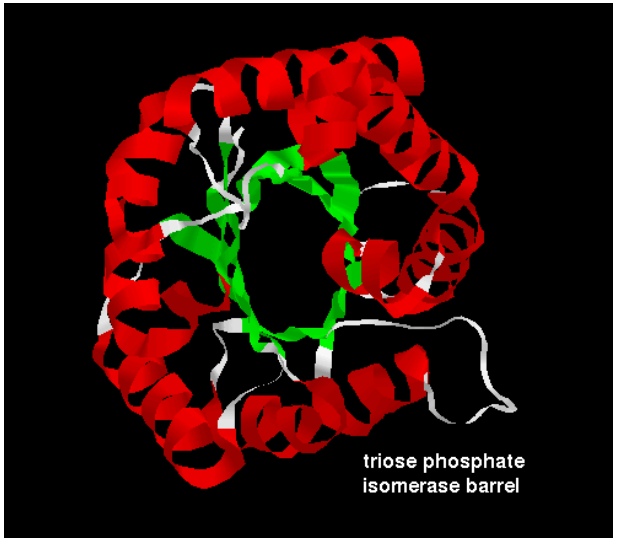
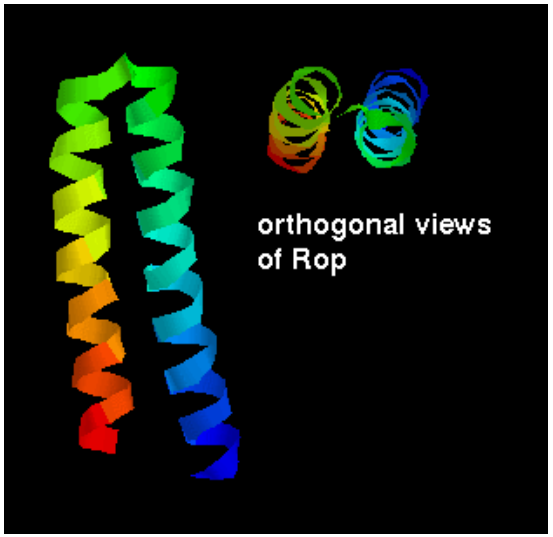
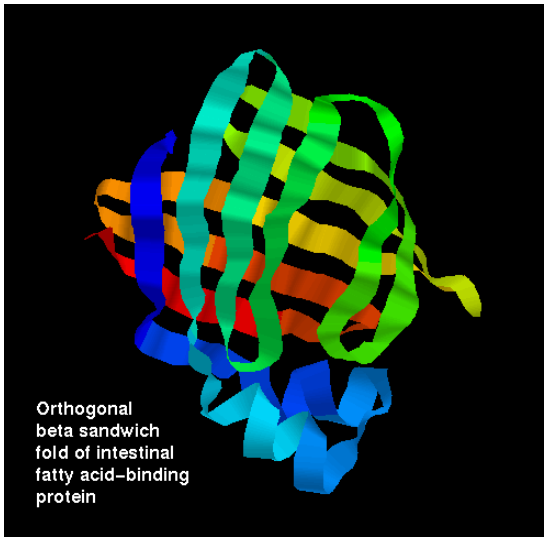
3D Structure

Prediction & Assessment Pt. 2







David Wishart

3-41 Athabasca Hall

david.wishart@ualberta.ca

Objectives

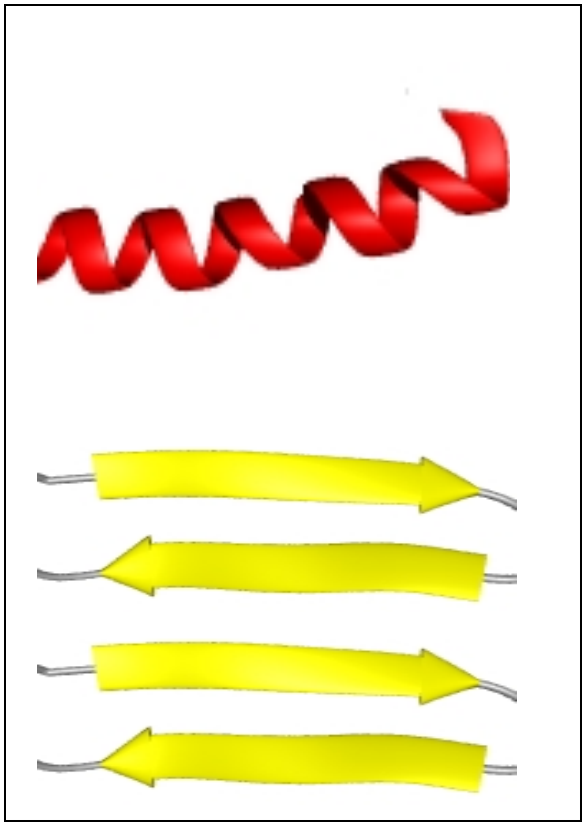
- Become familiar with methods and algorithms for secondary Structure Prediction
- Become familiar with protein Threading (2D and 3D threading)
- Become acquainted with Ab initio protein structure prediction

3D Structure Generation*

- X-ray Crystallography
- NMR Spectroscopy
- Homology or Comparative Modelling
- Secondary Structure Prediction
- Threading (2D and 3D threading)
- Ab initio Structure Prediction

Secondary (2°) Structure

Phi & Psi angles for Regular Secondary Structure Conformations		
Structure	Phi (Φ)	Psi(Ψ)
Antiparallel β-sheet	-139	+135
Parallel β-Sheet	-119	+113
Right-handed α-helix	- 64	- 40
3_{10} helix	-49	-26
π helix	-57	-70
Polyproline I	-83	+158
Polyproline II	-78	+149
Polyglycine II	-80	+150



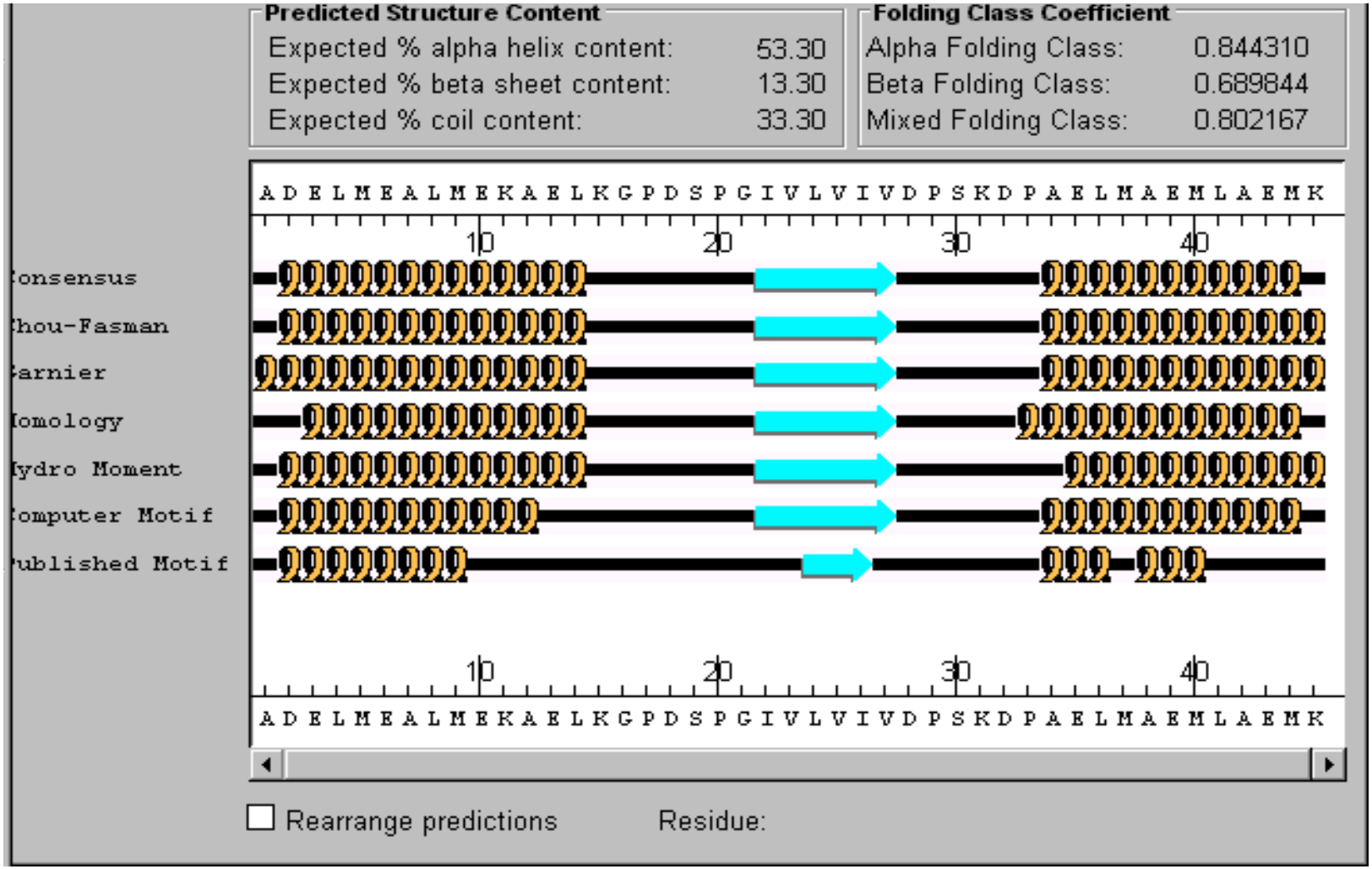
Secondary Structure Prediction*

- One of the first fields to emerge in bioinformatics (~1967)
- Grew from a simple observation that certain amino acids or combinations of amino acids seemed to prefer to be in certain secondary structures
- Subject of hundreds of papers and dozens of books, many methods…

2° Structure Prediction*

- **Statistical** (Chou-Fasman, GOR)
- **Homology or Nearest Neighbor** (Levin)
- **Physico-Chemical** (Lim, Eisenberg)
- **Pattern Matching** (Cohen, Rooman)
- **Neural Nets** (Qian & Sejnowski, Karplus)
- **Evolutionary Methods** (Barton, Niemann)
- **Combined Approaches** (Rost, Levin, Argos)

Secondary Structure Prediction



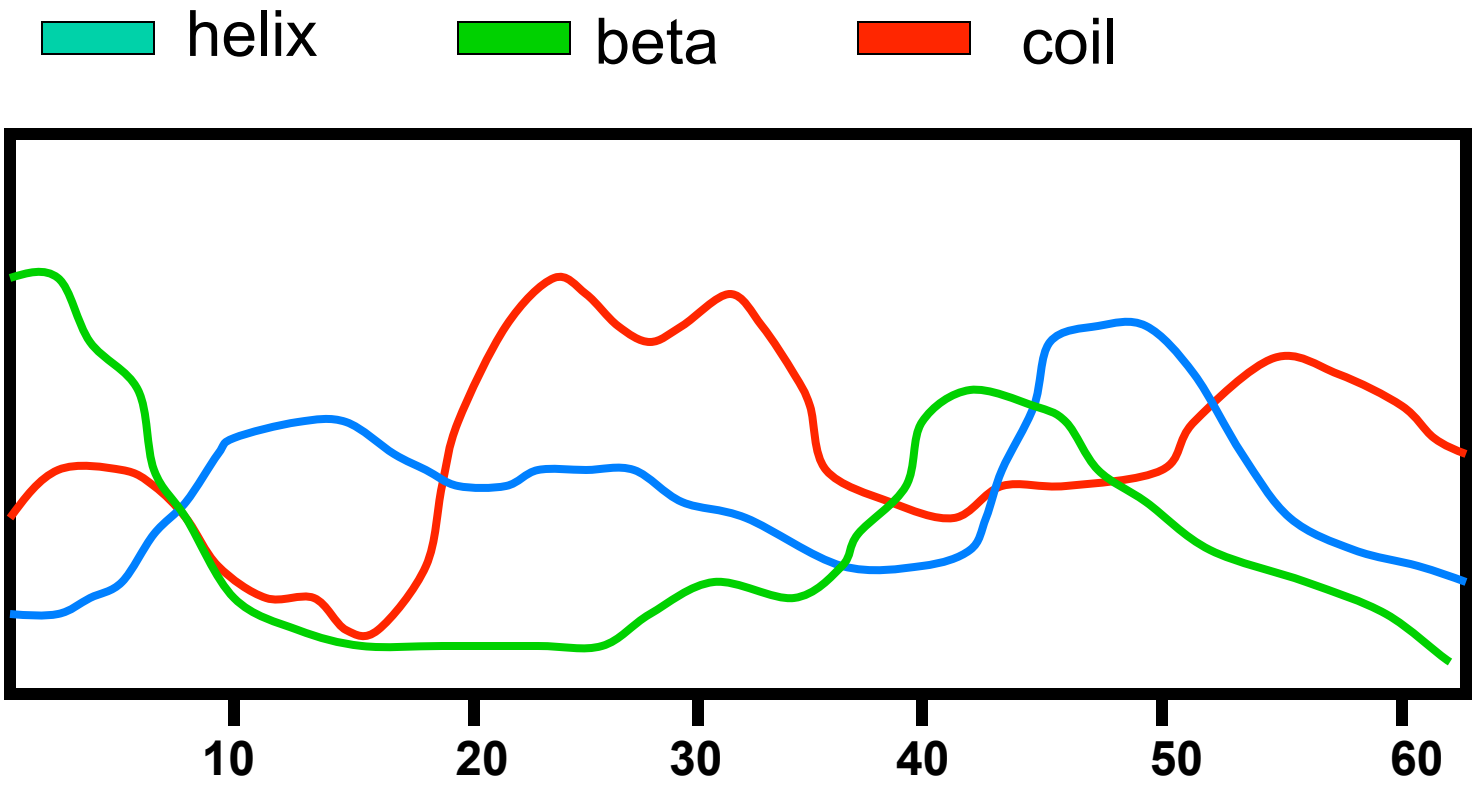
Chou-Fasman Statistics*

Table 8							
Chou & Fasman Secondary Structure Propensity of the Amino Acids							
	Pα	Pβ	Pc		Pα	Pβ	Pc
A	1.42	0.83	0.75	M	1.45	1.05	0.5
C	0.7	1.19	1.11	N	0.67	0.89	1.44
D	1.01	0.54	1.45	P	0.57	0.55	1.88
E	1.51	0.37	1.12	Q	1.11	1.1	0.79
F	1.13	1.38	0.49	R	0.98	0.93	1.09
G	0.57	0.75	1.68	S	0.77	0.75	1.48
H	1	0.87	1.13	T	0.83	1.19	0.98
I	1.08	1.6	0.32	V	1.06	1.7	0.24
K	1.16	0.74	1.1	W	1.08	1.37	0.45
L	1.21	1.3	0.49	Y	0.69	1.47	0.84

Simplified C-F Algorithm*

- Select a window of 7 residues
- Calculate average P_α over this window and assign that value to the central residue
- Repeat the calculation for P_β and P_c
- Slide the window down one residue and repeat until sequence is complete
- Analyze resulting "plot" and assign secondary structure (H, B, C) for each residue to highest value

Simplified C-F Algorithm



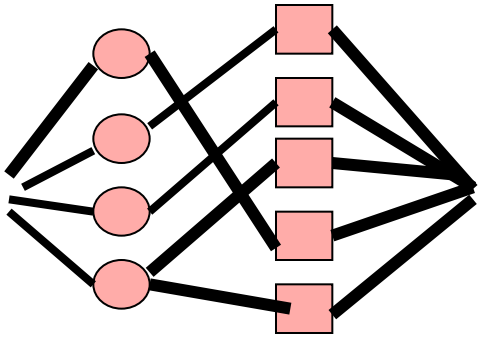
Limitations of Chou-Fasman

- Does not take into account long range information (>3 residues away)
- Does not take into account sequence content or probable structure class
- Assumes simple additive probability (not true in nature)
- Does not include related sequences or alignments in prediction process
- Only about 55% accurate (on good days)

The PhD Approach

Consensus:	CSNLSTCVLGKLSQDLHKLQTFPRT--GAG-P
1: sockeye	CSNLSTCVLGKLSQDLHKLQTFPRTNTGAGVP
2: chum	CSNLSTCVLGKLSQDLHKLQTFPRTNTGAGVP
3: pink	CSNLSTCVLGKLSQDLHKLQTFPRTNTGAGVP
4: coho	CSNLSTCMLGKLSQDLHKLQTFPRTNTGAGVP
5: pig	CSNLSTCVLSAYWRNLNNFHRFSGMGFGPETP
6: bovine	CSNLSTCVLSAYWKDLNNYHRFSGMGFGPETP
7: eel	CSNLSTCVLGKLSQELHKLQTYPRTDVGAGTP

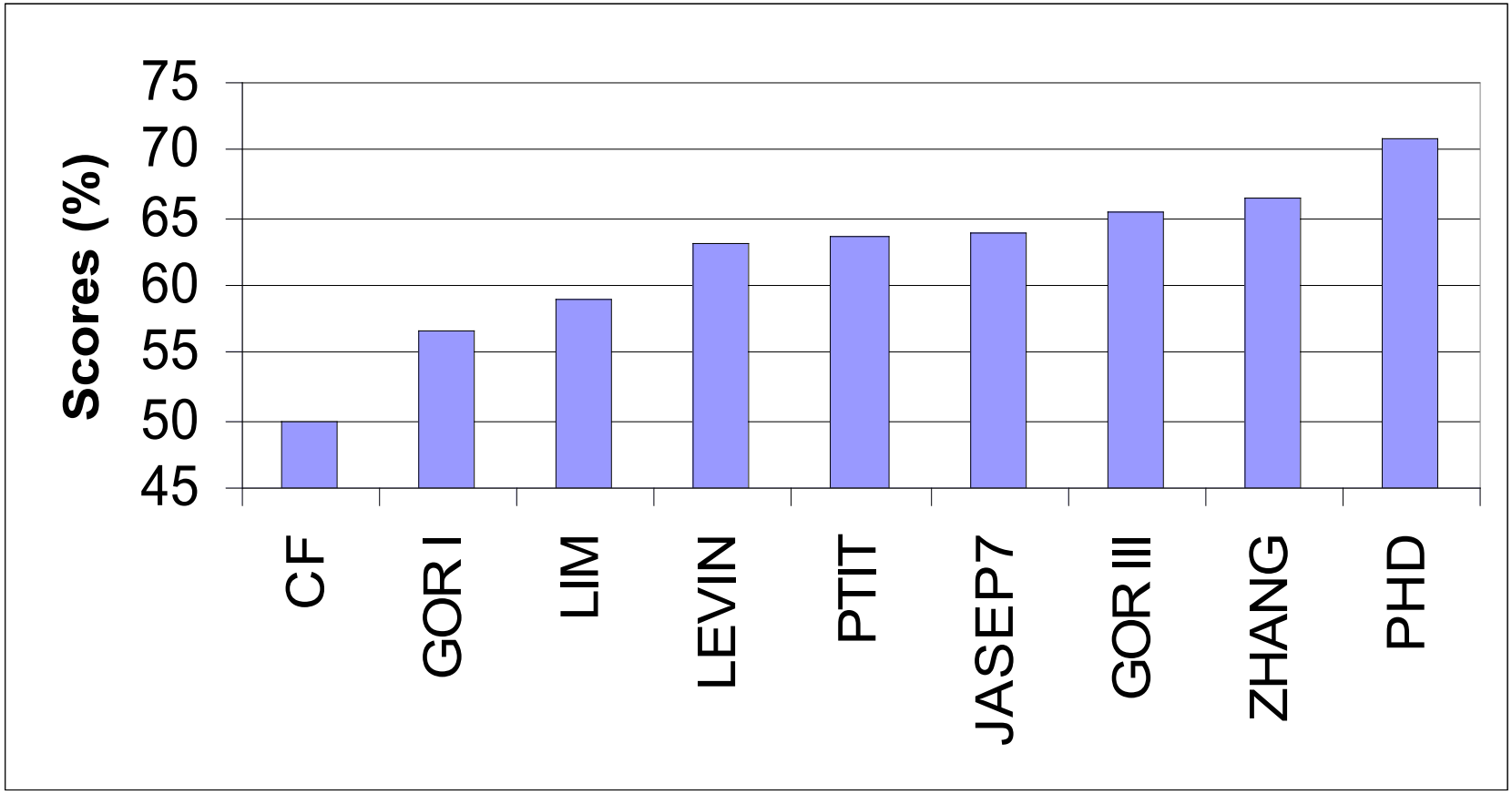
PRFILE...



