Quantitative modeling of non-additive protein-DNA energy landscapes: simulation and analysis

Michael P. Murrell, Phillip Dextras, Lenoid Mirny, Mehran Kardar

Prokaryotic transcription factors are known to find their target sequences to activate or repress transcription rapidly, and bind to their cognate sequences with high affinity. Previous studies have postulated that these proteins associate with their operator sequences by cytoplasmic diffusion coupled with one-dimensional sliding along the DNA. Attempts at modeling this process have focused on a "two-state" model, which seeks to explain how it is plausible for a protein to search a highly heterogeneous, sequence-dependent potential driven purely by thermal fluctuations. Here, we explore an alternative to the two state model, by proposing that the energy of interaction between the DNA and the protein binding domain is, in most cases, an overestimate, in that the overall contribution to the free energy of association is not independent and additive, but compensate; that adjacent base pairs when interacting with a residue of the binding domain of the protein affect the free energy of their neighbors. By using this assumption, we generated characteristic energy landscapes of P22 bacteriophage/Mnt interaction, and tested the overall search times by dynamic simulation. In the end, we compared the results to what could be expected from a two-state model, and propose conditions for yielding 'faster than diffusion' association with factors such as LacI and Integration Host Factor.

I. Introduction

In the bacterial transcriptional network, a small number of DNA-binding proteins that have been identified that find their target sequences to active or repress transcription faster than what is allowed by three-dimensional diffusion. Bacterial transcription factors are known to find their sites passively (i.e. in the absence of any ATP utilization), and therefore the maximal rate of association to a binding site is the diffusion limit. To this end, most of the known transcription factors are known to find their cognate sites along the DNA on the order of seconds, and bind with high affinity, on or near this limit. They do this however, in large concentrations, up to hundreds of molecules in the cell at any given time. Moreover, there can be anywhere between one and dozens of binding sites for any given transcription factor[1]. Consequently, a natural question that arises is how do proteins that appear in low concentrations, and have only a single target ('cognate') site reach their targets on the order of seconds, or less – faster than the rate of diffusion?

It has been postulated and widely accepted that these proteins undergo successive rounds of three dimensional diffusion followed by one-dimensional sliding along the DNA [2, 3]. In general, the protein dissociates from one strand, jumps into three dimensional space, and then re-associates with another strand nearby. This process continues until the protein has found its target sequence, which it binds with high affinity. In this sense, there is a change in the dimension of the transport of DNA binding proteins; three-dimensional diffusion in solution switches to one-dimensional searching along DNA. Various *in vitro* experiments add credit to this hypothesis [4], but inconsistencies present themselves when comparing the theoretical energies required making these searches feasible and the energies of known protein-DNA interactions– in particular, a statistical argument known as the "search-stability paradox" [5]. To preface, it is known that the stability of the protein-DNA complexes is high, on the order of 15kT. Yet, to efficiently search space in three dimensions, the variation in specific binding energy for sites along the genome (the "energy landscape") cannot exceed 1-2 kT. Moreover, presuming that this energy landscape is distributed normally, the probability of finding a site that is ~15kT is negligible. To yield a near unity probability of finding the site, the deviation in the distribution would need to be at least 3.5 kT. Thus it is apparent that with this model and these assumptions as a framework, we cannot have both stability of the transcriptional complex, and have fast search times.

In this study, we address this paradox by proposing the converse; we should understand and justify the transport mechanisms of proteins that are on the order of diffusion, based on key properties such as transcription factor structure and genome sequence, and determine which properties they do not share in common with faster DNA-binding proteins. In this sense, we aim to reconstruct a model of transcription factor translocation where physical paradoxes do not exist; then, by applying the same framework to an entirely different species, we can understand which differences can account for the disparate transport rates.

