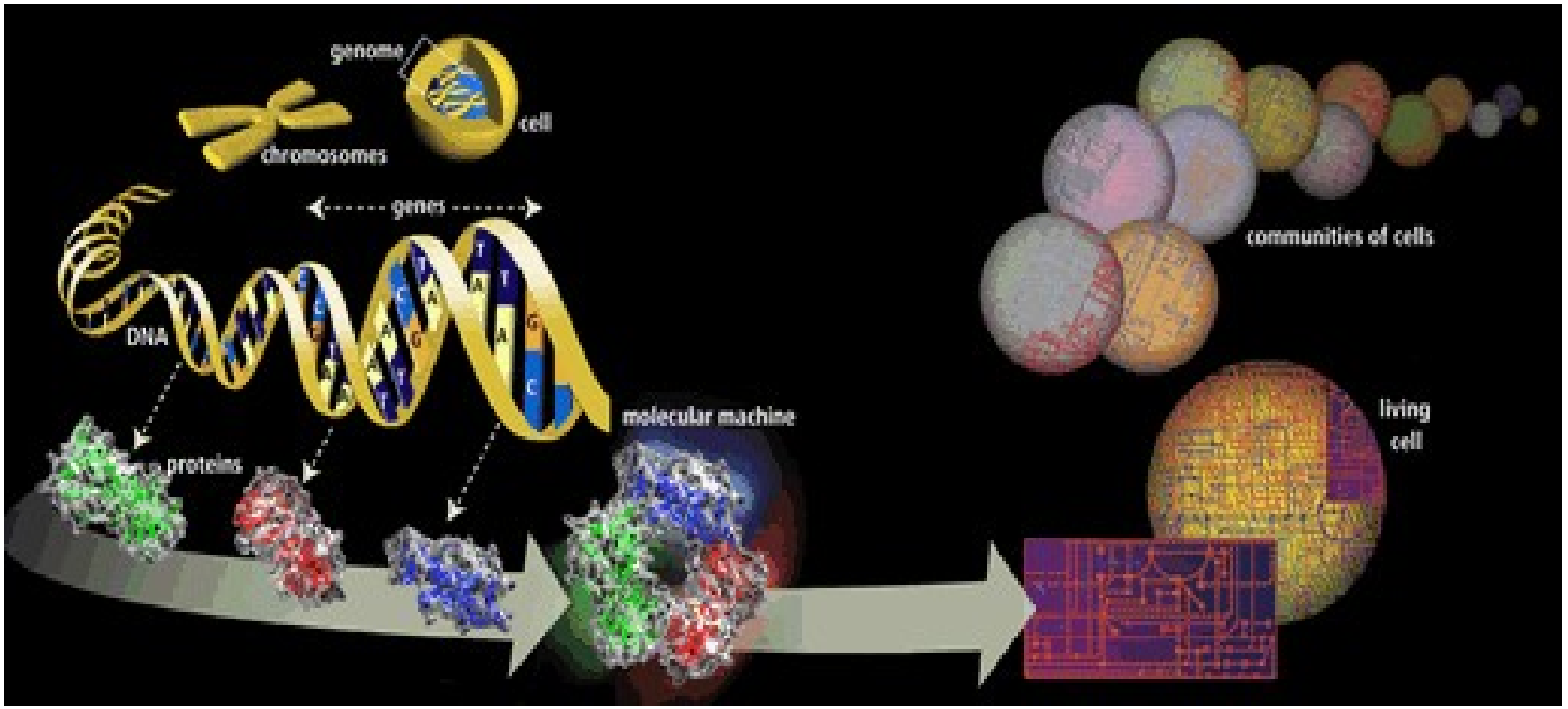
Comparative Modeling



Comparative Modeling: Learning Goals

- Main Steps in Comparative Modeling for Structure Prediction
  - Sequence alignment and template(s) selection
  - Model building from template(s)
  - Dealing with loops and side chains to complete the structure
  - Final refinement of the completed structure
- Hands-on modeling with SwissModel/DeepView
- Other applications of comparative modeling
  - Domain prediction
  - Binding sites and interaction interface prediction
  - Structural motif prediction

What is Comparative Modeling?

- Comparative modeling is modeling of the unknown based on comparison to what is known

- In the context of modeling or computing the structure s_x assumed by a sequence x of amino acids:
  - Structure is a function of sequence: So, s_x = f(x)
  - The function f encodes how the sequence x determines the structure s_x
  - Given another protein of sequence y and known structure s_y, we can infer: **IF** x $\cong$ y **THEN** s_x $\cong$ s_y
    - It is important that x and y be *similar enough*
    - An important question: how similar?

Comparative Modeling: Some Terminology

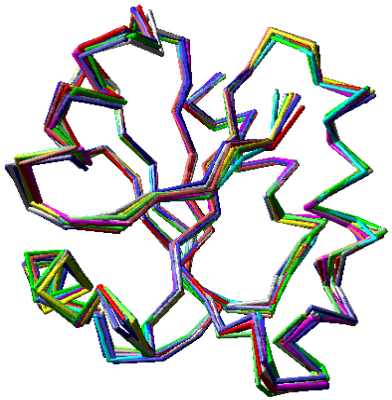
- The protein of unknown structure is the *query* or the *target*
- The protein of known structure whose sequence is similar to that of the target is the *template*
- The process of inferring the coordinates for the target is called *model building*
- Comparative modeling builds the model, completes it, refines it, and then evaluates it

Why Use Comparative Modeling?

- Structures of proteins in a given functional family are more conserved than their sequences
- About a third of all sequences assume known structures
- The number of unique protein folds is limited
  - If not applicable to yield a high-resolution structure, comparative modeling can at least yield the fold for a sequence
- Currently, comparative modeling is both faster and more accurate (as long as the sequence identity is high) than ab initio or de novo methods for structure prediction

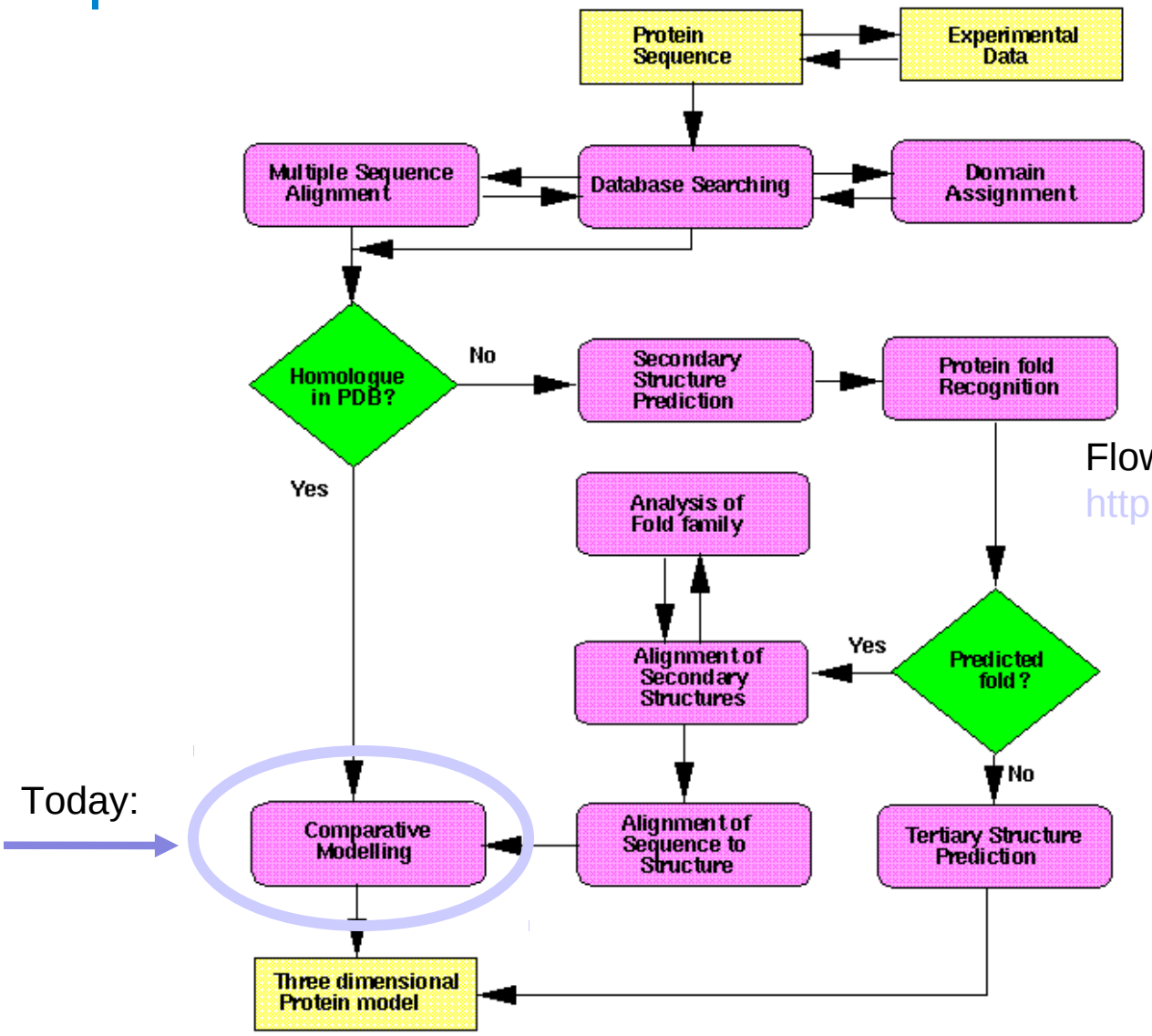
When to Use Comparative Modeling?

- How similar do x and y have to be to infer that the structure assumed by the sequence x is similar to that assumed by the sequence y?
- Statistical analysis of sequences with known structure reveals:
  - Sequences with no less than 50% sequence identity assume structures no more than 1 A different in lRMSD
  - Minimum sequence identity for structural similarity: 25-30%



Higher than 30% sequence identity often results in very similar structures

Comparative Modeling in Structure Prediction



Flowchart by Rob Russell
http://speedy.embl-heidelber

A Simplistic View of Comparative Modeling

(Query) SQKKSIQFHWKNSNQIKILGNQGSFLTKGP

(Template) GQSLTLTLESPPGSSPSVQCRSPRGKNI

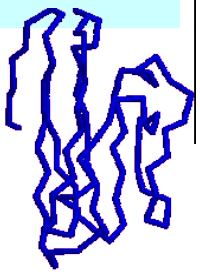
Sequence Alignment Program

ELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGP

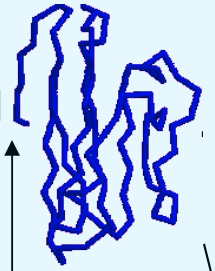
---GQSLTLTLESPPGSSPSVQCRS----

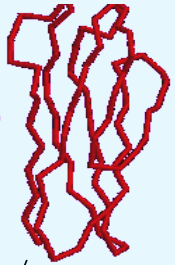
P-----RGKNI--

Build a model structure for the query sequence with the coordinates of the template, as guided by the sequence alignment