The PhD Algorithm*

- ***Search the SWISS-PROT database and select high scoring homologues***
- ***Create a sequence "profile" from the resulting multiple alignment***
- ***Include global sequence info in the profile***
- ***Input the profile into a trained two-layer neural network to predict the structure and to "clean-up" the prediction***

Prediction Performance



Evaluating Structure 2° Predictions*

- Historically problematic due to tester bias (developer trains and tests their own predictions)
- Some predictions were up to 10% off
- Move to make testing independent and test sets as large as possible
- EVA – evaluation of protein secondary structure prediction

EVA

Method	Number of proteins	Average ModZscore for all proteins	Comments
APSSP	1312	76.5	
APSSP2	672	82.9	
JNet	?	?	no longer tested
JPred	1218	73.8	
PHD	1599	71.0	
PHDpsi	1598	74.4	
PROF_king	1276	74.6	
PROFsec	1554	76.7	
Prospect	103	71.7	
PSIpred	1461	77.9	
PSSP	?	?	no longer tested
SAM-T99sec	543	75.6	
SSpro1	?	?	no longer tested
SSpro2	1348	76.9	

- **~10 different methods evaluated in real time as new structures arrive at PDB**
- **Results posted on the web and updated weekly**
- **http://www.pdg.cnb.uam.es/eva/**

EVA- http://www.pdg.cnb.uam.es/eva/

EVA measures for secondary structure prediction accuracy

http://www.rostlab.org/eva/doc/measure_sec.html

Department o...ell Biology | Login- Depar... of Alberta | Audiobaba Music Search | Bioinformati... the U of A! | Coilgun Basics 2

Per-residue accuracy

The following scores are used to measure per-residue accuracy (reference, note: most tables contain only a subset of these scores):

1. *Prediction accuracy matrix:*

   M_{ij} = number of residues observed in state i and predicted in state j, with i and j $\in$ {H, E, L}

   note: the total number of residues observed in state i is:

   $$obs_i = \sum_{j=1}^{3} M_{ij} \quad \text{, with } j \in \{H, E, L\}$$

   note: the total number of residues predicted in state i is (helix, strand, other)

   $$prd_i = \sum_{j=1}^{3} M_{ji} \quad \text{, with } i, j \in \{H, E, L\}$$

   and the total number of residues is simply:

   $$N_{res} = \sum_i obs_i = \sum_i prd_i = \sum_{i,j} M_{ij}$$

2. *Three-state prediction accuracy: Q_3*

   Thus, the three-state per residue accuracy Q_3 becomes:

   $$Q_3 = 100 \cdot \frac{1}{N_{res}} \cdot \sum_{i=1}^{3} M_{ii}$$

3. *Per-state percentages:*

   To define accuracy for a particular state (helix, strand, other), there are two possible varia

   - How many of the observed helix residues (strandlother) were correctly predicted? Given are the correctly predicted residues as percentage of all residues OBSERVE

     $$Q_i^{\%obs} = 100 \cdot \frac{M_{ii}}{obs_i}$$

   - How many of the predicted helix (strandlother) residues were correctly predicted? Given are the correctly predicted residues as percentage of all residues PREDICTI

     $$Q_i^{\%prd} = 100 \cdot \frac{M_{ii}}{prd_i}$$

4. *Information index:*

EVA measures for secondary structure prediction accuracy

http://www.rostlab.org/eva/doc/measure_sec.html

Department o...ell Biology | Login- Depar... of Alberta | Audiobaba Music Search | Bioinformati... the U of A! | Coilgun Basics 2

Per-segment accuracy (SOV)

The Segment OVerlap measure for prediction accuracy is defined by (for more details):

- *Per-stage segment overlap:*

  $$SOV_i = \frac{1}{N_i} \sum_{S_i} \frac{MINOV(S1;S2) + DELTA(S1;S2)}{MAXOV(S1;S2)}$$

  with the following definitions:

  S1 and S2 — are the observed and predicted secondary structure segments (in state i, which can be either H, E c
  LEN(S1) — is the number of residues in the segments S1
  MINOV(S1;S2) — is the length of actual overlap of S1 and S2, i.e. the extent for which both segments have residues
  MAXOV(S1;S2) — is the length of the total extent for which either of the segments S1 or S2 has a residue in state
  DELTA(S1;S2) — is the integer value defined as being equal to the following

  $$DELTA(S1;S2) = \min \begin{pmatrix} MAXOV(S1;S2) - MINOV(S1;S2) \\ MINOV(S1;S2) \\ INT(0.5 \cdot LEN(S1)) \\ INT(0.5 \cdot LEN(S2)) \end{pmatrix}$$

  THE SUM — is taken over S, all the pairs of segments {S1;S2}, where S1 and S2 have at least one residue in st
  N(i) — is the number of residues in state i defined as follows:

  $$N_i = \sum_{S(i)} LEN(S1) + \sum_{S'(i)} LEN(S1)$$

  The two sums are taken over S and S':
  S(i) — is the number of all the pairs of segments {S1;S2}, where S1 and S2 have at least one residue in st
  S'(i) — is the number of segments S1 that do not produce any segment pair

- *Segment OVerlap quantity measure for all three states:*

  $$SOV = SOV_3 = \frac{1}{N} \cdot \sum_{S(i)} \frac{MINOV(S1;S2) + DELTA(S1;S2)}{MAXOV(S1;S2)} \cdot LEN(S1)$$

  with:

  $$N = \sum_i N_i$$

2° Structure Evaluation*

- Q3 score – standard method in evaluating performance, 3 states (H,C,B) evaluated like a multiple choice exam with 3 choices. Same as % correct
- SOV (segment overlap score) – more useful measure of how segments overlap and how much overlap exists

Best of the Best

- PredictProtein-PHD (74%)
  - http://www.predictprotein.org/meta.php
- Jpred (73-75%)
  - http://www.compbio.dundee.ac.uk/www-jpred/
- PSIpred (77%)
  - http://bioinf.cs.ucl.ac.uk/psipred/
- Proteus and Proteus2 (88%)
  - http://wks80920.ccis.ualberta.ca/proteus/
  - http://www.proteus2.ca/proteus2/

MetaPP – [Submission]

http://www.predictprotein.org/meta.php

Most Visited Getting Started Latest Headlines

MetaPP – [Submission]

MetaPredictProtein

About | Submission | Help | Downloads | Register | MetaPP

Welcome David My Queries Edit Account Logout

Google™ Custom Search Search

Description of field (click on description for help)	Type the required information into the fields (and select one or more services)
Your email address (watch for typos)	david.wishart@ualberta.ca
One-line name of protein	
Paste, or type your sequence	

- amino acids in one-letter code (any number of spaces allowed)
- other possible formats

For retrieving protein sequences from databases we recommend the Sequence Retrieval System SRS6

Available Services

Choose (at least one) checkbox(es) to request respective services for your protein

Homology-based prediction of 3D structure (not always possible)

☐	Homology Modelling	3D-JIGSAW	Go There	About

Done

Jpred – A Secondary Structure Prediction Server

http://www.compbio.dundee.ac.uk/www-jpred/

Most Visited Getting Started Latest Headlines

Jpred – A Secondary Structure Pre...

UNIVERSITY OF DUNDEE

Jpred 3

Incorporating Jnet

A Secondary Structure Prediction Server

Sequence:

Help
Advanced

Make Prediction Clear

The Barton Group - The University of Dundee

Citation: Cole C, Barber JD & Barton GJ. *Nucleic Acids Res.* **2008**. [Advanced Access]
More citations

W3C XHTML 1.0 W3C CSS level 2

Done

psipred : index

http://bioinf.cs.ucl.ac.uk/psipred/

Most Visited Getting Started Latest Headlines

psipred : index

UCL Department Of Computer Science

Bioinformatics Group

Bloomsbury Centre for Bioinformatics

UCL

Search Group

UCL Home ›› Departments of Computer Science ›› Bioinformatics Group ›› psipred

Site Navigation

Server Navigation

- PSIPRED Server
- PSIPRED help
- Server Overview
- Server Citation
- News
- History
- Build A TDB File
- Software Download
- Login

The PSIPRED Protein Structure Prediction Server

The PSIPRED Protein Structure Prediction Server aggregates several of our structure prediction methods into one location. Users can submit a protein sequence, perform the prediction of their choice and receive the results of the prediction via e-mail. You may select one of three prediction methods to apply to your sequence:
PSIPRED - a highly accurate method for protein secondary structure prediction
MEMSAT and MEMSAT-SVM - our widely used transmembrane topology prediction method
and one of GenTHREADER, pGenTHREADER and pDomTHREADER - sequence profile based fold recognition methods. More...

For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

Choose Prediction Method

- (•) Predict Secondary Structure (PSIPRED v3.0)
- ( ) Predict Transmembrane Topology (MEMSAT3 & MEMSAT-SVM)
- ( ) SVM Prediction of TM Topology and Helix Packing (MEMPACK) - NEW!
- ( ) Fold Recognition (GenTHREADER - quick)
- ( ) Fold Recognition (pGenTHREADER - with profiles and predicted secondary structure)
- ( ) Fold Recognition (pDomTHREADER - annotates multiple domain on chains)

Help...

Input Sequence (single letter amino acid code)

Help...
If you wish to test these services follow this link to retrieve a test fasta sequence.

Filtering Options

Done

Proteus

Submit a sequence for prediction - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://129.128.185.184/proteus/

Proteus Structure Prediction Server

Comprehensive Secondary Structure Predictions

HOME DOCUMENTATION SAMPLE OUTPUT CONTACT & DOWNLOAD

Welcome to Proteus

Proteus is a high-performing integrated web server and a s
prediction methods (PSIPRED, JNET and TRANSSEC [a l
a robust PDB-based structure alignment process to genera
protein Proteus is able to achieve a very high level of accur
query protein shows no similarity whatsoever to any known
Proteus is not restricted to generating accurate secondary
for integral membrane proteins (both helix-containing protei
homologues or a portion of a homologue in the PDB. The 2
updated PDB secondary structure database available for d

Submitting a sequence

You can submit your sequence in FASTA Format by pastin
to the server. A prediction will be returned to you.

Paste Single Sequence (FASTA for

Legend:

```
H = Helix
E = Beta Strand
C = Coil
Line 1 = sequence (single letter IUPAC code,
Line 2 = secondary structure (H, E or C)
Line 3 = confidence score (0-9, 0 = low, 9 =

A '*' character above the overall prediction
residue.
```

Predicted Secondary Structure:

```
  1
    MELWRQCTHWLIQCRVLPPSHRVTWDGAQVCELAQALRD
    CCHHHHHHHHHHCCCCCCCCCCCCCCHHHHHHHHHHHC
    9858999999876356777778877765899999999984

 61
    NLRPQMSQFLCLKNIRTFLSTCCEKFGLKRSELFEAFDI
    CCCCCCHHHHHHHHHHHHHHHHHHCCCHHHHCCCCCHH
    7887757999999999999999985776565478757

121
    AQNRGIMPFPTEEESVGDEDIYSGLSDQIDDTVEEDEDI
    CCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCC
    667777777667777777556555787777766667886

181  ************************************************************ 240
    EPVSMPPKMTEYDKRCCCLREIQQTEEKYTDTLGSIQQHFLKPLQRFLKPQDIEIIFINI
    CCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHCCCHHHHHHHCCH
    9889999999999999999999999999999999999999999999999999999999999

241  ************************************************************ 300
    EDLLRVHTHFLKEMKEALGTPGAANLYQVFIKYKERFLVYGRYCSQVESASKHLDRVAAA
    HHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
    9999999999999999999999999999999999999999999999999999999999999

301  ************************************************************ 360
    REDVQMKLEECSQRANNGRFTLRDLLMVPMQRVLKYHLLLQELVKHTQEAMEKENLRLAL
    HHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHHHHCCCCCCCHHHHHHHHHH
    9999999999999999999999999999999999999999999999999999999999999
```

Proteus Structure Prediction Server

Comprehensive Secondary Structure Predictions

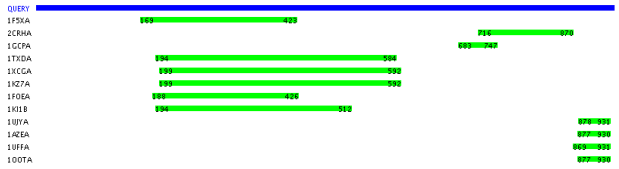
HOME DOCUMENTATION SAMPLE OUTPUT CONTACT & DOWNLOAD

Proteus prediction (ID=5689539) complete

Summary:

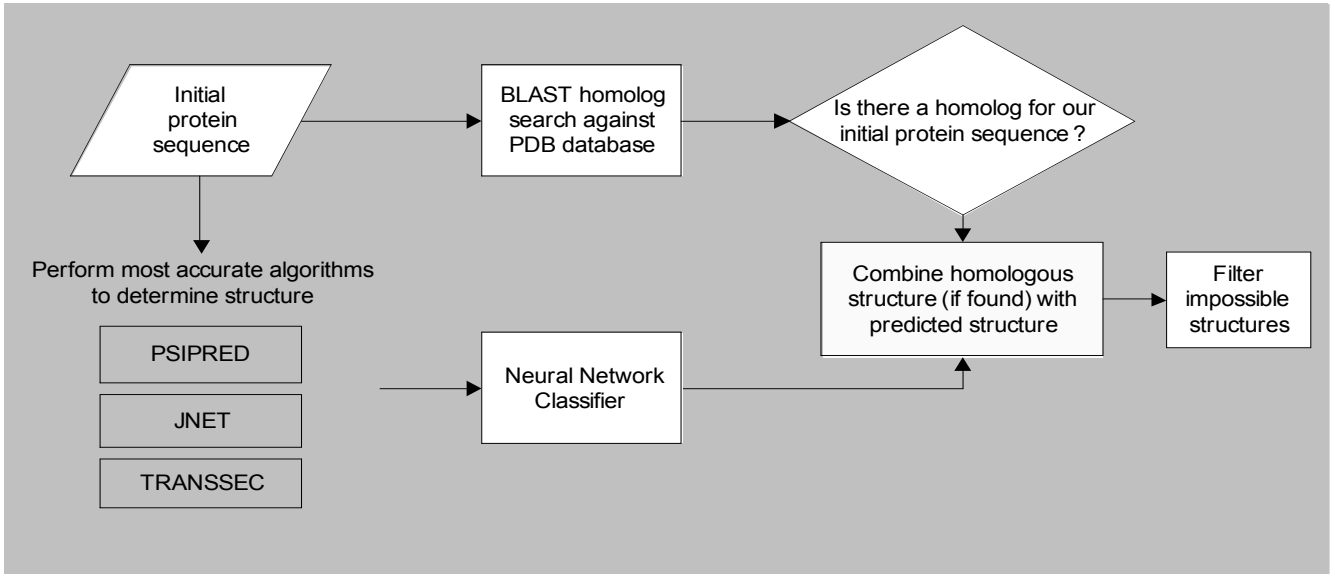
- Time of Submission: 21:34:16 Apr 01, 2006
- Sequence Name:
- Number of residues read in: 844
- Number of useable PDB homologs found: 12
  - 1F5XA, e-value = 1.0E-105 NMR STRUCTURE OF THE Y174 AUTOINHIBITED DBL HOMOLOGY DOMAIN
  - 2CRHA, e-value = 5.0E-65 SOLUTION STRUCTURE OF THE SH2 DOMAIN OF HUMAN PROTO-
  - 1GCPA, e-value = 8.0E-34 CRYSTAL STRUCTURE OF VAV SH3 DOMAIN
  - 1TXDA, e-value = 6.0E-17 CRYSTAL STRUCTURE OF THE DH/PH DOMAINS OF LEUKEMIA-
  - 1XCGA, e-value = 2.0E-13 CRYSTAL STRUCTURE OF HUMAN RHOA IN COMPLEX WITH DH/PH
  - 1KZ7A, e-value = 7.0E-12 CRYSTAL STRUCTURE OF THE DH/PH FRAGMENT OF MURINE DBS IN
  - 1F0EA, e-value = 5.0E-11 CRYSTAL STRUCTURE OF RAC1 IN COMPLEX WITH THE GUANINE
  - 1KI1B, e-value = 1.0E-9 GUANINE NUCLEOTIDE EXCHANGE REGION OF INTERSECTIN IN
  - 1UFYA, e-value = 7.0E-9 SOLUTION STRUCTURE OF SH3 DOMAIN IN RAC/CDC42 GUANINE
  - 1AZEA, e-value = 1.0E-8 NMR STRUCTURE OF THE COMPLEX BETWEEN THE C32S-Y7V MUTANT OF
  - 1UFFA, e-value = 5.0E-8 SOLUTION STRUCTURE OF THE FIRST SH3 DOMAIN OF HUMAN
  - 1OOTA, e-value = 6.0E-8 CRYSTAL STRUCTURE OF THE SH3 DOMAIN FROM A S. CEREVISIAE
- Number of sequence alignments used for ab-initio predictions: **49**
- Overall confidence value: 83.4%
- Predicted % Helix content: 37 % (313 residues)
- Predicted % Beta sheet content: 14 % (115 residues)
- Predicted % Coil content: 49 % (416 residues)

