6 | 17.1 | 7.5 | 2.9 | -1.2 | 0.5 | 2.8 |
| cJun | 1.3 | 44.9 | 76.8 | 13.1 | -1.4 | -0.8 | 25.9 | 6.9 |
| JunB | 0.0 | 29.1 | 38.0 | 6.4 | -0.7 | 0.0 | 16.5 | 6.0 |
| JunD | -0.3 | 37.2 | 51.3 | 9.3 | -1.4 | -0.5 | 21.6 | 6.4 |
| Fos | 32.3 | 2.9 | 2.9 | 1.2 | 4.5 | 0.5 | -0.5 | -0.8 |
| Fra2 | 23.7 | 3.6 | 1.2 | 2.3 | 0.1 | -0.6 | -2.0 | -1.8 |
| ATF-3 | 11.2 | 3.3 | 6.6 | -0.2 | 6.1 | -0.7 | 1.5 | 7.0 |
| ATF-4 | -1.6 | 5.4 | 1.0 | 3.9 | -1.9 | -0.5 | 2.9 | 2.9 |
| ATF-5 | -3.4 | -3.1 | -1.6 | -2.5 | -1.8 | -1.6 | 0.3 | 0.7 |
| B-ATF | 26.7 | -0.3 | -2.0 | 2.0 | 3.0 | 2.2 | -0.2 | 0.8 |
| p21SNFT | 16.5 | 0.2 | 0.4 | 7.5 | 3.9 | 1.0 | 1.1 | 2.2 |
| HLF | -1.1 | -1.5 | -1.2 | -1.7 | -0.4 | -0.7 | 1.0 | -0.5 |
| MafG | 1.8 | 1.3 | -1.0 | 0.0 | -0.1 | 0.6 | -0.8 | 1.1 |
| cMaf | 0.2 | 0.2 | -0.4 | -0.6 | 3.5 | -0.3 | -0.9 | -2.0 |
| MafB | 0.6 | 3.1 | 2.8 | -0.6 | -1.4 | -1.0 | -0.7 | -2.0 |
| NFE2 | -1.7 | -1.4 | -1.1 | -1.5 | -0.6 | -0.2 | -1.8 | -2.5 |
| NFE2L1 | -0.8 | 0.7 | 0.2 | -1.7 | 4.2 | 0.2 | -1.2 | -1.4 |
| NFE2L3 | -2.6 | -2.9 | -2.0 | -2.3 | -1.0 | -0.9 | -0.8 | -2.7 |
| BACH1 | 1.4 | 1.8 | 1.0 | -0.3 | 0.2 | 1.3 | 0.4 | -0.1 |

| Protein | HLF | MafG | cMaf | MafB | NFE2 | NFE2L1 | NFE2L3 | BACH1 |
|---|---|---|---|---|---|---|---|---|
| C/EBPa | 0.0 | -0.6 | -0.4 | -0.7 | 0.0 | 1.5 | 0.5 | -0.7 |
| C/EBPb | -0.4 | -0.2 | -0.2 | -0.3 | -1.0 | -1.2 | -0.7 | -1.7 |
| C/EBPd | -0.8 | -0.4 | -1.5 | -2.2 | -0.1 | 0.3 | -0.1 | -2.4 |
| C/EBPg | 0.2 | 0.0 | -0.4 | -0.2 | -1.0 | -1.2 | 0.6 | -1.3 |
| CHOP | 23.8 | 1.3 | -0.8 | 0.1 | 0.4 | -0.5 | -0.5 | 5.8 |
| ATF-1 | -1.4 | -0.2 | 0.1 | -0.3 | -0.8 | -0.8 | -1.0 | 2.3 |
| CREB | -3.1 | -2.0 | -0.9 | -2.0 | -2.9 | -1.5 | -4.2 | 2.6 |
| CREB-H | -1.3 | -0.2 | -0.9 | -0.6 | 0.4 | -0.4 | 0.5 | -1.9 |
| CREB3 | -0.3 | -0.1 | 0.2 | -0.1 | 0.4 | -0.6 | -0.1 | -1.3 |
| ATF-6 | -1.6 | -0.4 | -2.1 | -1.9 | -1.2 | -3.9 | -2.5 | -2.1 |
| ZF | -1.4 | 0.2 | 0.0 | 0.3 | 7.8 | 25.5 | 0.1 | 2.5 |
| XBP-1 | -1.8 | -0.3 | -1.3 | -2.1 | -2.4 | -2.7 | -0.9 | -1.8 |
| E4BP4 | -0.5 | 0.4 | -0.6 | -0.3 | -1.1 | -1.8 | 0.0 | 0.1 |
| ATF-2 | 0.6 | 0.1 | 0.7 | 1.1 | -0.1 | 2.7 | -0.6 | 6.3 |
| ATF-7 | 0.6 | 0.4 | 0.0 | 0.2 | 3.9 | 6.5 | 0.9 | 9.2 |
| cJun | 2.7 | -0.4 | 0.2 | 0.0 | -2.4 | -0.8 | 0.0 | -0.3 |
| JunB | 1.1 | -0.8 | 0.4 | -0.1 | -0.7 | -1.6 | -0.1 | 0.3 |
| JunD | 1.5 | -0.2 | -0.9 | -0.3 | -0.9 | -0.7 | -0.1 | -0.6 |
| Fos | 1.2 | 0.4 | 1.4 | 6.3 | -0.3 | 5.2 | 1.0 | 4.7 |
| Fra2 | 0.4 | 0.2 | -0.4 | 4.4 | -1.1 | 1.1 | -0.6 | 1.3 |
| ATF-3 | 0.8 | 0.5 | 0.4 | 3.5 | -0.7 | -1.2 | -2.0 | -0.2 |
| ATF-4 | 3.0 | -0.3 | 5.6 | 0.5 | 0.6 | 17.6 | 5.5 | 0.1 |
| ATF-5 | -0.2 | -3.2 | -0.9 | -2.2 | 1.5 | 0.5 | -2.0 | 0.4 |
| B-ATF | 8.3 | 0.1 | -2.2 | 0.2 | 0.6 | 4.1 | 1.7 | 5.1 |
| p21SNFT | 8.2 | 2.0 | -0.5 | 0.7 | -1.0 | 0.6 | -0.5 | 2.5 |
| HLF | 1.9 | -1.0 | -1.1 | -2.3 | -0.3 | -0.9 | -0.4 | -0.3 |
| MafG | -1.4 | 3.7 | -0.5 | 0.4 | 21.3 | 36.1 | 71.6 | 41.6 |
| cMaf | -1.4 | -0.1 | 16.3 | 13.3 | -0.7 | 3.9 | 1.1 | 6.9 |
| MafB | -2.4 | -0.6 | 15.7 | 8.2 | -0.1 | 5.2 | 0.0 | 12.7 |
| NFE2 | -1.5 | 1.2 | -0.7 | -0.4 | 37.6 | -0.4 | 22.3 | -1.4 |
| NFE2L1 | -0.8 | 38.2 | 0.7 | 2.2 | 3.4 | 4.4 | 0.7 | -0.7 |
| NFE2L3 | -2.3 | 46.4 | -0.4 | -1.2 | 33.0 | -1.9 | 1.1 | -2.4 |
| BACH1 | 1.8 | 18.1 | 5.9 | 10.3 | -0.5 | 0.8 | 0.2 | 3.9 |

## REFERENCES

1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006a;45(38):11324-32.

2. Acharya A, Rishi V, Moll J, Vinson C. Experimental identification of homodimerizing B-ZIP families in homo sapiens. Journal of Structural Biology. 2006b;155(2):130-9.

3. Acharya A, Ruvinov SB, Gal J, Moll JR, Vinson C. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: Amino acids I, V, L, N, A, and K. Biochemistry. 2002;41(48):14122-31.

4. Ahn S, Olive M, Aggarwal S, Krylov D, Ginty DD, Vinson C. A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. Mol Cell Biol. 1998 Feb;18(2):967-77.

5. Aloy P, Russell RB. Interrogating protein interaction networks through structural biology. Proc Natl Acad Sci U S A. 2002 Apr 30;99(9):5896-901.

6. Amoutzias GD, Veron AS, Weiner J,3rd, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL. One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. Mol Biol Evol. 2007 Mar;24(3):827-35.

7. Apgar JR, Gutwin KN, Keating AE. Predicting helix orientation for coiled-coil dimers. Proteins: Structure, Function, and Bioinformatics. 2008;72(3):1048-65.

8. Apgar JR, Hahn S, Grigoryan G, Keating AE. Cluster expansion models for flexible-backbone protein energetics. J Comput Chem. 2009 Nov 30;30(15):2402-13.

9. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell. 2008 Jun 27;133(7):1266-76.

10. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G. Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol. 2008 Sep;26(9):1041-5.

11. Das R, Baker D. Macromolecular modeling with rosetta. Annu Rev Biochem. 2008;77:363-82.

12. Fong J, Keating A, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biology. 2004;5(2):R11.

13. Friedland GD, Linares AJ, Smith CA, Kortemme T. A simple model of backbone flexibility improves modeling of side-chain conformational variability. J Mol Biol. 2008 Jul 18;380(4):757-74.

14. Fu X, Apgar JR, Keating AE. Modeling backbone flexibility to achieve sequence diversity: The design of novel alpha-helical ligands for bcl-xL. J Mol Biol. 2007 Aug 24;371(4):1099-117.

15. Gonzalez L,Jr, Woolfson DN, Alber T. Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat Struct Biol. 1996 Dec;3(12):1011-8.

16. Grigoryan G, Zhou F, Lustig SR, Ceder G, Morgan D, Keating AE. Ultra-fast evaluation of protein energies directly from sequence. PLoS Comput Biol. 2006 Jun 16;2(6):e63.

17. Grigoryan G, Keating AE. Structure-based prediction of bZIP partnering specificity. Journal of Molecular Biology. 2006;355(5):1125-42.

18. Hadley EB, Testa OD, Woolfson DN, Gellman SH. Preferred side-chain constellations at antiparallel coiled-coil interfaces. Proc Natl Acad Sci U S A. 2008 Jan 15;105(2):530-5.

19. Hou T, Zhang W, Case DA, Wang W. Characterization of domain-peptide interaction interface: A case study on the amphiphysin-1 SH3 domain. J Mol Biol. 2008 Feb 29;376(4):1201-14.

20. Hou T, Chen K, McLaughlin WA, Lu B, Wang W. Computational analysis and prediction of the binding motif and protein interacting partners of the abl SH3 domain. PLoS Comput Biol. 2006 Jan;2(1):e1.

21. Jamal Rahi S, Virnau P, Mirny LA, Kardar M. Predicting transcription factor specificity with all-atom models. Nucleic Acids Res. 2008 Nov;36(19):6209-17.

22. John DM, Weeks KM. Van't hoff enthalpies without baselines. Protein Sci. 2000 Jul;9(7):1416-9.

23. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006 Jan 12;439(7073):168-74.

24. Kaplan T, Friedman N, Margalit H. Ab initio prediction of transcription factor targets using structural knowledge. PLoS Comput Biol. 2005 Jun;1(1):e1.

25. Kingsford CL, Chazelle B, Singh M. Solving and analyzing side-chain positioning problems using linear and integer programming. Bioinformatics. 2005 Apr 1;21(7):1028-36.

26. Kortemme T, Baker D. Computational design of protein-protein interactions. Curr Opin Chem Biol. 2004 Feb;8(1):91-7.

27. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol. 2004 Apr;11(4):371-9.

28. Landgraf C, Panni S, Montecchi-Palazzi L, Castagnoli L, Schneider-Mergener J, Volkmer-Engert R, Cesareni G. Protein interaction networks by proteome peptide scanning. PLoS Biol. 2004 Jan;2(1):E14.

29. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. TRANSFAC: Transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 2003 Jan 1;31(1):374-8.

30. McClain DL, Gurnon DG, Oakley MG. Importance of potential interhelical salt-bridges involving interior residues for coiled-coil stability and quaternary structure. J Mol Biol. 2002 Nov 22;324(2):257-70.

31. Moitra J, Szilak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. Biochemistry. 1997;36(41):12567-73.

32. Morozov AV, Havranek JJ, Baker D, Siggia ED. Protein-DNA binding specificity predictions with structural models. Nucleic Acids Res. 2005 Oct 24;33(18):5781-98.

33. Newman JRS, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 June 27, 2003;300(5628):2097-101.

34. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell. 2008 Jun 27;133(7):1277-89.

35. Oakley MG, Kim PS. A buried polar interaction can direct the relative orientation of helices in a coiled coil. Biochemistry. 1998 Sep 8;37(36):12603-10.

36. O'Shea EK, Rutkowski R, Kim PS. Mechanism of specificity in the fos-jun oncoprotein heterodimer. Cell. 1992 Feb 21;68(4):699-708.

37. Paillard G, Deremble C, Lavery R. Looking into DNA recognition: Zinc finger binding specificity. Nucleic Acids Res. 2004 Dec 21;32(22):6673-82.

38. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C. Computer-aided design of a PDZ domain to recognize new target sequences. Nat Struct Biol. 2002 Aug;9(8):621-7.

39. Remy I, Michnick SW. A highly sensitive protein-protein interaction assay based on gaussia luciferase. Nat Methods. 2006 Dec;3(12):977-9.

40. Sanchez IE, Beltrao P, Stricher F, Schymkowitz J, Ferkinghoff-Borg J, Rousseau F, Serrano L. Genome-wide prediction of SH2 domain targets using structural information and the FoldX algorithm. PLoS Comput Biol. 2008 Apr 4;4(4):e1000052.

41. Siggers TW, Honig B. Structure-based prediction of C2H2 zinc-finger binding specificity: Sensitivity to docking geometry. Nucleic Acids Res. 2007;35(4):1085-97.

42. Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. J Mol Biol. 2008 Jul 18;380(4):742-56.

43. Spaller MR. Act globally, think locally: Systems biology addresses the PDZ domain. ACS Chem Biol. 2006 May 23;1(4):207-10.

44. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. Science. 2007 July 20, 2007;317(5836):364-9.

45. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, Malitskaya Y, Vogel J, Bussey H, Michnick SW. An in vivo map of the yeast protein interactome. Science. 2008 Jun 13;320(5882):1465-70.

46. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS. A specificity map for the PDZ domain family. PLoS Biol. 2008 Sep 30;6(9):e239.

47. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, Bonovich M. Classification of human B-ZIP proteins based on dimerization properties. Mol Cell Biol. 2002 Sep;22(18):6321-35.

48. Vinson C, Acharya A, Taparowsky EJ. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2006;1759(1-2):4-12.

49. Wiedemann U, Boisguerin P, Leben R, Leitner D, Krause G, Moelling K, Volkmer-Engert R, Oschkinat H. Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. J Mol Biol. 2004 Oct 22;343(3):703-18.

50. Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, DeGrado WF. Computational design of peptides that target transmembrane helices. Science. 2007 March 30, 2007;315(5820):1817-22.

51. Zhou F, Grigoryan G, Lustig SR, Keating AE, Ceder G, Morgan D. Coarse-graining protein energetics in sequence variables. Phys Rev Lett. 2005 Sep 30;95(14):148103.

52. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 2002 Nov;11(11):2714-26.

# APPENDIX C

# Supplementary Information for "A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering"

**Reproduced with permission from:**
Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J Am Chem Soc. 2010 May 5;132(17):6025-31.

**Collaborator notes:**
Robert Grant helped in solving the two SYNZIP crystal structures.

**SUPPLEMENTARY EXPERIMENTS**

A

```
SYNZIP peptides
SYNZIP1*   NLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE
SYNZIP2*   ARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASHEQ
SYNZIP3*   NEVTTLENDAAFIENENAYLEKEIARLRKEKAALRNRLAHKK
SYNZIP4*   QKVAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE
SYNZIP5*   NTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAAHKFE
SYNZIP6*   QKVAQLKNRVAYKLKENAKLENIVARLENDNANLEKDIANLEKDIANLERDVAR
SYNZIP7*   KEIEYLEKEIERLKDLREHLKQDNAAHRQELNALRLEEAKLEFILAHLLST
SYNZIP8*   KEIANLEKEIASLEKKVAVLKQRNAAHKQEVAALRKEIAYVEDEIQYVEDE
SYNZIP9*   QKVESLKQKIEELKQRKAQLKNDIANLEKEIAYAET
SYNZIP10*  NLLATLRSTAAVLENENHVLEKEKELRKEKEQLLNKLEAYK
SYNZIP11*  ELTDELKNKKEALRKDNAALLNELASLENEIANLEKEIAYFK
SYNZIP12*  NEDLVLENRLAALRNENAALENDLARLEKEIAYLEKEIEREK
SYNZIP13*  QKVEELKNKIAELENRNAVKKNRVAHLKQEIAYLKDELAAHEFE
SYNZIP14*  NDLDAYEREAEKLEKKNEVLRNRLAALENELATLRQEVASMKQELQS
SYNZIP15*  FENVTHEFILATLENENAKLRRLEAKLERELARLRNEVAWL
SYNZIP16*  NILASLENKKEELKKLNAHLLKEIENLEKEIANLEKEIAYFK
SYNZIP17*  NEKEELKSKKAELRNRIEQLKQKREQLKQKIANLRKEIEAYK
SYNZIP18*  SIAATLENDLARLENENARLEKDIANLERDLAKLEREEAYF
SYNZIP19*  NELESLENKKEELKNRNEELKQKREQLKQKLAALRNKLDAYKNRL
SYNZIP20*  STVEELLRAIQELEKRNAELKNRKEELKNLVAHLRQELAAHKYE
SYNZIP21*  NEVAQLENDVAVIENENAYLEKEIARLRKEIAALRDRLAHKK
SYNZIP22*  KRIAYLRKKIAALKKDNANLEKDIANLENEIERLIKEIKTLENEVASHEQ
SYNZIP23*  ALRAELKAKIALLRADNWALKRKAKDLRRLLRRLRNKAEELK
SYNZIP24   QKLQTLRDLLAVLENRNQELKQLRQHLKDLLKYLEDELATLEKE
SYNZIP25   NETEQLINKKEQLKNDNAALEKDAASLEKEIANLEKEIAYFK
SYNZIP26   EKIQELKRRLAYFRRENATLKNDNATLENELASVEAENEALRK
SYNZIP27   QKIQYLKQRIAELRKKIANLRKDIANLEDDAAVKEDELVHL
SYNZIP28   EKIEYLKDRIAELRSKIAALRNDLTHLKNDKAHKENELAHLA
SYNZIP29   NDIENLKDKIEELKQRKEELKQKIEYLKQKIEALRQKLAALKQRIA
SYNZIP30   EKIEELKDKIAELRSRNAALRNKIEALKQKLEALRQKIEYLKDRIA
SYNZIP31   AENQYVEDLIQYLEKENARLKKEVQRLVRELSYFRRRIAELA
SYNZIP32   AENQSVEDIIAKKEDENAHLKNEVKTLINELETLRKKIEYLA
SYNZIP33   RDLQNVEREIQSLEKKNESLKKKIASLENELATLKQEIAYFKRELAY
SYNZIP34   DRLAVKENRVAVLKNENAKLRNIIANLKDRIAYFRRELAYLELEEEQLA
SYNZIP35   NKVEQLKNKVEQLKNRNAALKNDLARLEREIAYAEE
SYNZIP36   EKNQELKNRLAVLENDNAALRNDLARLEREIAYME
SYNZIP37   KDIANLKKEIAHLKNDLQRLESIRERLKFDILNHEQEEYALE
SYNZIP38   NKNETLKNINARLRNDVARLKNRIARLKDDIENVEDEIQYLE
SYNZIP39   LENAQIKKEIAQLRKEVAQLKQKIEELKNDNARVEREIQYLE
SYNZIP40   QKRQQLKQKLAALRRDIENLQDEIAYKEDEIANLKDKIEQLLS
SYNZIP41   QKIESLKDKLANKRDKIALLRSEVASFEKEIAYLEKEIANLEN
SYNZIP42   EKIEYLKDKLAHKRNEVAQLRKEVTHKVDELTSLENEVAQLLK
SYNZIP43   QKVEQLKNKVEQKLKENESLENKVAELKNRNEYLKNKIENLINDITNLENDVAR
SYNZIP44   QKVAQLKNIIAKKEDENAVLENLVAVLENENAYLEKELARLERDIARAERDVKV
SYNZIP45   NRLQELENKNEVLEKRKAELRNEVATLEQELAAHRYELAAIEKEIA
SYNZIP46   KEIERLEKEIKTLINLLTTLRQDNAAHRKEAAALEKEEANLERDIQNLLRY
SYNZIP47   SKYDALRNKLEALKNRNAQLRKENEQLRLEEAVLEVRNEVL
SYNZIP48   QKIAYLRDRIAALKAENEALRAKNEALRSKIEELKKEKEELRDKIAQKKDR

Human peptides
BATF*      QKADTLHLESEDLEKQNAALRKEIKQLTEELKYFTSVLNSHE
FOS*       ELTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHR
ATF4*      AEQEALTGECKELEKKNEALKERADSLAKEIQYLKDLIEEVRKARGKKRV
ATF3       EKTECLQKESEKLESVNAELKAQIEELKNEKQHLIYMLNLHR
BACH1      DCIQNLESEIEKLQSEKESLLKERDHILSTLGETKQNLTGL
JUND       ERISRLEEKVKTLKSQNTELASTASLLREQVAQLKQKVLSHV
NFE2L3     DIILNLEDDVCNLQAKKETLKREQAQCNKAINIMKQKLHDL
Heptad     fgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabc
```

B

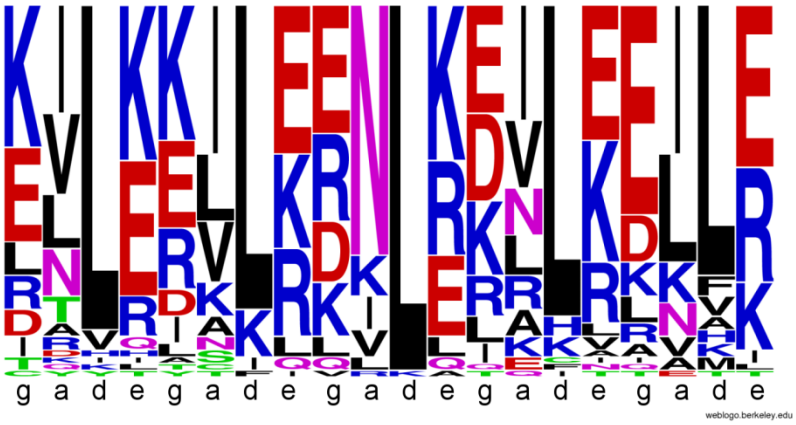


Figure C.S1. Sequences and sequence features of the 55 peptides measured. **(**A**)** Multiple-sequence alignment of the coiled-coil regions of the 55 peptides. Sequences start at an **f** position. Positions are colored as follows: **b**, **c,** and **f** positions (black), **g** (orange), **a** (blue), **d** (green), and **e** (purple). Peptides that form at least one hetero-specific interaction are indicated with an asterisk. **(**B**)** Sequence logo constructed using **a**, **d**, **e**, and **g** positions of the first 5 heptads of each peptide. See (Grigoryan, et al. 2009) for details. Sequence logo created with http://weblogo.berkeley.edu/ (Crooks, et al. 2004).


Figure C.S2. Array measurements for all 55 peptides. Peptides printed on the surface are listed in rows, and fluorescently labeled peptides in solution are listed in columns. Color indicates the strength of the array fluorescence signal, given as *arrayscore* values (see Methods) according to the color bar with 0 (black) indicating the strongest signal and >1 (white) indicating the weakest.

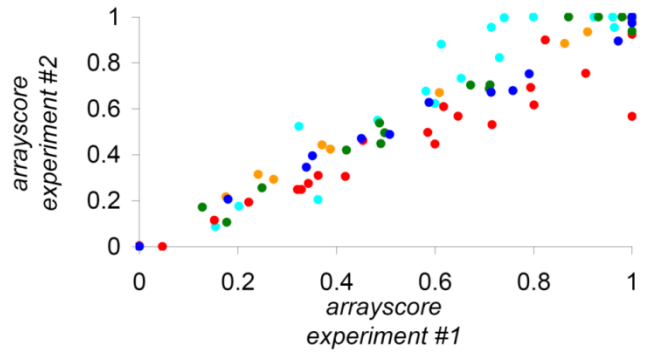



Figure C.S3. Reproducibility of the array experiments. Five solution probes measured in separate experiments are shown as a scatter plot. *Arrayscore* values >1 are set to 1. Blue, SYNZIP5 ($R^2$=.99). Orange, SYNZIP6 ($R^2$=.99). Teal, SYNZIP37 ($R^2$=.91). Red, FOS ($R^2$=.95). Green, ATF4 ($R^2$=.99).

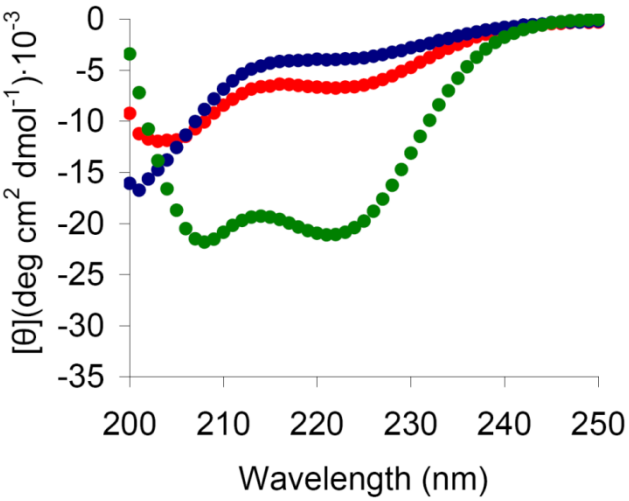



Figure C.S4. CD spectra for heterospecific pair SYNZIP6 + SYNZIP5. The mixture of SYNZIP5 with SYNZIP6 (4 μM each peptide) is in green**.** SYNZIP6 alone (4 μM) is in blue, SYNZIP5 alone (4 μM) is in red.

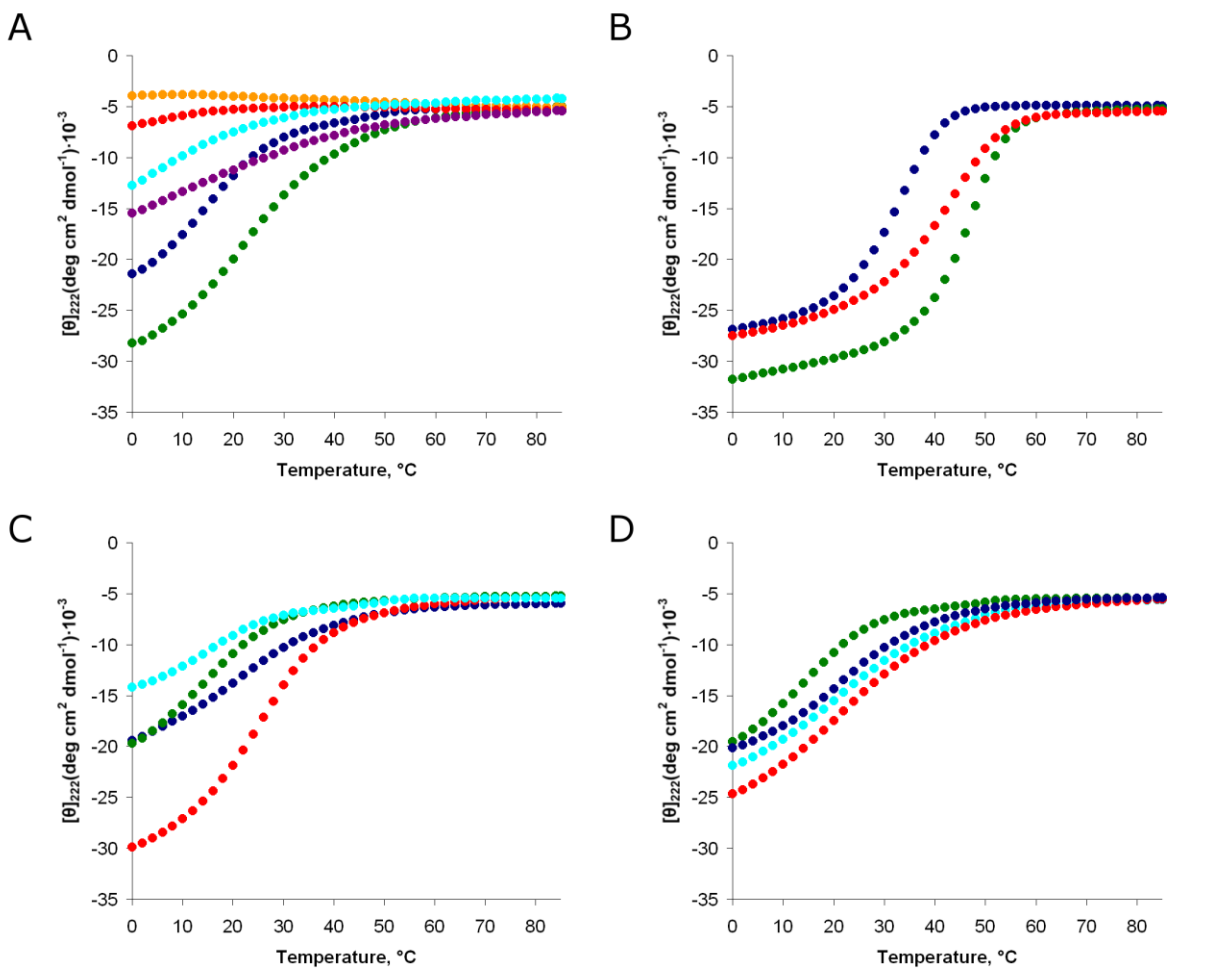



Figure C.S5. CD-monitored thermal melts of peptide pairs that form orthogonal sets. **(**A**)** Isolated peptides. ATF4-2 (green), SYNZIP1 (blue), SYNZIP3 (purple), SYNZIP5 (teal), SYNZIP4 (red), and SYNZIP6 (orange). **(**B**)** Interacting complexes**:** SYNZIP2 + SYNZIP1 (green), SYNZIP4 + SYNZIP3 (red), SYNZIP6 + SYNZIP5 (blue). **(**C**)** Non-interactions for orthogonal pair [SYNZIP2:SYNZIP1, SYNZIP6:SYNZIP5]: SYNZIP2 + SYNZIP5 (red), SYNZIP2 + SYNZIP6 (blue), SYNZIP1 + SYNZIP5 (green) + SYNZIP1 + SYNZIP6 (teal). **(**D**)** Non-interactions for orthogonal pair [SYNZIP2:SYNZIP1, SYNZIP4:SYNZIP3]: SYNZIP2 + SYNZIP3 (red), SYNZIP2 + SYNZIP4 (blue), SYNZIP1 + SYNZIP3 (teal) + SYNZIP1 + SYNZIP4 (green). Each individual peptide concentration was 4 μM, or 4 μM each (8 μM total peptide concentration) for mixtures.