Model

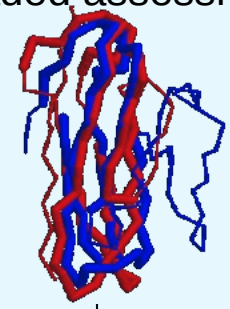




Actual Structure

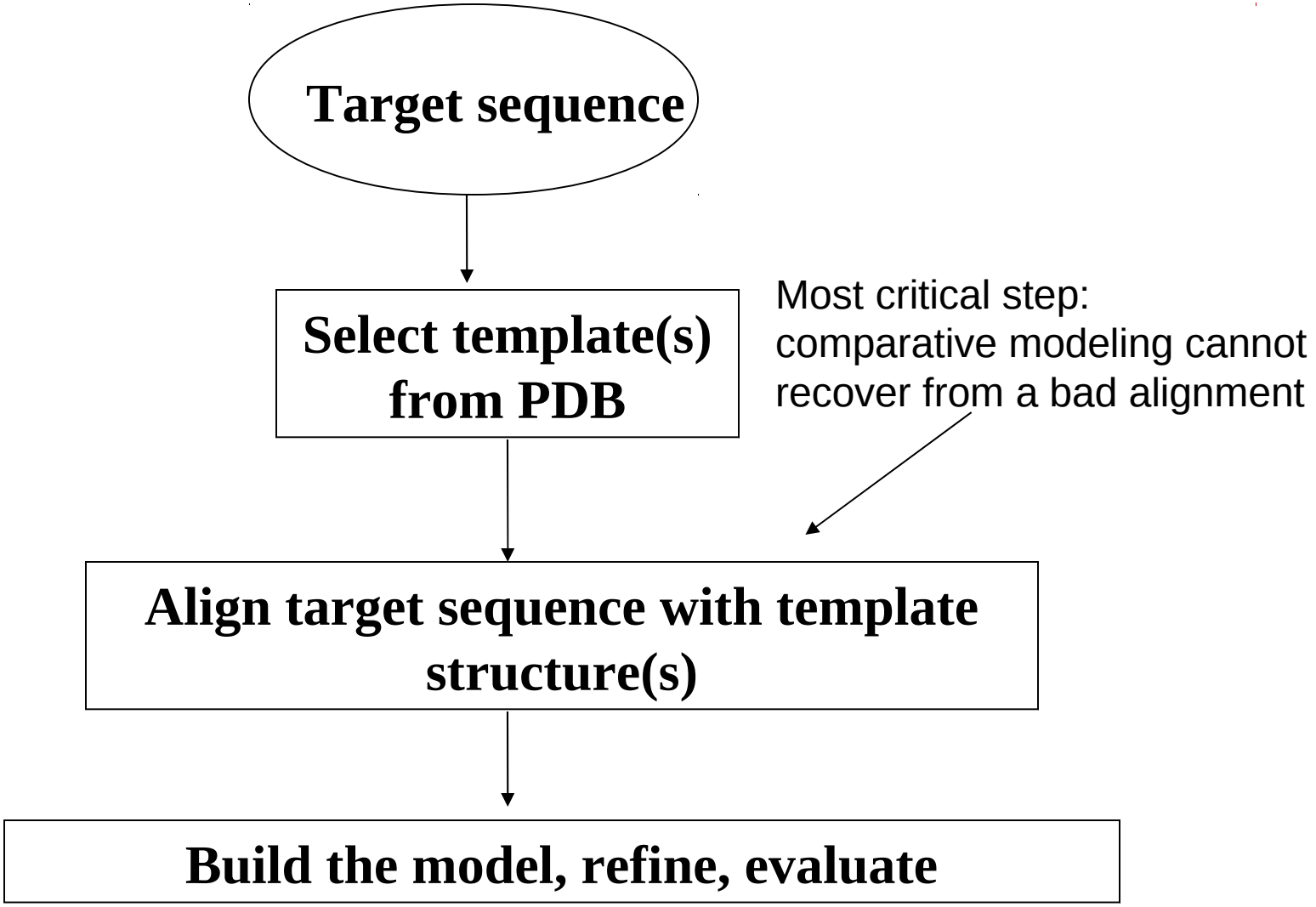
Structural Superimposition and Alignment

Dreaded assessment!



Evaluate the Model using RMSD, LGScore, GDT_TS Score etc

A Simplistic View of Comparative Modeling



A More Detailed View of Comparative Modeling

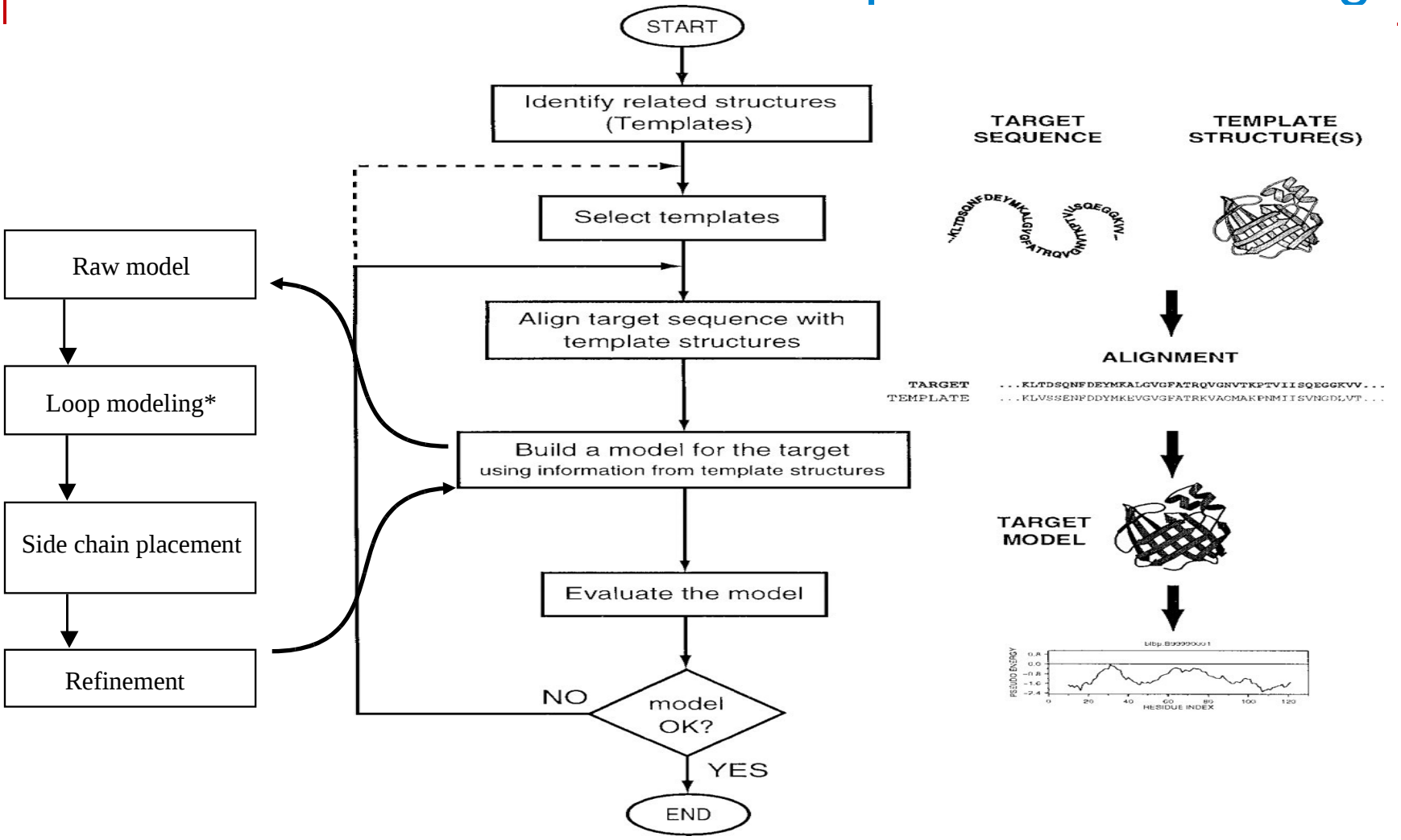


Figure 5.1.1 from M. A. Marti-Renom and A. Sali "Modeling Protein Structure from Its Sequence" *Current Prototocols in Bioinformatics (2003).* 5.1.1-5.1.32

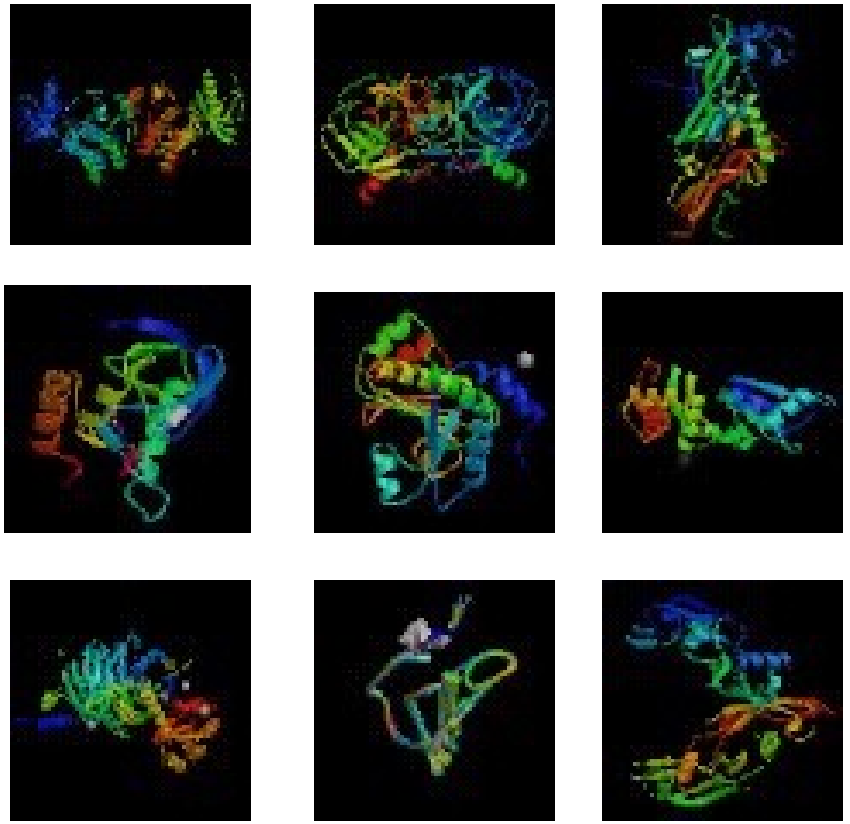
An Algorithmic View of Comparative Modeling

- **Step 1:** Query a database of protein sequences with known structures with the target sequence, focusing on those with >= 30% seq. identity to the target sequence
- **Step 2:** Align obtained sequences to target to choose templates
- **Step 3:** Identify structurally conserved (SC) and variable (SV) regions
- **Step 4:** Generate coordinates for the core region of the target
- **Step 5:** Complete the structure of the target
  - a) generate coordinates for loop regions
  - b) generate coordinates for side-chains
- **Step 6:** Refine the completed structure using energy minimization
- **Step 7:** Validate/evaluate completed structure

Comparative Modeling: Step 1 – Query PDB

PRTEINSEQENCEPRTEINSEQUENC
EPRTEINSEQNCEQWERYTRASDFHG
TREWQIYPASDFGHKLMCNASQERWW
PRETWQLKHGFDSADAMNCVCNQWER
GFDHSDASFWERQWK

Query Sequence



PDB

Comparative Modeling: Step 1 – Query PDB

PRTEINSEQENCEPRTEINSEQUENC
EPRTEINSEQNCEQWERYTRASDFHG
TREWQIYPASDFGHKLMCNASQERWW
PRETWQLKHGFDSADAMNCVCNQWER
GFDHSDASFWERQWK

Query Sequence

PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQNCEQWERYTRASDFH
G
TREWQIYPASDFGHKLMCNASQERW
W
PRETWQLKHGFDSADAMNCVCNQWE
R
GFDHSDASFWERQWK
PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQQWEWEWQWEWEQWEW
EWQRYEYEWQWNCEQWERYTRASDF
HG
TREWQIYPASDWERWEREWRFDSFG

Hit #1

PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQNCEQWERYTRASDFH
G
TREWQIYPASDFGHKLMCNASQERW
W
R
GFDHSDASFWERQWK
PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQQWEWEWQWEWEQWEW
EWQRYEYEWQWNCEQWERYTRASDF
HG
TR

PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQNCEQWERYTRASDFH
G

Hit #2

PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQNCEQWERYTRASDFH
G
TREWQIYPASDFGPRTEINSEQENC
EPRTEINSEQUENCEPRTEINSEQN
CEQWERYTRASDFHGTREWQIYPAS
DFG
TREWQIYPASDFGPRTEINSEQENC
EPRTEINSEQUENCEPRTEINSEQN
CEQWERYTRASDFHGTREWQ
PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQNCEQWERYTRASDFH
G
TREWQIYPASDFG

PRTEINSEQENCEPRTEINSEQUEN
C
EPRTEINSEQNCEQWERYTRASDFH
G
TREWQIYPASDFGPRTEINSEQENC

PDB

Comparative Modeling: Step 2 – Alignment

	G	E	N	E	T	I	C	S
G	10	0	0	0	0	0	0	0
E	0	10	0	10	0	0	0	0
N	0	0	10	0	0	0	0	0
E	0	0	0	10	0	0	0	0
S	0	0	0	0	0	0	0	10
I	0	0	0	0	0	10	0	0
S	0	0	0	0	0	0	0	10

	G	E	N	E	T	I	C	S
G	60	40	30	20	20	0	10	0
E	40	50	30	30	20	0	10	0
N	30	30	40	20	20	0	10	0
E	20	20	20	30	20	10	10	0
S	20	20	20	20	20	0	10	10
I	10	10	10	10	10	20	10	0
S	0	0	0	0	0	0	0	10