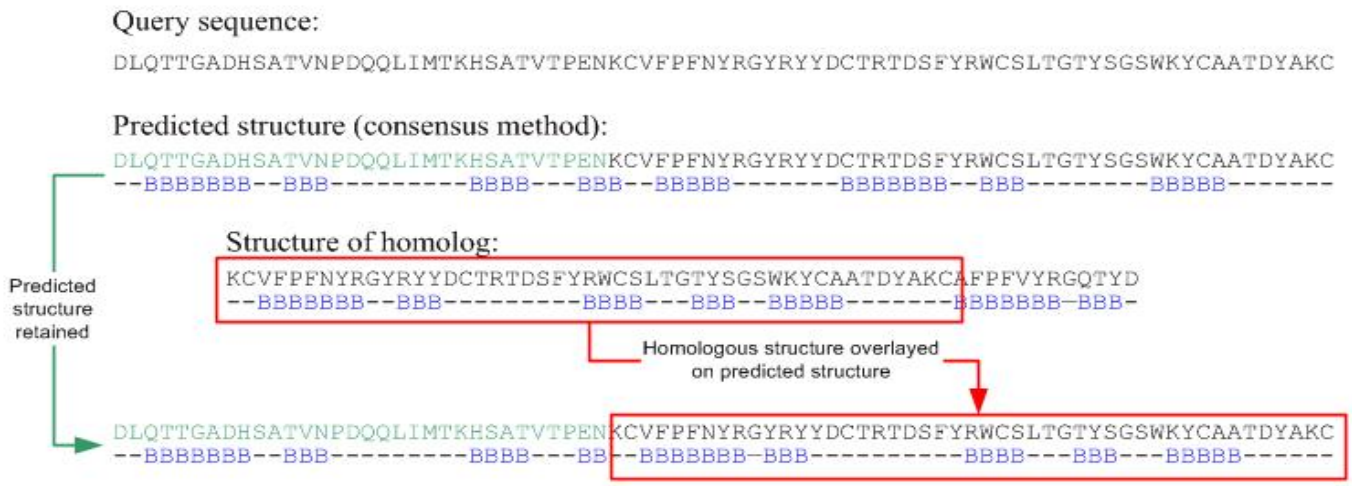
Graphical Alignment of PDB Homologs:



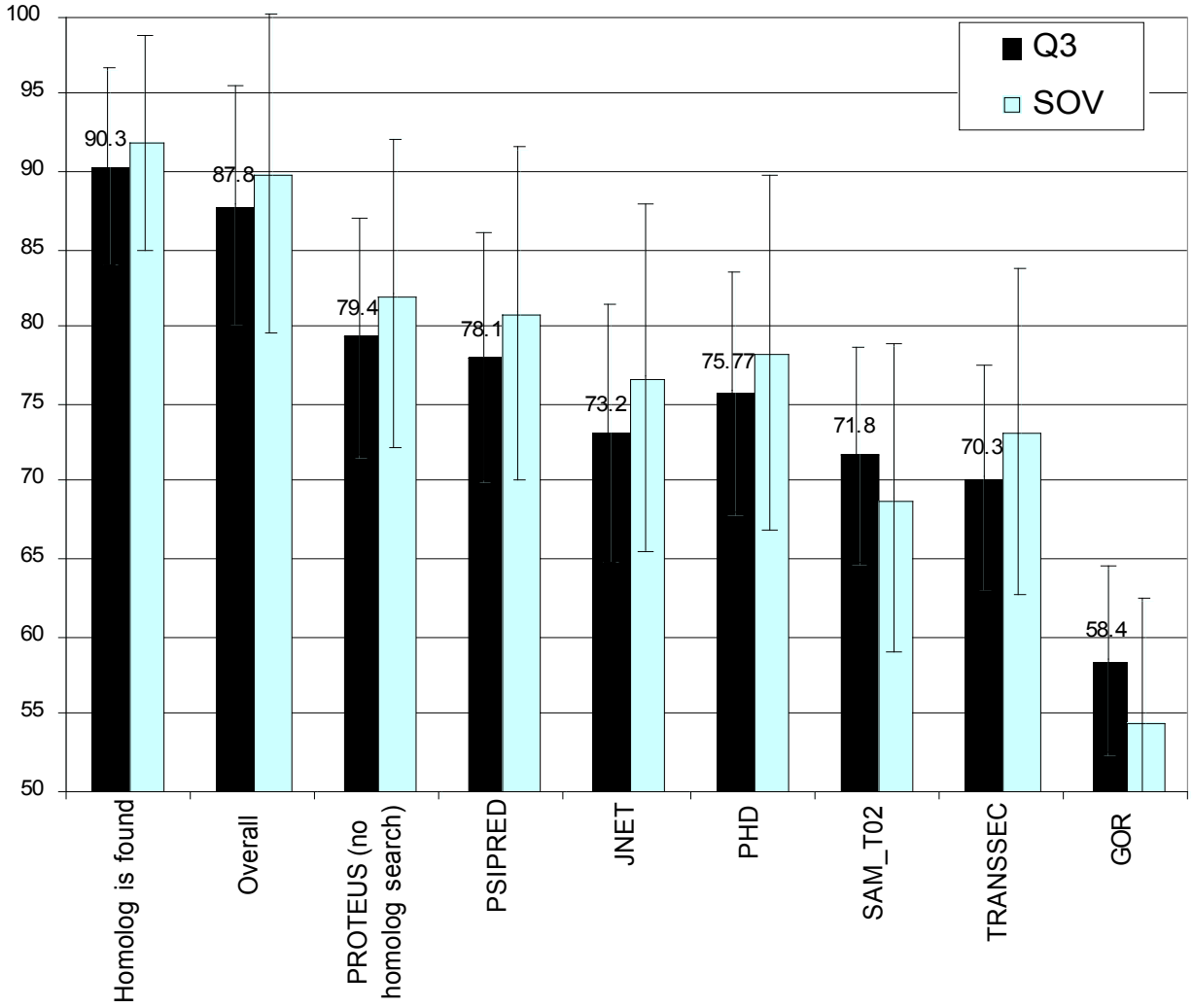
start SnagIt Submit a sequen... Local Dis

Proteus Methods*

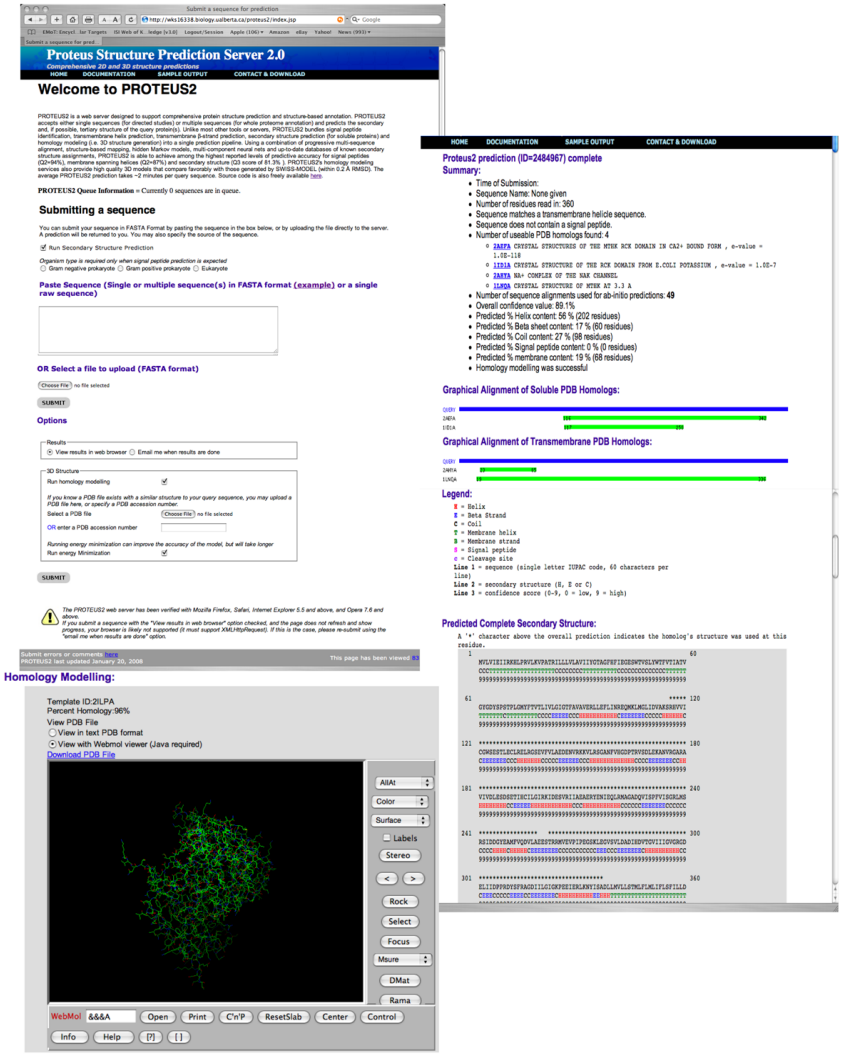


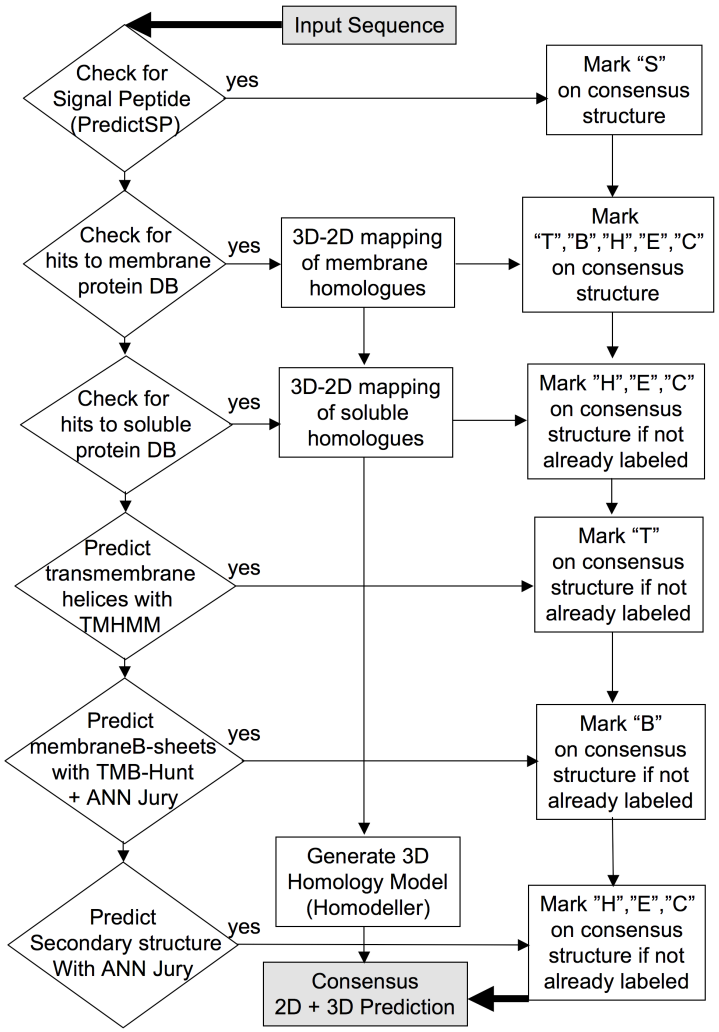


Performance Comparison



Proteus2*


Proteus Structure Prediction Server 2.0
Welcome to PROTEUS2
Submitting a sequence
Homology Modelling:


Input Sequence
Check for Signal Peptide (PredictSP)
yes
Mark "S" on consensus structure
Check for hits to membrane protein DB
yes
3D-2D mapping of membrane homologues
Mark "T","B","H","E","C" on consensus structure
Check for hits to soluble protein DB
yes
3D-2D mapping of soluble homologues
Mark "H","E","C" on consensus structure if not already labeled
Predict transmembrane helices with TMHMM
yes
Mark "T" on consensus structure if not already labeled
Predict membraneB-sheets with TMB-Hunt + ANN Jury
yes
Mark "B" on consensus structure if not already labeled
Generate 3D Homology Model (Homodeller)
Predict Secondary structure With ANN Jury
yes
Mark "H","E","C" on consensus structure if not already labeled
Consensus 2D + 3D Prediction

Proteus2 Performance*

Transmembrane Helix Prediction Performance (TMH Benchmark test set)		
Program or Server	**Q2**	**# False positives**
PROTEUS2	91%	0
TMHMM	80%	1
HMMTOP	80%	6
DAS	72%	16
Transmembrane Helix Prediction Performance (PPT-DB-TMH test set)		
Program or Server	**Q2**	**# False neg. (Missed Prots)**
PROTEUS2	87%	0
TMHMM	82%	8
Transmembrane Beta Barrel Detection Performance (PPT-DB "All" protein data set)		
Program or Server	**Q2**	**Accuracy (TMB vs glob)**
PROTEUS2	100%	100%
TMB-Hunt	78%	99%
Transmembrane Beta Strand Prediction Performance (PPT-DB -TMB test set)		
Program or Server	**Q2**	
PROTEUS2	86%	
Pred-TMBB	73%	
Non-membrane Secondary Structure Prediction Performance (EVA Test Set)		
Program or Server	**Q3**	**SOV**
PROTEUS2	81	82
Porter	77	76
JNET	72	73
PSIPred	77	78

Definition*

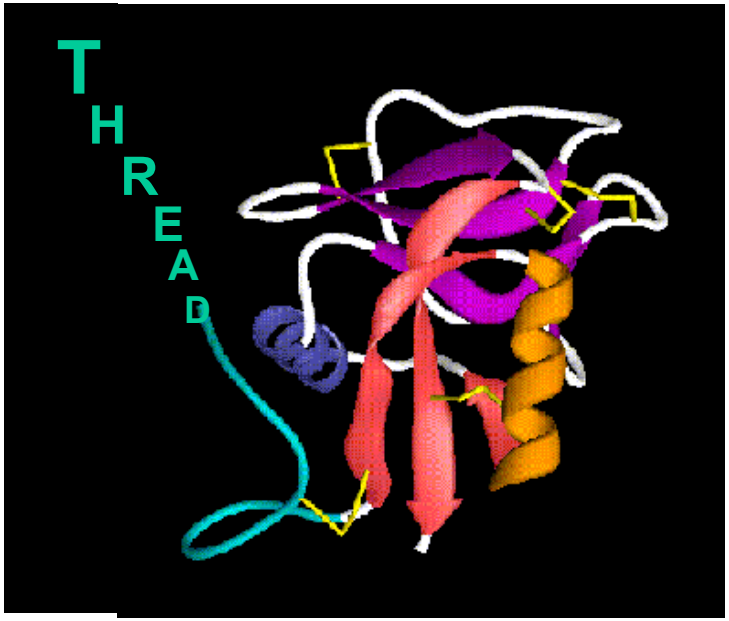
- **Threading - A protein fold recognition technique that involves incrementally replacing the sequence of a known protein structure with a query sequence of unknown structure. The new "model" structure is evaluated using a simple heuristic measure of protein fold quality. The process is repeated against all known 3D structures until an optimal fit is found.**

