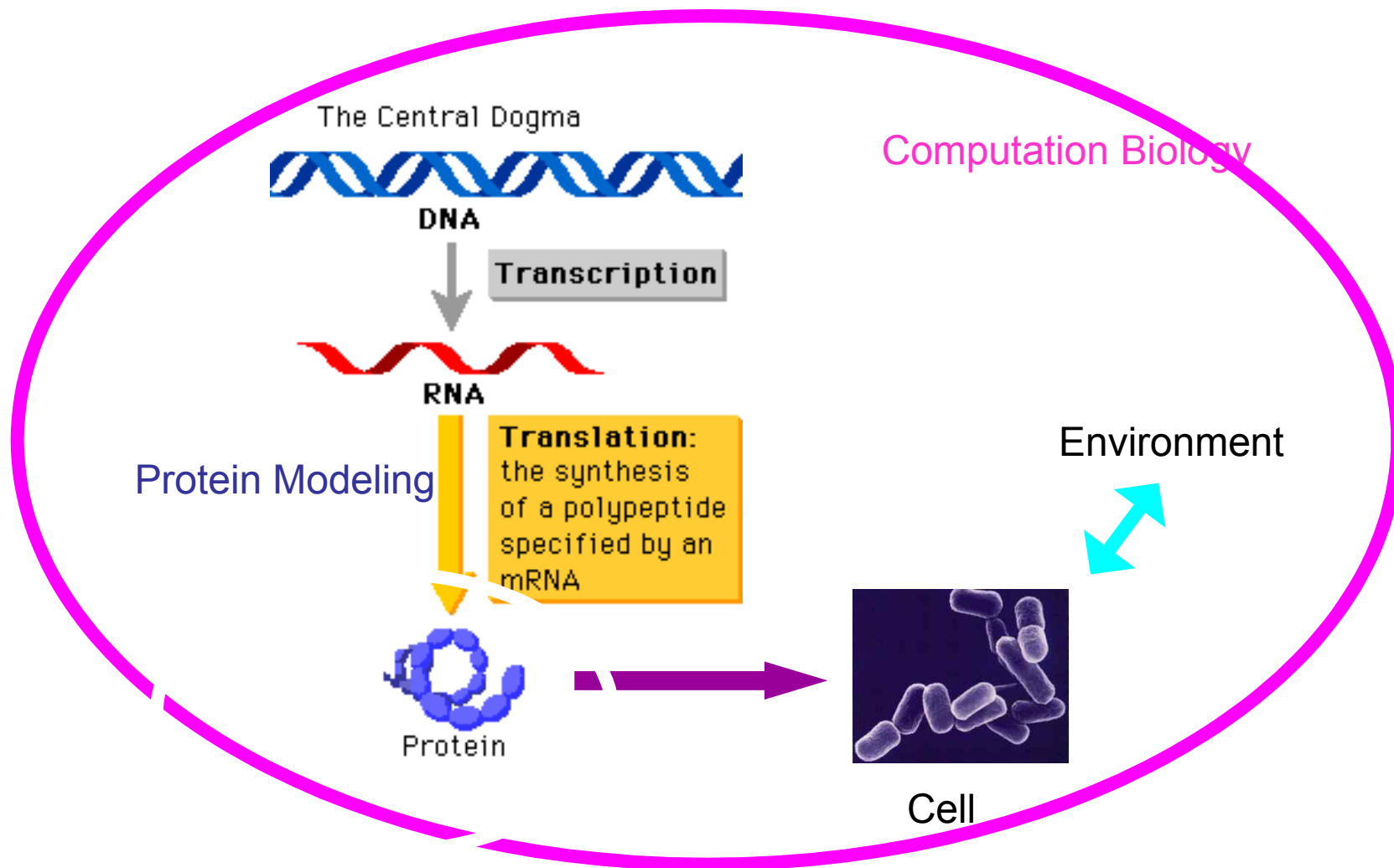


Enzyme Engineering

10. Softwares for protein engineering

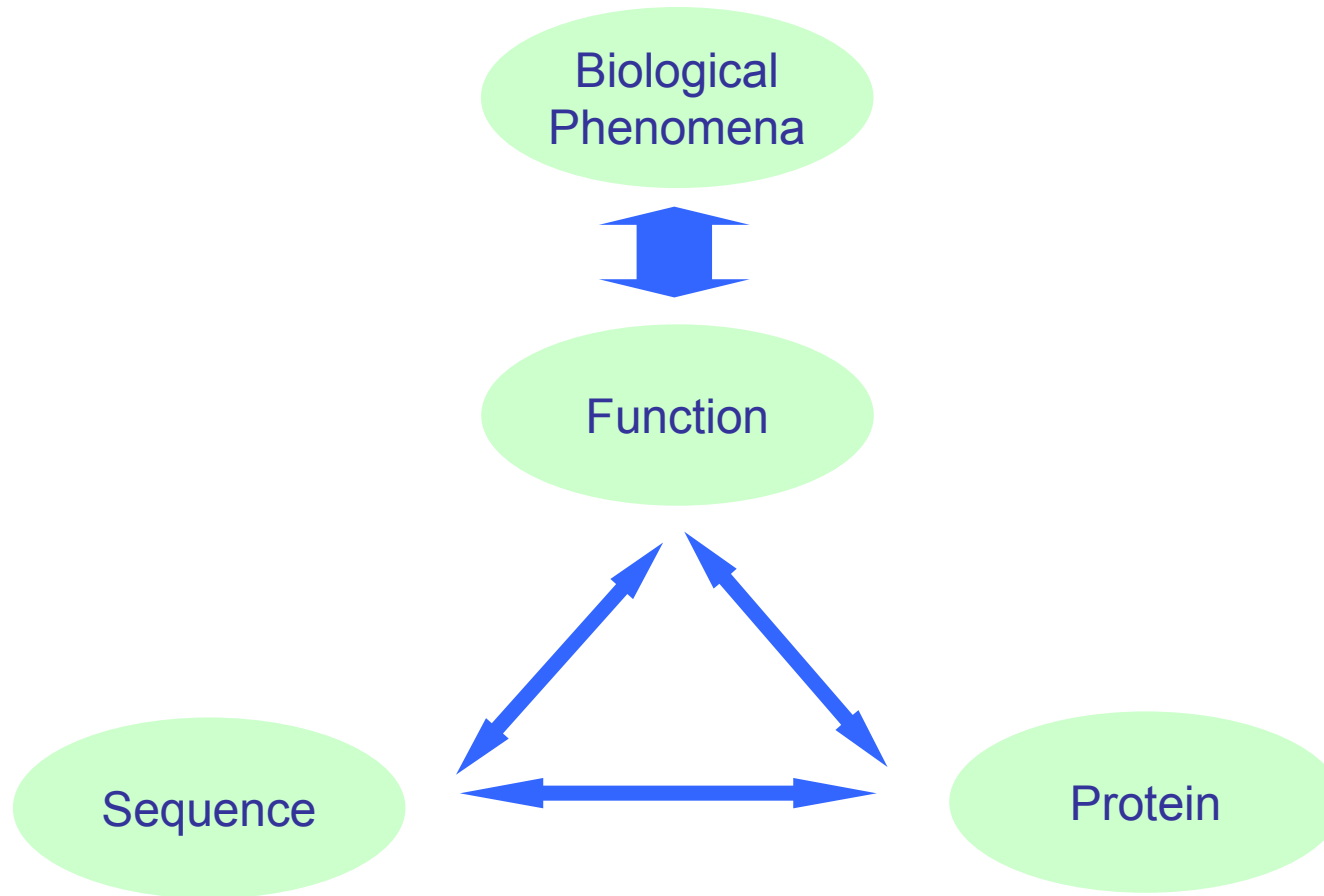
Central dogma

○ Central dogma of molecular biology



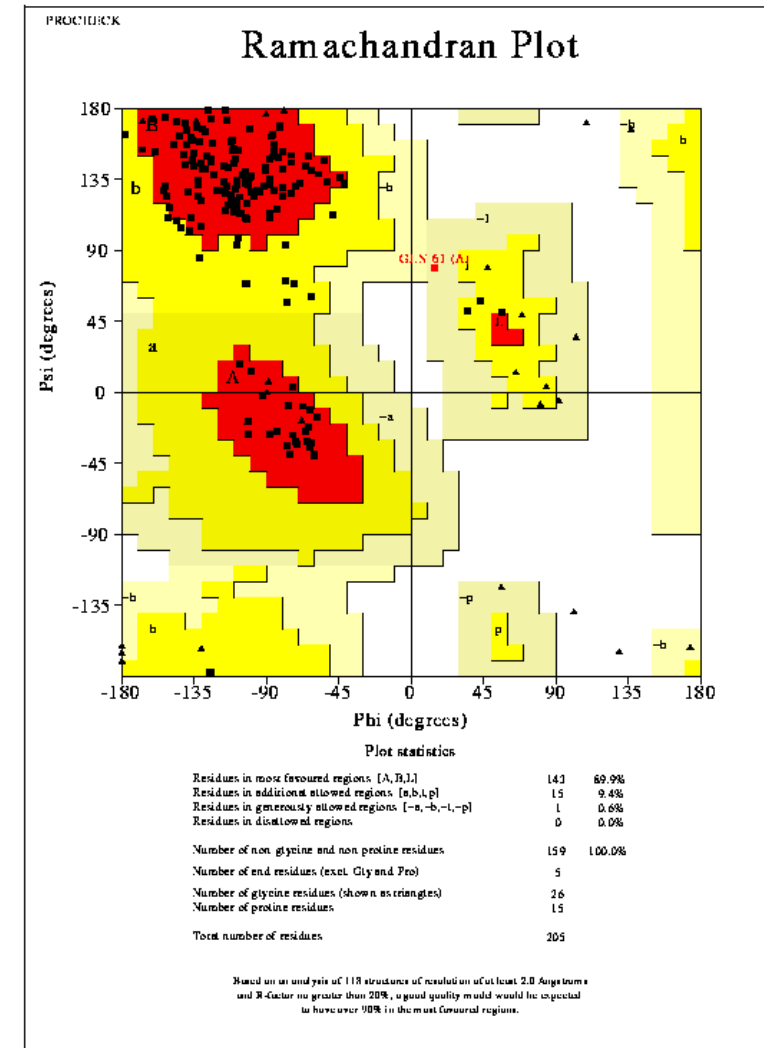
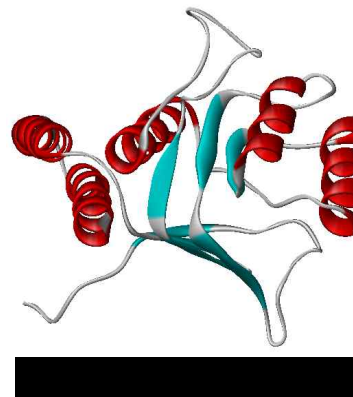
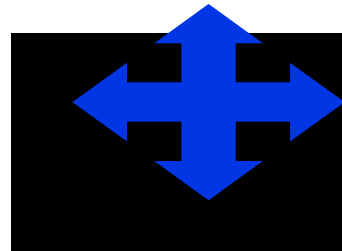
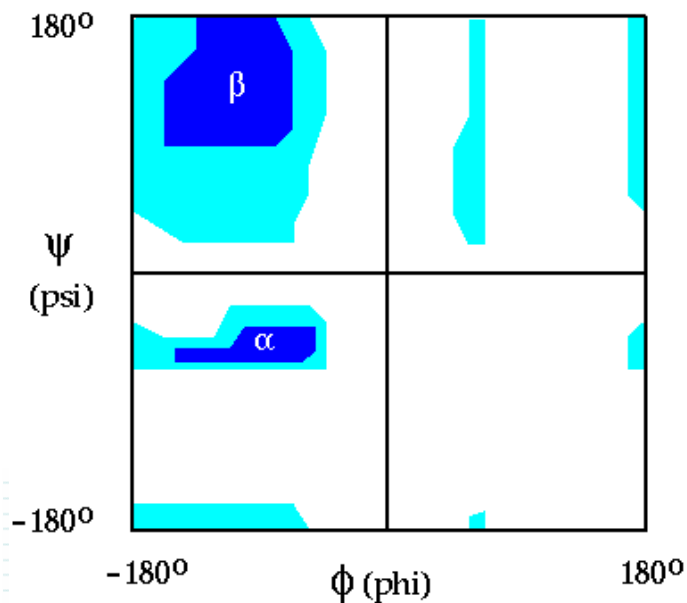
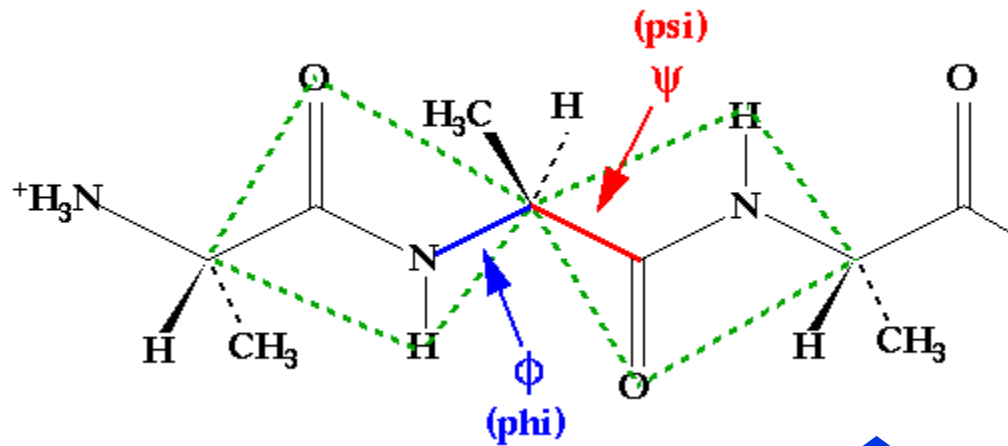
Central dogma