II. Model and Analysis

Most efforts that attempted to analyze the plausibility of one-dimensional sliding as part of transcription factor transport focused on characterizing the landscape distribution that would facilitate a faster search. This has involved generating normally distributed landscapes centered about an arbitrary mean and only a theoretically plausible standard deviation. Beginning with known single base pair mutations from equilibrium binding experiments, many authors assume that each base pair contributes independently to the overall binding free energy, and the entire landscape can be calculated from these values. While many agree this is in general a good assumption[6], it may in fact be untrue for many proteins, and may be the reason why 'faster than diffusion' proteins are still unexplained (this assumption has always been in question, and there have been numerous studies that aimed at justifying its use[6, 7]). Nevertheless, the energy landscape will still remain predominantly hypothetical, as these binding experiments can surely not be performed for all possible combinations of binding sequences 10-30 base pairs long on genomes on the order of 10^4-10^7 long.

Any attempt at constructing hypothetical energy landscapes will at best, result equally hypothetical results. However, the additivity assumption remains the most evident assumption in current models of transcription factor translocation, and consequently is the most likely assumption, that when removed, may yield different results. Thus far, there exists little effort to create a more realistic landscape, which would take into account interactions between base pairs. To be specific, it is possible, that multiple residues act on a single base pair, or that adjacent base pairs can contribute to each other's stability. Our method therefore, uses precisely this premise to adjust the energy landscape. To do so, we use knowledge of the frequency at which various amino acid – base pair interactions occur in 53 known transcription factor/DNA pairs (acquired from published crystal structures) [7, 8], and weight the individual interaction energy of non-consensus base pairs in combination. This information is summarized in a Contact Matrix (CM), the scored frequency of occurrence of each amino acid with each individual base pair. In addition, we use the same single base pair mutation data to construct what is termed a Position Weight Matrix (PWM), the specific $\Delta\Delta G$ s of mutations (See Appendix)[9].

In this study, we chose to focus on the Mnt Bacteriophage P22, an 80 residue protein that binds and represses transcription at a 17 base pair operator [10, 11]. The recognition of this target sequence depends on the interactions of four amino acids, Arg2, His6, Asn8, and Arg9, with the base pairs in the bound sequence (Table 1)[12]. The binding domain of the protein is a dimer, and the binding energy to the cognate sequence is symmetric about position 11 (Table 1). There are over 30 known NMR structures that capture Mnt/DNA complexes (or Arc/DNA complexes which is homologous), as well as the full binding experiments for all single base pair mutations. We can therefore use all of the structural information and mutational studies to construct binding energies for any combination of consensus and non-consensus base pairs in a given binding site.

Table 1. Mnt Consensus Sequence and Residue Interactions [10]

Nomencl.*	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
5' end	A	G	G	T	C	C	A	C	C	G	T	G	G	A	C	C	T
3' end	T	C	C	A	G	G	T	G	G	C	A	C	C	T	G	G	A
Position**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Residues		**N8**	**H6**	**H6**	**N8**	**R10**		**R2**		**R2**		**R10**	**N8**	**H6**	**H6**	**N8**	

*numbering convention used in literature; **numbering we use for our simulations

We tested a wide variety of landscapes, varying in their free energy per combination base pair interdependence, including a set of landscapes where we considered an additional state of the protein, a proposed "conformational change" where the protein senses only a specific "non-specific" energy for binding site mismatches (MM) that exceed a threshold value. As the goal of this project is to explain transcription factor translocation as a function of binding free energy, creating landscapes with different distributions is the key to defining a set of protein-DNA interactions, that when incorporated into an energy distribution, minimize the overall search time.

In generating over binding free energy, we consider two types of energy: the residue/base pair whose contribution to the overall energy (summed over j, Equation 1) depends on the consensus of its neighbors, and those that do not, and whose contribution is purely additive (summed over k). In this sense, all of the base pairs that have residue interactions are part of this first category, and those that do not fall into the latter. We therefore specify that any adjacent, non-

consensus residue/base pair interactions do not merely contribute their individual $\Delta\Delta G$ s as specified by the position weight matrix, but that the contribution from each non-consensus base pair is weighted by their respective frequency in the contact matrix.

$$(1) \qquad \Delta\Delta G_{site} = \Delta\Delta G_{\substack{residue \\ interactions}} + \Delta\Delta G_{\substack{non-residue \\ interactions}} = \frac{\sum_j CM_{j,\alpha} PWM_{j,\alpha}}{\sum_i CM_{i,\alpha}} + \sum_k PWM_{k,\alpha}$$

In this sense, instead of adding up the individual contributions from each of the elements in the position-weight matrix, we calculate a new energetic cost, which is will be less than what would have been calculated if we considered the interactions independent. There have been estimates that this difference grows with the number of mismatched pairs[6].