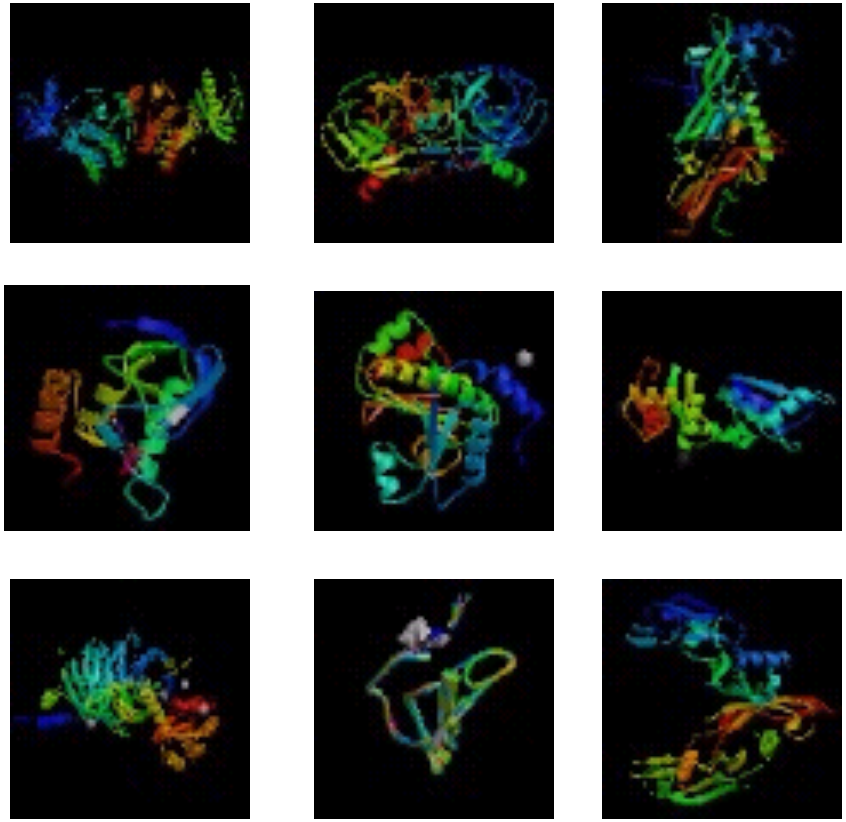
Why Threading?*

- Secondary structure is more conserved than primary structure
- Tertiary structure is more conserved than secondary structure
- Therefore very remote relationships can be better detected through 2° or 3° structural homology instead of sequence homology

Visualizing Threading

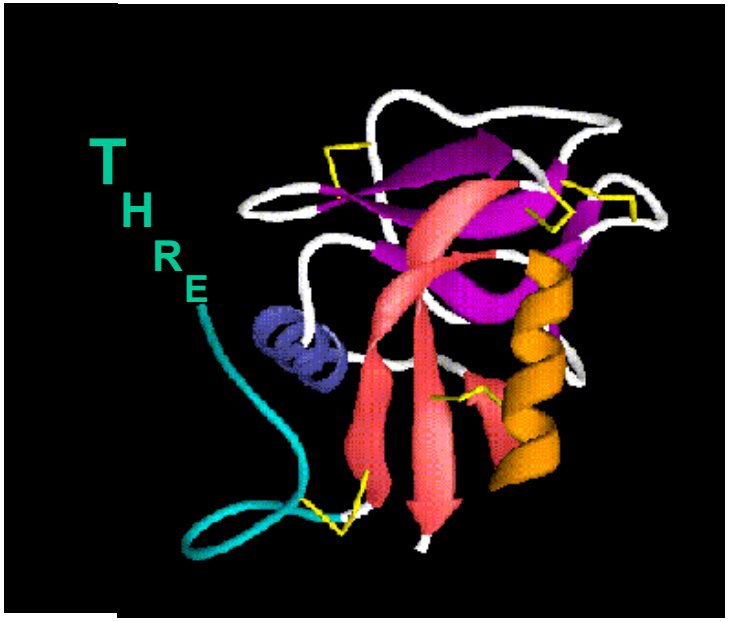
THREADINGSEQNCEECNQESGNI
ERHTHREADINGSEQNCETHREAD
GSEQNCEQCQESGIDAERTHR...

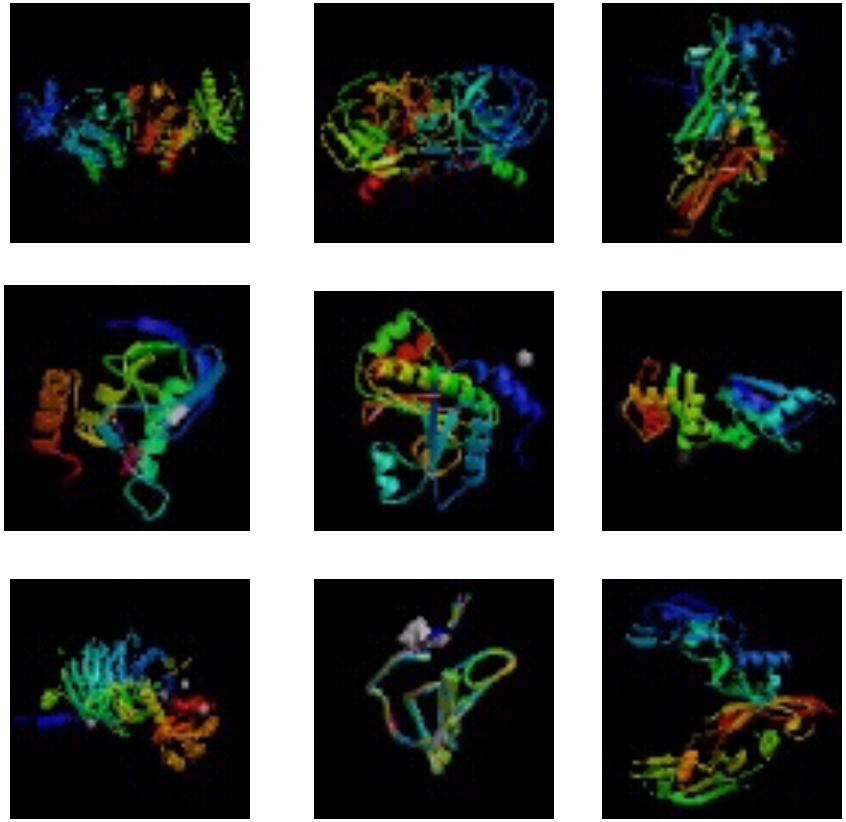




Visualizing Threading

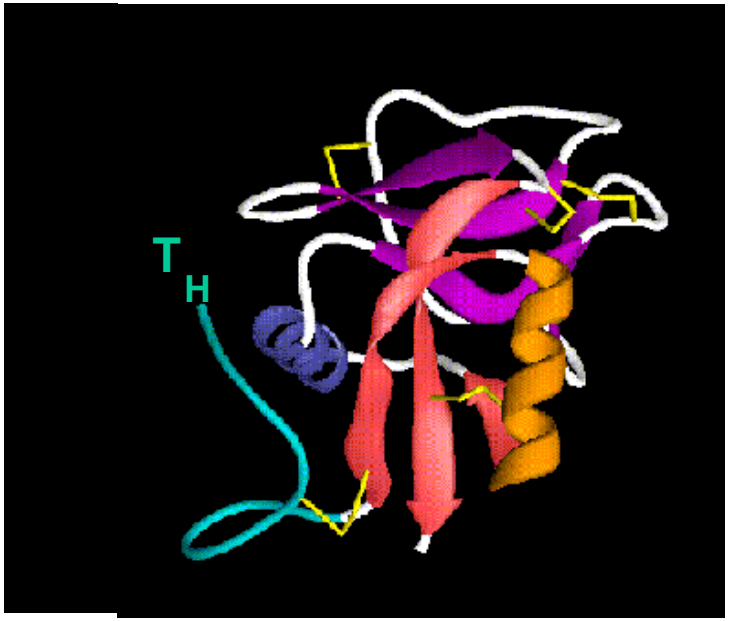
THREADINGSEQNCEECNQESGNI
ERHTHREADINGSEQNCETHREAD
GSEQNCEQCQESGIDAERTHR...

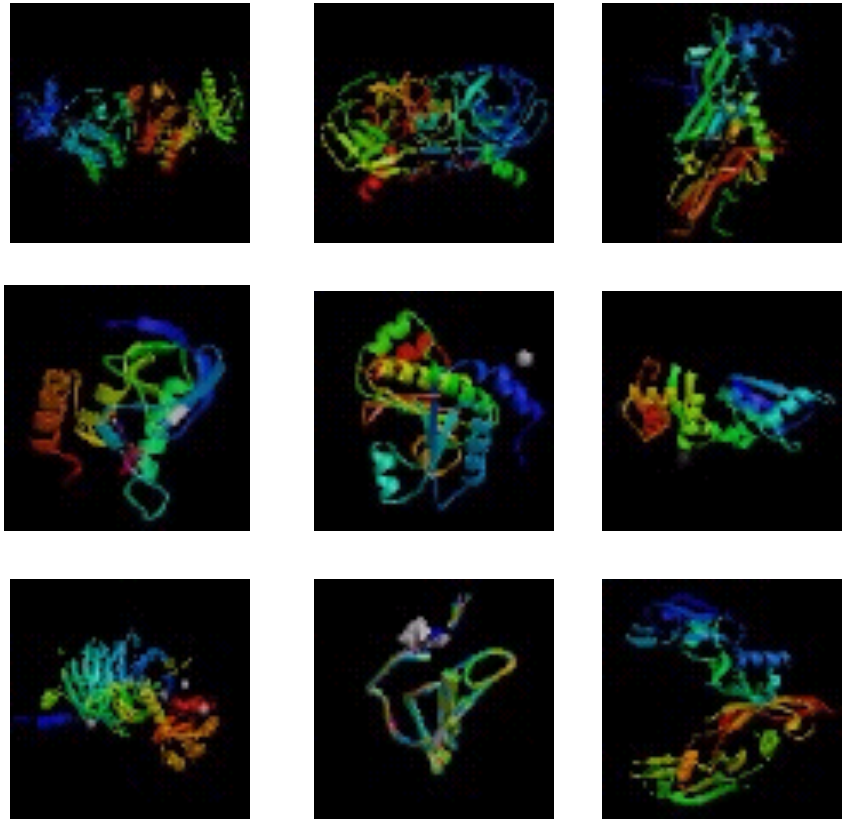




Visualizing Threading

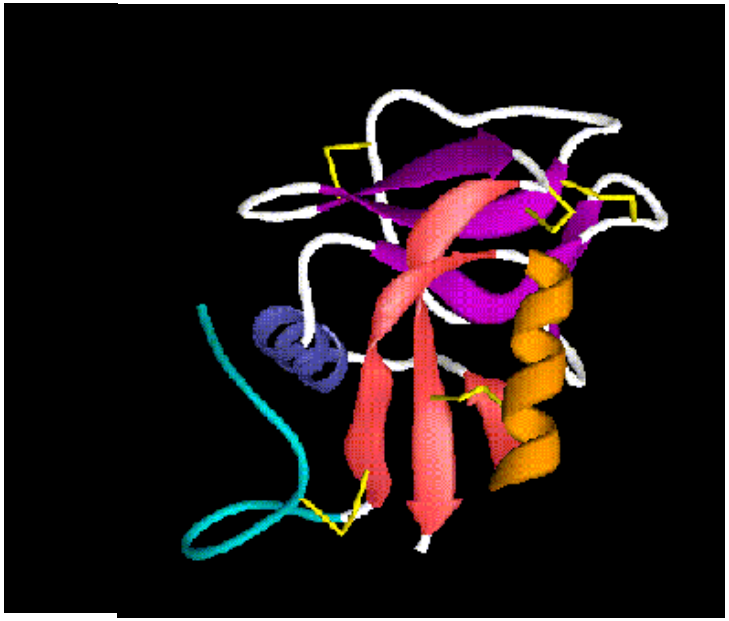
THREADINGSEQNCEECNQESGNI
ERHTHREADINGSEQNCETHREAD
GSEQNCEQCQESGIDAERTHR...

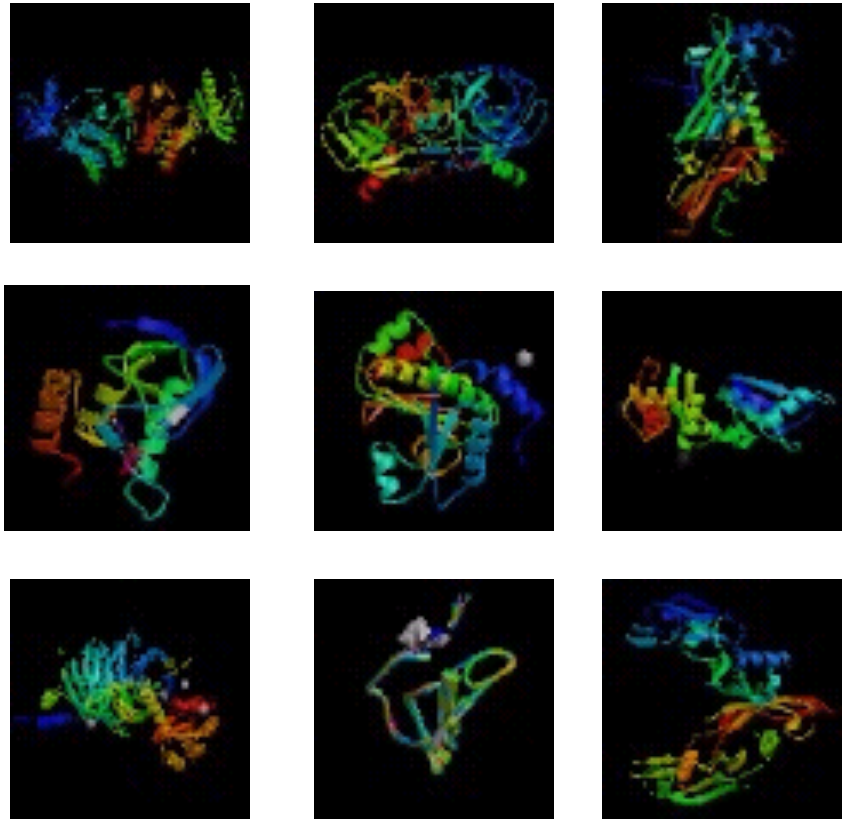




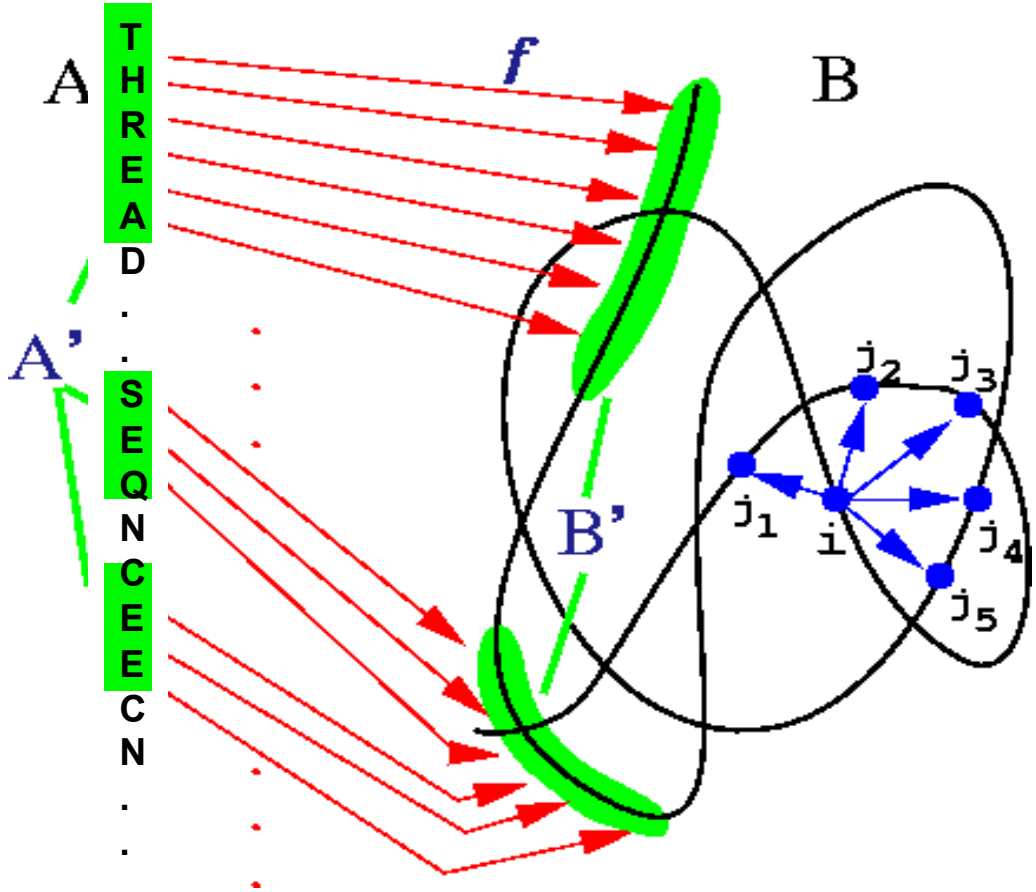
Visualizing Threading

THREADINGSEQNCEECNQESGNI
ERHTHREADINGSEQNCETHREAD
GSEQNCEQCQESGIDAERTHR...