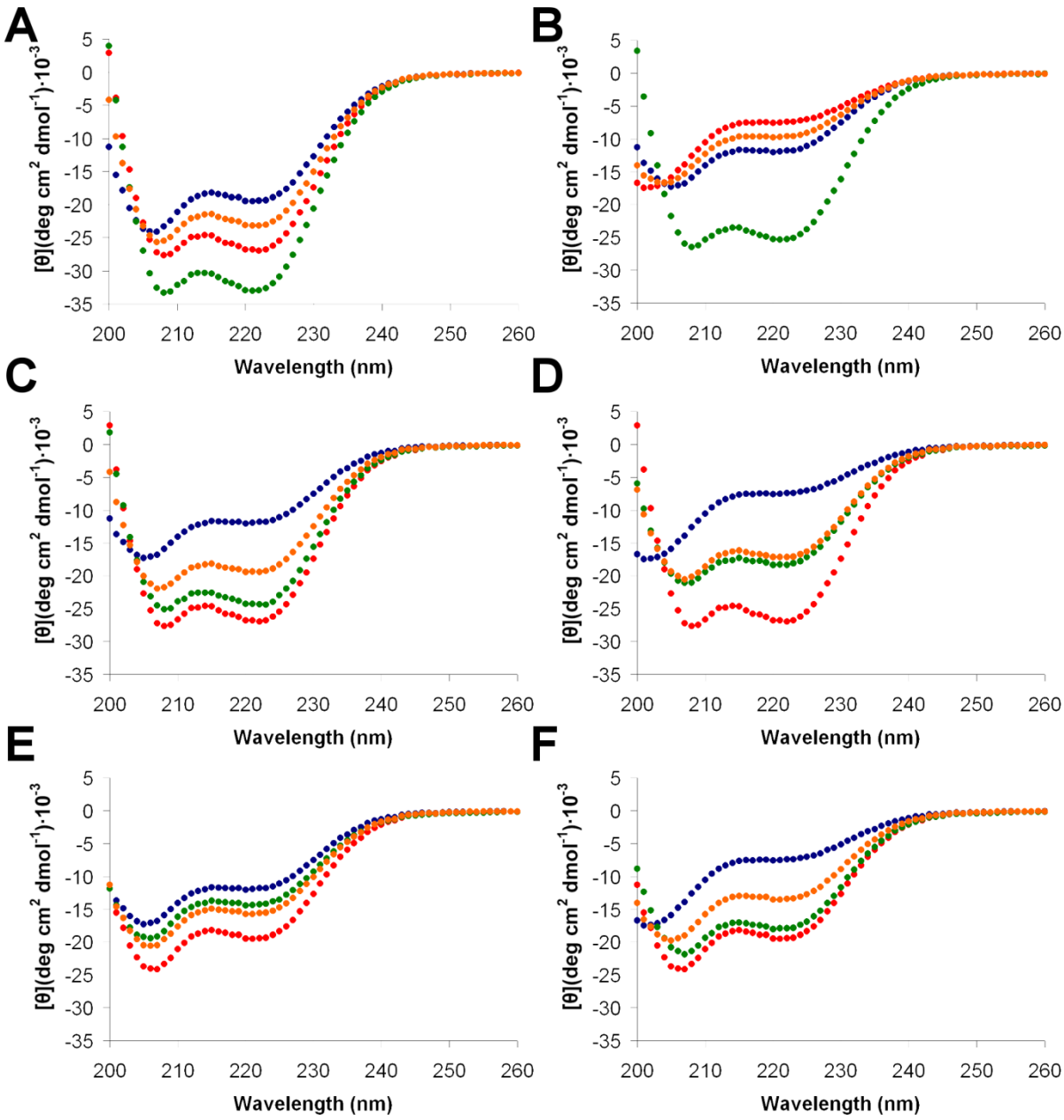



Figure C.S6. CD spectra characterizing an orthogonal set consisting of FOS:SYNZIP9 and SYNZIP3:SYNZIP4. (A, B) Characterization of 'on' interactions. **(**C-F**)** Characterization of 'off' interactions. **(**A**)** FOS (blue), SYNZIP9 (red), mixture of FOS + SYNZIP9 (green), and the mathematical average of the individual spectra (orange). **(**B**)** SYNZIP3 (blue), SYNZIP4 (red), mixture of SYNZIP3 + SYNZIP4 (green), and the average of the individual spectra (orange). **(**C**)** SYNZIP3 (blue), SYNZIP9 (red), mixture of SYNZIP3 + SYNZIP9 (green), and the average of the individual spectra (orange). **(**D**)** SYNZIP4 (blue), SYNZIP9 (red), mixture of SYNZIP4 + SYNZIP9 (green), and average of the individual spectra (orange). **(**E**)** SYNZIP3 (blue), FOS (red), mixture of SYNZIP3 + FOS (green), and average of the individual spectra (orange). **(**F**)** SYNZIP4 (blue), FOS (red), mixture of SYNZIP4 + FOS (green), and average of the individual spectra (orange). Spectra were measured at 25 °C at peptide concentrations of 40 µM or 20 µM of each peptide in mixtures (40 µM total peptide concentration).

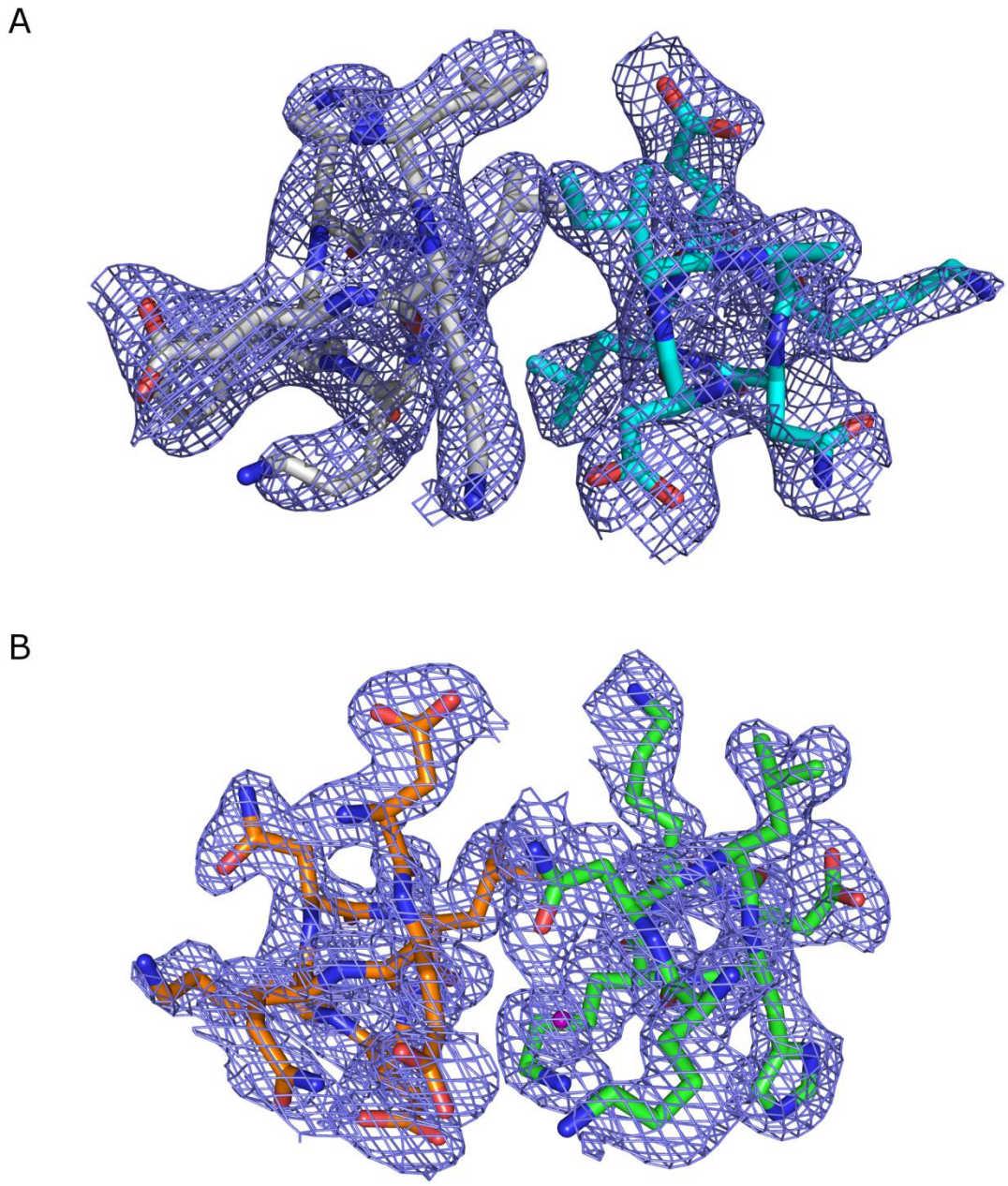



Figure C.S7. Electron density maps of SYNZIP5:SYNZIP6 and SYNZIP2:SYNZIP1. (A) The fourth heptad of SYNZIP5 (residues 23-29):SYNZIP6 (residues 37-43). (B) The fourth heptad of SYNZIP2 (residues 23-29):SYNZIP1(residues 23-29). These correspond to the heptads shown in Figure 3 G and H.

Table C.S1. Protein and DNA sequences used in this study.

| Proteins used in array assay. | | | |
|---|---|---|---|
| Name | Protein | DNA[e] | Source |
| SYNZIP1[a] | SYYHHHHHHLESTSLYKKAGSGSNLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE | GGATCCAACCTGGTTGCGCAGCTCGAAAACGAAGTTGCGTCTCTGGAAAATGAGAACGAAACCCTGAAGAAAAAGAACCTGCACAAAAAAGACCTGATCGCGTACCTGGAGAAAGAAATCGCGAATCTGCGTAAGAAAATCGAAGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP2[a] | SYYHHHHHHLESTSLYKKAGSGSARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASHEQ | GGATCCGCGCGTAACGCGTATCTGCGTAAGAAAATCGCACGTCTGAAAAAAGACAACCTGCAGCTGGAACGTGATGAACAGAACCTGGAAAAAATCATCGCGAACCTGCGTGACGAAATCGCGCGTCTCGAAAACGAAGTTGCGTCTCACGAACAGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP3[a] | SYYHHHHHHLESTSLYKKAGSGSNEVTTLENDAAFIENENAYLEKEIARLRKEKAALRNRLAHKK | GGATCCAACGAAGTTACCACTCTGGAGAATGACGCTGCGTTCATCGAAAATGAAAACGCTTACCTGGAAAAAGAAATCGCGCGTCTGCGTAAAGAAAAAGCGGCGCTGCGCAACCGTCTGGCGCACAAAAAATGATAACTCGAG | This study |
| SYNZIP4[a] | SYYHHHHHHLESTSLYKKAGSGSQKVAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE | GGATCCCAGAAAGTTGCGGAACTCAAAAACCGTGTTGCGGTTAAACTGAATCGTAACGAACAGCTGAAAAACAAAGTTGAAGAGCTGAAGAACCGTAACGCTTACCTCAAGAACGAACTGGCGACCCTGGAGAACGAGGTTGCGCGTCTGGAAAACGACGTTGCAGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP5[a] | SYYHHHHHHLESTSLYKKAGSGSNTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAAHKFE | GGATCCAACACCGTTAAAGAACTGAAAAACTACATCCAGGAGCTGGAAGAGCGTAACGCTGAACTCAAAAACCTGAAGGAACACCTGAAATTCGCAAAAGCGGAACTGGAATTCGAACTGGCGGCTCACAAATTCGAGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP6[a] | SYYHHHHHHLESTSLYKKAGSGSQKVAQLKNRVAYKLKENAKLENIVARLENDNANLEKDIANLEKDIANLERDVAR | GGATCCCAAAAAGTTGCGCAGCTGAAAAACCGTGTTGCGTACAAACTGAAAGAAAACGCGAAGCTGGAGAACATCGTGGCGCGTCTGGAAAACGACAATGCGAACCTGGAGAAAGACATTGCGAATCTCGAAAAGGACATCGCAAATCTGGAACGTGACGTTGCGCGTTGATAACTCGAG | Grigoryan, et al. 2009 |

| | | | |
|---|---|---|---|
| SYNZIP7[a] | SYYHHHHHHLESTSLY<br>KKAGSGSKEIEYLEKEI<br>ERLKDLREHLKQDNA<br>AHRQELNALRLEEAKL<br>EFILAHLLST | GGATCCAAAGAGATCGAATACCTGGAAAAA<br>GAAATTGAACGTCTGAAAGACCTGCGTGAA<br>CACCTGAAACAGGACAACGCGGCTCACCGT<br>CAGGAACTGAACGCGCTGCGTCTGGAAGAA<br>GCGAAACTGGAATTCATCCTGGCGCACCTG<br>CTGTCTACCTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP8[a] | SYYHHHHHHLESTSLY<br>KKAGSGSKEIANLEKEI<br>ASLEKKVAVLKQRNA<br>AHKQEVAALRKEIAY<br>VEDEIQYVEDE | GGATCCAAAGAGATCGCTAACCTGGAGAAA<br>GAAATTGCGTCTCTGGAAAAAAAGGTTGCG<br>GTTCTGAAACAGCGTAACGCTGCGCACAAA<br>CAGGAAGTTGCGGCTCTGCGTAAGGAAATC<br>GCTTACGTGGAGGACGAAATCCAGTACGTT<br>GAAGACGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP9[a] | SYYHHHHHHLESTSLY<br>KKAGSGSQKVESLKQ<br>KIEELKQRKAQLKNDI<br>ANLEKEIAYAET | GGATCCCAGAAGGTTGAATCTCTGAAACAG<br>AAAATCGAAGAACTGAAGCAGCGTAAAGC<br>GCAGCTGAAAAACGACATCGCGAACCTGGA<br>AAAAGAAATCGCGTATGCGGAAACCTGATA<br>ACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP10[a] | SYYHHHHHHLESTSLY<br>KKAGSGSNLLATLRST<br>AAVLENENHVLEKEK<br>EKLRKEKEQLLNKLEA<br>YK | GGATCCAACCTGCTGGCGACCCTGCGTTCT<br>ACCGCTGCGGTTCTGGAAAACGAAAACCAC<br>GTACTGGAGAAGGAGAAAGAGAAACTGCG<br>CAAAGAAAAAGAACAGCTGCTGAACAAAC<br>TGGAAGCGTACAAATGATAACTCGAG | This study |
| SYNZIP11[a] | SYYHHHHHHLESTSLY<br>KKAGSGSELTDELKNK<br>KEALRKDNAALLNEL<br>ASLENEIANLEKEIAYF<br>K | GGATCCGAACTGACCGATGAACTGAAAAAC<br>AAAAAAGAAGCTCTGCGTAAAGACAACGCT<br>GCGCTGCTGAACGAACTGGCGTCTCTGGAA<br>AACGAAATTGCGAACCTGGAGAAAGAAATC<br>GCGTACTTCAAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP12[a] | SYYHHHHHHLESTSLY<br>KKAGSGSNEDLVLENR<br>LAALRNENAALENDL<br>ARLEKEIAYLEKEIERE<br>K | GGATCCAATGAAGACCTGGTTCTGGAAAAC<br>CGCCTTGCGGCGCTGCGTAACGAAAACGCT<br>GCGCTTGAGAATGACCTGGCGCGTCTGGAG<br>AAAGAGATCGCGTACTTGGAGAAGGAAATC<br>GAACGTGAAAAATGATAACTCGAG | This study |
| SYNZIP13[a] | SYYHHHHHHLESTSLY<br>KKAGSGSQKVEELKN<br>KIAELENRNAVKKNRV<br>AHLKQEIAYLKDELAA<br>HEFE | GGATCCCAGAAAGTTGAAGAACTGAAAAAC<br>AAAATCGCGGAACTGGAAAACCGTAACGCG<br>GTTAAAAAGAACCGTGTTGCGCACCTGAAA<br>CAGGAAATCGCTTATCTGAAAGACGAACTG<br>GCGGCTCACGAATTTGAATGATAACTCGAG | Grigoryan, et al. 2009 |

| | | | |
|---|---|---|---|
| SYNZIP14[a] | SYYHHHHHHLESTSLYKKAGSGSNDLDAYEREAEKLEKKNEVLRNRLAALENELATLRQEVASMKQELQS | GGATCCAACGACCTGGACGCGTACGAACGTGAAGCGGAAAAACTGGAAAAGAAAAACGAAGTTCTGCGTAACCGTCTGGCGGCTCTCGAAAACGAGCTGGCGACCCTGCGTCAGGAAGTTGCGTCTATGAAACAGGAACTGCAATCTTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP15[a] | SYYHHHHHHLESTSLYKKAGSGSFENVTHEFILATLENENAKLRRLEAKLERELARLRNEVAWL | GGATCCTTTGAAAACGTTACCCACGAATTCATCCTGGCGACCCTGGAAAACGAAAACGCTAAACTGCGTCGTCTGGAAGCGAAACTGGAACGTGAACTGGCTCGTCTGCGTAACGAAGTTGCGTGGCTGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP16[a] | SYYHHHHHHLESTSLYKKAGSGSNILASLENKKEELKKLNAHLLKEIENLEKEIANLEKEIAYFK | GGATCCAACATCCTGGCGTCTCTCGAAAACAAAAAAGAAGAACTGAAAAAACTGAACGCGCACCTGCTGAAAGAAATCGAAAATCTGGAGAAAGAGATCGCAAACCTGGAAAAGGAAATCGCGTACTTCAAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP17[a] | SYYHHHHHHLESTSLYKKAGSGSNEKEELKSKKAELRNRIEQLKQKREQLKQKIANLRKEIEAYK | GGATCCAACGAAAAAGAAGAACTGAAATCCAAAAAAGCGGAACTGCGCAACCGTATCGAACAGCTGAAACAGAAACGTGAACAACTGAAGCAGAAAATCGCGAACCTGCGTAAAGAAATCGAAGCTTACAAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP18[a] | SYYHHHHHHLESTSLYKKAGSGSSIAATLENDLARLENENARLEKDIANLERDLAKLEREEAYF | GGATCCAGCATCGCGGCGACCCTGGAGAACGATCTGGCGCGTCTGGAAAACGAAAACGCTCGTCTCGAAAAAGACATCGCGAACCTGGAACGTGACCTGGCGAAACTGGAGCGTGAAGAAGCGTACTTCTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP19[a] | SYYHHHHHHLESTSLYKKAGSGSNELESLENKKEELKNRNEELKQKREQLKQKLAALRNKLDAYKNRL | GGATCCAACGAACTGGAATCTCTGGAGAACAAAAAAGAAGAACTGAAGAACCGTAACGAAGAGCTGAAGCAGAAACGTGAACAGCTGAAACAGAAACTGGCGGCTCTGCGTAACAAACTGGACGCGTACAAAAACCGTCTGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP20[a] | SYYHHHHHHLESTSLYKKAGSGSSTVEELLRAIQELEKRNAELKNRKEELKNLVAHLRQELAAHKYE | GGATCCAGCACTGTTGAAGAACTGCTGCGTGCGATCCAGGAGCTGGAAAAACGTAACGCGGAACTCAAAAACCGTAAAGAGGAACTGAAAAATCTGGTTGCGCACCTGCGTCAAGAGCTGGCAGCGCACAAATACGAATGATAACTCGAG | Grigoryan, et al. 2009 |

| | | | |
|---|---|---|---|
| SYNZIP21[a] | SYYHHHHHHLESTSLYKKAGSGSNEVAQLENDVAVIENENAYLEKEIARLRKEIAALRDRLAHKK | GGATCCAACGAAGTTGCGCAGCTGGAAAACGACGTTGCGGTTATCGAAAATGAAAACGCGTACCTGGAGAAGGAGATCGCGCGTCTGCGTAAAGAAATTGCGGCGCTGCGTGACCGTCTGGCGCACAAAAAATGATAACTCGAG | This study |
| SYNZIP22[a] | SYYHHHHHHLESTSLYKKAGSGSKRIAYLRKKIAALKKDNANLEKDIANLENEIERLIKEIKTLENEVASHEQ | GGATCCAAACGTATCGCGTACCTGCGTAAGAAAATCGCGGCACTGAAAAAAGACAACGCGAACCTCGAAAAAGATATCGCAAACCTGGAAAACGAAATCGAACGTCTGATCAAAGAAATCAAAACCCTGGAGAACGAAGTTGCGTCTCACGAACAGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP23[a] | SYYHHHHHHLESTSLYKKAGSGSALRAELKAKIALLRADNWALKRKAKDLRRLLRRLRNKAEELK | GGATCCGCACTCCGTGCGGAACTGAAAGCGAAAATCGCGCTCCTGCGTGCTGACAACTGGGCGCTGAAACGTAAAGCTAAAGACCTGCGTCGTCTGCTGCGCCGTCTGCGTAACAAAGCGGAAGAGCTGAAATGATAACTCGAG | This study |
| SYNZIP24[a] | SYYHHHHHHLESTSLYKKAGSGSQKLQTLRDLLAVLENRNQELKQLRQHLKDLLKYLEDELATLEKE | GGATCCCAGAAACTGCAGACCCTGCGTGATCTGCTGGCGGTTCTGGAGAACCGTAATCAGGAACTGAAACAGCTGCGTCAGCACCTGAAAGACCTGCTGAAATACCTGGAAGACGAACTGGCGACCCTGGAAAAAGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP25[a] | SYYHHHHHHLESTSLYKKAGSGSNETEQLINKKEQLKNDNAALEKDAASLEKEIANLEKEIAYFK | GGATCCAACGAAACCGAACAGCTGATCAACAAAAAAGAGCAGCTGAAAAACGACAACGCAGCGCTCGAAAAAGATGCGGCGTCTCTGGAAAAGGAAATCGCGAACCTGGAGAAAGAAATTGCGTACTTCAAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP26[a] | SYYHHHHHHLESTSLYKKAGSGSEKIQELKRRLAYFRRENATLKNDNATLENELASVEAENEALRK | GGATCCGAAAAAATCCAGGAACTGAAACGTCGTCTGGCGTACTTCCGTCGTGAAAACGCGACCCTGAAAAACGACAACGCTACCCTGGAGAACGAACTGGCGTCTGTTGAAGCGGAAAACGAAGCGCTGCGTAAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP27[a] | SYYHHHHHHLESTSLYKKAGSGSQKIQYLKQRIAELRKKIANLRKDIANLEDDAAVKEDELVHL | GGATCCCAGAAAATCCAGTACCTGAAACAGCGTATCGCGGAACTGCGTAAAAAGATTGCGAACCTGCGCAAAGACATCGCTAACCTGGAAGATGACGCTGCGGTTAAAGAAGACGAACTGGTTCACCTGTGATAACTCGAG | Grigoryan, et al. 2009 |

| | | | |
|---|---|---|---|
| SYNZIP28[a] | SYYHHHHHHLESTSLYKKAGSGSEKIEYLKDRIAELRSKIAALRNDLTHLKNDKAHKENELAHLA | GGATCCGAAAAAATCGAATACCTGAAAGACCGTATCGCGGAACTGCGTTCTAAAATCGCTGCGCTGCGTAACGACCTGACCCACCTGAAGAACGACAAAGCGCACAAAGAAAACGAACTGGCGCACCTGGCGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP29[a] | SYYHHHHHHLESTSLYKKAGSGSNDIENLKDKIEELKQRKEELKQKIEYLKQKIEALRQKLAALKQRIA | GGATCCAACGACATCGAAAACCTGAAAGACAAGATCGAAGAACTCAAACAGCGTAAAGAAGAGCTGAAACAGAAAATCGAATACCTCAAGCAGAAGATTGAAGCGCTGCGTCAGAAACTGGCGGCTCTGAAGCAGCGTATCGCGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP30[a] | SYYHHHHHHLESTSLYKKAGSGSEKIEELKDKIAELRSRNAALRNKIEALKQKLEALRQKIEYLKDRIA | GGATCCGAAAAAATCGAAGAACTGAAAGACAAAATCGCGGAACTGCGTTCTCGTAACGCTGCGCTGCGTAACAAAATTGAAGCGCTGAAACAGAAACTGGAAGCTCTGCGTCAGAAGATCGAATACCTCAAAGACCGTATCGCGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP31[a] | SYYHHHHHHLESTSLYKKAGSGSAENQYVEDLIQYLEKENARLKKEVQRLVRELSYFRRRIAELA | GGATCCGCTGAAAACCAGTACGTTGAAGACCTGATCCAGTACCTGGAAAAAGAGAACGCTCGTCTGAAAAAAGAAGTTCAGCGTCTGGTTCGTGAACTGTCTTACTTCCGTCGTCGTATCGCGGAACTGGCGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP32[a] | SYYHHHHHHLESTSLYKKAGSGSAENQSVEDIIAKKEDENAHLKNEVKTLINELETLRKKIEYLA | GGATCCGCTGAAAACCAGTCTGTTGAAGACATCATCGCGAAAAAAGAAGATGAAAACGCGCACCTGAAAAACGAAGTTAAAACCCTGATCAACGAACTGGAAACTCTGCGTAAGAAAATCGAATACCTGGCGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP33[a] | SYYHHHHHHLESTSLYKKAGSGSRDLQNVEREIQSLEKKNESLKKKIASLENELATLKQEIAYFKRELAY | GGATCCCGTGACCTGCAGAACGTTGAACGTGAAATCCAGTCCCTGGAAAAGAAAAACGAATCTCTGAAGAAGAAAATCGCTTCTCTGGAGAACGAACTGGCGACCCTGAAACAGGAAATCGCGTACTTCAAACGTGAGCTGGCTTACTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP34[a] | SYYHHHHHHLESTSLYKKAGSGSDRLAVKENRVAVLKNENAKLRNIIANLKDRIAYFRRELAYLELEEEQLA | GGATCCGACCGTCTGGCGGTTAAAGAAAACCGTGTTGCGGTTCTGAAAAACGAAAACGCGAAACTGCGTAACATCATCGCGAACCTGAAAGACCGTATCGCGTACTTCCGTCGTGAACTGGCGTACCTGGAACTGGAAGAAGAACAGCTGGCGTGATAACTCGAG | Grigoryan, et al. 2009 |

| | | | |
|---|---|---|---|
| SYNZIP35[a] | SYYHHHHHHLESTSLYKKAGSGSNKVEQLKNKVEQLKNRNAALKNDLARLEREIAYAEE | GGATCCAACAAGGTTGAGCAGCTCAAAAACAAAGTTGAACAGCTGAAAAACCGTAACGCTGCGCTGAAGAACGACCTGGCGCGTCTGGAACGTGAAATCGCGTATGCGGAAGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP36[a] | SYYHHHHHHLESTSLYKKAGSGSEKNQELKNRLAVLENDNAALRNDLARLEREIAYME | GGATCCGAAAAAAACCAGGAACTGAAAAACCGTCTGGCGGTTCTGGAAAACGACAACGCTGCTCTGCGTAACGACCTGGCGCGTCTGGAACGTGAAATCGCGTACATGGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP37[a] | SYYHHHHHHLESTSLYKKAGSGSKDIANLKKEIAHLKNDLQRLESIRERLKFDILNHEQEEYALE | GGATCCAAAGACATCGCGAACCTCAAAAAAGAAATCGCGCACCTGAAAAACGACCTGCAGCGTCTGGAATCTATCCGTGAACGTCTGAAATTCGACATTCTGAACCACGAACAGGAAGAATACGCACTGGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP38[a] | SYYHHHHHHLESTSLYKKAGSGSNKNETLKNINARLRNDVARLKNRIARLKDDIENVEDEIQYLE | GGATCCAACAAAAACGAAACTCTGAAGAACATCAACGCACGTCTGCGTAACGATGTTGCTCGTCTCAAAAACCGTATCGCGCGTCTGAAAGACGACATCGAAAACGTTGAAGACGAAATCCAGTACCTGGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP39[a] | SYYHHHHHHLESTSLYKKAGSGSLENAQIKKEIAQLRKEVAQLKQKIEELKNDNARVEREIQYLE | GGATCCCTGGAAAACGCTCAGATCAAAAAAGAAATCGCTCAGCTGCGTAAAGAAGTTGCACAGCTGAAACAGAAAATCGAAGAACTGAAAAACGATAACGCACGTGTTGAACGTGAAATCCAGTACCTGGAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP40[a] | SYYHHHHHHLESTSLYKKAGSGSQKRQQLKQKLAALRRDIENLQDEIAYKEDEIANLKDKIEQLLS | GGATCCCAGAAACGTCAGCAACTGAAACAGAAACTGGCGGCTCTGCGTCGTGACATCGAAAACCTGCAAGATGAAATCGCGTACAAAGAAGACGAAATTGCGAACCTGAAAGACAAAATCGAACAGCTGCTGTCTTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP41[a] | SYYHHHHHHLESTSLYKKAGSGSQKIESLKDKLANKRDKIALLRSEVASFEKEIAYLEKEIANLEN | GGATCCCAGAAAATCGAATCTCTGAAAGACAAACTGGCGAACAAACGTGACAAAATCGCGCTGCTGCGTTCTGAAGTTGCGTCTTTTGAAAAAGAAATCGCATACCTGGAGAAAGAGATCGCAAACCTGGAAAACTGATAACTCGAG | Grigoryan, et al. 2009 |

