Foundations of Computational and Systems Biology
7.36, 7.91, BE.390, BE.490
Prereq: (7.05 or 5.07) or (7.01 and 6.001) or permission of the instructors, but see Prerequisites below

Course Meetings
Tuesday and Thursday 1-2:30
plus optional weekly recitations
Room 32-141

Spring, 2004
3-0-9
Introduction to computational biology emphasizing the fundamentals of nucleic acid and protein sequence and structural analysis, also including an introduction to the analysis of complex biological systems. Covers principles and methods used for sequence alignment, motif finding, structural modeling, structure prediction and network modeling. Includes exposure to currently emerging research areas. Subject designed for advanced undergraduates and graduate students with strong backgrounds in either molecular biology or computer science, but not necessarily both. Foundational material covering basic programming skills will be provided. Enrollment limited to 75.

Instructors
Chris Burge cбурге@mit.edu
Amy Keating keating@mit.edu
Michael Yaffe myaffe@mit.edu

Teaching Assistants
Gevorg Grigoryan gevorg@mit.edu
Alex Mallet amallet@mit.edu

Python/Probability/Statistics Instructor
Graham Ruby jimigr@mit.edu

Recitations
Two recitation sections will be offered each week. Times will be determined on the first day of class. These will be led by the TAs and will provide students an opportunity to ask questions about material presented in lecture, the readings or the homeworks. Recitations are optional. However, if you are having difficulty in the class, you are strongly encouraged to attend regularly.

Office Hours
Meetings with the instructors or the TAs can be arranged by email request. If there is sufficient interest, regular office hours will be scheduled.

Prerequisites
A background in basic genetics, biochemistry and molecular biology (as taught in 7.01, 7.03 and 7.05) is required for this course. For example, we assume that you are familiar with chapters 3, 4 and 6 of the textbook Molecular Biology of the Cell, Alberts et al. If you are having problems with this introductory material, please see your teaching assistant. Computer programming experience is NOT required.

Web site
All course materials, including copies of the lecture slides, will be distributed via Stellar. Access to the course Stellar site is linked from the course home page:
http://mit.edu/7.91/

**Auditors/Listeners**
The class may be audited only by permission of one of the instructors. Please meet us after class and tell us who you are and why you want to audit. Students are strongly encouraged to take the course for credit.

**Text Books** (with 2005 prices from barnesandnobel.com):
The following text is *highly recommended* for this course and is available at the COOP. Bioinformatics: *Sequence and Genome Analysis* (2nd Edition), David W. Mount
$84.55

Assigned readings will be from the Mount book, as well as from the primary literature. However, we have also selected the following texts as particularly useful in selected areas, if you are looking for further information. A subset of these references will be on reserve in the Hayden Science library.

*Background in python programming (introductory):*
useful web tutorial
http://www.python.org/doc/current/tut/tut.html

**Learning Python** (2nd Edition), David Ascher and Mark Lutz
this is available on-line for free for MIT students at http://safari.oreilly.com/

**Python in a Nutshell**, Alex Martelli, Paula Ferguson (Editors)
$34.95

*Background in basic molecular biology, cell biology, biochemistry:*

**Molecular Biology of the Gene**, Watson, Baker, Bell, Gann, Levine & Losick
$124.45

**Molecular Biology of the Cell**, Alberts, Johnson, Lewis, Raff, Roberts & Walter
$104.50

**Biochemistry**, Stryer, Tymoczko, Berg
$132.76

**Introduction to Protein Structure**, 2nd edition Carl-Ivar Branden, John Tooze
$56.00

**Protein Structure and Function**, Petsko and Ringe
$47.45

*Other useful reference books for the class:*

**Molecular Modelling: Principles and Applications** (2nd edition), Andrew R. Leach
$85.50

In-depth treatment of molecular mechanics and other modeling methods.

$45.60

Contains more in-depth discussion of many of the algorithms and methods discussed in class.