Visualizing Threading



Threading*

- Database of 3D structures and sequences
  - Protein Data Bank (or non-redundant subset)
- Query sequence
  - Sequence < 25% identity to known structures
- Alignment protocol
  - Dynamic programming
- Evaluation protocol
  - Distance-based potential or secondary structure
- Ranking protocol

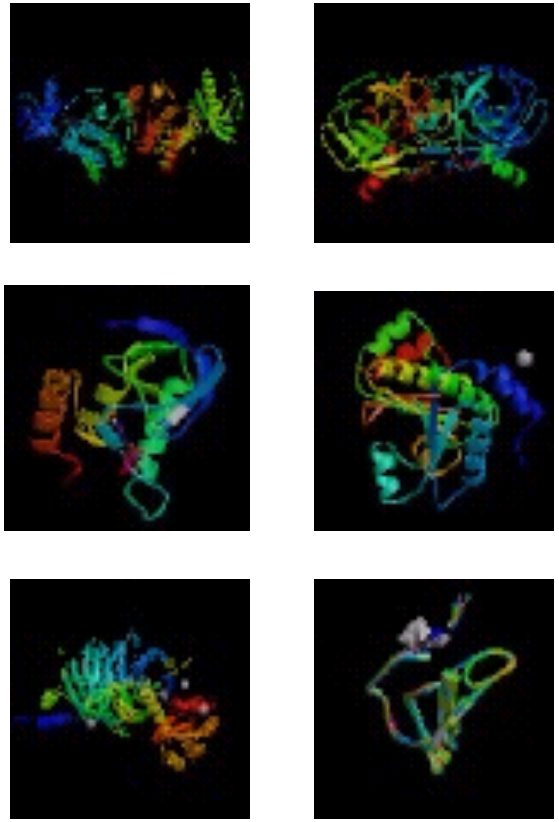
2 Kinds of Threading*

- **2D Threading or Prediction Based Methods (PBM)**
  - **Predict secondary structure (SS) or ASA of query**
  - **Evaluate on basis of SS and/or ASA matches**
- **3D Threading or Distance Based Methods (DBM)**
  - **Create a 3D model of the structure**
  - **Evaluate using a distance-based "hydrophobicity" or pseudo-thermodynamic potential**

2D Threading Algorithm*

- ***Convert PDB to a database containing sequence, SS and ASA information***
- ***Predict the SS and ASA for the query sequence using a "high-end" algorithm***
- ***Perform a dynamic programming alignment using the query against the database (include sequence, SS & ASA)***
- ***Rank the alignments and select the most probable fold***

Database Conversion



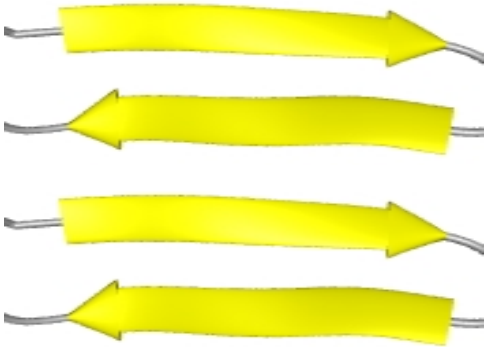
```
>Protein1
THREADINGSEQNCEECNQESGNI
HHHHHHCCCCEEEEECCCHHHHHH
ERHTHREADINGSEQNCETHREAD
HHCCEEEEECCCCCHHHHHHHHHH

>Protein2
QWETRYEWQEDFSHAECNQESGNI
EEEEECCCCHHHHHHHHHHHHHHH
YTREWQHGFDSASQWETRA
CCCCEEEEECCCEEEEECC

>Protein3
LKHGMNSNWEDFSHAECNQESG
EEECCEEEECCCEEECCCCCCC
```

Secondary Structure

Phi & Psi angles for Regular Secondary Structure Conformations		
Structure	Phi (Φ)	Psi(Ψ)
Antiparallel β-sheet	-139	+135
Parallel β-Sheet	-119	+113
Right-handed α-helix	+64	+40
3_{10} helix	-49	-26
π helix	-57	-70
Polyproline I	-83	+158
Polyproline II	-78	+149
Polyglycine II	-80	+150

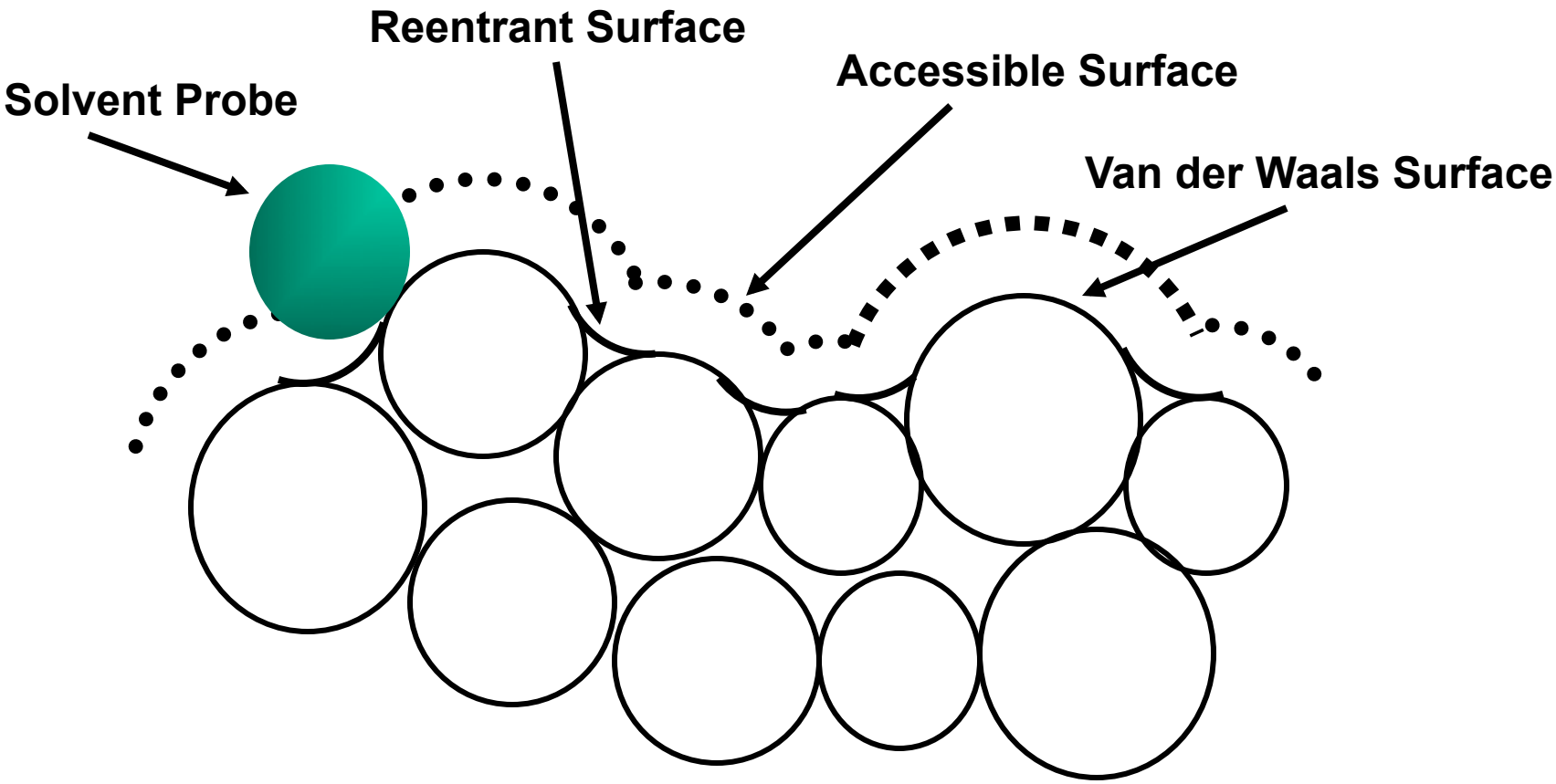


2° Structure Identification*

- **DSSP** - Database of Secondary Structures for Proteins (http://swift.cmbi.ru.nl/gv/start/index.html)
- **VADAR** - Volume Area Dihedral Angle Reporter (http://vadar.wishartlab.com/)
- **PDB** - Protein Data Bank (www.rcsb.org)
- **STRIDE** (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py)

QHTAWCLTSEQHTAAVIWDCETPGKQNGAYQEDCA
HHHHHHCCEEEEEEEEEEECCHHHHHHHCCCCCCC

Accessible Surface Area



ASA Calculation*

- **DSSP** - Database of Secondary Structures for Proteins (http://swift.cmbi.ru.nl/gv/start/index.html)
- **VADAR** - Volume Area Dihedral Angle Reporter (http://vadar.wishartlab.com/)
- **GetArea** - http://curie.utmb.edu/getarea.html

```
QHTAWCLTSEQHTAAVIWDCETPGKQNGAYQEDCAMD
BBPPBEEEEEEPBPBPBPBBPEEEPBPEPEEEEEEEEE
10562987994152515104789414969899999999
```

Other ASA sites

- Connolly Molecular Surface Home Page
  - http://www.biohedron.com/
- Naccess Home Page
  - http://www.bioinf.manchester.ac.uk/naccess/
- MSMS
  - http://www.scripps.edu/~sanner/html/msms_home.html
- Surface Racer
  - http://apps.phar.umich.edu/tsodikovlab/

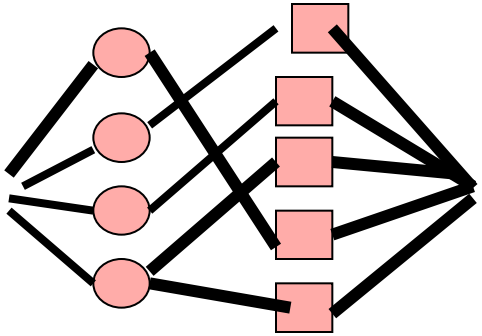
2D Threading Algorithm

- ***Convert PDB to a database containing sequence, SS and ASA information***
- ***Predict the SS and ASA for the query sequence using a "high-end" algorithm***
- ***Perform a dynamic programming alignment using the query against the database (include sequence, SS & ASA)***
- ***Rank the alignments and select the most probable fold***

ASA Prediction*

- **NetSurfP (70%)**
  - http://www.cbs.dtu.dk/services/NetSurfP/
- **PredAcc (70%?)**
  - http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=PredAcc

QHTAW...



QHTAWCLTSEQHTAAVIW

BBPPBEEEEEPBPBPBPB

2D Threading Algorithm

- ***Convert PDB to a database containing sequence, SS and ASA information***
- ***Predict the SS and ASA for the query sequence using a "high-end" algorithm***
- ***Perform a dynamic programming alignment using the query against the database (include sequence, SS & ASA)***
- ***Rank the alignments and select the most probable fold***

Dynamic Programming

	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	10	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

```
G   E   N   E   T   I   C   S
|   |   |   |   *   |       |
G   E   N   E   S   I       S
```

S_{ij} (Identity Matrix)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

A Simple Example...*

	A	A	T	V	D
A	1				
V					
V					
D					

	A	A	T	V	D
A	1	1			
V					
V					
D					

	A	A	T	V	D
A	1	1	0	0	0
V					
V					
D					

	A	A	T	V	D
A	1	1	0	0	0
V	0				
V					
D					

	A	A	T	V	D
A	1	1	0	0	0
V	0	1	1		
V					
D					

	A	A	T	V	D
A	1	1	0	0	0
V	0	1	1	2	
V					
D					

A Simple Example..*.

	A	A	T	V	D
A	1	1	0	0	0
V	0	1	1	2	1
V					
D					

	A	A	T	V	D
A	1	1	0	0	0
V	0	1	1	2	1
V	0	1	1	2	2
D	0	1	1	1	3

	A	A	T	V	D
A	1	1	0	0	0
V	0	1	1	2	1
V	0	1	1	2	2
D	0	1	1	1	3

```
A A T V D
|   | | |
A - V V D
```

```
A A T V D
  | | | |
  A V V D
```

```
A A T V D
| |   | |
A V - V D
```

Let's Include 2° info & ASA*

S_{ij}^{strc}

	H	E	C
H	1	0	0
E	0	1	0
C	0	0	1

S_{ij}^{asa}

	E	P	B
E	1	0	0
P	0	1	0
B	0	0	1

$$S_{ij}^{total} = k_1 S_{ij}^{seq} + k_2 S_{ij}^{strc} + k_3 S_{ij}^{asa}$$

A Simple Example...*

	E	E	E	C	C
	A	A	T	V	D
E	A	2			
E	V				
C	V				
C	D				

	E	E	E	C	C
	A	A	T	V	D
E	A	2	2		
E	V				
C	V				
C	D				

	E	E	E	C	C
	A	A	T	V	D
E	A	2	2	1	0	0
E	V				
C	V				
C	D				

	E	E	E	C	C
	A	A	T	V	D
E	A	2	2	1	0	0
E	V	1				
C	V				
C	D				

	E	E	E	C	C
	A	A	T	V	D
E	A	2	2	1	0	0
E	V	1	3	3		
C	V				
C	D				

	E	E	E	C	C
	A	A	T	V	D
E	A	2	2	1	0	0
E	V	1	3	3	3	
C	V				
C	D				

A Simple Example...

		E	E	E	C	C
		A	A	T	V	D
E	A	2	2	1	0	0
E	V	1	3	3	3	2
C	V					
C	D					

		E	E	E	C	C
		A	A	T	V	D
E	A	2	2	1	0	0
E	V	1	3	3	3	2
C	V	0	2	3	5	4
C	D	0	2	3	4	7

		E	E	E	C	C
		A	A	T	V	D
E	A	2	2	1	0	0
E	V	1	3	3	3	2
C	V	0	2	3	5	4
C	D	0	2	3	4	7

```
A A T V D
|   | | |
A - V V D
```

```
A A T V D
  | | | |
A V V D
```

```
A A T V D
| |   | |
A V - V D
```

2D Threading Performance

- In test sets 2D threading methods can identify 30-40% of proteins having very remote homologues (i.e. not detected by BLAST) using "minimal" non-redundant databases (<700 proteins)
- If the database is expanded ~4x the performance jumps to 70-75%
- Performs best on true homologues as opposed to postulated analogues

2D Threading Advantages*

- Algorithm is easy to implement
- Algorithm is very fast (10x faster than 3D threading approaches)
- The 2D database is small (<500 kbytes) compared to 3D database (>1.5 Gbytes)
- Appears to be just as accurate as DBM or other 3D threading approaches
- Very amenable to web servers

Secondary structure element alignment

http://protein.cribi.unipd.it/ssea/

Most Visited Getting Started Latest Headlines

Secondary structure element align...

Secondary Structure Element Alignment (SSEA)

Database alignment mode

One vs. One Alignment | Quick Help and References | Scoring and Benchmarking | Download the Software

Please see our *acceptable usage policy*.

Name of sequence (optional)	Fold Library
	PDB 95 fold library

Secondary Structure

Options:

Predict secondary structure from sequence	☐
Local Alignment	◉
Global Alignment	○
Z-Score Cutoff:	0

Submit Query Reset

(c) Silvio Tosatto for Biocomputing GRUP 08 / 2004

Done

http://protein.cribi.unipd.it/ssea/

PDBeFold

http://www.ebi.ac.uk/msd-srv/ssm/

Most Visited Getting Started Latest Headlines

PDBeFold

EMBL-EBI Enter Text Here Find Help Feedback

Databases Tools Research Training Industry About Us Help Site Index

PDBe HOME

Structure Similarity

pdbe.org/fold

Bringing Structure to Biology

FEEDBACK PDB EMDB

Documentation and links

PDBeFold links

- Tips
- Visualisation
- Performance
- Privacy
- FAQs
- Version log
- Fold Links
- Comparisons
- Publications
- PDBeFOLD tutorial

Other links

- PDBePISA
- CCP4 CoorLib
- Rasmol
- Rastop
- Jmol
- PDB
- SCOP
- PDBeMotif
- GeneCensus
- FSSP
- CATH
- PDBSum
- UniProt

PDBeFold (Structure Similarity)

PDBeFold (SSM) is an interactive service for comparing protein structures in 3D.

PDBeFold functionality:

- pairwise comparison and 3D alignment of protein structures
- multiple comparison and 3D alignment of protein structures
- examination of a protein structure for similarity with the whole PDB archive or SCOP archive
- best Cα-alignment of compared structures
- download and visualisation of best-superposed structures using Rasmol (Unix/Linux platforms), Rastop (Windows machines) and Jmol(platform-independent server-side java viewer)
- linking the results to other services - PDBeMotif, SCOP, GeneCensus, FSSP, CATH, PDBSum, UniProt

Launch PDBeFold

PDBeFold: A comparison with other protein matching services.
PDBeFold is used as a structure search engine in PDBePISA.
PDBeFold queries may be launched from any web site (instructions).
PDBeFold is based on the CCP4 Coordinate Library.

We are having hardware issues that may occasionally affect the PISA and SSM services.
If you are experiencing problems with the service please wait for 20 minutes and try again.
We apologise for any inconvenience.

We welcome your feedback! Please send any questions, comments, suggestions and bug reports using the FEEDBACK button on the top of the page.

Done

http://www.ebi.ac.uk/msd-srv/ssm/

Servers - HHPred

HHpred – Homology detection & structure prediction by HMM-HMM comparison

http://toolkit.tuebingen.mpg.de/hhpred

Most Visited Getting Started Latest Headlines

HHpred – Homology detection & s...

HOME | Login | PDBalert | Personal Databases | Contact | Imprint | Disclaimer | Help

MAX-PLANCK-GESELLSCHAFT

Bioinformatics Toolkit

Max-Planck Institute for Developmental Biology

Quickfinder

Search | Alignment | Sequence Analysis | 2ary Structure | 3ary Structure | Classification | Utils

CS-BLAST | FHMMER | HHpred | HHsenser | NucBLAST | PSI-BLAST | PatternSearch | ProtBLAST | SimShiftDB

Show results of job:

Show results

Recent jobs:

Select all Deselect all

Clear sel. Jobs

Delete sel. Jobs

queued
running
done
error

HHpred - Homology detection & structure prediction by HMM-HMM comparison

Help

HHpred now runs with HHsearch 1.6.0.0

Input

Paste protein sequence or multiple alignment

or upload a local file — Browse...

