

CS174: Other Topics in Bioinformatics

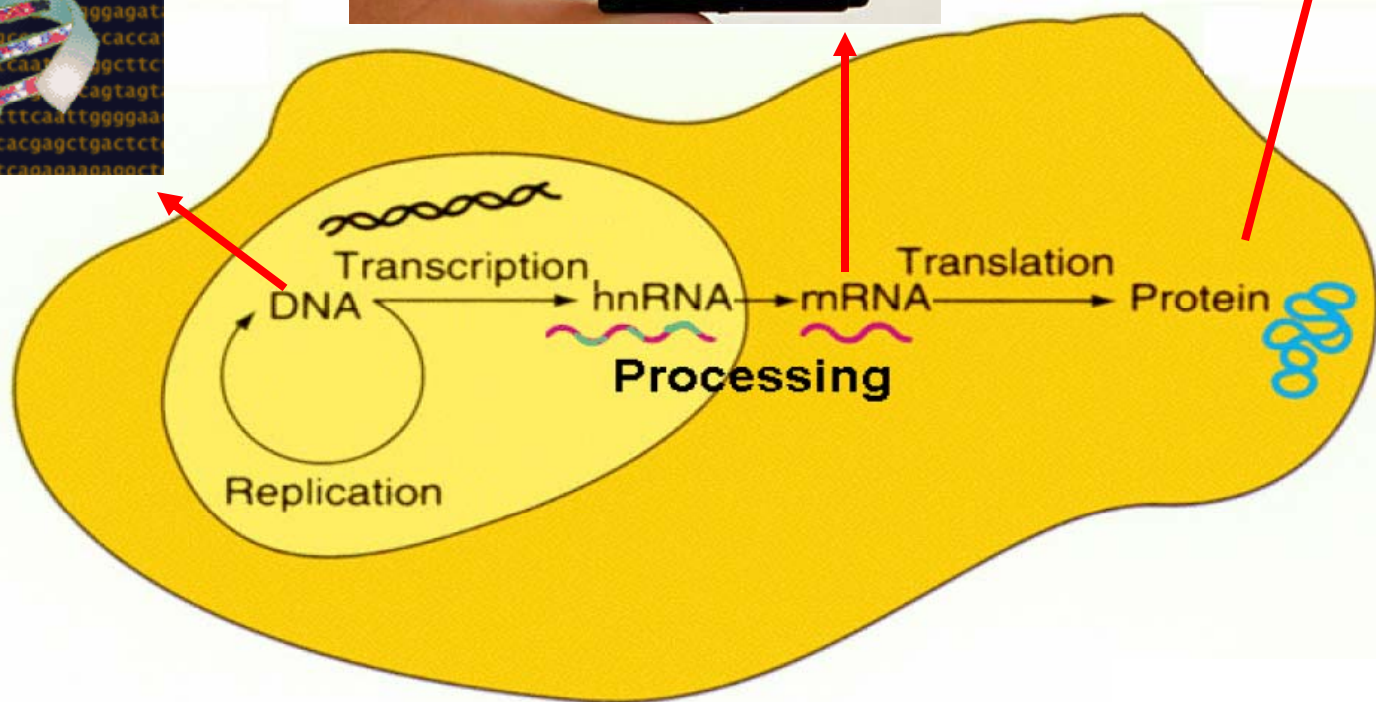
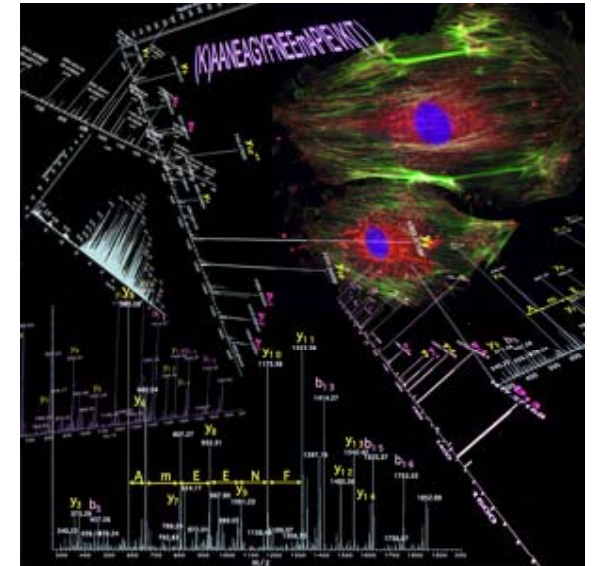
Topics covered so far

- Gene (ORF) discovery
- Regulatory motif discovery
- Sequence alignments
- Genome assembly
- Hidden Markov model

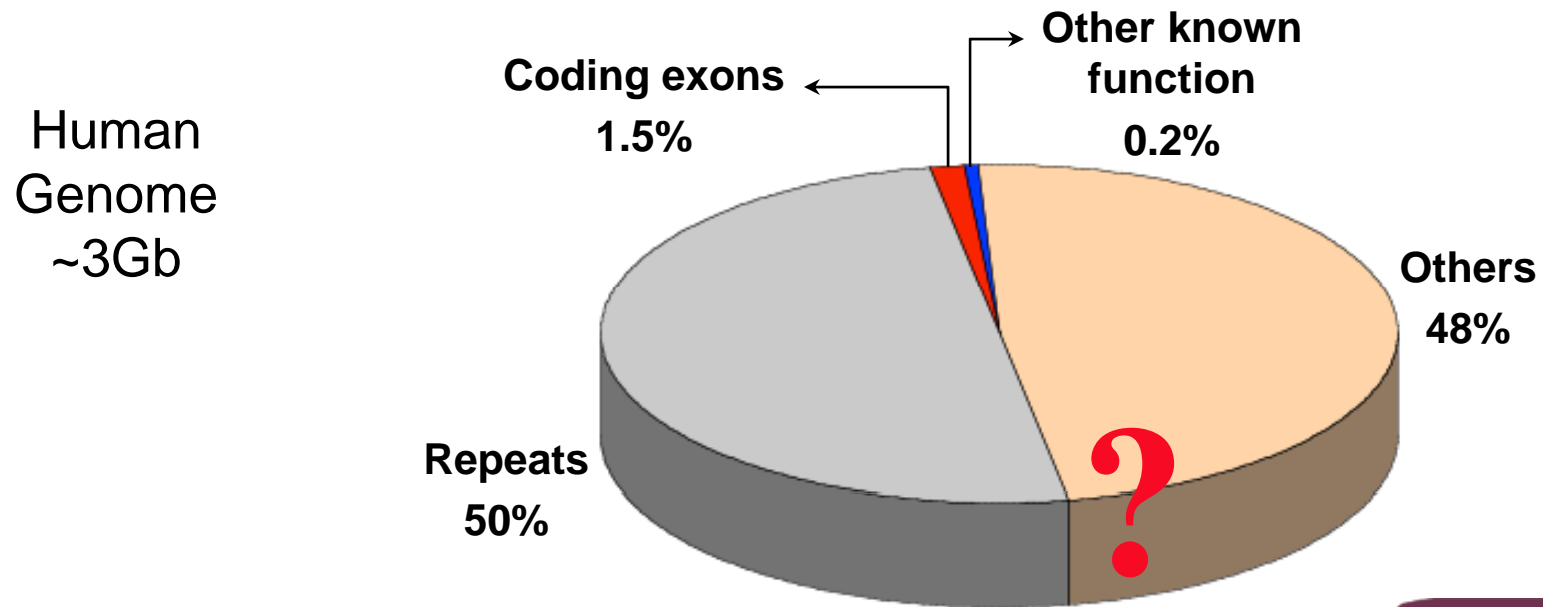
Other Topics:

- Comparative Genomics
- Protein structure prediction
- Systems biology: clustering, reverse-engineering approaches
- Evolutionary theory
- Population dynamics

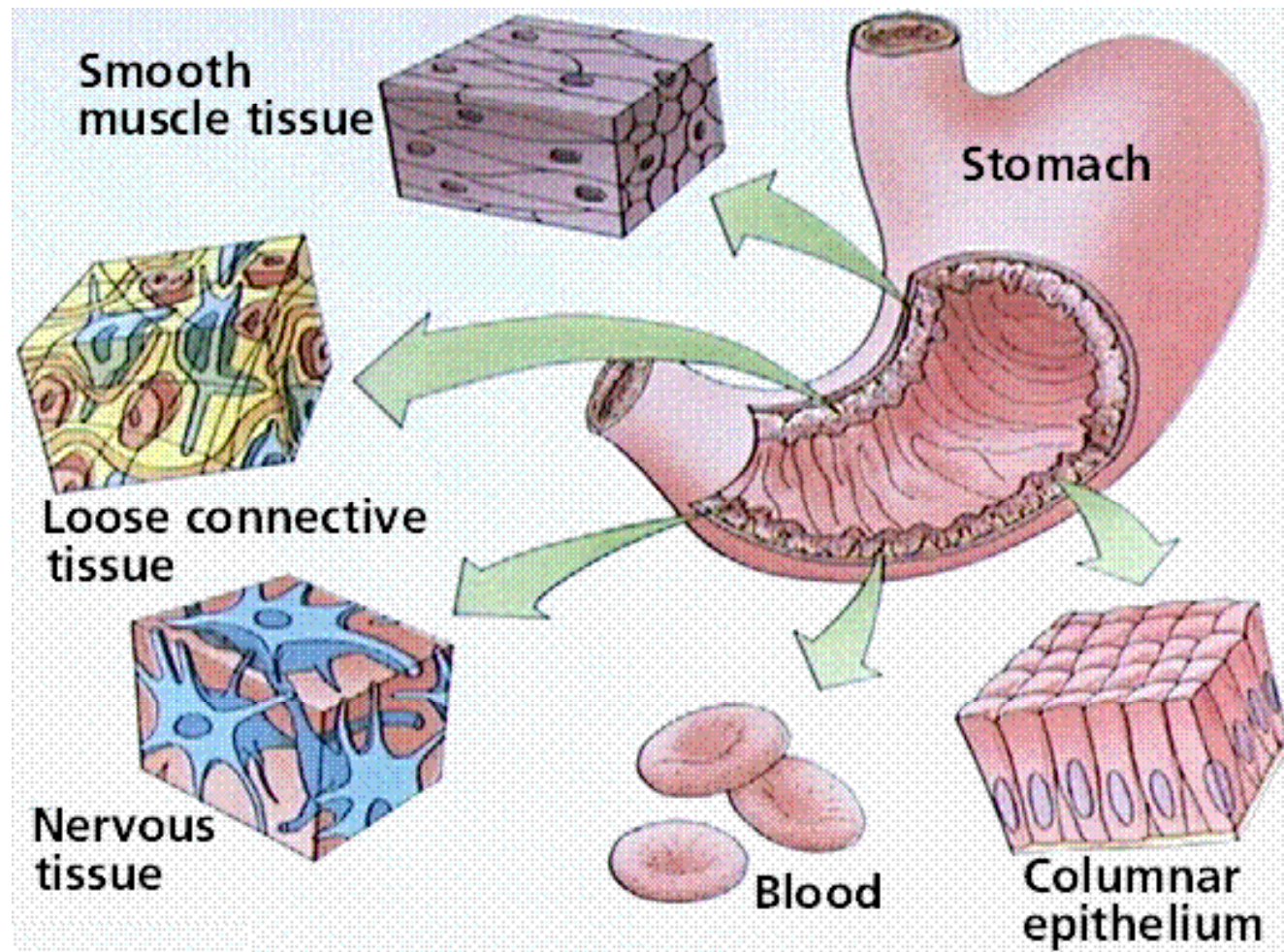
Readout from the genome



98% of the human genome unknown



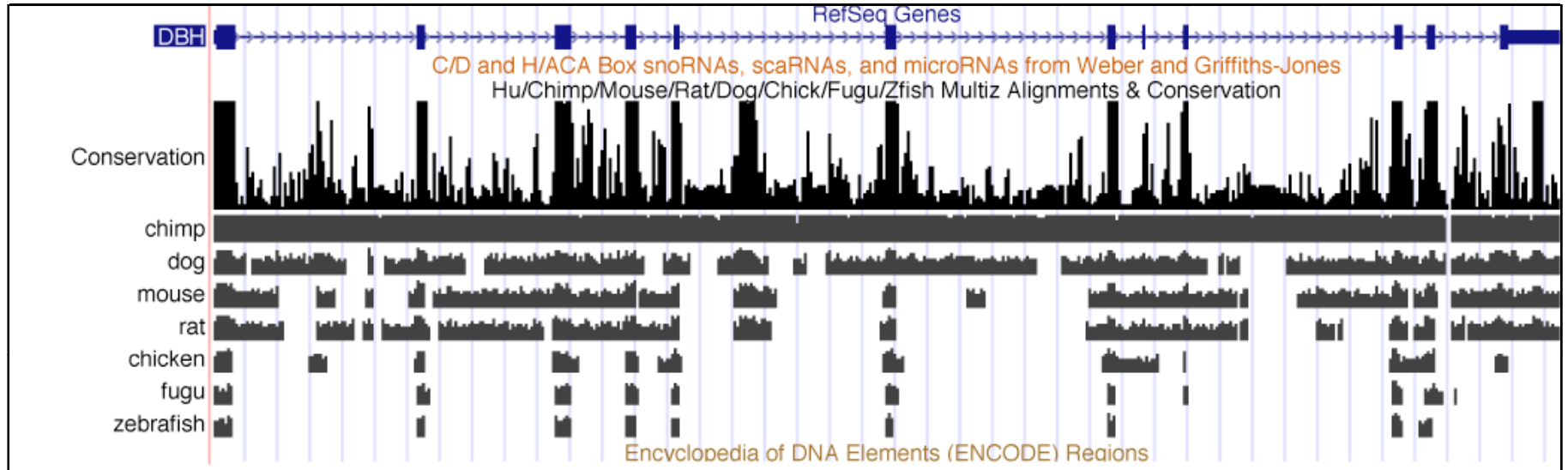
Example: Tissues in Stomach



How is this variety encoded and expressed ?

