COMPUTATIONAL STRUCTURAL BIOLOGY

STRUCTURE, SIMULATION, FUNCTION & PREDICTION

Lecture 9

Michael Levitt
Structural Biology, Stanford

http:/csb.stanford.edu/class

STRUCTURE PREDICTION II

Fold Recognition.

Fold Recognition at CASP

Progress at CASP.

Early ab Initio Prediction.

Modern New Fold Prediction.

Winning NF Methods at CASP.

Fold Recognition
Concept 9.1

FOLD RECOGNITION

Comparative Modeling vs. Threading.

What is Threading?

Profile Matching.

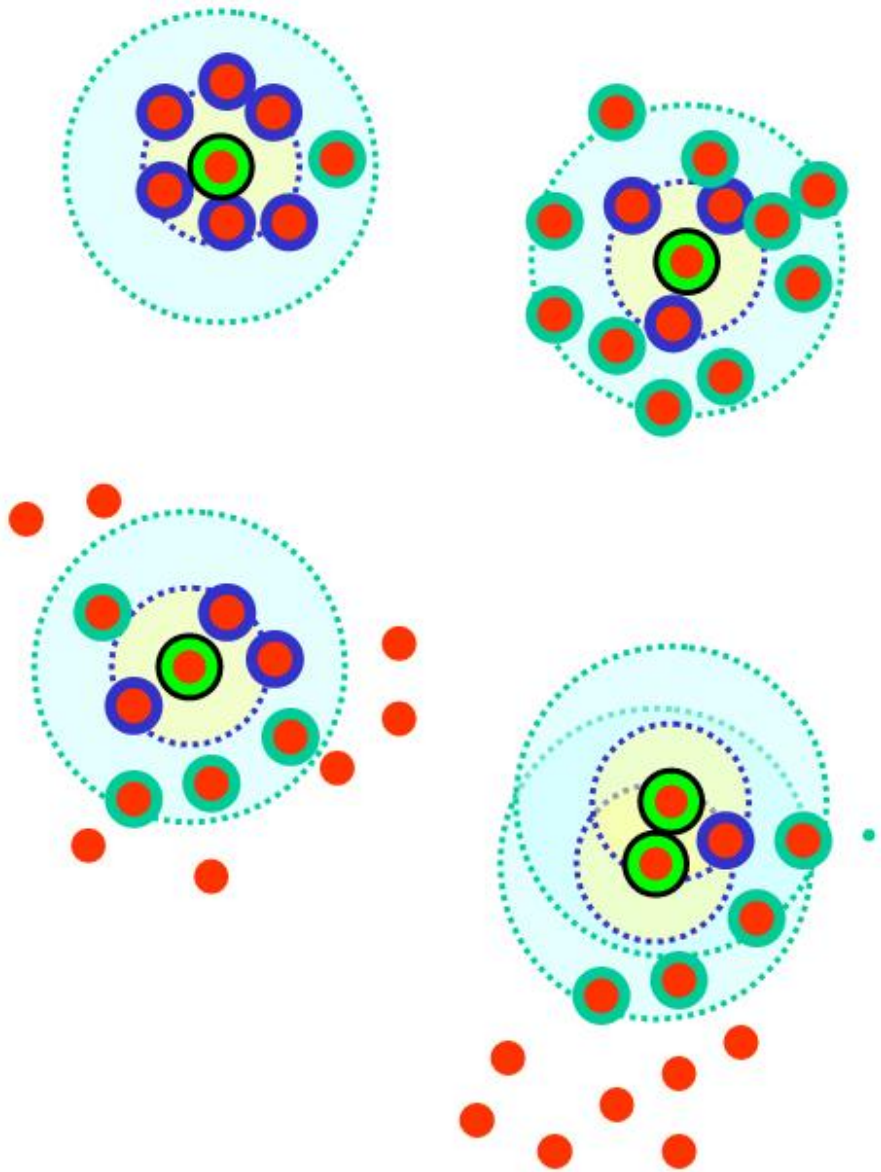
Deletions and insertions using structure.

Getting good alignments.

COMPARATIVE MODELING vs. THREADING

- Both Comparative Modeling and Threading use a known structure as a template.
- In Comparative Modeling the target sequence is very close to the template sequence.
  - Recognition of the template structure is obvious.
- In Threading the target sequence is not very close to the template.
  - Recognition of the template is a problem in itself, hence the term "Fold Recognition".

THREADING COVERS MORE OF PROTEIN SPACE



A protein sequence is a point in sequence space.

Structure is known for these sequences.

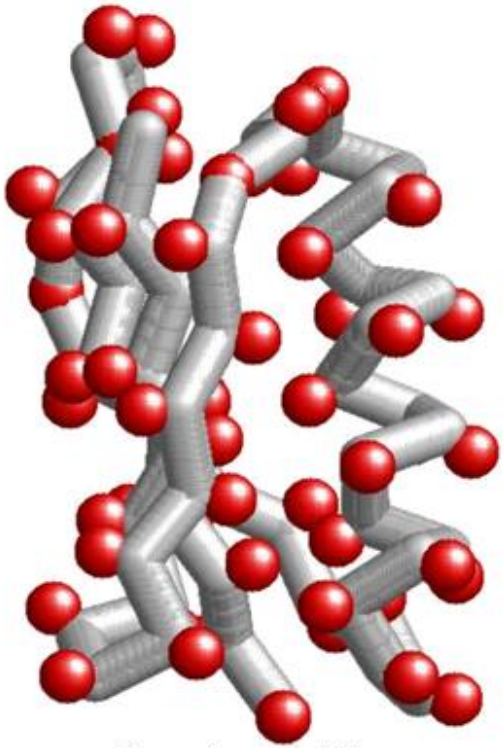
Range of Comparative Modeling.

Range of Threading.

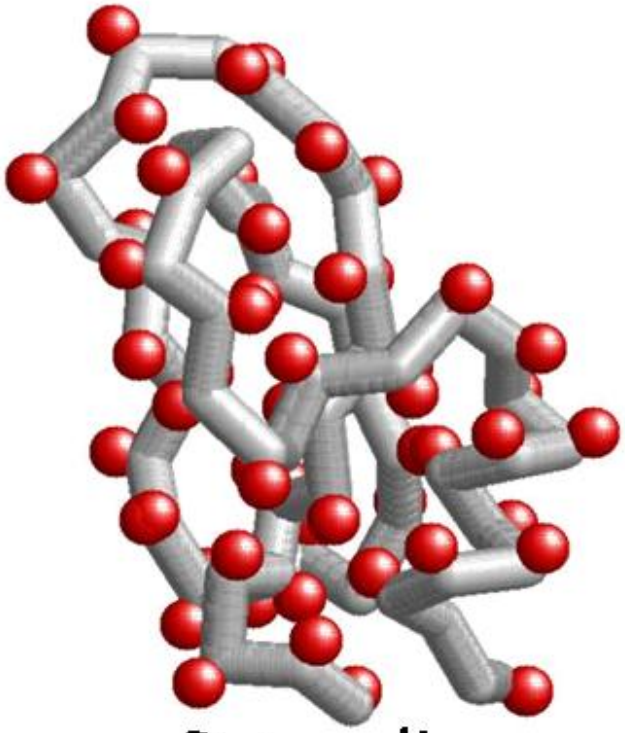
WHAT IS FOLD RECOGNITION

Query Sequence: R V L G F I P T W F A L S K Y

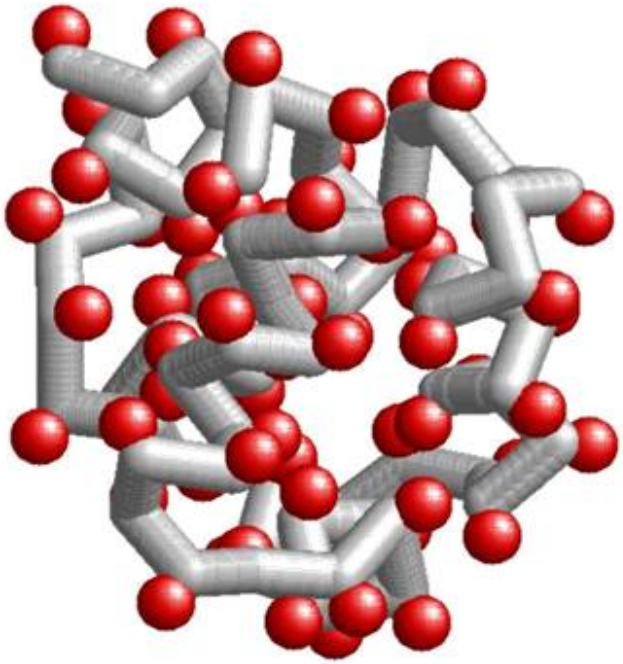
Find the known fold that best fits the query sequence.