| | | | |
|---|---|---|---|
| SYNZIP42[a] | SYYHHHHHHLESTSLYKKAGSGSEKIEYLKDKLAHKRNEVAQLRKEVTHKVDELTSLENEVAQLLK | GGATCCGAAAAAATCGAATACCTGAAAGACAAACTGGCGCACAAACGTAACGAAGTTGCTCAGCTGCGTAAAGAAGTTACCCACAAAGTTGACGAACTGACCTCTCTGGAAAACGAGGTTGCACAGCTGCTGAAATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP43[a] | SYYHHHHHHLESTSLYKKAGSGSQKVEQLKNKVEQKLKENESLENKVAELKNRNEYLKNKIENLINDITNLENDVAR | GGATCCCAGAAAGTGGAACAGCTGAAGAACAAGGTTGAACAGAAACTGAAAGAGAACGAGTCTCTGGAGAACAAAGTTGCGGAGCTGAAAAACCGTAACGAGTACCTCAAAAACAAAATCGAGAACCTGATCAACGACATCACCAACCTGGAAAACGACGTTGCGCGTTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP44[a] | SYYHHHHHHLESTSLYKKAGSGSQKVAQLKNIIAKKEDENAVLENLVAVLENENAYLEKELARLERDIARAERDVKV | GGATCCCAGAAAGTTGCGCAGCTGAAAAACATCATCGCGAAAAAAGAAGATGAGAACGCTGTTCTGGAAAACCTGGTTGCGGTGCTGGAGAACGAAAACGCGTACCTCGAAAAGGAACTGGCGCGTCTGGAACGCGACATCGCGCGTGCGGAACGTGATGTTAAAGTTTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP45[a] | SYYHHHHHHLESTSLYKKAGSGSNRLQELENKNEVLEKRKAELRNEVATLEQELAAHRYELAAIEKEIA | GGATCCAACCGTCTGCAGGAACTGGAAAACAAAAACGAGGTTCTGGAGAAACGTAAAGCGGAACTGCGCAACGAAGTTGCGACCCTGGAACAGGAGCTGGCTGCGCACCGTTACGAACTGGCGGCGATCGAAAAAGAAATCGCATGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP46[a] | SYYHHHHHHLESTSLYKKAGSGSKEIERLEKEIKTLINLLTTLRQDNAAHRKEAAALEKEEANLERDIQNLLRY | GGATCCAAAGAAATCGAACGTCTGGAAAAAGAGATCAAAACCCTGATCAACCTCCTGACCACCCTGCGTCAGGACAACGCGGCACACCGTAAAGAAGCAGCGGCACTGGAGAAAGAAGAAGCGAACCTGGAACGTGACATCCAGAACCTGCTGCGTTACTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP47[a] | SYYHHHHHHLESTSLYKKAGSGSSKYDALRNKLEALKNRNAQLRKENEQLRLEEAVLEVRNEVL | GGATCCAGCAAATACGACGCGCTGCGTAACAAACTGGAAGCGCTGAAAAACCGTAACGCGCAGCTCCGTAAAGAAAACGAACAGCTGCGTCTGGAAGAAGCGGTTCTGGAGGTTCGTAACGAAGTTCTGTGATAACTCGAG | Grigoryan, et al. 2009 |
| SYNZIP48[a] | SYYHHHHHHLESTSLYKKAGSGSQKIAYLRDRIAALKAENEALRAKNEALRSKIEELKKEKEELRDKIAQKKDR | GGATCCCAGAAAATTGCGTACCTGCGTGATCGTATCGCGGCACTGAAAGCTGAAAACGAAGCTCTGCGTGCGAAAAATGAAGCGCTGCGTTCTAAAATCGAGGAACTGAAGAAAGAAAAAGAAGAACTGCGCGACAAAATCGCTCAGAAAAAAGACCGTTGATAACTCGAG | Grigoryan, et al. 2009 |

| | | |
|---|---|---|
| BATF[b] | SYYHHHHHHLESTSLYKKAGSGSQKADTLHLESEDLEKQNAALRKEIKQLTEELKYFTSVLNSHELE | Newman, et al. 2003 |
| FOS[b] | SYYHHHHHHLESTSLYKKAGSEFFRRERNKMAAAKCRNRRRELTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDDLGFPEEMSLE | Newman, et al. 2003 |
| ATF4[b] | SYYHHHHHHLESTSLYKKAGSEFNKTAATRYRQKKRAEQEALTGECKELEKKNEALKERADSLAKEIQYLKDLIEEVRKARGKKRVP | Newman, et al. 2003 |
| ATF3[b] | SYYHHHHHHLESTSLYKKAGSGSCPEEDERKKRRRERNKIAAAKCRNKKKEKTECLQKESEKLESVNAELKAQIEELKNEKQHLIYMLNLHRPTCIVRAQNGRTPEDLE | Newman, et al. 2003 |
| BACH1[b] | SYYHHHHHHLESTSLYKKAGSEFGCRKRKLDCIQNLESEIEKLQSEKESLLKERDHILSTLGETKQNLTGLCQKVCKEAALSQEQNLE | Newman, et al. 2003 |
| JUND[b] | SYYHHHHHHLESTSLYKKAGSGSERISRLEEKVKTLKSQNTELASTASLLREQVAQLKQKVLSHVLE | Newman, et al. 2003 |
| NFE2L3[b] | SYYHHHHHHLESTSLYKKAGSEFGDIRRRGKNKVAAQNCRKRKLDIILNLEDDVCNLQAKKETLKREQAQCNKAINIMKQKLHDLYHDIFSRLRDDQGRPVLE | Newman, et al. 2003 |

| | | | |
|---|---|---|---|
| Proteins used in circular dichroism and crystallography studies. | | | |
| SYNZIP1[c] | GSNLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE | This study | |
| SYNZIP2[c] | GSARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASH EQ | This study | |
| SYNZIP3[c] | GSNEVTTLENDAAFIENENAYLEKEIARLRKEKAALRNRLAHKK | This study | |
| SYNZIP4[c] | GSQKVAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLENEVA RLENDVAE | This study | |
| SYNZIP5[c] | GSNTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAAHKFE | This study | |
| SYNZIP6[c] | GSQKVAQLKNRVAYKLKENAKLENIVARLENDNANLEKDIANLEKDIAN LERDVAR | This study | |

| | | | | |
|---|---|---|---|---|
| SYNZIP9[c] | GSQKVESLKQKIEELKQRKAQLKNDIANLEKEIAYAET | | This study | |
| FOS LZ[c] | GSELTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHR | GGATCCGAACTGACCGACACTCTGCAGGCGGAAACCGACCAGCTCGAAGATGAAAAATCTGCGCTGCAGACCGAAATCGCGAACCTGCTGAAAGAAAAAGAGAAACTGGAATTCATCCTGGCTGCTCACCGTTGATAACTCGAG | This study | |
| SYNZIP4(1-42)[c] | GSQKVAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLE | | This study | |
| SYNZIP4(15-54)[c] | GSNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE | | This study | |
| Proteins used in pull-down assays. | | | | |
| SYNZIP1(1-47)[d] | GSCGSNLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE | | This study | |
| SYNZIP2(1-47)[d] | GSCGSARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVAS | | This study | |

| | | | |
|---|---|---|---|
| SYNZIP3(1-40)[d] | GSCGSNEVTTLENDAAF IENENAYLEKEIARLRKEKAALRNRLAH | This study | |
| SYNZIP4(15-54)[d] | GSCGSNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE | This study | |
| SYNZIP5(1-40)[d] | GSCGSNTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAA | This study | |
| SYNZIP6(15-54)[d] | GSCGSKENAYLENIVARLENDNANLEKDIANLEKDIANLERDVAR | This study | |

[a] SYNZIP protein constructs used for array measurments have the following linker at the N-Terminus including the BamHI site: SYYHHHHHHLESTSLYKKAGSGS

[b] The coiled-coil region of the human sequences is in green. Additional human protein sequence is in red. Cloning sequence is in black.

[c] Constructs used for circular dichroism and crystallography studies include a GS at the N-Terminus after cleavage by TEV.

[d] Constructs used in pull-down assays include a GS at the N-Terminus after cleavage by TEV and a cysteine followed by a short GS linker.

[e] DNA sequence is of the insert and includes BamHI and XhoI sites that were used for cloning. DNA sequence for proteins used in array studies is the same for the proteins used in other assays unless otherwise indicated.

Table C.S2. Average background-corrected fluorescence values from the array experiment. Peptides in solution are in columns and those on the surface are in rows. Duplicates are indicated with a number 2 in parentheses.

| Protein | SYNZIP1 | SYNZIP2 | SYNZIP3 | SYNZIP4 | SYNZIP5 | SYNZIP6 | SYNZIP7 | SYNZIP8 | SYNZIP9 | SYNZIP10 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP1 | 2727.0 | 47850.8 | 605.9 | 1205.6 | 22.9 | 199.5 | 4675.4 | 299.5 | 38.0 | -30.5 |
| SYNZIP2 | 49531.5 | 4907.1 | 1023.0 | 1079.1 | 224.6 | 491.0 | 3565.9 | 3800.1 | 1065.6 | -103.4 |
| SYNZIP3 | -484.4 | 1852.9 | -583.3 | 34221.4 | 7000.9 | 99.4 | -113.8 | 1257.0 | 467.9 | -591.9 |
| SYNZIP4 | -314.8 | 2693.8 | 36766.9 | -52.5 | -8.3 | 15580.8 | 210.1 | 1145.5 | -220.3 | 2395.4 |
| SYNZIP5 | 85.3 | 7669.4 | 9659.6 | 2677.5 | -176.3 | 35912.5 | 118.0 | 4261.3 | 1387.8 | 712.8 |
| SYNZIP6 | -898.5 | 8038.1 | 1564.0 | 40661.4 | 37774.0 | -90.5 | -143.1 | 720.1 | 3008.5 | 4376.5 |
| SYNZIP7 | 3491.8 | 6240.3 | -319.3 | 285.4 | -309.1 | -559.8 | 281.8 | 2372.4 | -118.8 | 135.4 |
| SYNZIP8 | -26.3 | 7685.3 | 1528.5 | 1877.3 | 452.3 | -27.0 | 3148.9 | 1916.0 | 3176.3 | 5871.5 |
| SYNZIP9 | -178.8 | 3007.3 | 1339.8 | -612.9 | 854.6 | 1983.5 | 58.5 | 4728.9 | 653.9 | 4627.0 |
| SYNZIP10 | 310.3 | 5178.4 | 211.6 | 12234.0 | 1521.5 | 4070.0 | 1541.6 | 17712.1 | 7305.1 | -601.0 |
| SYNZIP11 | 6169.8 | 4973.8 | 5287.0 | 10656.3 | 7699.0 | 497.4 | 1255.1 | 1310.3 | 5522.8 | 766.3 |
| SYNZIP12 | 359.0 | 5306.0 | 4520.8 | 9755.4 | 13703.5 | 288.4 | 295.3 | 911.6 | 4612.0 | 2802.9 |
| SYNZIP13 | -345.9 | 32090.0 | 10973.8 | -2270.0 | 360.9 | 1113.0 | -74.5 | 167.9 | -653.5 | 12047.6 |
| SYNZIP14 | 308.1 | 37913.3 | 7528.3 | 5078.0 | 686.9 | 3957.0 | 194.5 | 1646.5 | 1597.3 | -41.9 |
| SYNZIP15 | 608.5 | 409.9 | 2844.3 | -1704.5 | -328.4 | 516.0 | -207.8 | -233.0 | -1521.0 | -1778.5 |
| SYNZIP16 | 1200.8 | 14994.5 | 5220.3 | 4013.3 | 17575.6 | 199.6 | 538.1 | 1751.4 | 3404.1 | 2502.8 |
| SYNZIP17 | 26.8 | 9431.3 | 14422.1 | 1381.8 | 759.1 | 5368.6 | 300.9 | 4604.0 | 136.1 | 1475.5 |
| SYNZIP18 | -424.4 | 2586.3 | -4.3 | 23264.6 | 12069.0 | -291.6 | -150.5 | 1529.4 | 9335.6 | -397.8 |
| SYNZIP19 | 5755.8 | 45176.9 | 2068.1 | 2440.9 | -115.9 | 20840.0 | 680.1 | 11074.8 | 2522.1 | -532.1 |
| SYNZIP20 | 2040.5 | 49162.4 | 6805.8 | 3834.1 | -348.5 | 19414.9 | 3515.0 | 2430.6 | 2857.6 | 6797.8 |
| SYNZIP21 | -563.0 | 6872.3 | 2657.1 | 30320.8 | 25120.1 | -154.9 | 4659.9 | 4335.9 | 1424.1 | 13937.8 |
| SYNZIP22 | 20784.0 | 19647.1 | 3614.3 | 2006.1 | 17091.3 | 392.4 | 6791.8 | 11137.5 | 1996.0 | 27744.9 |
| SYNZIP23 | 4067.9 | 20087.8 | 2195.3 | 13143.9 | 491.8 | 24128.6 | 2195.0 | 50967.8 | 16428.9 | 242.6 |
| SYNZIP24 | -757.1 | 46231.9 | -287.0 | -2798.0 | -794.3 | 649.1 | -352.4 | 3943.9 | -546.4 | -624.5 |
| SYNZIP25 | 1703.0 | -228.9 | -495.8 | 4192.5 | 4468.4 | -290.3 | -310.8 | -304.9 | 244.8 | -391.4 |
| SYNZIP26 | 243.3 | 157.0 | -176.8 | 348.1 | -292.8 | 2291.9 | 32.6 | 734.3 | -365.4 | -740.6 |
| SYNZIP27 | 389.1 | 5852.0 | 855.1 | 5952.4 | 2178.8 | 1826.3 | 1518.3 | 3726.3 | 988.3 | 313.0 |
| SYNZIP28 | -451.6 | 6751.8 | 14.1 | 2456.4 | 331.0 | 2174.5 | 822.1 | 2946.0 | 482.8 | -569.5 |

| Protein | SYNZIP11 | SYNZIP12 | SYNZIP13 | SYNZIP14 | SYNZIP15 | SYNZIP16 | SYNZIP17 | SYNZIP18 | SYNZIP19 | SYNZIP20 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP1 | 3129.4 | 208.9 | 674.0 | 460.6 | 1026.0 | 2663.5 | 1067.3 | -239.8 | 5059.6 | 4305.8 |
| SYNZIP2 | 602.4 | 1427.1 | 17562.1 | 42445.5 | -2040.4 | 16600.4 | 5490.0 | 306.0 | 24005.9 | 33929.3 |
| SYNZIP3 | 4448.8 | 6912.4 | 14412.8 | 6351.1 | 6177.4 | 11614.3 | 15054.1 | -255.3 | 1281.9 | 16144.0 |
| SYNZIP4 | 489.6 | 3136.9 | -167.4 | 2033.0 | -1616.4 | 6396.4 | 501.5 | 12474.3 | 1452.5 | 8718.3 |
| SYNZIP5 | 12771.6 | 22853.4 | 2599.4 | 2408.9 | -638.5 | 33866.9 | 5001.6 | 19688.1 | -356.3 | 953.3 |
| SYNZIP6 | 1258.3 | 2618.0 | 4032.5 | 8001.0 | 4907.5 | 1410.1 | 19399.5 | -173.1 | 44070.8 | 33644.3 |
| SYNZIP7 | -227.5 | -838.5 | 179.0 | -841.0 | -2143.1 | -879.5 | 115.0 | -773.4 | 377.1 | 5980.6 |
| SYNZIP8 | -138.4 | -43.6 | -283.1 | -37.1 | -307.6 | 2545.9 | 2857.4 | 233.4 | 4890.4 | 3113.5 |
| SYNZIP9 | 1250.9 | 2207.6 | -836.0 | 1389.6 | 1595.3 | 5740.9 | 1029.8 | 5380.9 | 2561.5 | 6657.9 |
| SYNZIP10 | 2878.3 | 3540.0 | 22523.8 | 4706.6 | -1154.4 | 16454.4 | 4745.5 | 1651.8 | -286.4 | 17169.3 |
| SYNZIP11 | 2033.9 | 122.8 | 3307.6 | 8680.3 | 4114.5 | 2228.9 | 10994.0 | 934.3 | 23409.5 | 30053.0 |
| SYNZIP12 | 99.1 | -367.3 | 21923.0 | 21412.9 | 6919.6 | 3774.3 | 15590.8 | 1723.0 | 24047.9 | 22937.9 |
| SYNZIP13 | 270.8 | 14448.4 | -785.0 | 1039.0 | 35569.3 | 12968.1 | -1576.9 | 10401.1 | -1291.1 | -849.0 |
| SYNZIP14 | 4518.1 | 21521.9 | 2938.5 | 3194.6 | 20648.3 | 16603.8 | 22442.5 | 10510.9 | 2516.8 | 3112.0 |
| SYNZIP15 | 1994.3 | 4332.4 | 36848.6 | 13204.5 | -1832.3 | 11318.8 | 14731.9 | 537.6 | 1312.0 | 834.8 |
| SYNZIP16 | 564.0 | 1005.4 | 8469.8 | 7710.0 | 5484.4 | 9851.5 | 12009.5 | 2442.3 | 15363.3 | 30927.3 |
| SYNZIP17 | 4598.0 | 9530.8 | 382.3 | 12973.0 | 31507.6 | 16546.9 | 2430.9 | 29468.9 | 51.5 | 4942.6 |
| SYNZIP18 | 2356.9 | 2288.9 | 17480.3 | 10362.9 | 2825.4 | 10579.0 | 34443.3 | 195.0 | 34738.6 | 19722.3 |
| SYNZIP19 | 15415.9 | 25068.0 | 955.0 | 470.0 | -974.5 | 38380.9 | 2777.3 | 45298.1 | -969.0 | -607.1 |
| SYNZIP20 | 41208.5 | 43431.4 | 628.3 | 641.1 | 1738.3 | 39021.5 | 5128.9 | 43702.4 | -806.3 | 5974.6 |
| SYNZIP21 | 19684.5 | 9979.8 | 20943.3 | 15840.0 | 27662.0 | 24776.9 | 22328.3 | 3974.3 | 36175.3 | 22019.4 |
| SYNZIP22 | 4447.5 | 6560.8 | 13293.5 | 44903.9 | 22642.0 | 17881.1 | 8033.3 | 5696.3 | 39174.6 | 29506.9 |
| SYNZIP23 | 16873.3 | 34517.6 | 4931.1 | 23576.8 | 26741.0 | 48678.0 | 12428.3 | 37390.1 | -662.5 | 14160.9 |
| SYNZIP24 | 1719.3 | 8750.4 | -639.8 | -1220.8 | 325.4 | 21574.3 | -1315.0 | 2602.5 | -1666.1 | -2221.8 |
| SYNZIP25 | -351.8 | -629.1 | 1032.6 | 510.9 | -2561.6 | -1163.6 | 1746.8 | -606.8 | 12561.6 | 20570.9 |
| SYNZIP26 | 1319.8 | 150.6 | 1725.4 | -95.6 | -402.1 | 1.4 | -368.6 | 12.6 | 1905.4 | 3065.8 |
| SYNZIP27 | 5058.0 | 2295.8 | 501.8 | 5619.5 | 15248.9 | 9462.1 | 3089.6 | 1459.4 | 629.0 | 6337.5 |
| SYNZIP28 | 4881.4 | 945.3 | -94.9 | 2150.8 | 8602.9 | 11086.9 | 1900.8 | 744.0 | -1211.1 | -564.6 |

| Protein | SYNZIP21 | SYNZIP22 | SYNZIP23 | SYNZIP24 | SYNZIP25 | SYNZIP26 | SYNZIP27 | SYNZIP28 | SYNZIP29 | SYNZIP30 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP1 | 1348.8 | 13166.0 | 11992.6 | 2089.1 | 571.4 | 741.9 | 985.3 | 392.4 | 2686.8 | 5131.9 |
| SYNZIP2 | 7554.8 | 5972.3 | 20889.4 | 31406.8 | 114.3 | 581.1 | 8668.3 | 18089.3 | 8859.1 | 5370.0 |
| SYNZIP3 | 9319.3 | 1659.8 | 10970.8 | 2170.9 | -27.9 | 243.9 | 170.6 | 12.0 | 1133.0 | 1757.5 |
| SYNZIP4 | 38542.3 | 4748.5 | 33133.0 | -997.5 | -209.3 | 799.6 | 6243.8 | 2106.1 | 2143.0 | 6827.6 |
| SYNZIP5 | 33212.9 | 21966.1 | 3954.5 | 1907.4 | 2968.6 | 738.9 | 19386.3 | 7975.8 | 6678.5 | 7283.1 |
| SYNZIP6 | 526.5 | -1250.0 | 41241.5 | 16534.5 | 68.4 | 165.6 | 5514.3 | 8949.3 | 13667.1 | 14659.5 |
| SYNZIP7 | 9034.1 | 3329.9 | 2323.5 | -861.0 | -217.1 | 552.9 | 5001.4 | 3690.3 | 2588.6 | 5274.6 |
| SYNZIP8 | 8510.0 | 3386.6 | 32177.9 | 8915.8 | -123.4 | 407.9 | 7415.3 | 9124.1 | 5094.0 | 6270.1 |
| SYNZIP9 | 4588.6 | -214.6 | 29461.3 | 644.4 | 207.8 | 558.1 | 6560.5 | 7202.9 | 2679.0 | 224.3 |
| SYNZIP10 | 28564.3 | 22701.6 | 2167.0 | 465.4 | 455.0 | 686.0 | 4247.4 | 239.1 | 10781.9 | 7398.9 |
| SYNZIP11 | 32459.1 | 3536.0 | 34355.9 | 12299.8 | 584.3 | 2561.1 | 22745.5 | 36794.6 | 23519.8 | 21946.9 |
| SYNZIP12 | 16377.6 | 3419.5 | 33987.9 | 15220.6 | -41.3 | 862.1 | 14693.3 | 13717.1 | 12273.5 | 13080.4 |
| SYNZIP13 | 24996.9 | 9577.1 | 9489.6 | -210.5 | -160.8 | 558.8 | -1246.1 | -706.8 | 3881.1 | 2904.6 |
| SYNZIP14 | 26253.5 | 21724.0 | 36041.9 | -114.0 | 204.3 | 970.6 | 11060.1 | 6766.0 | 8380.4 | 11584.4 |
| SYNZIP15 | 35396.0 | 7230.5 | 17382.3 | 5704.0 | 217.5 | 963.3 | 26415.9 | 19114.1 | 9017.8 | 8936.1 |
| SYNZIP16 | 27027.5 | 7271.8 | 42243.6 | 21729.1 | -83.8 | 602.9 | 15216.6 | 20446.1 | 9539.0 | 10450.5 |
| SYNZIP17 | 35158.8 | 1022.4 | 22352.0 | 3058.1 | 403.5 | 750.9 | 5285.8 | 7445.3 | 2337.8 | 1693.5 |
| SYNZIP18 | 9645.8 | 1627.8 | 33912.3 | 10686.9 | -1031.1 | 533.1 | 14059.4 | 9643.9 | 12798.0 | 16171.6 |
| SYNZIP19 | 49971.5 | 31356.6 | 1179.3 | -182.4 | 2028.3 | 958.3 | 2979.1 | 34.9 | 5165.9 | 6621.8 |
| SYNZIP20 | 36110.6 | 15364.5 | 18270.1 | 1188.9 | 8618.6 | 1781.6 | 15086.0 | -648.8 | 1156.8 | 2391.3 |
| SYNZIP21 | 5635.0 | 3822.4 | 38270.0 | 21588.3 | 572.8 | 632.1 | 220.3 | 766.9 | 2907.3 | 2997.3 |
| SYNZIP22 | 16302.6 | 1366.0 | 33168.3 | 22768.5 | 631.6 | 961.1 | 7654.8 | 18178.3 | 8721.3 | 11848.6 |
| SYNZIP23 | 46508.3 | 18793.4 | -635.3 | 9376.5 | 1206.6 | 2496.4 | 126.5 | -1312.8 | 3667.1 | 2347.3 |
| SYNZIP24 | 42194.3 | 8413.8 | 420.8 | 10344.3 | 70.1 | -32.1 | 3985.3 | -221.5 | -11.9 | -343.6 |
| SYNZIP25 | 1140.3 | -993.6 | 7693.1 | 2440.8 | -70.5 | 377.1 | 4227.5 | 7108.3 | 15053.8 | 13857.0 |
| SYNZIP26 | 3487.4 | -1153.1 | 9555.6 | -571.5 | 3731.3 | 1079.4 | -2087.3 | -1663.6 | 700.4 | 895.3 |
| SYNZIP27 | 1197.6 | -120.0 | 1155.8 | 5261.0 | 373.0 | 946.3 | 15214.5 | 11332.9 | 2437.9 | 1684.1 |
| SYNZIP28 | 715.1 | -157.5 | -1390.5 | 1587.6 | 245.6 | 530.3 | 10344.4 | 4448.1 | 931.0 | 521.3 |

| Protein | SYNZIP31 | SYNZIP32 | SYNZIP33 | SYNZIP34 | SYNZIP35 | SYNZIP36 | SYNZIP37 | SYNZIP38 | SYNZIP39 | SYNZIP40 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP1 | 1478.4 | 38.5 | 305.8 | 549.5 | 446.1 | 412.6 | 710.5 | 541.3 | 429.3 | 1523.1 |
| SYNZIP2 | 16270.3 | 385.6 | 38042.5 | 2768.0 | -40.9 | 139.0 | 16860.3 | 790.3 | 1660.4 | 1912.6 |
| SYNZIP3 | 9286.6 | 889.3 | 2402.3 | 9764.8 | 13850.0 | 3096.6 | 701.9 | 2401.1 | 617.8 | -132.6 |
| SYNZIP4 | 10758.4 | 41.5 | 5789.6 | 764.1 | 575.1 | 341.0 | 1183.6 | 1496.0 | 1091.1 | 3005.9 |
| SYNZIP5 | 35487.3 | 30555.3 | 5312.8 | 3299.5 | 1725.6 | 1474.9 | 1406.3 | 740.4 | 4688.1 | 30876.8 |
| SYNZIP6 | 15801.4 | 623.9 | 7950.1 | 2593.4 | 6735.6 | 2904.6 | 1100.0 | 1450.6 | -392.0 | 1167.9 |
| SYNZIP7 | 41.9 | -370.3 | 2247.3 | 424.5 | 91.8 | 89.1 | 1726.8 | 1522.6 | 1153.4 | 11141.5 |
| SYNZIP8 | 1217.8 | -82.0 | 3136.1 | 809.5 | 421.9 | 116.9 | 915.5 | 2925.8 | 3316.3 | 10745.1 |
| SYNZIP9 | 2950.8 | 200.9 | 1236.5 | -117.5 | 327.1 | 285.8 | 24072.8 | 442.3 | 1522.3 | 2002.4 |
| SYNZIP10 | 12440.5 | 664.8 | 14952.5 | 4863.1 | 7532.4 | 810.1 | 1806.0 | 1580.3 | 9100.6 | 2918.8 |
| SYNZIP11 | 8404.6 | -173.1 | 31036.3 | 9121.3 | 1905.4 | 620.5 | 4844.1 | 12993.5 | 10133.6 | 3213.0 |
| SYNZIP12 | 22124.5 | 2328.1 | 36988.6 | 6454.5 | 4732.9 | 958.4 | 1956.9 | 4639.9 | 961.5 | 5161.3 |
| SYNZIP13 | 2772.3 | 81.5 | 2221.1 | 16.8 | -181.4 | -50.1 | 166.6 | -13.9 | -2.0 | 4561.0 |
| SYNZIP14 | 18649.3 | 916.3 | 24270.0 | 22491.3 | 2464.9 | 2040.5 | 5534.3 | 3229.3 | 1917.9 | 11333.5 |
| SYNZIP15 | 48837.4 | 13359.5 | 43436.3 | 36460.3 | 638.5 | 336.8 | 3790.0 | 1328.9 | -287.4 | 1684.6 |
| SYNZIP16 | 7936.6 | 484.0 | 34066.0 | 7310.6 | 1855.4 | 407.3 | 9738.0 | 3304.0 | 4137.0 | 14315.5 |
| SYNZIP17 | 9065.3 | 360.9 | 7464.9 | 724.3 | 655.8 | 1238.0 | 33626.5 | 624.4 | 4377.6 | 1317.4 |
| SYNZIP18 | 16593.4 | 401.9 | 10363.3 | 9376.3 | 19906.8 | 3726.4 | 5216.5 | 3350.5 | 551.9 | 12116.0 |
| SYNZIP19 | 16170.0 | 1115.5 | 43399.3 | 11110.5 | 2646.0 | 1352.3 | 7652.1 | 2448.3 | 6755.6 | 32438.9 |
| SYNZIP20 | 15291.3 | 2172.9 | 9773.4 | 4216.1 | 4062.1 | 7885.3 | -383.3 | 307.0 | 456.4 | 9689.3 |
| SYNZIP21 | 19554.9 | 6103.4 | 13281.3 | 8181.8 | 9571.6 | 5215.9 | 1600.0 | 1910.0 | 170.8 | 698.3 |
| SYNZIP22 | 15375.9 | 1028.1 | 31330.0 | 5762.5 | 7432.1 | 511.8 | 6781.0 | 5398.8 | 2886.9 | 5431.8 |
| SYNZIP23 | 47534.9 | 4024.0 | 20924.1 | 22911.3 | 11860.4 | 25743.9 | 357.3 | 854.0 | 2202.8 | 11532.5 |
| SYNZIP24 | 25124.9 | 6517.9 | 1311.6 | 3093.5 | -520.3 | -458.9 | 1583.4 | 1664.0 | 533.4 | 17240.6 |
| SYNZIP25 | 1985.3 | -32.4 | 26088.8 | -31.4 | -6.0 | 9.0 | 347.9 | 945.5 | 438.0 | 492.9 |
| SYNZIP26 | 918.8 | -265.9 | 590.9 | 214.9 | 333.3 | 890.4 | 100.1 | 88.1 | -298.3 | -239.6 |
| SYNZIP27 | 5682.9 | 676.1 | 2856.3 | 1714.9 | 922.8 | 888.3 | 21756.1 | 1749.4 | 4126.5 | 6724.8 |
| SYNZIP28 | 1914.5 | 61.8 | 2431.0 | 530.3 | -86.6 | 522.9 | 12196.5 | 631.3 | 4580.3 | 6693.4 |