Dynamic Programming

Comparative Modeling: Step 2 – Alignment

Goal: Find a template or templates

pairwise sequence alignment - finds high homology sequences BLAST

http://www.ncbi.nlm.nih.gov/BLAST/

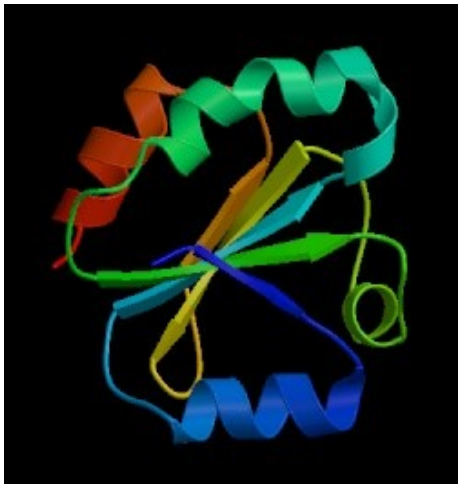
Improved Multiple sequence alignment methods improves sensitivity - remote homologs PSIBLAST, CLUSTAL

Comparative Modeling: Step 2 – Alignment

- Pairwise sequence alignment: BLAST, FASTA, WU-BLAST, SSEARCH, and more
- Available as web servers and standalone software
- Basic functionality needed: compare target sequence with sequences in the PDB (or any other comprehensive structural database)
- BLAST scans the sequence for 3-letter words (wmers, where w = 3) and expands alignments from 3-mers
- Statistically significant alignments are hits
- Templates are hits with no lower than 30% sequence identity

Comparative Modeling: Step 2 – Alignment

```
Query  ACDEFGHIKLMNPQRST--FGHQWERT-----TYREWYEG
Hit #1 ASDEYAHLRILDPQRSTVAYAYE--KSFAPPGSFKWEYEA
Hit #2 MCDEYAHIRLMNPERSTVAGGHQWERT----GSFKEWYAA
```



Hit #1

Hit #2

Comparative Modeling: Step 2 – Alignment

- Global (Needleman-Wunsch) alignment can be used
- Alignment is the most crucial step, as comparative modeling can never recover from a bad alignment
- A small error in the alignment can translate to a significant error in the reconstructed model
- Multiple sequence alignments (that also align the templates to one another) is often better than pairwise alignment

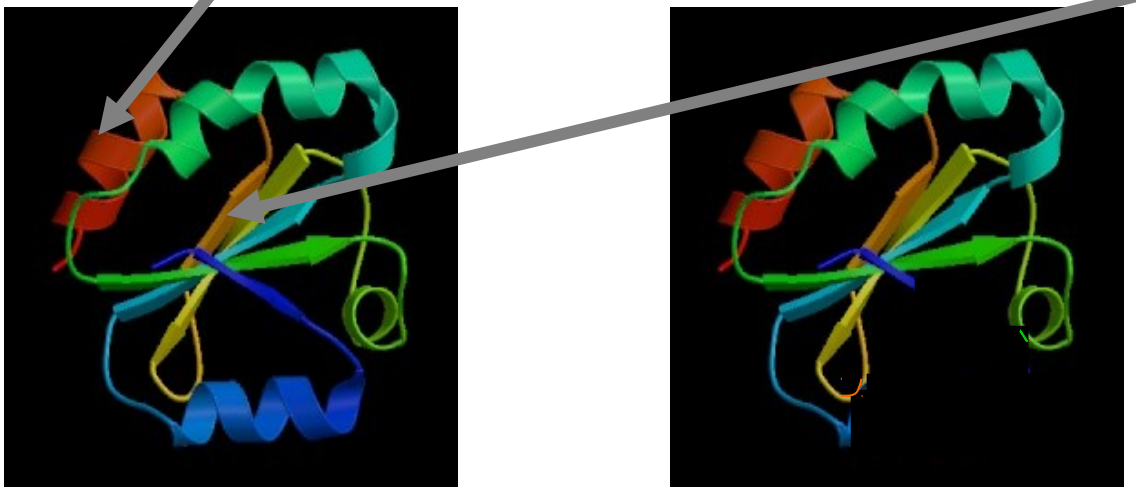
Comparative Modeling: Step 2 – Get Templates

- A good template is closest to the target in terms of subfamilies
- This means that high overall sequence similarity is needed
- The template environment like pH, ligands, etc., should be the same as that of the target
- The quality of the experimentally-available template structure - the resolution, R-factor, etc. - should be high
- When choosing a template for a protein-ligand model, it is preferred that the template have the same ligand
- When modeling an active site – a high resolution template structure with ligand is important

Comparative Modeling: Step 3 – SCRs

```
Query   ACDEFGHIKLMNPQRST--FGHQWERT----TYREWYEG
Hit #1  ASDEYAHLRILDPQRSTVAYAYE--KSFAPPGSFKWEYEA
Hit #2  MCDEYAHIRLMNPERSTVAGGHQWERT----GSFKEWYAA
        HHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCBBBBBBBBB
```

SCR #1 *SCR #2*



Hit #1 **Hit #2**

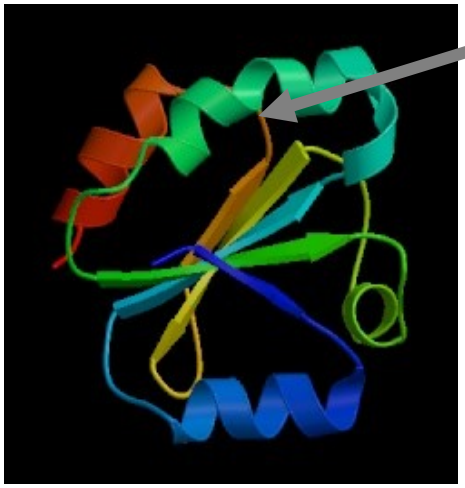
Comparative Modeling: Step 3 – SCRs

- SCRs correspond to the most stable structures or regions (usually in the interior/core) of the protein
- SCRs also often correspond to sequence regions with the lowest level of gapping and highest level of sequence conservation
- SCRs are often the secondary structures

Comparative Modeling: Step 3 – SVRs

```
Query   ACDEFGHIKLMNPQRST--FGHQWERT----TYREWYEG
Hit #1  ASDEYAHLRILDPQRSTVAYAYE--KSFAPEGSFKWEYEA
Hit #2  MCDEYAHIRLMNPERSTVAGGHQWERT----GSFKEWYAA
        HHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCBBBBBBBBB
```

SVR (loop)

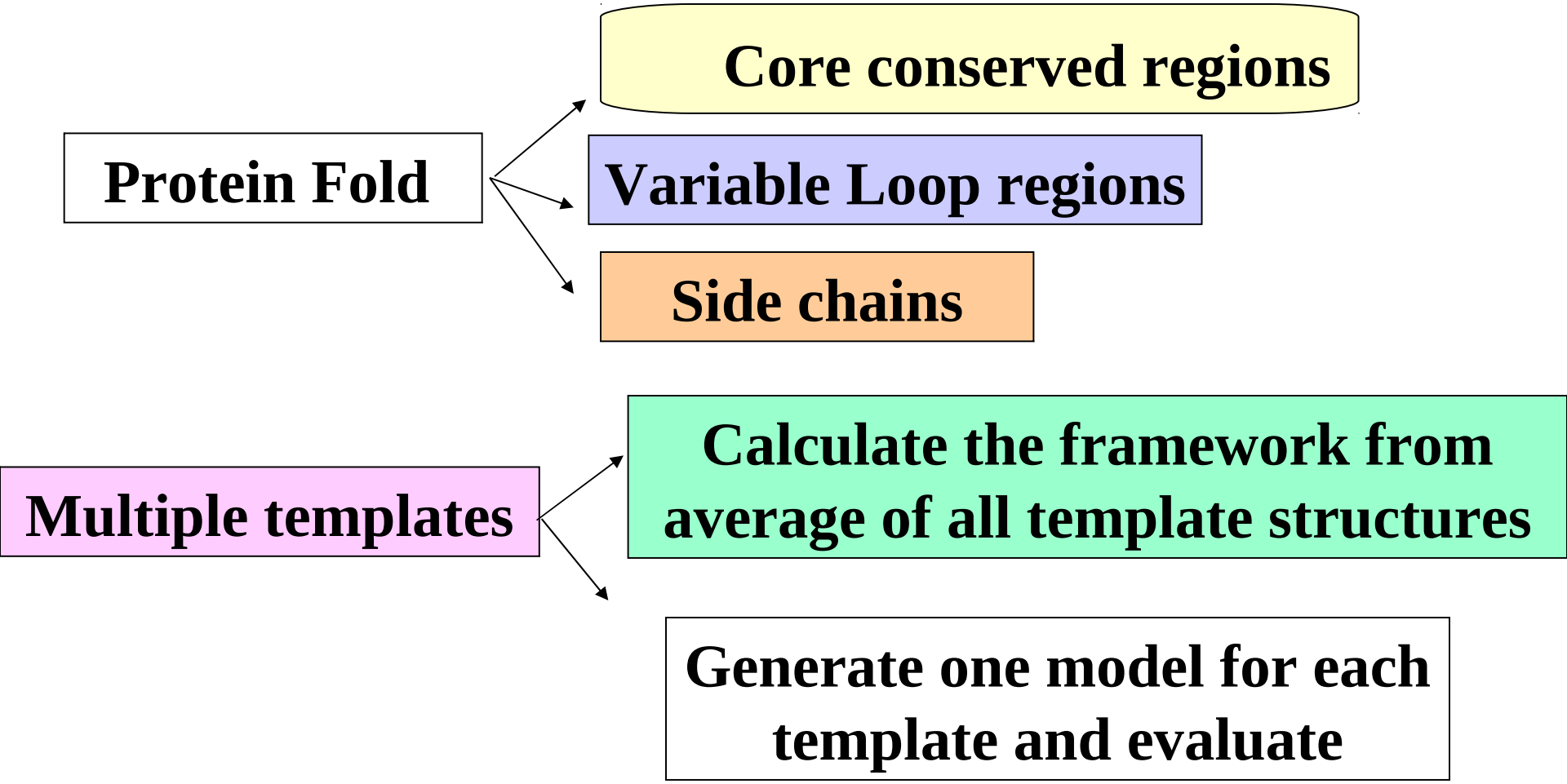


Hit #1 Hit #2

Comparative Modeling: Step 3 – SVRs

- SVRs correspond to the least stable or the most flexible regions (usually in the exterior/surface) of the protein
- SVRs correspond to sequence regions with the highest level of gapping and lowest level of sequence conservation
- SVRs are usually loops and turns

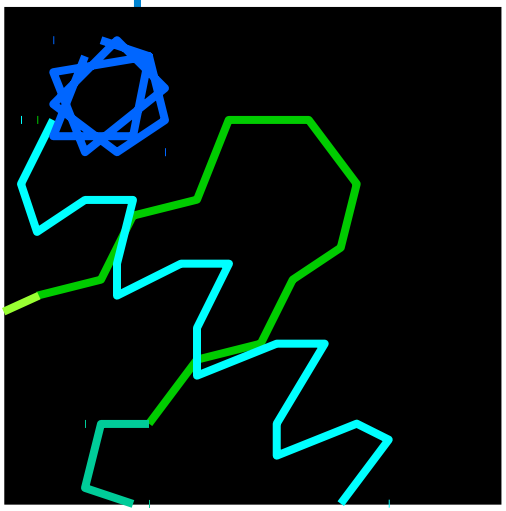
Comparative Modeling: Step 4 – Threading



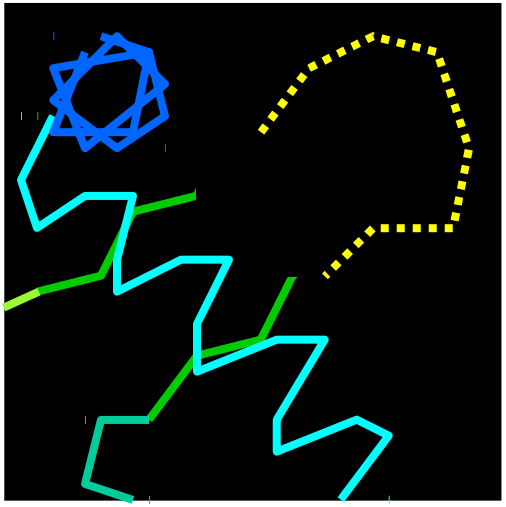
Comparative Modeling: Step 4 – Core Coords