1pgb.pdb



5pti.pdb

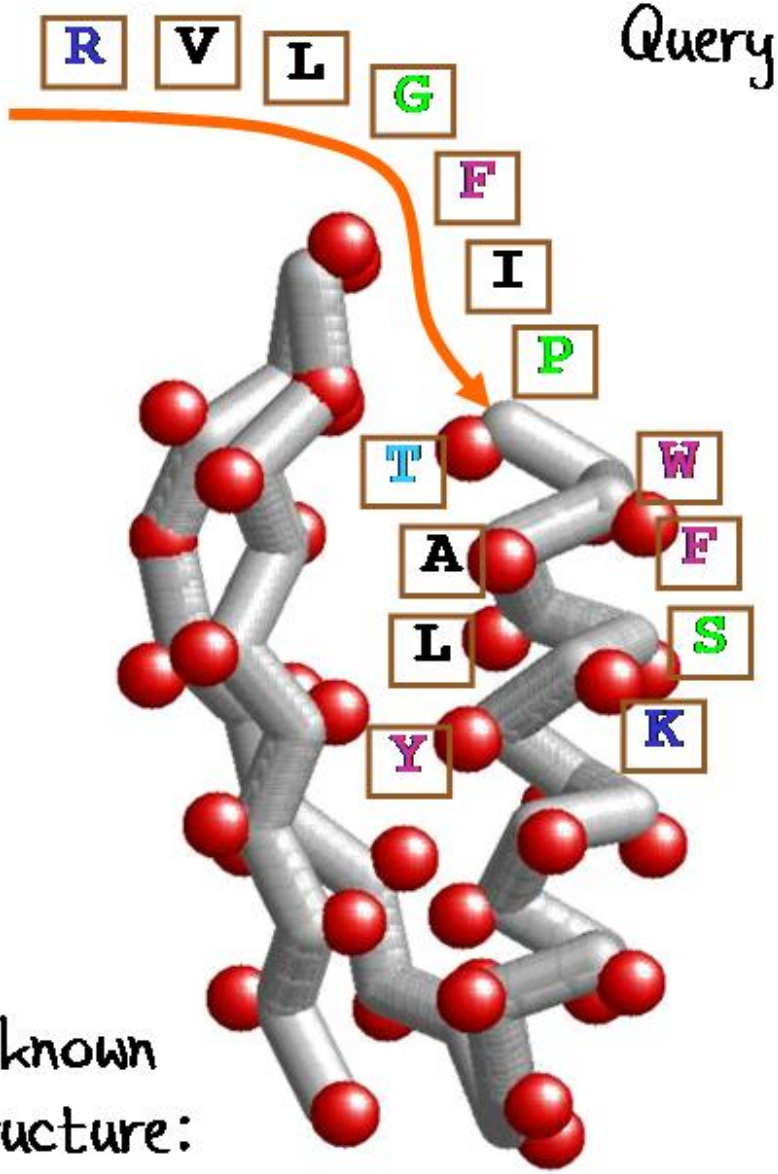


2cro.pdb

Plus 1000 more folds.

WHAT IS THREADING

Query Sequence: RVLGFIPTWFALSKY



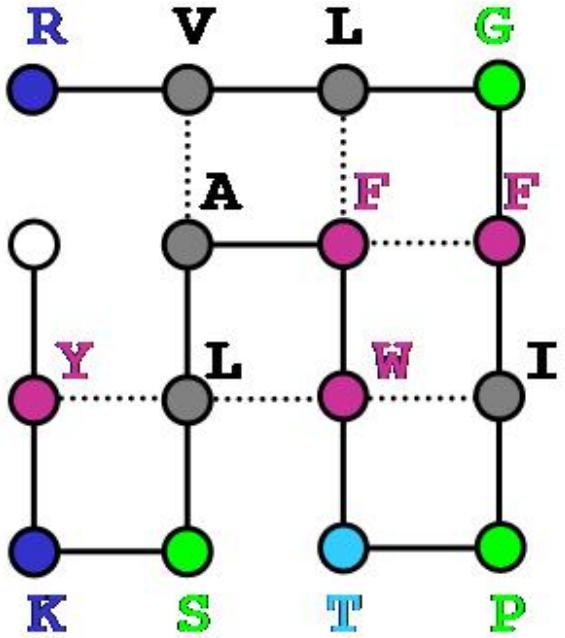
A known structure:

- Thread the sequence onto the structure.
- Use structural properties to evaluate the fit:
  - Local structure
  - Environment
  - Pairwise interactions

SIMPLE EXAMPLE

Query Sequence: **RVLGFIPTWFALSKY**

Many good interactions in this threading:



(1) **G**, **P** & **S** at turns.

(2) **R** & **K** are exposed.

(3) Hydrophobic residues buried.

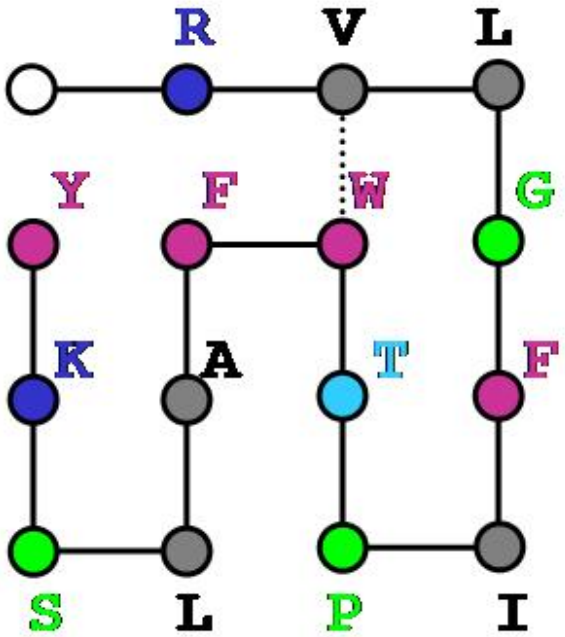
(4) Good pairwise interactions:

V..A, L..F, F..F,
W..I, Y..L, L..W.

A DIFFERENT THREADING

Query Sequence: **RVLGFIPTWFALSKY**

Not many good interactions in this threading:



(1) **G**, **P** & **S** not all at turns.

(2) **R** & **K** are buried.

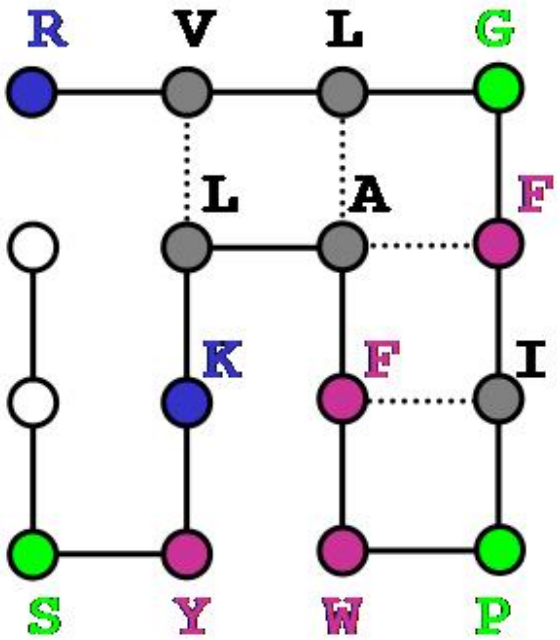
(3) Polar **T** is buried.

(4) Just one pair-wise interaction:

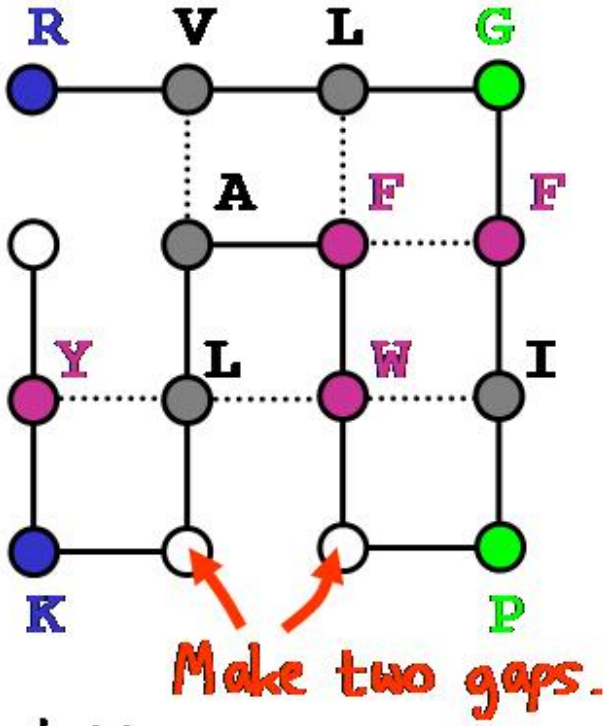
W..V.

GAPS IN THREADING

Query Sequence: **RVLGFIPWFALKYS**



Not good.
Expose Y & W, bury K
and have just 3 good pairs.



Much better now.
All hydrophobic residues are
buried and have 6 good pairs.

FROZEN APPROXIMATION

- Threading is like structural alignment.
- It cannot be solved by dynamic programming.

  No optimal solution is guaranteed.
- Use the Frozen Approximation:
  - Use the current alignment to define the environment for the next cycle.
  - Continue to convergence.

SEQUENCE PROFILE

```
1    1bpi     ..RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGA.
2    1bpi     ..RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGA.
3    1bzxI    ..RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGA.
4    1fakI    ..APDFCLEPPYDGPCRALHLRYFYNAKAGLCQTFYYGGCLAKRNNFESAEDCMRTC....
5    1bunB    ..RHPDCDKPPDTKICQTVVRAFYYKPSAKRCVQFRYGGCNGNGNHFKSDHLCRCECLEY.
6    1bf0     ..PPWYCKEPVRIGSCKKQFSSFYFKWTAKKCLPFLFSGCGGNANRFQTIGECRKKCLGK.
```

		F	C	L	E	P	P	Y	T	G
1	1bpi									
2	1bpi									
3	1bzxI									
4	1fakI									
5	1bunB									
6	1bf0									

Number of	A	0	0	0	0	0	0	0	0	0
Number of	C	0	6	0	0	0	0	0	0	0
Number of	D	1	0	1	0	0	0	1	1	0
Number of	E	0	0	0	5	0	0	0	0	0
Number of	F	4	0	0	0	0	0	0	0	0
Number of	G	0	0	0	0	0	0	0	0	5
Number of	H	0	0	0	0	0	0	0	0	0
Number of	I	0	0	0	0	0	0	0	1	0
Number of	K	0	0	1	1	0	0	0	0	1
Number of	L	0	0	4	0	0	0	0	0	0
Number of	M	0	0	0	0	0	0	0	0	0
Number of	N	0	0	0	0	0	0	0	0	0
Number of	P	0	0	0	0	6	5	0	0	0
Number of	Q	0	0	0	0	0	0	0	0	0
Number of	R	0	0	0	0	0	0	1	0	0
Number of	S	0	0	0	0	0	0	0	0	0
Number of	T	0	0	0	0	0	0	0	4	0
Number of	V	0	0	0	0	0	0	0	0	0
Number of	W	0	0	0	0	0	0	0	0	0
Number of	Y	1	0	0	0	0	1	4	0	0
Number of	.	0	0	0	0	0	0	0	0	0

- For each position along the sequence, tabulate how often each of the 20 amino acids occur (also count the gap denoted ".").
- Convert to frequencies.
- A profile is always 21 by N no matter how many sequences are compared.