- Central dogma of molecular biology



Protein structure

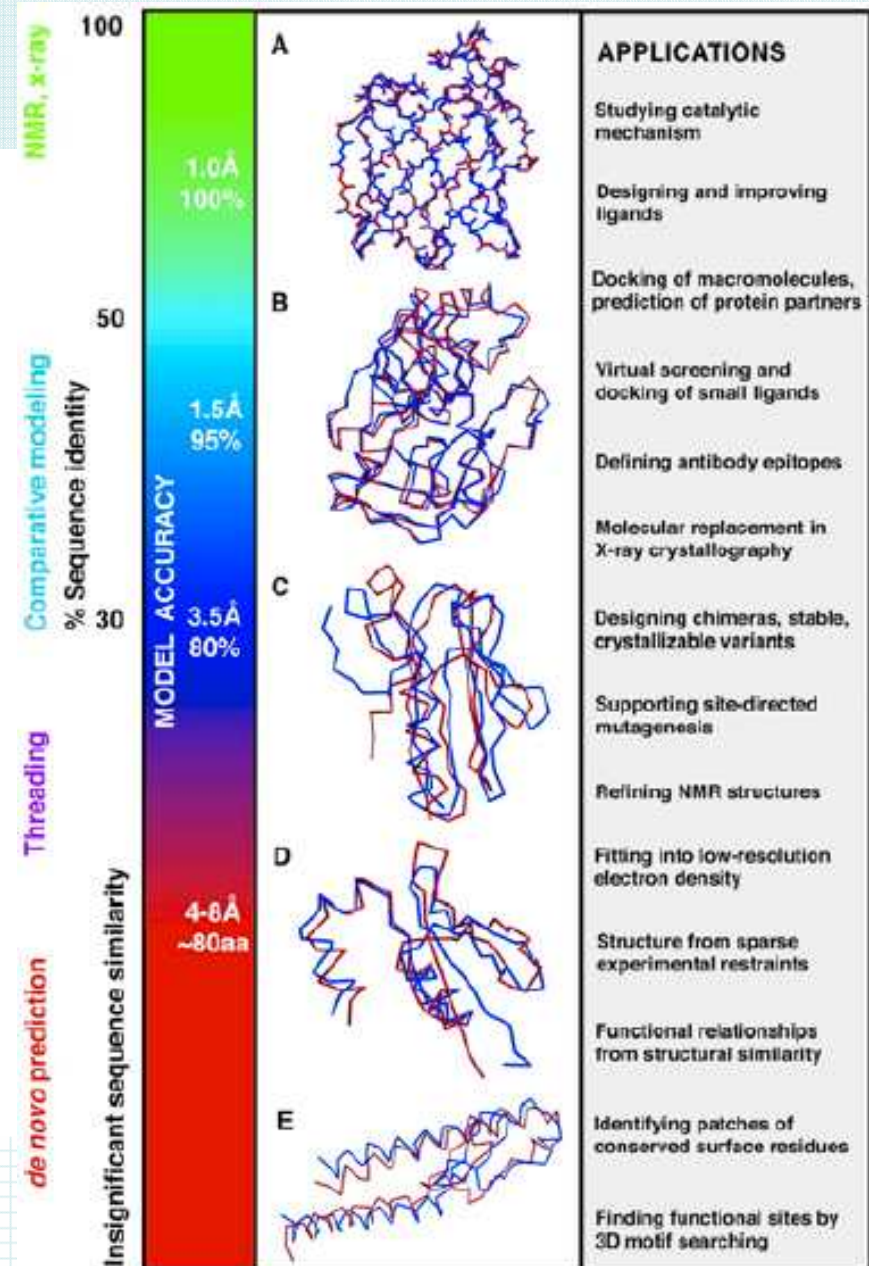
○ Protein structure validation: Ramachandran plot



Protein structure

○ How to predict protein structure?

- ✓ Experimental
 - X-ray crystallography
 - NMR
- ✓ Simulational
 - Comparative method
 - Threading method
 - *Ab initio* method



Protein structure

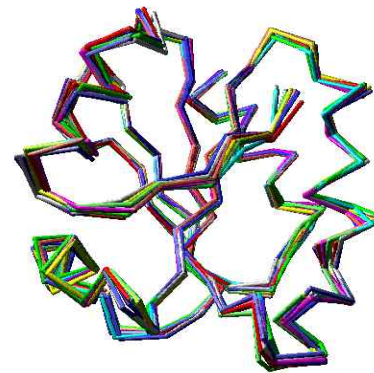
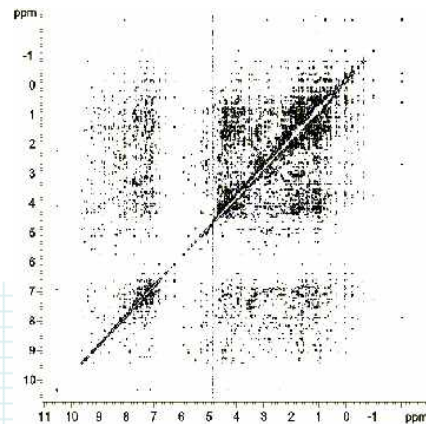
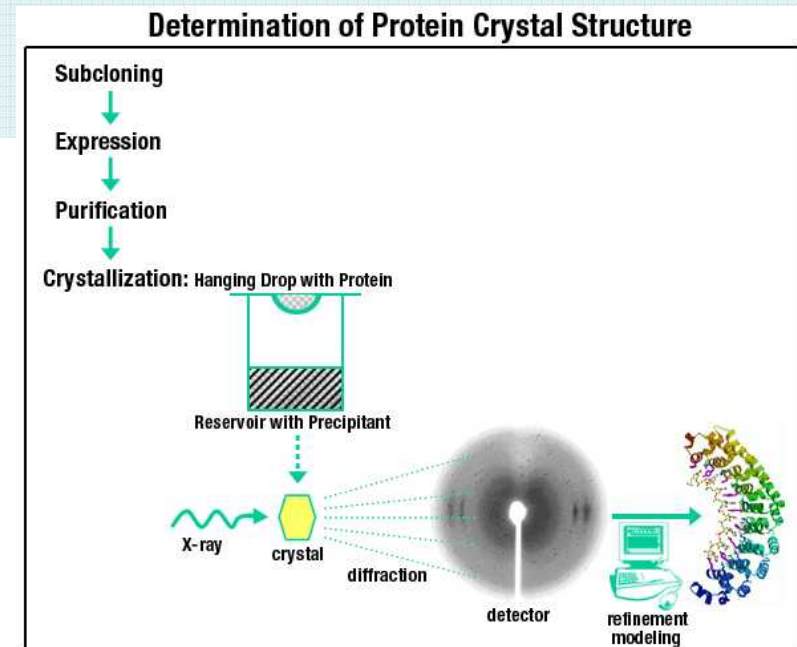
○ Experimental

✓ X-ray crystallography

- No size limitation
- Single structure

✓ NMR

- Structure determination from chemical shift
- Multiple ensemble structures
- less than 300 a.a



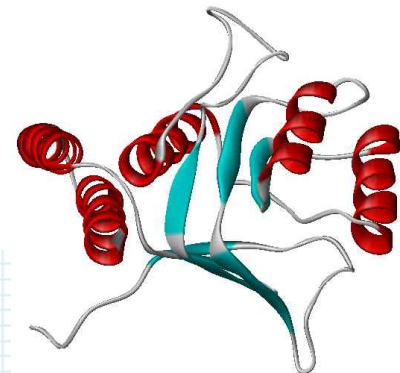
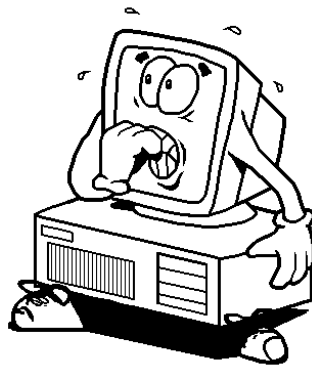
Protein structure

○ Simulational

✓ *Ab initio* method

- Ab initio = “from the beginning”; in strictest sense uses first principle, not information about other protein structure
- Size limitation : less than 50~60 a.a
- expensive computing time
- Good for finding new fold structure
- Not exact method for already existed fold structure

AKWDSTF
KKGGLLL



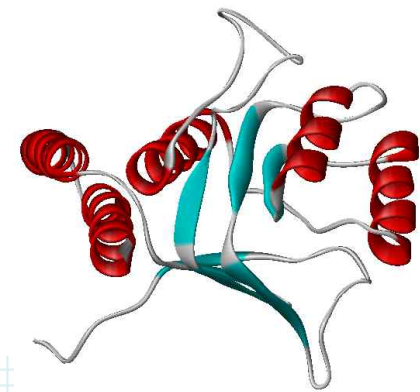
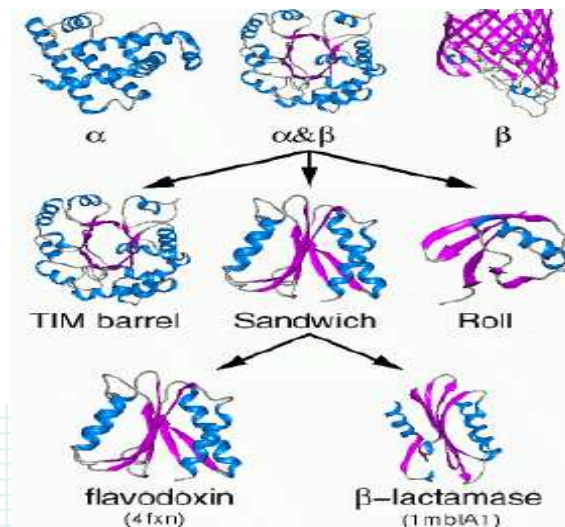
Protein structure

○ Simulational

✓ Homology modeling

- Comparative protein modeling
- Knowledge – based modeling
- Extrapolation of the structure for a new sequence (target) form the known 3D structure of related family members (template)