Comparative Genomics

Comparative genomics and evolutionary signatures



- ✓ **Comparing genomes can reveal functional elements**
- ✓ **Can we also pinpoint specific functions of each elements? Yes!**
 - Patterns of change distinguish different types of functional elements
 - Specific function ↔ Selective pressures ↔ Patterns of mutations/indels
- ✓ **Develop evolutionary signatures characteristic of each function**

hg18.chr15	AGACTAGTGATTTTGTGTTGTCTTAC--TGC-GCTCAACAACAAATCCCAGTCT
panTro1	AGACTAGTGATTTTGTGTTGTCTTAC--TGC-GCTCAACAACAAATCCCAGTCT
rheMac2	AGACTAGTGATTTTGTGTTGTCTTAC--TGC-GCTCAACAACAAATCCCAGTCT
mm8	AGACTAGTGATTTTGTGTTGTGTCTC--TGT-ATCCAACAACAAGTCCCAGTCT
oryCun1	AGACTAGTGATTTTGTGTTGTCTCGC--TGT-GCTCAACAACAAGTCCCAGTCT
rn4	AGACTAGTGATTTTGTGTTGTG--TC--TGT-GTCCAACAACAAGTCCCAGTCT
bosTau2	AGACTAGTGATTTTGTGTTGTCTC-C--TGC-GCTCAACAACAAGTCCCAGTCT
canFam2	AGACTAGTGATTTTGTGTTGTCTCAC--TGC-ATCCAACAACAAGTCCCAGTCT
dasNov1	AGACTAGTGATTTTGTGTTGTCTCACTGTGC-GCTCAACAACAAGTCCCAGTCT
loxÅfr1	AGACTAGTGATTTTGTGTTGTCTCAT--TACCGTTCAACAACAAGTCCCAGTCT
echTel1	AGACTAGTGATTTTGTGTTGTCTCGC--TACTGCTCAACAAC.....
monDom4	AGACTAGTGATTTTGTGTTCTCTAACGTAAA-GATTGACAACAATCCCAGTCT

Signatures specific to RNA genes:

- Stem conservation >> loop conservation
- Compensatory changes for paired bases
- Gaps allowed

has-mir-7

Evolutionary Signatures: RNA genes

Known motifs are preferentially conserved

human	CTCTTAATGGTACACGTTCTGCCT----	AAGTAGCCTAGACGCTCCCGTGCGCCC-GGGG
dog	CTCTTA-CGGGGCACATTCTGCTTTCAACAGTGGGGCAGACGGTCCCGCGCGCCCCAAGG	
mouse	GTCTTAGGAGGCT-CGATCGCC-----	GCCTGCATTATT-----
rat	GTCTTAGTTGGCCACGACCTGC-----	TCATGCATAATT-----
	***** * * * *	* *

Errα

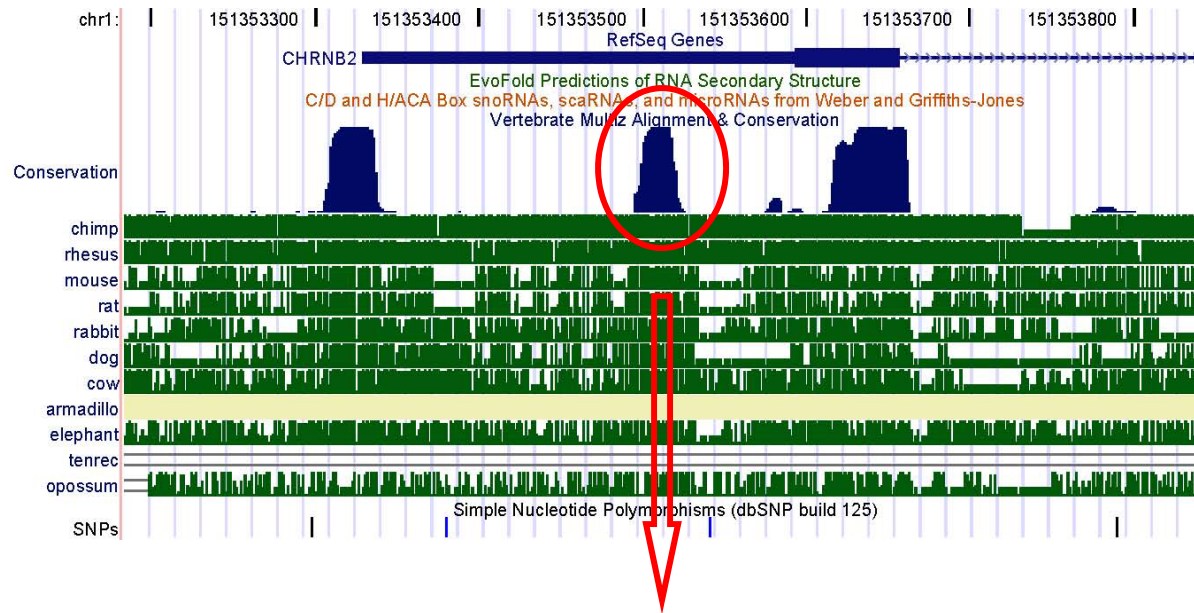
human	CGGGTAGGCCTGGCCGAAAATCTCTCCCGCGCGCC	TGACCTTG	GGTTGCCCCAGCCAGGC
dog	CAGGC---CCGGGCTGCAGACCTGCCCTGAGGGAA	TGACCTTG	GGCGGCCGCAGCGGGGC
mouse	-----CACAAGCCTGTGGCGCGC-CG	TGACCTTG	GGCTGCCCCAGGCGGGC
rat	-----CACAAGTTTCTC---TGC-CC	TGACCTTG	GGTTGCCCCAGGCGAG-
	* * *	*****	** *** ** *

human	TGCGGGCCCGAGACCCCCG-----	GGCCTCCCTGCCCCCGCGCCG
dog	CGCGGGCCCGAGCCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTGCCCCCGGACCG	
mouse	TGCAGGCTCACCACCCCGTCTTTTCT-----	GCTTTTCGAGTCG
rat	-GCATACACCCCGCCTTTTTTTTTTTTTTTT-----	TTTTTTTTTTGCCGTTCAAG-AG
	** * * **	** * *

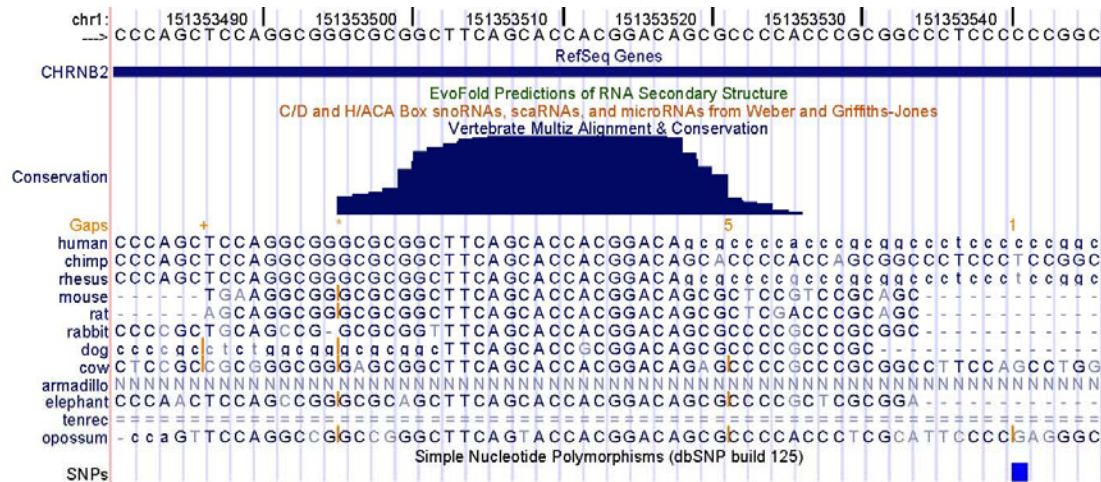
Gabpa



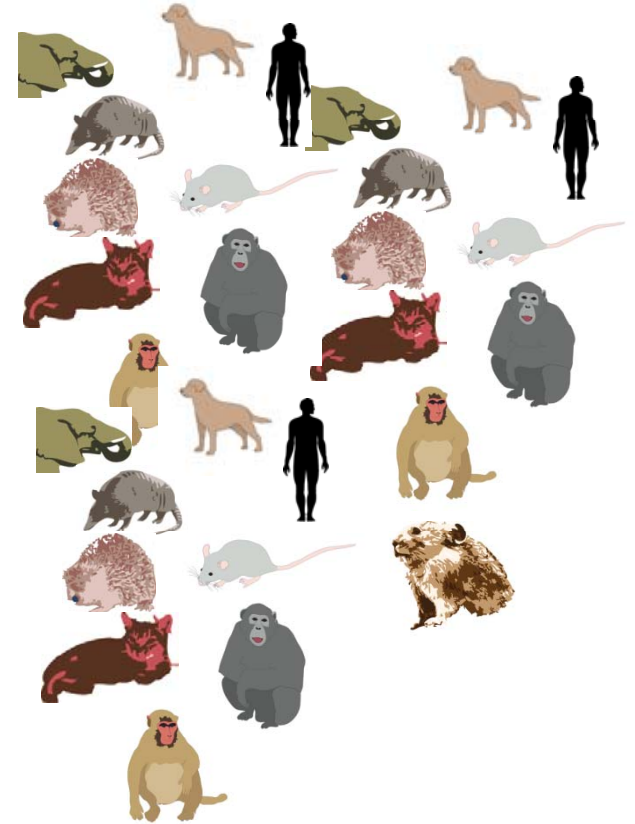
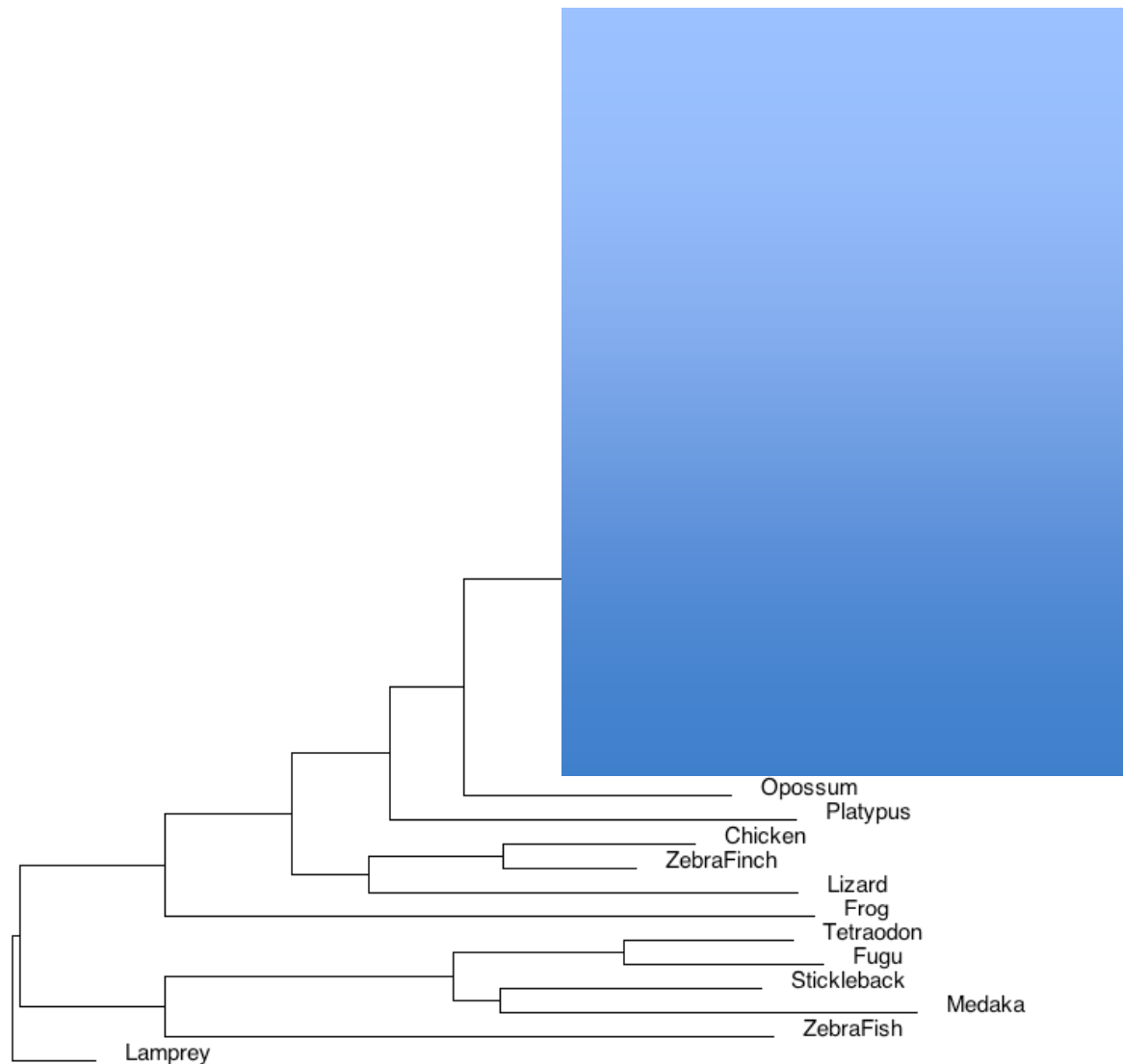
Evolutionary Signatures: Regulatory Motifs