Also called Position Sensitive Scoring Matrix.

PROFILE MATCHING

Profile 1

#	1	2	3	4	5	6	7	8	9	10
A	1	0	0	4	1	1	0	0	3	4
E	0	2	0	0	0	7	0	0	1	1
K	4	7	0	0	0	2	0	0	1	0
L	0	0	0	1	2	0	4	0	1	0
P	0	0	9	0	0	0	0	0	1	0
S	3	0	0	4	0	0	1	9	1	4
Y	1	0	0	0	6	0	4	0	1	0

Profile 2

#	1	2	3	4	5	6	7	8	9	10
A	0	0	2	3	1	0	0	3	4	1
E	1	0	0	1	7	0	0	1	1	0
K	8	0	1	0	2	5	2	1	0	4
L	0	0	1	2	0	0	2	1	0	0
P	0	9	0	0	0	0	4	0	0	0
S	0	0	4	0	0	0	1	1	4	3
Y	0	0	1	3	0	4	0	2	0	1

- We have two profiles each built from a multiple-sequence alignment.
- One is built around a sequence of a known structure (from SCOP).
- The other is built around the query sequence.

PROFILE MATCHING

Profile 2

#	1	2	3	4	5	6	7	8	9	10
A	0	0	2	3	1	0	0	3	4	1
E	1	0	0	1	7	0	0	1	1	0
K	8	0	1	0	2	5	2	1	0	4
L	0	0	1	2	0	0	2	1	0	0
P	0	9	0	0	0	0	4	0	0	0
S	0	0	4	0	0	0	1	1	4	3
Y	0	0	1	3	0	4	0	2	0	1

Profile 1

#	A	E	K	L	P	S	Y
1	1	0	4	0	0	3	1
2	0	2	7	0	0	0	0
3	0	0	0	0	9	0	0
4	4	0	0	1	0	4	0
5	1	0	0	2	0	0	6
7	0	0	0	4	0	1	4
6	1	7	2	0	0	0	0
8	0	0	0	0	0	9	0
9	3	1	1	1	1	1	1
10	4	1	0	0	0	4	0

4	0	2	0	1	3	1	1	2	3
7	0	0	0	3	4	1	1	0	3
0	9	0	0	0	0	4	0	0	0
0	0	3	1	0	0	0	2	4	2
0	0	1	3	0	3	0	2	0	0
2	0	0	1	6	1	0	1	1	1
0	0	1	2	0	2	1	1	0	0
0	0	4	0	0	0	1	1	4	3
1	1	1	1	1	1	1	1	2	1
0	0	3	1	1	0	0	2	4	2

- Compare Profile 1 with Profile 2 using a dot product. For the first element this is:

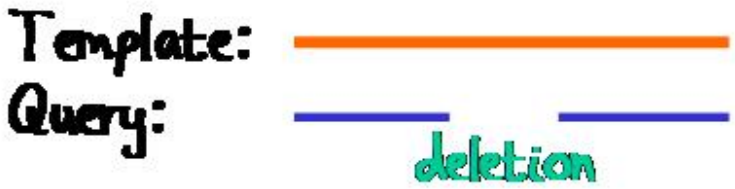
  (1 0 4 0 0 3 1) •
  (0 1 8 0 0 0 0) =
  (0 0 32 0 0 0 0) = 32

  normalized to 4.

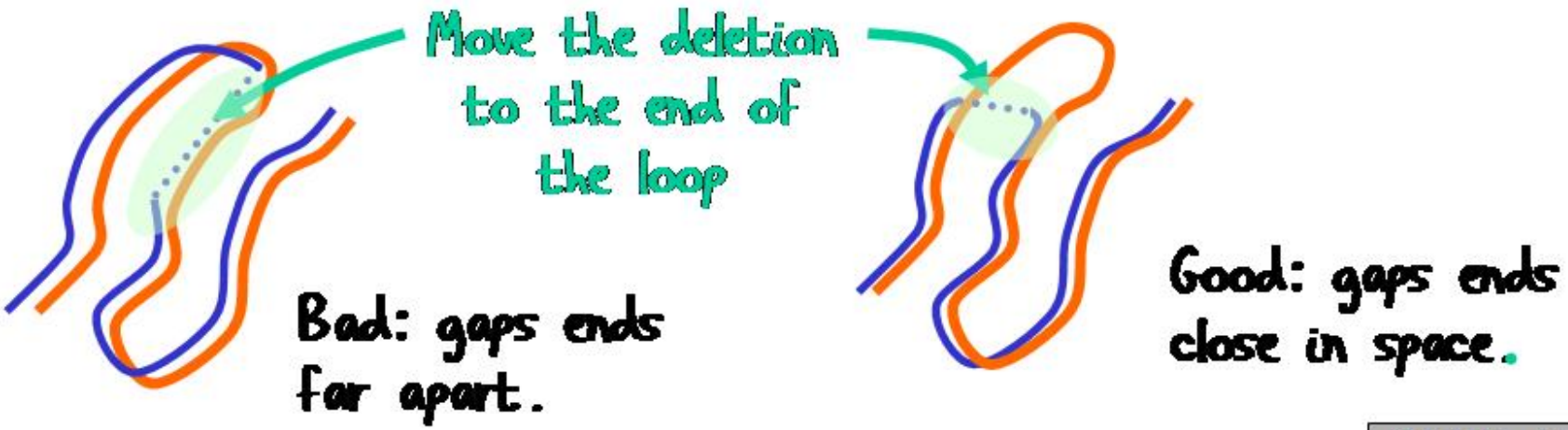
- Use the resulting similarity matrix for dynamic programming.

DELETIONS USING STRUCTURE

- Normally gap penalty increases linearly with size of gap for both insertions or deletions.



- With a known structure, deletions should leave the ends of the gap close in space so that the ends can easily be joined.



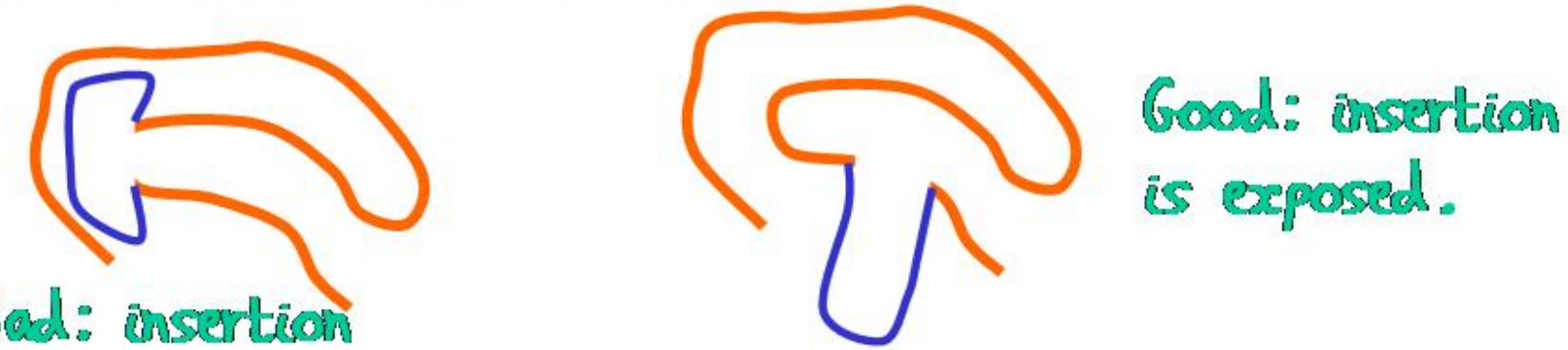
INSERTIONS USING STRUCTURE

Template: —— ————————

Query: ————————————

- No insertions in the middle of segments of secondary structure.

- Insertions should not be buried.



GETTING GOOD ALIGNMENTS SUMMARY

Combine many different components:

- Profiles from multiple sequence and structure alignments.
- Profiles from observed and predicted secondary structure.
- Residue environment. Exposed/Buried, Polar/Nonpolar.
- Adjust positions of gaps using structure.

Fold Recognition at CASP
Concept 9.2

FOLD RECOGNITION AT CASP

CASP4 results.

CASP5 results.

Fold Recognition Summary.

CASP4 THREADING: SERVERS & CREATORS

Threading Targets

RANK	GROUP	N	ZSUM
4	GODZIK	9	15.28
34	FFAS	2	3.16
93	PDB-BLAST	0	0.42
5	STERNBERG	9	14.80
20	3D-PSSM	4	4.80
7	KARPLUS	9	13.91
23	SAM-T99	4	4.64
10	BLUNDELL	5	11.37
75	FUGUE-CAM	1	1.19
14	FISCHER	6	6.91
22	INBGU	4	4.65
36	JONES	3	2.88
46	MGENTHREADER	3	2.39
53	GENTHREADER	2	2.10

- The servers generally come from people who do well.
- The creators are always better than their servers.

FOLD RECOGNITION META SERVERS ARE WINNERS

- Meta-Servers help win fold recognition at CASP5

RANK	GROUP	N	ZSCORE
1	Ginalski	9	24.2
4	Bionfo.pl	6	16.8
48	BIOINFO.PL-BASICC	1	2.7
3	Baker	8	19.5
6	BAKER-ROBETTA	6	14.5
12	Fischer	5	9.0
15	3D-SHOTGUN-3DS5	5	8.0
22	3D-SHOTGUN-INBGU	4	6.7
20	Bujnicki-Janusz	5	6.7
25	GeneSilico	4	6.2
37	GENESILICO.PL	3	3.5

- Meta-servers better than individual servers.
- Experts do better than meta-servers with the meta-server output.
- SHGU is best self-contained meta-server.