AKWDSTF
KKGGLLL



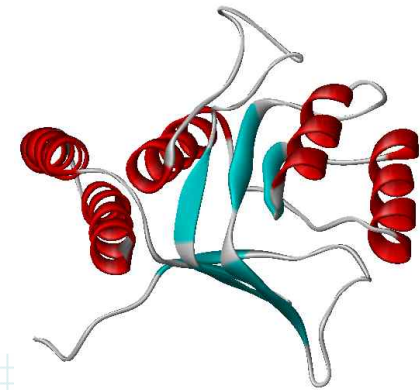
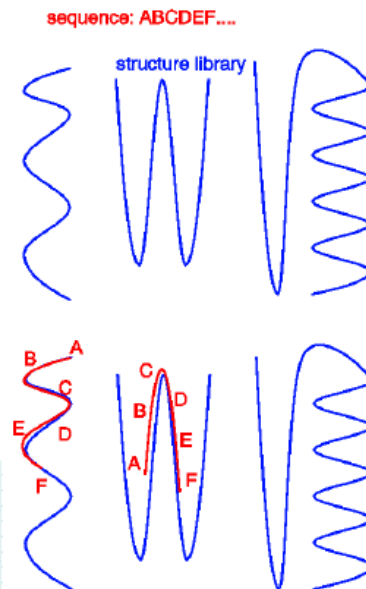
Protein structure

○ Simulational

✓ Threading method (Folding recognition)

- Find a compatible fold for a given sequence
- Take all possible segments from all PDB structures, “Thread” a target sequence through all conformations and compute some energy value
- Lowest energy conformations are predicted conformations for that segment

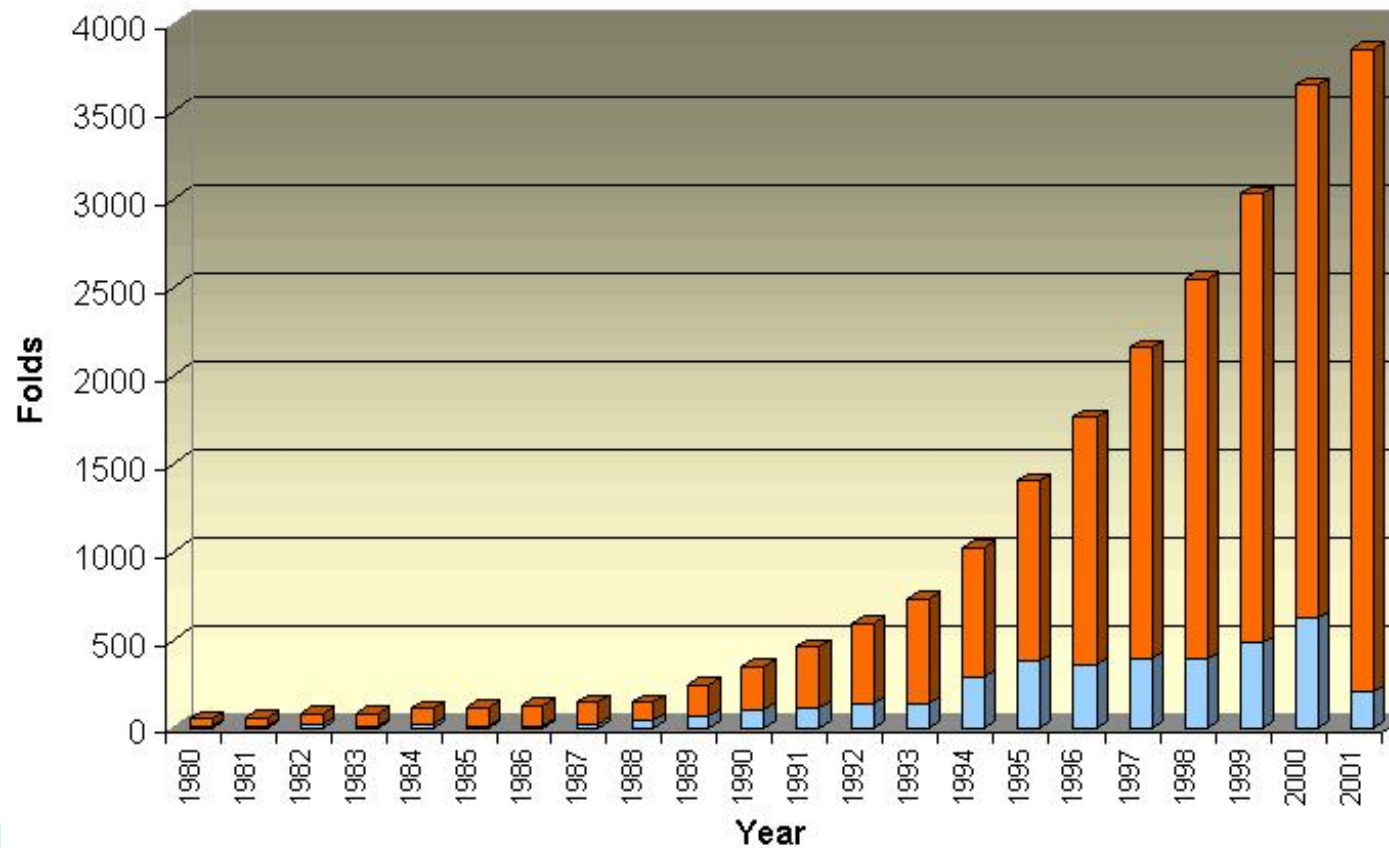
AKWDSTF
KKGGLLL



Protein structure

- PDB growth in new fold

PDB Growth in New Folds

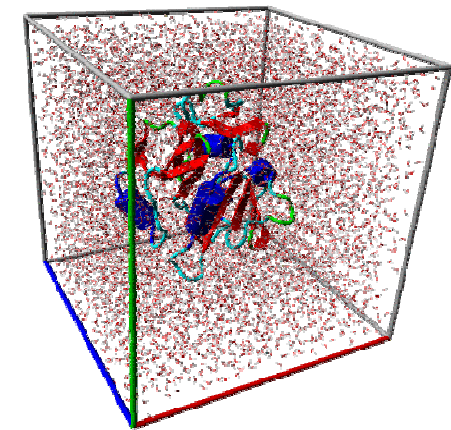
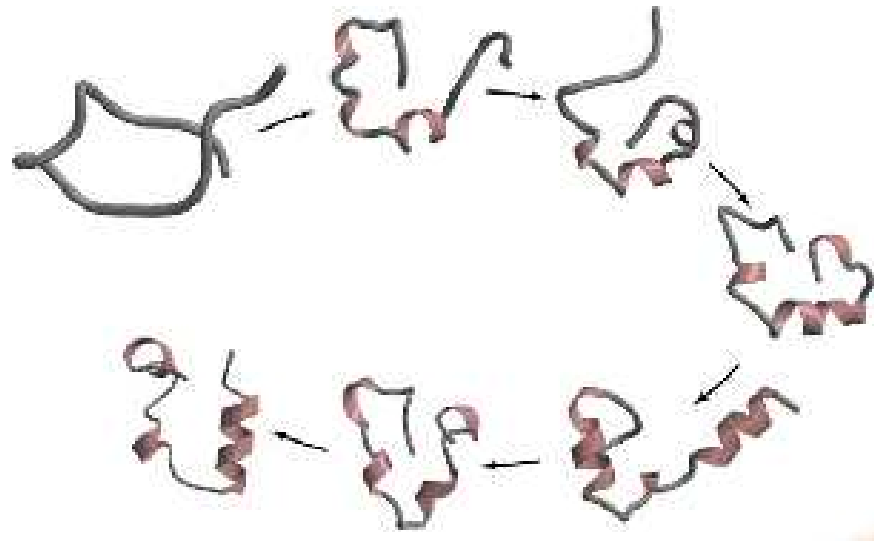
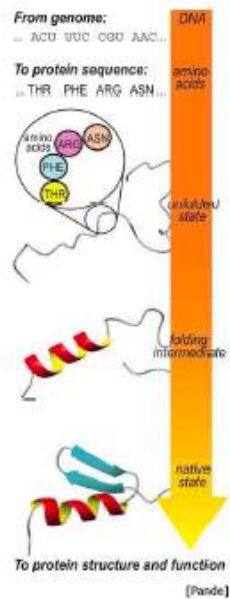


structures submitted per year; new folds per year

Protein modeling

○ Protein modeling

- ✓ How can we use computation methods to analyze, design, or predict protein structures and sequences?
- ✓ How can we predict the motion of a given protein structure?