In the algorithm for calculating the overall free energy, we consider the cognate site to have the lowest energy, set to zero, and all other sites will be evaluated by the total number and type of mutations to this consensus sequence. For each mismatch, a characteristic $\Delta\Delta G$ will be added to the consensus energy, creating positive and highly unfavorable energies. As previously stated, for each non-cognate binding site, the algorithm identifies which base pairs are unmatched. In the case that one of the residue/DNA interactions in mismatched, it checks to see if one adjacent to it is also mismatched, and if true, adjusts their contributions to the overall free energy by Equation (1) above. If not, then include the overall $\Delta\Delta G$ as found in the position weight matrix. Since we only evaluate whether or not the immediate nearest neighbor is also non-consensus, mismatches more than one base pair apart do not contribute to each other's binding energy. The overall energy therefore, is highly specific to the protein used, and can be adjusted to incorporate additional experimental observations. In the case of Mnt Bacteriophage P22, it has been characterized that the base pairs at positions 3 and 4, as well as 14 and 15 are highly dependent on each other. From crystal structures, we know that there is no direct contact or bonding known to exist between positions 4 and 14 and any residue. Since they are known to effect the free energy contributions of their neighbors however, we presumed that these two positions interact with the H6 residue, and thus count it into the residue/DNA energetic contributions. This means that positions 2 through 6 and 12 through 16 are all dependent on each other, and provide the significant contribution to the sequence specific binding energy. Moreover, as positions 8 and 10 have no neighbors that interact with residues, we chose to include these as contributing additively, non-weighted, to the overall energy for each binding site.

The energy landscape calculated that includes all residue/DNA interactions as interdependent (minus positions 8 and 10) can be seen in Figure 1, where we see both the overall range and distribution of free energies for each position, as well as a plot of energies that surround the cognate site, which exists at the lowest energy state.

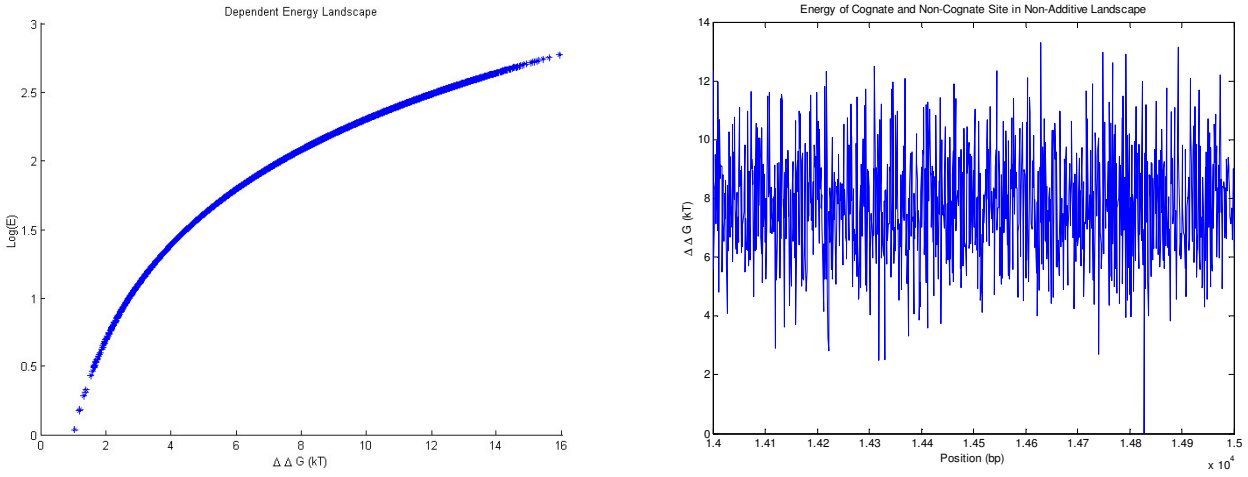

Figure 1. Free Energies of Association for Cognate and Non-Cognate Sites

As previously mentioned, we generated additional landscapes that varied in their degree of energetic interdependence of base pair combinations. Thus, in addition to testing non-additivity for *all* residues, we restricted this metric to only residues 2 and 3, as well as 12 and 13, the residues which have been experimentally proven to be the 'master residues', the control much of the specificity of binding[10, 13]. This was an effort to test different potential distributions that might yield an energy landscape that has a variance small enough to yield realistic search times.