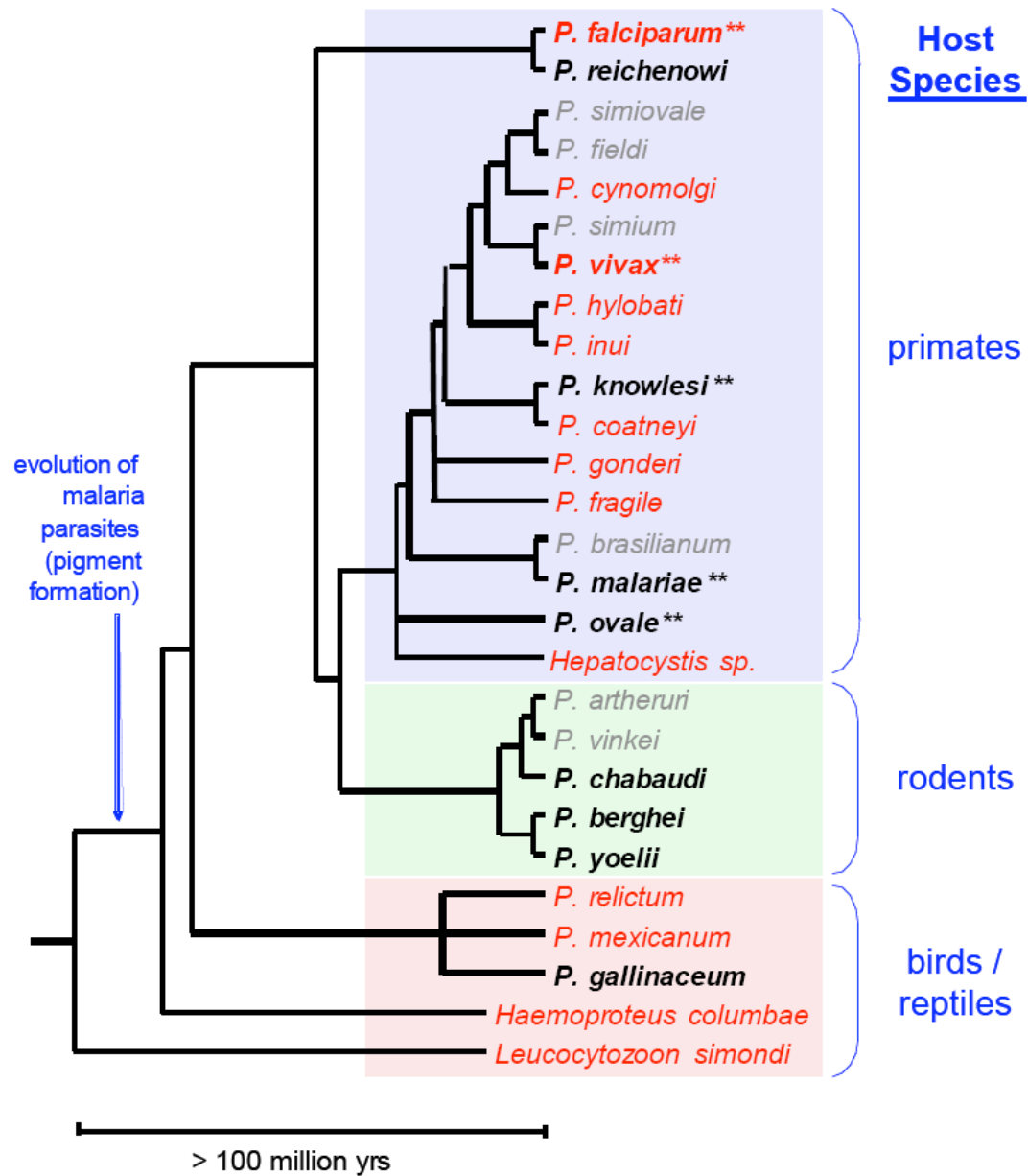
cholinergic receptor,
nicotinic, beta 2
(neuronal)



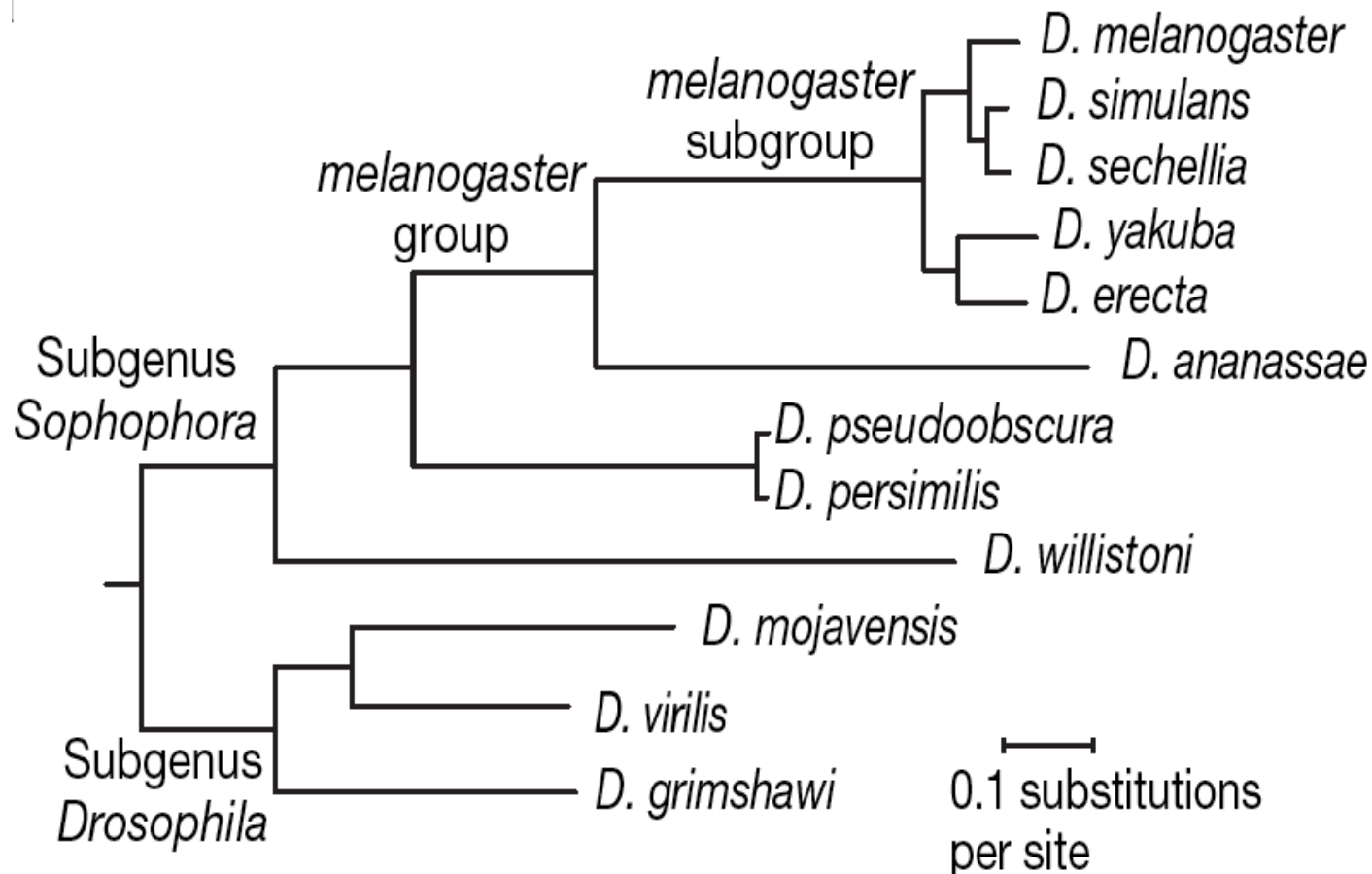
2009: 29 of the 44 vertebrates sequenced are eutherian mammals