Progress at CASP
Concept 9.3

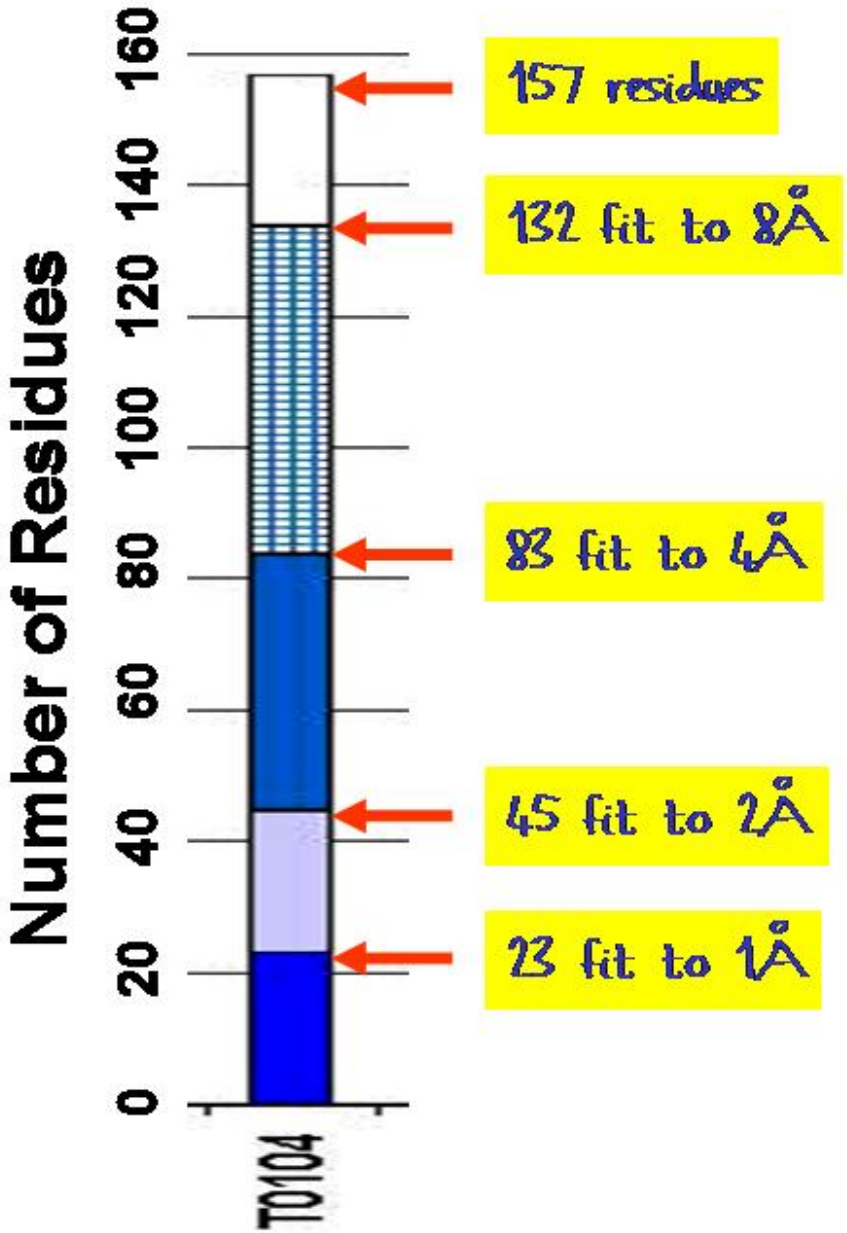
PROGRESS AT CASP

GDT Results at CASP1 to CASP4.

Measuring Target Difficulty.

Progress at CASP.

EXPLANATION OF SCORING INDEX



- Give the predictor maximum leeway.
- Find the largest subset of residues that match the native structure to thresholds of 1Å, 2Å, 4Å and 8Å.
- If 83 residues match with no CA deviation greater than 4Å, the RMS deviation will be about 2Å, a very close fit.

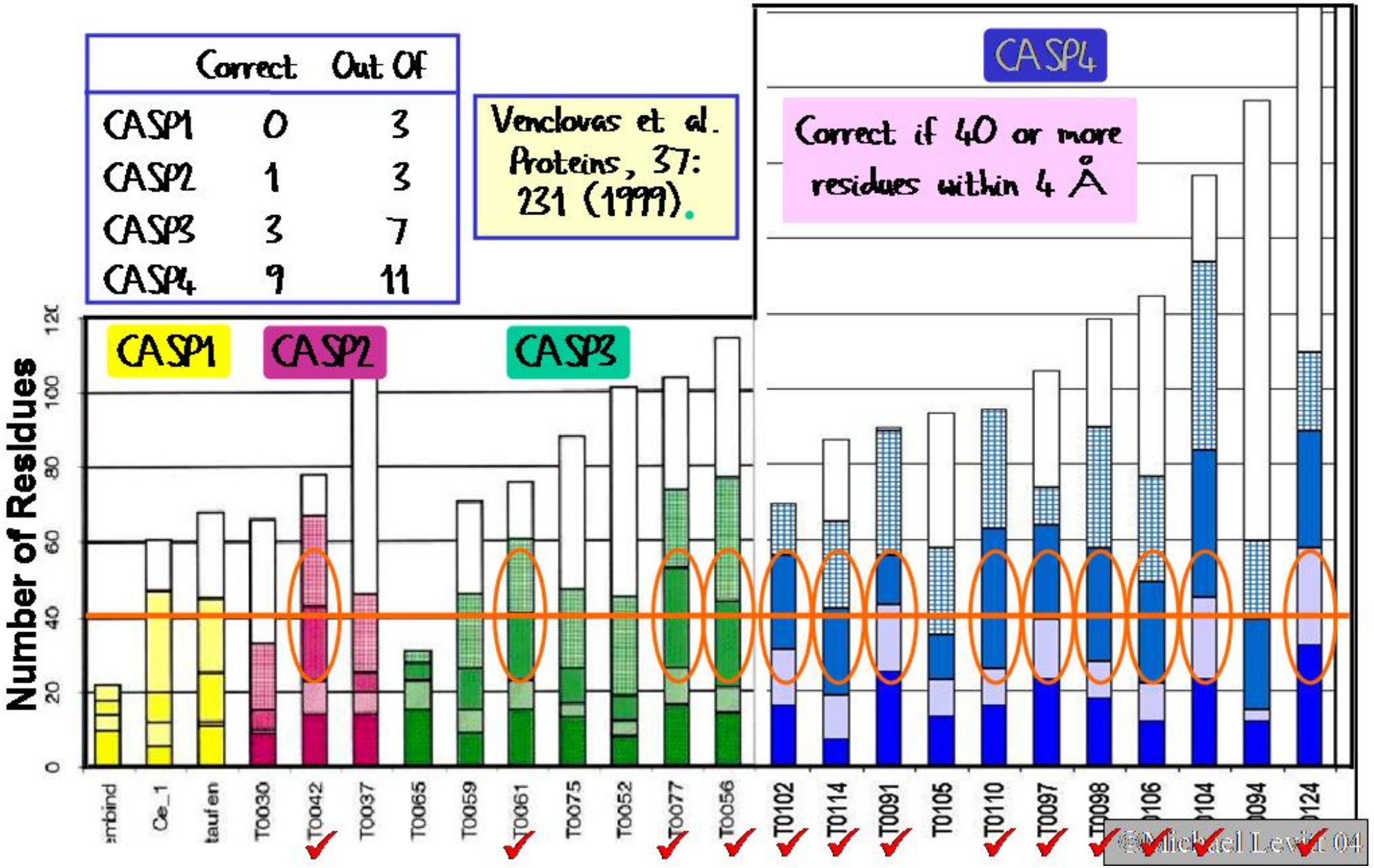
BEST RESULTS AT CASP1 TO CASP4

	Correct	Out Of
CASP1	0	3
CASP2	1	3
CASP3	3	7
CASP4	9	11

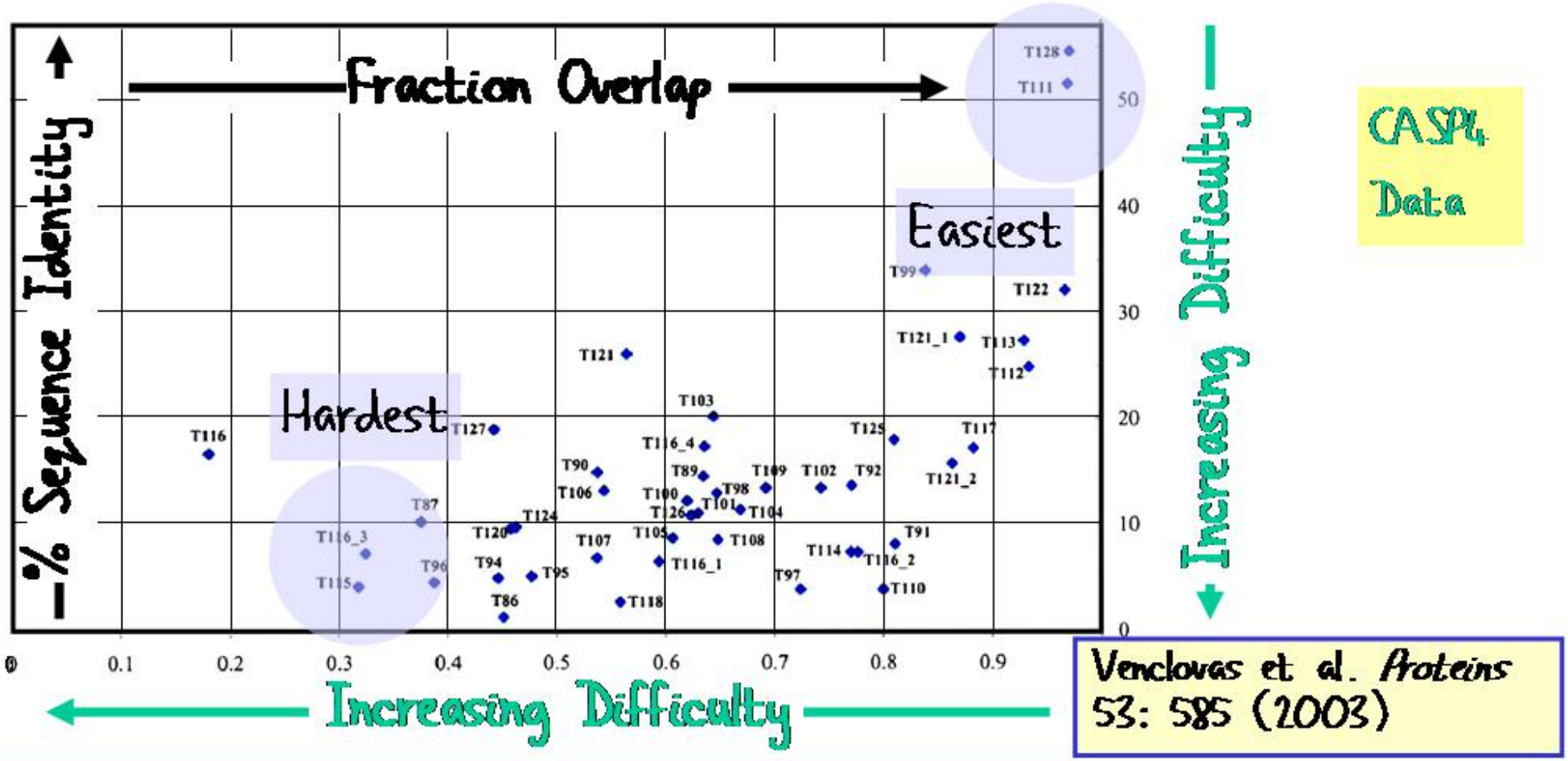
Venclovas et al. Proteins, 37: 231 (1999).