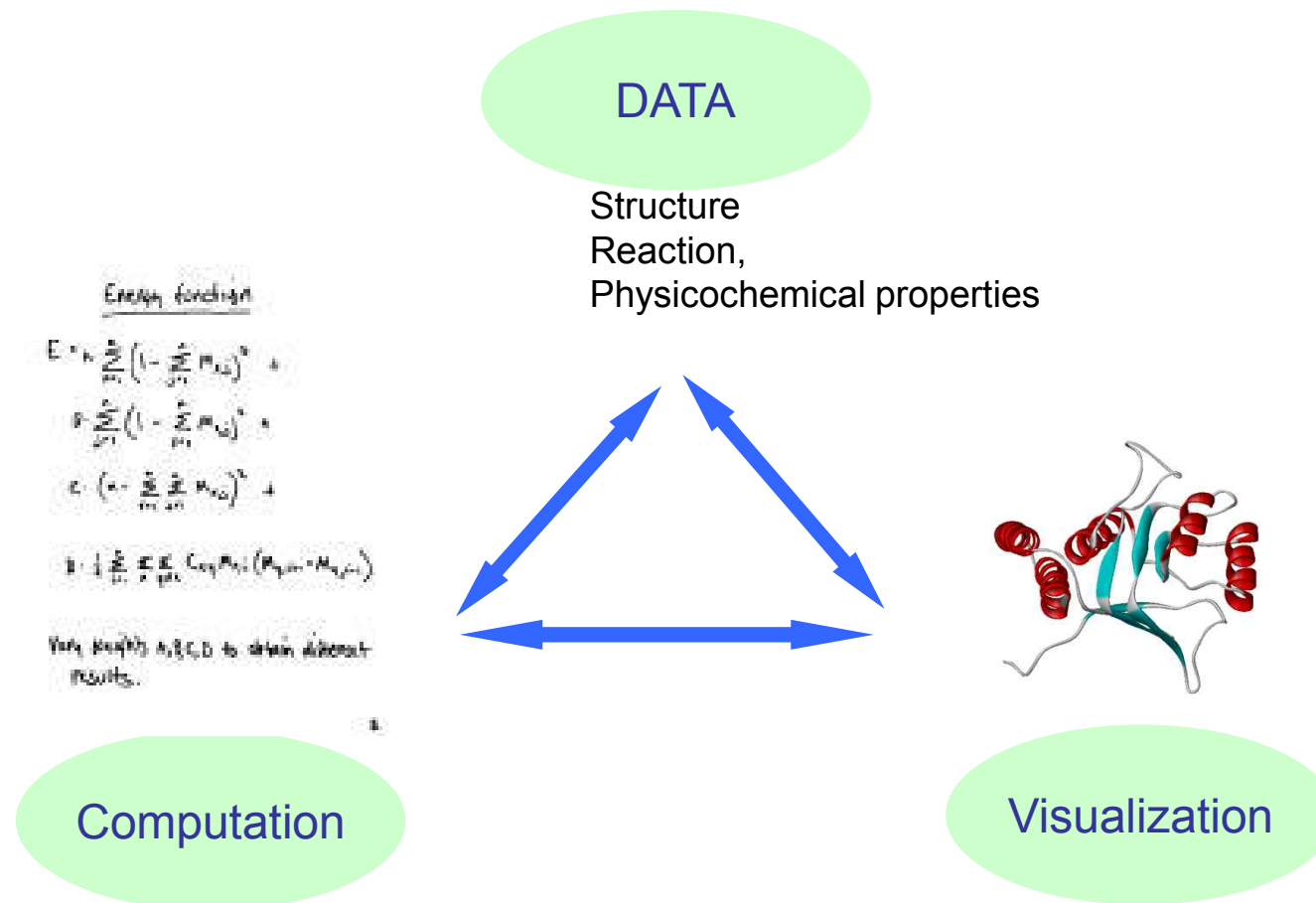
Native structure

Folding pathway

Protein dynamics

Protein modeling

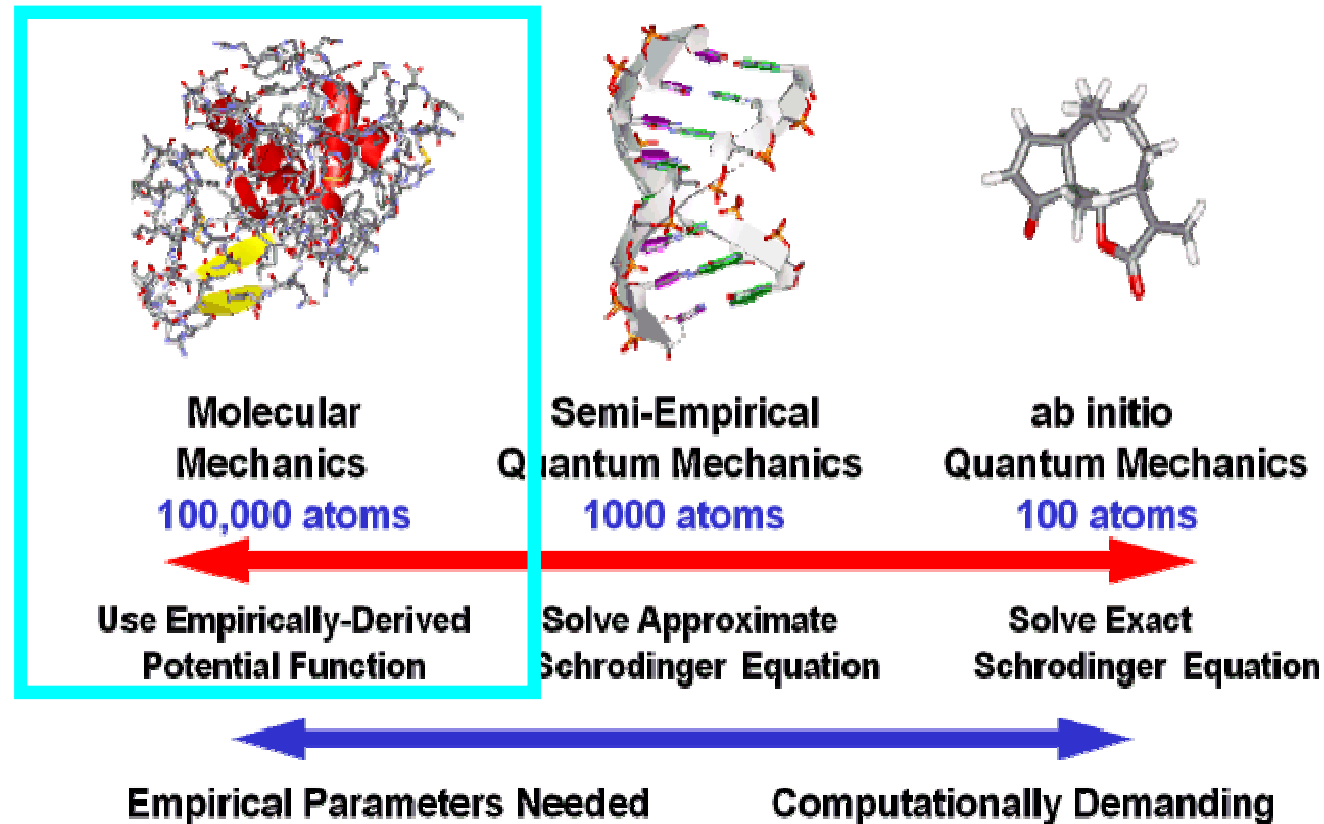
○ Major components of protein modeling



Protein modeling

- Power of computation methods

Computational Chemistry Wide Range of Methods with Different Strengths



Protein modeling

○ Software for protein modeling

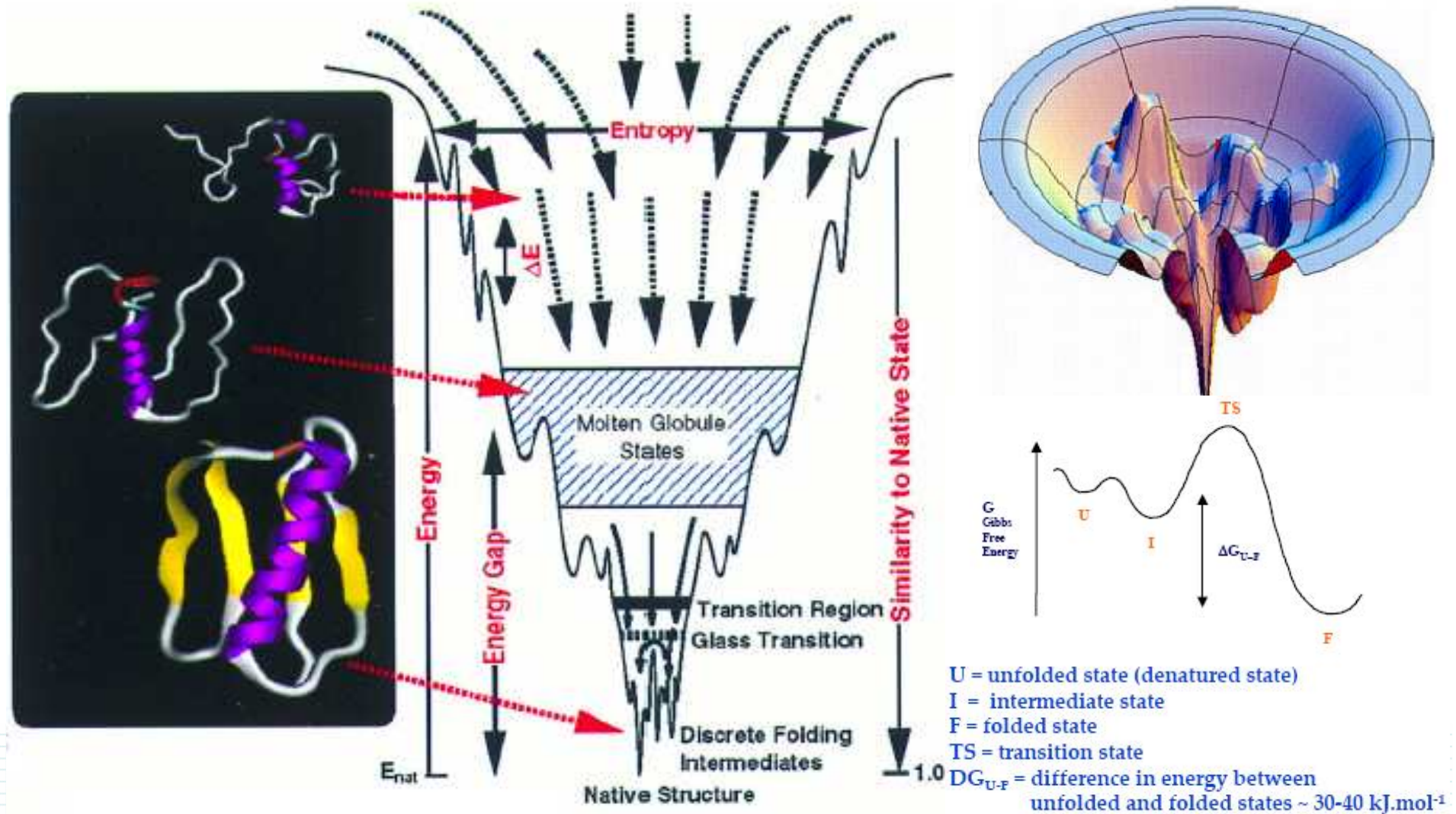
✓ *Ab initio* method : ACES, CADPAC, Gaussian, MOLPRO...

✓ Semi-empirical method : CNDO, MNDO, NNDO...

✓ Empirical method : AMBER, CAChe, CHARMM, GROMOS, X-PLOR

Protein modeling

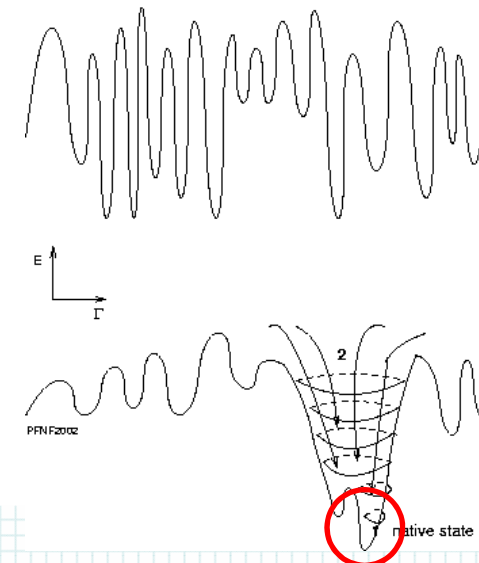
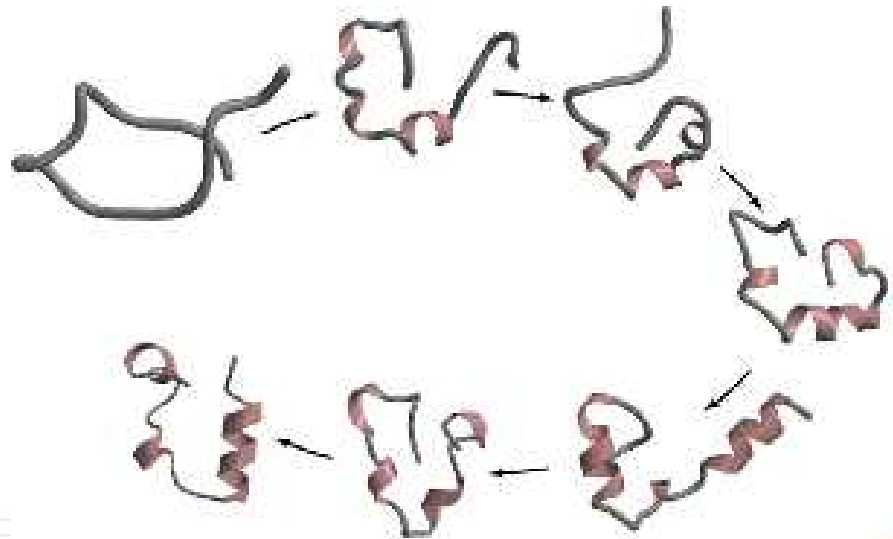
○ Protein folding landscape



Protein modeling

○ Protein folding : global minimal

- ✓ Energy minimization is frequently used to find out global minimal state of protein, especially mutant protein.
- ✓ But there are many local minima, so energy function should be accurate to find out global minima among them.



Protein modeling

○ Protein folding : Energy function

✓ To predict exact protein structures, more exact energy function is needed.

$$\begin{aligned} E \quad U(\vec{R}) = & \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{UB}} K_{UB} (S - S_0)^2 + \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}} (\phi - \phi_0)^2 \\ & + \sum_{\text{nonbond}} \epsilon \left[\left(\frac{R_{\text{min}_{ij}}}{r} \right)^{12} - \left(\frac{R_{\text{min}_{ij}}}{r} \right)^6 \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}} \end{aligned}$$

$K_b, K_{UB}, K_\theta, K_\chi, K_{\text{imp}}$ are constants, b is the bond length, b_0 is the equilibrium bond length, S is the UB 1,3-distance, S_0 is the ideal UB 1,3-distance, θ is the angle value, θ_0 is the equilibrium angle value, χ is the dihedral angle value, n is the periodicity, ϕ is the improper angle value, ϕ_0 is the ideal improper angle value, ϵ is the Lennard-Jones well depth, $R_{\text{min}_{ij}}$ is the distance at the Lennard Jones minimum, q_i and q_j are the atoms' charge ϵ_1 is the effective dielectric constant, r_{ij} is the distance between the atoms

Protein modeling

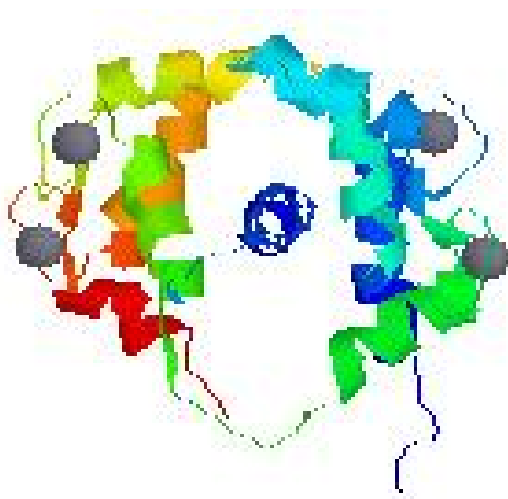
○ Molecular dynamics

- ✓ Cvff : peptide, protein
 - ✓ AMBER : protein, DNA
 - ✓ CHARMM : large macromolecules
 - ✓ ECEPP : peptide, protein
 - ✓ Merck : organic molecules
 - ✓ ...
- FF is dependant upon the kind of target molecules.