Select input format — FASTA

Reset form | Submit job

Search Options

Select HMM databases (hold Ctr to select several)

Standard	Genomes
pdb70_24Oct09	Arabidopsis_thaliana
pdb_on_hold_23Oct09	Drosophila_melanogaster
scop70_1.71	Homo_sapiens
scop70_1.75	Mus_musculus
cdd_23Oct09	Plasmodium_falciparum

Max. PSI-BLAST iterations — 8

Score secondary structure — yes / no / predicted vs predicted only

Alignment mode — local / global

Realign with MAC

Done

http://toolkit.tuebingen.mpg.de/hhpred

Servers - GenThreader

psipred : index

http://bioinf.cs.ucl.ac.uk/psipred/

Most Visited Getting Started Latest Headlines

psipred : index

UCL Department Of Computer Science

Bioinformatics Group

Bloomsbury Centre for Bioinformatics

UCL

Search Group

UCL Home ›› Departments of Computer Science ›› Bioinformatics Group ›› psipred

Site Navigation

Server Navigation

- PSIPRED Server
- PSIPRED help
- Server Overview
- Server Citation
- News
- History
- Build A TDB File
- Software Download
- Login

The PSIPRED Protein Structure Prediction Server

The PSIPRED Protein Structure Prediction Server aggregates several of our structure prediction methods into one location. Users can submit a protein sequence, perform the prediction of their choice and receive the results of the prediction via e-mail. You may select one of three prediction methods to apply to your sequence:
PSIPRED - a highly accurate method for protein secondary structure prediction
MEMSAT and MEMSAT-SVM - our widely used transmembrane topology prediction method
and one of GenTHREADER, pGenTHREADER and pDomTHREADER - sequence profile based fold recognition methods. More...

For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

Choose Prediction Method

- Predict Secondary Structure (PSIPRED v3.0)
- Predict Transmembrane Topology (MEMSAT3 & MEMSAT-SVM)
- SVM Prediction of TM Topology and Helix Packing (MEMPACK) - NEW
- Fold Recognition (GenTHREADER - quick)
- Fold Recognition (GenTHREADER - with predicted secondary structure)
- Fold Recognition (pDomTHREADER - annotates multiple domain on chains)

Help...

Input Sequence (single letter amino acid code)

Help...
If you wish to test these services follow this link to retrieve a test fasta sequence.

Filtering Options

Done

http://bioinf.cs.ucl.ac.uk/psipred/

2D Threading Disadvantages*

- Reliability is not 100% making most threading predictions suspect unless experimental evidence can be used to support the conclusion
- Does not produce a 3D model at the end of the process
- Doesn't include all aspects of 2° and 3° structure features in prediction process
- PSI-BLAST may be just as good (faster too!)

Making it Better

- Include 3D threading analysis as part of the 2D threading process -- offers another layer of information
- Include more information about the "coil" state (3-state prediction isn't good enough)
- Include other biochemical (ligands, function, binding partners, motifs) or phylogenetic (origin, species) information

3D Threading Servers

- Generate 3D models or coordinates of possible models based on input sequence
- Loopp (version 4)
  - http://cbsuapps.tc.cornell.edu/loopp.aspx
- Phyre
  - http://www.sbg.bio.ic.ac.uk/~phyre/index.cgi
- All require email addresses since the process may take hours to complete

PHYRE Protein Fold Recognition Server

http://www.sbg.bio.ic.ac.uk/~phyre/index.cgi

Most Visited Getting Started Latest Headlines

PHYRE Protein Fold Recognition S...

phyre

Protein Homology/analogY Recognition Engine

Version 0.2

New Phyre server scores highly in CASP8 competition. Results

Phyre has been highlighted in June 2009 Nature PSI Knowledgebase

The Phyre webserver is for Academic use only
For in-house and/or commercial use please click here

Other tools available from our lab (function prediction, docking, etc.)

E-mail Address

Optional Job description

Amino Acid Sequence

Quick Phyre Search

Done

LOOPP @ BioHPC

http://cbsuapps.tc.cornell.edu/loopp.aspx

Most Visited Getting Started Latest Headlines

LOOPP @ BioHPC

APPLICATIONS
(click on a category below to access programs)

Show all Hide all

- Sequence analysis
- Sequence alignment
- Population genetics
- Protein structure
- LOOPP
- MODELLER
- MSR Biomedical
- Other
- Links

MISCELLANEOUS

- Subscribe
- Apps Home
- Clusters Status
- Applications Statistics
- BioHPC Home
- CBSU Home
- CBSU ftp server
- CBSU SeqDB
- CTC Windows Bioinformatics Applications
- DISTRUCT
- T-REX (T-RFLP manager)

LOOPP @ BioHPC

version 4.0

The **LOOPP** (Learning, Observing and Outputting Protein Patterns) server. LOOPP is a fold recognition program based on the collection of numerous signals, merging them into a single score, and generating atomic coordinates based on an alignment into a homologue template structure. The signals we are using include straightforward sequence alignment, sequence profile, threading, secondary structure and exposed surface area prediction. For more information please refer to the LOOPP home page.

LOOPP HAS BEEN UPGRADED. We have upgraded LOOPP to the newest version just in time for CASP8. The new version is much better than the previous one, it uses new scoring technique as well as upgraded database. The improvemnets are descibed in the upcoming *Proteins* paper.

NOTE ABOUT CASP8. This server is participating in CASP8 experiment. All LOOPP predictions for CASP8 targets with the default server parameters are available online here. Please don't submit CASP8 targets again with default parameters - just use the link!

If you have any comments or questions about LOOPP please contact us at loopp@tc.cornell.edu.

It may take several hours to run this program.

Calculations will be carried out on the BioHPC compute cluster at CBSU. You will receive e-mail notifications when the job is submitted, when it starts, and when it is finished. Output will be available via links embedded in the notification e-mails. **For more information about this program and BioHPC interface in general, please visit our Frequently Asked Questions page.**

Please acknowledge us in all publications and presentation of work that used our resources using the following text.

E-mail: (only guests need to use this field, registered users should log in)

Job name: protein (please, no spaces, special characters etc., uderscore is OK)

☐ *Please notify me about LOOPP updates*

Input sequence (amino acids one-letter code only, no names, numbers etc):

Done

Outline

- Secondary Structure Prediction
- Threading (1D and 3D threading)
- Ab initio Structure Prediction

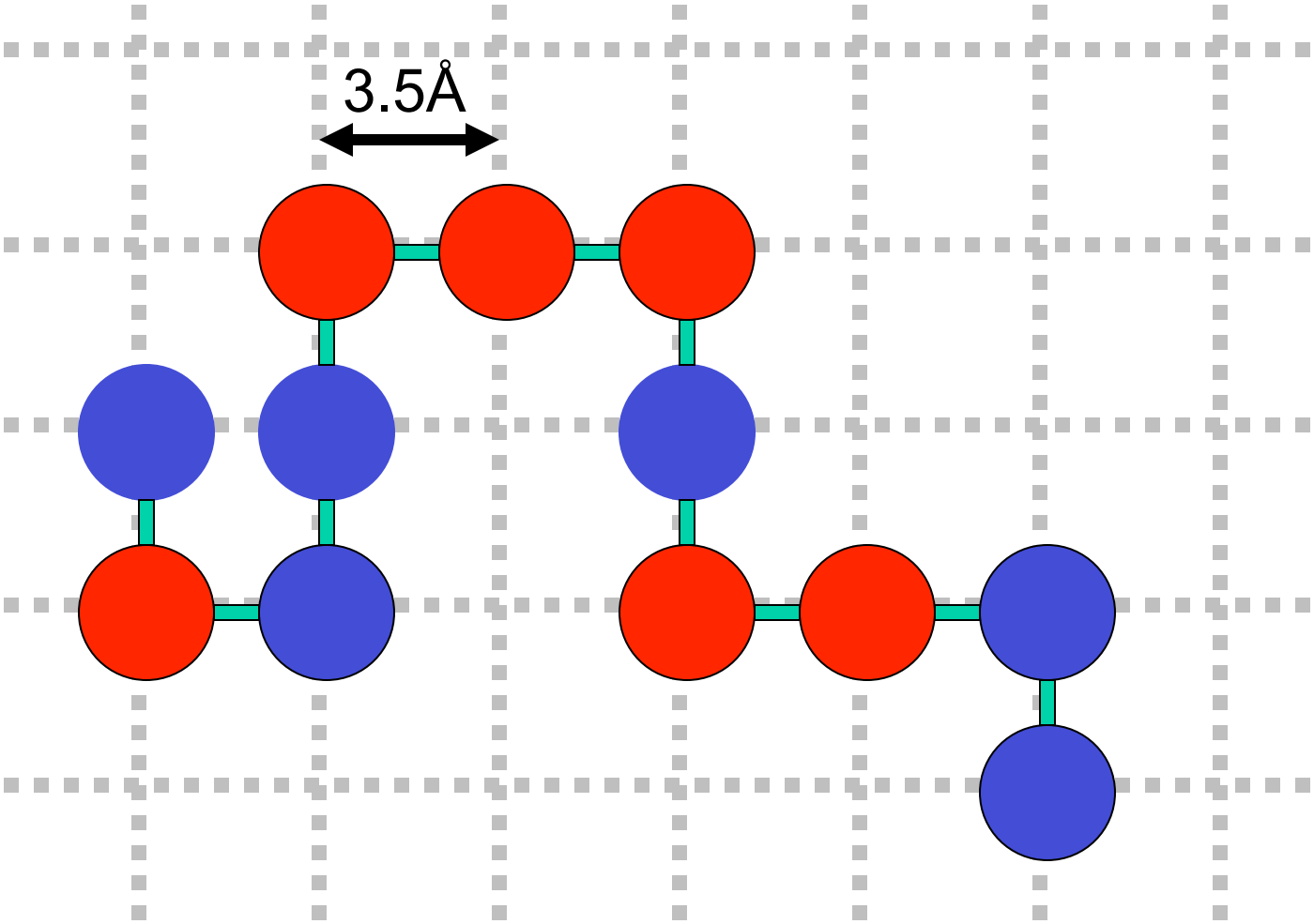
Ab Initio Prediction*

- Predicting the 3D structure without any "prior knowledge"
- Used when homology modelling or threading have failed (no homologues are evident)
- Equivalent to solving the "Protein Folding Problem"
- Still a research problem

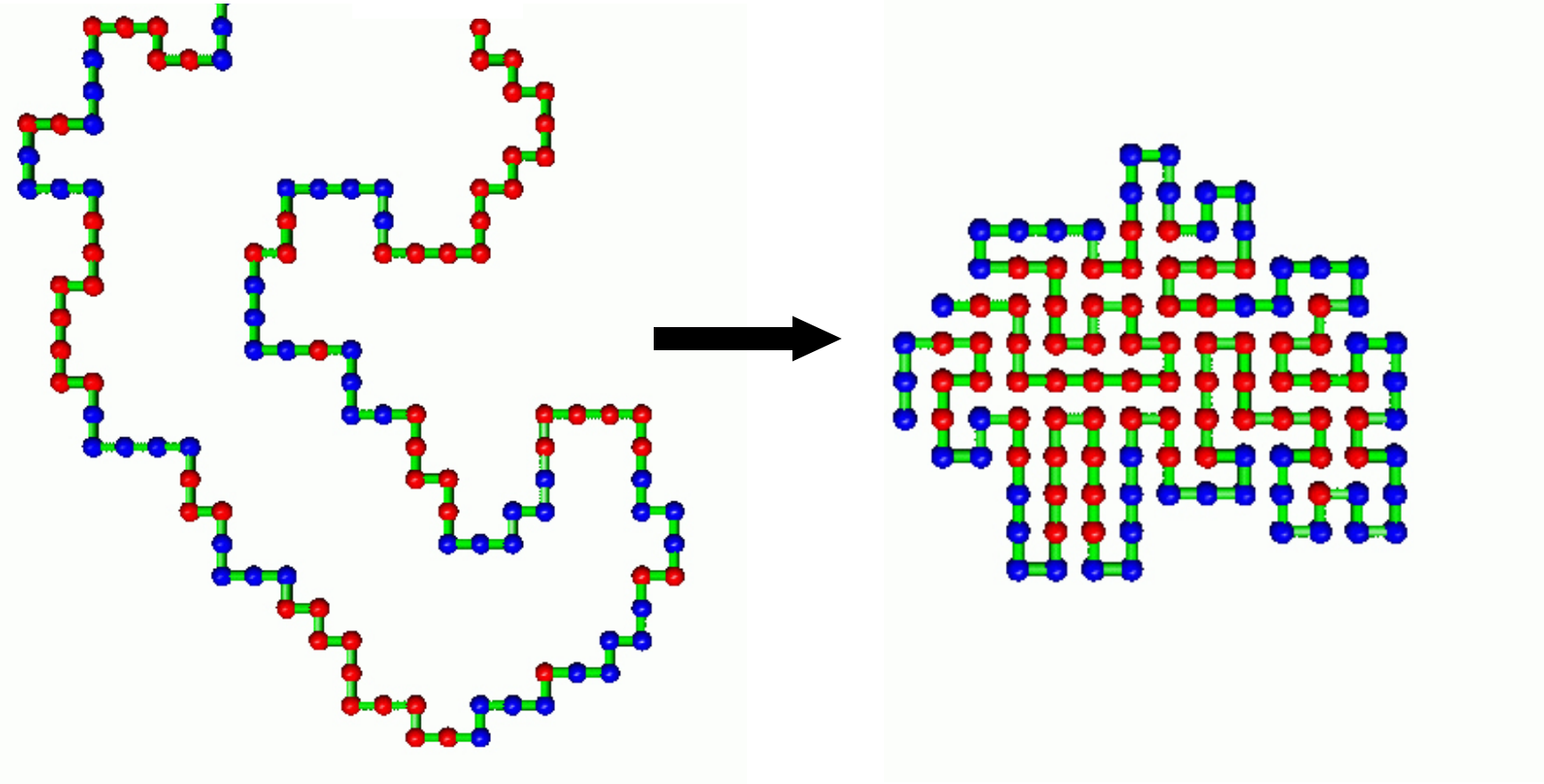
Ab Initio Folding*

- **Two Central Problems**
  - Sampling conformational space (10^{100})
  - The energy minimum problem
- **The Sampling Problem (Solutions)**
  - Lattice models, off-lattice models, simplified chain methods, parallelism
- **The Energy Problem (Solutions)**
  - Threading energies, packing assessment, topology assessment

A Simple 2D Lattice



Lattice Folding



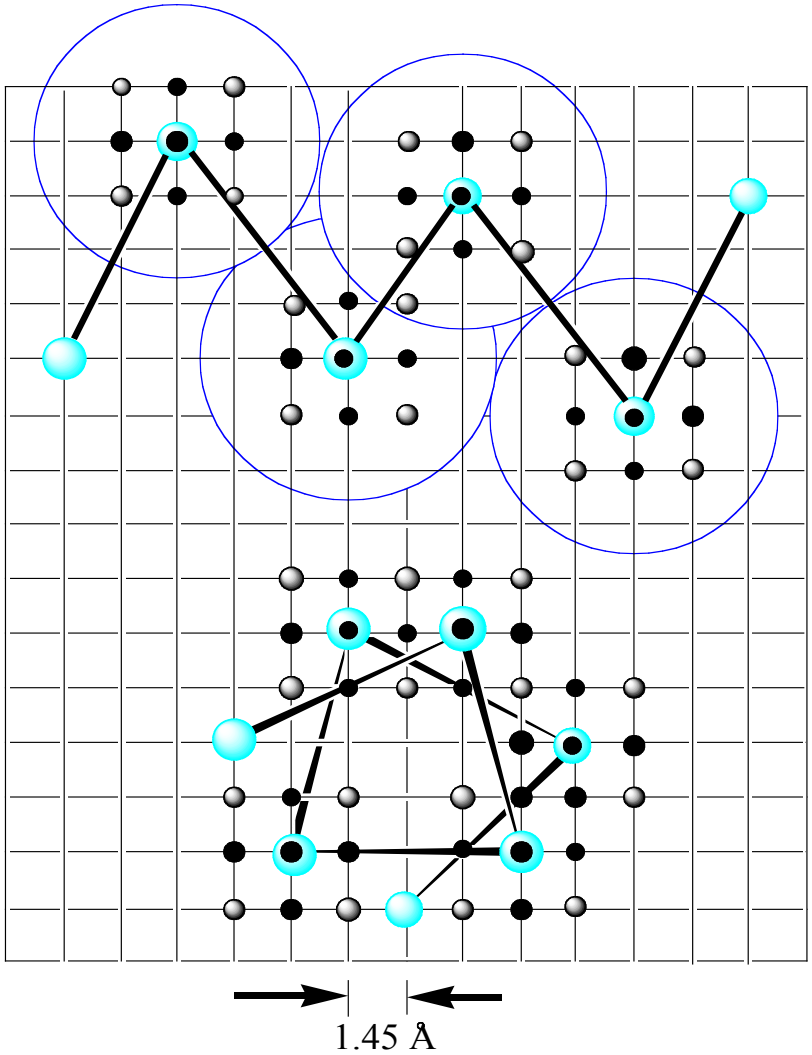
Lattice Algorithm

- *Build a "n x m" matrix (a 2D array)*
- *Choose an arbitrary point as your N terminal residue (start residue)*
- *Add or subtract "1" from the x or y position of the start residue*
- *Check to see if the new point (residue) is off the lattice or is already occupied*
- *Evaluate the energy*
- *Go to step 3) and repeat until done*