CASP4

Correct if 40 or more residues within 4 Å



MEASURING TARGET DIFFICULTY

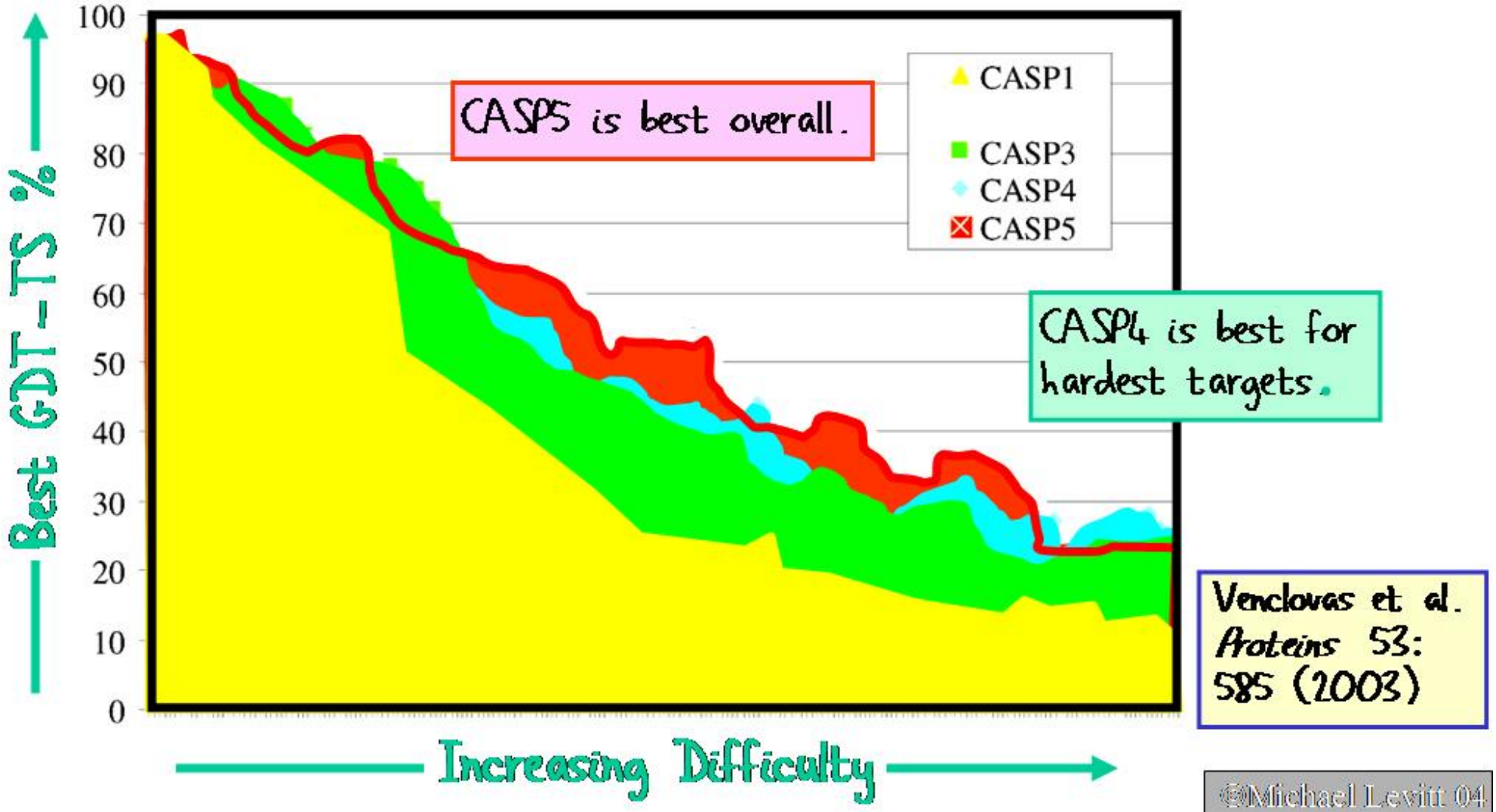


Difficulty (in one dimension) =

[rank(Sequence_Indentity) + rank(Fraction_Overlap)]/2

CASP PROGRESS

- General progress from CASP1 to CASP5.



Early Ab Initio Prediction
Concept 9.4

AB INITIO EARLY HISTORY

Coordinate and distance deviation.

Folding as a random walk.

Very simple lattice models.

Minimization with restraints.

EARLY HISTORY

- 1965: Gibson & Scheraga Minimize Small Peptides.

  Gibson & Scheraga. Minimization of Polypeptide Energy, I. Preliminary Studies of Bovine Pancreatic Ribonuclease S-Peptide. PNAS, 58: 420–426 (1967).

- 1969: Levitt and Lifson minimize entire protein structure.

  Levitt & Lifson. Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. J. Mol. Biol. 46: 269–279 (1969).

- 1973: Ptitsyn Packs Wooden Rod Helices for Myoglobin.

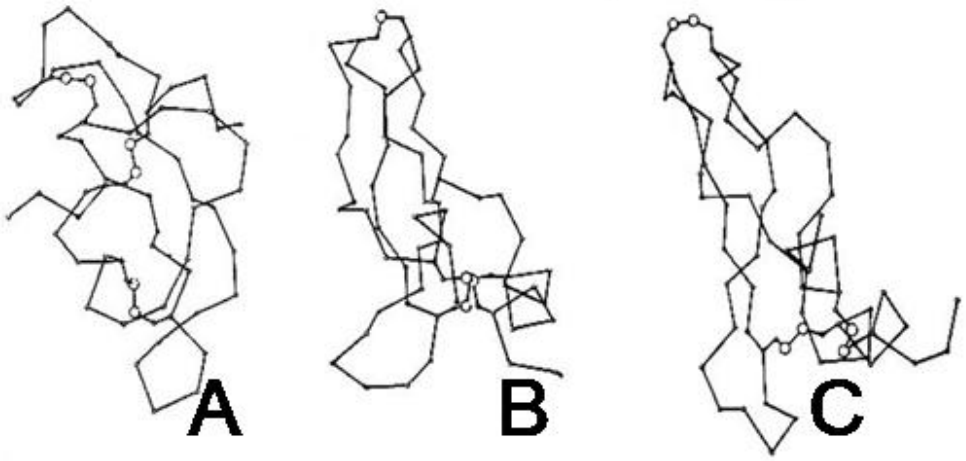
  Ptitsyn. Stagewise Mechanism for Protein Self-Organization. Vestnik Akad. Nauk. S.S.S.R. 5: 57–68 (1973).

- 1975: Levitt & Warshel use simplified model to fold chain.

  Levitt & Warshel. Computer Simulation of Protein Folding. Nature, 253: 694–698 (1975).

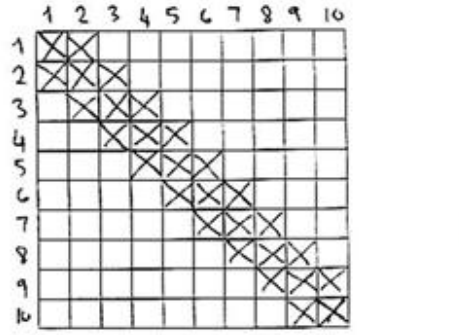
COORDINATE AND DISTANCE DEVIATION

RMS Deviation of corresponding atoms after best superposition.

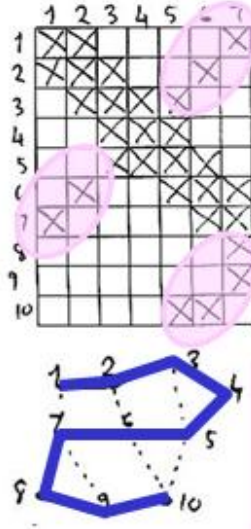


- Coordinate Deviation, known as cRMS, measures the deviation of superimposed atomic positions as a root mean square.

Contact Map marks pairs of atoms close in space. No superposition.



Extended. No off-diagonal contacts



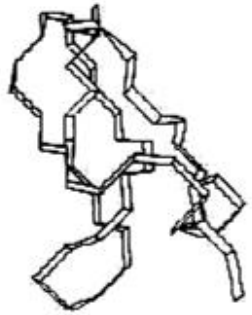
Folded. Many off-diagonal contacts

- Distance Deviation, known as dRMS, measures the deviation of corresponding interatomic or inter-CA distances as a root mean square.