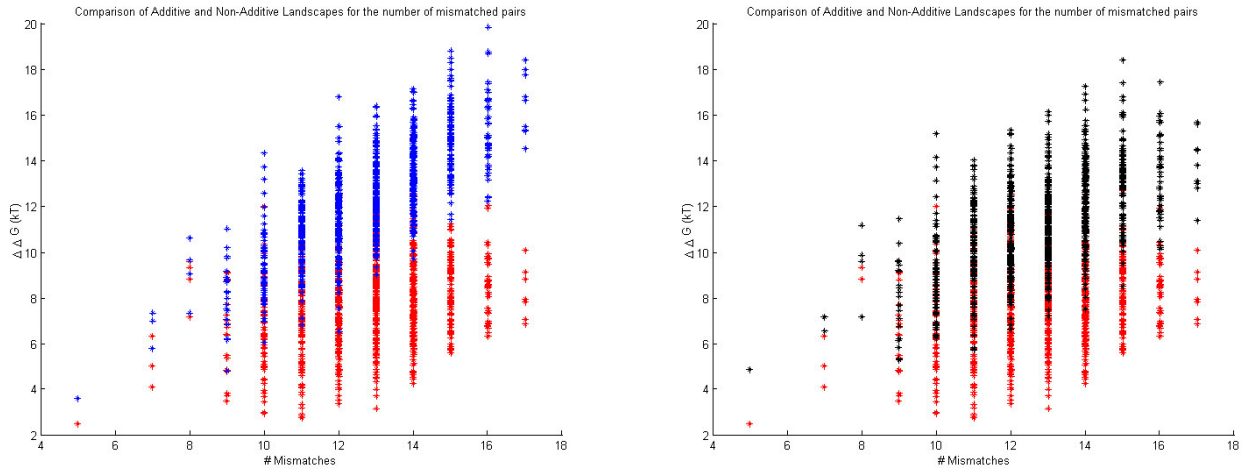

Figure 2. Comparison of the energetic costs per number of base pair mismatches. The non-additive landscape (red) is compared to the completely independent and additive sequence (blue) on the left. On the right, we compare the same non-additive landscape to a landscape where only the 'master residues' of the Mnt protein share free energy contributions dependently (black).

From Figure 2, we can see that for any number of mismatches between the binding site and the cognate site, the non-additive landscape has a lower ΔΔG. Again, this would be true, because as we presume that the total contribution to each overall binding affinity is not the sum of the individual energies, when more than one base pair is mismatched with the residues in the binding domain, they each share a portion of the overall ΔG, and thus the contribution from each of them is less than the sum of each individual ΔΔG that corresponds to a base pair deletion. Not surprisingly, when comparing the independent energy landscape to the landscape with energetic dependencies only between positions 2 and 3, and 12 and 13, the overall profile is very similar to the purely additive one. The primary reason is that for every binding sequence, the total enthalpic cost, or loss of binding specificity is a function of every base pair. Modulating the energetic contribution from four mispairing bases yields only a minor difference in the overall free energy, as 13 other bases still contribute. Unless we presume that these positions are principally responsible for the affinity and their mispairing results in a complete loss of specificity, then their matching or not does not significantly change the energy landscape. This is an important result, as it implies that to achieve a landscape with ubiquitously lower energy, nearly every combination of base pair mutations from the cognate sequence should be interdependent. Moreover, as the number of mispaired combinations increase, it should be that the total ΔΔG moves further away from the sum over all individual ΔΔGs.