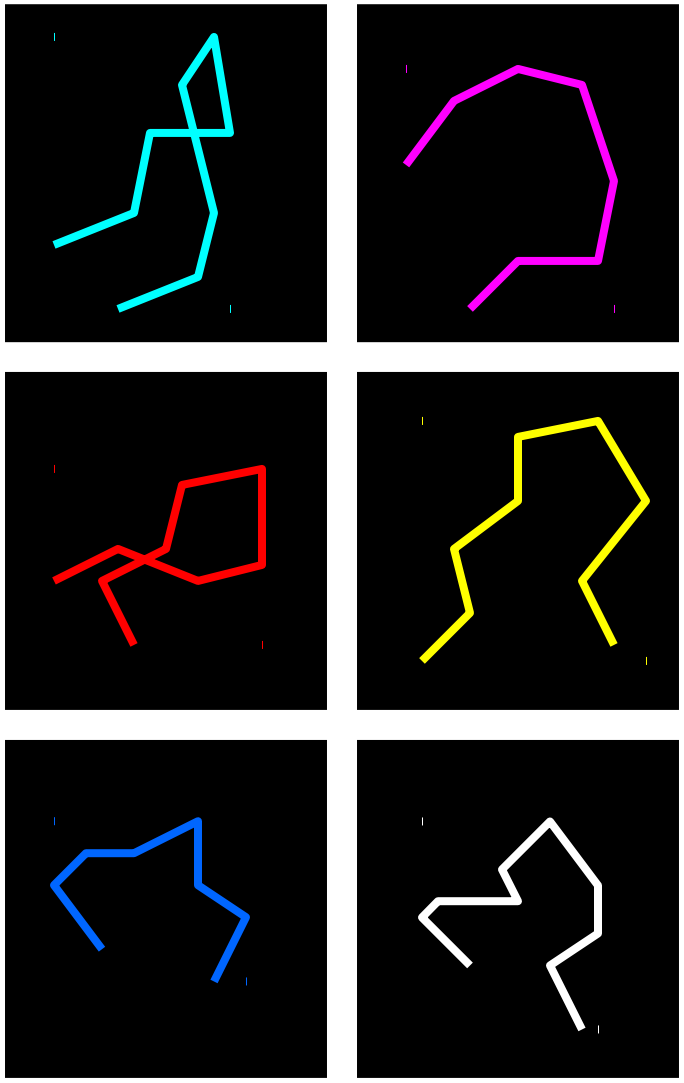
- For *identical* amino acids, just transfer all atom coordinates (x, y, z) to the query protein (both backbone and side-chain atoms are identical)
- For *similar* amino acids, transfer the backbone coordinates and replace side-chain atoms while respecting χ angles
- For *different* amino acids, one can only transfer the backbone coordinates (x, y, z) to query sequence
- The side chains of different amino acids have to be built at a later stage, when completing the model

Comparative Modeling: Step 5 – SVRs



Query **FGHQWERT**
Hit #1 **YAYE--KS**





Comparative Modeling: Step 5 – Loop Modeling

Loops result from substitutions and indels in same family

Mini protein folding problem- loops can be very long in membrane proteins

Ab-initio methods - generate various *random* loop conformations and evaluate/score

Compare the loop sequence string to PDB, get hits, and evaluate/score

Some comparative modeling methods have fewer loops to be added because of extensive multiple sequence alignment of profiles

Comparative Modeling: Step 5 – Loop Modeling

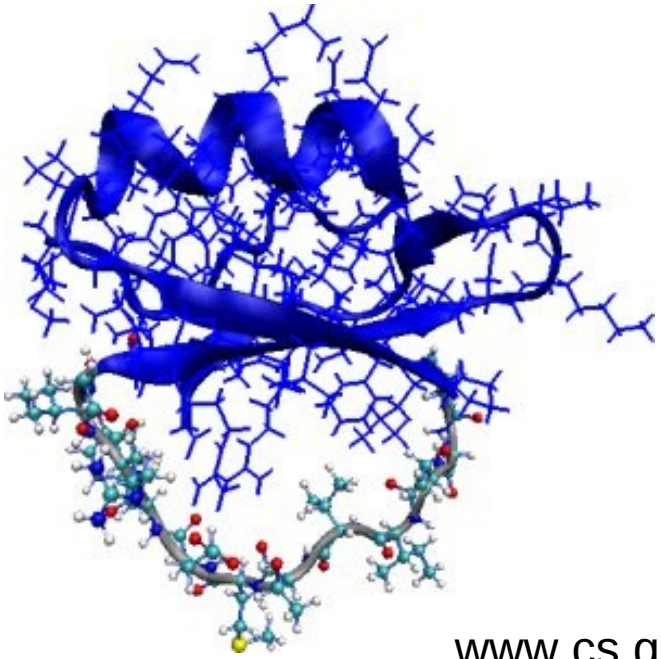
- *Ab-initio* loop modeling - Monte Carlo, Monte Carlo with simulated annealing, MD, main chain dihedral angle search biased with the data from PDB, inverse kinematics-based, etc.

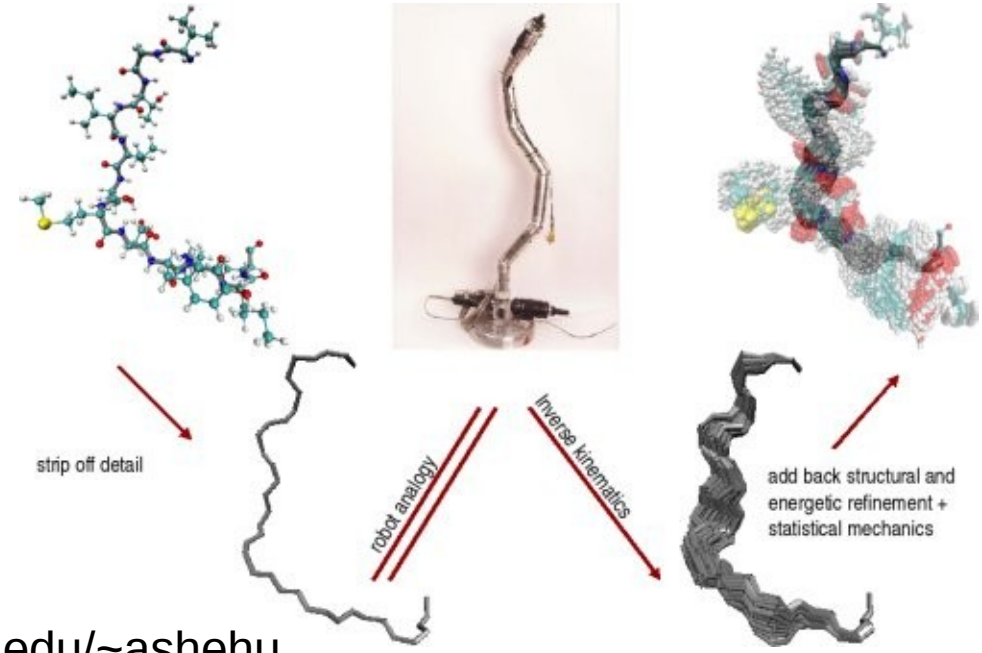
- Energy functions used: physics-based (CHARMM, AMBER, etc.) or knowledge-based (built with statistics obtained from PDB)

- *Ab-initio* methods - allow simultaneous addition of several loops, which yields a conformational ensemble view for the loop

Comparative Modeling: Step 5 – Loop Modeling

- **Step 5:** Loop modeling to complete the model
  - Ab-initio loop modeling is an active research area
  - One can start at a robotics-inspired approach (Shehu et al. Proteins 2006)





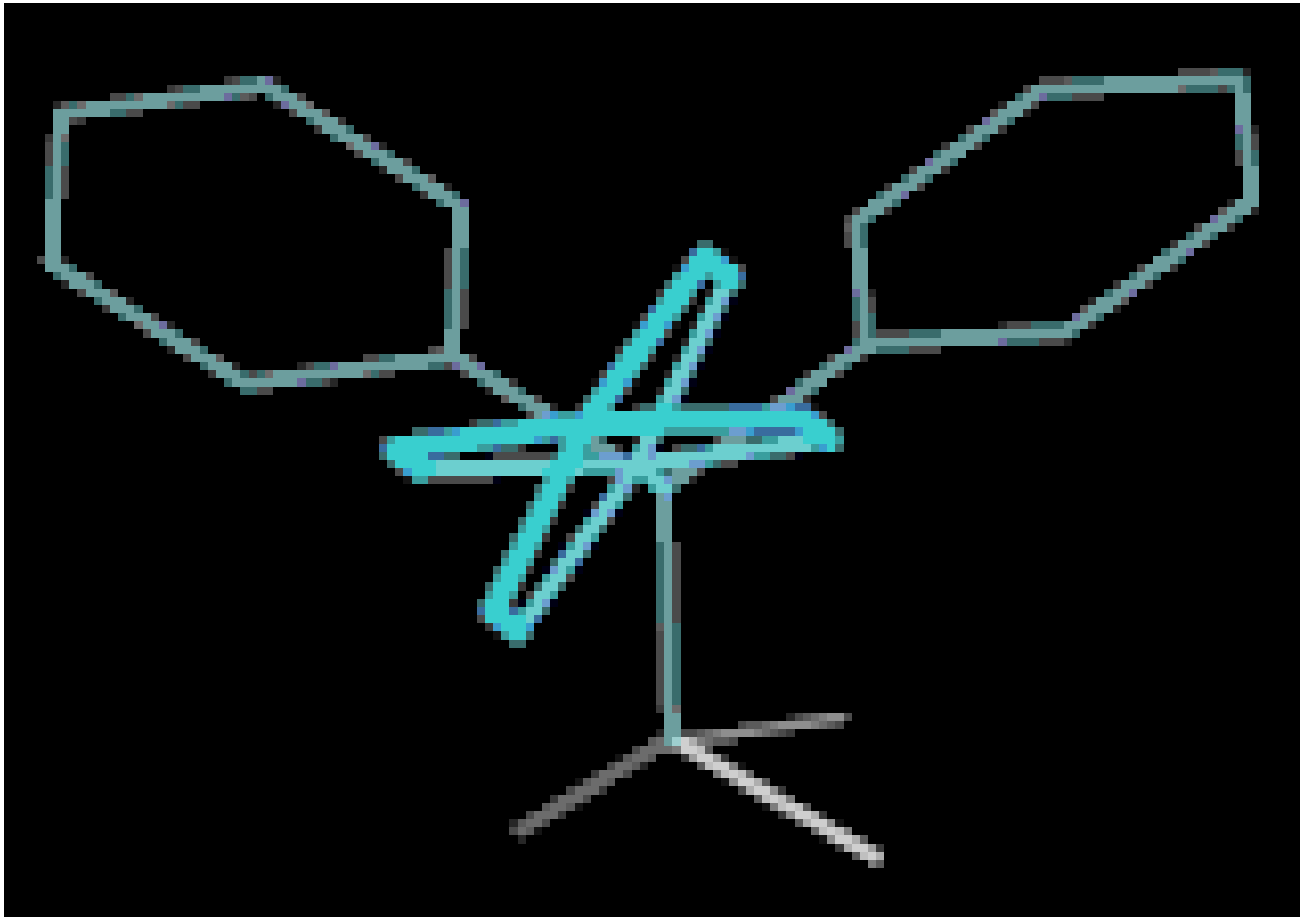
Comparative Modeling: Step 5 – Loop Modeling

- Loops can be modeled with comparative modeling as well
- Comparison of loop sequence to PDB - pick sequence hits
- Sort hits through geometric restraints (often, termini constraints) or a more detailed energy function
- Works well for special loops like β-hairpins
- Normally only the main chain is modeled
- Limited by the lengths of the loops it can complete

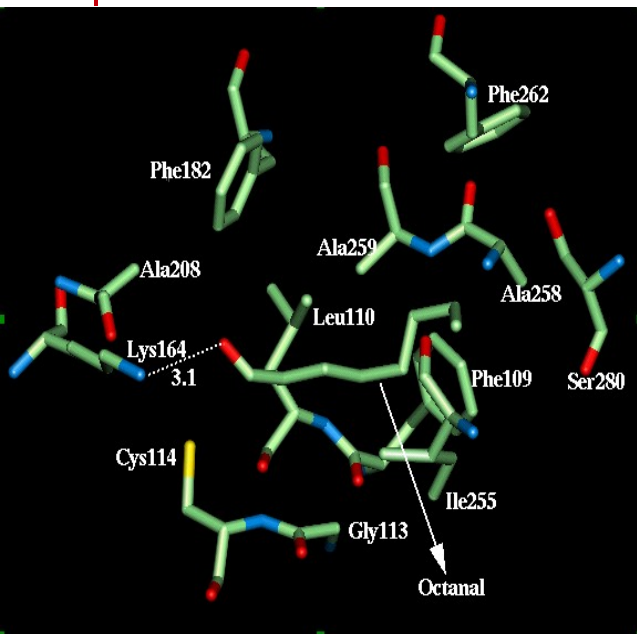
Comparative Modeling: Step 5 – Loop Modeling