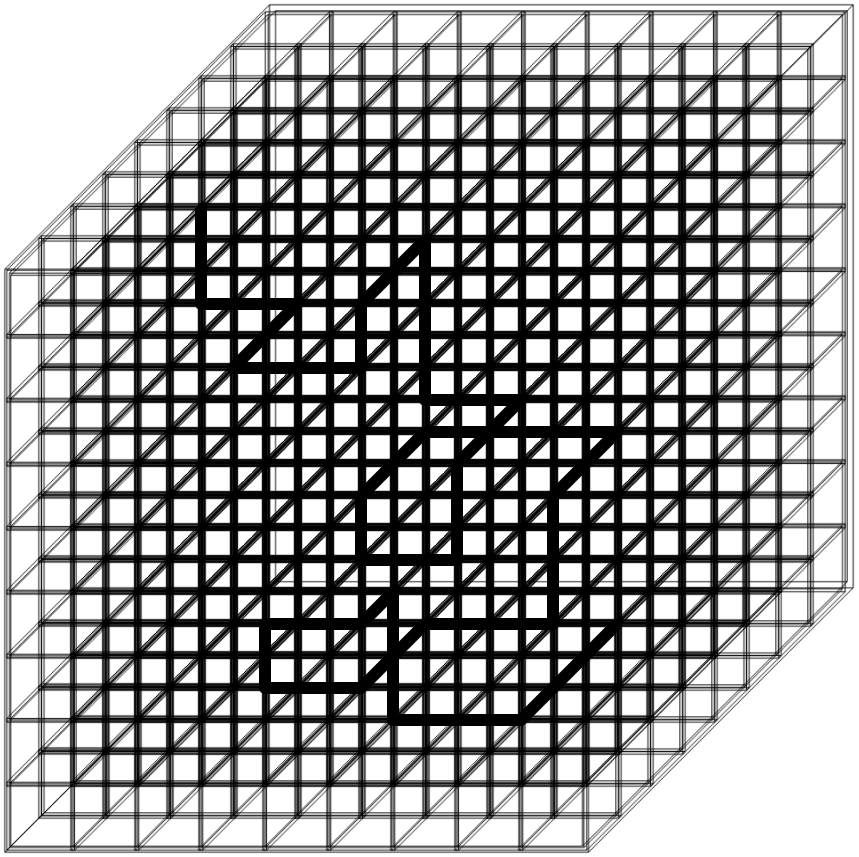
Lattice Energy Algorithm

- *Red = hydrophobic, Blue = hydrophilic*
- *If Red is near empty space E = E+1*
- *If Blue is near empty space E = E-1*
- *If Red is near another Red E = E-1*
- *If Blue is near another Blue E = E+0*
- *If Blue is near Red E = E+0*

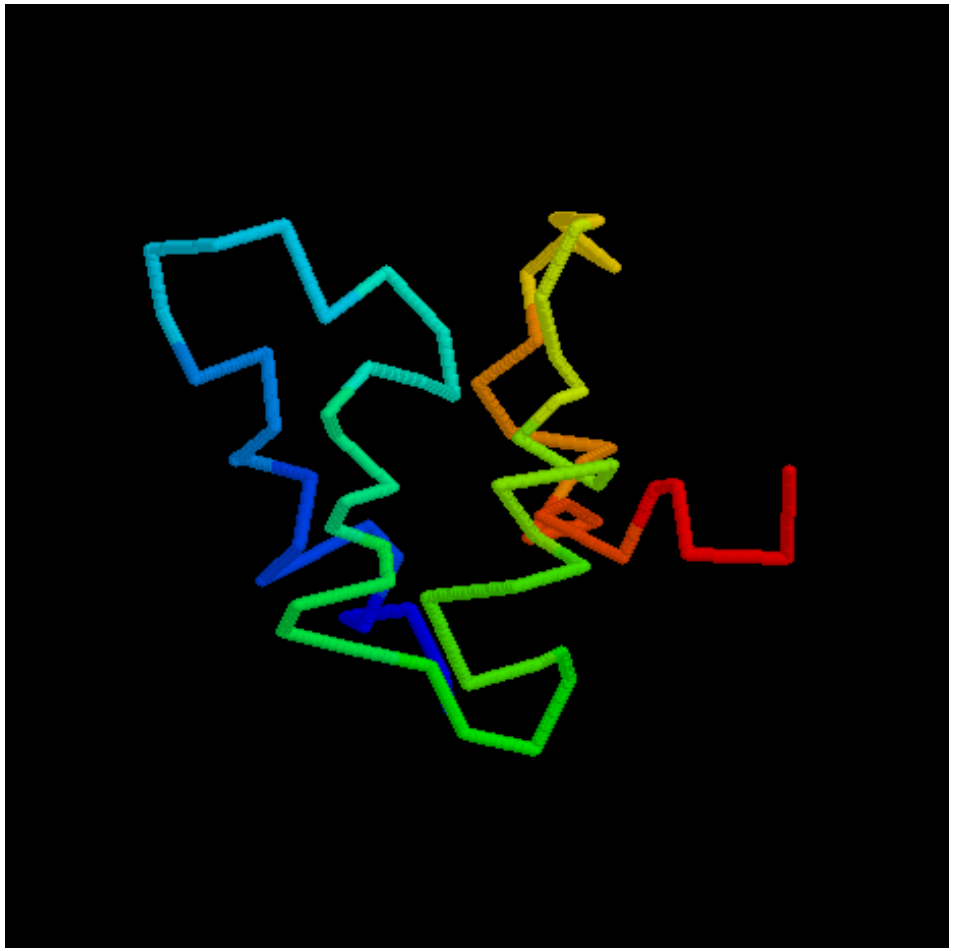
More Complex Lattices



3D Lattices



Really Complex 3D Lattices



J. Skolnick

Lattice Methods*

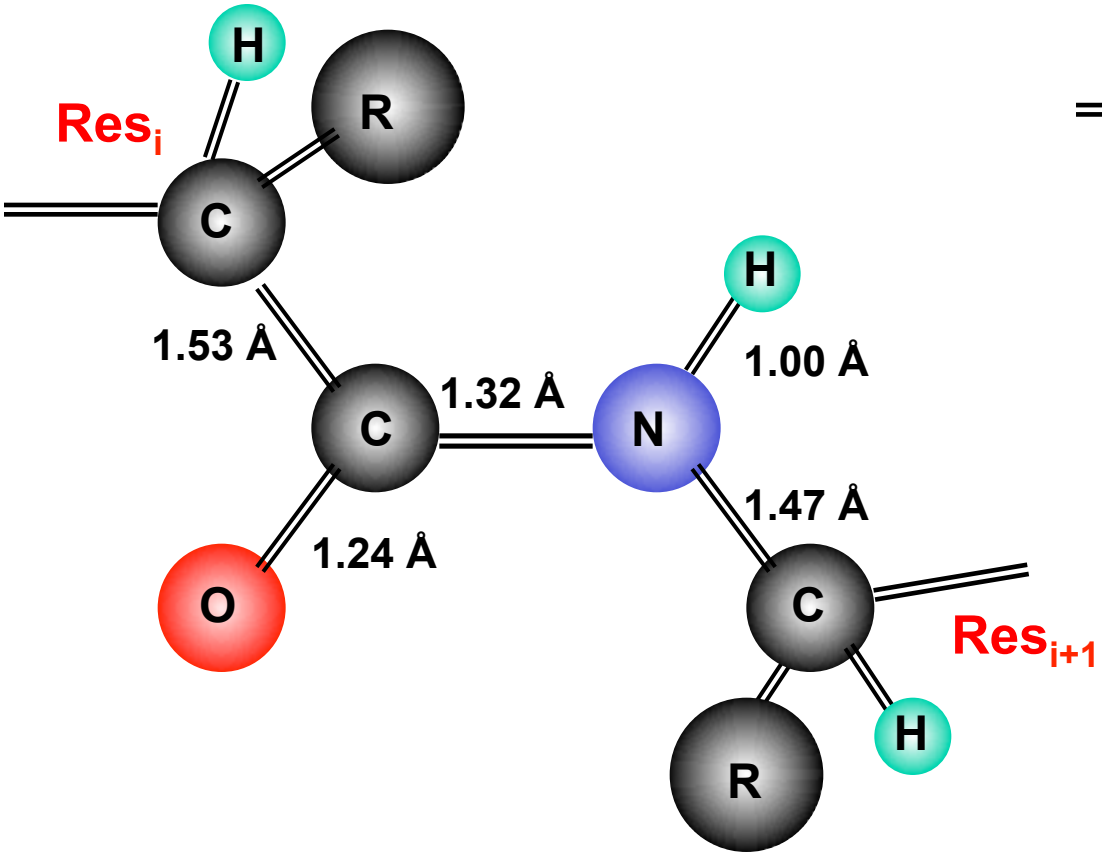
Advantages

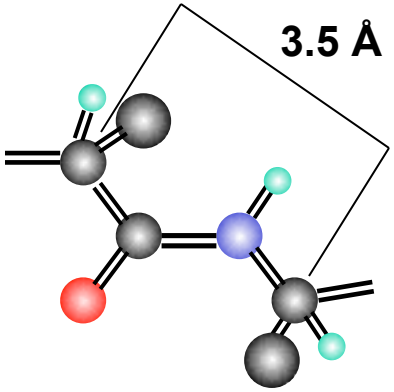
- Easiest and quickest way to build a polypeptide
- Implicitly includes excluded volume
- More complex lattices allow reasonably accurate representation

Disadvantages

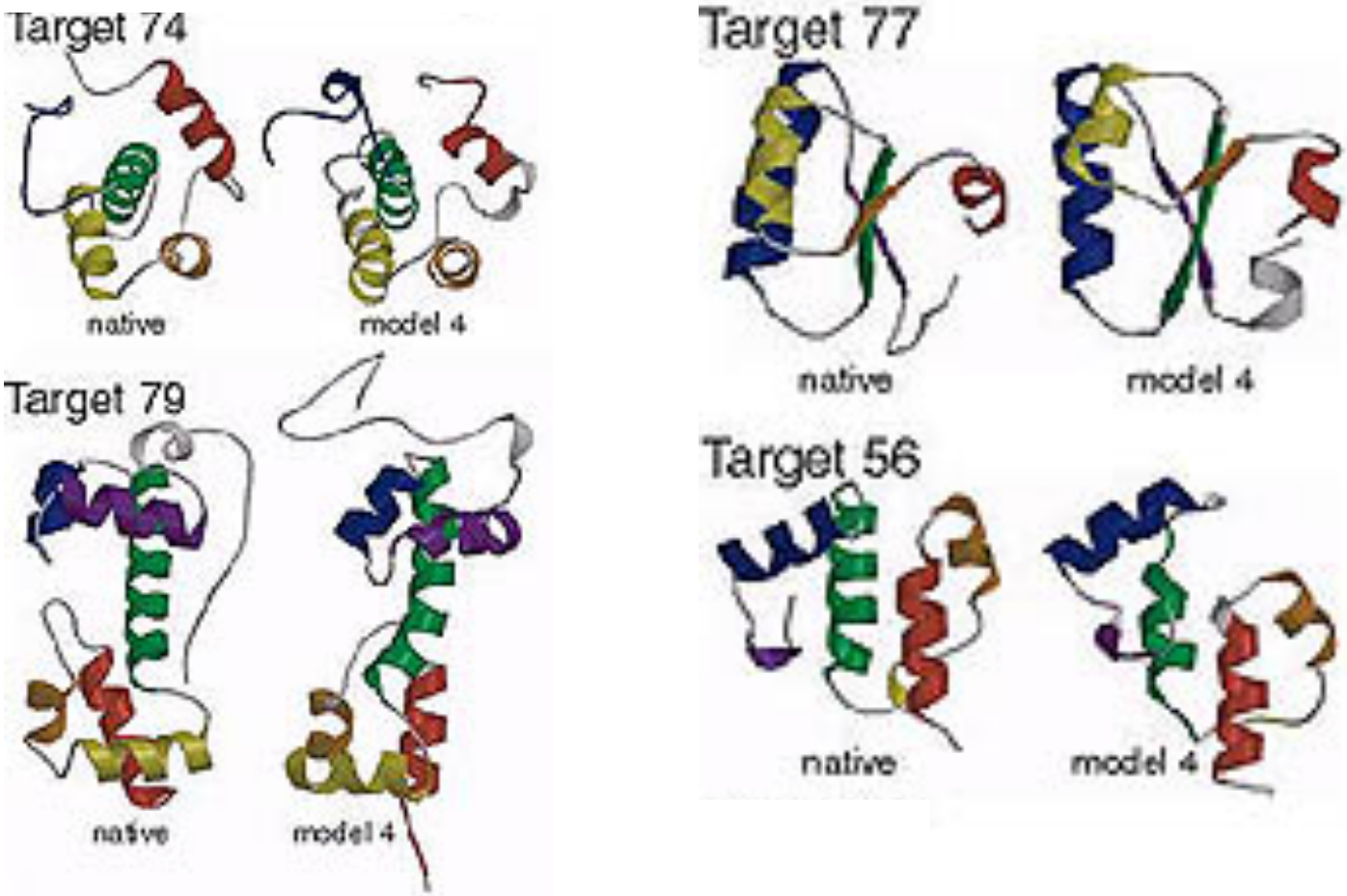
- At best, only an approximation to the real thing
- Does not allow accurate constructs
- Complex lattices are as "costly" as the real thing

Non-Lattice Models





Best Method So Far...*

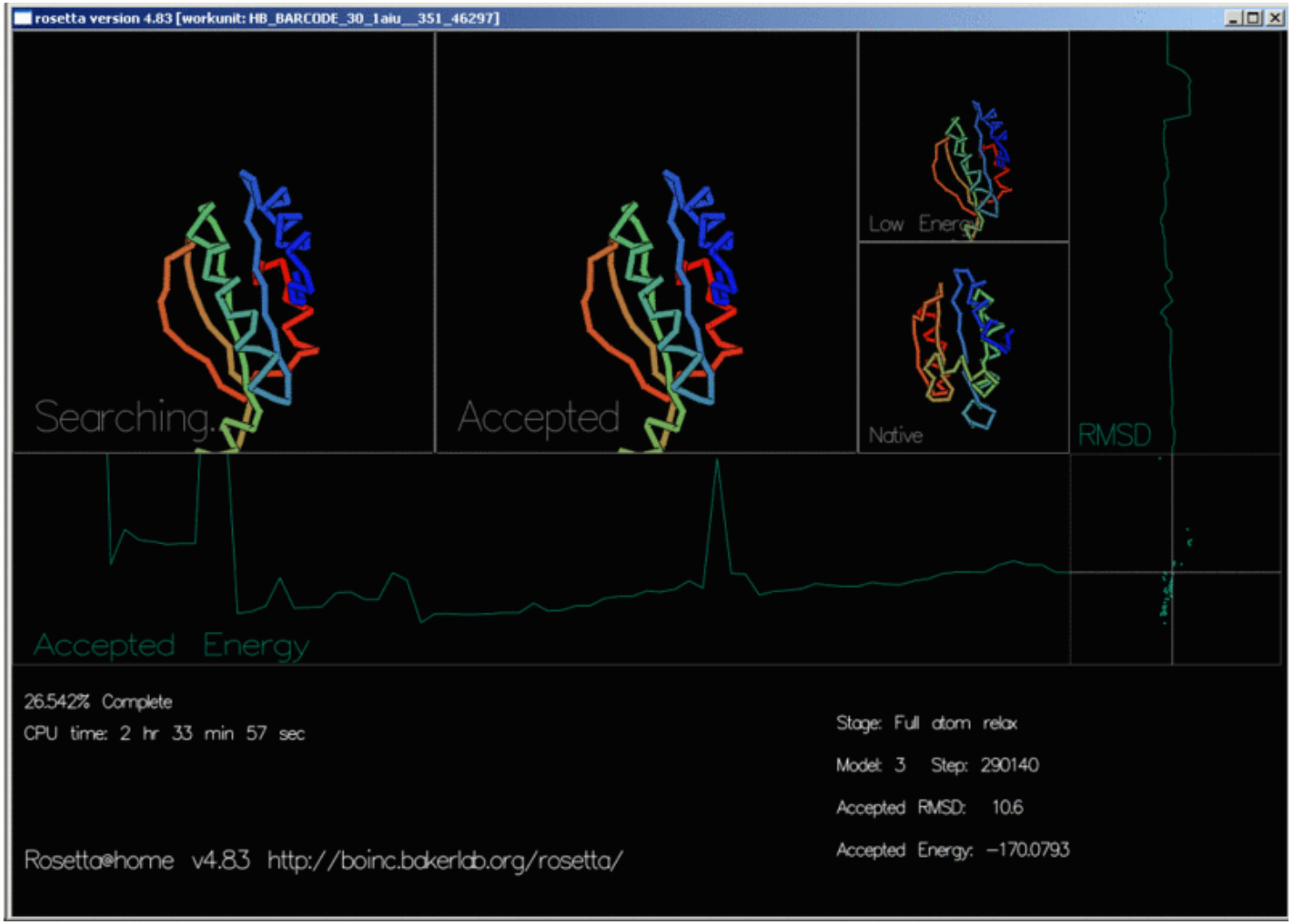


Rosetta - David Baker

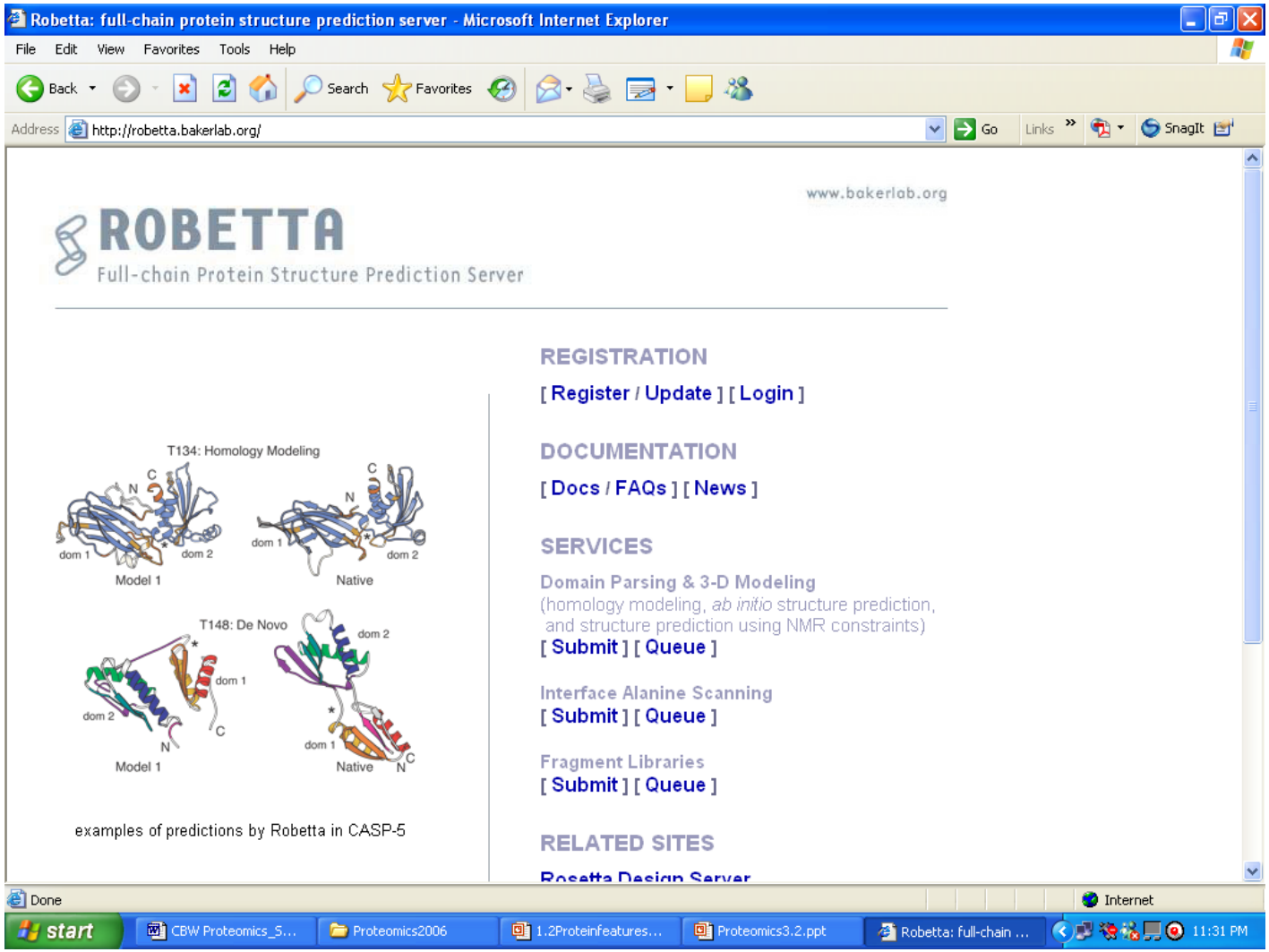
Rosetta Outline*

- Assembles proteins using "fragment assembly" of known protein fragments
- Fragments are 3-9 residues long
- Fragments identified via PSI-BLAST
- Starts with extended chain and then randomly changes conformation of selected regions based on fragment matches
- Evaluates energy using Monte Carlo