**The Cartoon Guide to Statistics**, Larry Gonick (Author), Woollcott Smith
$11.86

Better than it sounds, and fun too! A very useful introduction or refresher.

$37.95
Covers basic statistics, with applications to biology and medicine.

**Structure of Class**
The majority of sessions will consist of lectures by the instructors. On occasion, class time will be used to discuss current research literature. Modern papers will be used to illustrate how the methods you are learning about in class are used every day to address important biological questions, and will also serve to integrate different parts of the curriculum. On each of the literature discussion days one or two papers will be assigned, and a student discussion will be led by the instructors. The assigned papers will often provide the basis for homework and exam problems.

**Homework**
Seven written or computer-based homework assignments will be posted on the course web site. These are designed to promote deeper understanding of the algorithms discussed in class and to provide hands-on experience with bioinformatics and computational biology tools. Students may discuss the homework problems amongst themselves, but all students must complete their own assignment. Duplicate or nearly-identical homework from different students will not be accepted.

The dates that the assignments are due are included in the syllabus below. Homework must be turned in within 15 minutes of the beginning of class on the due date to be eligible for full credit. Assignments turned in to the Biology education office (68-120) by noon two work days following the due date (Thursday for a Tuesday due date, Monday for a Thursday due date, Tuesday for a Friday due date) will be eligible for 50% credit. NO HOMEWORKS WILL BE ACCEPTED AFTER THIS TIME. NO EXCEPTIONS. Your lowest homework grade will be dropped, but no make-up assignments will be given.

**Exams**
There will be an in-class midterm exam on Tuesday, April 11th covering all of the topics up to that point. There will also be a final exam that will be comprehensive but biased toward material covered after the midterm.

**Grading**
Grades will be assigned based on the following scheme:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework assignments</td>
<td>42%</td>
</tr>
<tr>
<td>In-class midterm</td>
<td>20%</td>
</tr>
<tr>
<td>Final exam</td>
<td>35%</td>
</tr>
<tr>
<td>Class participation</td>
<td>3%</td>
</tr>
</tbody>
</table>

**Python Programming & Instruction**
The homework assignments will include problems that involve writing simple programs in the scripting language Python. Python and Perl are widely used for bioinformatics and computational biology. Because many students may have little or no programming experience, hands-on tutorials in Python will be offered by the Python Instructor, Graham Ruby, during the first and second weeks of classes (see times below). Graham will also be available to answer questions regarding programming, as well as basic probability and statistics, during scheduled office hours for the first several weeks of class.

The TAs will additionally be available in section to answer questions about writing Python scripts, as well as about other materials in the course. If you are having difficulty after attending all of these sessions and studying in the books recommended above, see your TA. Note that specific questions about the homework assignments – including the programming assignments - should be addressed to your TAs and not to
the Python Instructor. Graham is available to help those with little or no programming background learn the basics of Python programming during the first few weeks of the course only.

There will be two “Intro to Python” sessions, targeting those with little or no previous programming experience.

**Session I**  
Wednesday February 8, 7-8:30 pm  
Thursday February 9, 7-8:30 pm

**Session II**  
Tuesday February 14, 3-4:30 pm  
Wednesday February 15, 3-4:30 pm

All meetings will be held in room 26-204.

Sessions II will be a repeat of session I, although you are welcome to attend both if you wish to have extra reinforcement of the concepts introduced. Notes outlining the materials covered in these sessions, as well as short exercises designed to help you get up to speed (NOT for credit), will be posted on Stellar.

**Basic Probability and Statistics**  
A primer covering basic concepts in probability and statistics that are useful for this class is available at the class web site. The workbook is designed for students to complete at their own pace, and/or to use as a reference. It includes explanatory material and many examples drawn from biology. Students who have less background in these areas or need a refresher are strongly encouraged to read the primer and do the examples. Graham Ruby will be available to answer questions about this material for the first month of class and your TAs will be able to answer questions throughout the semester.