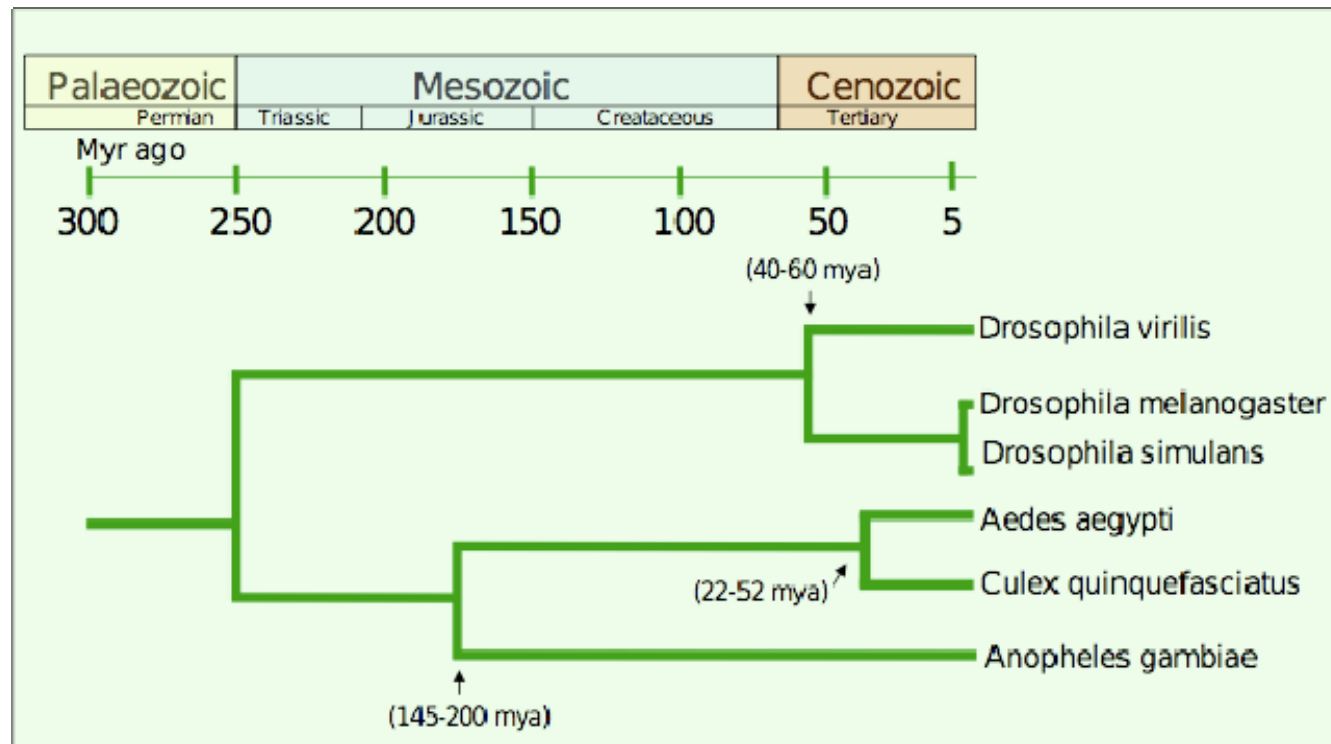
Plasmodium genomes



12 fly genomes



Mosquito genomes



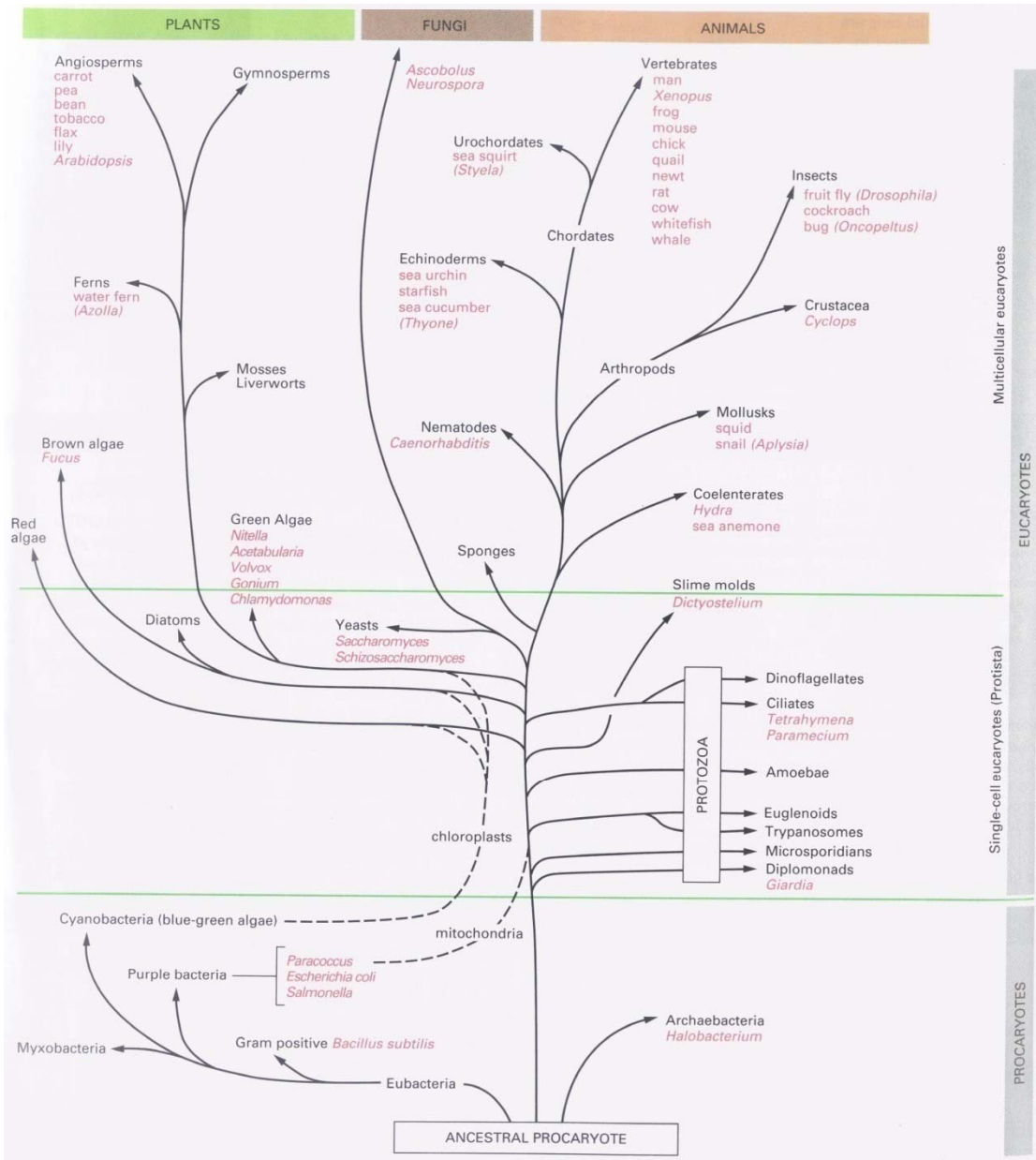
Anthony James Lab at UCI

Sieglaff et al., PNAS, 200

Evolution

- Related organisms have similar DNA
 - Similarity in sequences of proteins
 - Similarity in organization of genes along the chromosomes
- Evolution plays a major role in biology
 - Many mechanisms are shared across a wide range of organisms
 - During the course of evolution existing components are adapted for new functions

The Tree of Life



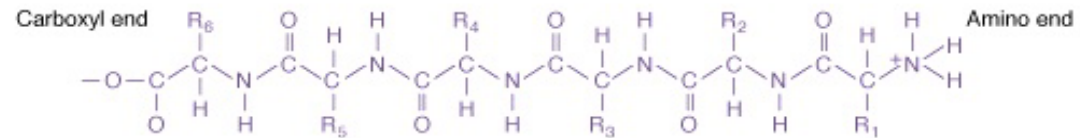
Source: Alberts et al

Protein Structure Prediction

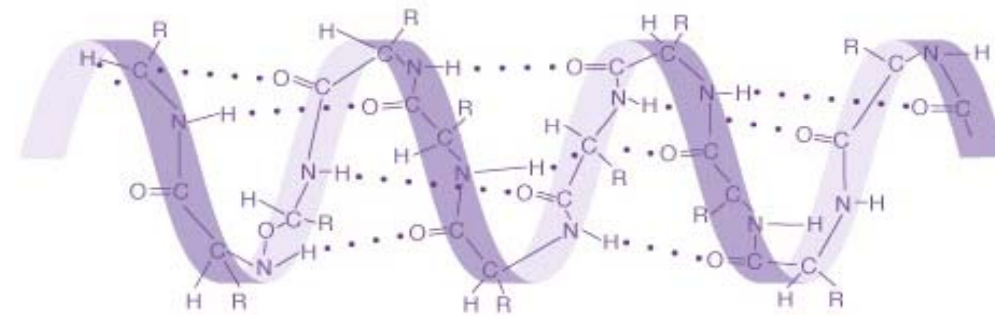
Protein Structure

- Proteins are poly-peptides of 70-3000 amino-acids
- This structure is (mostly) determined by the sequence of amino-acids that make up the protein

(a) Primary structure

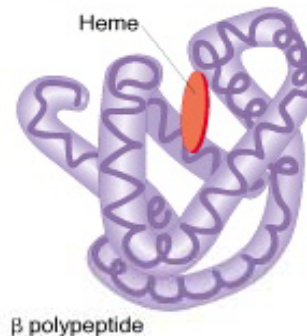


(b) Secondary structure

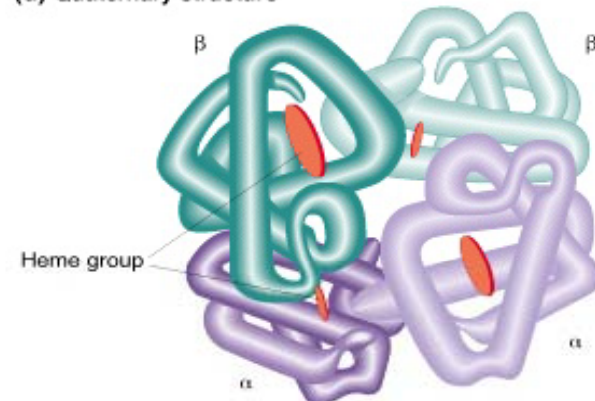


Hydrogen bonds between amino acids at different locations in polypeptide chain

(c) Tertiary structure

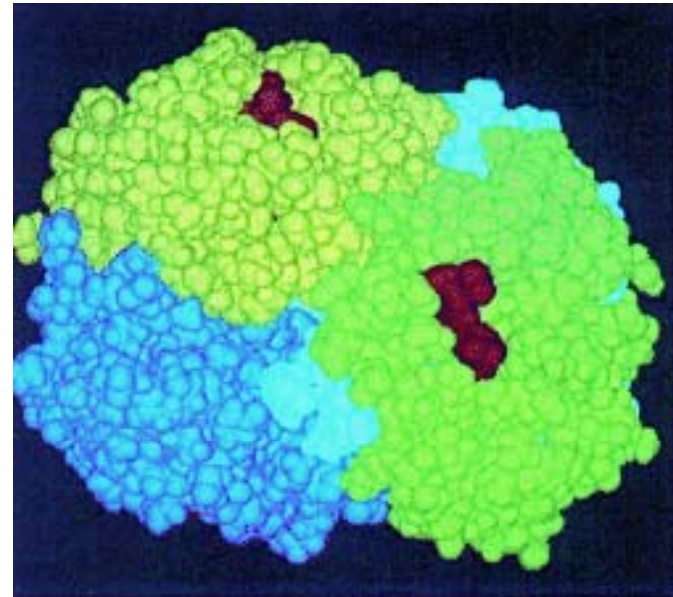


(d) Quaternary structure



Hemoglobin

- protein built from 4 polypeptides
- responsible for carrying oxygen in red blood cells



Protein structures

Why do we need protein structures?

Protein structure is the key to understand:

- Sequence-function relationships
- Evolution (a history of changes at the level of sequence and function)

Based on the knowledge of protein structure we can attempt to:

- Design new drugs
- Design proteins with new functions (e.g. industrial enzymes)

Protein structure determination / prediction



Protein model accuracy vs time required for its construction

Experimental protein structure solution provides high resolution models,
BUT:

...it is costly and time-consuming,
...fails for many proteins

Drug design requires hyper-accurate protein structures.

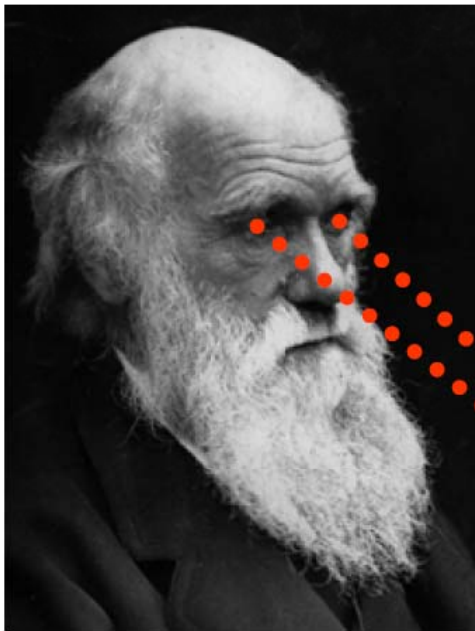
NONETHELESS:

to understand the protein function or to engineer proteins by genetic methods it is often sufficient to know only an approximate structure (at the level of amino acid residues, not necessarily all atoms)

Theoretical modeling can quickly provide approximate structures

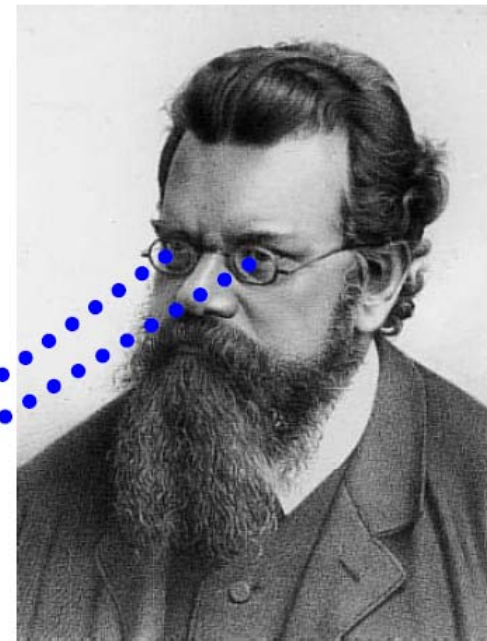
TWO DIFFERENT PERSPECTIVES ON PROTEIN STRUCTURE PREDICTION

EVOLUTIONARY BIOLOGY

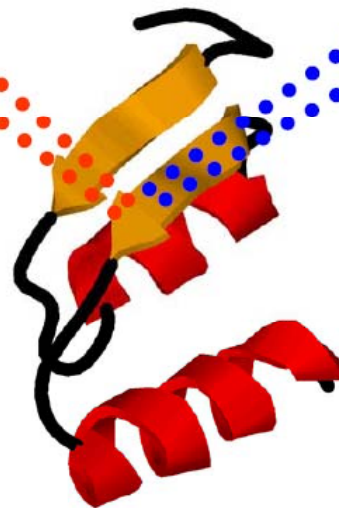


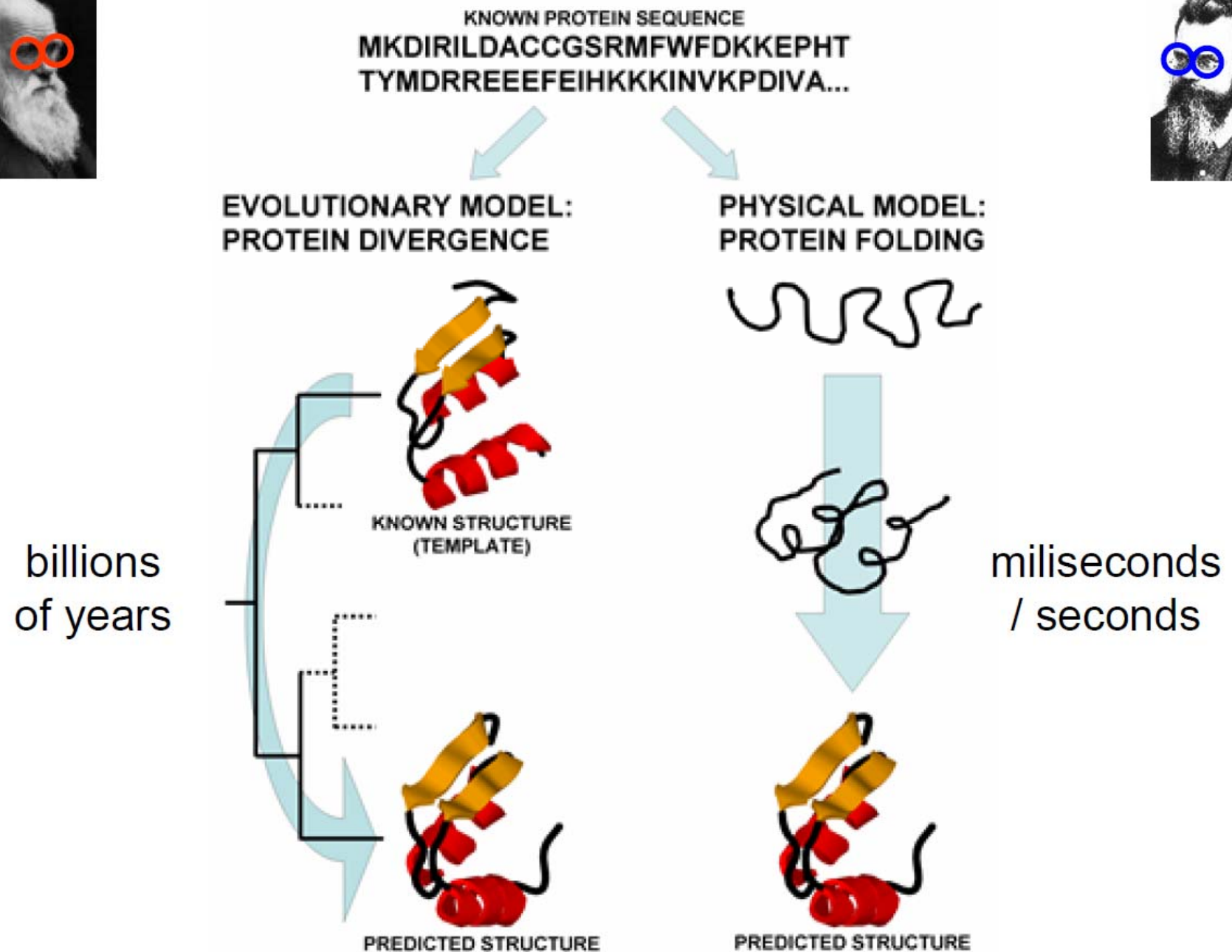
Charles
Darwin
(1809-1882)