| Protein | SYNZIP41 | SYNZIP42 | SYNZIP43 | SYNZIP44 | SYNZIP45 | SYNZIP46 | SYNZIP47 | SYNZIP48 | BATF | FOS |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP1 | 999.1 | 432.6 | 134.0 | 23.9 | 620.8 | 14386.0 | 575.9 | 14520.0 | -75.8 | 1417.4 |
| SYNZIP2 | 671.6 | 648.6 | 7749.1 | -270.6 | 2049.0 | 2085.1 | 244.0 | 7664.8 | 1308.0 | 17359.6 |
| SYNZIP3 | 164.9 | 359.8 | 8791.1 | -106.8 | 370.3 | 28.4 | 950.0 | -123.6 | 447.8 | -556.3 |
| SYNZIP4 | 2911.1 | 488.1 | 6217.1 | 37336.0 | 70.3 | 4227.3 | 195.1 | 2593.9 | 890.6 | 4944.8 |
| SYNZIP5 | 667.4 | 616.0 | 9197.4 | 7318.1 | 370.4 | 960.0 | 333.4 | 8558.8 | 6890.5 | 3569.3 |
| SYNZIP6 | -194.9 | 506.3 | 17352.1 | 5085.8 | 2047.1 | -227.0 | 144.9 | 1530.6 | 44.8 | 5079.8 |
| SYNZIP7 | 302.9 | 368.8 | 5864.4 | -716.6 | 311.6 | 1454.4 | 255.9 | 3570.8 | -529.0 | 2602.3 |
| SYNZIP8 | 1110.5 | 536.4 | -190.0 | -44.8 | 1177.3 | 2348.9 | 567.0 | 2013.9 | 819.6 | 2767.0 |
| SYNZIP9 | 7895.5 | 688.4 | -627.6 | 1500.6 | 531.6 | 2978.1 | -632.4 | 169.9 | 214.5 | 34170.9 |
| SYNZIP10 | 1378.1 | 738.0 | 4152.5 | 1352.3 | 598.3 | 2150.9 | 1645.6 | 5208.0 | 1003.9 | 1562.6 |
| SYNZIP11 | 4424.5 | 1424.5 | 2567.8 | 813.3 | 3601.9 | 2973.3 | 1379.1 | 8323.0 | 2010.9 | 15830.4 |
| SYNZIP12 | 1610.9 | 1178.6 | 8583.1 | -1728.1 | 3103.0 | 799.8 | 914.5 | 5387.5 | 3180.3 | 9262.3 |
| SYNZIP13 | 1019.4 | 452.1 | -1177.1 | -100.3 | -145.0 | 480.4 | 59.5 | 351.1 | 73.6 | 7393.3 |
| SYNZIP14 | 2611.1 | 759.3 | 3923.6 | 3037.9 | 668.8 | 1356.4 | 442.1 | 19711.0 | 1634.6 | 6924.8 |
| SYNZIP15 | -313.1 | 224.6 | 17729.9 | -1455.6 | -297.4 | 1080.5 | 649.8 | 15768.8 | 5377.8 | 8950.4 |
| SYNZIP16 | 22046.6 | 942.6 | 1694.3 | 1519.6 | 5918.5 | 1485.0 | 1025.6 | 5514.1 | 1104.9 | 11950.5 |
| SYNZIP17 | 13695.9 | 551.6 | 0.5 | 10334.5 | 1269.3 | 7741.8 | 198.1 | -279.5 | 1423.4 | 31131.5 |
| SYNZIP18 | -254.0 | 671.3 | 18844.9 | 883.4 | 2713.9 | 309.4 | 1966.6 | 6394.5 | -55.3 | 2034.0 |
| SYNZIP19 | 3142.4 | 542.1 | 533.8 | 6987.9 | 457.4 | 2429.1 | 106.0 | 9506.1 | 20530.6 | 11655.9 |
| SYNZIP20 | 2605.5 | 439.9 | -142.4 | 3469.3 | 271.4 | 1832.6 | 1090.3 | 5166.9 | 5052.0 | 17962.3 |
| SYNZIP21 | 351.0 | 566.8 | 2579.8 | 699.3 | 900.5 | 1037.3 | 7398.3 | 3473.3 | 2297.3 | 4793.9 |
| SYNZIP22 | 14364.3 | 7311.9 | 6067.9 | 1502.6 | 10906.8 | 6275.9 | 979.0 | 2258.1 | 6420.5 | 18413.1 |
| SYNZIP23 | 262.4 | 459.0 | 20967.5 | 25367.3 | 722.1 | 5010.5 | 1245.4 | 4291.0 | 41325.6 | 25403.4 |
| SYNZIP24 | -121.4 | 301.3 | -906.8 | -1086.9 | -153.4 | 1198.9 | -424.0 | 478.1 | 1389.4 | -557.1 |
| SYNZIP25 | 193.6 | 621.8 | -339.6 | -1308.8 | 992.1 | -852.4 | 673.0 | 4632.4 | 125.4 | 14939.0 |
| SYNZIP26 | -115.8 | 323.6 | -114.8 | 359.5 | -178.8 | -19.0 | 531.8 | 9262.0 | 721.9 | 1261.9 |
| SYNZIP27 | 24764.8 | 982.3 | 271.4 | 2822.1 | 3689.9 | 5945.8 | 258.4 | 224.3 | 1233.4 | 11190.9 |
| SYNZIP28 | 6882.5 | 577.9 | -385.6 | 1292.8 | 947.5 | 6163.0 | -24.6 | -361.5 | 796.6 | 7196.6 |

| Protein | ATF4 | ATF3 | BACH1 | JUND | NFE2L3 | SYNZIP5(2) | SYN37(2) | SYNZIP6(2) | ATF4(2) | FOS(2) |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP1 | 3053.8 | 1420.0 | 1577.6 | 3627.1 | 2925.1 | 290.4 | 380.3 | 163.9 | 1997.3 | 1168.5 |
| SYNZIP2 | 28082.8 | 18064.8 | 1266.9 | 7929.5 | 1109.6 | 1068.3 | 7298.6 | 763.0 | 22237.3 | 18017.8 |
| SYNZIP3 | 5823.0 | 3581.0 | 134.5 | 28656.3 | -86.9 | 11407.4 | -180.0 | 418.0 | 4817.3 | -658.3 |
| SYNZIP4 | 3458.6 | 17232.4 | 696.4 | 5558.6 | 681.0 | 4.5 | 278.5 | 14845.0 | 3406.3 | 6151.4 |
| SYNZIP5 | 6154.3 | 7063.1 | 146.1 | 11543.6 | 147.0 | -121.6 | -26.6 | 40014.4 | 5716.6 | 4467.9 |
| SYNZIP6 | 4243.6 | 12134.8 | 2673.8 | 16589.0 | -258.6 | 50451.1 | -107.1 | -112.6 | 3719.8 | 4670.8 |
| SYNZIP7 | 32347.3 | 360.5 | 4649.4 | 6636.9 | 27047.4 | -185.5 | 201.1 | -246.8 | 30001.6 | 3755.4 |
| SYNZIP8 | 1243.1 | 2317.9 | 2708.6 | 6235.3 | 1894.3 | 1311.3 | -7.4 | 125.3 | 1264.1 | 2566.5 |
| SYNZIP9 | 1541.8 | 6073.8 | 24587.6 | 5738.9 | 3399.1 | 1439.0 | 19199.0 | 2013.0 | 1329.5 | 31369.4 |
| SYNZIP10 | 5436.6 | 4179.5 | -150.5 | 12891.0 | 470.9 | 2624.5 | 702.9 | 4658.4 | 5082.5 | 2021.8 |
| SYNZIP11 | 17929.4 | 26643.8 | 10092.0 | 16369.5 | 4545.5 | 11578.1 | 2212.0 | 533.5 | 15494.5 | 18153.5 |
| SYNZIP12 | 12482.6 | 40057.8 | 4713.6 | 11054.3 | 828.3 | 17760.9 | 447.9 | 850.5 | 10379.1 | 10370.5 |
| SYNZIP13 | 1628.3 | 11177.0 | -271.5 | 1128.6 | 576.6 | 659.3 | -230.9 | 1800.0 | 1851.0 | 6599.4 |
| SYNZIP14 | 9191.0 | 9794.9 | 2727.0 | 8261.8 | 303.5 | 1466.0 | 2471.6 | 5248.0 | 5608.0 | 7193.1 |
| SYNZIP15 | 35747.3 | 14239.9 | -629.9 | 1597.1 | 342.9 | -426.4 | 1037.8 | 393.6 | 26273.6 | 8321.8 |
| SYNZIP16 | 20306.6 | 10669.4 | 10058.1 | 16018.1 | 4126.0 | 23360.8 | 5900.3 | 253.3 | 16289.1 | 14487.6 |
| SYNZIP17 | 951.1 | 3487.1 | 20467.5 | 4762.5 | 351.8 | 1223.6 | 23343.1 | 5787.1 | 1121.8 | 31600.1 |
| SYNZIP18 | 8274.4 | 5824.8 | 4007.5 | 19155.4 | 956.1 | 17117.4 | 2225.8 | 109.9 | 5166.6 | 1663.5 |
| SYNZIP19 | 4515.0 | 7485.1 | 259.4 | 265.0 | 388.0 | -200.1 | 3888.1 | 19731.6 | 4373.4 | 11508.9 |
| SYNZIP20 | 4251.1 | 15726.5 | -88.9 | 3124.3 | 1787.9 | -52.0 | -718.1 | 20718.6 | 3381.3 | 19099.0 |
| SYNZIP21 | 3314.1 | 32347.4 | 1111.0 | 27319.5 | 1097.1 | 31805.6 | -131.3 | 651.1 | 3480.8 | 4989.3 |
| SYNZIP22 | 46531.6 | 16363.6 | 15474.1 | 26119.5 | 28539.3 | 20949.0 | 2210.4 | -128.5 | 37329.8 | 19989.5 |
| SYNZIP23 | 3541.8 | 19932.6 | 7074.4 | 2978.0 | 2526.4 | 688.6 | -243.1 | 24579.4 | 3239.4 | 25406.4 |
| SYNZIP24 | 1939.9 | 365.5 | -1420.4 | -284.6 | 171.9 | -707.0 | 308.3 | 771.3 | 2117.4 | -651.6 |
| SYNZIP25 | 3907.1 | 7493.3 | 357.8 | 3463.0 | 246.9 | 7360.1 | -180.4 | -351.0 | 3354.0 | 14167.5 |
| SYNZIP26 | 1364.1 | 743.8 | 1003.0 | 1360.8 | 300.1 | 848.8 | -158.8 | 58.5 | 1260.4 | 407.6 |
| SYNZIP27 | 3004.3 | 3749.1 | 20822.9 | 1764.3 | 1659.9 | 3435.1 | 15729.6 | 2018.0 | 2248.6 | 12173.0 |
| SYNZIP28 | 1807.1 | 2316.8 | 10386.3 | -42.0 | 1139.0 | 716.5 | 6897.8 | 2023.4 | 1833.1 | 8138.3 |

| Protein | SYNZIP1 | SYNZIP2 | SYNZIP3 | SYNZIP4 | SYNZIP5 | SYNZIP6 | SYNZIP7 | SYNZIP8 | SYNZIP9 | SYNZIP10 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP29 | 1896.9 | 30870.3 | 1920.8 | 1975.0 | 3266.1 | 9225.5 | 3999.0 | 12211.0 | 1676.0 | 7345.8 |
| SYNZIP30 | 4521.9 | 24828.0 | 2650.0 | 13367.0 | 3763.5 | 15017.8 | 8682.0 | 19723.5 | 661.9 | 9620.9 |
| SYNZIP31 | -227.6 | 4698.1 | 192.9 | -476.5 | 3162.6 | 92.5 | -91.4 | -141.6 | -156.5 | 490.9 |
| SYNZIP32 | -1281.1 | 1743.4 | -1031.4 | -4299.9 | 6521.3 | 9.4 | -152.8 | -259.0 | -593.0 | -526.4 |
| SYNZIP33 | -354.9 | 25369.3 | 1083.9 | 334.6 | 316.5 | 845.6 | 1766.4 | 2265.9 | 317.0 | 4585.0 |
| SYNZIP34 | 119.5 | 4531.6 | 3470.8 | 1309.3 | 514.3 | 513.4 | 3.1 | 350.1 | -146.4 | 612.9 |
| SYNZIP35 | 427.6 | 1246.8 | 16456.8 | 1286.9 | 655.4 | 3118.8 | 255.4 | 588.0 | 87.4 | 2182.1 |
| SYNZIP36 | -71.9 | -324.3 | 311.9 | -1904.9 | -132.8 | 772.3 | -62.1 | 92.0 | -371.0 | -196.8 |
| SYNZIP37 | -141.6 | 8759.5 | -522.0 | 220.9 | -456.3 | -663.6 | 544.5 | 148.5 | 8519.1 | -415.4 |
| SYNZIP38 | 57.0 | 404.9 | 1296.6 | 2669.0 | 37.5 | 103.9 | 983.8 | 1996.6 | -247.9 | -24.6 |
| SYNZIP39 | -191.0 | 3869.1 | 847.9 | 2954.3 | 365.9 | -3.8 | 1788.6 | 3831.8 | 916.3 | 1073.1 |
| SYNZIP40 | 137.4 | 2158.0 | -158.0 | 2992.3 | 10116.6 | 237.3 | 8195.4 | 8907.0 | -131.9 | -811.6 |
| SYNZIP41 | -1118.5 | 1083.1 | -182.9 | 2248.1 | -303.8 | -229.0 | 415.6 | 544.5 | 2035.4 | -991.5 |
| SYNZIP42 | 35.9 | 788.4 | 408.6 | -90.4 | -124.8 | 102.8 | 26.9 | 300.4 | 34.1 | -209.5 |
| SYNZIP43 | -934.0 | 15741.5 | 10102.1 | 4654.3 | 2996.9 | 4873.8 | 4742.4 | 68.0 | -238.5 | 1651.9 |
| SYNZIP44 | -190.0 | 441.8 | -258.4 | 38735.1 | 3533.3 | 1397.5 | -129.6 | -180.3 | 3271.5 | 306.5 |
| SYNZIP45 | -175.6 | 4051.9 | 672.5 | 707.8 | -535.0 | 1052.4 | 67.8 | 1818.8 | 430.4 | -403.3 |
| SYNZIP46 | 2115.8 | -620.0 | -932.8 | -914.3 | -506.4 | -383.4 | -100.4 | -75.9 | -126.0 | -163.3 |
| SYNZIP47 | 27.9 | -97.6 | 306.1 | -1210.6 | -298.8 | 1398.6 | -277.5 | 590.9 | -366.6 | -503.8 |
| SYNZIP48 | 5590.1 | 10726.9 | 904.1 | 1250.6 | 2799.6 | 1541.6 | 4291.0 | 7820.9 | -245.5 | 5330.5 |
| BATF | 169.0 | 4108.0 | 289.5 | 3800.6 | 1982.6 | 122.5 | 483.4 | 640.1 | 273.4 | 188.6 |
| FOS | 355.9 | 25811.1 | 438.1 | 4373.0 | 353.0 | 823.4 | 1809.4 | 1107.1 | 31699.6 | -516.1 |
| ATF4 | 341.3 | 45201.3 | 3634.5 | 1430.3 | 703.5 | 443.0 | 36041.1 | 476.3 | -387.3 | 481.1 |
| ATF3 | -158.8 | 17343.6 | 891.5 | 8663.8 | 369.3 | 1460.9 | 363.5 | 598.0 | 670.9 | 220.3 |
| BACH1 | 452.4 | 2642.6 | -1.4 | 4098.1 | 39.0 | 366.8 | 2991.0 | 1942.3 | 8760.5 | -308.9 |
| JUND | 271.3 | 2510.0 | 12860.5 | 2623.4 | 832.3 | 3651.6 | 795.6 | 3186.0 | 878.1 | 1003.8 |
| NFE2L3 | 134.3 | -858.3 | -336.5 | -580.9 | -133.3 | 83.8 | 4285.5 | 165.0 | 327.6 | -171.6 |

| Protein | SYNZIP11 | SYNZIP12 | SYNZIP13 | SYNZIP14 | SYNZIP15 | SYNZIP16 | SYNZIP17 | SYNZIP18 | SYNZIP19 | SYNZIP20 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP29 | 20577.6 | 14436.9 | 3188.5 | 10125.3 | 14839.0 | 25111.4 | 3769.8 | 16366.8 | 581.5 | 6595.0 |
| SYNZIP30 | 21824.9 | 15677.3 | 5439.3 | 15343.8 | 29943.3 | 23308.1 | 3590.8 | 23744.4 | 6878.9 | 6579.3 |
| SYNZIP31 | 248.3 | 3610.6 | -657.3 | 2890.5 | 17842.8 | 2402.3 | -785.0 | 1358.6 | 803.9 | 5209.0 |
| SYNZIP32 | -227.8 | 279.5 | -1209.8 | 323.4 | 11618.3 | 162.5 | -1720.1 | -391.9 | 1243.3 | 1323.3 |
| SYNZIP33 | 12673.0 | 22749.5 | 307.3 | 22381.8 | 27053.1 | 24659.4 | 5735.1 | 4350.4 | 13702.9 | 7584.0 |
| SYNZIP34 | 3235.9 | 2524.3 | 748.4 | 20234.4 | 40815.0 | 8797.0 | 927.9 | 3038.1 | 2890.9 | 4347.8 |
| SYNZIP35 | 976.8 | -495.8 | 374.3 | 1183.8 | 1481.4 | 3338.9 | 2659.6 | 10191.5 | 3651.4 | 7530.1 |
| SYNZIP36 | -139.6 | 128.4 | -523.4 | -231.8 | -996.9 | -405.0 | 16.4 | 490.8 | 506.0 | 2182.9 |
| SYNZIP37 | 44.3 | -426.4 | 14.1 | 39.5 | 627.4 | 8424.4 | 15728.6 | 366.5 | 3770.3 | -1960.5 |
| SYNZIP38 | 3164.3 | -37.1 | 181.9 | 7925.3 | 2440.6 | 4952.9 | 879.8 | -34.9 | 1560.0 | -555.3 |
| SYNZIP39 | 3212.0 | 266.1 | -453.1 | 1242.1 | 762.1 | 6160.0 | 3434.4 | 19.1 | 4787.5 | 1473.9 |
| SYNZIP40 | 692.9 | 236.8 | 648.8 | 7773.0 | 1464.0 | 18851.3 | 301.8 | 2569.1 | 13835.0 | 7316.4 |
| SYNZIP41 | 54.0 | 128.3 | -254.6 | -831.8 | -1632.6 | 15347.4 | 6736.6 | -357.0 | 1778.9 | 960.4 |
| SYNZIP42 | 150.6 | -51.1 | 161.8 | -673.9 | -871.5 | 1826.1 | 831.5 | -423.5 | -838.6 | -548.3 |
| SYNZIP43 | 132.8 | 3818.1 | -337.1 | 2192.5 | 26601.9 | 2690.3 | -751.5 | 6669.6 | -707.1 | 511.5 |
| SYNZIP44 | 813.3 | -423.1 | 1070.5 | 4856.4 | -544.8 | 2946.3 | 11164.0 | 38.1 | 9649.9 | 7054.6 |
| SYNZIP45 | 1308.8 | 1357.0 | -60.5 | -119.1 | -2049.0 | 10866.9 | 1885.8 | 742.6 | 464.6 | -100.4 |
| SYNZIP46 | -700.5 | -987.5 | -701.1 | -1094.1 | -5031.0 | -1059.5 | -806.0 | -777.1 | -261.9 | -160.6 |
| SYNZIP47 | 765.3 | 20.1 | 487.1 | -302.0 | -592.0 | -337.3 | -1327.6 | -42.8 | -591.4 | 1623.8 |
| SYNZIP48 | 7251.1 | 5793.0 | 184.0 | 6076.4 | 10075.8 | 4610.3 | -1247.5 | 6124.1 | 3850.3 | 3134.5 |
| BATF | 1846.9 | 673.9 | 539.3 | 883.0 | 5957.1 | 2534.0 | 3008.6 | -212.5 | 16793.1 | 7000.8 |
| FOS | 6157.9 | 1934.0 | 2685.0 | 6466.4 | 5057.4 | 28581.8 | 37108.3 | 117.8 | 6972.8 | 25938.1 |
| ATF4 | 10480.9 | 6243.9 | 520.3 | 8382.4 | 49513.4 | 41931.1 | 346.1 | 1062.1 | 294.1 | 838.9 |
| ATF3 | 6402.1 | 19974.8 | 4257.4 | 4087.5 | 4948.4 | 12070.3 | 2376.5 | 214.0 | 1093.4 | 13519.4 |
| BACH1 | 1788.8 | -636.5 | 615.5 | 2117.1 | -2430.8 | 7640.8 | 10141.1 | -480.0 | 543.8 | -1276.6 |
| JUND | 506.9 | -1547.8 | 1265.4 | -289.3 | -458.3 | 14361.9 | 2505.5 | 1165.8 | -1055.6 | 2079.0 |
| NFE2L3 | 30.8 | -417.5 | 98.1 | -792.1 | -804.4 | -560.6 | -524.0 | -538.9 | -1534.5 | -1125.4 |

| Protein | SYNZIP21 | SYNZIP22 | SYNZIP23 | SYNZIP24 | SYNZIP25 | SYNZIP26 | SYNZIP27 | SYNZIP28 | SYNZIP29 | SYNZIP30 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP29 | 10481.0 | 9108.3 | 12240.8 | 5028.9 | 14431.5 | 424.0 | 8083.5 | 6141.9 | 416.4 | 640.5 |
| SYNZIP30 | 14194.4 | 10006.6 | 16090.4 | 10835.1 | 12132.4 | 778.3 | 1441.4 | 229.9 | 426.5 | 1630.0 |
| SYNZIP31 | 9598.8 | 435.6 | 23329.6 | 16403.5 | -92.0 | 177.1 | 419.5 | 669.8 | 798.0 | 375.9 |
| SYNZIP32 | 7954.6 | -1280.0 | 4434.9 | 9960.4 | -93.5 | 964.4 | 4821.9 | 3274.6 | 2862.3 | 1435.4 |
| SYNZIP33 | 14104.1 | 9135.1 | 14963.9 | 5618.4 | 7911.0 | 331.9 | 5119.9 | 10270.8 | 3513.1 | 1894.1 |
| SYNZIP34 | 10535.1 | 2697.3 | 28266.9 | 8234.6 | 227.5 | 743.0 | 2034.1 | 658.4 | 2390.8 | 3495.5 |
| SYNZIP35 | 17653.9 | 1099.1 | 26860.6 | 923.0 | 151.3 | 874.5 | 2715.4 | -240.6 | 2466.9 | 7036.1 |
| SYNZIP36 | 1569.8 | -1066.0 | 11293.4 | -1201.4 | 34.8 | 388.1 | 791.3 | 173.5 | 1053.4 | 2300.5 |
| SYNZIP37 | 987.4 | -1737.9 | -893.8 | 1978.5 | 61.4 | 596.8 | 32892.1 | 18435.0 | 4601.3 | 1977.9 |
| SYNZIP38 | 2810.9 | -388.9 | 2498.4 | 11068.4 | 171.5 | 967.3 | 6707.0 | 6157.9 | 2708.8 | 2190.4 |
| SYNZIP39 | 1328.0 | -109.0 | 5792.3 | 3799.4 | -305.5 | 572.8 | 9844.3 | 12439.9 | 3262.1 | 4044.1 |
| SYNZIP40 | -276.5 | -659.1 | 8391.6 | 26446.1 | 138.4 | 849.1 | 10091.6 | 17681.6 | 2758.9 | 1056.0 |
| SYNZIP41 | -130.5 | 2339.1 | -1279.1 | 555.0 | 68.0 | 315.4 | 30882.3 | 13176.8 | 4472.8 | 2349.4 |
| SYNZIP42 | -69.4 | 6331.9 | -171.4 | 652.4 | 192.3 | 986.4 | 4210.1 | 237.4 | 5146.6 | 2024.4 |
| SYNZIP43 | 11665.8 | 2572.1 | 34636.5 | 12303.6 | 112.9 | 339.0 | -1211.5 | -1321.0 | 1187.5 | 2325.8 |
| SYNZIP44 | 1918.0 | 23.6 | 28527.5 | 2598.5 | 548.8 | 574.6 | 11327.6 | 10547.1 | 7743.5 | 8754.5 |
| SYNZIP45 | 1837.8 | 3990.3 | 1966.3 | -392.3 | -268.3 | 754.8 | 12878.0 | 5165.8 | 2977.6 | 5765.9 |
| SYNZIP46 | -1490.8 | -2292.1 | -1062.4 | -545.8 | -196.1 | 63.4 | 2680.3 | 3731.4 | 1810.3 | 1776.8 |
| SYNZIP47 | 9564.6 | -1087.5 | 871.3 | -1214.1 | 4208.8 | 255.1 | -2068.1 | -636.1 | 338.0 | 352.6 |
| SYNZIP48 | 5566.9 | 686.1 | 2989.5 | 483.4 | 3965.9 | 12198.8 | -2002.1 | -1452.5 | -217.1 | -60.0 |
| BATF | 3581.4 | 3339.6 | 38364.0 | 5683.3 | 187.6 | 757.3 | 7114.9 | 8121.6 | 5987.0 | 6497.4 |
| FOS | 7486.6 | 12546.8 | 26601.6 | 534.0 | 1413.4 | 1068.9 | 24859.9 | 19649.8 | 8339.8 | 5178.5 |
| ATF4 | 4479.1 | 30585.9 | 7694.1 | 11346.5 | 288.0 | 777.4 | 3533.3 | 1365.6 | 8279.6 | 6124.8 |
| ATF3 | 29434.8 | 7975.8 | 24547.0 | 1577.0 | 106.0 | 1185.8 | 2495.0 | 1887.5 | 3678.9 | 4279.8 |
| BACH1 | 919.4 | 3805.9 | 6787.6 | -99.0 | 1584.8 | 1866.3 | 30298.1 | 22918.8 | 4367.1 | 4725.8 |
| JUND | 29895.6 | 10619.0 | 7296.3 | 697.5 | 20.4 | 1306.9 | -1786.1 | -1311.0 | 4660.0 | 4605.3 |
| NFE2L3 | -1093.9 | 3456.1 | -1738.8 | -185.5 | 186.8 | -105.9 | -491.3 | -1419.5 | 269.5 | 603.3 |