An alternate method to computing weighted energy contributions for all binding sites in the genome is to establish a cutoff, or 'non-specific' energy for all binding sites where the number of mismatches exceeds a predetermined number. This energy, E_{ns}, represents a loss of all residue-base pair interactions, and is only the energy of interaction between the binding domain of the protein, and the backbone of the DNA. The assumption behind this method is that a certain number of contacts must be kept in order to maintain specificity, and that a majority of mismatches cannot provide sufficient searches. Moreover, by this method, one can directly alter the variance in free energies by modulating the number of mismatches would be necessary for loosing all specificity. We therefore chose different cutoffs and with an known nonspecific energy (17kT) to change the landscape [12, 14]. The rationale is that the binding affinity is not an entirely continuous, but depends upon an additional, internal energy state of the protein itself - a frequently modeled second state,

that reads the same free energy for any binding site with a high number of mismatches. Likewise, the variance in the landscape is nearly zero, taking into account the majority of constant free energy sites. We generated landscapes with varying cutoffs (Figure 3). It can be seen that for high cutoffs, there is a significant proportion of the overall population of binding sites that have non E_{ns} free energies, which then decreases exponentially as the cutoff decreases. In the event that translocation cannot be reproduced without resorting to a 2-state hypothesis, we see by Figure 3, that we would need to establish a cutoff energy at roughly 8 or less mismatched pairs before switching to a non-specific energy.

Table 2. Distribution parameters for Mnt-Specific Energy Landscapes

Energy Landscapes	Mean ΔΔG (kT)	Variance (ΔΔG) (k^2T^2)
Independent (Additive)	12.3876	5.1988
Non-Additive (all Residue pairs)	7.8730	3.7980
Non-Additive (Master Residue)	11.3394	4.8539

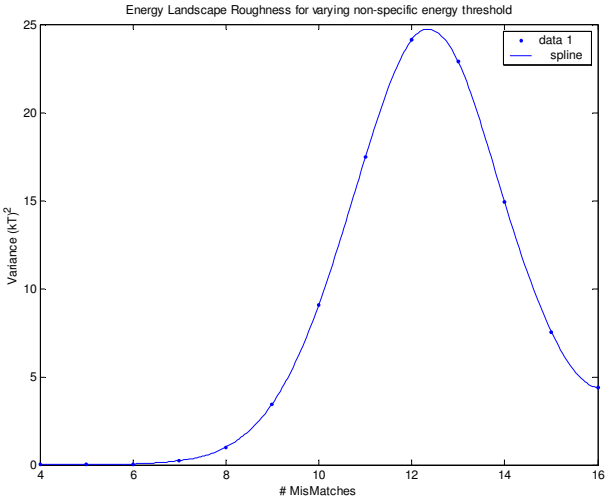

Figure 3. Landscapes with varying cutoffs for switching to nonspecific energy (17kT).

After defining the ranges of feasibility for realistic energy landscapes, we run Monte Carlo simulations of a protein diffusing in the cytoplasm and sliding along the DNA. First, we randomly walk a polymer through a 3D lattice, with a lattice constant equal to the length one binding site, on the length scale of 5.7nm. This occupies the space with the DNA, through a volume of 1 μm^3, the typical volume of an Ecoli cell. Then, beginning the walk in the center of the lattice, the particle makes has the choice of diffusing in three dimensions, or associating with the DNA. For the latter, there is an initial probability of collision, and an additional probability of association, given by the Boltzmann factor for the difference in energy between binding at the site α, and remaining in solution. Likewise, when associated with the DNA, the probability of moving in either direction, or jumping back into free solution is given by the Boltzmann factor, with the probability:

$$P_\alpha = \frac{e^{-\Delta\Delta G_\alpha / k_B T}}{Z} \tag{3}$$

where each energy term is the overall free energy of the sites along the DNA, the elements of the calculated energy landscape. The overall partition function Z, is the sum over all energy states,

$$Z = 1 + \sum_i e^{-\Delta\Delta G_i / k_B T} \tag{4}$$

where we set the cognate state to zero.

This framework was used for simulating landscapes to determine precisely what distribution parameter ranges would yield realistic search times. Each of the landscapes is normally distributed with different mean and variance parameters. Moreover, for each of these landscapes, we could check the proportion of time spent diffusing in one-dimension versus the time diffusing in three-dimensional space. The results of the Monte Carlo simulations on the hypothesized (randomly generated) landscapes are shown in Figure 4.