STATISTICAL THERMODYNAMICS

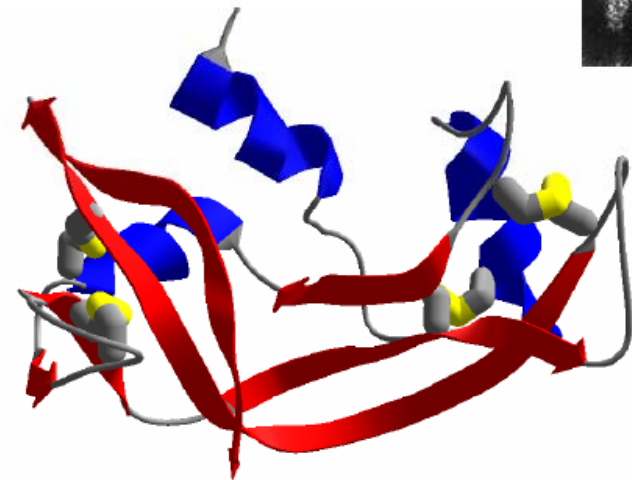
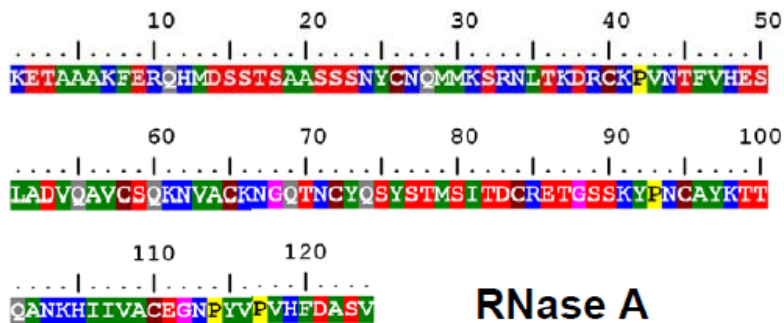


Ludwig Edward
Boltzmann
(1844-1906)

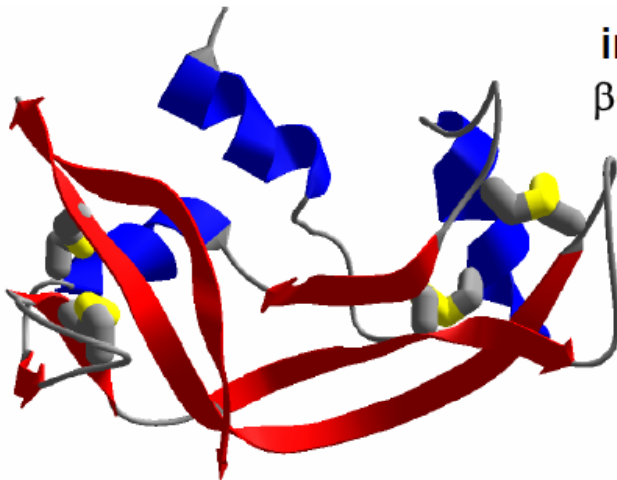




Protein refolding:



renaturation
in the presence of
 β -mercaptoethanol



Anfinsen et al, *PNAS* 47, 1309 (1961)

Physics of protein folding

... Anfinsen's thermodynamic hypothesis:

"The three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, etc.) is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment."



Christian B.
Anfinsen
(1916-1995)

Structure prediction based on physics



1. Assume Anfinsen's hypothesis is correct (native protein structure is thermodynamically stable and located at the global free energy minimum)
2. **SAMPLING:** Within a reasonable time, effectively sample the conformational space available to the polypeptide chain of the target protein to generate a large number of „decoys“, of which at least one should be sufficiently similar to the native structure to be in the same energy minimum.
3. **ENERGY:** Use an accurate energy function to rank the decoys and identify the native structure as one of the lowest energy.

WHAT IS POSSIBLE

So, is it possible to predict the protein structure based solely on the principles of physics?

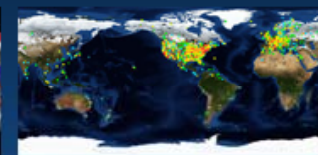
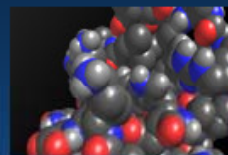
Not yet. But hope remains... ;-)

It is not yet possible to sample all conformations within reasonable time to guarantee that one of them will be sufficiently similar to the native structure. It is also not yet possible to guarantee that the native structure (or the corresponding closest decoy) would have the lowest energy.



Folding@home

distributed computing



[Home](#)
[Download](#)
[Guides](#)
[FAQ](#)
[Stats](#)
[Science](#)
[Results](#)
[Awards](#)
[About Us](#)

[Main](#)
[News](#)
[Forum](#)
[\(Help\)](#)
[Donate](#)

العربية

中文

Dansk

Deutsch

English

Español

Français

Italiano

Ελληνικά

日本語

한국말

Lietuvių

Magyar

Nederlands

Norsk

Occitan

فارسی

Polski

Português

Русский

Suomeksi

Svenska

Turkish

தமிழ்

Our goal: to understand protein folding, misfolding, and related diseases

What is protein folding and how is folding linked to disease?

Proteins are biology's workhorses -- its "nanomachines." Before proteins can carry out these important functions, they assemble themselves, or "fold." The process of protein folding, while critical and fundamental to virtually all of biology, in many ways remains a mystery.

Moreover, when proteins do not fold correctly (i.e. "misfold"), there can be serious consequences, including many well known [diseases](#), such as Alzheimer's, Mad Cow (BSE), CJD, ALS, Huntington's, Parkinson's disease, and many Cancers and cancer-related syndromes.

You can help by simply running a piece of software.

Folding@home is a distributed computing project -- people from throughout the world [download](#) and run software to band together to make one of the largest supercomputers in the world. Every computer takes the project closer to our goals. Folding@home uses novel computational methods coupled to distributed computing, to simulate problems millions of times more challenging than previously achieved.

What have we done so far?

We have had several successes. You can read about them on our [Science](#) page, on our [Awards](#) page, or go directly to our [Results](#) page.

Want to learn more?

Click on the links on the left for downloads or more information. You can also download our [Executive Summary](#), which is a PDF suitable for distribution. Also, you can learn more by watching recent seminars ([Stanford BMI](#) ; [Xerox PARC](#)). One can also help by donating funds to [the project](#), via Stanford University.

[Download Folding@home](#)



Make a gift now! ♦

[Click here to make a gift >](#)

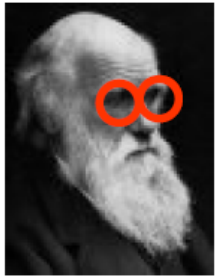


THE EVOLUTIONARY SCHOOL OF PROTEIN STRUCTURE PREDICTION

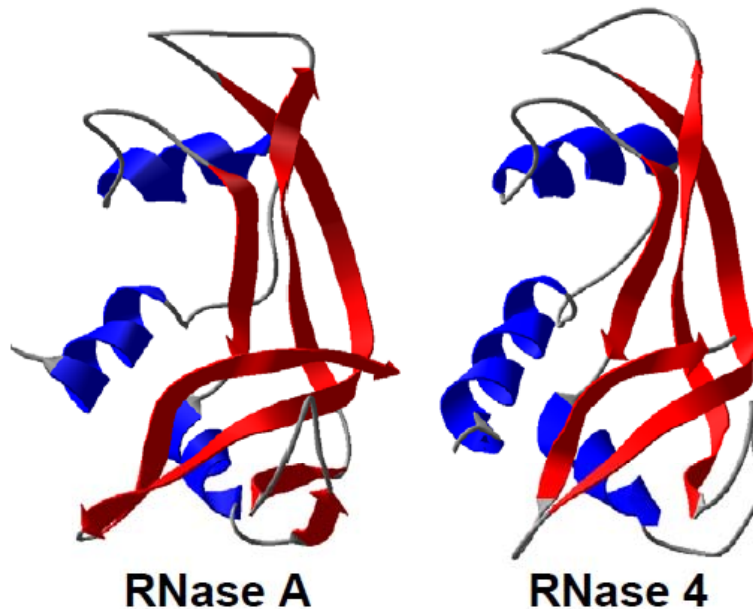


nothing in biology
makes sense
except in the light of
evolution

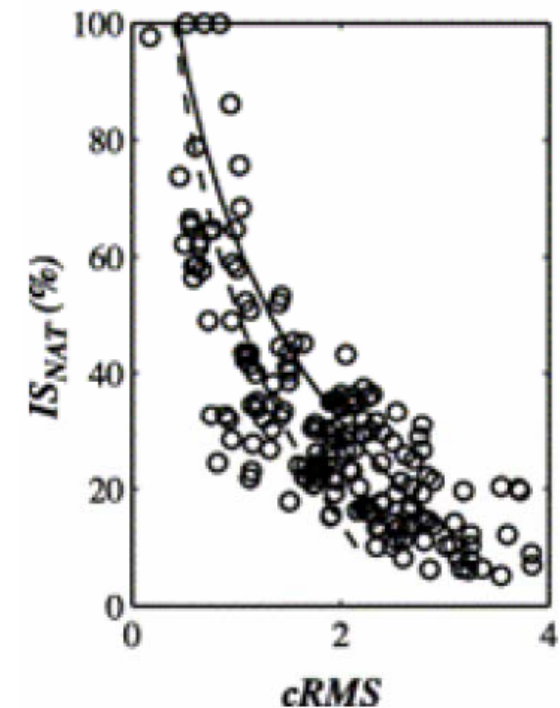
Theodosius
Dobzhansky
(1900-1975)



Evolution of proteins: sequence-structure-function



RNase A KETAAAKFEROHMDSSTSAASSNYCNOMMKSRNLTKDRCKPVTTFVHESLADVQAVC
 RNase 4 QDGMYPRLRQHVHPPEE-TGGSDRYCNDMMORRKMTLYHCKRENTETIHEDIWNIRSIC
 SQKNVACKNGQTNCYQSYSTMSTIDCRETGSSKYPNCAYKTTQANKHITVACEGNPPYVPVHFDASV
 STTNIQCKNGKMNCHEG--VVKVTDGCDTGSSRAPNCRVRAIASTRRVVIACEGNPPQVPVHFDG--



Evolutionarily related proteins retain very similar tertiary structure despite the accumulation of amino acid substitutions, deletions and insertions

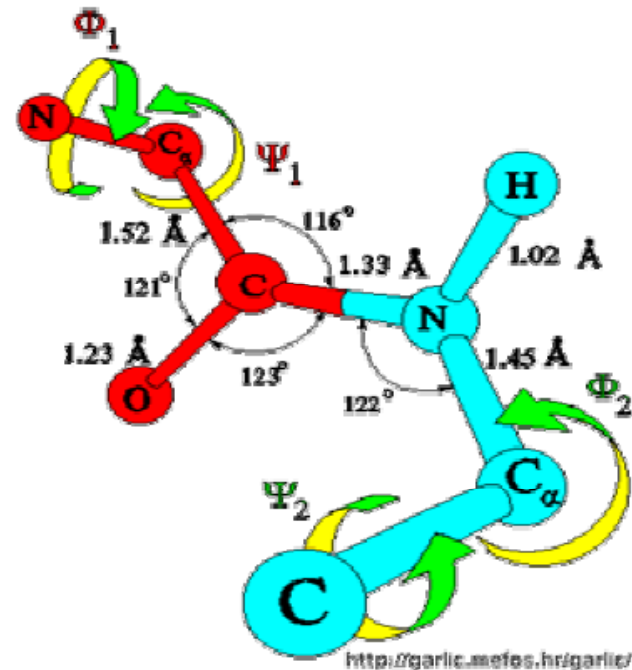
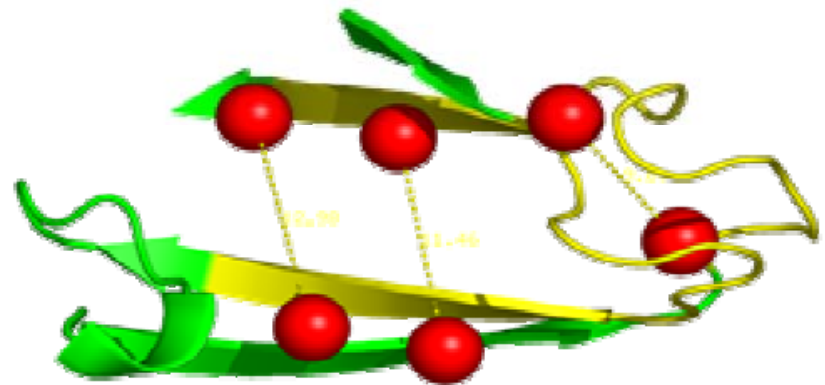