- Must match the desired number of residues in the loop
- Must match the Ca-Ca distance (<0.5 Å)
- Must not collide into other parts of protein (no Ca-Ca distance <3.0 Å)
- Preceding and following Ca's (3 residues) from loop should match well with corresponding Ca coordinates in template structure
- Loop placement and positioning is done using superposition algorithms: loop fits are evaluated using RMSD calculations and standard collision checking.
- If no "good" loop is found, resort to ab-initio: compute loops using randomly generated backbone dihedral angles (with termini constraints)

Comparative Modeling: Step 5 – Side chains



Comparative Modeling: Step 5 – Side chains



Side chain packing is critical to studying ligand binding to proteins

Side Chain Builder

- Predicted from similar structures
- Built from steric and energetic considerations with robust conformational search algorithms
- Combination of rotamer library and energy evaluations

Comparative Modeling: Step 5 – Side chains

- Rotamer libraries have been created (statistical analysis of torsion angles of side chains of amino acids) from structures in the PDB
- Two main effects in predicting side chains
  - how it sits on top of the main chain(very critical)
  - continuous variation of side chain torsions - only 6% varies +/- 40^{o} from the rotamer libraries
- Current techniques predict side chains up to 1.5 Å accuracy for a fixed backbone for the core residues
- Solvation and H-bond terms are very important in modeling exposed side chains

Comparative Modeling: Step 5 – Side chains

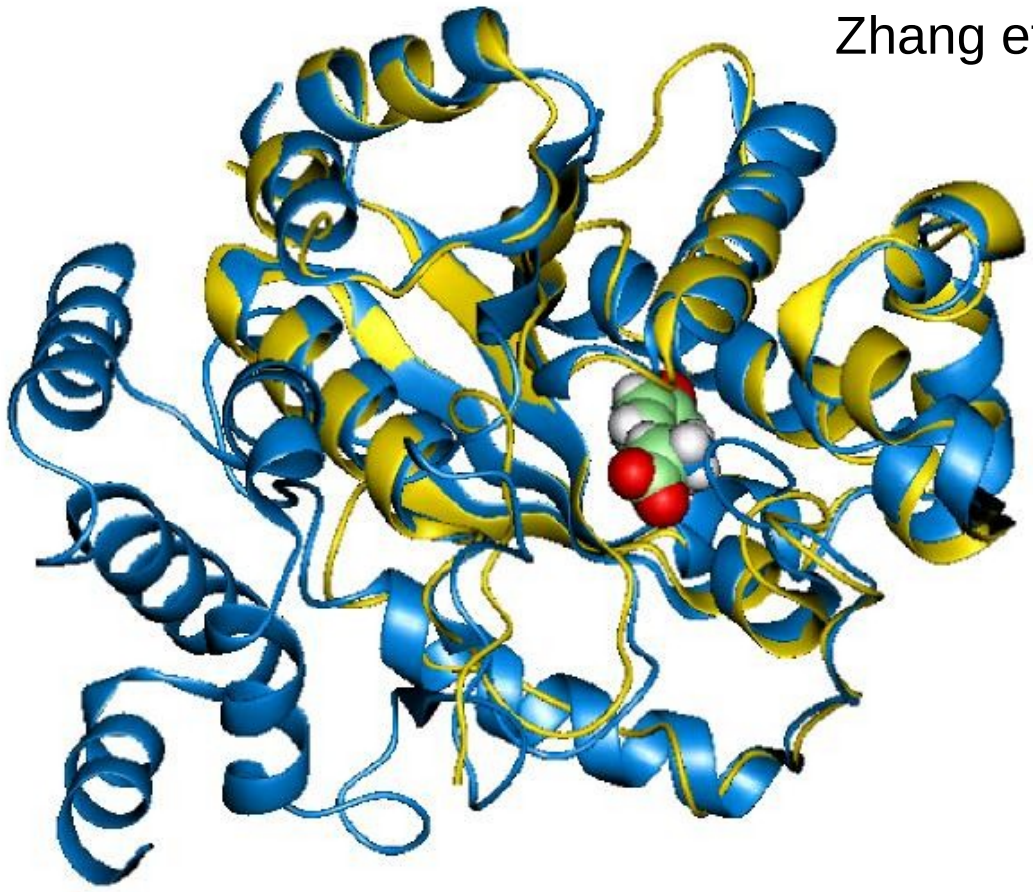
- Methods available - SCWRL, SCAP, MODELLER, Insight II, WhatIf, SCREAM, and recent ones like Heath et al. Proteins 2007
- Evaluation of all three methods for backbone < 4 Å lRMSD to native all work equally - 50% of χ_1 and 35% of χ_2 and χ_3
- SCREAM – works well – accurate energy analysis – computationally intensive
- Heath et al. 2007 reconstructs thoursands of all-atom conformations from backbone conformations

Comparative Modeling: Step 6 – Refinement

- Completed model may undergo a short energy minimization
- Physics-based or knowledge-based functions may be used
- The minimization may help remove steric clashes and improve favorable interactions in the completed model prior to the final evaluation of the built model for the target

Comparative Modeling: Showcase

Predicted Structure of *M.Jann* TyrRS
Zhang et al PNAS



Yellow - *M. Jann.* Predicted

Blue - *Bac. Thermophillus* 4ts1

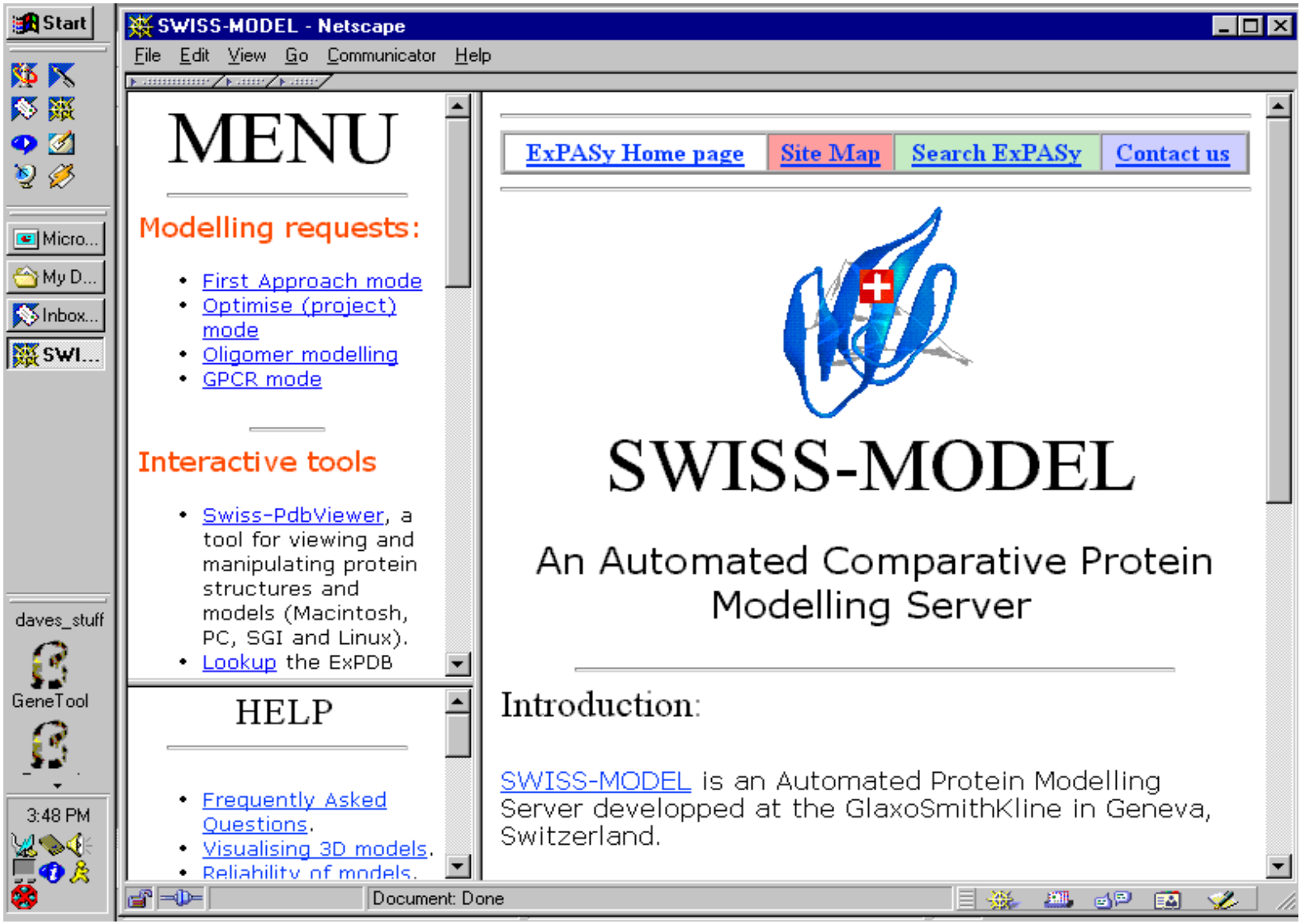
Comparative Modeling: Step 7 – Evaluation

- Given a predicted structure
  - Ramachandran plot - allowed regions for backbone torsions
  - Calculate the H-bond network - use Quanta or WhatIf or InsightII - normally calculated for heteroatoms with distance cutoff
  - Identify hydrophobic residues on the surface
  - Identify hydrophilic residues in the core - satisfied with salt bridges?
  - Voids in the core are typically small two water cluster?

Comparative Modeling on the WWW

- Prior to 1998, comparative modeling could only be done with commercial software or command-line freeware
- The process was time-consuming and labor-intensive
- The past few years has seen an explosion in automated web-based comparative modeling servers
- Now anyone can!

Comparative Modeling Hands-on



http://www.expasy.ch/swissmod/SWISS-MODEL.html

Comparative Modeling with SwissModel

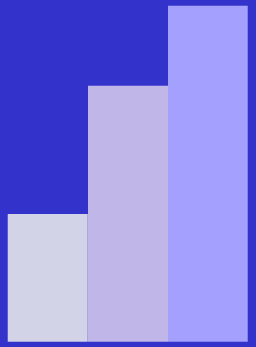
- **Swiss-Model - an automated homology modeling server developed at Glaxo Welcome Experimental Research in Geneva.**

  http://www.expasy.ch/swissmod/

- **Closely linked to Swiss-PdbViewer, a tool for viewing and manipulating protein structures and models.**

- **Will likely take 24 hours to get results returned!**

Comparative Modeling with SwissModel



1) Search for suitable templates
2) Check sequence identity with target
3) Create ProModII jobs
4) Generate models with ProModII
5) Energy minimization with Gromos96

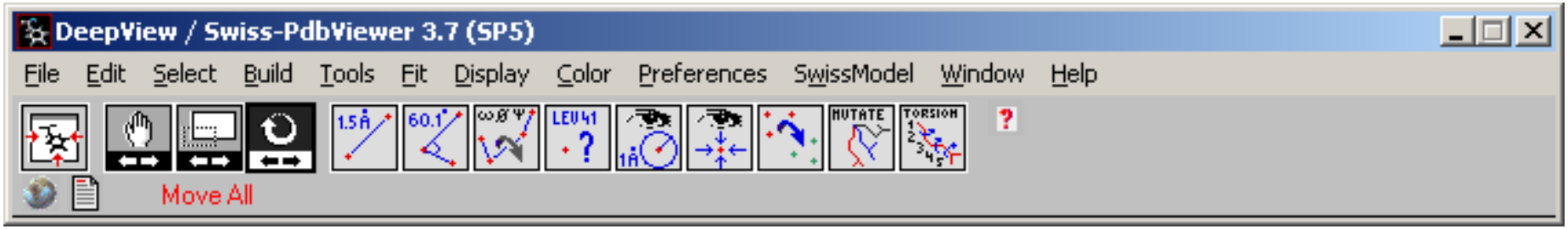
First approach mode (regular)
First approach mode (with user-defined template)
Optimize mode

Comparative Modeling with SwissModel

Program	Database	Action
BLASTP2	ExNRL-3D	Find homologous sequences of proteins with known structure
SIM	--	Will select all templates with sequence identities above 25%
--	--	Generate ProModII input files
ProModII	ExPDB	Generate all models
Gromos96	--	Energy minimization of all models