**Intellectual Honesty**  
We hope and trust that academic misconduct will not occur during this course. We nevertheless want to emphasize that we will be rigorous in our enforcement of Institute rules. It is the policy of the Biology Department to keep a record of all cases of academic misconduct and to forward cases to the Dean of Undergraduate and Student Affairs.
Syllabus

February 2006
7 Tuesday:  Intro/Sequence Comparison and Dynamic Programming (Yaffe)

9 Thursday  Multiple Sequence Alignments I (Yaffe)

14 Tuesday  Multiple Sequence Alignments II (Yaffe)
16 Thursday  Phylogenetic Analysis (Yaffe)

HOMEWORK #1 AVAILABLE ON-LINE

18 Tuesday  Monday schedule of classes

21 Tuesday  Protein secondary structure prediction (Yaffe)

23 Thursday:  Introduction to protein structure and structural classification

HOMEWORK #2 DUE IN CLASS
HOMEWORK #3 AVAILABLE ON-LINE

March 2006
2 Thursday  Principles of molecular mechanics (Keating)

7 Tuesday  Solving structures using X-ray crystallography and NMR (Keating)
9 Thursday  Modeling mutants and homologs (Keating)

HOMEWORK #3 DUE IN CLASS
HOMEWORK #4 AVAILABLE ON-LINE

March 10 is ADD DATE

14 Tuesday  Threading and ab initio structure prediction (Keating)
16 Thursday  Computational protein design/Proteomics I – experimental methods

21 Tuesday  Proteomics II – Analysis of Biological Networks (Keating)

HOMEWORK #4 DUE IN CLASS

23 Thursday  wrap up Proteomics/Literature discussion on protein structure/proteomics (Keating)

Monday March 27 through Friday March 31 is Spring Break

April 2006
4 Tuesday  Genome Sequencing and DNA Sequence Analysis (Burge)

HOMEWORK #5 AVAILABLE ON-LINE

6 Thursday  DNA Sequence Comparison and Alignment (Burge)
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Tuesday</td>
<td>MIDTERM EXAM – in class</td>
</tr>
<tr>
<td>13 Thursday</td>
<td>DNA Motif Modeling and Discovery (Burge)</td>
</tr>
<tr>
<td>14 Friday</td>
<td>HOMEWORK #5 DUE TO BIOLOGY EDUCATION OFFICE, 68-120, BY 3 PM</td>
</tr>
<tr>
<td>18 Tuesday</td>
<td>In Institute holiday – no class (Patriots’ Day)</td>
</tr>
<tr>
<td>20 Thursday</td>
<td>Markov and Hidden Markov Models for DNA Sequences (Burge)</td>
</tr>
<tr>
<td>25 Tuesday</td>
<td>DNA Sequence Evolution (Burge)</td>
</tr>
<tr>
<td>27 Thursday</td>
<td>RNA Secondary Structure Prediction (Burge)</td>
</tr>
<tr>
<td>11 Thursday</td>
<td>Feedback systems and coupled differential equations I (Yaffe)</td>
</tr>
<tr>
<td>12 Friday</td>
<td>HOMEWORK #7 DUE</td>
</tr>
<tr>
<td>16 Tuesday</td>
<td>Feedback systems and coupled differential equations II (Yaffe)</td>
</tr>
<tr>
<td>18 Thursday</td>
<td>Discussion of Current Literature and Course Topics Review (all)</td>
</tr>
</tbody>
</table>

April 27 is drop date

May 2006:

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Tuesday</td>
<td>Data for Systems Biology (Burge/Yaffe)</td>
</tr>
<tr>
<td>4 Thursday</td>
<td>Clustering and Principal Component Analysis (Burge/Yaffe)</td>
</tr>
<tr>
<td>9 Tuesday</td>
<td>Bayesian Networks (Burge)</td>
</tr>
<tr>
<td>11 Thursday</td>
<td>Feedback systems and coupled differential equations I (Yaffe)</td>
</tr>
<tr>
<td>12 Friday</td>
<td>HOMEWORK #7 DUE</td>
</tr>
</tbody>
</table>

Final exams are May 22-26