Structure prediction based on evolution

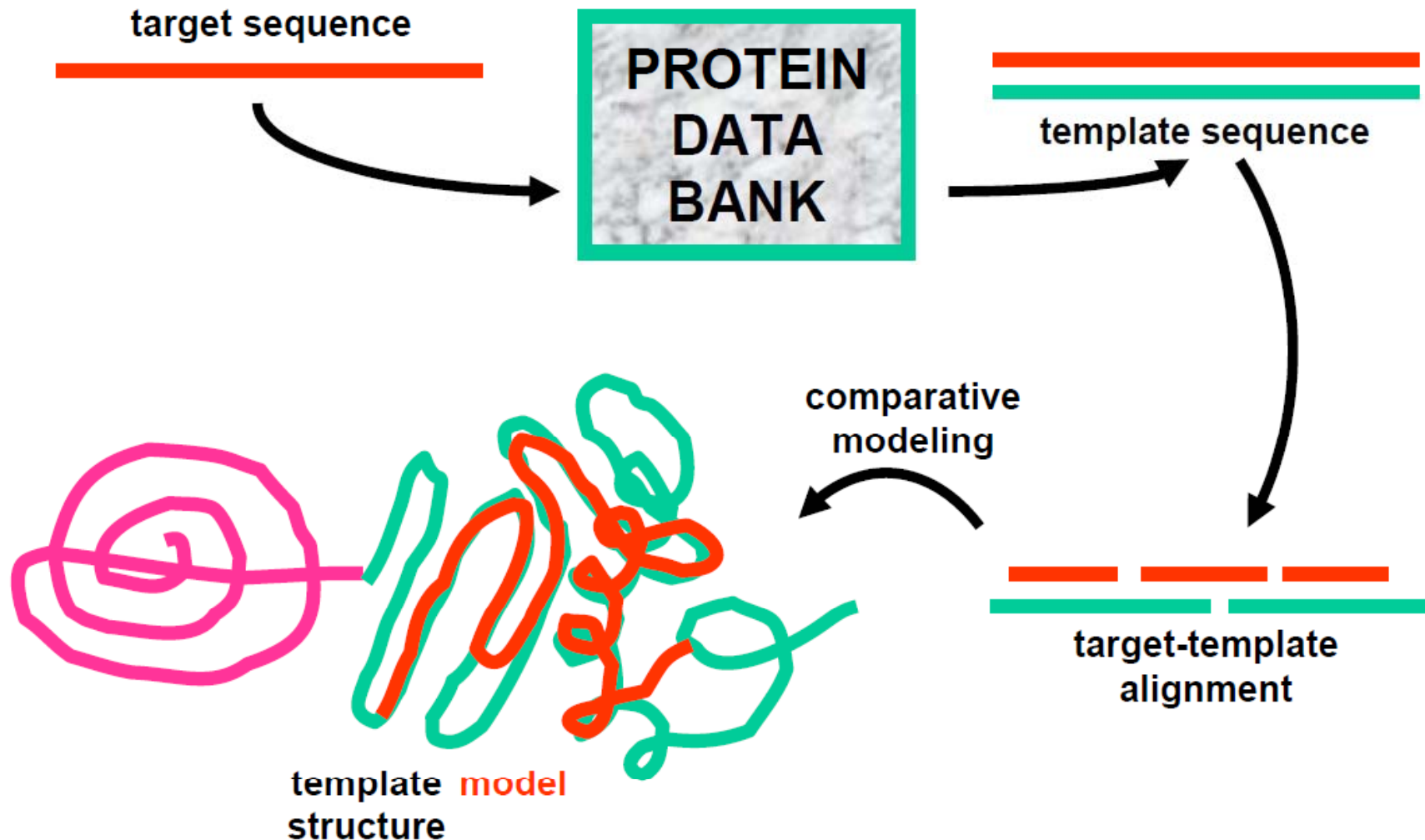
1. Assume that extant proteins evolved from common ancestors
2. Assume that evolutionarily related (homologous) proteins are structurally similar
3. Find a structure of a protein homologous to the target protein and model the evolutionary changes that occurred since these two proteins diverged from their common ancestor (**provided that an accurate model of divergent evolution of protein sequences and structures is available**)

Template-based modeling

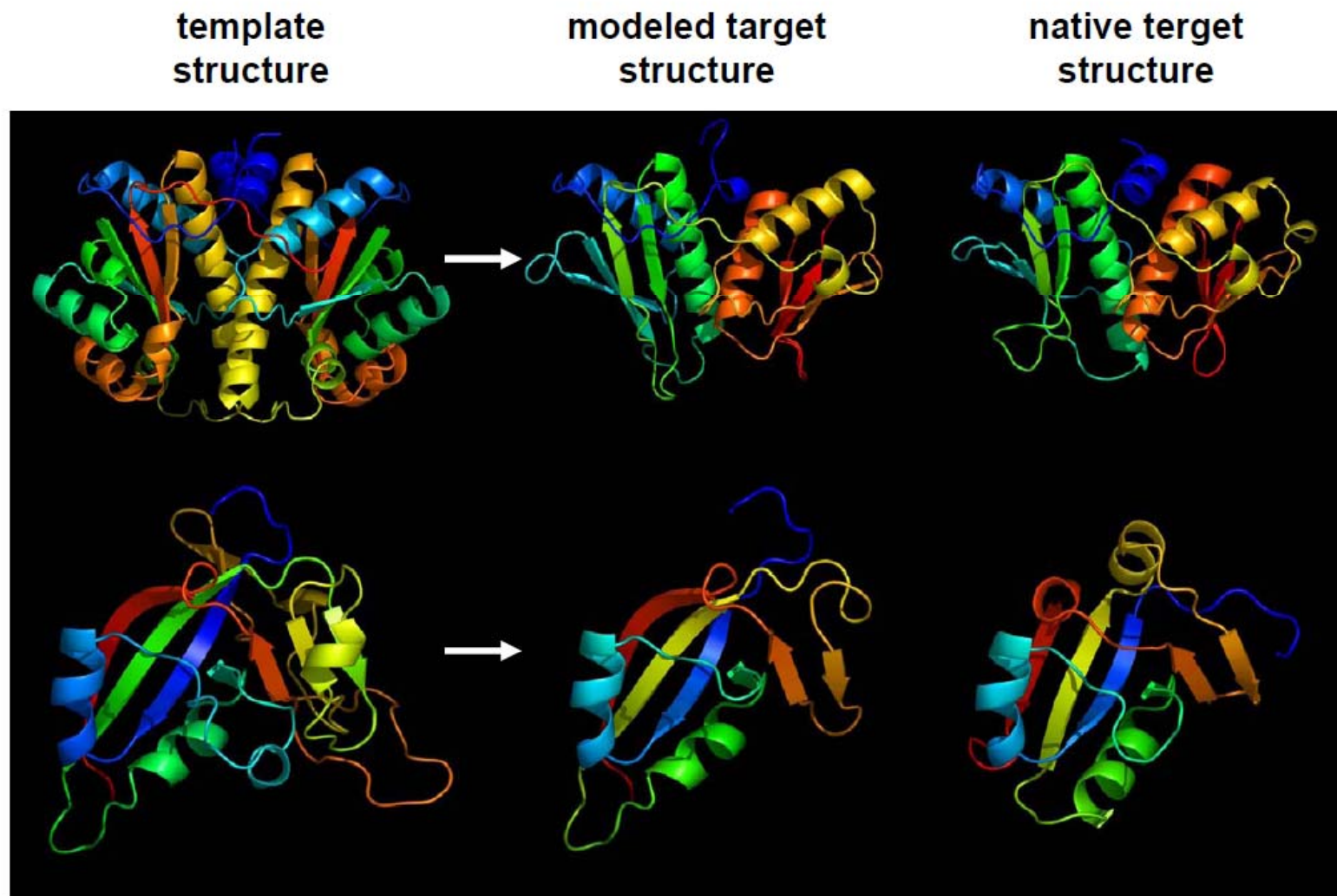
1. For all residues in the target protein assign restraints derived from the template structure:
 - bond lengths, angles
 - interatomic distances
2. Minimization of the target function usually comprising:
 - a term describing the agreement of the model with the restraints
 - a term describing the ideal values of various structural parameters



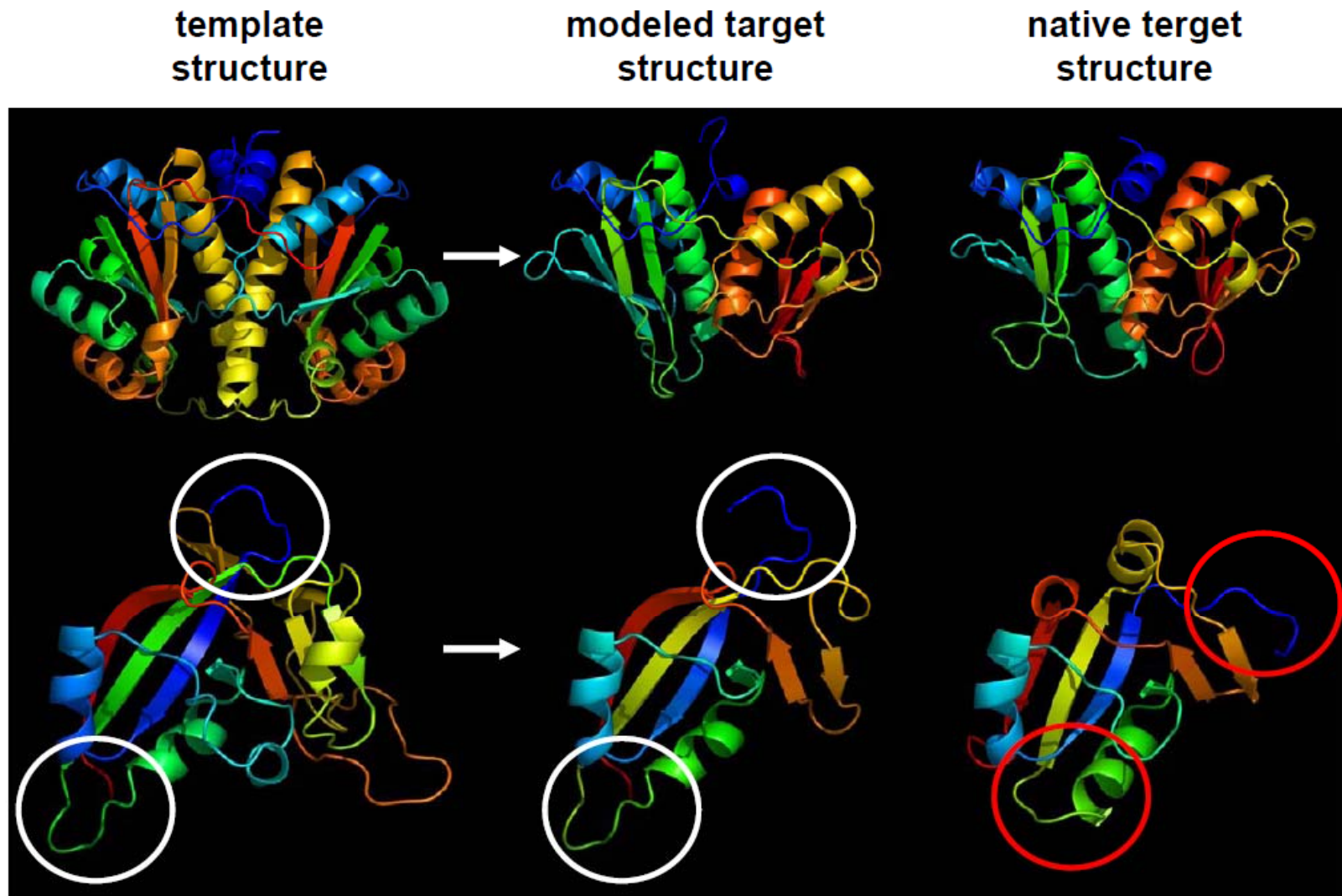
And get a model derived from the template (with some modifications)



Template-based models can be quite good



Template-based models can be quite good ...but devil is in the details



SO, WHAT IS POSSIBLE?

Is it possible to predict the protein structure based solely on the principles of physics?

Not yet. But hope remains... ;-)

It is not yet possible to sample all conformations within reasonable time to guarantee that one of them will be sufficiently similar to the native structure.

It is also not yet possible to guarantee that the native structure (or the corresponding closest decoy) would have the lowest energy.

Is it possible to predict the protein structure based solely on the principles of evolution?

Yes...But only if 1) there is a homologous protein with known structure, 2) if we can correctly identify it among all proteins with known structures and 3) if we can approximate the evolutionary changes at the level of sequences (alignment) and structures (3D coordinates).

Protein data bank

[File](#) [Edit](#) [View](#) [History](#) [Bookmarks](#) [Tools](#) [Help](#)

[PDB](#) [http://www.rcsb.org/pdb/home/home.do](#) [Google](#)

[Most Visited](#) [Getting Started](#) [Latest Headlines](#) [PubMed](#) [ICS](#) [Yahoo](#) [New York Times](#) [UCSC](#) [Java](#)

RCSB
PDB
PROTEIN DATA BANK

An Information Portal to Biological Macromolecular Structures
As of **Tuesday Jun 02, 2009** there are 57944 Structures | [PDB Statistics](#)

[MyPDB Login](#) A MEMBER OF THE **PDB**

[WHAT'S NEW](#) | [CONTACT](#) | [FEEDBACK](#) | [HELP](#) | [PRINT](#)

[Search](#) [Adv. Search](#)

[Home](#) [Search](#)

- Home
- Getting Started
- Structural Genomics
- BioSync
- Electron Microscopy
- Download Files
- Deposit and Validate
- Dictionaries & File Formats
- Software Tools
- General Education
- General Information
- Acknowledgements
- Frequently Asked Questions

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the [wwPDB](#), the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Molecule of the Month: Vaults



Our cells are filled with compartments, each performing a specific function. Some of these compartments, such as mitochondria and lysosomes, are very large and enclose many different molecular machines. Other intracellular compartments are smaller, such as the transport vesicles that shuttle proteins from site to site inside the cell. Most of these compartments, including mitochondria, lysosomes and transport vesicles, are surrounded by membranes. However, in special cases, cells build smaller compartments surrounded by a protein shell. In our own cells, vaults are a spectacular example of these protein-enclosed compartments.