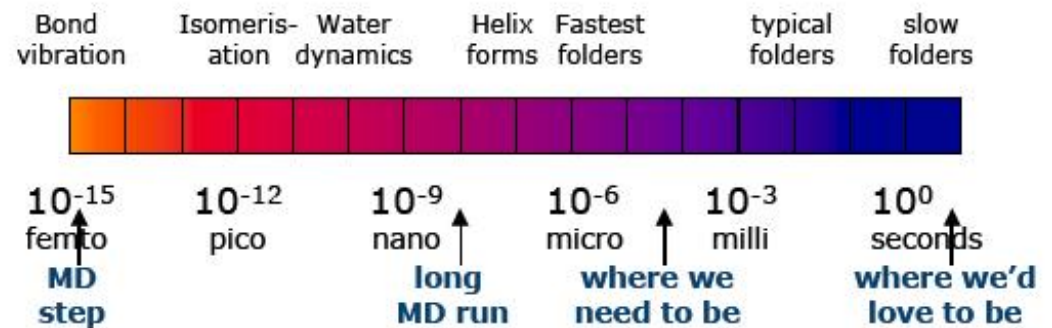
Protein modeling

○ Protein motion : Molecular dynamics

- ✓ Molecular dynamics can show how the protein will move.
- ✓ Protein motion can give much more information on protein function rather than static 3D structure.



Relevant time scales



✓ Femtosecond timesteps

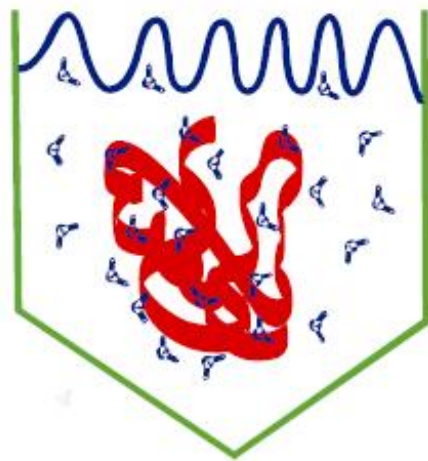
✓ Need to simulate micro to milliesconds

Protein modeling

○ Protein motion : Water dynamics

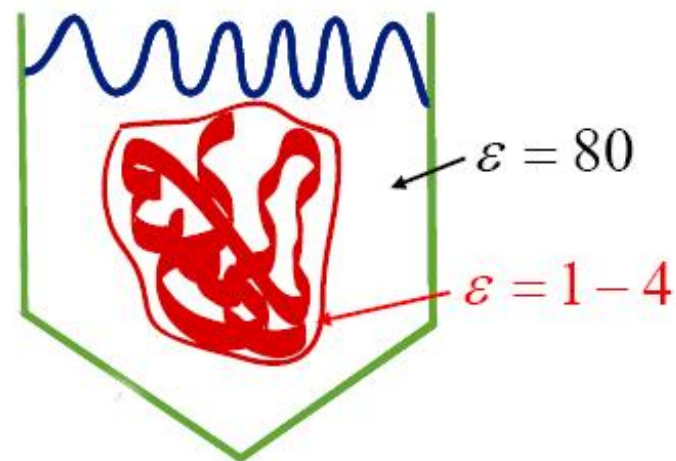
- ✓ Water is important for the biological phenomena.
- ✓ But it is difficult to reflect exact water dynamics.

Microscopic treatment



Explicit model

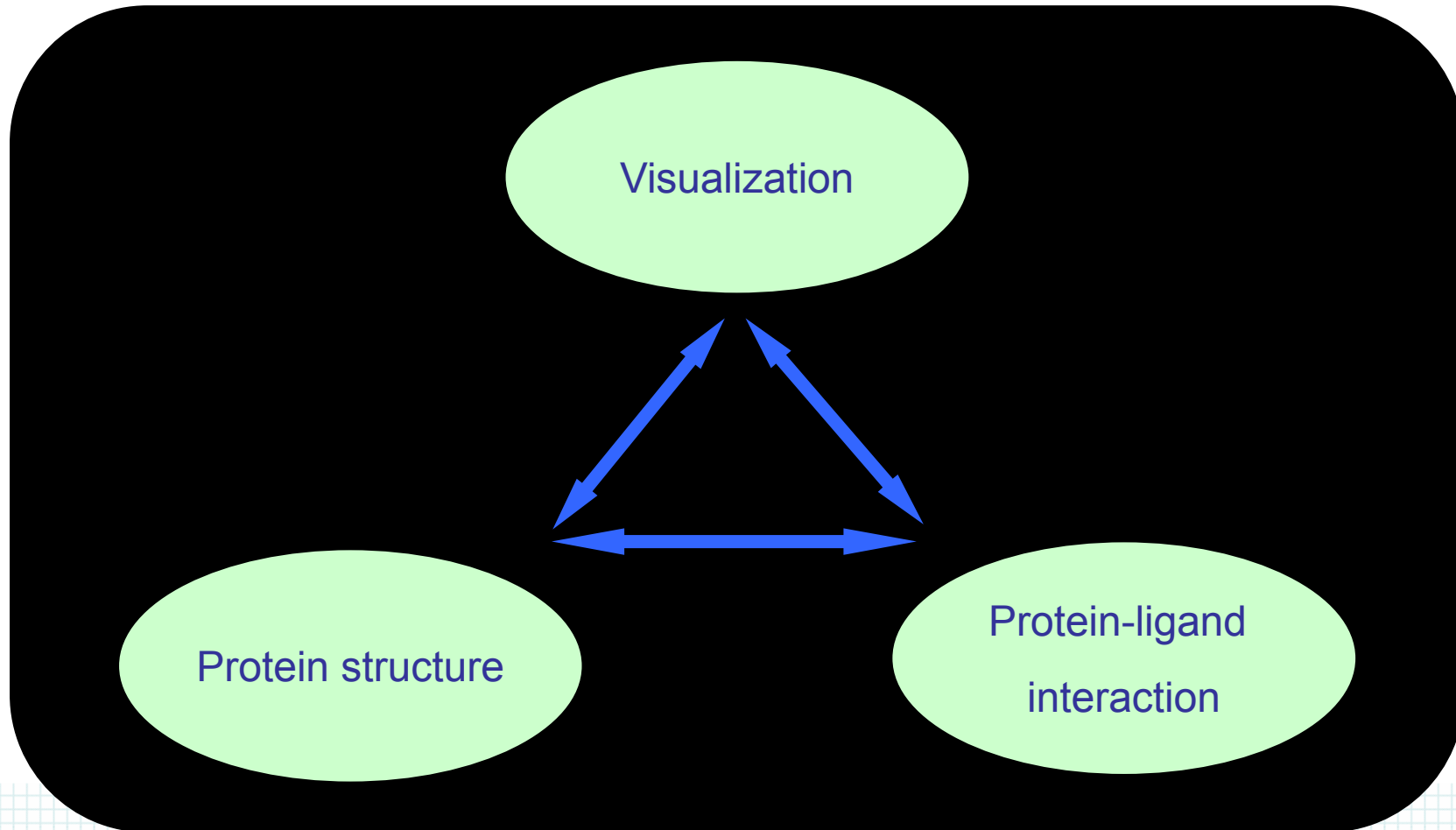
Macroscopic treatment



Implicit model

Protein modeling

- Softwares for protein modeling

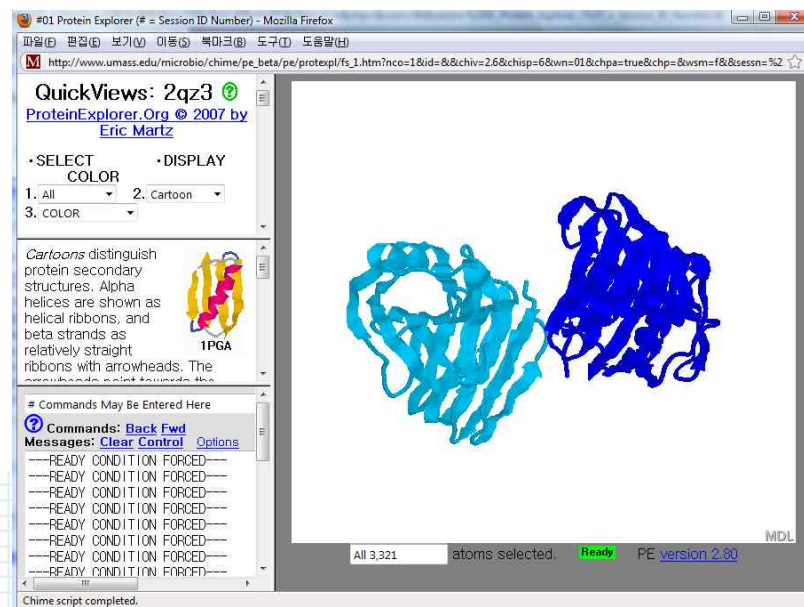
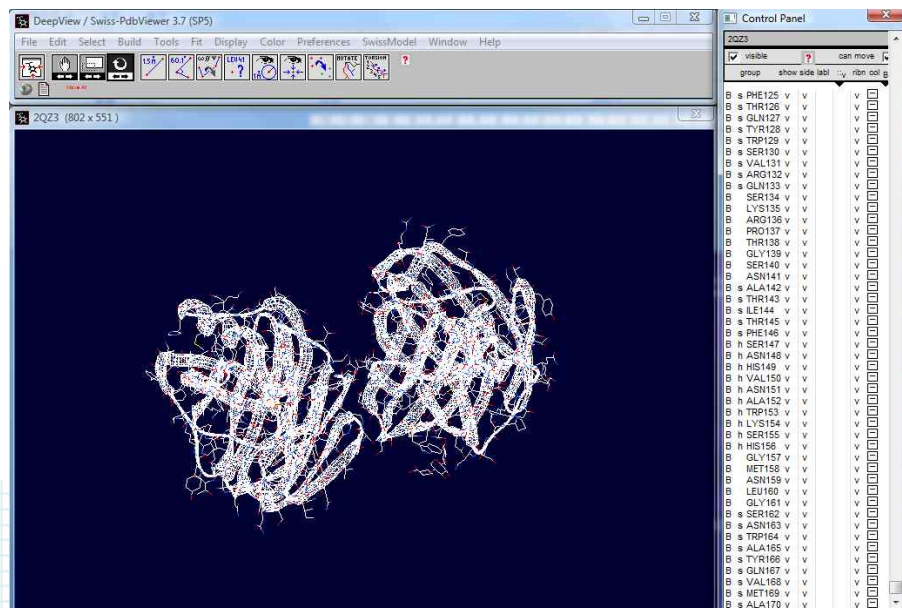
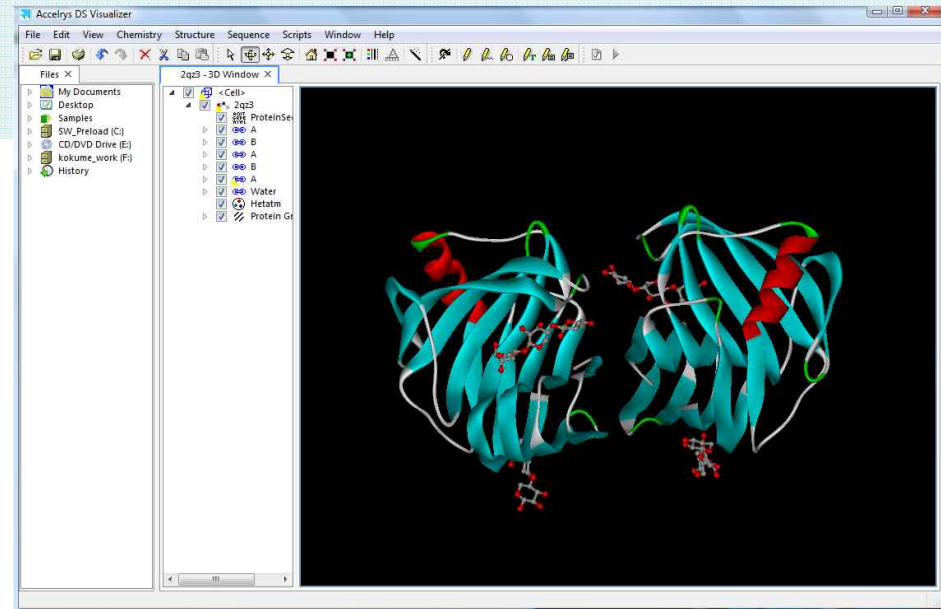
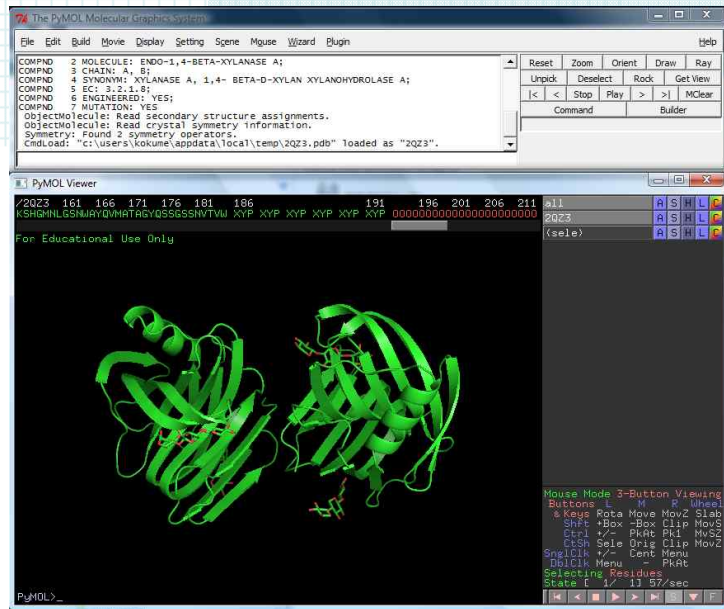