| Protein | SYNZIP31 | SYNZIP32 | SYNZIP33 | SYNZIP34 | SYNZIP35 | SYNZIP36 | SYNZIP37 | SYNZIP38 | SYNZIP39 | SYNZIP40 |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP29 | 6525.9 | 4252.6 | 12063.4 | 3571.1 | 103.4 | 758.8 | 15566.4 | 4197.3 | 4578.9 | 9246.5 |
| SYNZIP30 | 11833.9 | 3705.1 | 12311.5 | 6951.1 | 7601.3 | 7576.1 | 5188.4 | 2818.1 | 5833.5 | 6277.9 |
| SYNZIP31 | 12845.8 | 820.8 | 5088.3 | 22225.5 | 96.0 | 848.0 | 1488.0 | 474.9 | -209.1 | 247.1 |
| SYNZIP32 | 5780.6 | 303.6 | 1929.5 | 21185.6 | -404.6 | 249.9 | 764.9 | 1464.5 | 2267.8 | -236.5 |
| SYNZIP33 | 13516.8 | 2042.8 | 18294.6 | 21285.5 | 947.5 | 2015.9 | 2558.4 | 1247.9 | 3484.6 | 4232.4 |
| SYNZIP34 | 41747.5 | 36836.0 | 39042.9 | 10390.4 | 274.5 | 734.3 | 3025.0 | 1614.3 | 925.4 | 1853.9 |
| SYNZIP35 | 4433.4 | -63.1 | 2486.9 | 677.1 | 1358.8 | 677.0 | 1555.6 | 1583.3 | 1357.1 | 3599.3 |
| SYNZIP36 | 2821.9 | 171.8 | 1090.0 | 587.6 | 174.5 | 843.1 | 572.0 | 471.3 | 65.6 | -103.4 |
| SYNZIP37 | 1434.0 | 39.3 | 596.3 | 1546.3 | 406.4 | 75.0 | 2358.9 | 2354.4 | 4380.6 | 4384.5 |
| SYNZIP38 | 4101.5 | -46.3 | 4548.4 | 3334.1 | 746.9 | 452.9 | 9506.3 | 3193.1 | 2086.8 | 3038.4 |
| SYNZIP39 | 1774.6 | 219.6 | 2998.3 | 1040.9 | 1274.8 | 333.0 | 8809.6 | 2277.1 | 9723.1 | 3486.0 |
| SYNZIP40 | 1426.0 | -143.1 | 3572.3 | 243.3 | 2184.6 | 173.8 | 7521.8 | 2104.4 | 1557.4 | 14896.5 |
| SYNZIP41 | 794.1 | -38.5 | 2688.3 | 321.5 | 2607.5 | 169.4 | 1699.9 | 1982.1 | 1782.6 | 3529.0 |
| SYNZIP42 | 1770.0 | -530.9 | 1984.8 | 322.3 | 290.3 | 316.6 | 1136.0 | 540.1 | 592.0 | 19532.3 |
| SYNZIP43 | 9910.3 | 179.6 | 4927.8 | 366.6 | 416.1 | 701.1 | 984.9 | 246.9 | -91.4 | 1951.3 |
| SYNZIP44 | 16103.5 | 1072.0 | 5818.9 | 2281.1 | 7600.0 | 4859.6 | 716.0 | 1479.3 | 96.1 | 2382.0 |
| SYNZIP45 | 4293.4 | 149.9 | 7143.4 | 1168.0 | 87.4 | 290.3 | 1092.3 | 3241.4 | 2472.4 | 3650.8 |
| SYNZIP46 | -1576.0 | -56.1 | -1184.9 | -1889.4 | -99.8 | -168.1 | 249.0 | 230.6 | 193.0 | -247.8 |
| SYNZIP47 | 576.3 | 51.8 | -43.1 | -216.5 | -200.0 | 64.3 | 356.9 | 369.0 | -1236.5 | -264.6 |
| SYNZIP48 | 2622.1 | 178.0 | 6620.4 | 639.5 | 1486.1 | 3665.0 | 768.4 | 806.0 | 709.6 | 1509.9 |
| BATF | 13330.5 | 1540.5 | 3127.1 | 17304.9 | 584.5 | 582.6 | 1721.9 | 10870.4 | 1694.4 | 1011.4 |
| FOS | 1853.4 | 816.4 | 5876.9 | 10360.4 | 2255.4 | 724.9 | 7880.9 | 13262.1 | 12213.9 | 5632.3 |
| ATF4 | 32811.0 | 1242.9 | 27808.3 | 28253.0 | 260.4 | 569.8 | 1829.8 | 349.1 | 528.4 | 21756.3 |
| ATF3 | 9108.1 | 1161.8 | 3026.9 | 14628.5 | 1130.9 | 994.0 | 2252.9 | 2349.8 | 3975.0 | 1357.8 |
| BACH1 | 2206.8 | -234.6 | 1996.3 | 1235.1 | 3108.3 | 801.1 | 10083.3 | 2518.9 | 6443.4 | 6296.5 |
| JUND | 18241.6 | 300.0 | 8043.9 | 2521.3 | 224.5 | 403.5 | -13.3 | 292.3 | 418.4 | 1418.8 |
| NFE2L3 | 624.8 | -494.8 | 821.6 | -249.6 | 229.9 | 349.5 | 126.9 | -152.3 | -109.3 | 326.8 |

| Protein | SYNZIP41 | SYNZIP42 | SYNZIP43 | SYNZIP44 | SYNZIP45 | SYNZIP46 | SYNZIP47 | SYNZIP48 | BATF | FOS |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP29 | 13843.0 | 4461.0 | 1072.9 | 10681.0 | 4194.5 | 12912.9 | 99.0 | -131.4 | 5111.3 | 9189.1 |
| SYNZIP30 | 12834.6 | 1001.4 | 5983.5 | 16354.3 | 5876.8 | 23510.6 | 808.1 | 859.1 | 7876.8 | 12206.1 |
| SYNZIP31 | -221.9 | 296.0 | 1153.4 | 1996.8 | 230.1 | 9.0 | 102.8 | -633.5 | 2210.5 | -26.0 |
| SYNZIP32 | -987.5 | 395.0 | -1065.3 | -552.4 | 228.3 | 319.1 | 116.1 | -60.9 | 1737.3 | 4674.1 |
| SYNZIP33 | 3187.5 | 954.3 | 2582.5 | 860.8 | 5600.1 | 959.5 | 147.5 | 3504.8 | 1597.6 | 4427.9 |
| SYNZIP34 | 1148.8 | 491.9 | 547.8 | 535.1 | 378.3 | 1169.3 | 228.6 | 693.1 | 13095.3 | 8182.8 |
| SYNZIP35 | 5301.6 | 545.3 | 708.5 | 3017.9 | 142.1 | 465.9 | 299.5 | 1453.1 | 295.9 | 3024.1 |
| SYNZIP36 | 418.4 | 461.4 | 662.1 | 1624.4 | 75.5 | -98.6 | 52.8 | 959.3 | -0.9 | 638.8 |
| SYNZIP37 | 849.1 | 551.0 | -685.5 | -691.4 | 601.5 | 1419.5 | 483.1 | -43.3 | -23.9 | 4497.6 |
| SYNZIP38 | 23897.9 | 650.6 | -91.5 | -75.6 | 2717.3 | 692.5 | -26.8 | 484.8 | 9668.5 | 18673.1 |
| SYNZIP39 | 2346.8 | 596.3 | 25.5 | -97.0 | 951.8 | 5209.3 | 325.9 | 762.0 | 227.5 | 10208.5 |
| SYNZIP40 | 3498.4 | 6775.5 | -4647.8 | -1083.8 | 831.5 | 2152.0 | 430.8 | 38.3 | -121.4 | 3875.4 |
| SYNZIP41 | 13639.9 | 1222.6 | -454.9 | -401.9 | 787.5 | -185.9 | -12.8 | -303.8 | -158.8 | 1074.6 |
| SYNZIP42 | 4762.6 | 695.3 | 0.5 | -99.3 | 80.4 | -108.0 | 140.0 | 85.9 | -292.1 | 2172.3 |
| SYNZIP43 | 822.1 | 420.8 | -811.3 | 30243.9 | 352.3 | 6761.5 | 63.3 | -58.3 | 706.6 | 3319.3 |
| SYNZIP44 | -230.1 | 335.6 | 33033.3 | 9029.1 | 512.1 | 1088.6 | 2209.6 | 1393.5 | 202.0 | 1035.5 |
| SYNZIP45 | 490.5 | 526.3 | 1257.0 | -192.4 | 709.8 | 328.3 | 70.5 | 819.4 | 68.3 | 3057.6 |
| SYNZIP46 | -522.3 | 227.5 | 436.6 | -2569.1 | 338.4 | 1545.9 | 10.5 | 5615.1 | -214.1 | -678.8 |
| SYNZIP47 | -222.6 | 273.6 | -104.3 | 398.9 | -310.8 | -296.8 | 128.0 | 8248.0 | 332.4 | 746.9 |
| SYNZIP48 | -348.5 | 458.6 | -195.3 | 2779.8 | 1187.4 | 12065.8 | 21439.1 | 2571.6 | 1855.4 | 2462.5 |
| BATF | 715.1 | 480.8 | 1645.9 | -484.8 | 421.5 | 664.3 | 346.8 | 949.0 | 671.9 | 1091.0 |
| FOS | 1418.5 | 973.3 | -145.6 | -163.3 | 614.9 | 3601.5 | 661.8 | 383.3 | 183.0 | 2417.6 |
| ATF4 | 125.1 | 926.0 | 941.5 | 3857.8 | 412.1 | 2929.0 | 258.3 | 2005.5 | 5058.0 | 5382.0 |
| ATF3 | 908.3 | 512.9 | 2961.4 | 1974.9 | 1449.8 | 370.6 | 678.5 | -242.5 | 1625.8 | 3603.0 |
| BACH1 | 1611.0 | 930.0 | -0.4 | -1432.5 | 1162.5 | 3202.6 | 11239.8 | 1899.9 | 966.4 | 2435.1 |
| JUND | -31.4 | 601.3 | -140.1 | -861.4 | 225.8 | 345.0 | 303.1 | 1184.3 | 17584.1 | 22293.1 |
| NFE2L3 | -994.8 | 291.3 | -21.9 | -611.5 | 0.8 | 265.9 | 55.8 | -742.8 | -260.3 | -1926.1 |

| Protein | ATF4 | ATF3 | BACH1 | JUND | NFE2L3 | SYNZIP5(2) | SYN37(2) | SYNZIP6(2) | ATF4(2) | FOS(2) |
|---|---|---|---|---|---|---|---|---|---|---|
| SYNZIP29 | 8674.4 | 4866.3 | 11467.6 | 9058.4 | 5495.8 | 4790.6 | 14756.8 | 9066.6 | 6620.0 | 11400.4 |
| SYNZIP30 | 12458.9 | 7376.4 | 17800.8 | 11935.9 | 9128.6 | 6310.1 | 2989.5 | 15460.1 | 10116.8 | 13449.4 |
| SYNZIP31 | 6120.8 | 2839.4 | 206.8 | 3165.5 | 13.1 | 3248.3 | 223.0 | 44.8 | 4529.0 | 364.6 |
| SYNZIP32 | 5014.3 | 7568.0 | 66.9 | 6813.9 | 632.1 | 9805.4 | 63.9 | -606.8 | 4410.9 | 12184.0 |
| SYNZIP33 | 18008.4 | 4339.4 | 4169.5 | 8353.5 | 8524.8 | 824.9 | 817.8 | 978.0 | 13227.0 | 4968.4 |
| SYNZIP34 | 17674.8 | 19232.9 | 675.0 | 5185.3 | 1235.3 | 717.4 | 1362.0 | 505.0 | 14255.3 | 9653.9 |
| SYNZIP35 | 1989.5 | 2626.6 | 5027.8 | 2326.6 | 1194.5 | 935.6 | 728.3 | 3840.1 | 1924.3 | 3736.6 |
| SYNZIP36 | 1753.6 | 3381.1 | 1060.8 | 1372.9 | 507.9 | -205.5 | -58.3 | 1580.8 | 1919.3 | 863.5 |
| SYNZIP37 | 1973.4 | 1903.8 | 4773.3 | 2179.6 | 1959.8 | -202.5 | 439.5 | 176.4 | 2504.4 | 4514.6 |
| SYNZIP38 | 2727.0 | 4894.4 | 4388.8 | 1231.8 | 772.3 | 275.1 | 3456.8 | -287.0 | 2642.4 | 19993.1 |
| SYNZIP39 | 2422.0 | 3492.3 | 7314.3 | 4994.8 | 77.8 | 907.6 | 4679.5 | 366.8 | 958.0 | 12813.9 |
| SYNZIP40 | 13239.0 | -896.0 | 7118.1 | 2705.5 | 1785.1 | 12696.8 | 2756.1 | -333.6 | 10121.5 | 6037.3 |
| SYNZIP41 | 669.0 | 536.8 | 948.8 | 424.5 | 459.6 | -156.8 | -288.1 | -115.5 | 1025.1 | 1602.4 |
| SYNZIP42 | 4351.1 | 746.3 | 338.3 | 358.0 | 813.3 | 95.1 | 455.0 | 17.4 | 3206.3 | 2475.9 |
| SYNZIP43 | 2988.9 | 9388.4 | -141.5 | 1247.6 | 221.0 | 3747.8 | 145.0 | 5225.3 | 2600.9 | 3998.1 |
| SYNZIP44 | 5835.3 | 8621.6 | 403.6 | 5087.1 | 71.4 | 5305.9 | -94.8 | 798.0 | 5829.6 | 1247.5 |
| SYNZIP45 | 3345.9 | 7287.3 | 1439.1 | 4763.0 | 199.6 | 77.6 | 239.0 | 1796.3 | 1648.9 | 4256.1 |
| SYNZIP46 | -419.8 | -462.6 | 355.1 | 1211.5 | 833.6 | -538.0 | -79.3 | -553.3 | 400.1 | -244.4 |
| SYNZIP47 | 686.8 | 660.0 | 4774.5 | 453.3 | -300.0 | 211.8 | -673.0 | 122.0 | 521.9 | 77.4 |
| SYNZIP48 | 5761.4 | 2081.0 | 10982.6 | 1879.3 | 1272.8 | 4041.9 | 115.4 | 1665.5 | 4930.1 | 2420.0 |
| BATF | 8019.1 | 7824.9 | 4464.8 | 35898.8 | 1824.6 | 2706.3 | 768.1 | -666.8 | 7322.9 | 2072.8 |
| FOS | 6565.6 | 3277.9 | 2061.5 | 26790.8 | 676.1 | 660.6 | 2984.5 | 876.5 | 5317.8 | 3855.9 |
| ATF4 | 881.5 | 15257.6 | 2381.3 | 1593.0 | 8034.4 | 1142.5 | 1138.3 | 119.3 | 683.9 | 6049.3 |
| ATF3 | 9937.0 | 1793.8 | 491.5 | 20496.4 | -409.3 | 1160.8 | 1131.5 | 1854.1 | 6762.9 | 3957.6 |
| BACH1 | 3553.5 | 472.5 | 3295.6 | 7532.3 | 524.0 | 170.1 | 5262.9 | 283.0 | 3410.0 | 4463.4 |
| JUND | 987.4 | 25712.8 | -84.5 | 2800.8 | 25.6 | 1688.8 | 87.5 | 2166.8 | 1413.1 | 22039.5 |
| NFE2L3 | 1456.4 | -1875.4 | -295.3 | -1931.3 | 382.1 | -293.5 | 272.5 | -146.8 | 1664.4 | -1536.9 |

Table C.S3. List of the proteins composing each of the subnetworks identified.

| **2 nodes** | **Motif** | **A** | **B** | | | | |
|---|---|---|---|---|---|---|---|
| pairs | A-B | SYNZIP5 | SYNZIP6 | | | | |
| pairs | A-B | SYNZIP20 | SYNZIP11 | | | | |
| pairs | A-B | SYNZIP20 | SYNZIP16 | | | | |
| pairs | A-B | SYNZIP20 | SYNZIP2 | | | | |
| pairs | A-B | SYNZIP13 | SYNZIP15 | | | | |
| pairs | A-B | SYNZIP16 | SYNZIP23 | | | | |
| pairs | A-B | SYNZIP12 | SYNZIP23 | | | | |
| pairs | A-B | SYNZIP22 | SYNZIP19 | | | | |
| pairs | A-B | SYNZIP22 | SYNZIP14 | | | | |
| pairs | A-B | SYNZIP22 | SYNZIP10 | | | | |
| pairs | A-B | SYNZIP22 | ATF4 | | | | |
| pairs | A-B | SYNZIP2 | SYNZIP14 | | | | |
| pairs | A-B | SYNZIP2 | SYNZIP1 | | | | |
| pairs | A-B | SYNZIP19 | SYNZIP18 | | | | |
| pairs | A-B | SYNZIP19 | SYNZIP21 | | | | |
| pairs | A-B | SYNZIP15 | ATF4 | | | | |
| pairs | A-B | SYNZIP9 | FOS | | | | |
| pairs | A-B | SYNZIP17 | SYNZIP18 | | | | |
| pairs | A-B | SYNZIP17 | FOS | | | | |
| pairs | A-B | SYNZIP18 | SYNZIP23 | | | | |
| pairs | A-B | SYNZIP23 | SYNZIP21 | | | | |
| pairs | A-B | SYNZIP23 | SYNZIP6 | | | | |
| pairs | A-B | SYNZIP23 | SYNZIP8 | | | | |
| pairs | A-B | SYNZIP23 | BATF | | | | |
| pairs | A-B | SYNZIP21 | SYNZIP4 | | | | |
| pairs | A-B | SYNZIP3 | SYNZIP4 | | | | |
| pairs | A-B | SYNZIP7 | ATF4 | | | | |
| **3 nodes** | **Motif** | **A** | **B** | **C** | | | |
| line | A-B-C | SYNZIP5 | SYNZIP6 | SYNZIP5 | | | |
| line | A-B-C | SYNZIP11 | SYNZIP20 | SYNZIP16 | | | |
| line | A-B-C | SYNZIP11 | SYNZIP20 | SYNZIP2 | | | |
| line | A-B-C | SYNZIP20 | SYNZIP2 | SYNZIP14 | | | |
| line | A-B-C | SYNZIP20 | SYNZIP2 | SYNZIP1 | | | |
| line | A-B-C | SYNZIP13 | SYNZIP15 | ATF4 | | | |
| line | A-B-C | SYNZIP16 | SYNZIP23 | SYNZIP16 | | | |
| line | A-B-C | SYNZIP16 | SYNZIP23 | SYNZIP16 | | | |
| line | A-B-C | SYNZIP16 | SYNZIP23 | SYNZIP6 | | | |
| line | A-B-C | SYNZIP16 | SYNZIP23 | SYNZIP8 | | | |
| line | A-B-C | SYNZIP16 | SYNZIP23 | BATF | | | |
| line | A-B-C | SYNZIP12 | SYNZIP23 | SYNZIP12 | | | |
| line | A-B-C | SYNZIP12 | SYNZIP23 | SYNZIP6 | | | |
| line | A-B-C | SYNZIP12 | SYNZIP23 | SYNZIP8 | | | |
| line | A-B-C | SYNZIP12 | SYNZIP23 | BATF | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| line | A-B-C | SYNZIP19 | SYNZIP22 | SYNZIP14 | | | |
| line | A-B-C | SYNZIP19 | SYNZIP22 | SYNZIP10 | | | |
| line | A-B-C | SYNZIP19 | SYNZIP22 | ATF4 | | | |
| line | A-B-C | SYNZIP14 | SYNZIP22 | SYNZIP10 | | | |
| line | A-B-C | SYNZIP10 | SYNZIP22 | ATF4 | | | |
| line | A-B-C | SYNZIP14 | SYNZIP2 | SYNZIP1 | | | |
| line | A-B-C | SYNZIP19 | SYNZIP18 | SYNZIP19 | | | |
| line | A-B-C | SYNZIP19 | SYNZIP18 | SYNZIP23 | | | |
| line | A-B-C | SYNZIP18 | SYNZIP19 | SYNZIP21 | | | |
| line | A-B-C | SYNZIP19 | SYNZIP21 | SYNZIP19 | | | |
| line | A-B-C | SYNZIP19 | SYNZIP21 | SYNZIP4 | | | |
| line | A-B-C | SYNZIP15 | ATF4 | SYNZIP15 | | | |
| line | A-B-C | SYNZIP9 | FOS | SYNZIP9 | | | |
| line | A-B-C | SYNZIP18 | SYNZIP17 | FOS | | | |
| line | A-B-C | SYNZIP18 | SYNZIP23 | SYNZIP21 | | | |
| line | A-B-C | SYNZIP18 | SYNZIP23 | SYNZIP6 | | | |
| line | A-B-C | SYNZIP18 | SYNZIP23 | SYNZIP8 | | | |
| line | A-B-C | SYNZIP18 | SYNZIP23 | BATF | | | |
| line | A-B-C | SYNZIP21 | SYNZIP23 | SYNZIP6 | | | |
| line | A-B-C | SYNZIP21 | SYNZIP23 | SYNZIP8 | | | |
| line | A-B-C | SYNZIP21 | SYNZIP23 | BATF | | | |
| line | A-B-C | SYNZIP6 | SYNZIP23 | SYNZIP8 | | | |
| line | A-B-C | SYNZIP6 | SYNZIP23 | BATF | | | |
| line | A-B-C | SYNZIP8 | SYNZIP23 | BATF | | | |
| line | A-B-C | SYNZIP21 | SYNZIP4 | SYNZIP21 | | | |
| **4 nodes** | **Motif** | **A** | **B** | **C** | **D** | | |
| 2 pairs | A-B,C-D | SYNZIP5 | SYNZIP6 | SYNZIP13 | SYNZIP15 | | |
| 2 pairs | A-B,C-D | SYNZIP5 | SYNZIP6 | SYNZIP1 | SYNZIP2 | | |
| 2 pairs | A-B,C-D | SYNZIP5 | SYNZIP6 | ATF4 | SYNZIP15 | | |
| 2 pairs | A-B,C-D | SYNZIP5 | SYNZIP6 | FOS | SYNZIP9 | | |
| 2 pairs | A-B,C-D | SYNZIP5 | SYNZIP6 | SYNZIP7 | ATF4 | | |
| 2 pairs | A-B,C-D | SYNZIP20 | SYNZIP11 | SYNZIP15 | SYNZIP13 | | |
| 2 pairs | A-B,C-D | SYNZIP22 | SYNZIP19 | SYNZIP3 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP22 | ATF4 | SYNZIP3 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP2 | SYNZIP1 | SYNZIP21 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP2 | SYNZIP1 | SYNZIP3 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP15 | ATF4 | SYNZIP3 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP9 | FOS | SYNZIP21 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP9 | FOS | SYNZIP3 | SYNZIP4 | | |
| 2 pairs | A-B,C-D | SYNZIP9 | FOS | SYNZIP7 | ATF4 | | |
| 2 pairs | A-B,C-D | SYNZIP17 | FOS | SYNZIP7 | ATF4 | | |
| 2 pairs | A-B,C-D | SYNZIP23 | SYNZIP6 | SYNZIP7 | ATF4 | | |
| 2 pairs | A-B,C-D | SYNZIP23 | SYNZIP8 | SYNZIP7 | ATF4 | | |
| 2 pairs | A-B,C-D | SYNZIP3 | SYNZIP4 | SYNZIP7 | ATF4 | | |
| hub | A-B,A-C,A-D | SYNZIP2 | SYNZIP1 | SYNZIP20 | SYNZIP14 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP16 | SYNZIP12 | SYNZIP18 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP6 | SYNZIP12 | SYNZIP16 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP12 | SYNZIP16 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP12 | SYNZIP16 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP6 | SYNZIP18 | SYNZIP16 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP18 | SYNZIP16 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP18 | SYNZIP16 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP16 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP16 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP16 | SYNZIP8 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP6 | SYNZIP18 | SYNZIP12 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP18 | SYNZIP12 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP18 | SYNZIP12 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP12 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP12 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP12 | SYNZIP8 | | |
| hub | A-B,A-C,A-D | SYNZIP22 | SYNZIP10 | SYNZIP19 | SYNZIP14 | | |
| hub | A-B,A-C,A-D | SYNZIP22 | ATF4 | SYNZIP19 | SYNZIP10 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP6 | SYNZIP18 | SYNZIP21 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP18 | SYNZIP21 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP18 | SYNZIP21 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP18 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP18 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP18 | SYNZIP8 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | SYNZIP8 | SYNZIP21 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP21 | SYNZIP6 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP21 | SYNZIP8 | | |
| hub | A-B,A-C,A-D | SYNZIP23 | BATF | SYNZIP6 | SYNZIP8 | | |
| line | A-B-C-D | SYNZIP5 | SYNZIP6 | SYNZIP23 | SYNZIP8 | | |
| line | A-B-C-D | SYNZIP13 | SYNZIP15 | ATF4 | SYNZIP7 | | |
| line | A-B-C-D | SYNZIP19 | SYNZIP21 | SYNZIP4 | SYNZIP3 | | |
| box | A-B-C-D-A | SYNZIP21 | SYNZIP23 | SYNZIP18 | SYNZIP19 | | |
| **5 nodes** | **Motif** | **A** | **B** | **C** | **D** | **E** | |
| pair+line | A-B,C-D-E | SYNZIP5 | SYNZIP6 | SYNZIP13 | SYNZIP15 | ATF4 | |
| pair+line | A-B,C-D-E | SYNZIP5 | SYNZIP6 | SYNZIP15 | ATF4 | SYNZIP7 | |
| pair+line | A-B,C-D-E | SYNZIP7 | ATF4 | SYNZIP5 | SYNZIP6 | SYNZIP23 | |
| pair+line | A-B,C-D-E | SYNZIP3 | SYNZIP4 | SYNZIP19 | SYNZIP22 | ATF4 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| pair+line | A-B,C-D-E | SYNZIP2 | SYNZIP1 | SYNZIP21 | SYNZIP4 | SYNZIP3 | |
| pair+line | A-B,C-D-E | SYNZIP3 | SYNZIP4 | SYNZIP15 | ATF4 | SYNZIP7 | |
| pair+line | A-B,C-D-E | SYNZIP7 | ATF4 | SYNZIP9 | FOS | SYNZIP17 | |
| pair+line | A-B,C-D-E | SYNZIP9 | FOS | SYNZIP21 | SYNZIP4 | SYNZIP3 | |
| pair+line | A-B,C-D-E | SYNZIP7 | ATF4 | SYNZIP8 | SYNZIP23 | SYNZIP6 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP18 | SYNZIP16 | SYNZIP6 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP18 | SYNZIP16 | SYNZIP8 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP18 | SYNZIP16 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP16 | SYNZIP6 | SYNZIP8 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP16 | SYNZIP6 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP16 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP16 | SYNZIP6 | SYNZIP8 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP16 | SYNZIP6 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP16 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP16 | SYNZIP6 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP12 | SYNZIP6 | SYNZIP8 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP12 | SYNZIP6 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP12 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP12 | SYNZIP6 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP21 | SYNZIP6 | SYNZIP8 | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP21 | SYNZIP6 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP21 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP18 | SYNZIP6 | SYNZIP8 | BATF | |
| hub | A-B,A-C,A-D,A-E | SYNZIP23 | SYNZIP21 | SYNZIP6 | SYNZIP8 | BATF | |
| **6 nodes** | **Motif** | **A** | **B** | **C** | **D** | **E** | **F** |
| line + pair | A-B-C-D, E-F | SYNZIP5 | SYNZIP6 | SYNZIP23 | SYNZIP8 | ATF4 | SYNZIP7 |
| line + pair | A-B-C-D, E-F | SYNZIP13 | SYNZIP15 | ATF4 | SYNZIP7 | SYNZIP5 | SYNZIP6 |
| 3 pairs | A-B,C-D,E-F | SYNZIP9 | FOS | ATF4 | SYNZIP7 | SYNZIP5 | SYNZIP6 |
| 3 pairs | A-B,C-D,E-F | SYNZIP9 | FOS | ATF4 | SYNZIP7 | SYNZIP4 | SYNZIP3 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP16 | BATF | SYNZIP8 | SYNZIP6 | SYNZIP18 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP16 | BATF | SYNZIP8 | SYNZIP6 | SYNZIP12 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP16 | BATF | SYNZIP8 | SYNZIP18 | SYNZIP12 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP16 | BATF | SYNZIP6 | SYNZIP18 | SYNZIP12 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP16 | SYNZIP8 | SYNZIP6 | SYNZIP18 | SYNZIP12 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP12 | BATF | SYNZIP8 | SYNZIP6 | SYNZIP18 |
| hub | A-B,A-C,A-D,A-E,A-F | SYNZIP23 | SYNZIP18 | BATF | SYNZIP8 | SYNZIP6 | SYNZIP21 |

Table C.S4. Crystallographic data collection and refinement statistics.

| Protein<br>Data Set | SYNZIP6:<br>SYNZIP5 | SYNZIP1:<br>SYNZIP2 |
|---|---|---|
| Space Group | P 63 | P 31 |
| Cell dimensions<br>*a*, *b*, *c*(Å)<br>α, β, γ(°) | <br>82.7, 82.7, 150.6<br>90, 90, 120 | <br>49.9, 49.9, 113.2<br>90, 90, 120 |
| λ (Å) | 0.97927 | 1.5418 |
| Resolution (Å) | 50 - 2.46 | 50 - 1.75 |
| Rsym (%)[a,b] | 10.9 (54.8) | 3.8 (29.4) |
| # ref | 21204 | 31354 |
| Completeness (%)[a] | 99.7 (99.3) | 98.2 (90.7) |
| Redundancy[a] | 5.8 (5.4) | 4.6 (2.8) |
| # dimers/ASU | 4 | 3 |
| Twin law | h,-h-k,-l | -k,-h,-l |
| Twin fraction | 0.324 | 0.392 |
| $R_{work}/R_{free}$(%)[c] | 21.2/25.8 | 19.0/22.8 |

[a]Values in parentheses refers to data in the highest resolution shell

[b]$R_{sym} = \Sigma_h\Sigma_j |I_j(h) - \langle I(h)\rangle| / \Sigma_h\Sigma_j \langle I(h)\rangle$, where $I_j(h)$ is the $j^{th}$ reflection of index h and $\langle I(h)\rangle$ is the average intensity of all observations of I(h)

[c]$R_{work} = \Sigma_h |F_{obs}(h) - F_{calc}(h)| | / \Sigma_h |F_{obs}(h)|$, calculated over the 95% of the data in the working set. $R_{free}$ equivalent to $R_{work}$ except calculated over the 5% of the data assigned to the test set

## REFERENCES

1. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009 Apr 16;458(7240):859-64.

2. Crooks GE, Hon G, Chandonia J, Brenner SE. WebLogo: A sequence logo generator. Genome Research. 2004 June 2004;14(6):1188-90.

# APPENDIX D

## Design of peptide inhibitors that bind the bZIP domain of Epstein-Barr virus protein BZLF1

**Reproduced with permission of Elsevier B.V. from**

Chen, T. S., Reinke, A. W. & Keating, A. E. Design of peptide inhibitors that bind the bZIP domain of Epstein-Barr virus protein BZLF1. *J Mol Biol* 408, 304-20

**Collaborator notes**

Aaron Reinke performed the electrophoretic mobility shift assay

## ABSTRACT

Designing proteins or peptides that bind native protein targets can aid the development of novel reagents and/or therapeutics. Rational design also tests our understanding of the principles underlying protein recognition. This article describes several strategies used to design peptides that bind to the basic region leucine zipper (bZIP) domain of the viral transcription factor BZLF1, which is encoded by the Epstein-Barr virus. BZLF1 regulates the transition of the Epstein-Barr virus from a latent state to a lytic state. It shares some properties in common with the more studied human bZIP transcription factors, but also includes novel structural elements that pose interesting challenges to inhibitor design. In designing peptides that bind to BZLF1 by forming a coiled-coil structure, we considered both affinity for BZLF1 and undesired self-association, which can weaken the effectiveness of an inhibitor. Several designed peptides exhibited different degrees of target-binding affinity and self-association. Rationally engineered molecules were more potent inhibitors of DNA binding than a control peptide corresponding to the native BZLF1 dimerization region itself. The most potent inhibitors included both positive and negative design elements and exploited interaction with the coiled-coil and basic DNA-binding regions of BZLF1.