Comparative Modeling with SwissModel

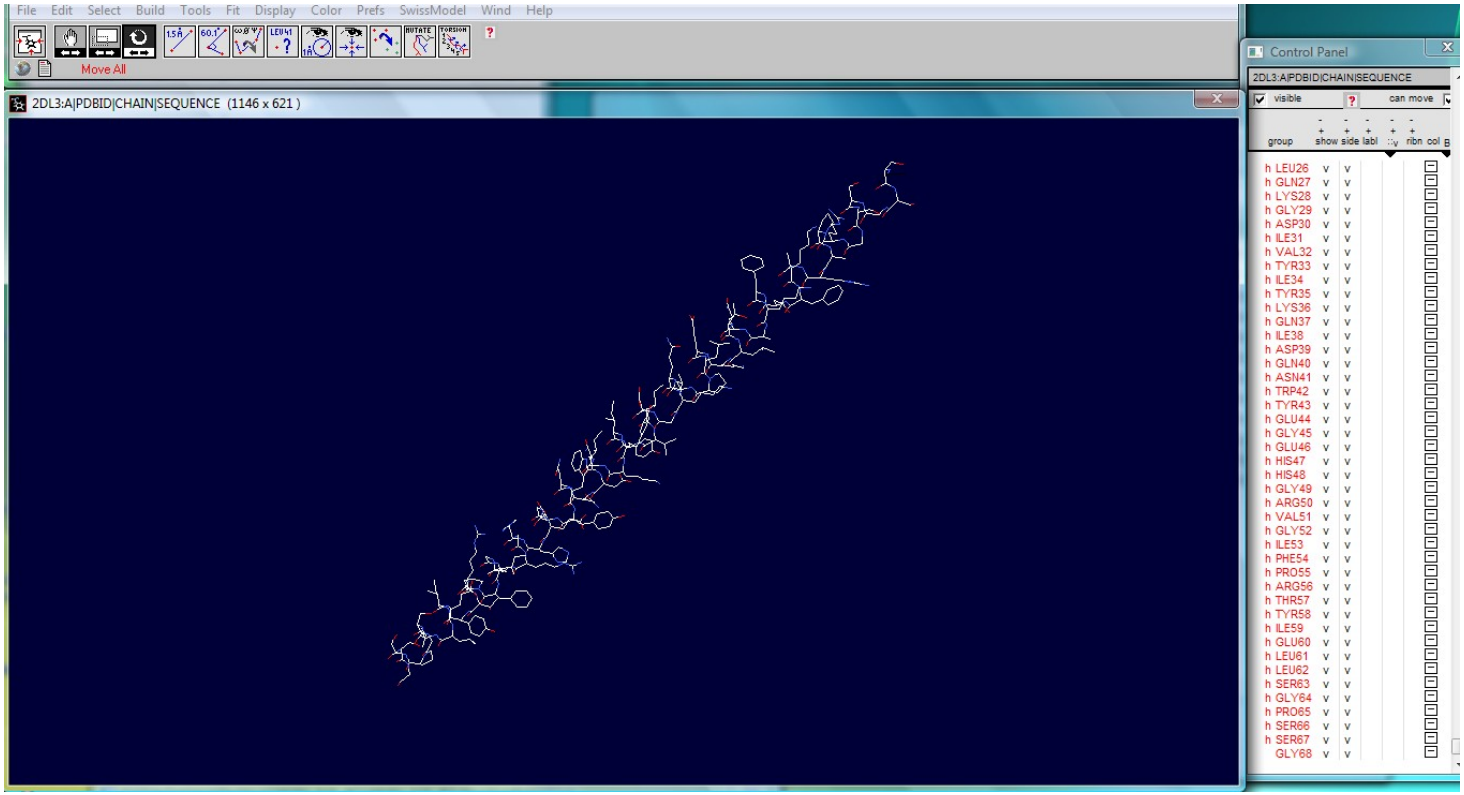
- Illustrate all these steps through a case-study:
  - Modeling the structure of the SH3 sequence
  - The Swiss-Model web server or *DeepView* program can be used
    - Web server: http://swissmodel.expasy.org/workspace/
    - Program: http://spdbv.vital-it.ch/download.html



- Target sequence: Save the sequence of PDB ID 2DL3 to a FastA file
- Housekeeping: under **Preferences/Swiss-Model** menu enter name and e-mail address. Make sure that **Preferences/Network Server** is www.swissmodel.unibas.ch Port: 27000

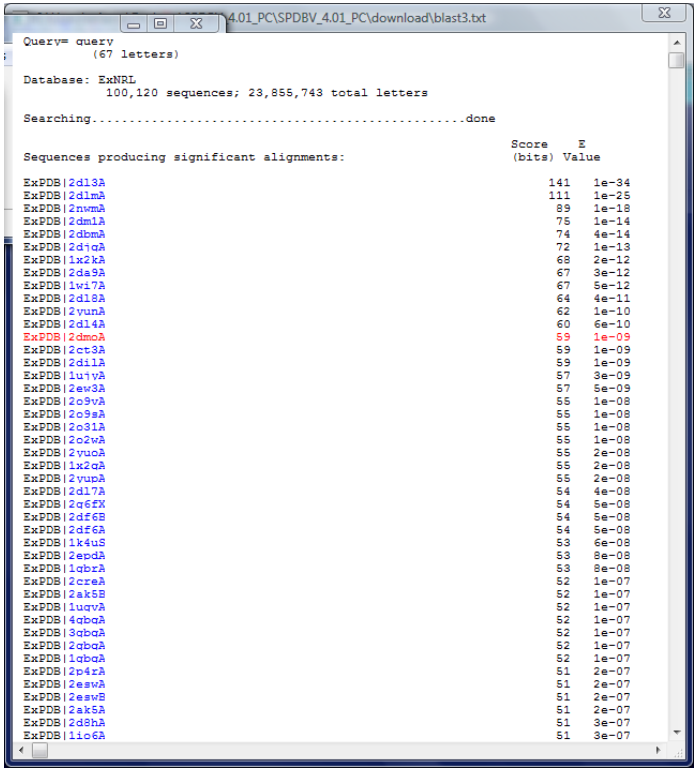
Comparative Modeling with DeepView

- **SwisModel/Load Raw Sequence** to load the target FastA file
  - A default long alpha helix is associated with target



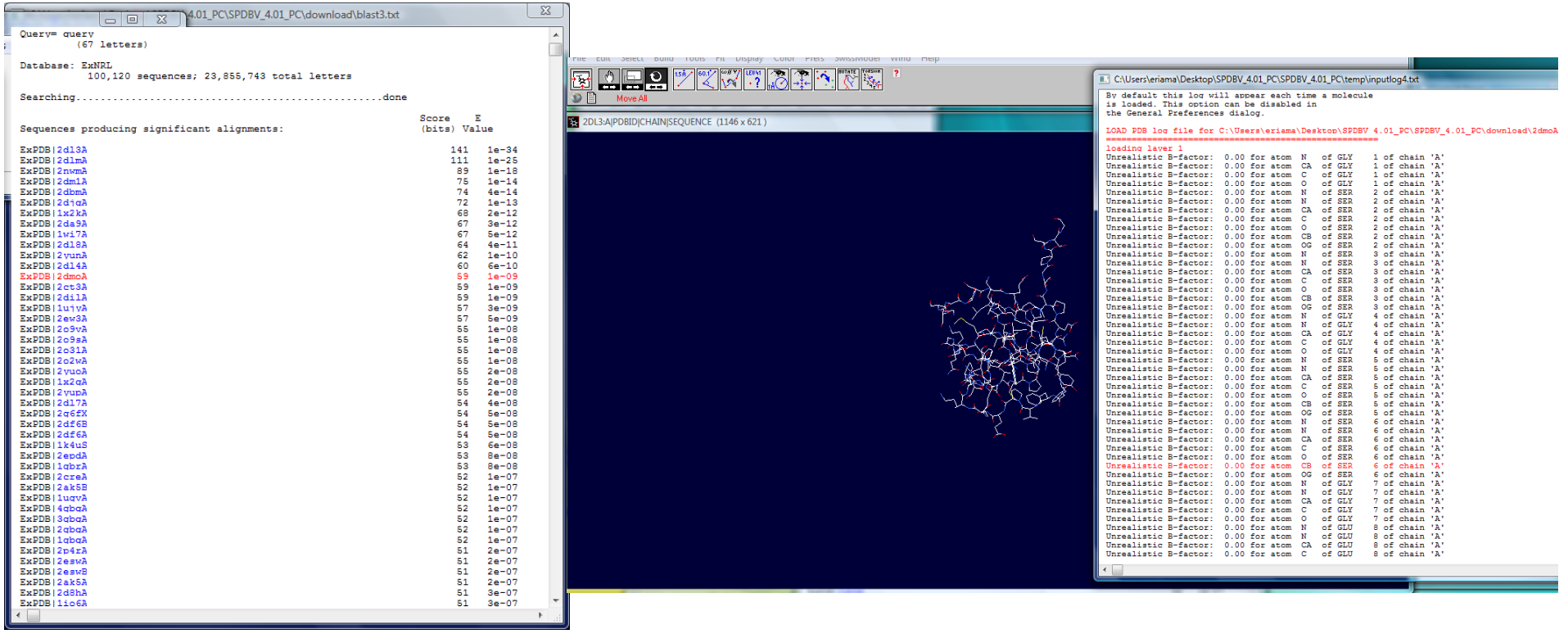
Comparative Modeling with DeepView

- **Select/all** residues in your target sequence then use **Edit/Blast Selection vs. ExPDB** to generate a BLAST search
  - Sequences of potential templates are identified and ranked
  - One can load one or more templates to build the target structure



Comparative Modeling with DeepView

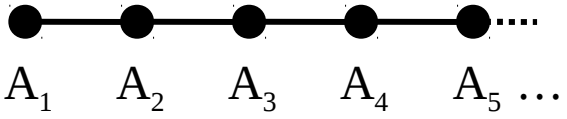
- **Select/all** residues in your target sequence then use **Edit/Blast Selection vs. ExPDB** to generate a BLAST search
  - Loading one template does not need step 2: no comparison of templates



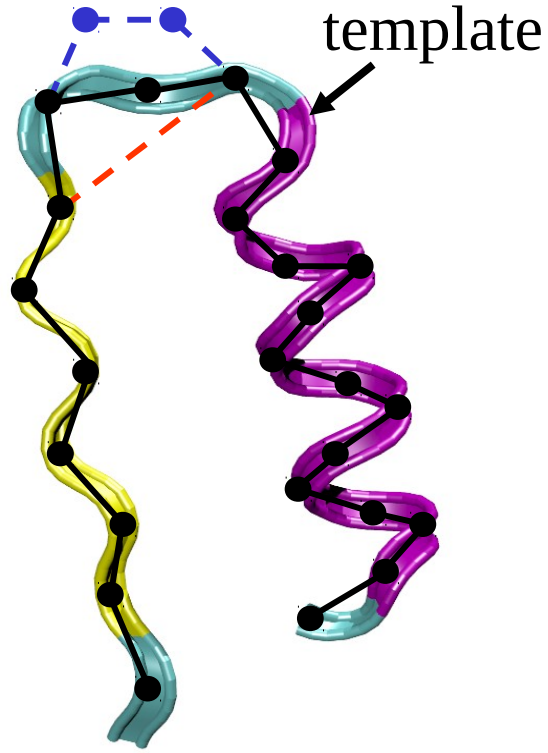
Comparative Modeling with DeepView

- Loading more than one template
  - One can look at the energy and identify high-energy regions
  - One can look at the gaps or no alignments over entire sequence length

Target sequence



Alignment between target(s) and scaffold(s)



$$H = E_{contact} + E_{profile} + E_{H-bonds} + E_{gap}$$

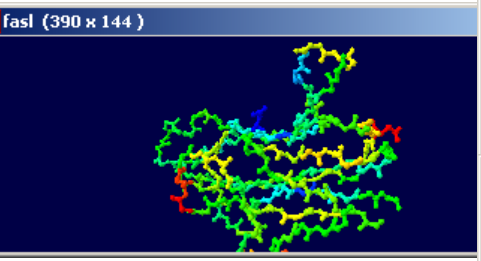
$$E_{profile} = \sum_{i}^{n} \gamma^{(p)}\left(A_i, SS_i, SA_i\right)$$