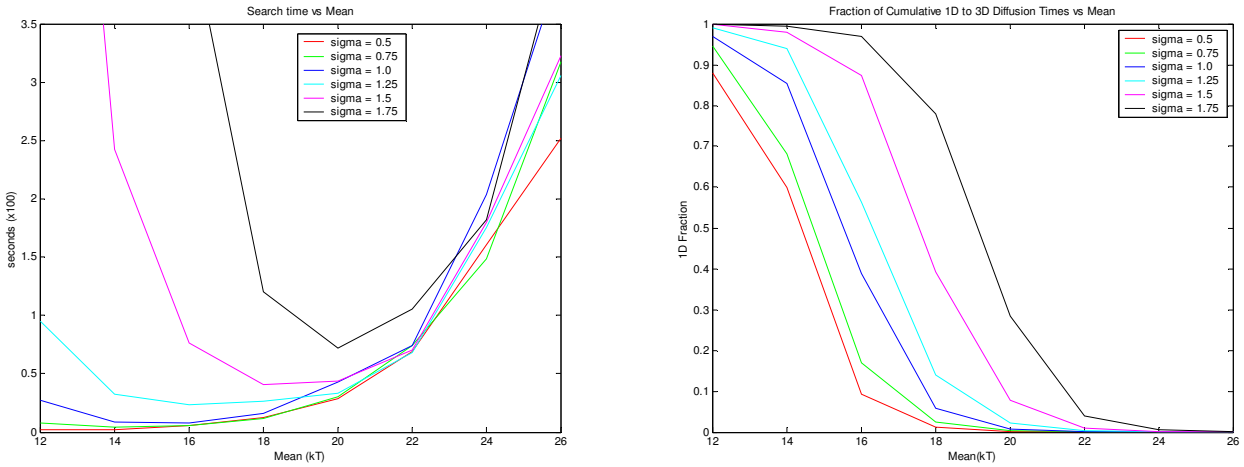

Figure 4. Monte Carlo simulations of particle diffusion in a three dimensional, polymer filled lattice were performed for hypothetical distributions. Each distribution had a separate mean and sigma, and was used to outline the requirements for feasible search times.

From Figure 4, we see that realistic search times only occur for σ less than about 1.25kT. This is consistent with previous studies, where estimates ranged between 0.5 and 2 kT. For the same distributions, we can also see in the fraction of time spent in 1D versus 3D. We understand by this figure, that the larger the mean, the less time is spent in 1D diffusion, as remaining in solution becomes more energetically favorable. However, for any given mean, the higher the deviation, the longer you remain on the strand. Thus, while we would prefer a high mean to accompany a low deviation, it is much more important to have a low deviation; as the deviation decreases, the sensitivity to the mean becomes less pronounced. This sheds light on the previously generated Mnt-specific landscapes. Incorporating the interdependence of free energy in the Mnt binding domain produces an σ ~ 1.9, and a mean of 7.5. Extrapolating to a mean of 7.5 in Figure 4 yields unrealistic search times (we can still run simulations for greater neighbor interactions, but ran out of time!). In fact, it would even take too long to simulate computationally. This however, was done for one protein, and one cognate site. In the event that there are hundreds of copies of the protein per cell, it may be possible, but in the absence of that information, we can make no further conclusion.

III. Discussion

It has been established that small differences in binding free energy are not consistent with the required stability of the cognate/protein complex, so pursuing this framework questions the role of one-dimensional diffusion, or alternatively, concludes that current energy landscapes are poorly approximated. Choosing not to reject the last thirty years of theory, we chose to investigate the latter scenario, that the possibility of explaining translocation of a protein across the cell on the order of seconds can be explained by a change in dimension during transport. Consequently, the task becomes evaluating the feasibility of this change in dimension – determining the critical parameters that affect the one-dimensional search and that minimize the overall search time. The primary obstacle to this is what has been described previously as the "search-stability paradox", a problem we now understand to be associated with more than just "faster than diffusion" proteins. A first argument would challenge the existence of a normally distributed landscape. To argue the extent of Gaussian character in the derived landscapes however, is inconsequential; for fast and efficient searching in one-dimension, there needs to be low binding site to binding site variation in free energy. Whether or not the

entire genome follows a normal vs. slightly non-normal distribution is irrelevant to a diffusing particle. We therefore chose to restrict our focus to the mechanisms that might reduce this overall variation. As part of this analysis, we generated various potential protein-DNA interactions with our model protein, the Mnt P22 repressor, and generate energy landscapes that might facilitate a fast search of the viral genome. In addition, we contrasted this more in depth analysis, with the standard two-state hypothesis, the assumption that specificity is only maintained when most binding site base pairs are consensus.