Protein modeling

○ Softwares for visualization

- ✓ Pymol: <http://pymol.sourceforge.net>
- ✓ Swiss-PdbViwer: <http://spdbv.vital-it.ch>
- ✓ Ds visualizer: <http://accelrys.com>
- ✓ Protein explorer: [http:// proteinexplorer.org](http://proteinexplorer.org)
- ✓ ...

Protein modeling



Protein modeling

○ Softwares for structure prediction

- ✓ Scwrl 4.0: <http://dunbrack.fccc.edu/scwrl4/SCWRL4.php>
- ✓ Swiss Model: <http://swissmodel.expasy.org>
- ✓ RosettaDesign: <http://rosettadesign.med.unc.edu>
- ✓ FoldX: <http://foldx.crg.es>
- ✓ ...

Protein modeling

SCWRL4 - Protein Side Chain Prediction: Dunbrack Lab - Windows Internet Explorer

Home
People
Projects
Publications
Software
Links

SCWRL3
PISCES
BBDep
SIC
MolDE
ArbIDraw
BioDownloader
ProtBuD

News:
SCWRL4

Dunbrack Lab

>>SCWRL4

Registered Scwrl users: 3407

[A baby squirrel from the Fox Chase section of Philadelphia]

License for SCWRL4 for non-profit users. Click here

Description
SCWRL4 Availability
Installation
Usage
Contact

Description:

SCWRL4 is based on a new algorithm and new potential function that results in improved accuracy at reasonable speed. This has been achieved through: 1) a new backbone-dependent rotamer library based on local description of the protein surface and

SIB
Swiss Institute of Bioinformatics

BIOCENTRUM
University of Basel

SWISS-MODEL

Modelling

myWorkspace
Automated Mode
Alignment Mode
Project Mode

Tools

Template Identification
Domain Annotation
Structure Assessment
Template Library

Repository

Search by Sequence
Search by AC

Documentation

SWISS-MODEL Workspace
SWISS-MODEL Repository
Structures & Models
Helpdesk

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists WorldWide.

SWISS-MODEL Team

Torsten Schwede: Project Leader
Florian Kiefer: SWISS-MODEL Repository
Lorenza Bordoli: Method Development and user support
Konstantin Arnold: SWISS-MODEL Workspace

References:
When you publish or report results using SWISS-MODEL, please cite the relevant publications:

- Arnold K, Bordoli L, Kopp J, and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, 37, D387-D392.
- Peitsch, M. C. (1995) Protein modeling by E-mail *BioTechnology* 13: 658-660.

What's new?

- New automated modeling pipeline with improved hierarchical approach for template selection.
- Increased sensitivity of template detection (sequence to profile search using an adapted HHSearch protocol)
- New tools for model and structure quality assessment: Dfire and Qmean global scores; ProQres residue based assessment scores

Rosetta Design - Windows Internet Explorer

http://rosetta-design.mrc-lmb.cam.ac.uk/rosetta-design

File Edit View Favorites Tools Help

Home
Documentation
Register
Log in
Queue

RosettaDesign

Welcome to the RosettaDesign web server.

RosettaDesign identifies low energy sequences for specified protein backbones, and has been used previously to stabilize proteins and create new protein structures.

Please login to begin using RosettaDesign.

Crystal structure of top7 -- A novel protein structure created with RosettaDesign.

FoldX - Windows Internet Explorer

http://folds.org/

File Edit View Favorites Tools Help

Home
Documentation
Register
Log in
Queue

FoldX

A force field for energy calculations and protein design

News

- FoldX 3.0 Beta3 released with new mutation capabilities. You can download it after log in on this page.
- Warning: Dear Users, The version 3.0 is a beta version, the final one will be released soon. Please be aware that there are some things not yet fully operational. For the moment, avoid mutating DNA to the methylated bases. Please if you find some problems, send us an email.
- Warning: The energy values of the version 3.0 beta have been fitted to correlate with differences in energies between wild type and mutant. Do not take them as absolute values of the stability (or affinity) of your protein (or complex).

Links

- FoldX: Fragmenting the protein space
- SmartCell: A Cell Network Simulation Program
- Tango: A computer algorithm for prediction of aggregating regions in unfolded polypeptide chains

We have developed a computer algorithm, **FoldX** to provide a fast and quantitative estimation of the importance of the interactions contributing to the stability of proteins and protein complexes.

The predictive power of FOLDEF has been tested on a very large set of point mutants (1088 mutants) spanning most of the structural environments found in proteins.

FoldX uses a full atomic description of the structure of the proteins. The different energy terms taken into account in **FoldX** have been weighted using empirical data obtained from

Protein modeling

○ Softwares for protein-ligand interaction

✓ GRAMM-X :

<http://vakser.bioinformatics.ku.edu/resources/gramm/grammx>

✓ ZDOCK: <http://zdock.bu.edu>

✓ Autodock: <http://autodock.scripps.edu/>

✓ ...

Protein modeling

Vakser Lab

GRAMM-X Protein Docking Web Server v.1.2.0

This is the Web interface to our current protein docking software made available to the public. This software is different from the original GRAMM, except that both packages use FFT for the global search of the best rigid body conformations.

You can submit input files and parameters to this web server and the docking simulation will be run on our computer cluster.

When the results are ready, they will be saved in a temporary directory on the web server and the link to that directory will be sent to you.

Please, read the [Conditions of Use](#) before proceeding.

Questions

Send questions or comments to Andrey.Tovchigrechko.

Start new GRAMM-X simulation

Main Input

Receptor PDB file •

Select the PDB file on your computer to use as the receptor. This file will be uploaded to our server.

Chain Ids of the Receptor

Specify chain id's for the receptor as one or several letters. Examples: LH - chains L and H; C - chain C; leaving this field empty will mean to use all chains from the receptor file.

 AutoDock

[site map](#) [accessibility](#)

[home](#) [downloads](#) [resources](#) [faqs & help](#) [forum](#) [wiki](#) [contact](#)

[로그인](#)

현재 위치: 홈

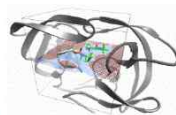


AutoDock

by [morris](#) - 최종 수정일: 2009-10-08 09:35
Contributors: Sargis Dallakyan

Welcome!

- What is [AutoDock?](#)
- What's new?
- What is [AutoDockTools?](#)
- Where is [AutoDock](#) used?
- Why use [AutoDock?](#)
- Run [AutoDock](#) on World Community Grid!



[Garrett M. Morris](#)
[David S. Goodsell](#)
[Ruth Huey](#)
[William Lindstrom](#)
[William E. Hart](#)
[Scott Kurowski](#)
[Scott Halliday](#)
[Rik Belew](#)
[Arthur J. Olson](#)

새소식

- [AutoDock Vina 1.0 released!](#)
2009-06-10
- [AutoDock Vina new beta release.](#)
2009-02-19
- [Run your AutoDock Research project on World Community Grid!](#)
2009-02-10
- [AutoDock's role in Developing the First Clinically-Approved HIV Integrase Inhibitor](#)
2007-12-07

What is AutoDock?

[AutoDock](#) is a suite of automated docking tools. It is

ZDOCK SERVER

[Zdock](#)

[Zlab](#)

[Help](#)

[Contact](#)

Upload file 1:

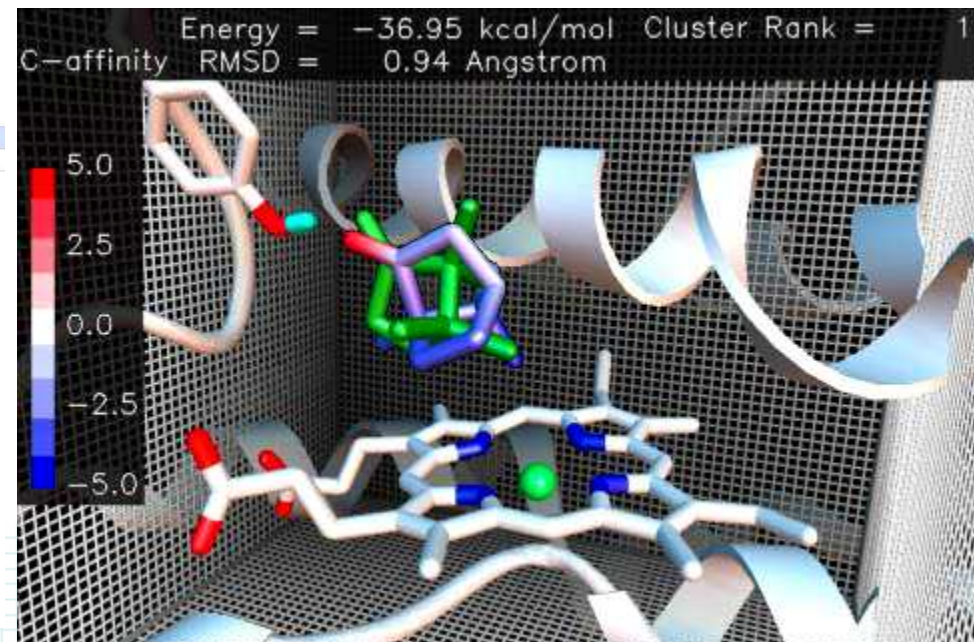
Upload file 2:

Enter your email:

First time using this server? Try the [tutorial](#).

Note: Please refrain from submitting more than approximately 10 jobs per day as this can back up the server and prevent other users' jobs from running.