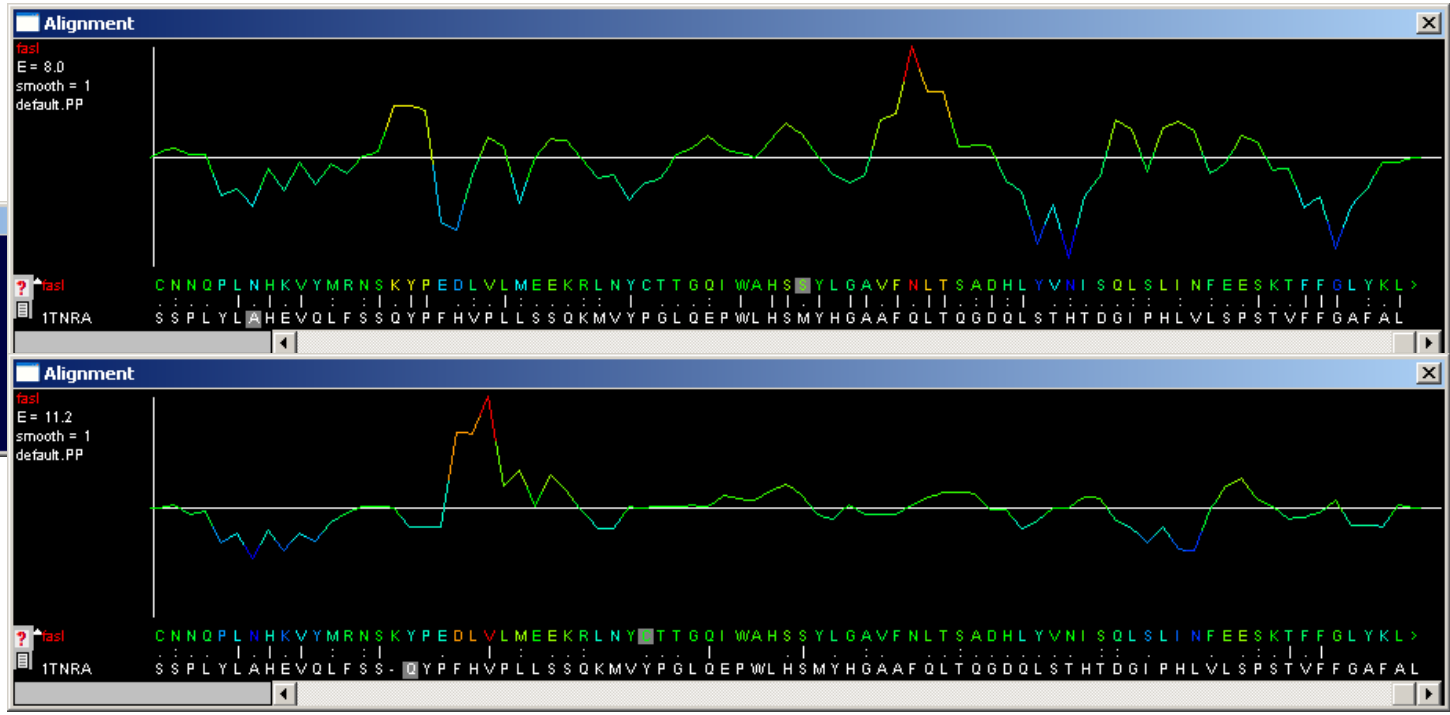
$$E_{contact} = \sum_{i,j} \sum_{k=1}^{2} \gamma_k^{(ct)}\left(A_i, A_j\right) * U\left(r_k - r_{ij}\right)$$

*R. Goldstein, Z. Luthey-Schulten, P. Wolynes (1992, PNAS), K. Koretke et.al. (1996, Proteins)

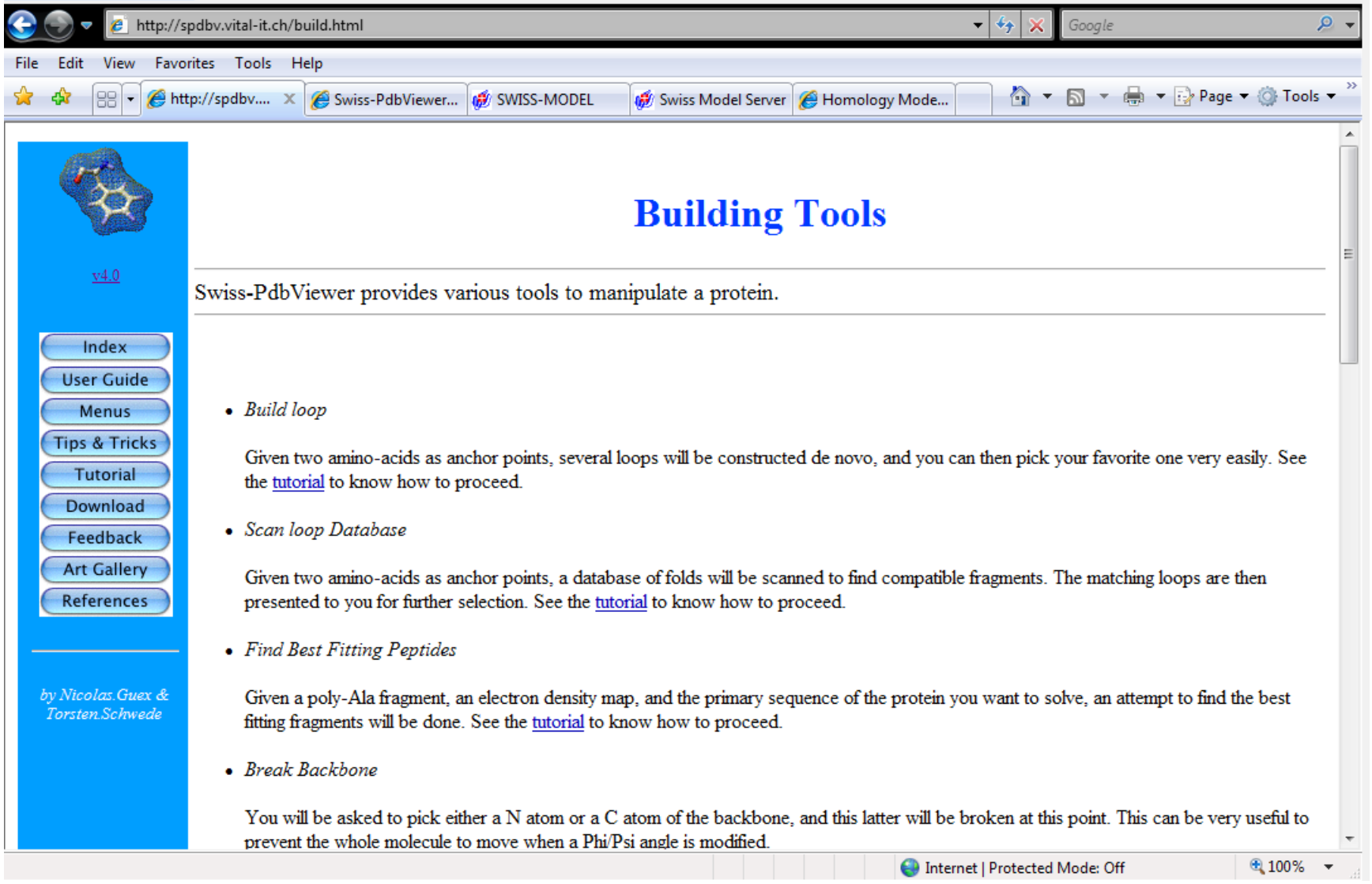
Comparative Modeling with DeepView

- Loading more than one template
  - One can look at the threading energy and identify high-energy regions
  - One can look at the gaps or no alignments over entire sequence length