FOLDING AS A RANDOM WALK

Random fold generation gives native-like folds.

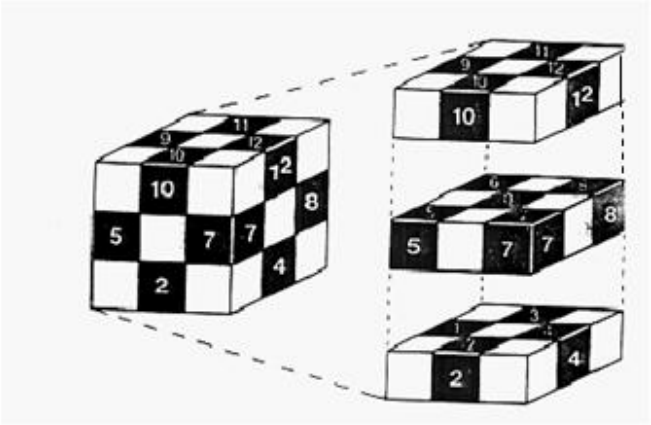
Fit to Native
dRMS=2.8Å



Random walk on lattice.

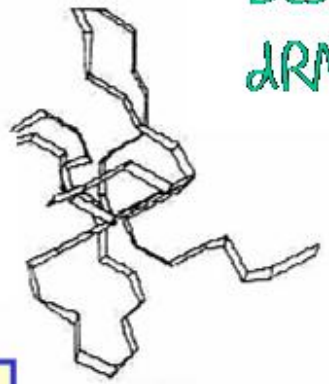
Self-avoiding and bounded.

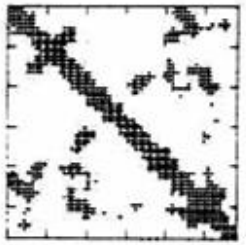
Can extend the chain in 4 ways all with a bond angle of 120°.



Use three-dimensional face-centered cubic lattice.

Best Random
dRMS=5.4Å

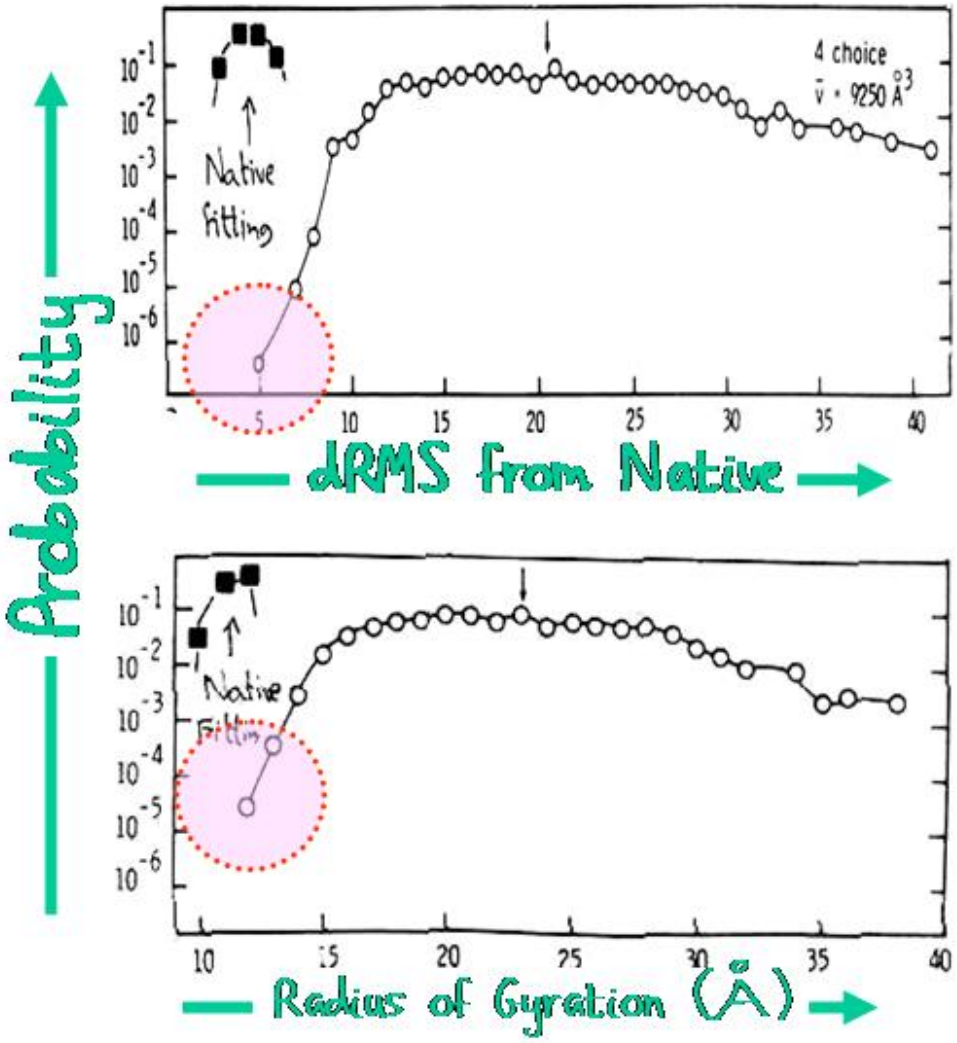


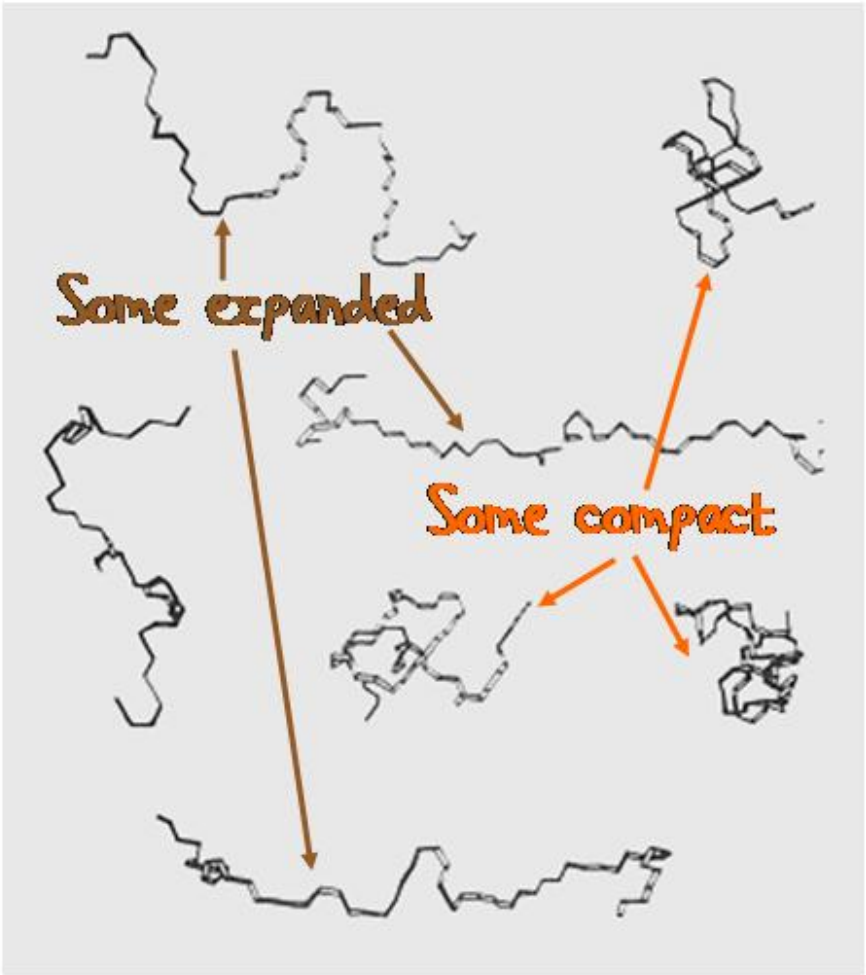


Levitt. Protein Folding as a Random Walk. Taniguichi Symposium. (1971).

WIDE DISTRIBUTION OF FOLDS

The best fold of the random set (1 in 1,000,000) has as good a dRMS as folds that are fit native structure on a lattice.

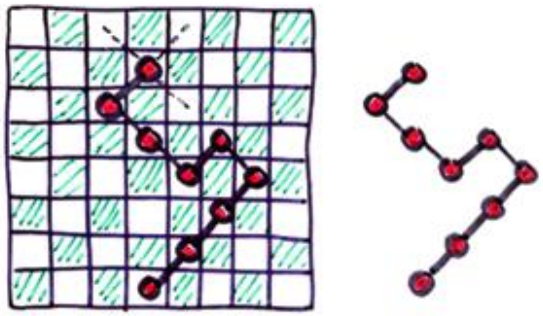




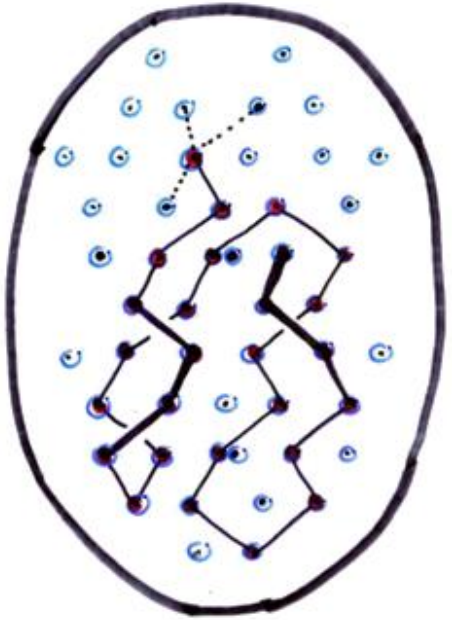
VERY SIMPLE LATTICE MODEL

- Self-avoiding walk on tetrahedral lattice

  This gives:
  - A chain
  - Self-avoidance
  - Compactness



- Put every second residue on the lattice



- Can represent real proteins to 5 Å

Hinds, D.A. and M. Levitt. A Lattice Model for Protein Structure Prediction at Low Resolution. PNAS. 89, 2536-2540 (1992).