Our initial aim was to reproduce search times on the order of the diffusion limit, as the Mnt repressor is not one of the "faster than diffusion" proteins. We generated hypothetical, randomly generated landscapes which confirmed previous hypotheses, that realistic search times occur when the variance is low, on the order of 0.5-2 kT, and increase exponentially with increasing variation in base pair binding energy. Interestingly, we also find that with decreasing landscape variance, we get a decreased sensitivity of the total search time to mean values. This in essence, relieves us from the necessity to retain a high mean, in order to favor the dissociation from the DNA strand, and reinforces that the principal determinant of biologically relevant association rates is the variance of the binding energy landscape. Using this conclusion to address the Mnt-specific landscapes, we see that an increased interdependence in base pair to base pair free energy of non-cognate binding sites results in decreased variation in free energy across different binding site. Apparently, the imposed interactions between the DNA binding domain of the protein and the DNA template itself prove that the additivity assumption is an over-approximation. By this analysis however, it appears that while including 'nearest-neighbor' effects in the quantification of overall free energy will result in faster association time, it appears to not be fast enough to generate realistic search times. In this case, it is more likely that the Mnt repressor can be characterized by different internal states, conformation changes that can allow the protein to sample the DNA 'non-specifically', reading only a specified, non-specific energy for nearly all non-cognate sequences. From Table 2, we see that when the number of mismatches between the binding site and the cognate site falls below five base pairs, the deviation in the energy landscape goes to zero, and there is almost no resistance to thermally driven translocation in one-dimension. Nevertheless, we propose that the interaction between base pairs in the binding domain of 'faster than diffusion proteins' may in part explain their behavior. From Table 1, we see that roughly half of the base pairs in the cognate sequence will interact with residues in the binding domain of the protein. Moreover, there are no shared residues between base pairs. However, in the Lac repressor for example, there exist numerous base pair-residue interactions [15, 16]. In fact, many base pairs interact with multiple residues, theoretically enhancing the interdependence effect that we showed can occur with Mnt.

Ultimately, the justification for proposing methods of reducing binding energy variation to facilitate diffusion hinges on the knowledge of actual in vivo search times. Thus far, there exists little verification that in vivo one-dimensional sliding actually occurs. Thus, while our initial decision to explore approximations to the landscapes as a solution to the paradox; it may in fact exist, because one-dimensional diffusion is not a significant mechanism in the translocation of transcription factors. Gross over-estimates may have resulted in the apparent 'faster than diffusion' times, as they were all in vitro, and only differ by roughly an order of magnitude. Thus, while we would all like to believe that thermal processes are able to utilize interesting physical processes to control global mechanisms such as transcription, without further experimental evidence, its existence may in fact, be a myth.

IV. Appendix

Contact Matrix

A	C	G	T	A.A.
-3.93	-3.72	-3.93	0.66	A
0.07	0.07	-2.23	-2.23	C
-3.37	1.01	-3.93	-3.93	D
-1.24	0.55	-3.93	-3.93	E
-3.93	-0.12	-3.93	-0.81	F
-3.93	-3.93	-3.93	-3.93	G
0.46	-0.23	1.56	0.87	H
-3.93	-3.44	-3.93	0.65	I
-0.08	-3.93	2.16	0.21	K
-3.93	-3.93	-3.93	-0.94	L
-0.28	-0.28	-2.58	0.42	M
1.93	0.71	0.48	0.71	N
-3.93	-3.29	-3.93	-0.30	P
1.16	-3.09	-3.93	-0.3	Q
0.34	-3.93	2.74	1.25	R
-0.68	-0.68	0.42	-0.28	S
-0.06	-1.16	-3.46	-0.06	T
-3.93	-3.93	-1.96	-1.96	V
-2.87	0.13	-2.87	0.54	Y