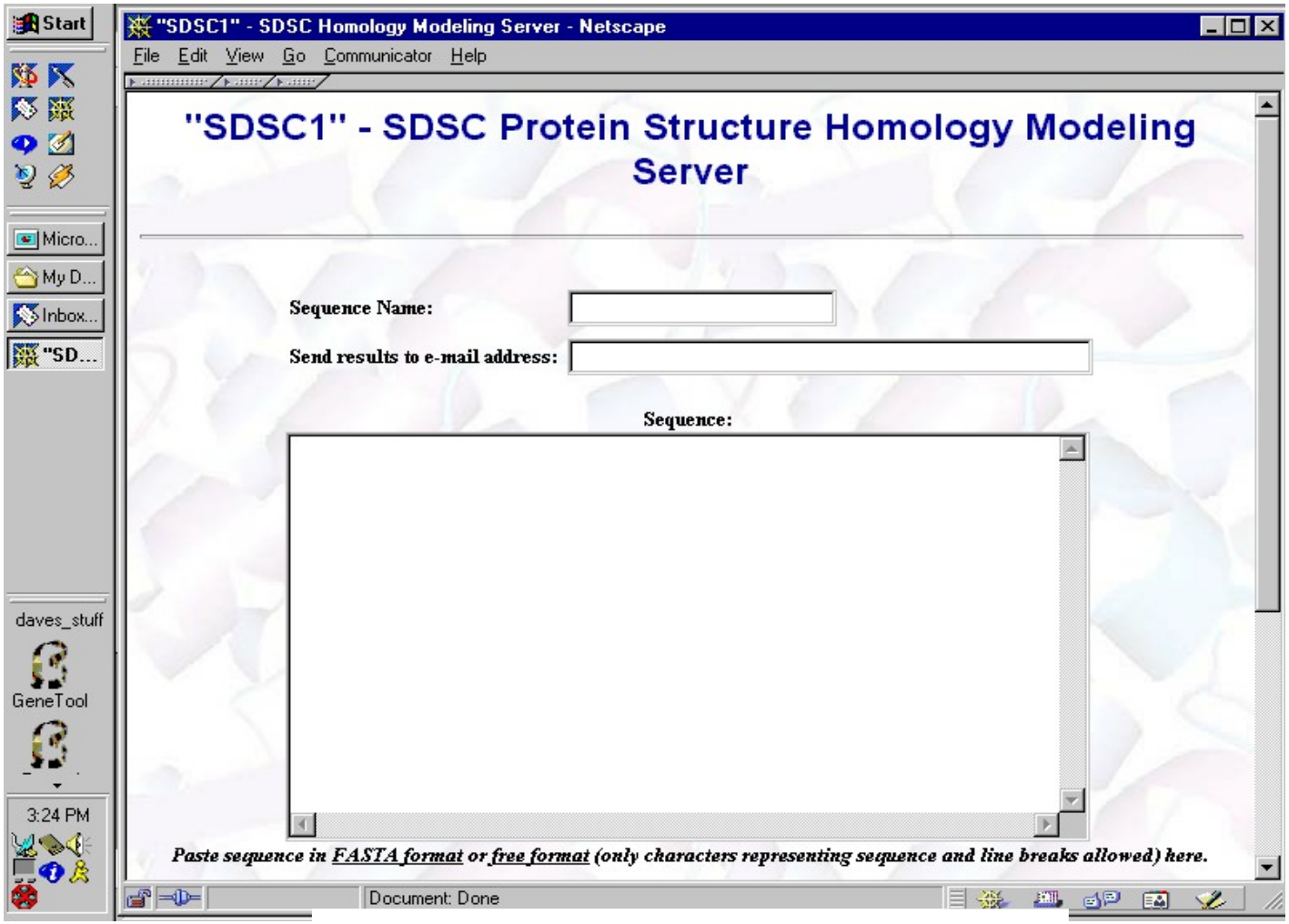




Completing Model with DeepView

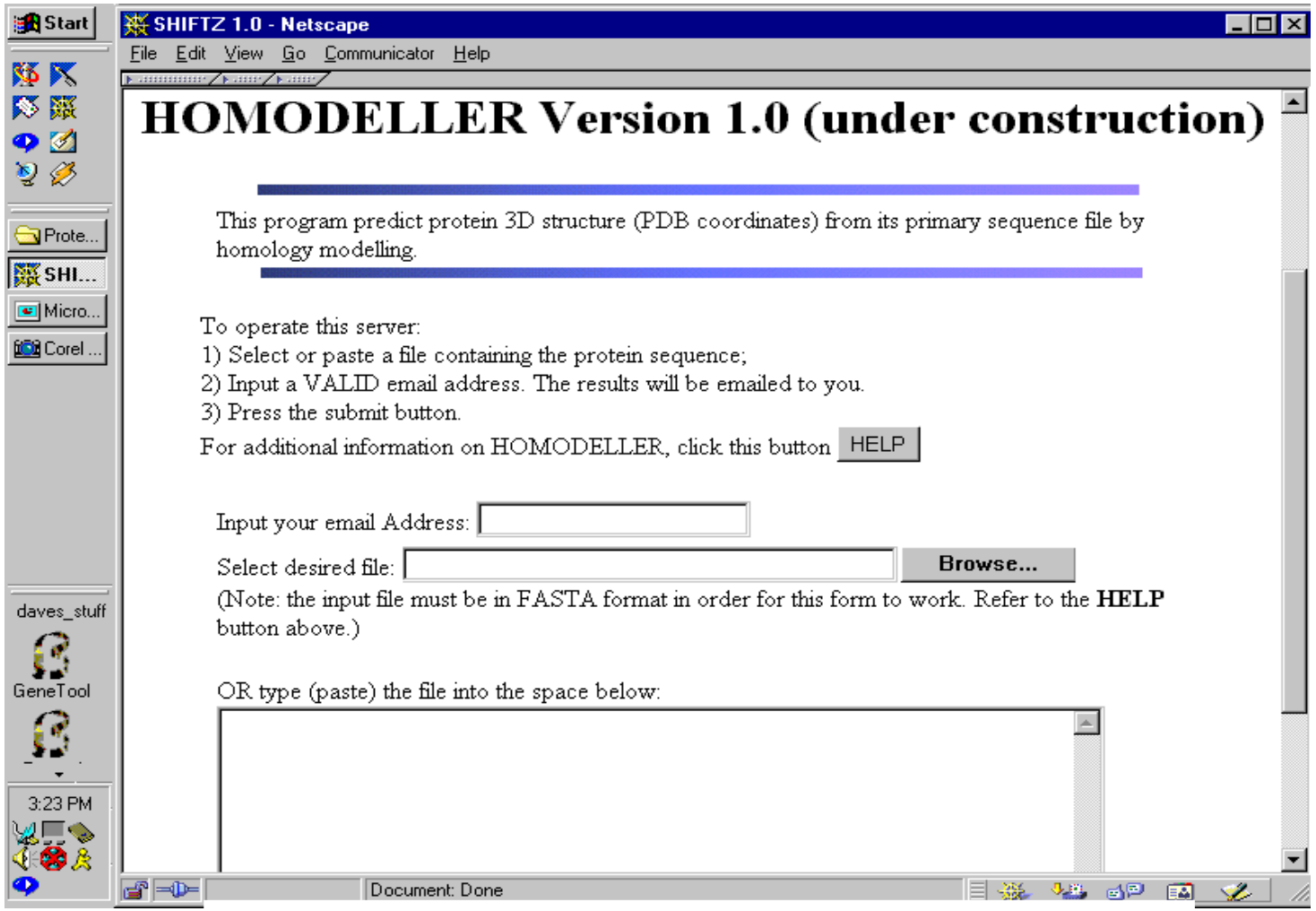


Comparative Modeling with SDSC



http://cl.sdsc.edu/hm.html

Comparative Modeling with Homodeller



http://redpoll.pharmacy.ualberta.ca

Comparative Modeling: Tools

BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
FastA	http://www.ebi.ac.uk/fasta33/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
PredictProtein	http://www.predictprotein.org/
123D; SARF2; PDP	http://123d.ncifcrf.gov/
GenTHREADER	http://bioinf.cs.ucl.ac.uk/psipred/
UCLA-DOE	http://fold.doe-mbi.ucla.edu/

##	Scoring			Rmsd	N_{algn}	N_g	$\%_{seq}$	Query	Target (PDB entry)				
	Q	P	Z					$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	1.00	48.7	21.0	0.00	343	0	100	100	1nj4:A	100	343	☐	CRYSTAL STRUCTURE OF A DEACYLATION-DEFECTIVE MUTANT OF PENICILLIN-BINDING PROTEIN 5 AT 1.9 A RESOLUTION
2	0.92	40.1	19.0	0.54	337	4	99	96	1nzu:A	89	347	☐	WILD-TYPE PENICILLIN-BINDING PROTEIN 5 FROM E. COLI MODIFIED BY BETA-MERCAPTOETHANOL
3	0.92	40.1	19.0	0.61	333	3	100	96	1hd8:A	96	337	☐	CRYSTAL STRUCTURE OF A DEACYLATION-DEFECTIVE MUTANT OF PENICILLIN-BINDING PROTEIN 5 AT 2.3 A RESOLUTION
4	0.89	35.1	17.7	0.82	338	3	100	92	1sdn:A	92	347	☐	CRYSTAL STRUCTURE OF A DEACYLATION-DEFECTIVE MUTANT OF PENICILLIN-BINDING PROTEIN 5 MODIFIED BY MERCURY
5	0.83	35.0	17.7	0.91	331	3	100	96	1z6f:A	86	354	☐	CRYSTAL STRUCTURE OF PENICILLIN-BINDING PROTEIN 5 FROM E. COLI IN COMPLEX WITH A BORONIC ACID INHIBITOR
6	0.79	29.5	16.2	1.16	331	3	100	96	1nzo:A	86	352	☐	THE CRYSTAL STRUCTURE OF WILD TYPE PENICILLIN-BINDING PROTEIN 5 FROM E. COLI
7	0.36	7.5	9.9	2.03	256	15	26	72	1xp4:A	82	369	☐	CRYSTAL STRUCTURE OF A PEPTIDOGLYCAN SYNTHESIS REGULATORY FACTOR (PBP3) FROM STREPTOCOCCUS PNEUMONIAE

Welcome to Secondary Structure Matching

a tool for protein structure comparison

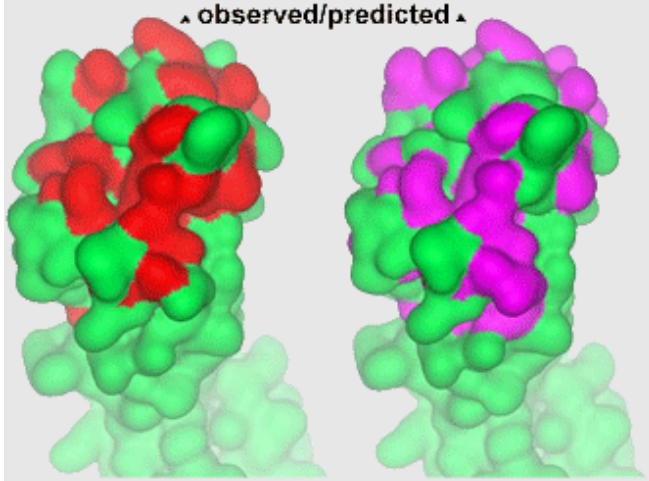
Now available for download and in-house installation, check here .

Please cite SSM when you publish or otherwise present any results obtained from this service.

Current version: 2.34 built 07/08/2007

Comparative Modeling: Tools

BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
FastA	http://www.ebi.ac.uk/fasta33/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
PredictProtein	http://www.predictprotein.org/
123D; SARF2; PDP	http://123d.ncifcrf.gov/
GenTHREADER	http://bioinf.cs.ucl.ac.uk/psipred/
UCLA-DOE	http://fold.doe-mbi.ucla.edu/



About PredictProtein

PredictProtein is a service for sequence analysis, structure and function prediction. When you **submit** any protein sequence PredictProtein retrieves similar sequences in the database and predicts aspects of protein structure and function (**more**)

News

10/04/2007

PredictProtein upgrade PredictProtein has been upgraded! We have integrated many new methods into the system; you can now get predictions of disordered/natively unstructured regions, of inter-residue contacts, of domain assignments, and protein-protein interaction and protein-DNA binding residues unsing our newer and **faster** server. The new system requires **registration**. Note that registration is free, and the use of PredictProtein remains free for academia.

Comparative Modeling: Tools

BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
FastA	http://www.ebi.ac.uk/fasta33/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
PredictProtein	http://www.predictprotein.org/
123D; SARF2; PDP	http://123d.ncifcrf.gov/
GenTHREADER	http://bioinf.cs.ucl.ac.uk/psipred/
UCLA-DOE	http://fold.doe-mbi.ucla.edu/

Thread 1D to 3D with 123D+

123D+ combines sequence profiles, secondary structure prediction, and contact capacity potentials to thread a protein sequence through the set of 3D structures

Comparative Modeling: Tools

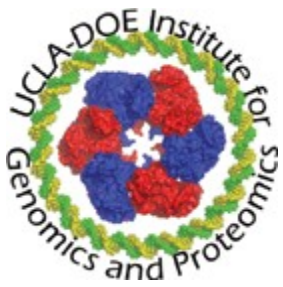
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
FastA	http://www.ebi.ac.uk/fasta33/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
PredictProtein	http://www.predictprotein.org/
123D; SARF2; PDP	http://123d.ncifcrf.gov/
GenTHREADER	http://bioinf.cs.ucl.ac.uk/psipred/
UCLA-DOE	http://fold.doe-mbi.ucla.edu/

Bioinformatics Unit UCL

PSIPRED – for secondary structure prediction
MEMSAT2 - for transmembrane topology prediction
GenTHREADER – fold recognition based on a sequence profile

Comparative Modeling: Tools

BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
FastA	http://www.ebi.ac.uk/fasta33/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
PredictProtein	http://www.predictprotein.org/
123D; SARF2; PDP	http://123d.ncifcrf.gov/
GenTHREADER	http://bioinf.cs.ucl.ac.uk/psipred/
UCLA-DOE	http://fold.doe-mbi.ucla.edu/



Sequence → Structure

Fold Recognition
MOMENT Transmembrane Helix Prediction
Motif-based Fold Assignment
DPANN: Sequence to Structure Alignment
Profile Search Software: Bowie et al. 1991

Comparative Modeling: Tools

Even more…

EMBOSS	http://www.ebi.ac.uk/emboss/align/
Tcoffee	http://www.igs.cnrs-mrs.fr/Tcoffee
ClustalW	http://www.ebi.ac.uk/clustalw/
BCM	http://searchlauncher.bcm.tmc.edu/multi-align/
POA	http://www.bioinformatics.ucla.edu/poa/
STAMP	http://www.ks.uiuc.edu/Research/vmd/
SwissModel	http://www.expasy.org/spdbv/

Comparative Modeling: Conclusions

- Comparative modeling is a key component to structural proteomics
- Allows 3D coordinates of proteins to be generated where no prior experimental structure exists
- Allows hypotheses to be tested and working models to be developed
- Complex process requiring many components (alignment, superposition, energy minimization, structure evaluation)
- Now mostly automated through several freely available web servers
- Reasonably good structures can be generated using known structures with sequence identities >= 25%
- Models based on lower sequence identity are not always valid
- *A good alignment is key to good models*

Comparative Modeling: Additional Applications

Identification of homologous interacting pairs
Search whether two sequences have homologues which form a complex in a database of known structures of complexes.

The alignments are analyzed to check whether contact residues in the known complex are conserved in the alignment.

Identification of structural patterns
Build a library of known protein-protein interfaces from the PDB, defining interfaces as pairs of fragments that are no further than a threshold distance.

Residues that have a high frequency for a given position are considered hotspots.

This library is then used to identify potential interactions between pairs of targets, providing that they have a known structure.

Further Reading on Comparative Modeling for Structure Prediction

Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E., and Thornton, J.M. (1987) Knowledge-Based Prediction of Protein Structures and the Design of Novel Molecules. Nature 326: 347-352.

Fetrow, J.S. and Bryant, S.H. (1993) New Programs for Protein Tertiary Structure Prediction. Bio/Technology 11: 479-484.

Greer, J. (1991) Comparative Modeling of Homologous Proteins. Meth. Enzymol. 202: 239-252.

Johnson, M.S., Srinivasan, N., Sowdhamini, R., and Blundell, T.L. (1994) Knowledge-Based Protein Modeling. Crit. Rev. Biochem. Mol. Biol. 29: 1-68.

Sali, A., Overington, J.P., Johnson, M.S., and Blundell, T.L. (1990) From Comparisons of Protein Sequences and Structures to Protein Modelling and Design. Trends Biochem. Sci. 15: 235-240.

Kabsch, W. and Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers 22: 2577.

Sali, A. and Blundell, T.L. (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. J. Mol. Biol. 234: 779-815.

Luthy, R., Bowie, J.U., and Eisenberg, D. (1992) Assessment of Protein Models with Three-Dimensional Profiles. Nature 356: 83-85.

Sources Cited

- A. Shehu, “Conformational Search for the Protein Native State,” *Protein Structure Prediction*, Wiley Series in Bioinformatics 2009.
- A. Shehu, Ph.D. Defense Talk, 2008.
- S. Sontum, Chem324, Spring 2008, Middlebury College