POTENTIAL ENERGY IN TORSION SPACE

$$U = \sum K_\phi [1 - \cos(n\phi + \delta)]$$

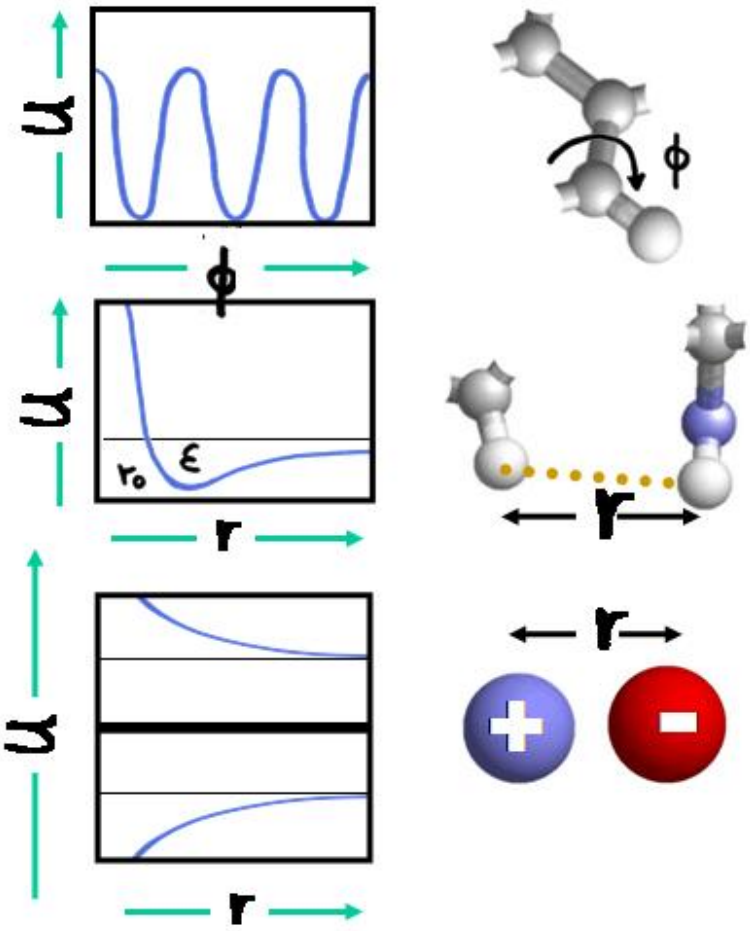
All Torsion Angles

$$+ \sum \varepsilon \left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^{6} \right]$$

All nonbonded pairs

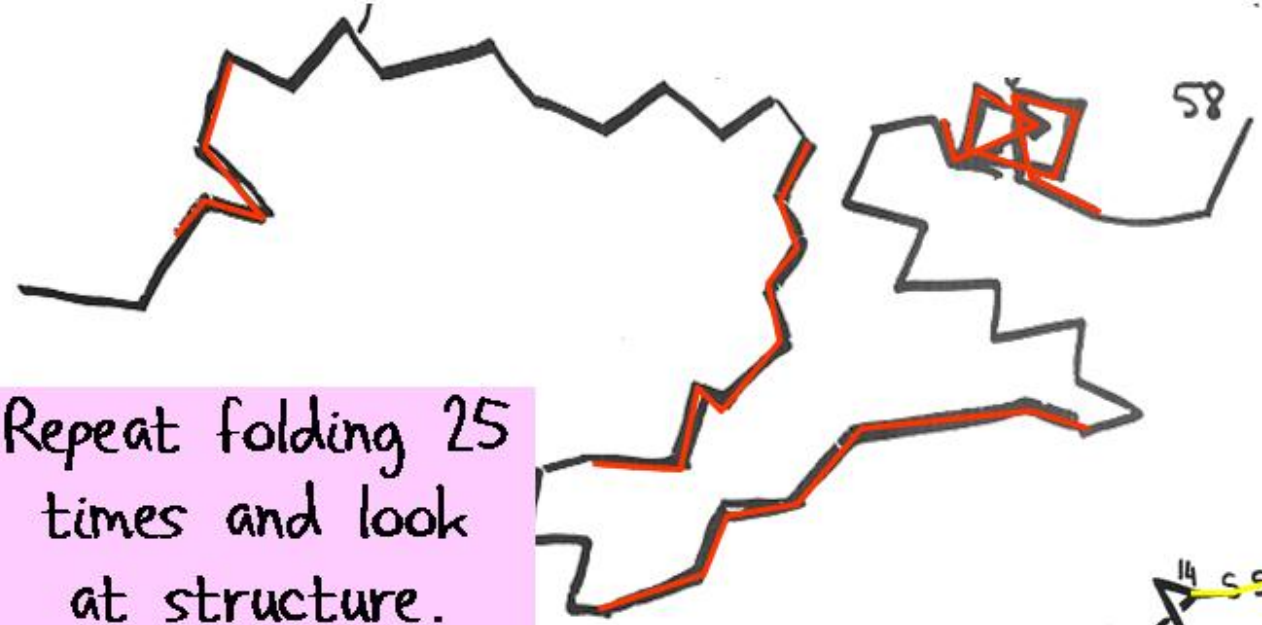
$$+ \sum 332 q_i q_j / r$$

All partial charges



- A protein with N residues has about 4N (φ, ψ, χ) single bond torsion angles.
- The same protein has about 5ON Cartesian coordinates (x, y, z).

FOLDING WITH RESTRAINTS

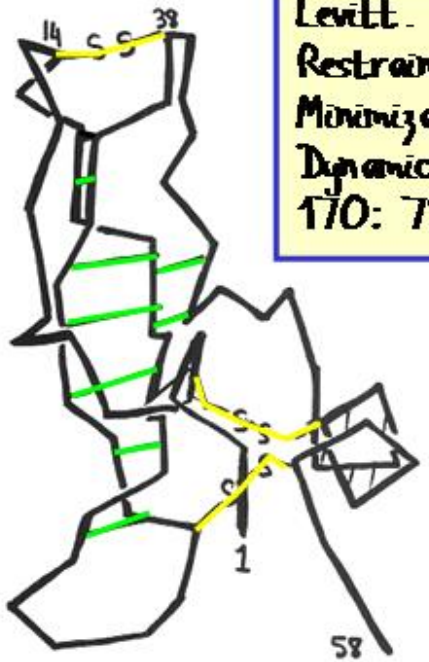


Start folding at open chain with preformed segments of secondary structure

Repeat folding 25 times and look at structure.

Test on BPTI (58 residues, 515 atoms, 208 torsion degrees of freedom).

Use 16 hydrogen bonds and 3 SS bridges as restraints.

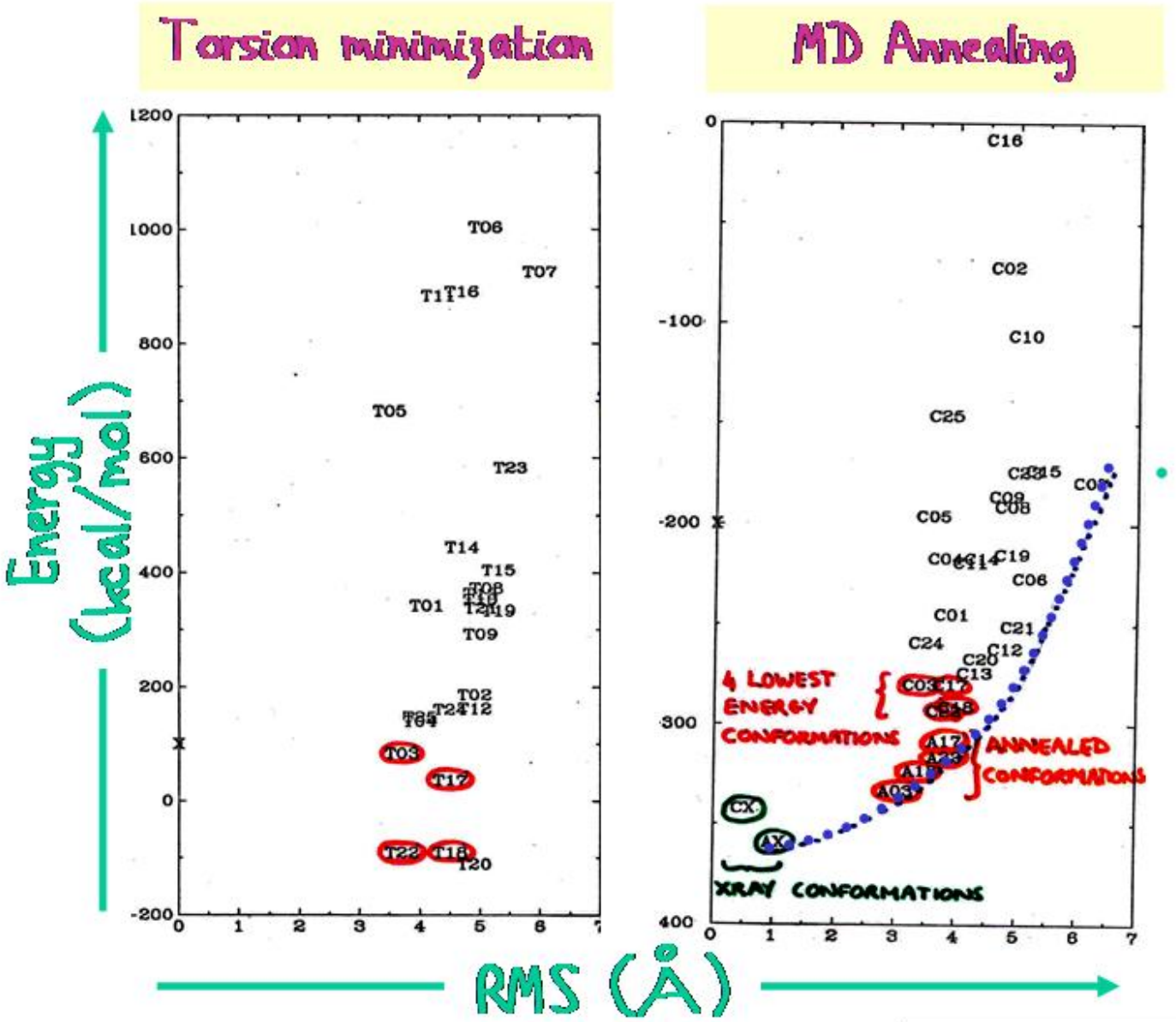


Levitt. Protein Folding by Restrained Energy Minimization and Molecular Dynamics. *J. Mol. Biol.* 170: 723 (1983).