Position Weight Matrix

A	C	G	T	Nomenc.*	Position
0	0.61	-0.17	0.55	3	1
0.44	0.63	0	0.47	4	2
1.46	2.02	0	1.27	5	3
0.42	0.74	0.18	0	6	4
1.41	0	1.3	0.618	7	5
1.36	0	2.58	0.99	8	6
0	0.99	1.21	1.54	9	7
1.88	0	1.55	1.05	10	8
0.42	0	0	0.42	11	9
1.05	1.55	0	1.88	12	10
1.54	1.21	0.99	0	13	11
0.99	2.58	0	1.36	14	12
0.618	1.3	0	1.41	15	13
0	0.18	0.74	0.42	16	14
1.27	0	2.02	1.46	17	15
0.47	0	0.63	0.44	18	16
0.55	-0.17	0.61	0	19	17

* Refers to the numbering convention used in literature.

V. Acknowledgements

We have benefited a great deal from our many conversations with Michael Slutsky. We wish to acknowledge him for all of his help.

---

1. Robison, K., A.M. McGuire, and G.M. Church, *A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome.* J Mol Biol, 1998. **284**(2): p. 241-54.
2. Berg, O.G., *Selection of DNA binding sites by regulatory proteins. Functional specificity and pseudosite competition.* J Biomol Struct Dyn, 1988. **6**(2): p. 275-97.
3. Berg, O.G., *Selection of DNA binding sites by regulatory proteins: the LexA protein and the arginine repressor use different strategies for functional specificity.* Nucleic Acids Res, 1988. **16**(11): p. 5089-105.
4. Halford, S.E. and M.D. Szczelkun, *How to get from A to B: strategies for analysing protein motion on DNA.* Eur Biophys J, 2002. **31**(4): p. 257-67.
5. Slutsky, M. and L.A. Mirny, *Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential.* Biophys J, 2004. **87**(6): p. 4021-35.
6. Benos, P.V., M.L. Bulyk, and G.D. Stormo, *Additivity in protein-DNA interactions: how good an approximation is it?* Nucleic Acids Res, 2002. **30**(20): p. 4442-51.
7. Benos, P.V., A.S. Lapedes, and G.D. Stormo, *Is there a code for protein-DNA recognition? Probab(ilistical)ly.* Bioessays, 2002. **24**(5): p. 466-75.

8. Mandel-Gutfreund, Y. and H. Margalit, *Quantitative parameters for amino acid-base interaction: implications for prediction of protein-DNA binding sites.* Nucleic Acids Res, 1998. **26**(10): p. 2306-12.
9. Stormo, G.D. and D.S. Fields, *Specificity, free energy and information content in protein-DNA interactions.* Trends Biochem Sci, 1998. **23**(3): p. 109-13.
10. Silbaq, F.S., S.E. Ruttenberg, and G.D. Stormo, *Specificity of Mnt 'master residue' obtained from in vivo and in vitro selections.* Nucleic Acids Res, 2002. **30**(24): p. 5539-48.
11. Vershon, A.K., et al., *Bacteriophage P22 Mnt repressor. DNA binding and effects on transcription in vitro.* J Mol Biol, 1987. **195**(2): p. 311-22.
12. Raumann, B.E., K.L. Knight, and R.T. Sauer, *Dramatic changes in DNA-binding specificity caused by single residue substitutions in an Arc/Mnt hybrid repressor.* Nat Struct Biol, 1995. **2**(12): p. 1115-22.
13. Fields, D.S., et al., *Quantitative specificity of the Mnt repressor.* J Mol Biol, 1997. **271**(2): p. 178-94.
14. Gerland, U., J.D. Moroz, and T. Hwa, *Physical constraints and functional characteristics of transcription factor-DNA interaction.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12015-20.
15. Kalodimos, C.G., R. Boelens, and R. Kaptein, *Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system.* Chem Rev, 2004. **104**(8): p. 3567-86.
16. Kalodimos, C.G., et al., *Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes.* Science, 2004. **305**(5682): p. 386-9.