## INTRODUCTION

The basic-region leucine-zipper (bZIP) transcription factors are a large class of proteins conserved in eukaryotes and several viruses that regulate a wide range of biological processes. The structure of bZIP-DNA complexes is very simple: a helical and positively charged DNA-binding region is contiguous with a coiled coil that mediates protein homo- or hetero-dimerization (O'Shea, et al. 1991). The bZIP coiled-coil helices wrap around one another in a parallel orientation with “knobs-into-holes” side-chain packing geometry, and a 7-amino-acid

heptad repeat characterizes the structure, in which each residue can be assigned a register position labeled *a* through *g* (Figure D.1). High-affinity binding of bZIP transcription factors to DNA requires protein dimerization.

Given the many important biological roles of the bZIPs, molecules that selectively disrupt bZIP-DNA interactions could be valuable reagents and even potential therapeutics. Several strategies have been reported for identifying inhibitors. Small molecules have been discovered via high-throughput screening, (Rishi, et al. 2005)and peptides that bind to the coiled-coil regions of the bZIPs and disrupt dimer formation have been selected from targeted combinatorial libraries (Mason, et al. 2009, Mason, et al. 2007, Mason, et al. 2006).A particularly effective strategy for blocking bZIP-DNA interactions was developed by Vinson and co-workers, who created a series of dominant-negative peptide inhibitors by replacing the basic regions of certain bZIP proteins with a sequence enriched in negatively charged residues (the "acidic extension"), giving so-called A-ZIPs (Acharya, et al. 2006b, Ahn, et al. 1998, Olive, et al. 1997, Krylov, et al. 1995). The A-ZIPs bind tightly and selectively to bZIPs and have been used to study the effects of inhibiting dimerization and hence DNA binding in both cell culture and animal models (Oh, et al. 2007, Gerdes, et al. 2006).

Current understanding of bZIP coiled-coil interactions has also enabled the computational design of synthetic peptides to block bZIP dimerization. Significant effort has been dedicated to elucidating sequence determinants governing the interactions of bZIP coiled coils, and to developing predictive computational models that capture these. Several types of residue-pair interactions that are important for specificity have been characterized in detail over the past 20 years, and models derived from physics-based calculations, machine learning, and experimentally measured coupling energies have been developed to explain and predict bZIP

coiled-coil interactions (Mason, et al. 2006, Grigoryan and Keating. 2006, Fong, et al. 2004, Krylov, et al. 1994, Acharya, et al. 2006a, Steinkruger, et al. 2010). Using such binding models, Grigoryan et al. recently designed a series of peptides that bind to targets in 19 out of 20 human bZIP families (Grigoryan, et al. 2009).

An interesting issue in the study of bZIP interactions is specificity. Given the similarities among sequences, and the many bZIPs in most eukaryotes, a large number of homo- and heterodimers can potentially form. Interactions among human bZIPs have been shown to be highly selective when assayed *in vitro*,(Vinson, et al. 2006, Newman and Keating. 2003) but it can be difficult to achieve specificity in designed bZIP-like peptides. In particular, peptides engineered to bind to bZIP coiled-coil regions have been shown to self-associate strongly and also interact with undesired partners, (Mason, et al. 2007, Grigoryan, et al. 2009) In this work we address considerations of both affinity and anti-homodimer specificity in the design of peptide inhibitors for a viral bZIP protein, BZLF1.

BZLF1 (Zta, ZEBRA, EB1) is encoded by the Epstein-Barr virus (EBV) and triggers the virus's latent to lytic switch by functioning as a transcription factor and regulator of DNA replication. (Countryman, et al. 1987, Schepers, et al. 1993, Feederle, et al. 2000, Liu and Speck. 2003)Infection by EBV has been linked to several human malignancies such as Hodgkin's disease and Burkitt's lymphoma (Young and Rickinson. 2004). The basic region of BZLF1 is highly homologous to that of human bZIPs and is responsible for direct contact with DNA; a coiled-coil region immediately C-terminal to the basic helix mediates dimerization. However, a recent crystal structure and other biochemical studies have revealed several unique features of BZLF1 (Figure D.1a) (Petosa, et al. 2006, Schelcher, et al. 2007). The coiled-coil region at the dimerization interface is only 4 heptads long, whereas the coiled-coil regions of human bZIPs

typically contain at least 5 heptads. Furthermore, only one of the four BZLF1 coiled-coil heptads includes a leucine residue at the *d* position; this residue occurs with much higher frequency in human bZIP sequences (hence the name "leucine zipper"). The stability of the BZLF1 homodimer is significantly enhanced by a unique C-terminal (CT) region that folds back on the coiled coil to form additional contacts; (Schelcher, et al. 2007)the CT region is only partially observed in the crystal structure. Prior work using peptide arrays showed that BZLF1 constructs corresponding to the coiled coil or the coiled coil plus the CT region homo-associate in preference to binding any of 33 representative human bZIP proteins (Reinke, et al. 2010b).

It has been shown that a peptide corresponding to the coiled-coil region of BZLF1, lacking the DNA binding residues, inhibits BZLF1 binding to DNA at high micromolar concentrations (Hicks, et al. 2003). In this work, we sought new peptides that would mimic the coiled-coil interface of the native structure yet provide more potent inhibition of DNA binding. As a design target, BZLF1 is both simpler and more complex than human and viral bZIPs that have been the subjects of previous computational design studies (Grigoryan, et al. 2009, Reinke, et al. 2010b). It is simpler because of its unique structural features, which make coiled-coil inhibitors designed to target it unlikely to interact broadly with other bZIP proteins. However, it is more complex because the CT region and unusually tight helix packing make the interface unlike the dimerization domains of better-understood bZIPs (Petosa, et al. 2006). Here we explore the extent to which previously applied design strategies can be used successfully in the context of BZLF1. Throughout our analyses, we explicitly addressed two design criteria: affinity for BZLF1 and design self-association, which is an undesirable trait for an inhibitor. The best inhibitor incorporated both elements and included modifications of BZLF1 in both the coiled-

coil and DNA-binding regions. As assessed using DNA-binding gel-shift assays, this designed peptide was much more potent than one corresponding to the native dimerization domain.

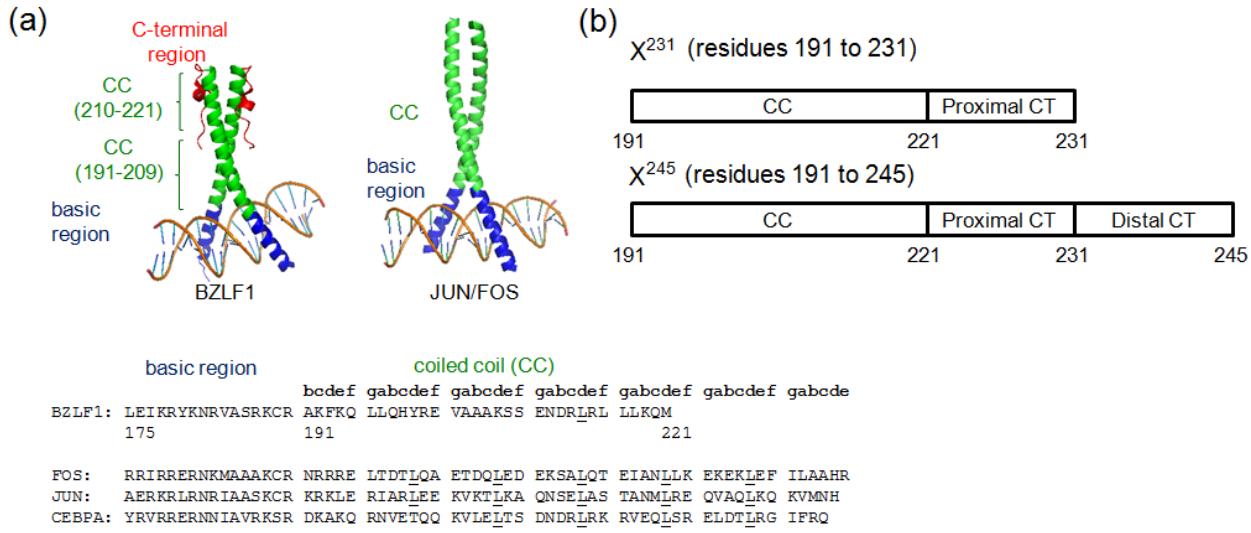
## RESULTS

*Computational design of a peptide to bind the N-terminal part of the BZLF1 coiled coil*

Our goal was to identify variants of the BZLF1 dimerization domain that would function as more effective dominant negative inhibitors of DNA binding. As described in the Introduction, BZLF1 possesses several unique features as a bZIP design target. These include the unconventional, short coiled-coil segment and the CT region. The CT can be divided into the proximal CT (residues 222 - 231) and the less structured distal CT (residues 232 – 246), as shown in Figure D.1b. We began by re-designing the N-terminal two and a half heptads of the BZLF1 coiled coil (residues 191 – 209, Fig 2.2b), because we anticipated that this segment would provide the greatest opportunity to improve affinity and heterodimer specificity over the native sequence. Residues 210 – 221 also form part of the coiled-coil structure, but additionally engage in non-coiled-coil hydrophobic contacts with the proximal CT as observed in the crystal structure (Figure D.1a). Thus, in order to maintain this stabilizing interaction, these residues were not changed in the design.

Both the desired design-target heterodimer and the undesired design homodimer were modeled as parallel, blunt ended coiled coils. We used the CLASSY protein-design algorithm to choose residues at 10 sites in the design, optimizing the predicted affinity of the design-target complex (Grigoryan, et al. 2009). The scoring function used was based on a hybrid model that included both physics-based and experimentally derived terms and is described further in the

Methods. The optimal-affinity design, which we call $BD_{cc}$ (BZLF1 design against the coiled-coil region, shown in Figure D.2c), was predicted to be hetero-specific. In design energy units the predicted stabilities were as follows: BZLF1 homodimer: -29 kcal/mol, $BD_{cc}$ homodimer: -32 kcal/mol, BZLF1/$BD_{cc}$ heterodimer: -44 kcal/mol. Although the score for the design self-interaction was close to that for native BZLF1 coiled-coil homodimerization, the score for the design-target interaction was significantly better. Thus, although CLASSY can be used to improve specificity against undesired states as well as affinity for a target, (Grigoryan, et al. 2009) this was predicted not to be necessary in this case.




```
                basic region                coiled coil (CC)
                             bcdef gabcdef gabcdef gabcdef gabcdef gabcdef gabcde
BZLF1: LEIKRYKNRVASRKCR AKFKQ LLQHYRE VAAAKSS ENDRLRL LLKQM
       175               191                          221

FOS:   RRIRRERNKMAAAKCR NRRRE LTDTLQA ETDQLED EKSALQT EIANLLK EKEKLEF ILAAHR
JUN:   AERKRLRNRIAASKCR KRKLE RIARLEE KVKTLKA QNSELAS TANMLRE QVAQLKQ KVMNH
CEBPA: YRVRRERNNIAVRKSR DKAKQ RNVETQQ KVLELTS DNDRLRK RVEQLSR ELDTLRG IFRQ
```

Figure D.1 Sequence and structure of the BZLF1 bZIP domain.
(a) Crystal structure of BZLF1 bound to DNA26 (PDB ID 2C9L, left) compared to human JUN/FOS bound to DNA (Glover and Harrison. 1995) (PDB ID 1FOS, right). The basic region is blue, the coiled coil is green, and the C-terminal (CT) region is red. At the bottom are sequence alignments for the basic and coiled-coil regions of BZLF1 and representative human bZIPs. Leucines at d positions in the coiled coils are underlined. (b) Scheme of constructs used in this study. The "231" construct includes the coiled coil (CC) and the proximal C-terminal (CT) region; the "245" construct includes the coiled coil (CC) and the full-length C-terminal (CT) region.

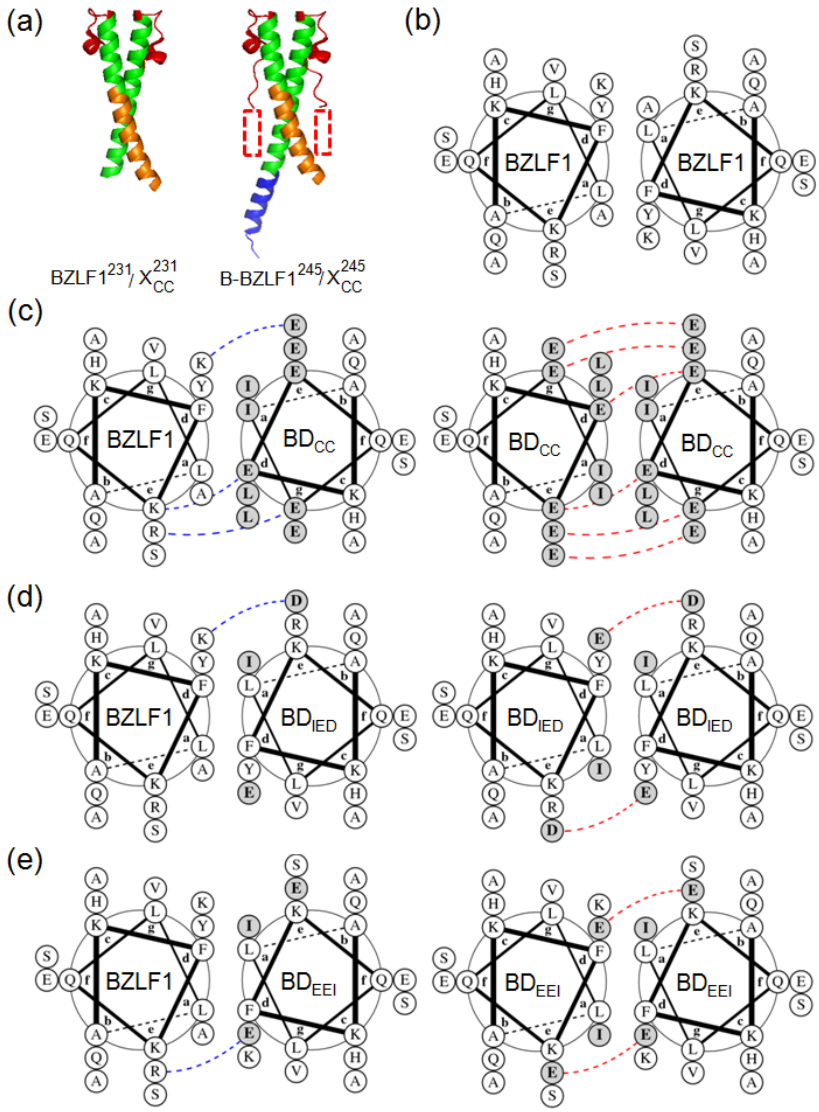



Figure D.2 Designed inhibitors.
(a) Structural models representing two types of design-BZLF1 complexes tested in this work. At left, the "231" constructs, and at right, the "245" constructs. "X" is a placeholder for the name of a design, e.g. $BD_{cc}$. Color is as in Figure D.1a except that the designed region is shown in orange. The dashed boxes in the "245" complex indicate that part of the distal CT (237-245) is not resolved in the crystal structure. (b) Helical wheel diagram for the BZLF1 homodimer. (c-e) Helical wheel diagrams for the designs. On the left are design-target heterodimers and on the right are design homodimers. Design residues are highlighted in bold and with a grey background. Potential electrostatic interactions are indicated in blue if attractive and red if repulsive. (c) Design $BD_{cc}$, (d) Design $BD_{IED}$, (e) Design $BD_{EEI}$. In all helical wheel diagrams, only residues from *b* position 191 (Ala) to *f* position 209 (Ser) are shown (this region is orange in Figure D.2a), with the helix proceeding from N-to-C terminus into the page. Diagrams generated using DrawCoil 1.0 (http://www.gevorggrigoryan.com/drawcoil/).

The $BD_{cc}$ solution populated most *a* and *d* positions (coiled-coil "core" positions) with Ile and Leu respectively, which are very common in conventional bZIP sequences (Figure D.2c). A single *d*-position Glu residue at the extreme N terminus of the coiled coil is uncharacteristic of bZIP sequences, but was predicted to interact favorably with an *e*-position Lys on BZLF1. The five designed *e* and *g* positions (coiled-coil "edge" positions) were all populated with glutamate for improved electrostatic interactions with the target, where three residues in this region are positively charged. Interestingly, predicted charged interactions involved both edge-to-edge (e.g. *g* to *e'*) and core-to-edge (*d* to *e'*) residues in the BZLF1 target, as was previously observed for anti-human bZIP designs (Grigoryan, et al. 2009). Although core sites occupied by Ile and Leu favor design self-interaction, the charged residues at *e* and *g* are predicted to disfavor it. Charge repulsion is a commonly observed negative design element in many native and model coiled coils (O'Shea, et al. 1993, Vinson, et al. 2002, Woolfson. 2005, Grigoryan and Keating. 2008).

The anti-BZLF1 peptide was cloned in the context of residues 191- 231 of BZLF1. This construct, $BD_{cc}^{231}$, includes the entire coiled-coil domain and the proximal CT (Figure D.1b, D.2a, Table D.1), potentially retaining native interactions observed in the X-ray structure between the C-terminal part of the coiled coil and the CT region. Because the residues optimized in the design calculations are more than 8 Å away from residue I231 in the modeled structure (Figure D.2a), the proximal CT excluded from the calculations was not expected to significantly influence the results. Potential interactions between the designed residues and the distal CT, which are not evident in the crystal structure but are suggested by prior studies (Schelcher, et al. 2007), are addressed in experiments described below.

We used circular dichroism (CD) spectroscopy to study the interaction properties of $BD_{cc}^{231}$. Thermal denaturation experiments showed that the $BD_{cc}^{231}$ homo-oligomer is destabilized

compared to the target homodimer in the same sequence context ($BZLF1^{231}$, residues 191 to 231); $T_m$ values were 38 °C vs. 43 °C (Figure D.3a and Table D.1). The hetero-complex between $BD_{CC}^{231}$ and $BZLF1^{231}$ ($T_m$ of 53 °C, Table D.2) was significantly stabilized compared to the $BZLF1^{231}$ homodimer. We conclude that the $BD_{CC}^{231}$ design is very hetero-specific, consistent with expectations based on the design algorithm. The agreement indicates success of the automated CLASSY approach even on a target with a sequence quite different from the human bZIPs.

*Designs with weaker self-association*

The $BD_{cc}$ design achieved hetero-specificity mostly by improving design-target affinity compared to the native BZLF1 complex. We also sought solutions that achieved hetero-specificity against the same target (the N-terminal part of the BZLF1 coiled coil) by weakening design self-interaction. Toward this end, we tested a negative design strategy that placed charged residues at a core *d* position and the adjacent *e* position such that they would create a local cluster of 4 negative charges in the modeled design coiled-coil homodimer. There are 3 close inter-chain pair contacts in such a cluster (2 *d-e'* interactions and one *d-d'* interaction). We observed variations of this strategy in design solutions obtained using the CLASSY algorithm when optimizing affinity for the target under increasingly stringent constraints limiting the stability of the design homodimer.

We picked two sets of amino-acid changes, (K207E, S208D) and (Y200E, R201E), each corresponding to the (*d, e'*) negative design strategy described above. We also included one stabilizing design element present in the $BD_{CC}^{231}$ solution, A204I (substituting Ile for Ala at an *a* position), to compensate for a potential loss in stability due to the introduction of charge in the

core. The resulting two designs were cloned, expressed and purified, again in the context of BZLF1 residues 191 to 231 (204I, 207E, 208D, referred to as $BD^{231}_{IED}$, and 200E, 201E, 204I, referred to as $BD^{231}_{EEI}$, Fig 2.2d-e).

**Table D.1 Sequences[a] and melting temperatures (°C)[b] for BZLF1 and design constructs.**

| | basic/acid | coiled coil<br>191 221<br>bcdefgabcdefgabcdefgabcdefgabcd | proximal CT<br>231 | distal CT<br>245 | $T_m$ |
|---|---|---|---|---|---|
| **$BZLF1^{231}$** | | AKFKQLLQHYREVAAAKSSENDRLRLLLKQM | CPSLDVDSII | | 43 |
| **A-$BZLF1^{231}$** | QRAEELARENEELEKEA | EELEQELLKYREVAAAKSSENDRLRLLLKQM | CPSLDVDSII | | 33 |
| **B-$BZLF1^{231}$** | LEIKRYKNRVASRKCR | AKFKQLLQHYREVAAAKSSENDRLRLLLKQM | CPSLDVDSII | | 31 |
| **$BZLF1^{245}$** | | AKFKQLLQHYREVAAAKSSENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | 71 |
| **A-$BZLF1^{245}$** | QRAEELARENEELEKEA | EELEQELLKYREVAAAKSSENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | 43 |
| **B-$BZLF1^{245}$** | LEIKRYKNRVASRKCR | AKFKQLLQHYREVAAAKSSENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | 67 |
| **$BD^{231}_{CC}$** | | AKEEQEIQHLEEEIAALESENDRLRLLLKQM | CPSLDVDSII | | 38 |
| **$BD^{245}_{CC}$** | | AKEEQEIQHLEEEIAALESENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | 40 |
| **A-$BD^{245}_{CC}$** | QRAEELARENEELEKEA | EELEQELLKLEEEIAALESENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | 40 |
| **$BD^{231}_{IED}$** | | AKFKQLLQHYREVIAAEDSENDRLRLLLKQM | CPSLDVDSII | | N/A[c] |
| **$BD^{245}_{IED}$** | | AKFKQLLQHYREVIAAEDSENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | 26 |
| **A-$BD^{245}_{IED}$** | QRAEELARENEELEKEA | EELEQELLKYREVIAAEDSENDRLRLLLKQM | CPSLDVDSII | PRTPDVLHEDLLNF | N/A[c] |
| **$BD^{231}_{EEI}$** | | AKFKQLLQHEEEVIAAKSSENDRLRLLLKQM | CPSLDVDSII | | N/A[c] |

[a] The sequences SHHHHHHGESKEYKKGSGS, or GYHHHHHHGSY (the latter for constructs with the acidic extension, A-) should be placed at each N terminus to obtain the full sequences of the recombinant proteins listed in the table. Sites with amino acids different from those of the native sequence (either introduced in the design or as part of the acidic extension) are underlined. Different regions of the sequence (basic region/acidic extension, coiled coil, proximal CT and distal CT) are separated by space. As explained in the text, the acidic extension overlaps the 9 N-terminal residues of the coiled coil. Coiled-coil heptads are indicated using shading.
[b] Total protein concentration was 4 μM.
[c] N/A indicates either lack of cooperative folding or that the observed melting curve indicated the presence of more than one species.

**Table D.2 Melting temperatures (°C) for different BZLF1/design hetero-interactions.**

| Target | Design | $T_m$ [a] | $\Delta T_m$ [b] |
|---|---|---|---|
| $BZLF1^{231}$ | $BD_{CC}^{231}$ | 53 | 12 (43/38) |
| | $BD_{IED}^{231}$ | N/A[c] | N/A[c] |
| $B\text{-}BZLF1^{245}$ | $BD_{CC}^{245}$ | 66 | 12 (67/40) |
| | $BD_{IED}^{245}$ | N/A[c] | N/A[c] |
| | $A\text{-}BD_{CC}^{245}$ | >80 | > 26 (67/40) |
| | $A\text{-}BZLF1^{245}$ | 74 | 19 (67/43) |
| $B\text{-}BZLF1^{231}$ | $A\text{-}BZLF1^{231}$ | 58 | 26 (31/33) |
| JUN | $BD_{CC}^{245}$ | 41 | 10 (23/40) |

[a] Total protein concentration was 4 μM.
[b] $\Delta T_m$ was obtained by taking the $T_m$ for the hetero-complex and subtracting from it the average of the $T_m$ values for each individual species (listed in parentheses for easy comparison, $T_m$ for the target is shown first, followed by that of the design) when applicable.
[c] N/A indicates either lack of cooperative folding or that the observed melting curve indicated the presence of more than one species.

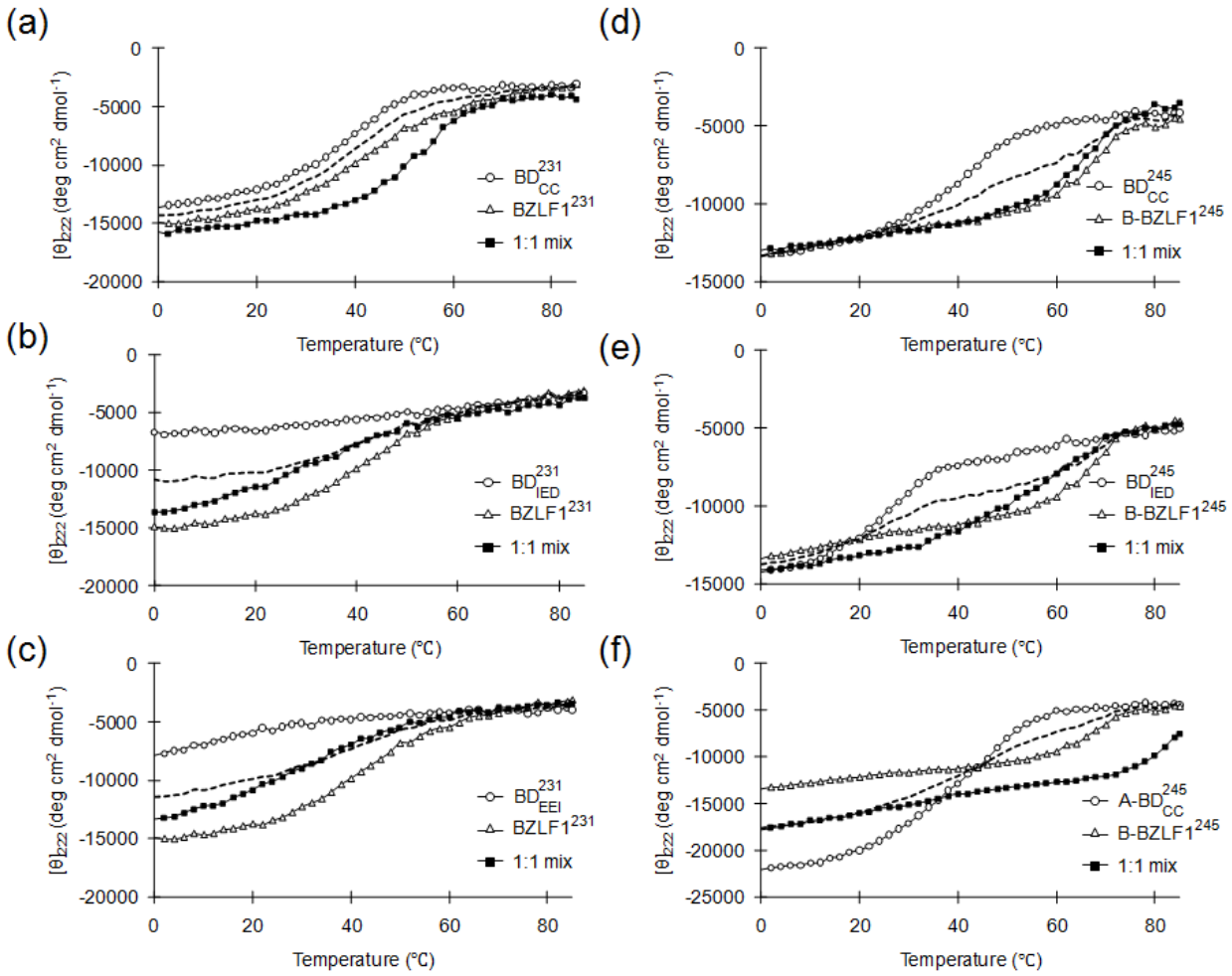



Figure D.3 Melting curves for targets, designs and complexes monitored by mean residue ellipticity at 222 nm.
Four curves are shown in each panel: the target at 4 μM (open triangles), the design at 4 μM (open circles), a mixture of the target and the design at 2 μM each (closed squares), and the numerical average of the individual melting curves for the target and the design (short dashed lines). The target is $BZLF1^{231}$ for panels (a) - (c) and $B\text{-}BZLF1^{245}$ for panels (d) - (f), as described in text, and the designs are: (a) $BD^{231}_{CC}$, (b) $BD^{231}_{IED}$, (c) $BD^{231}_{EEI}$, (d) $BD^{245}_{CC}$, (e) $BD^{245}_{IED}$, and (f) $A\text{-}BD^{245}_{CC}$.

Thermal denaturation experiments monitored by CD showed that both designed peptides, $BD^{231}_{IED}$ and $BD^{231}_{EEI}$, had relatively weak helical signals even at very low temperatures (Figure D.3b, c), illustrating the effectiveness of the negative design strategy. We compared the melting curve for the mixture of each design and $BZLF1^{231}$ with the numerical average of the individual melting

curves for each species (Figure D.3b, c). The difference between the two curves below ~22 °C reflects interaction between the designed peptides and BZLF1$^{231}$, and confirms that the designed peptides bind the target more strongly than they interact with themselves. However, an interaction is evident only at low temperatures, indicating that the stability of the design-target complex is lower than the BZLF1$^{231}$ target homodimer. Therefore, these 2 designed peptides represent a specificity profile distinct from that of $BD_{CC}^{231}$; one that achieves greater destabilization against design self-interaction at the expense of the stability of the design-target interaction.