Rosetta in Action



Robetta & Rosetta



http://robetta.bakerlab.org/

Robetta

- Allows users predict 3D structures using Rosetta ab-initio method and to do homology modelling too
- Requires considerable computational resources (now hosted at Los Alamos supercomputer facility)
- Requires that users register and login (to track mis-use and abuse)

Another Approach… Distributed Folding

- Attempt to harness the same computational power as BlueGene but by doing on thousands of PC's via a screen saver
- Three efforts underway:
  - http://folding.stanford.edu/
  - http://boinc.bakerlab.org/rosetta/
- You can be part of this exp't too!

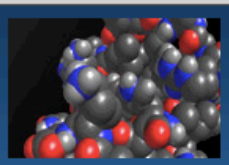
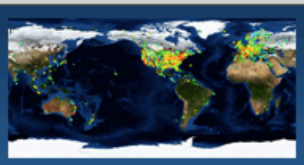
Folding@home – Main

http://folding.stanford.edu/

Most Visited Getting Started Latest Headlines

Folding@home – Main

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
연구말
Lietuvių
Magyar
Nederlands
Norsk
Occitan
فارسی
Polski
Português
Русский
Suomeksi
Svenska
Turkish
Tiếng Việt

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

What is protein folding?

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.

Download Folding@home

Protein folding is linked to disease, such as Alzheimer's, ALS, Huntington's, Parkinson's disease, and many Cancers

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 scientists studying these diseases 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. One can also help by donating funds to the project, via Stanford University.

Make a gift *now!*
Click here to make a gift >

Done

Rosetta@home

http://boinc.bakerlab.org/rosetta/

Most Visited Getting Started Latest Headlines

Rosetta@home

Rosetta@home

Protein Folding, Design, and Docking

Site search

What is Rosetta@home?

HHMI HOWARD HUGHES MEDICAL INSTITUTE

UNIVERSITY OF WASHINGTON

Rosetta@home needs your help to determine the 3-dimensional shapes of proteins in research that may ultimately lead to finding cures for some major human diseases. By running the Rosetta program on your computer while you don't need it you will help us speed up and extend our research in ways we couldn't possibly attempt without your help. You will also be helping our efforts at designing new proteins to fight diseases such as HIV, Malaria, Cancer, and Alzheimer's (See our Disease Related Research for more information). Please join us in our efforts! ***Rosetta@home is not for profit.***

[login/out]

Join Rosetta@home

1. Rules and policies
2. System requirements
3. Download, install, and run BOINC
   (enter the project URL: http://boinc.bakerlab.org/rosetta)
4. A welcome from David Baker
5. Donate

About

- 10 reasons why users crunch Rosetta@home
- Quick Guide to Rosetta@home and Its Graphics
- Play the interactive rosetta game, FoldIt!
- Rosetta@home FAQ
- Rosetta@home Science FAQ
- Disease Related Research
- Research Overview
- Publications
- News & Articles about Rosetta
- David Baker's Rosetta@home Journal
- Rosetta@home promo video
- Technical news

Returning participants

- Your account - view stats, modify preferences
- Results - view your results
- Teams - create or join a team
- Applications

User of the day

Herbert_surft

...nur damit die Rechner auch was zu tun haben...

Server Status as of 26 Oct 2009 22:52:09 UTC

[Scheduler running]

Total queued jobs: **360,803**

In progress: 396,146

Successes last 24h: 232,018

Users (last day) : 269,289 (+79)

Hosts (last day) : 804,167 (+380)

Credits last 24h : 9,712,698

Total credits : 8,333,557,026

TeraFLOPS estimate: 97.127

Oct 26, 2009

Predictor of the day: Congratulations to dookie_2k3 (Team Asturias-Team) for predicting the lowest energy structure for workunit *lr8_combine_smooth_torsion_it00_rama03_A_rlbn_1tig__14898_0* !

...more

XML Available as an RSS feed..

News

Oct,14, 2009

The minirosetta application has been updated to version 1.98. For details and to report bugs, go to this thread.

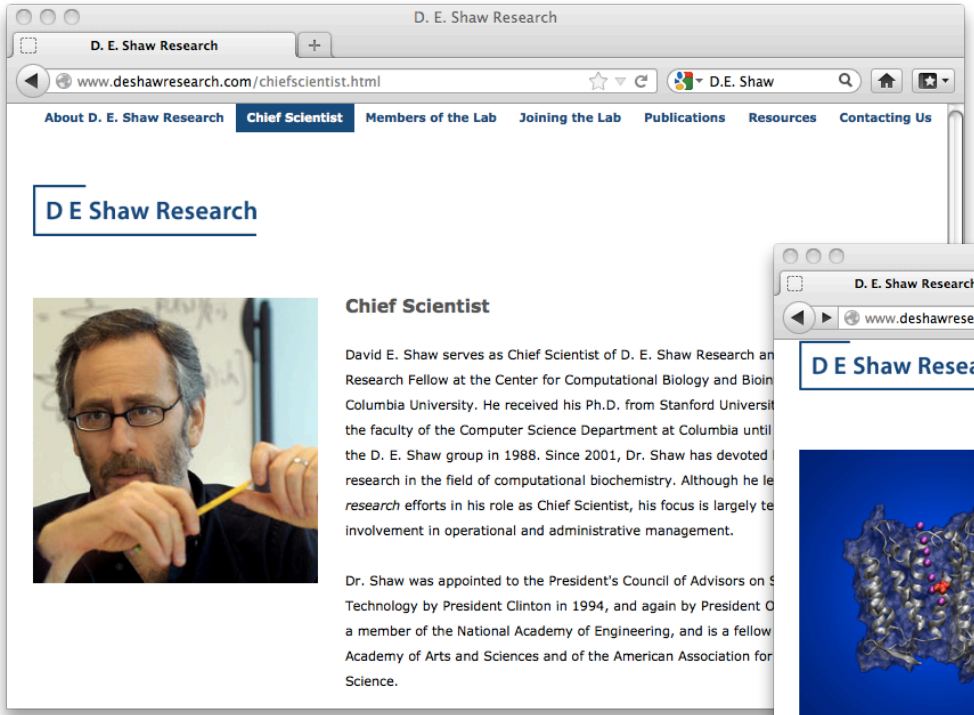
Sep 16, 2009

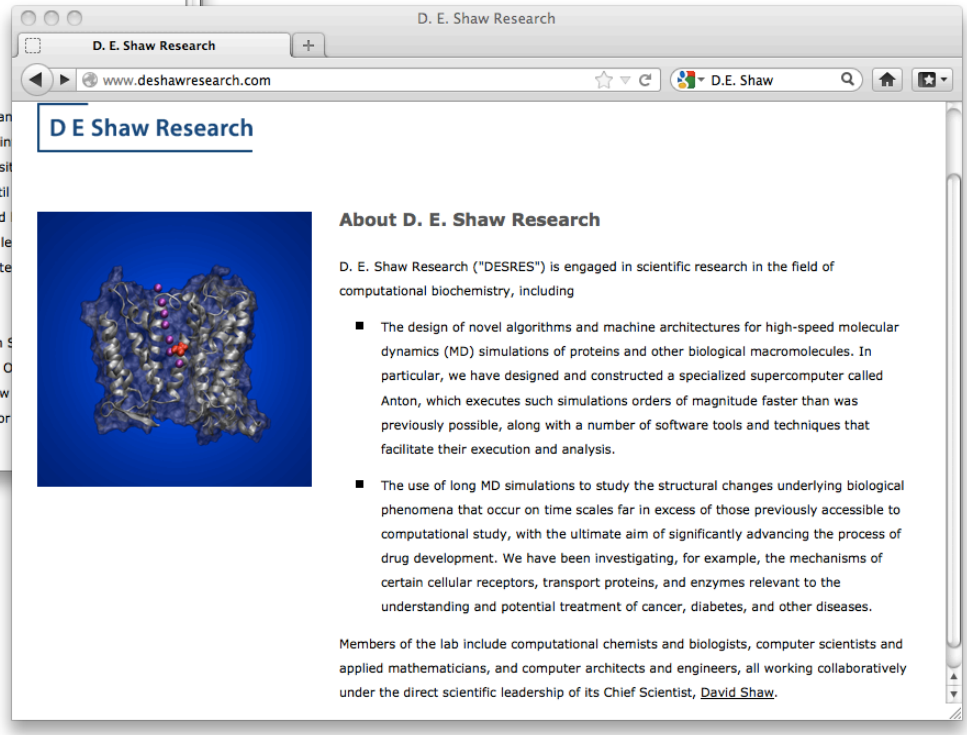
Our filesystem became bogged down late last night. Thanks to Keith, our systems administrator, the project is back online.

Sep 11, 2009

Done

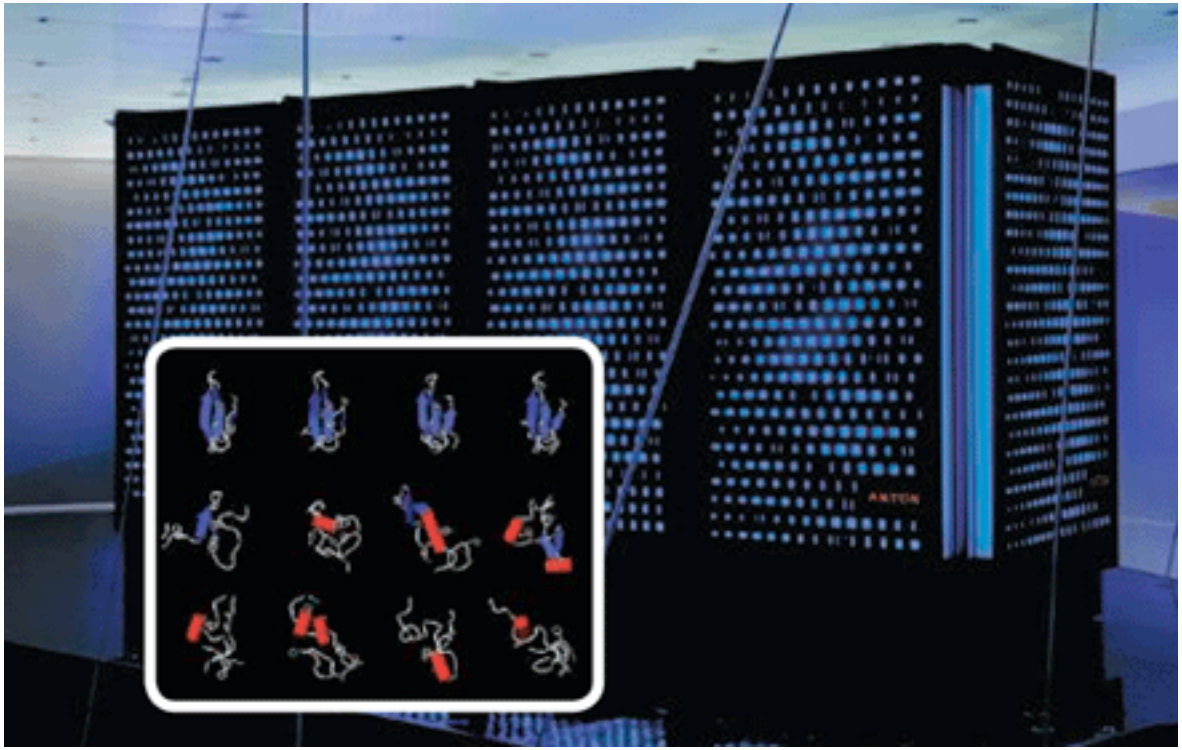
D.W. Shaw Research Institute (MD for 3D Structure Prediction)





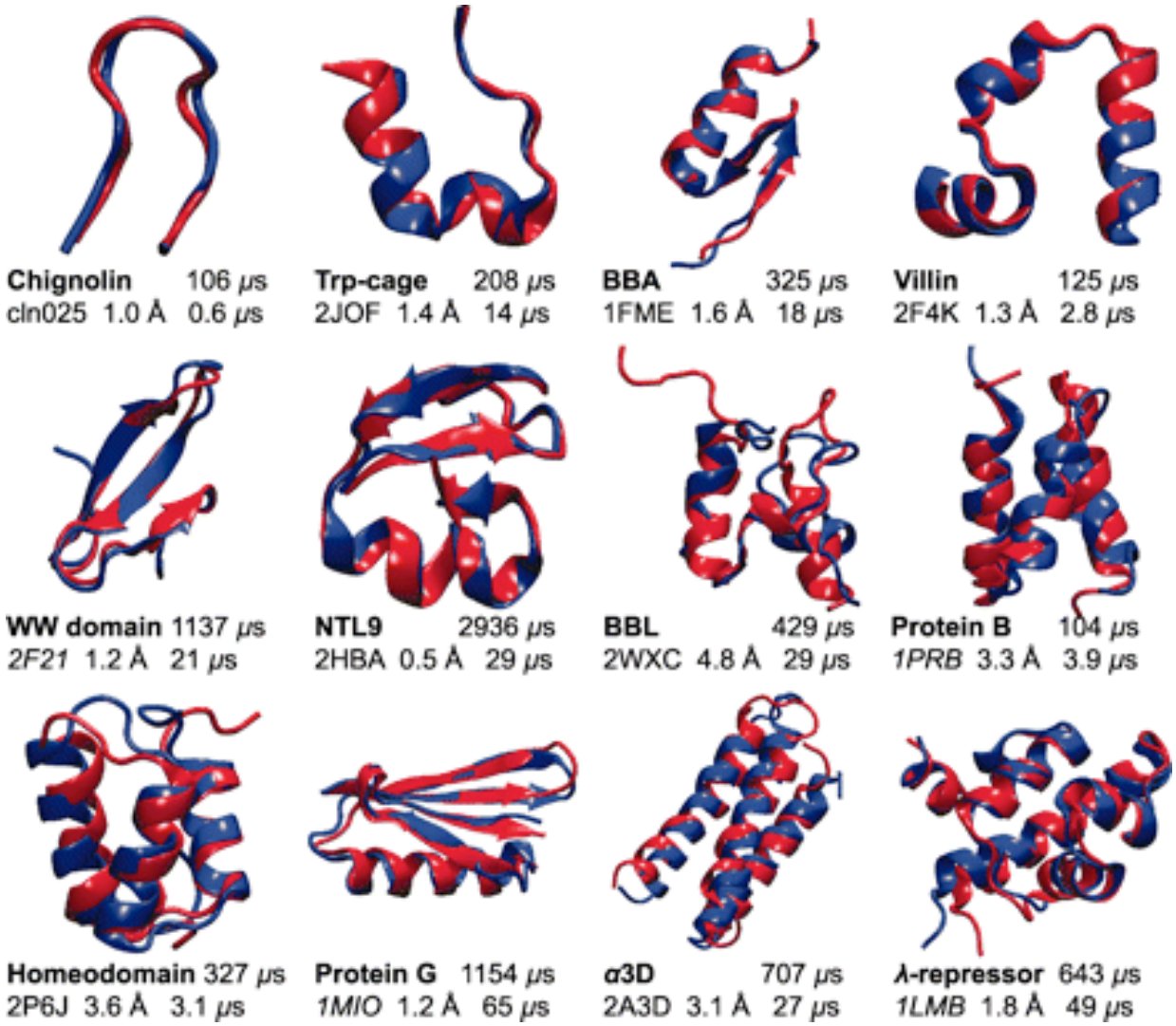
http://www.deshawresearch.com/

David E. Shaw Institute



The Anton Supercomputer – 100 X faster than any other supercomputer for protein folding simulations

How Well Does Anton Do?



Summary

- Structure prediction is still one of the key areas of active research in bioinformatics and computational biology
- Significant strides have been made over the past decade through the use of larger databases, machine learning methods and faster computers
- Ab initio structure prediction remains an unsolved problem (but getting closer)