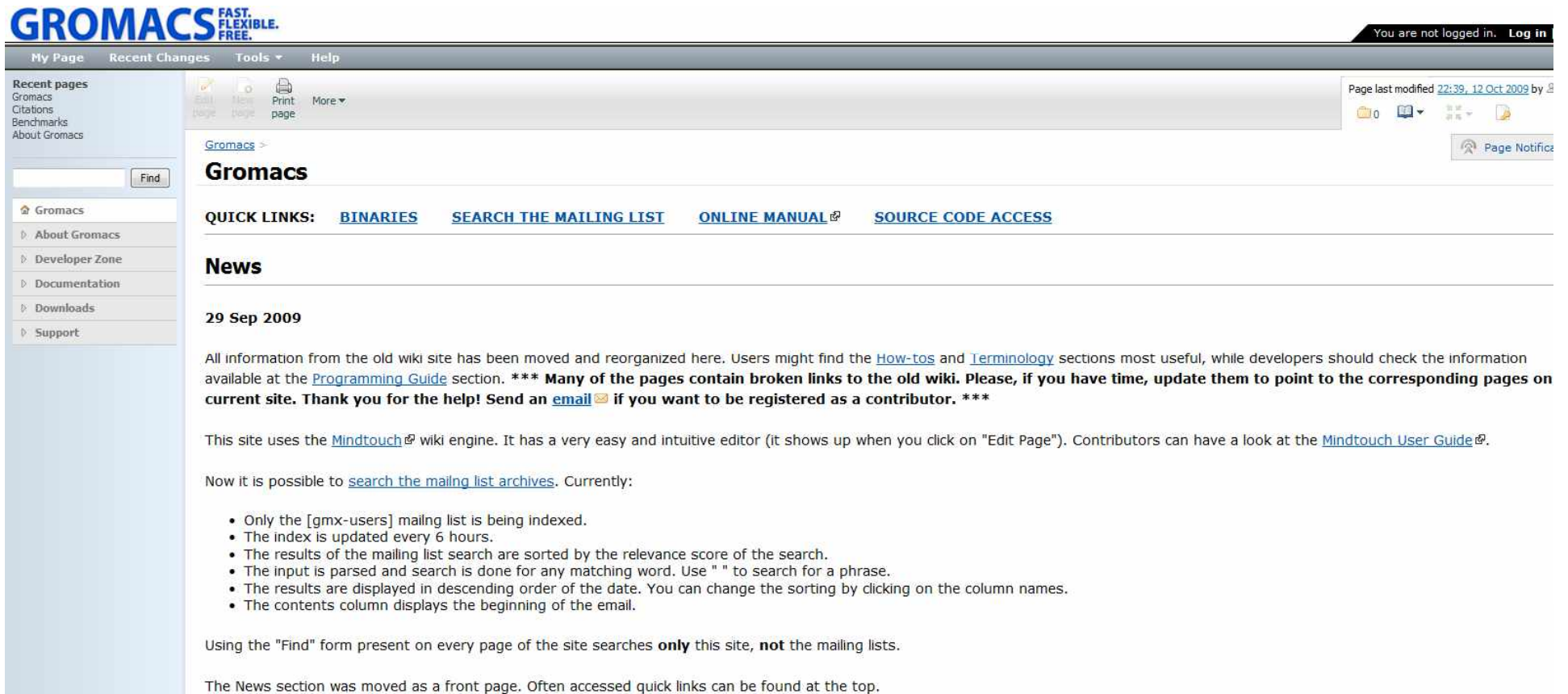
AVERAGE WAIT TIME IS: 01:26:35 (HH:MM:SS)



Protein modeling

○ Software for molecular dynamics

✓ Gromacs : <http://www.gromacs.org>



The screenshot shows the GROMACS website homepage. At the top left is the GROMACS logo with the tagline "FAST. FLEXIBLE. FREE.". Below the logo is a navigation menu with "My Page", "Recent Changes", "Tools", and "Help". On the right side of the header, it says "You are not logged in. Log in". Below the header is a sidebar on the left with "Recent pages" (Gromacs, Citations, Benchmarks, About Gromacs) and a search box with a "Find" button. The main content area has a "Gromacs" breadcrumb, the title "Gromacs", and "QUICK LINKS" for "BINARIES", "SEARCH THE MAILING LIST", "ONLINE MANUAL", and "SOURCE CODE ACCESS". Below this is a "News" section dated "29 Sep 2009" with a paragraph of text: "All information from the old wiki site has been moved and reorganized here. Users might find the [How-tos](#) and [Terminology](#) sections most useful, while developers should check the information available at the [Programming Guide](#) section. *** Many of the pages contain broken links to the old wiki. Please, if you have time, update them to point to the corresponding pages on current site. Thank you for the help! Send an [email](#) if you want to be registered as a contributor. ***". This is followed by a paragraph about the Mindtouch wiki engine and a paragraph about mailing list archives. A bulleted list follows: "• Only the [gmx-users] mailing list is being indexed. • The index is updated every 6 hours. • The results of the mailing list search are sorted by the relevance score of the search. • The input is parsed and search is done for any matching word. Use \" \" to search for a phrase. • The results are displayed in descending order of the date. You can change the sorting by clicking on the column names. • The contents column displays the beginning of the email." Below the list is a paragraph: "Using the \"Find\" form present on every page of the site searches **only** this site, **not** the mailing lists." and another paragraph: "The News section was moved as a front page. Often accessed quick links can be found at the top."

Protein modeling

○ How do we use this computation method?
: Case studies

✓ Design of protein for increasing stability

✓ Redesign of protein-ligand interactions

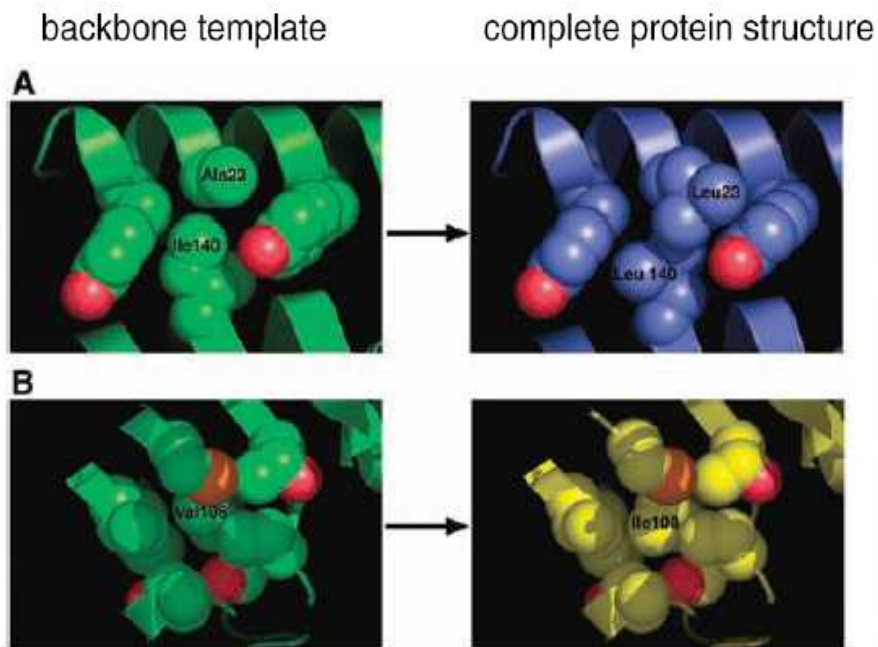
✓ Investigation on protein dynamics in organic solvents

✓ ...

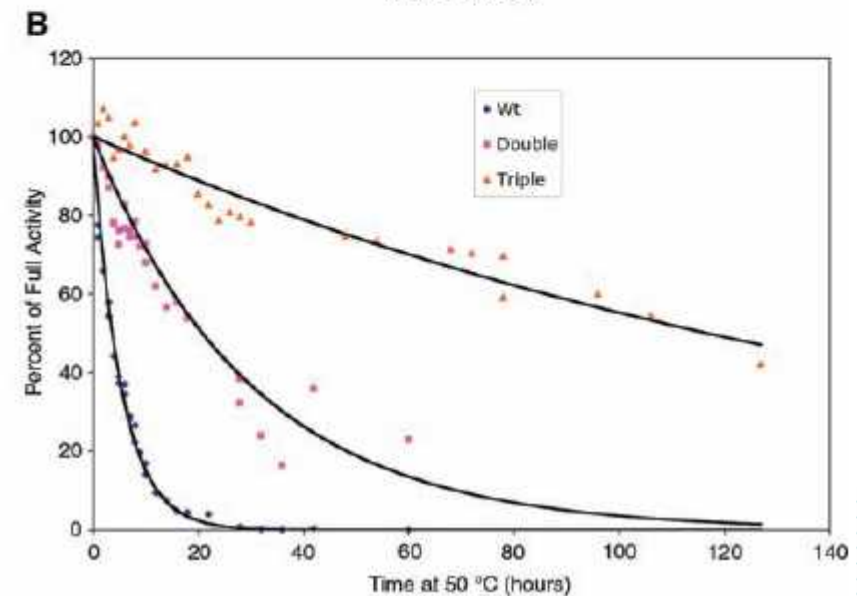
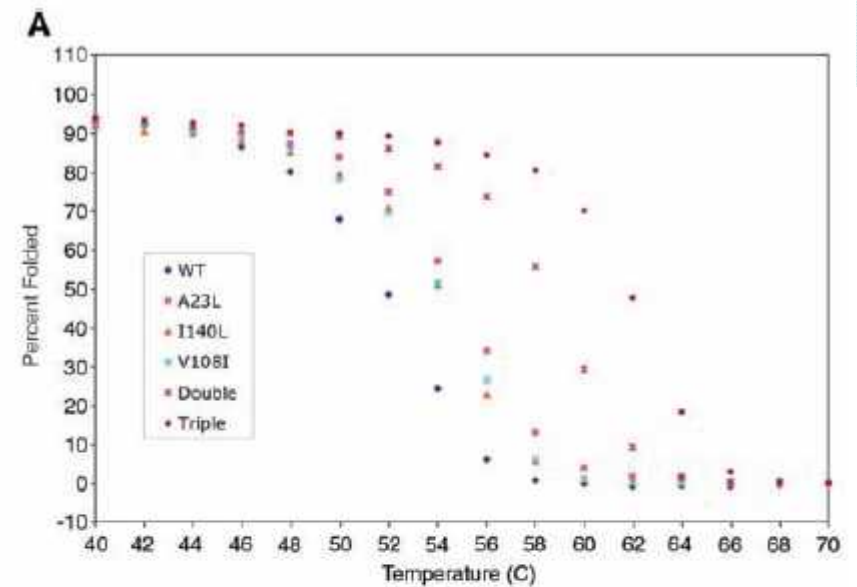
Protein modeling