■ [Read more ...](#) ■ [Previous Features](#)

PSI Featured Molecule: Hda and DNA Replication



When cells divide, they need to ensure that each daughter cell gets one copy of each chromosome. Bacteria contain one big circle of DNA, so they start replication in one place, then copy the DNA both ways around until it finishes on the other side. PSI Researchers have solved the first atomic structure at how bacteria use the Hda protein to initiate replication at this origin only once for each generation of the cell.

■ [Read more from PSI SGKB](#) ■ [Previous Features](#)

New user? Try the browser [compatibility check](#), information on [Getting Started](#), and see [What's New](#) on the site.

News

- Complete News
- Newsletter
- Discussion Forum
- Job Listings

02-June-2009

Literature View: Looking at Structures in PubMedCentral



The recently-released Literature View aims to provide a broad look at how a given structure has been analyzed and presented in open access publications, that is, where the full text of the article is available without copyright restriction. The overall intent is to make the PDB user more aware of publications associated with the structure under study.

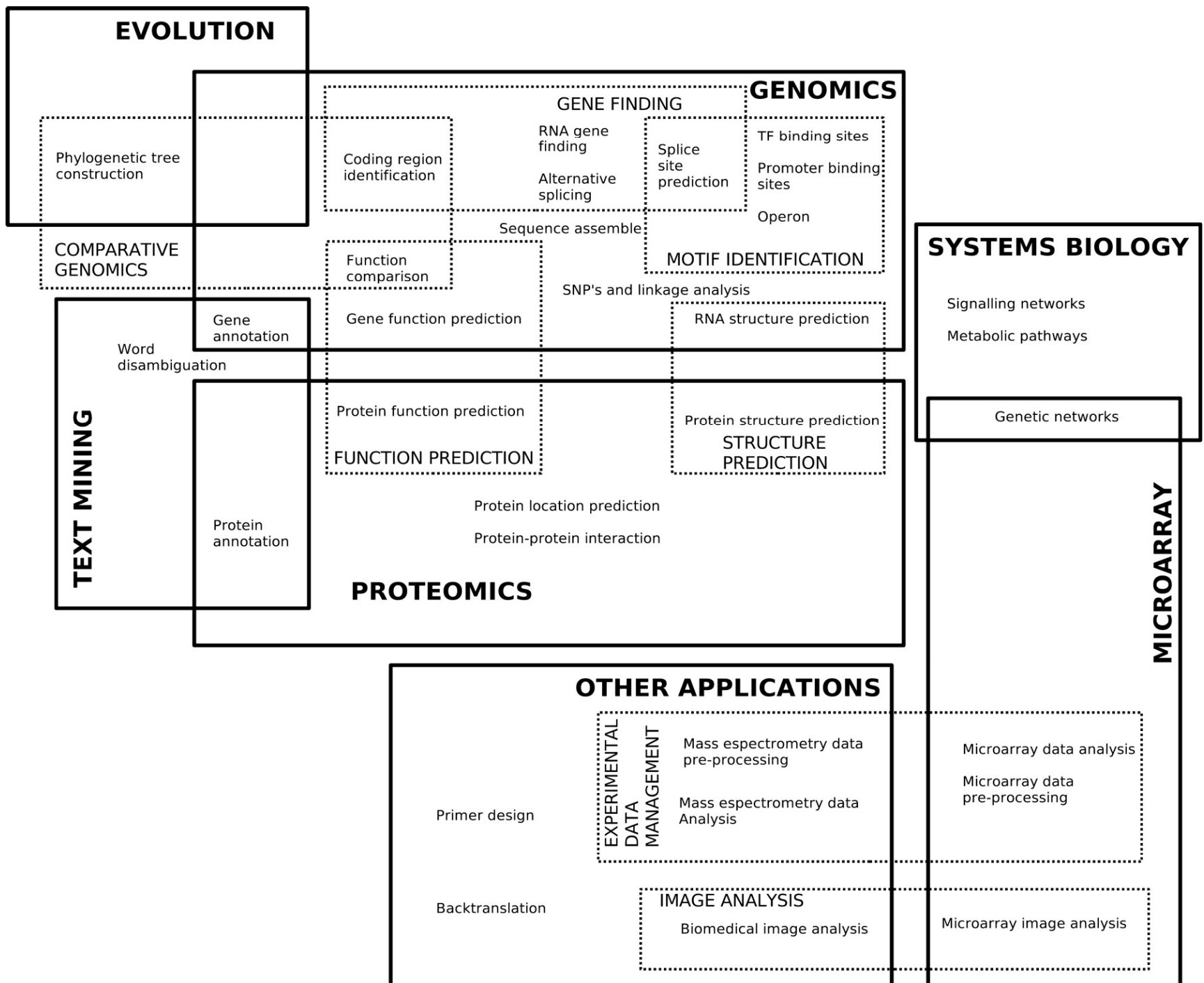
[More >>](#)

Data Snapshots

Time-stamped yearly snapshots of the PDB archive are available via FTP at: <ftp://snapshots.wwpdb.org>

The snapshots provide readily identifiable data sets for research on the PDB

Machine Learning in Bioinformatics



Online Resources

Protein Data Bank

Web browser interface showing the Protein Data Bank (PDB) homepage as of Tuesday, June 02, 2009. The address bar displays <http://www.rcsb.org/pdb/home/home.do>.

The page header includes the PDB logo (RCSB PROTEIN DATA BANK) and the text "An Information Portal to Biological Macromolecular Structures". It also states "As of Tuesday Jun 02, 2009 there are 57944 Structures" and provides a link to "PDB Statistics".

The navigation bar includes links for "WHAT'S NEW", "CONTACT", "FEEDBACK", "HELP", and "PRINT". A search bar is present with the placeholder text "PDB ID or keyword" and buttons for "Search" and "Adv. Search".

The left sidebar contains a "Home" tab and a "Search" tab. The "Home" tab is active, showing a list of links: Home, Getting Started, Structural Genomics, BioSync, Electron Microscopy, Download Files, Deposit and Validate, Dictionaries & File Formats, Software Tools, General Education, General Information, Acknowledgements, and Frequently Asked Questions.

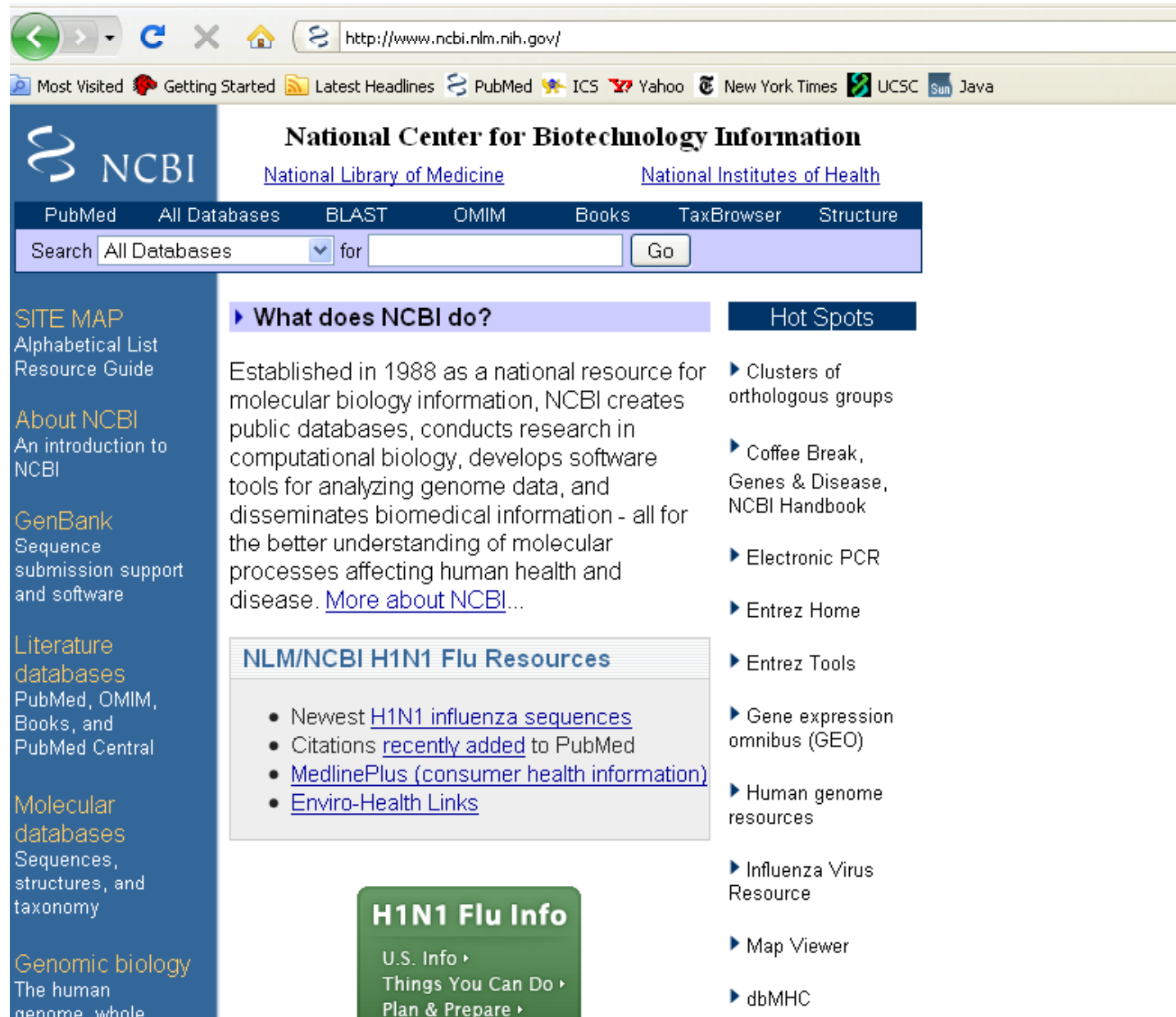
The main content area features the section "A Resource for Studying Biological Macromolecules". It describes the PDB archive as containing information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. It also mentions that the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The section "The RCSB PDB also provides a variety of tools and resources" explains that users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Below this text is the "Molecule of the Month: Vaults" section. It features a 3D visualization of a vault structure (a blue, barrel-like protein complex). The text describes vaults as protein-enclosed compartments found in cells, performing specific functions. It notes that vaults are surrounded by membranes and are a spectacular example of protein-enclosed compartments. Links for "Read more ..." and "Previous Features" are provided.

The right sidebar contains a "News" section. It lists links for "Complete News", "Newsletter", "Discussion Forum", and "Job Listings". Below this, it highlights a "Literature View: Looking at Structures in PubMedCentral" dated 02-June-2009. It includes a small image of a Kappa-alpha plot and text describing the plot's purpose in rapid search of protein structures. The text states: "The recently-released Literature View aims to provide a broad look at how a given structure has been analyzed and presented in open access publications, that is, where the full text of the article is available without copyright restriction. The overall intent is to make

NCBI



The image is a screenshot of the NCBI (National Center for Biotechnology Information) website homepage. At the top, there is a browser window showing the URL <http://www.ncbi.nlm.nih.gov/>. Below the browser window, there is a navigation bar with links to 'Most Visited', 'Getting Started', 'Latest Headlines', 'PubMed', 'ICS', 'Yahoo', 'New York Times', 'UCSC', and 'Java'. The main header features the NCBI logo and the text 'National Center for Biotechnology Information', with links to the 'National Library of Medicine' and 'National Institutes of Health'. Below the header, there is a search bar with a dropdown menu set to 'All Databases' and a 'Go' button. The left sidebar contains a 'SITE MAP' with links to 'Alphabetical List', 'Resource Guide', 'About NCBI', 'GenBank', 'Literature databases', 'Molecular databases', and 'Genomic biology'. The main content area has a section titled 'What does NCBI do?' which describes the center's mission and provides a link to 'More about NCBI...'. Below this is a section titled 'NLM/NCBI H1N1 Flu Resources' with a list of links: 'Newest H1N1 influenza sequences', 'Citations recently added to PubMed', 'MedlinePlus (consumer health information)', and 'Enviro-Health Links'. To the right of the main content area is a 'Hot Spots' section with links to 'Clusters of orthologous groups', 'Coffee Break, Genes & Disease, NCBI Handbook', 'Electronic PCR', 'Entrez Home', 'Entrez Tools', 'Gene expression omnibus (GEO)', 'Human genome resources', 'Influenza Virus Resource', 'Map Viewer', and 'dbMHC'. At the bottom center, there is a green box titled 'H1N1 Flu Info' with links to 'U.S. Info', 'Things You Can Do', and 'Plan & Prepare'.