*$BD_{cc}$ and BZLF1 form a heterodimer*

We modeled all coiled-coil interactions as parallel, symmetric dimers. Although the oligomerization states of coiled coils can be sensitive to very few amino-acid changes, (Harbury, et al. 1993, Taylor and Keating. 2005) in BZLF1 the presence of the CT region is expected to strongly favor the parallel dimer geometry observed in the crystal structure for BZLF1. The designed heterodimer also includes an Asn-Asn interaction at *a-a',* which has been shown to strongly favor dimers, and multiple charged residues at the *e* and *g* positions that are also more prevalent in dimers (Mason and Arndt. 2004). Nevertheless, we performed analytical ultracentrifugation (AUC) experiments to study the interaction between $BD_{CC}^{231}$ and BZLF1$^{231}$. Global analysis of sedimentation equilibrium runs performed at multiple concentrations and rotor speeds showed that the best-fit molecular weight for both $BD_{CC}^{231}$ and the 1:1 mixture of $BD_{CC}^{231}$ with BZLF1$^{231}$ corresponded to that expected for a dimer (representative data are shown along with the global fit in Figure D.4). For $BD_{CC}^{231}$ with BZLF1$^{231}$, the fitted molecular weight was 104% of that expected for the heterodimer, with a fitted RMS of 0.027

fringes. RMS values obtained by fixing an exact dimer or trimer weight were 0.029 or 0.090 fringes, respectively. For $BD_{CC}^{231}$, the fitted molecular weight is 102% of that expected for the homodimer, with a fitted RMS of 0.021 fringes. RMS values obtained by fixing a dimer or a trimer weight were 0.021 or 0.10 fringes, respectively. The AUC data thus confirm the validity of modeling these interactions as dimers.

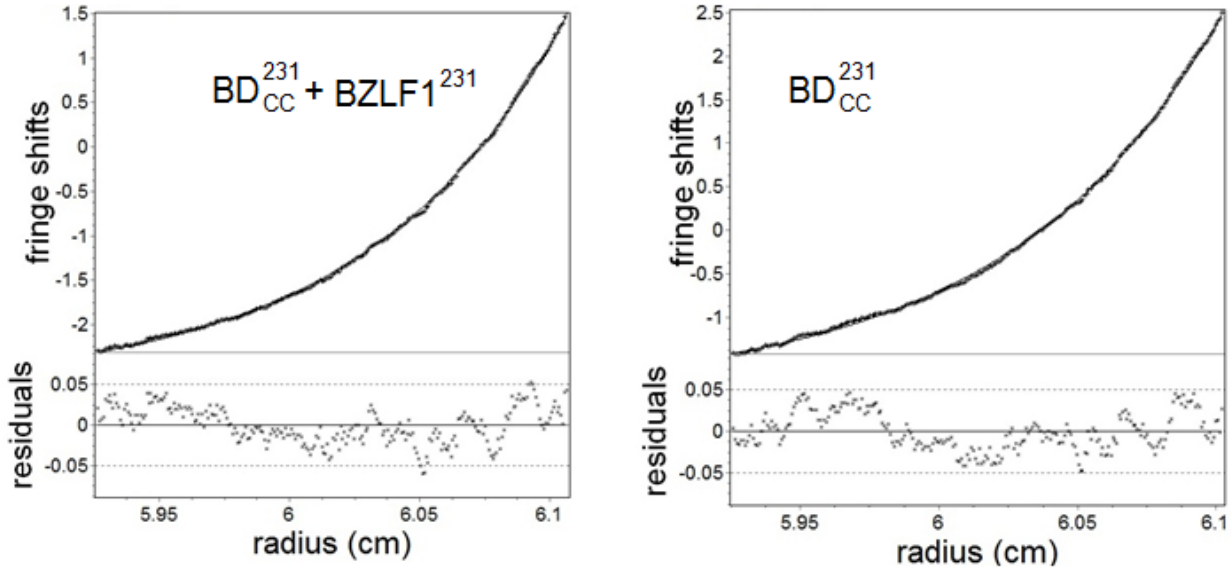



**Figure D.4 Representative analytical ultracentrifugation data for $BD_{CC}^{231}$ + $BZLF1^{231}$ (left) and $BD_{CC}^{231}$ (right).**
The fits shown were obtained with data collected at 2 concentrations and 3 different centrifuge speeds. At the bottom are the residuals to the fit.

*Testing designs in the full-length BZLF1 dimerization domain*

The designs described above targeted the BZLF1 coiled coil and were tested in the context of $BZLF1^{231}$. However, inhibitors of protein function must bind to the full-length protein. One difficulty with designing against the entire BZLF1 dimerization domain (residues 191 - 245) is that the crystal structure shows only the proximal and part of the distal CT region (up to residue 236), with the remaining part of the distal CT region contributing no electron density (Petosa, et

al. 2006). Nevertheless, the distal CT region (Figure D.1b) has been shown to contribute positively to BZLF1 dimer stability despite possibly being less structured (Schelcher, et al. 2007).

We tested whether our design procedures, which considered only the structured coiled coil, could provide molecules that bind the full-length BZLF1 dimerization domain. For this purpose, a BZLF1 construct that included both the DNA binding basic region and the full-length dimerization domain (termed B-BZLF1$^{245}$, residues 175-245, Table D.1) was used instead of BZLF1$^{231}$ as the target. The designed mutations in $BD_{CC}^{231}$ and $BD_{IED}^{231}$ were made in the context of the full-length BZLF1 dimerization domain without the basic region (residues 191-245) to create two new design constructs, $BD_{CC}^{245}$ and $BD_{IED}^{245}$ (Figure D.2a, Table D.1); the distal CT was included in the design constructs to exploit its potentially favorable interaction with the target.

The distal CT dramatically stabilized the BZLF1 homodimer (compare BZLF1$^{231}$ and BZLF1$^{245}$ $T_m$ values of 43 °C and 71 °C, respectively), consistent with prior reports (Schelcher, et al. 2007). In contrast, self-association of the $BD_{cc}$ design was not significantly stabilized by the distal CT (Table D.1). When $BD_{CC}^{245}$ and B-BZLF1$^{245}$ were mixed, there was clear evidence of interaction (Figure D.3d, Table D.2). However, the hetero-interaction between $BD_{CC}^{245}$ and B-BZLF1$^{245}$ did not appear to be stronger than the self-association of the target B-BZLF1$^{245}$ (Table D.1, D.2), which contrasts with the behavior of the shorter constructs, $BD_{CC}^{231}$ and BZLF1$^{231}$ (Figure D.3a, Table D.2). Differences in relative stabilities for the shorter and longer constructs suggest that residues in the design do not interact as favorably as the native residues with the distal CT.

In contrast to $BD_{CC}^{245}$, analysis of $BD_{IED}^{245}$ showed that both the design self-interaction and the design-target interaction were stabilized by the distal CT (compare Figure D.3b with Figure D.3e). As a result, $BD_{IED}^{245}$ was heterospecific at low temperature. Compared to $BD_{CC}^{245}$, $BD_{IED}^{245}$

showed weaker self-association but also displayed weaker affinity for B-BZLF1$^{245}$. Together, the results show that the effect of the distal CT is not negligible and depends on sequence in the coiled-coil region. The impact of the distal CT on the specificity profiles for different designs is considered further in the Discussion.

*Specificity of $BD_{cc}$ against human bZIPs*

Specificity against human bZIP proteins was not addressed explicitly in our design procedure because we reasoned that the CT region, which is unique to BZLF1, would likely stabilize interaction with BZLF1 but not with human proteins. To assess this, we selected a few human bZIPs and evaluated their interactions with $BD_{cc}$ using CD spectroscopy. To identify those human bZIP proteins most likely to associate with $BD_{cc}$, we calculated interaction scores with 36 representative human bZIP coiled coils using the scoring function employed in the CLASSY algorithm, which has been shown to be useful for evaluating bZIP coiled-coil associations (Figure D.5a) (Grigoryan and Keating. 2006). Interestingly, $BD_{cc}$ was predicted to interact more favorably with BZLF1 than with any of the human bZIPs, even though the scoring scheme used did not consider interactions involving the CT region. We chose 5 of the top 10 scoring complexes for experimental testing, selecting representative proteins that spanned 5 families and included JUN, the closest predicted competitor. We used constructs for the human proteins that included the basic region and the coiled coil (Figure D.5b-f). Analysis of melting curves for each human bZIP and each 1:1 mixture with $BD_{cc}^{245}$ showed that only JUN interacted with $BD_{cc}^{245}$. The $BD_{cc}^{245}$/JUN complex, however, was significantly weaker than that between $BD_{cc}^{245}$ and B-BZLF1$^{245}$ ($T_m$ values of 41°C vs. 66°C, Table D.2). Thus, $BD_{cc}$ is not a promiscuous design and binds preferentially to its target, BZLF1.

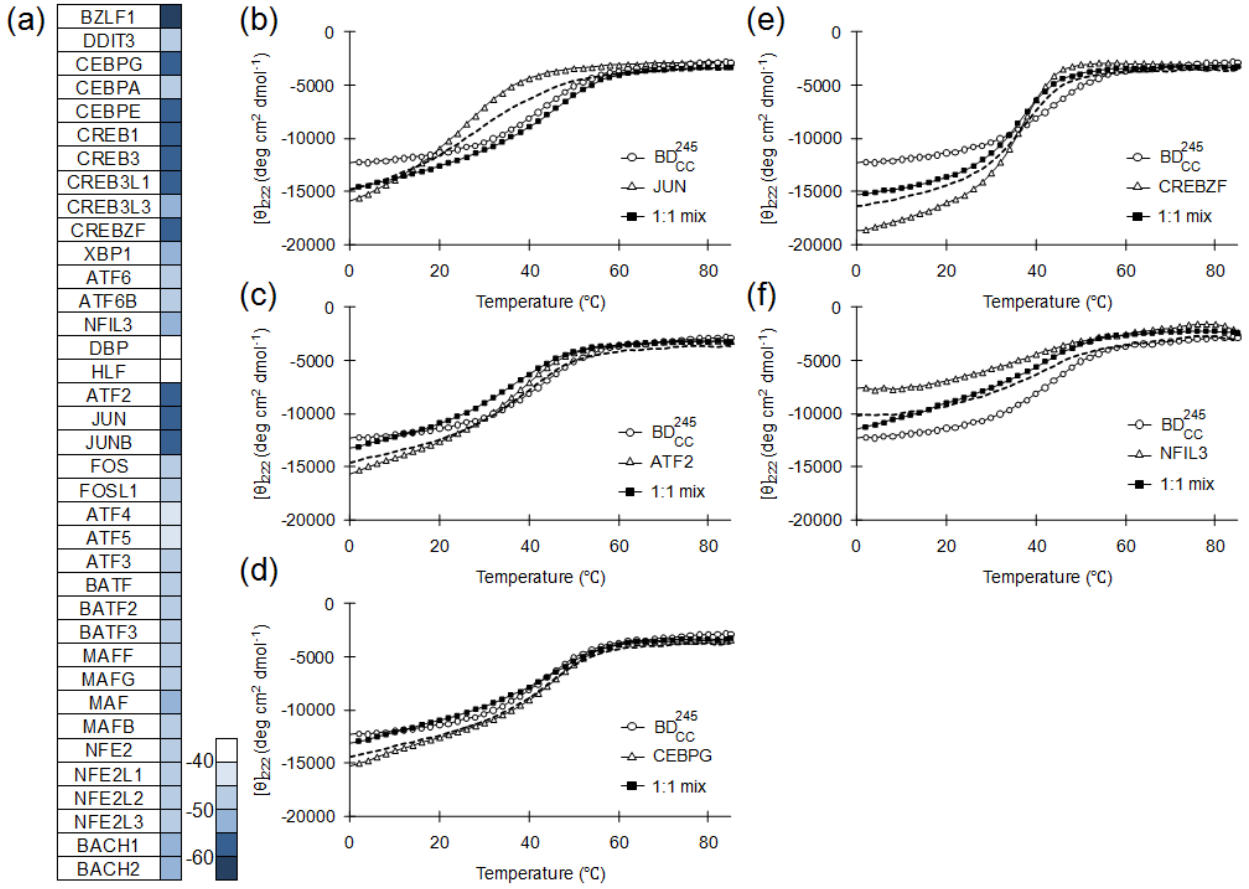



**Figure D.5 Specificity of design against human bZIPs**

(a) Predicted scores for $BD_{cc}$ interacting with BZLF1 or human bZIP peptides. (b-f) Melting curves for selected human bZIP peptides, $BD^{245}_{CC}$ or 1:1 mixtures of the two, monitored by mean residue ellipticity at 222 nm. Four curves are shown in each panel: the human bZIP at 4μM (open triangles), $BD^{245}_{CC}$ at 4 μM (open circles), a mixture of the human protein and $BD^{245}_{CC}$ at 2μM each (closed squares), and the numerical average of the individual melting curves for the human bZIP and the design (short dashed lines). The human bZIPs are: (b) JUN, (c) ATF2, (d) CEBPG, (e) CREBZF, and (f) NFIL3.

*Enhancing design performance with an N-terminal acidic extension*

Vinson and colleagues have shown that replacing the basic region of several native bZIPs with a designed sequence enriched in glutamates can provide potent dominant-negative inhibitors of bZIP dimerization and DNA binding (Acharya, et al. 2006b, Ahn, et al. 1998, Olive, et al. 1997). They also showed that such an acidic extension improved the affinity of a peptide rationally designed to heterodimerize with human bZIP CEBPA (Krylov, et al. 1995). Because the basic region of BZLF1 is highly similar to that of human bZIPs (Figure D.1a), we reasoned that incorporating an acidic extension into the N-terminus of our $BD_{CC}^{245}$ design might enhance its affinity for BZFL1.

Three acidic extension variants developed by Vinson et al. differ in 2 positions that could interact with the BZLF1 basic region, if the interaction occurred with a coiled-coil-like geometry as has been hypothesized for other systems (Acharya, et al. 2006b). We chose to use the "A"-extension, which introduced the possibility of an attractive Glu-Arg *g-e'* interaction and a Leu-Leu core-core *a-a'* interaction. Following prior work in the Vinson laboratory, (Olive, et al. 1997) we constructed A-$BD_{CC}^{245}$ (sequence in Table D.1). The modification added 17 residues at the N-terminus and replaced 6 out of 9 of the most N-terminal residues of the designed region (Table D.1). Interestingly, A-$BD_{CC}^{245}$ showed much greater helicity than $BD_{CC}^{245}$ and $BD_{CC}^{231}$, indicating that either some of the N-terminal 26 residues and/or the distal C-terminal region are likely helical in this context (Figure D.3f). The $T_m$ for A-$BD_{CC}^{245}$ was similar to those for $BD_{CC}^{231}$ and $BD_{CC}^{245}$ (Table D.1), whereas interaction with B-$BZLF1^{245}$ was significantly stabilized compared to the $BD_{CC}^{245}$/B-$BZLF1^{245}$ interaction as expected (Figure D.3f). The heterocomplex melted at > 80 °C (Table D.2). Together these observations indicate that changes made in A-$BD_{CC}^{245}$ did not stabilize the design homodimer, but further enhanced its interaction with B-$BZLF1^{245}$, as desired for inhibitor

design.

For comparison, we constructed several other peptides with acidic extensions and assessed their self-association (Table D.1). This modification dramatically destabilized BZLF1$^{245}$ by 28 °C (71 °C for BZLF1$^{245}$ vs. 43°C for A-BZLF1$^{245}$). A-BZLF1$^{231}$ was also destabilized, but by only 10 °C (43°C for BZLF1$^{231}$ vs. 33 °C for A-BZLF1$^{231}$). BD$^{245}_{IED}$ was destabilized by an amount that could not be quantified because A-BD$^{245}_{IED}$ did not exhibit a cooperative melt. A-BZLF1$^{245}$ was tested for interaction with B-BZLF1$^{245}$ and formed a heterocomplex with $T_m$ of 74 °C (compared to the $T_m$ for B-BZLF1$^{245}$ self-interaction, 67 °C, Tables D.1, D.2). The $T_m$ for the heterocomplex between A-BZLF1$^{231}$ and B-BZLF1$^{231}$ was 58 °C (compared to the $T_m$ for B-BZLF1$^{231}$ self-interaction, 31 °C). These results are consistent with an interaction between the acidic extension and the basic region stabilizing the heterocomplexes, and also with an unfavorable interaction between the distal CT and the acidic extension, which is considered further in the Discussion.

*Inhibiting DNA binding by BZLF1*

We used an electrophoretic mobility shift assay (EMSA) to assess inhibition of B-BZLF1$^{245}$ binding to DNA by different designed peptides (Figure D.6). The dimerization domain of BZLF1 lacking the basic region, BZLF1$^{245}$, was included for comparison purposes. All peptides tested showed concentration-dependent inhibition. BD$^{245}_{CC}$, A-BZLF1$^{245}$ and A-BD$^{245}_{CC}$ were more effective than BZLF1$^{245}$. Design BD$^{245}_{IED}$ was also an effective inhibitor. The most potent inhibitor was A-BD$^{245}_{CC}$, which completely inhibited B-BZLF1$^{245}$ binding to DNA at equi-molar concentration.

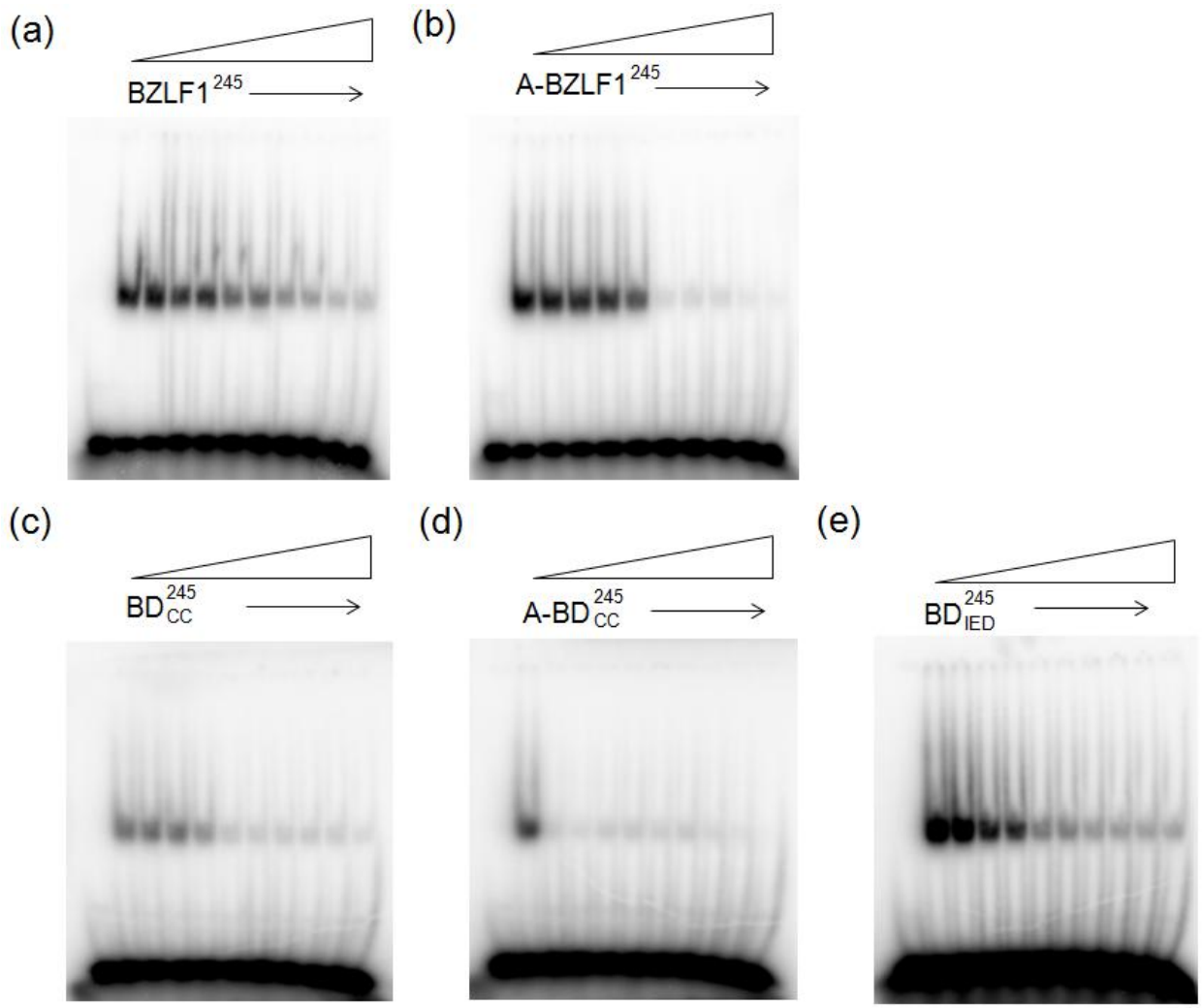



**Figure D.6 Peptide inhibition of B-BZLF1$^{245}$ binding to DNA.**
Representative gel-shift images were shown for: (a) BZLF1$^{245}$, (b) A-BZLF1$^{245}$, (c) $BD_{CC}^{245}$, (d) A-$BD_{CC}^{245}$, (e) $BD_{IED}^{245}$. The first two lanes for each gel include DNA only (first lane) and B-BZLF1$^{245}$with DNA (second lane). Inhibitor peptides were added in increasing concentrations from 10 nM to above 2 μM (left to right, 2-fold dilutions). Conditions are described in Materials and Methods in more detail, and were slightly different for panel (a)-(d) vs. panel (e).

## DISCUSSION

In this study, we employed different design strategies to create inhibitor peptides targeting the viral bZIP protein BZLF1. We sought peptides that achieved hetero-specificity through enhanced affinity for the target and/or reduced self-interaction. Below we discuss our different design approaches and the experimental behaviors of our designed peptides.

*Applying CLASSY to BZLF1*

As demonstrated earlier, (Grigoryan, et al. 2009) CLASSY is an algorithm that can be applied to design bZIP-like coiled coils. It was developed in conjunction with a specialized scoring function that includes computed structure-based terms, helix propensities, and experimentally determined coupling energies. The scoring function was validated on a large-scale dataset of human bZIP coiled-coil interactions (Grigoryan and Keating. 2006) and supported the successful design of numerous bZIP-binding peptides. It is not known to what extent the bZIP scoring function can be applied in design problems involving coiled-coil targets with features not observed in typical human bZIPs. Here, we explored whether the BZLF1 dimerization domain could be treated as a standard bZIP target for CLASSY design.

To treat BZLF1 as a coiled coil, we designed against the N-terminal part of the sequence and did experimental tests using constructs that did not include the distal CT (the "231" constructs, Fig 1b, 2a), much of which is not observed in the X-ray structure. The BZLF1 coiled-coil region is rather short (4 heptads), has only one Leu at position *d* among these heptads, and includes a

region with very narrow inter-helical distance (~4 Å Cα-Cα distance at *a*-position residue 204). These variations might be expected to compromise performance of the scoring function, as coiled-coil context is known to influence the contributions of residues and residue pairs to stability (Steinkruger, et al. 2010, Moitra, et al. 1997, Lu and Hodges. 2004). Thus, methods validated using human bZIPs might not generalize broadly to all coiled-coil dimers. However, we found that design $BD_{CC}^{231}$ incorporated elements very commonly employed in published anti-human bZIP designs (see below), and that these gave good experimental performance in this less canonical example. Success might be attributed to the fact that introducing more canonical residues at interfacial sites on one helix (the design) makes the design-target heterodimer more similar to the human bZIPs, e.g. the heterodimer likely has a more typical helix-helix separation.

*Features contributing to the stability and specificity of the designs*

Analysis of the designed sequences suggests that stability and specificity were achieved using different combinations of core, edge and core-edge interactions. For example, in the $BD_{CC}^{231}$ design, the *a* and *d* heptad positions were populated with hydrophobic Ile and Leu, respectively, (e.g. Y200L, A204I, K207L), which are expected to be exceptionally stabilizing in the design homodimer (Acharya, et al. 2002). Therefore, a strategy that used only these mutations to stabilize the design-target interaction would likely stabilize the design self-interaction even more, and fail to achieve heterospecificity. Negative design elements that likely compensate for over-stabilization of the design self-interaction come from interfacial *e* and *g* positions occupied by negatively charged amino acids. These negative charges make favorable interactions with positively charged residues in the target (e.g. 201R, 207K), consistent with improving the

stability of the design-target interaction. However, they also introduce repulsive *g-e'* or *e-g*' interactions in the design homodimer (e.g. 196E-201E (*g-e'*), 203E-208E (*g-e*'), 201E-203E (*e-g*')). Similar examples of using a highly hydrophobic core to achieve stability while modulating specificity using interfacial charge have been observed in many prior coiled-coil designs (Woolfson. 2005). One less familiar feature in the $BD_{CC}^{231}$ design is the presence of an N-terminal glutamate at a *d* position. Two glutamate residues at *d* and *d'* in a homodimer are destabilizing in coiled coils, (Tripet, et al. 2000) but this residue potentially interacts favorably with an *e'* lysine in BZLF1, via a core-to-edge type interaction that has previously been noted in CLASSY-derived designs and other studies (Steinkruger, et al. 2010, Grigoryan, et al. 2009, Havranek and Harbury. 2003, Reinke, et al. 2010a, Barth, et al. 2008).

Designs $BD_{IED}^{231}$ and $BD_{EEI}^{231}$ relied much more on core-to-edge interactions, which were placed close to the middle of the coiled coil in these designs. In contrast to *g-e'* interactions, no coupling energies have been measured for negatively charged residues at *d-d'* or *d-e'* sites. CLASSY performed poorly in predicting the relative stabilities of complexes involving $BD_{IED}^{231}$ and $BD_{EEI}^{231}$, most likely because experimental data describing such charged core-core and core-edge interactions were not available to guide the development of the scoring function (Grigoryan and Keating. 2006, Grigoryan, et al. 2009). Nevertheless, a cluster of 4 negatively charged residues in the design homodimer proved very effective as a negative design element; $BD_{IED}^{231}$ and $BD_{EEI}^{231}$ did not appreciably self-associate. Affinity for the target was also compromised, however. Substitution of alanine with isoleucine at *a* position 204 was introduced to compensate for some of the lost stability of the heterodimer, showing how a different combination of stabilizing and destabilizing elements can generate a hetero-specific design that inhibits DNA binding (Figure D.6).

Substitution of isoleucine for alanine at *a* position 204 is found in all 3 designs. In the native structure, alanine at this position fits well in the tight space between unusually close helices (~4 Å Cα-Cα distance between residue 204 on the two chains). Isoleucine cannot be built into this site in the crystal structure without severe clashes. Nonetheless, the larger Ile was accommodated in all three designs, and an alanine to isoleucine mutation is stabilizing in the context of BZLF1$^{245}$ (an increase of $T_m$ by 9 °C under the conditions of Table D.1, data not shown). These data suggest a change in the backbone structure upon making this substitution. Local rearrangement of the design-BZLF1 complex to a more typical backbone structure probably helps explain why the CLASSY bZIP scoring function worked well. To achieve good predictive ability for a wider range of backbone structures, backbone flexibility could be treated explicitly (Barth, et al. 2008, Mandell and Kortemme. 2009).

*The influence of the distal CT region*

Previous studies revealed that the distal CT, although unresolved in the BZLF1 crystal structure, might interact with the N-terminal part of the BZLF1 coiled-coil region, thereby stabilizing the dimer (Schelcher, et al. 2007). We confirmed a stabilizing role for this region (Table D.1, comparing BZLF1$^{231}$ and BZLF1$^{245}$). Interestingly, this effect depends on the sequence in the coiled-coil region (Table D.1, D.2). The distal CT does not stabilize the $BD^{245}_{CC}$ design self-interaction, and it enhances the stability of the $BD^{245}_{CC}$-target interaction only modestly. On the other hand, the distal CT significantly increased the stability of the $BD^{245}_{IED}$ design self-interaction, as well as the stability of the $BD^{245}_{IED}$-target interaction. There are more sequence changes in the $BD_{cc}$ design, and the number of negative charges introduced is larger than in the $BD_{IED}$ design. As discussed below, the influence of the distal CT is also sensitive to the acidic

extension included in some designs. Although the structure of the interaction between the distal CT and the N-terminal part of the coiled coil in the native protein is not known, repulsive electrostatics, or unfavorable desolvation of charges in the coiled-coil region are plausible mechanisms for disfavoring this interaction in the $BD_{cc}$ design.

*Specificity against human bZIPs*

We did not consider specificity against human bZIPs in our design procedure. However, we showed that the design $BD_{cc}$ is not promiscuous in binding human bZIP proteins. Computational analysis predicted that the coiled-coil region of $BD_{cc}$ would interact with the BZLF1 coiled coil moderately more favorably than with any other human bZIP coiled coil (but with a few close competitors). This is interesting, given the fairly canonical coiled-coil sequence features of $BD_{cc}$. The requirement to satisfy hydrogen bonding for Asn 204 at the *a* position in $BD_{cc}$, and the charge complementarity between the *e* and *g* positions of $BD_{cc}$ and BZLF1 helices but not most human proteins, contributed to the predicted binding preference.

Thermal stability studies confirmed that $BD_{cc}^{245}$ does not bind strongly to selected human bZIPs identified in the computational analysis. In addition to selectivity derived from the coiled-coil region (which was predicted to be modest), the CT region likely confers additional specificity. Interactions with $BD_{cc}^{245}$ and B-BZLF1$^{245}$ could benefit from native-like contacts between the CT region and the coiled coil domain, which are not conserved in complexes with human proteins. Thus, the interaction specificity of $BD_{cc}^{245}$ is likely encoded in both its coiled-coil domain and the CT region.