○ Redesign of protein for increasing stability

✓ Computational thermostabilization of an



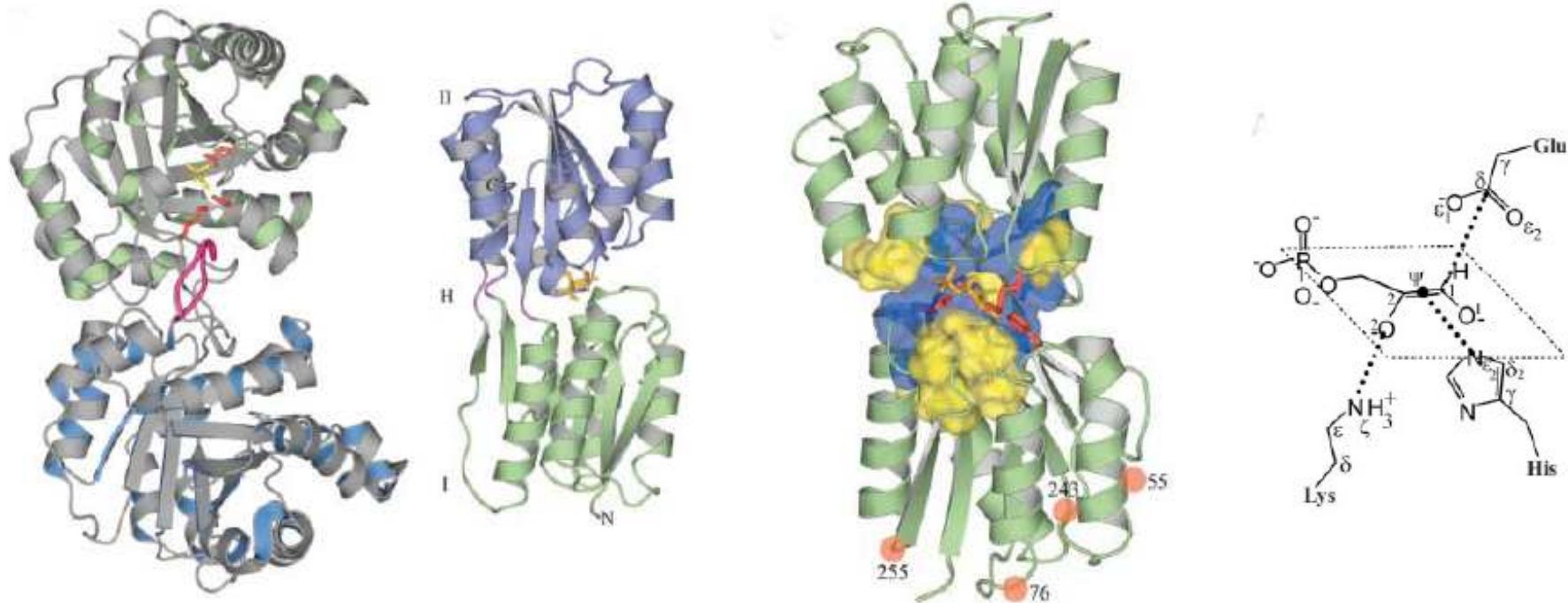
Redesign of yeast cytosine deaminase



Protein modeling

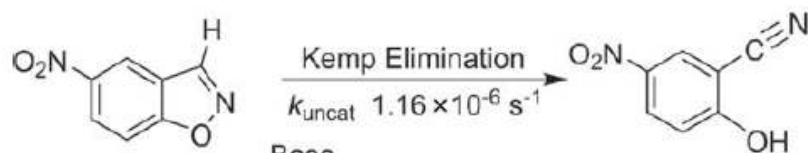
○ Redesign of protein to enzyme

- ✓ Computational design of a biologically active enzyme : Science , 304, 1967-1971

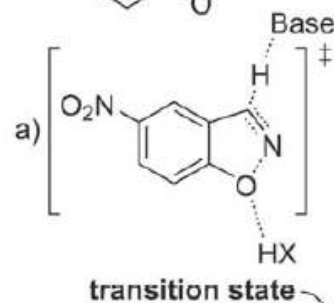


Comparison TIM (Triose phosphate isomerase) with RBP

Protein modeling



✓ Kemp elimination catalysts by
 computational enzyme design:
 Nature 453 (7192), pp. 190-195

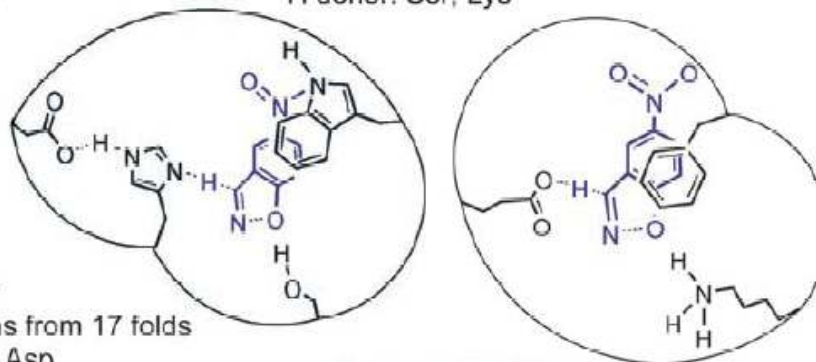


TIM barrel
 jelly roll
 lipocalin
 β propeller...
 protein folds

enzyme design

base: Glu, Asp, Asp + His
 π stacking: Trp, Phe
 H donor: Ser, Lys

RosettaMatch
 in silico enzyme design
 > 100000 candidates

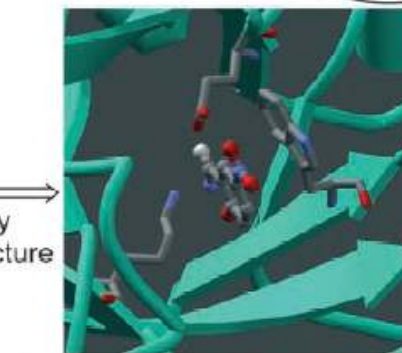


59 designs from 17 folds
 b) 39 Glu or Asp
 20 His + Glu or Asp

protein produced
 purified & screened
 c) 8 active artificial enzymes
 TIM barrel predominant

k_{cat}/K_m 163
 k_{cat}/k_{uncat} 2.5×10^5

X-ray structure



k_{cat}/K_m 12.2; k_{cat}/k_{uncat} 1.6×10^4

d)

site-directed mutagenesis

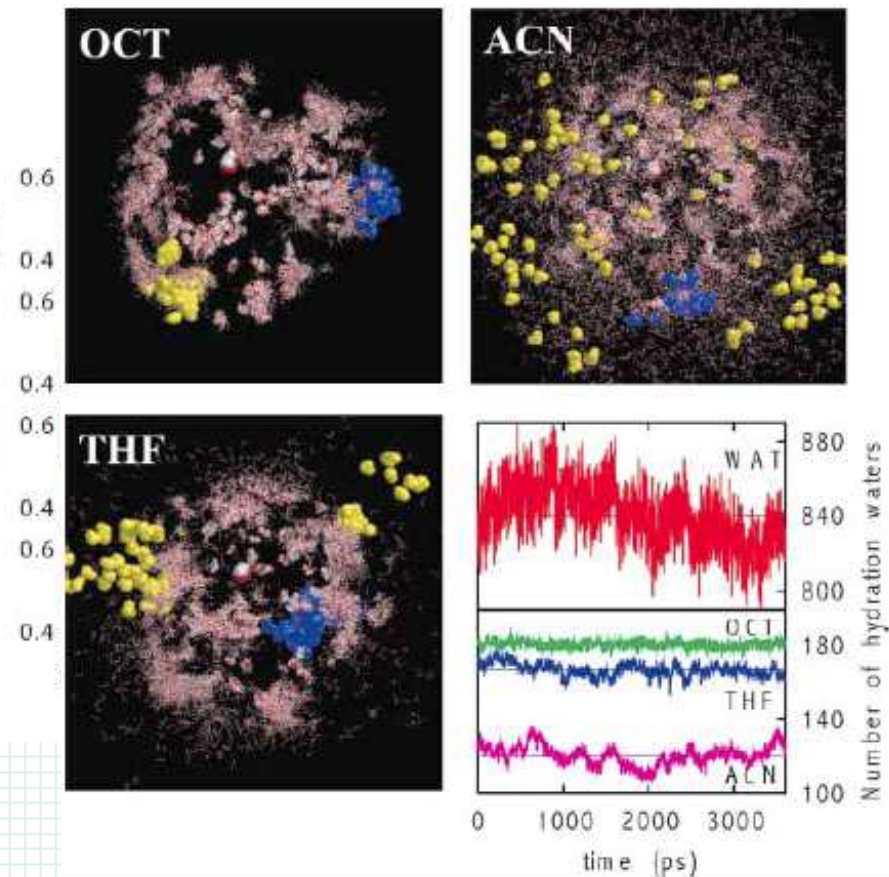
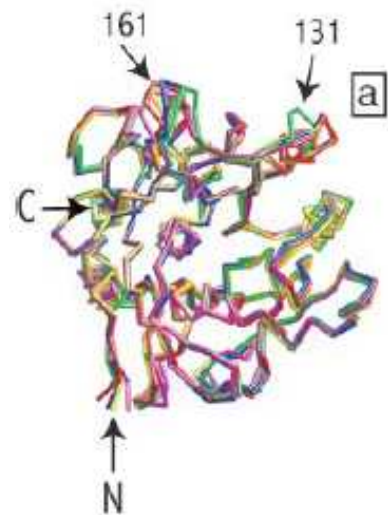
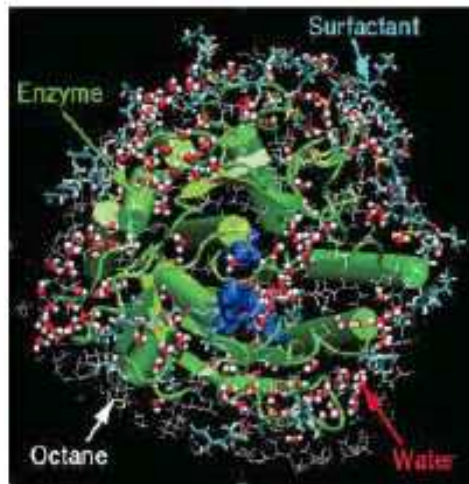
k_{cat}/K_m 22.7
 k_{cat}/k_{uncat} 2.6×10^4

in vitro directed evolution (7 rounds)

k_{cat}/K_m 2590
 k_{cat}/k_{uncat} 1.2×10^6
 TON > 1000
 e)

Protein modeling

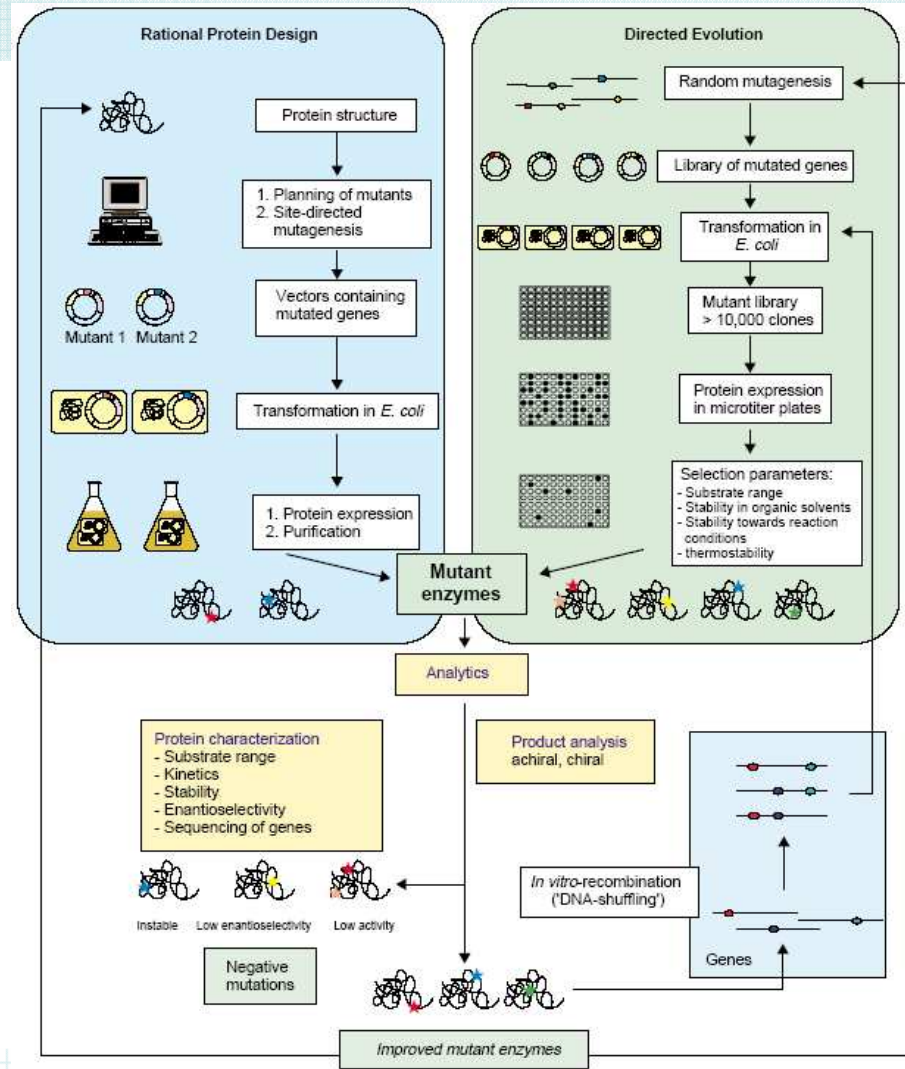
- Investigation on protein dynamics in organic solvents
 - ✓ Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity : Biophysical Journal 87, 812-821



Protein Function

Rational/
Computational
design

Directed
evolution



Protein stability

○ Rational/computational design vs Directed evolution