NCBI

National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for Go

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic biology
The human genome, whole

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More about NCBI...](#)

NLM/NCBI H1N1 Flu Resources

- Newest [H1N1 influenza sequences](#)
- Citations [recently added](#) to PubMed
- [MedlinePlus \(consumer health information\)](#)
- [Enviro-Health Links](#)

H1N1 Flu Info

U.S. Info ›
Things You Can Do ›
Plan & Prepare ›

Hot Spots

- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR
- ▶ Entrez Home
- ▶ Entrez Tools
- ▶ Gene expression omnibus (GEO)
- ▶ Human genome resources
- ▶ Influenza Virus Resource
- ▶ Map Viewer
- ▶ dbMHC

PubMed

Web browser address bar: <http://www.ncbi.nlm.nih.gov/sites/entrez>

Navigation bar: Most Visited, Getting Started, Latest Headlines, PubMed, ICS, Yahoo, New York Times, UCSC, Java

NCBI Logo:  **NCBI**

PubMed Logo:  **PubMed**
A service of the [U.S. National Library of Medicine](#) and the [National Institutes of Health](#)
www.pubmed.gov

My NCBI 
[\[Sign In\]](#) [\[Register\]](#)

Database tabs: All Databases, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, Books

Search: PubMed for hidden Markov model [Advanced Search](#) [Save Search](#)

Buttons: Limits, Preview/Index, History, Clipboard, Details

Display: Summary Show 20 Sort By Send to

All: 985 Review: 30 

Items 1 - 20 of 985 1 of 50 [Next](#)

☐ 1: [Fast statistical alignment.](#)
Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L.
PLoS Comput Biol. 2009 May;5(5):e1000392. Epub 2009 May 29.
PMID: 19478997 [PubMed - in process]
[Related Articles](#) [Free article at journal site](#)

☐ 2: [Identifying novel constrained elements by exploiting biased substitution patterns.](#)
Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X.
Bioinformatics. 2009 Jun 15;25(12):i54-62.
PMID: 19478016 [PubMed - in process]
[Related Articles](#) [Free article at journal site](#)

☐ 3: [Model-based clustering of array CGH data.](#)
Shah SP, Cheung KJ Jr, Johnson NA, Alain G, Gascoyne RD, Horsman DE, Ng RT, Murphy KP.
Bioinformatics. 2009 Jun 15;25(12):i30-8.
PMID: 19478003 [PubMed - in process]
[Related Articles](#) [Free article at journal site](#)

☐ 4: [Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data.](#)
Yin J, Jordan MI, Song YS.

Also try:
▶ **hidden markov model** review
▶ profile **hidden markov model**

Recent Activity 
[Turn Off](#) [Clear](#)

 **hidden Markov model** (985) **PubMed**

 Machine learning in bioinformatics.

Gene Expression Omnibus

The screenshot shows the NCBI GEO Profiles website. At the top, there's a browser address bar with the URL <http://www.ncbi.nlm.nih.gov/sites/entrez>. Below the address bar is a navigation bar with links to 'Most Visited', 'Getting Started', 'Latest Headlines', 'PubMed', 'ICS', 'Yahoo', 'New York Times', 'UCSC', and 'Java'. The main header features the NCBI logo on the left and the 'GEO PROFILES Gene Expression Omnibus' logo on the right. Below the header is a search bar with a dropdown menu set to 'GEO Profiles' and a search input field. To the right of the search bar are 'Go' and 'Clear' buttons. Below the search bar is a row of buttons: 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The left sidebar contains a list of links: 'The GEO site', 'GEO FAQ', 'List GEO Contents', 'Graph caption', 'Entrez Help | FAQ'. The main content area has two sections: 'GEO Profiles' and 'GEO DataSets'. The 'GEO Profiles' section contains a paragraph: 'This database stores individual gene expression and molecular abundance profiles assembled from the [Gene Expression Omnibus \(GEO\)](#) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics. GEO Profiles facilitates powerful searching and linking to additional information sources.' The 'GEO DataSets' section contains a paragraph: 'This database stores curated gene expression and molecular abundance DataSets assembled from the [Gene Expression Omnibus \(GEO\) repository](#). Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries. GEO DataSets can be searched [here](#).'

NCBI

GEO PROFILES
Gene Expression Omnibus

All Databases PubMed Nucleotide Protein Genome Structure

Search GEO Profiles for Go Clear

Limits Preview/Index History Clipboard Details

GEO Profiles

This database stores individual gene expression and molecular abundance profiles assembled from the [Gene Expression Omnibus \(GEO\)](#) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics. GEO Profiles facilitates powerful searching and linking to additional information sources.

GEO DataSets

This database stores curated gene expression and molecular abundance DataSets assembled from the [Gene Expression Omnibus \(GEO\) repository](#). Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries. GEO DataSets can be searched [here](#).

The GEO site
GEO FAQ
List GEO Contents
Graph caption
Entrez Help | FAQ

BLAST

BLAST *Basic Local Alignment Search Tool*

Home Recent Results Saved Strategies Help

NCBI/BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

New Designing or Testing PCR Primers? Try your search in **Primer-BLAST**.

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

<input type="checkbox"/> Human	<input type="checkbox"/> Oryza sativa	<input type="checkbox"/> Gallus gallus
<input type="checkbox"/> Mouse	<input type="checkbox"/> Bos taurus	<input type="checkbox"/> Pan troglodytes
<input type="checkbox"/> Rat	<input type="checkbox"/> Danio rerio	<input type="checkbox"/> Microbes
<input type="checkbox"/> Arabidopsis thaliana	<input type="checkbox"/> Drosophila melanogaster	<input type="checkbox"/> Apis mellifera

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a nucleotide database using a nucleotide query <i>Algorithms: blastn, megablast, discontinuous megablast</i>
protein blast	Search protein database using a protein query <i>Algorithms: blastp, psi-blast, phi-blast</i>
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query

UCSC Genome Browser

http://genome.ucsc.edu/cgi-bin/hgGateway

blast

Most Visited

Getting Started

Latest Headlines

PubMed

ICS

Yahoo

New York Times

UCSC

Java

Home

Genomes

Blat

Tables

Gene Sorter

PCR

Session

FAQ

Help

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade

Mammal

genome

Human

assembly

Mar. 2006

position or search term

chr4:111,155,075-113,453,517

image width

800

submit

[Click here to reset](#) the browser user interface settings to their defaults.

add custom tracks

configure tracks and display

clear position

About the Human Mar. 2006 (hg18) assembly [\(sequences\)](#)

The March 2006 human reference sequence (NCBI Build 36.1) was produced by the International Human Genome Sequencing Consortium.

Sample position queries


A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, or a cytological band, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:

chr7

Genome Browser Response:

Displays all of chromosome 7



Ensembl Genome Browser



Search Ensembl

Search: for

e.g. human gene BRCA2 or rat X:100000..200000 or insulin

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))



Human
NCBI36



Mouse
NCBIM37



Zebrafish
Zv8

All genomes

[View full list of all Ensembl species](#)

New to Ensembl?

Did you know you can:

 [Learn how to use Ensembl](#)

with our video tutorials and walk-throughs

 [Add custom tracks](#)

using our new Control Panel

 [Upload your own data](#)

and save it to your Ensembl account

 [Search for a DNA or protein sequence](#)

using BLAST or BLAT

 [Fetch only the data you want](#)

from our public database, using the Ensembl Perl API

 [Download our databases via FTP](#)

in FASTA, MySQL and other formats

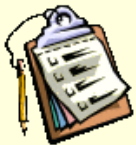
 [Mine Ensembl with BioMart](#)

and export sequences or tables in text, html, or Excel format

Still got questions? [Try our FAQs](#)

User Survey

Almost 6 months have passed since the release of the new website design. If you have a few minutes to spare, we would love to hear what you think of it:



- [Take the survey...](#)

What's New in Release 54 (5 May 2009)

Educational Opportunities at UCI

Biomedical Computing Major



BIOMEDICAL COMPUTING

Degrees available: B.S.

(open to freshmen in Fall 2009; open to transfer applicants in Fall 2011)

What is Biomedical Computing?

Biomedical Computing (BMC) is the intersection of computer science and information technology with biology and medicine. Students will receive a firm quantitative grounding in mathematics, statistics, and computation; become familiar with the basic foundations of physics, chemistry, and biology; master a rigorous and demanding Biomedical Computing year-long sequence; study theory, algorithms, data mining, and machine learning; and carry out a year-long immersive capstone Senior Project. The immense growth and impact of biomedical information has led to a critical need for people who can understand the languages, tools, and techniques of both life sciences and computational sciences. The Biomedical Computing program aims to create a new generation of professionals with these complementary cross-disciplinary skills.

Program Educational Objectives:

SUGGESTED CURRICULUM for the Biomedical Computing major

Freshman

Introduction to Computer Science I and II
Fundamental Data Structures
Single Variable Calculus
Multivariable Calculus
Biotech Basics²
From DNA to Organisms
From Organisms to Ecosystems
General Education¹ (one course)

Sophomore

Introduction to Computer Organization
Introduction to Software Engineering
Genetics
Introduction to Computational Biology
Discrete Mathematics
Boolean Algebra and Logic
Linear Algebra
Introduction to Probability and Statistics for Computer Science
Classical Physics (two courses)
Classical Physics Laboratory (two courses)
General Education (one course)

Junior

Representations and Algorithms for Molecular Biology
Probabilistic Modeling of Biological Data
Computational Systems Biology
Design and Analysis of Algorithms
Machine Learning and Data Mining
General Chemistry (two courses)
General Chemistry Laboratory (one course)
Biology Elective³
General Education (two courses)

Senior

Biomedical Computing Project (three courses)
Quantitative Elective⁴ (three courses)
Upper-division Writing
General Education (five courses)

BIT (Bioinformatics Training Program)



INSTITUTE FOR GENOMICS AND BIOINFORMATICS

University of California, Irvine

About IGB	▶
People	▶
Education	▶
Research	
Tools & Databases	▶
Facilities & Services	
Technology Transfer	
Events	▶
Sponsorship Programs	

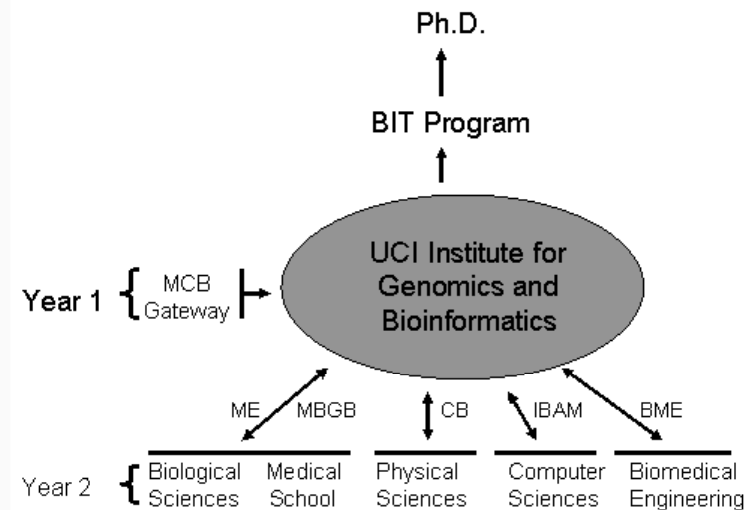
Predocutorial Program

[Predocutorial :: Postdoctoral](#)

Shortcuts: [Nomination Process](#) — [Mandatory Courses](#) — [Laboratory Rotations](#) — [Choosing a BIT Faculty Thesis Advisor](#) — [Continuing BIT Program Requirements](#).

Admission Procedures

Recruitment of first-year students directly into the BIT Program through the [MCSB](#) Gateway will be performed in a manner similar to recruitment of first-year students into other campus graduate programs.



The Biomedical Informatics Training (BIT) Program. [ME](#), Mol. Evolution track in Ecology and Evolution Dept.; [MBGB](#), Combined [Mol. Biol., Genet. and Biochem Program](#) in School of Biol. Sci.; [CB](#), Chemical Biol. Track in Chem Dept.; [IBAM](#), Informatics in Biology and Medicine track in ICS; [BME](#) Biomedical Engineering Program. Existing graduate programs and BIT cooperate to jointly admit first year students directly into the [MCSB](#) Gateway. Alternatively, second-year students can join the BIT Program at the beginning of their second year as in the past.

MCB program



UCIrvine | Graduate Program in **Mathematical, Computational and Systems Biology**

[About MCSB](#) [Admissions](#) [Research & Faculty](#) [Curriculum](#) [Participating Departments](#) [Contact MCSB](#) [Links](#)

Admissions & Application

[Overview & Application](#) ▼
[Application Checklist](#)
[Current Class](#)



[MCSB Home](#) > [Admissions and Applications](#) > [Overview & Application](#)

Admissions and Application:

Students are admitted via the MCB gateway program, as described below. After successful completion of the MCB year, students in good standing who have been accepted into the group of a participating thesis advisor are automatically admitted to the department Ph.D. program of that advisor.

Prerequisites:

The MCB curriculum is designed to teach students at the beginning of their graduate careers the necessary mathematical, computational, and biological knowledge for successful research at the interface between these disciplines. The needs of students with a variety of backgrounds can be met provided that they have had mathematical training comparable to a standard one-year university-level calculus course and a lower-division university course in elementary differential equations and linear algebra. Exceptional students not meeting these prerequisites may be admitted to the program on the condition that they fulfill these requirements during the first fall quarter of their graduate study or the summer preceding, and pass with a grade of B or better.

Admissions Information:

The MCB Executive Committee and Director act as the admissions committee. Following an initial screening of applications, the Executive Committee usually invites potential trainees for an interview at UCI. The Director makes final offers of admissions to applicants, based on the recommendations of the interviewing faculty and the Executive Committee.