*Improving inhibitor potency using an N-terminal acidic extension*

The Vinson group has demonstrated that dominant-negative inhibitors of bZIP dimerization and DNA binding can be created by replacing the basic region of native or modified native bZIPs with an acidic sequence (Acharya, et al. 2006b). In this study, we used this strategy to improve the potency of our designed peptides. The resulting A-$BD_{cc}^{245}$ peptide maintained specificity, showing little change in the $T_m$ for the design self-association. The small change in homodimer stability probably results from destabilization by the negative charges in the extension, countered by a stabilizing leucine residue introduced at *d* position 193 (this residue is Glu in $BD_{cc}$) (Olive, et al. 1997). A-$BD_{cc}^{245}$ formed a more stable complex with the target B-$BZLF1^{245}$ than did $BD_{cc}^{245}$ (an increase of $T_m$ > 14 °C at 4 μM, Table D.2). This indicates that the acidic extension, which targets the basic region of bZIPs, can be used in conjunction with computational design methods targeting the coiled coil. Given that the Vinson laboratory has demonstrated that the coiled-coil region of A-ZIPs governs interaction specificity, while the acidic extension provides much enhanced affinity, this is an appealing strategy for expanding the design of tight-binding and selective bZIP inhibitors (Acharya, et al. 2006b, Ahn, et al. 1998, Olive, et al. 1997, Krylov, et al. 1995, Grigoryan, et al. 2009).

Interestingly, modifying BZLF1 with an acidic extension did not stabilize interaction of A-$BZLF1^{245}$ with B-$BZLF1^{245}$ as much as expected ($T_m$ of 74 °C compared to 67 °C for the B-$BZLF1^{245}$ homodimer, Table D.1, D.2). In contrast, interaction of the shorter construct A-$BZLF1^{231}$ with B-$BZLF1^{231}$ was stabilized to a much greater extent ($T_m$ of 58 °C compared to 31 °C for the B-$BZLF1^{231}$ homodimer). Furthermore, the destabilizing effect of the acidic extension on design homodimer stability is quite different in $BZLF1^{245}$ vs. $BZLF1^{231}$ (decreasing $T_m$ values by 28 °C vs. 10 °C, Table D.1). These observations are consistent with a model where the distal

CT interacts unfavorably with the acid extension, much as it appears to interact unfavorably with negative charges in the N-terminal part of the $BD_{cc}$ design. Although not addressed in the present study, the performance of A-BZLF1$^{245}$ as an inhibitor could potentially be improved by redesigning the acidic extension so that interference from the distal CT is minimized, although this is difficult in the absence of structural information about this part of the protein.

*Analysis of inhibitor potency*

To test the designed peptides as inhibitors of BZLF1 DNA binding, we used an *in vitro* EMSA assay to monitor the population of B-BZLF1$^{245}$ bound to DNA in the presence of different peptides (Figure D.6). It is unsurprising that A-$BD^{245}_{CC}$, which formed the most thermo-stable complex with B-BZLF1$^{245}$ and exhibited the largest difference in homodimer vs. heterodimer stability, was the most potent inhibitor. The improved performance of $BD^{245}_{CC}$ and A-BZLF1$^{245}$ relative to the native peptide, BZLF1$^{245}$, could be rationalized by their improved affinity and/or anti-homodimer specificity (see below). $BD^{245}_{IED}$ inhibited DNA binding effectively and we estimate its potency is similar to that of BZLF1$^{245}$, although these two peptides could not be compared using identical assay conditions (see Materials and Methods). The effectiveness of $BD^{245}_{IED}$ resulted from a combination of reduced affinity but improved anti-homodimer specificity.

To explore more generally how affinity and specificity each influence potency, we constructed a simple computational model with the following assumptions: 1) the target bZIP, the DNA, and the designed peptide were the only components present, 2) the target bZIP homodimer was the only species that could bind DNA (i.e. complete cooperative binding), 3) non-specific DNA binding was neglected. Some of the assumptions made may not apply to all of our

experiments. We computed concentration dependent inhibition of DNA binding for a series of designs covering a spectrum of affinities and specificities. Affinity was described by the ratio between the dissociation constant of the target bZIP homodimer and that of the design-target heterodimer ($K_d^{T_2}/ K_d^{DT}$, D: design, T: target, see Materials and Methods), and specificity was described by the ratio between the dissociation constant for the design homodimer and that of the design-target heterodimer ($K_d^{D_2} / K_d^{DT}$). The efficacy of different inhibitors is illustrated in a heat map in Figure D.7 that indicates the improvement in $IC_{50}$ over a reference for which $K_d^{D_2} = K_d^{DT} = K_d^{T_2}$. The reference inhibitor with affinity and specificity of 1 was included to reflect the behavior of the dimerization domain of the target bZIP. We explored two scenarios that led to different inhibition landscapes: one where modeled dissociation constants for the target bZIP complex and bZIP-DNA interactions were lower than the target bZIP concentration (Figure D.7a), and another where they were higher (Figure D.7b)

The results in Figure D.7 support intuition about the importance of both affinity and specificity. Lines of constant color running across the plots in Figure D.7 show that equivalent potency can be achieved using different combinations of affinity and specificity. Clearly, neither affinity nor preference for hetero vs. homodimerization correlates directly with design performance. For the purposes of discussion, we label 3 regions on the plots: $H_{affinity}$:$L_{spec}$ indicates inhibitors with high affinity for the target but limited anti-homodimer specificity, $L_{affinity}$:$H_{spec}$ indicates inhibitors with affinity for the target that is comparable to or weaker than the reference inhibitor, but with weaker self-association, and $H_{affinity}$:$H_{spec}$ inhibitors have both tighter target-binding affinity and weaker self-association than the reference. Among our designs, and to the extent that approximate stabilities assessed by thermal denaturation under CD conditions can be extrapolated to the gel-shift assay, $BD^{245}_{CC}$ and $BD^{245}_{IED}$ are both $L_{affinity}$:$H_{spec}$

inhibitors that use anti-homodimer specificity to improve inhibitor potency. A-$BD_{CC}^{245}$ maintains anti-homodimer specificity but gains additional affinity via the acidic extension, making it a $H_{affinity}$:$H_{spec}$ inhibitor.

The model in Figure D.7 is useful for broadly guiding the computational design of specific inhibitors, so we conclude with a few general observations. First, heterospecificity is important, but not sufficient, for good performance. A design is hetero-specific if the ratio $K_d^{T_2} \bullet K_d^{D_2}/(K_d^{DT})^2$ is larger than 1. In the figure, this region is below the dashed line and all inhibitors with potency better than the reference lie in this region. Maintaining hetero-specificity for high affinity designs imposes a bound on design homodimer stability. This is relevant for parallel dimeric coiled-coil targets, because amino-acid changes that enhance interaction with the target often stabilize the design self-interaction even more (Acharya, et al. 2002). Second, the relative importance of improving affinity vs. specificity depends on the target and assay conditions. For panel a, improved hetero-specificity implies enhanced design performance regardless of whether affinity or specificity is the main contributor. On the other hand, if the target bZIP concentration is lower, as in panel b, improving specificity alone is no longer sufficient, and affinity must be optimized; very potent designs in panel b can only be achieved by optimizing along the path toward $H_{affinity}H_{spec}$. Finally, the overall diagonal trends for constant-$IC_{50}$ regions in both panels emphasize that improving either affinity or specificity can potentially lead to success, depending on the specific conditions and requirements for an application. Designs belonging to the $H_{affinity}H_{spec}$ class are the most effective. However, such designs might not exist, or could be hard to identify for a particular problem. In such cases, one could consider optimizing primarily affinity or specificity, depending on which is easier to achieve. Although not used extensively for this purpose here, the CLASSY algorithm is well suited for identifying designs with different

affinity vs. specificity trade-offs (Grigoryan, et al. 2009).

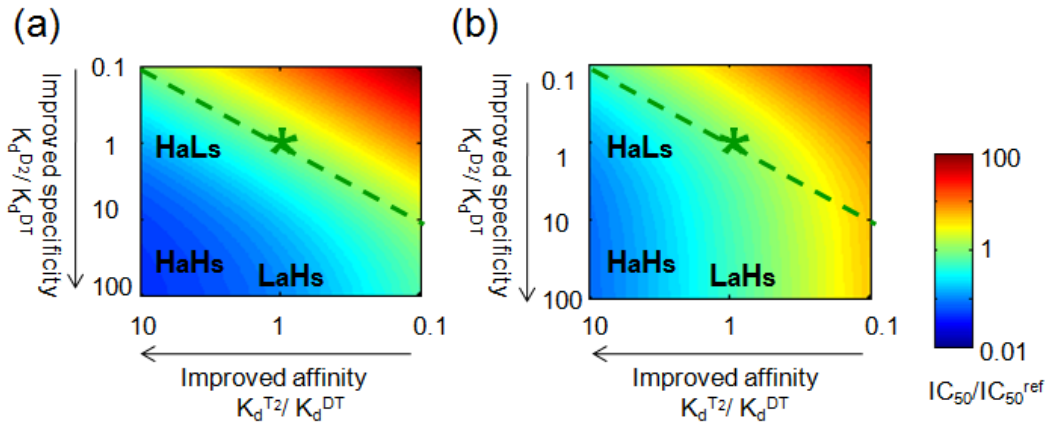



**Figure D.7 Inhibition of DNA binding as a function of the affinity and anti-homodimer specificity of the inhibitor.**
A description of the model is given in Methods. The ratio of the $IC_{50}$ for a design to the $IC_{50}$ for a reference inhibitor with affinity equal to the wild-type protein is used as an indicator of design potency (scale at right). This ratio is plotted as a function of the affinity and specificity of the inhibitor. In (a), the $K_d$ values for target dimerization and DNA binding are 10-fold lower than the bZIP concentration. In (b) the $K_d$ values for both associations are 10-fold higher than the bZIP concentration. Labeling on the graph ($H_aL_s$: $H_{affinity}L_{spec}$, $L_aH_s$: $L_{affinity}H_{spec}$ and $H_aH_s$: $H_{affinity}L_{spec}$) is described in Discussion. The dashed line represents designs with zero hetero-specificity. The reference inhibitor is indicated with a star.

## CONCLUSION: IMPLICATIONS FOR PROTEIN DESIGN

This study addresses three topics relevant to the design of peptides that inhibit native protein-protein interactions. First is the issue of specificity, which arises in many protein design problems and is acute for coiled-coil targets where self-association of the design can compete with target inhibition. Using BZLF1 as a target, we characterized peptides that balance affinity and specificity in different ways. This adds to the small number of examples where affinity and specificity have both been treated as design considerations (Grigoryan, et al. 2009, Havranek and Harbury. 2003, Barth, et al. 2008, Kortemme, et al. 2004, Ali, et al. 2005, Bolon, et al. 2005, Sammond, et al. 2010, Karanicolas and Kuhlman. 2009). Second, we explored a design problem

where features of the target that are not well described in an existing structure (the BZLF1 distal CT) nevertheless influence complex stability. We showed that different designs responded differently to the introduction of the distal CT. This argues for developing methods that broadly survey design solution space and discovering a large set of potentially good designs, rather than identifying only "the best" design according to some imperfect criteria. This can be accomplished in various ways, e.g. by exploring a range of tradeoffs between stability and specificity, or exploring a variety of related structural templates as design scaffolds (Grigoryan, et al. 2009, Fu, et al. 2007). Testing diverse solutions maximizes the chance of finding a design that interacts well with poorly characterized features of the target. Finally, our best design exploited a modular strategy where optimization of the coiled-coil dimerization interface was coupled with a more generic strategy developed previously for stabilizing inhibitor-bZIP complexes. Modularity is likely to be a key strategy for the design of ever more complex molecular parts.

## MATERIALS AND METHODS

*Cloning, protein expression and purification*

Synthetic genes encoding native or redesigned BZLF1 sequence, residues 175 or 191 to 245 (B-BZLF1$^{245}$, BZLF1$^{245}$, BD$^{245}_{CC}$, BD$^{245}_{IED}$), were constructed by gene synthesis. Primers were designed using DNAWorks, (Hoover and Lubkowski. 2002) and a two-step PCR procedure was used for annealing and amplification. Genes encoding the native or redesigned sequence in the context of residues 191 to 231 were made in a single-step PCR reaction using the longer constructs as templates. The genes were cloned via BamHI/XhoI restriction sites into a modified

version of a pDEST17 vector that encodes an N-terminal 6xHis tag and a GESKEYKKGSGS linker that improves the solubility of the recombinant protein (Reinke, et al. 2010b). To facilitate cloning of genes encoding the acidic extension, a pET16b vector (Novagen) was modified to encode an N-terminal 6xHis tag, followed by a GSY linker and the acidic extension sequence. Genes encoding BZLF1$^{231}$, BZLF1$^{245}$ and the designs BD$^{245}_{CC}$ and BD$^{245}_{IED}$ were subsequently cloned into the modified vector using AflII/XhoI restriction sites to make A-BZLF1$^{231}$, A-BZLF1$^{245}$, A-BD$^{245}_{CC}$ and A-BD$^{245}_{IED}$. Recombinant proteins were expressed in *E. coli* RP3098 cells. Cultures were grown at 37 °C to an OD of ~0.4-0.9, and expression was induced by addition of 1 mM IPTG.

Purification was performed under denaturing conditions (6M GdnHCl) using an Ni-NTA affinity column followed by reverse-phase HPLC. Human bZIP constructs containing the basic region and the coiled-coil domain were described previously (Reinke, et al. 2010b).

*Computational protein design using CLASSY*

The sequence BD$_{cc}$ was designed using the CLASSY algorithm as previously reported (Grigoryan, et al. 2009). In brief, the algorithm solves for the sequence predicted to interact most favorably with a target sequence (here, chosen to be the N-terminal part of the BZLF1 leucine zipper, residues 191 to 209) using integer linear programming. It is possible to impose constraints on the gap between the energy of interaction with the target and the energy of undesired states such as the design homodimer. No such constraint was applied in the design of BD$_{cc}$, which was predicted to favor the design-target interaction over design homodimerization without it. The scoring function used was HP/S/Cv. This function was derived by combining molecular mechanics calculations and experimentally determined coupling energies for many

core *a-a'* interactions .The Leu-Leu core *d-d*' interaction was modeled with an empirical value of –2 kcal/mol$^{-1}$. The HP/S/Cv structure-based energy function was transformed into a sequence-based expression using cluster expansion, and modified using empirical data, as described by Grigoryan et al (Grigoryan and Keating. 2006, Acharya, et al. 2006a, Grigoryan, et al. 2009).

*Predicting interactions between $BD_{cc}$ and human bZIPs*

BZLF1 was aligned with 36 human bZIPs using the conserved basic region, and interaction scores for residues 191-221 of $BD_{cc}$ with the correspondingly aligned 31 residues of each human bZIP were computed using the HP/S/Cv model as described above.

*Circular dichroism spectroscopy*

Circular dichroism experiments were performed and analyzed, and $T_m$ values fitted as described previously (Grigoryan, et al. 2009). Thermal melts from 0 ℃ to 85 ℃ were mostly reversible, regaining ≥95% of signal or giving closely similar $T_m$ values for the reverse melt (except for samples containing NFIL3, which precipitated upon heating to 85 ℃). Melting temperatures were estimated by fitting the data to a two-state equilibrium (unfolded/folded), assuming no heat capacity changes upon folding. A detailed description of the equation was described previously (Grigoryan, et al. 2009). In cases where high-temperature unfolding precluded accurate fitting of unfolded baselines, the $T_m$ was either defined as the mid-point of the unfolding transition after manually picking the baseline (for the 1:1 mixture of B-BZLF1$^{245}$ and A-BZLF1$^{245}$), or a lower bound on the $T_m$ value was estimated (for the 1:1 mixture of B-BZLF1$^{245}$

and A-$BD^{245}_{CC}$). The protein concentrations are given in the figure legends. All measurements were performed in PBS buffer containing 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 0.25 mM EDTA and 1 mM DTT. Samples were heated to 65 ℃ for 5 minutes before measurement to equilibrate peptide mixtures, and then cooled to and equilibrated at the starting temperature.

*Analytical ultracentrifugation*

Protein samples were dialyzed against the reference buffer (12.5 mM sodium phosphate, 150 mM NaCl, 1mM DTT, 0.25 mM EDTA, pH 7.4) three times (including once overnight) before measurements. Sedimentation equilibrium runs were performed with a Beckman XL-I analytical ultracentrifuge using interference optics. Two concentrations for each protein sample were prepared (50 and 100 μM), and runs at 3 different speeds (28,000, 35,000 and 48,000 rpm) were carried out at 20 ℃. Each run was ~ 20 h, and equilibrium was confirmed by negligible differences between the sample distribution in the cell over sequential scans. Data were analyzed globally with the program HeteroAnalysis (Cole and Lary. 2006) , using a calculated (Laue, et al. 1992) partial specific volume of 0.7275 ml/g (for the $BD^{231}_{CC}$/$BZLF1^{231}$ mixture) or 0.7245 ml/g (for $BD^{231}_{CC}$) and a solution density of 1.005 g/ml.

*Electrophoretic mobility shift assay (EMSA)*

Gel shift assays were performed as described previously (Reinke, et al. 2010b). Briefly, 10 nM B-$BZLF1^{245}$ was prepared either alone or mixed with each inhibitor at 9 concentrations ranging from 10 nM to 2560 nM in 2-fold dilutions. Gel-shift buffer ((150 mM KCl, 25 mM TRIS pH 8.0, 0.5 mM EDTA, 2.5 mM DTT, 1 mg/ml BSA, 10% (v/v) glycerol, 0.1 μg/ml

competitor DNA (Poly (I)·Poly (C) (Sigma))) was then added and incubated for 10 minutes at 42 °C. Closely similar results were obtained when incubating samples for 20 minutes at 42 °C. The competitor $BD^{245}_{IED}$ was not stable upon heating and was incubated for 2 hours at 18-22 °C. Radiolabeled annealed AP-1 site ,CGCTTGATGACTCAGCCGGAA (IDT), at a final concentration of 0.7 nM was added and incubated for 15 minutes at 18-22 °C. Complexes were separated on NOVEX DNA retardation gels (Invitrogen). Dried gels were imaged using a phosphorimaging screen and a Typhoon 9400 imager. ImageQuant software (Amersham Biosciences) was used to quantify band intensities.

*Simulating the impact of affinity and specificity on designed peptide behaviors*

The simulation treated the following species: The target bZIP monomer (T), the target bZIP homodimer ($T_2$), the design monomer (D), the design homodimer ($D_2$), the design-target bZIP heterodimer (DT), free DNA (DNA) and the complex formed between the target bZIP homodimer and DNA ($T_2$DNA). Species are linked by the following reactions:

$$2T \leftrightarrows T_2 \quad \frac{[T]^2}{[T_2]} = K_d^{T_2}$$

$$2D \leftrightarrows D_2 \quad \frac{[D]^2}{[D_2]} = K_d^{D_2}$$

$$D + T \leftrightarrows DT \quad \frac{[D][T]}{[DT]} = K_d^{DT}$$

$$T_2 + DNA \leftrightarrows T_2DNA \quad \frac{[T_2][DNA]}{[T_2DNA]} = K_d^{T_2DNA}$$

$$[T] + [DT] + 2[T_2] + 2[T_2DNA] = [T]_{total}$$

$$[D] + [DT] + 2[D_2] = [D]_{total}$$

$$[DNA] + [T_2DNA] = [DNA]_{total}$$

Affinity is defined as $K_d^{T_2} / K_d^{DT}$, and a value > 1 indicates the design-target bZIP heterodimer is more stable than the target bZIP homodimer (improved affinity). Specificity is defined as $K_d^{D_2} / K_d^{DT}$, and a value > 1 indicates the design-target bZIP heterodimer is more stable than design homodimer (improved specificity). A design with affinity and specificity equal to 1 was used as a reference. The $IC_{50}$ value was defined as the design concentration $[D]_{total}$ at which 50% less DNA is bound relative to zero design concentration. The total target bZIP concentration $[T]_{total}$ was fixed at 10 nM, and the total DNA concentration $[DNA]_{total}$ at 0.7 nM. Different combinations of $K_d^{T_2}$ and $K_d^{T_2DNA}$ values were explored ($10^{-9}$, $10^{-8}$, and $10^{-7}$ M for each), including when both are lower than $[T]_{total}$ ($10^{-9}$ M/$10^{-9}$ M, Figure D.7a) and when both are higher than $[T]_{total}$ ($10^{-7}$ M/$10^{-7}$ M, Figure D.7b). For each combination of fixed $K_d^{T_2}$ and $K_d^{T_2DNA}$, the $IC_{50}$ values for a range of designs with different affinities (0.1 to 10) and specificities (0.1 to 100) were calculated. The ratio $IC_{50}^{design}/IC_{50}^{ref}$, with a value < 1 implying greater potency than the reference, was plotted as a heat map. The dashed lines on the plots in Figure D.7 indicate points where the product of affinity and specificity ($(K_d^{T_2} * K_d^{D_2})/(K_d^{DT} * K_d^{DT})$) equals 1. All designs below the dashed line are hetero-specific. The simulation was carried out and heat maps were generated using Matlab (MathWorks).

## ACKNOWLEDGEMENTS

We thank K. E. Thompson for designing the acidic extension vector, making the A-BZLF1[245] construct, and providing valuable suggestions. We thank G. Grigoryan for assistance with the CLASSY algorithm, and members of the Keating lab, especially O. Ashenberg, C. Negron, S. Dutta, L. Reich, V. Potapov, K. Hauschild and J. DeBartolo for helpful discussion of the

manuscript. We thank D. Pheasant at the Biophysical Instrumentation Facility at MIT for assistance in analytical ultracentrifugation experiments. A.W.R. was supported by a Koch graduate fellowship. This work was funded by NIH award GM067681 and used computer resources provided by NSF award 0821391.

# REFERENCES


1. Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006a;45(38):11324-32.

2. Acharya A, Rishi V, Moll J, Vinson C. Experimental identification of homodimerizing B-ZIP families in homo sapiens. Journal of Structural Biology. 2006b;155(2):130-9.

3. Acharya A, Ruvinov SB, Gal J, Moll JR, Vinson C. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: Amino acids I, V, L, N, A, and K. Biochemistry. 2002;41(48):14122-31.

4. Ahn S, Olive M, Aggarwal S, Krylov D, Ginty DD, Vinson C. A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. Mol Cell Biol. 1998 Feb;18(2):967-77.

5. Ali MH, Taylor CM, Grigoryan G, Allen KN, Imperiali B, Keating AE. Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure. Structure. 2005 Feb;13(2):225-34.

6. Barth P, Schoeffler A, Alber T. Targeting metastable coiled-coil domains by computational design. J Am Chem Soc. 2008 Sep 10;130(36):12038-44.

7. Bolon DN, Grant RA, Baker TA, Sauer RT. Specificity versus stability in computational protein design. Proc Natl Acad Sci U S A. 2005 Sep 6;102(36):12724-9.

8. Cole JL, Lary JW. Heteroanalysis, analytical ultracentrifugation facility, university of connecticut, storrs, CT. . 2006

9. Countryman J, Jenson H, Seibl R, Wolf H, Miller G. Polymorphic proteins encoded within BZLF1 of defective and standard epstein-barr viruses disrupt latency. J. Virol. 1987 December 1, 1987;61(12):3672-9.

10. Feederle R, Kost M, Baumann M, Janz A, Drouet E, Hammerschmidt W, Delecluse HJ. The epstein-barr virus lytic program is controlled by the co-operative functions of two transactivators. EMBO J. 2000 Jun 15;19(12):3080-9.

11. Fong J, Keating A, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biology. 2004;5(2):R11.

12. Fu X, Apgar JR, Keating AE. Modeling backbone flexibility to achieve sequence diversity: The design of novel alpha-helical ligands for bcl-xL. J Mol Biol. 2007 Aug 24;371(4):1099-117.

13. Gerdes MJ, Myakishev M, Frost NA, Rishi V, Moitra J, Acharya A, Levy MR, Park SW, Glick A, Yuspa SH, Vinson C. Activator protein-1 activity regulates epithelial tumor cell identity. Cancer Res. 2006 Aug 1;66(15):7578-88.

14. Glover JN, Harrison SC. Crystal structure of the heterodimeric bZIP transcription factor c-fos-c-jun bound to DNA. Nature. 1995 Jan 19;373(6511):257-61.

15. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009;458(7240):859-64.

16. Grigoryan G, Keating AE. Structural specificity in coiled-coil interactions. Current Opinion in Structural Biology. 2008;18(4):477-83.

17. Grigoryan G, Keating AE. Structure-based prediction of bZIP partnering specificity. Journal of Molecular Biology. 2006;355(5):1125-42.

18. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science. 1993 November 26, 1993;262(5138):1401-7.

19. Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. Nat Struct Biol. 2003 Jan;10(1):45-52.

20. Hicks MR, Al-Mehairi SS, Sinclair AJ. The zipper region of epstein-barr virus bZIP transcription factor zta is necessary but not sufficient to direct DNA binding. J. Virol. 2003 July 15, 2003;77(14):8173-7.

21. Hoover DM, Lubkowski J. DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. Nucl. Acids Res. 2002 May 15, 2002;30(10):e43.

22. Karanicolas J, Kuhlman B. Computational design of affinity and specificity at protein-protein interfaces. Curr Opin Struct Biol. 2009 Aug;19(4):458-63.

23. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol. 2004 Apr;11(4):371-9.

24. Krylov D, Olive M, Vinson C. Extending dimerization interfaces: The bZIP basic region can form a coiled coil. EMBO J. 1995 Nov 1;14(21):5329-37.

25. Krylov D, Mikhailenko I, Vinson C. A thermodynamic scale for leucine zipper stability and dimerization specificity: E and g interhelical interactions. EMBO J. 1994 Jun 15;13(12):2849-61.

26. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Computer-aided interpretation of analytical sedimentation data for proteins. Analytical Ultracentrifugation in Biochemistry and Polymer Science. 1992:90-125.

27. Liu P, Speck SH. Synergistic autoactivation of the epstein-barr virus immediate-early BRLF1 promoter by rta and zta. Virology. 2003 Jun 5;310(2):199-206.

28. Lu SM, Hodges RS. Defining the minimum size of a hydrophobic cluster in two-stranded alpha-helical coiled-coils: Effects on protein stability. Protein Sci. 2004 Mar;13(3):714-26.

29. Mandell DJ, Kortemme T. Backbone flexibility in computational protein design. Curr Opin Biotechnol. 2009 Aug;20(4):420-8.

30. Mason JM, Hagemann UB, Arndt KM. Role of hydrophobic and electrostatic interactions in coiled coil stability and specificity. Biochemistry. 2009 Nov 3;48(43):10380-8.

31. Mason JM, Muller KM, Arndt KM. Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. Biochemistry. 2007 Apr 24;46(16):4804-14.

32. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. Chembiochem. 2004 Feb 6;5(2):170-6.

33. Mason JM, Schmitz MA, Müller KM, Arndt KM. Semirational design of jun-fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. Proceedings of the National Academy of Sciences. 2006 June 13, 2006;103(24):8989-94.

34. Moitra J, Szilak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. Biochemistry. 1997;36(41):12567-73.

35. Newman JRS, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003 June 27, 2003;300(5628):2097-101.

36. Oh WJ, Rishi V, Orosz A, Gerdes MJ, Vinson C. Inhibition of CCAAT/Enhancer binding protein family DNA binding in mouse epidermis prevents and regresses papillomas. Cancer Res. 2007 February 15, 2007;67(4):1867-76.

37. Olive M, Krylov D, Echlin DR, Gardner K, Taparowsky E, Vinson C. A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. J Biol Chem. 1997 Jul 25;272(30):18586-94.

38. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science. 1991 Oct 25;254(5031):539-44.

39. O'Shea EK, Lumb KJ, Kim PS. Peptide `Velcro': Design of a heterodimeric coiled coil. Current Biology. 1993;3(10):658-67.

40. Petosa C, Morand P, Baudin F, Moulin M, Artero J, Müller CW. Structural basis of lytic cycle activation by the epstein-barr virus ZEBRA protein. . 2006;21(4):565-72.

41. Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J Am Chem Soc. 2010a May 5;132(17):6025-31.

42. Reinke AW, Grigoryan G, Keating AE. Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. Biochemistry. 2010b Mar 9;49(9):1985-97.

43. Rishi V, Potter T, Laudeman J, Reinhart R, Silvers T, Selby M, Stevenson T, Krosky P, Stephen AG, Acharya A, Moll J, Oh WJ, Scudiero D, Shoemaker RH, Vinson C. A high-throughput fluorescence-anisotropy screen that identifies small molecule inhibitors of the DNA binding of B-ZIP transcription factors. Anal Biochem. 2005 May 15;340(2):259-71.

44. Sammond DW, Eletr ZM, Purbeck C, Kuhlman B. Computational design of second-site suppressor mutations at protein-protein interfaces. Proteins. 2010 Mar;78(4):1055-65.

45. Schelcher C, Al Mehairi S, Verrall E, Hope Q, Flower K, Bromley B, Woolfson DN, West MJ, Sinclair AJ. Atypical bZIP domain of viral transcription factor contributes to stability of dimer formation and transcriptional function. J Virol. 2007 Jul;81(13):7149-55.

46. Schepers A, Pich D, Hammerschmidt W. A transcription factor with homology to the AP-1 family links RNA transcription and DNA replication in the lytic cycle of epstein-barr virus. EMBO J. 1993 Oct;12(10):3921-9.

47. Steinkruger JD, Woolfson DN, Gellman SH. Side-chain pairing preferences in the parallel coiled-coil dimer motif: Insight on ion pairing between core and flanking sites. J Am Chem Soc. 2010 Jun 9;132(22):7586-8.

48. Taylor CM, Keating AE. Orientation and oligomerization specificity of the bcr coiled-coil oligomerization domain. Biochemistry. 2005 Dec 13;44(49):16246-56.

49. Tripet B, Wagschal K, Lavigne P, Mant CT, Hodges RS. Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". J Mol Biol. 2000 Jul 7;300(2):377-402.

50. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, Bonovich M. Classification of human B-ZIP proteins based on dimerization properties. Mol Cell Biol. 2002 Sep;22(18):6321-35.

51. Vinson C, Acharya A, Taparowsky EJ. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2006;1759(1-2):4-12.

52. Woolfson DN. The design of coiled-coil structures and assemblies. Adv Protein Chem. 2005;70:79-112.

53. Young LS, Rickinson AB. Epstein-barr virus: 40 years on. Nat Rev Cancer. 2004 Oct;4(10):757-68.