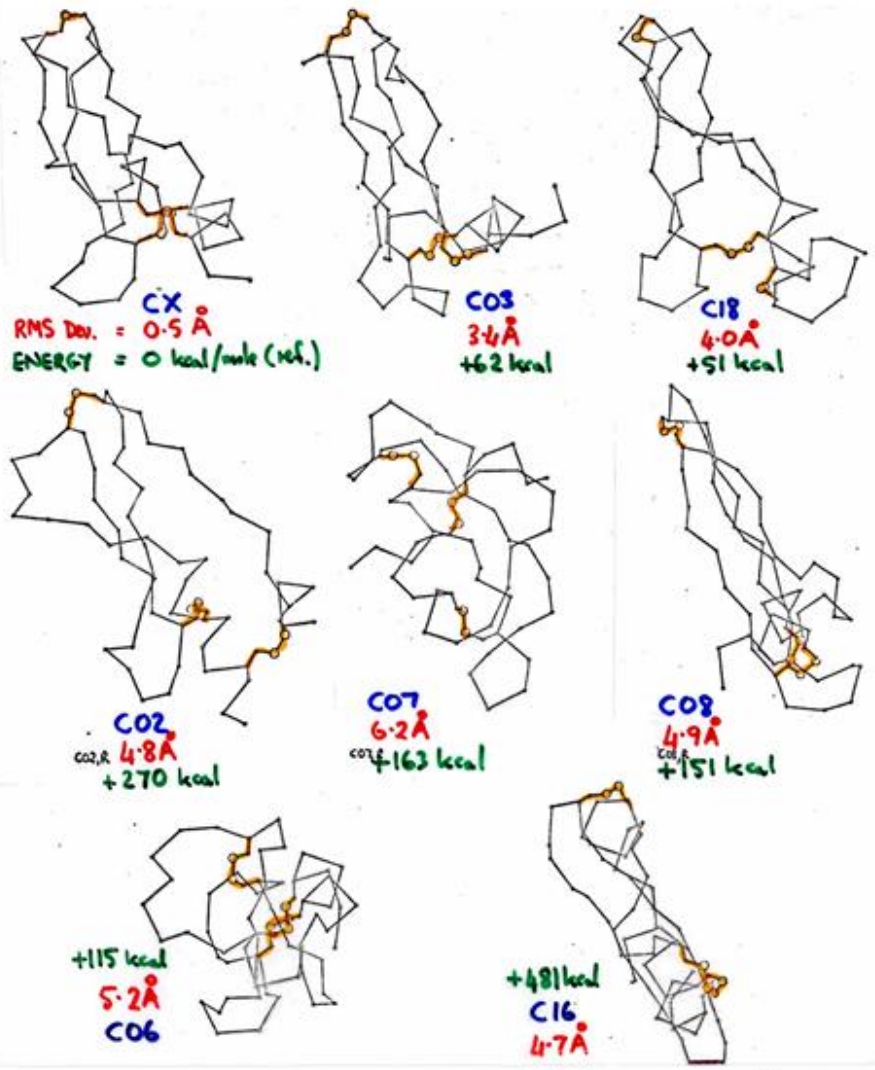
FOLDING WITH RESTRAINTS

Plot energy vs. RMS deviation from x-ray structure.

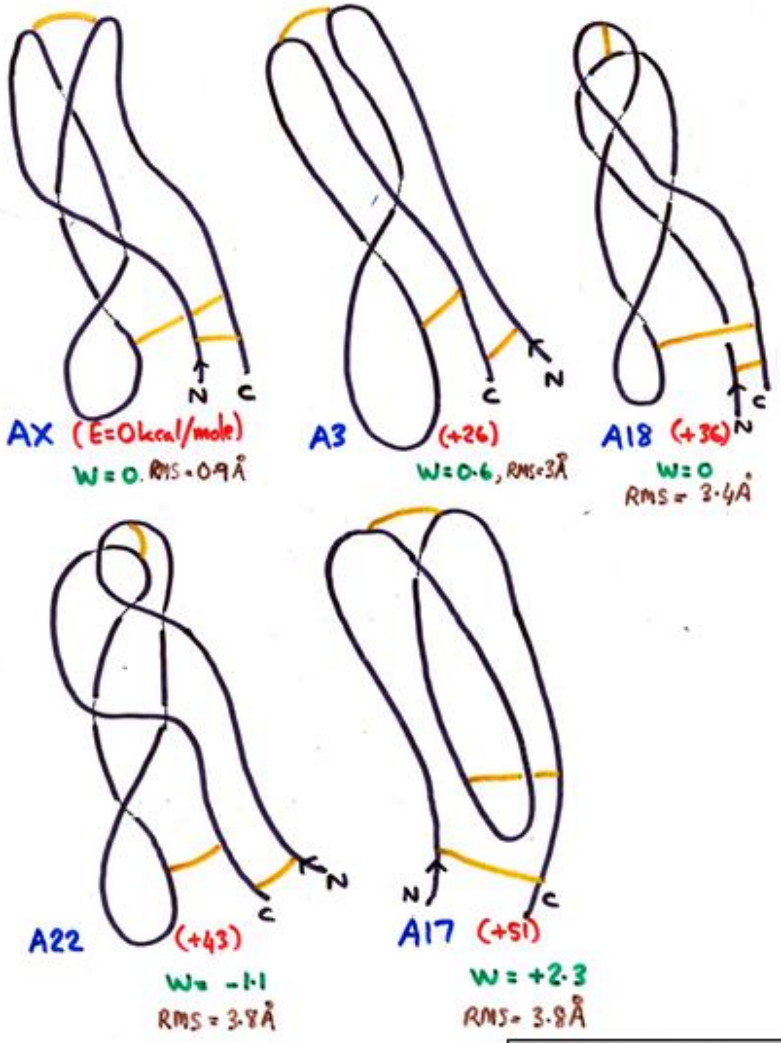
Want low energy to be at low RMS.



DIVERSITY OF FOLDS



Some are threaded properly.
Most are not.



Some are as close as 3.4 Å cRMS.
Others are as far as 6.2 Å cRMS.

Modern New Fold Prediction Concept 9.5

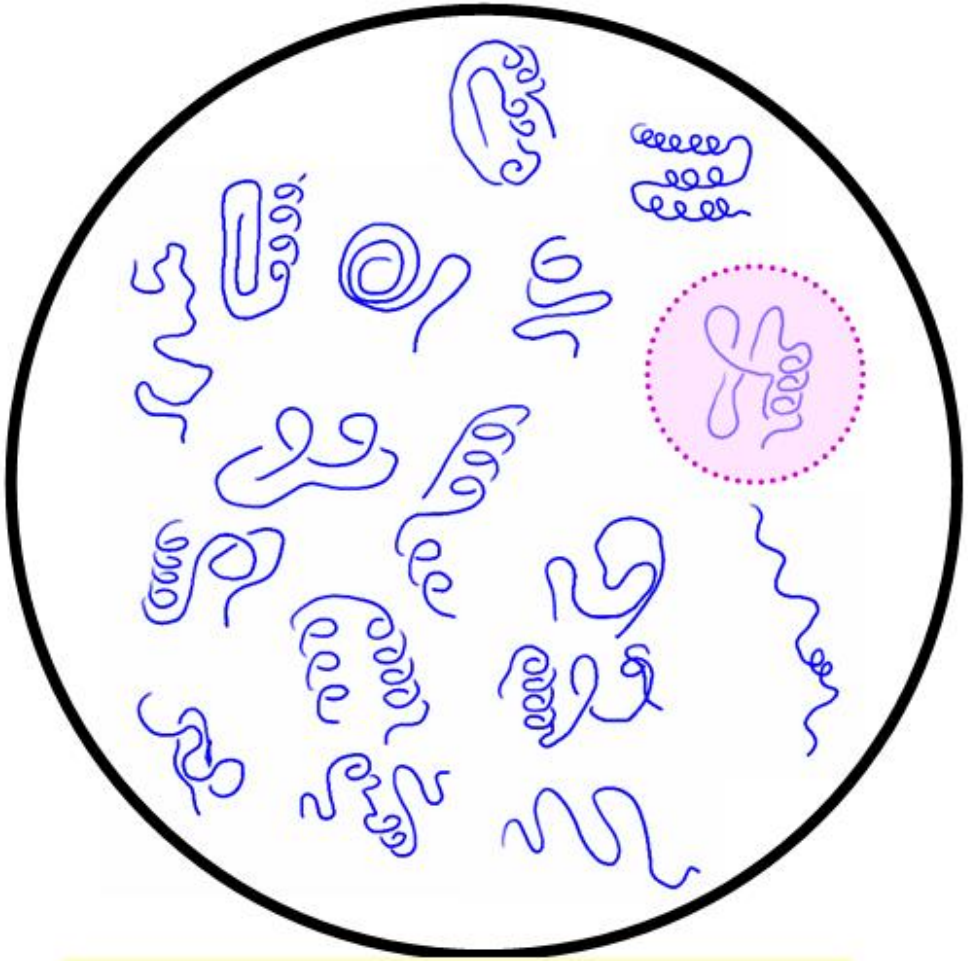
MODERN NEW FOLD PREDICTION

Discrimination Paradigm.

Energy vs. RMS Plots.

Hierarchical Methods.

A PARADIGM FOR PREDICTING STRUCTURE



DECOYS

- Construct a large number of possible folded shapes.

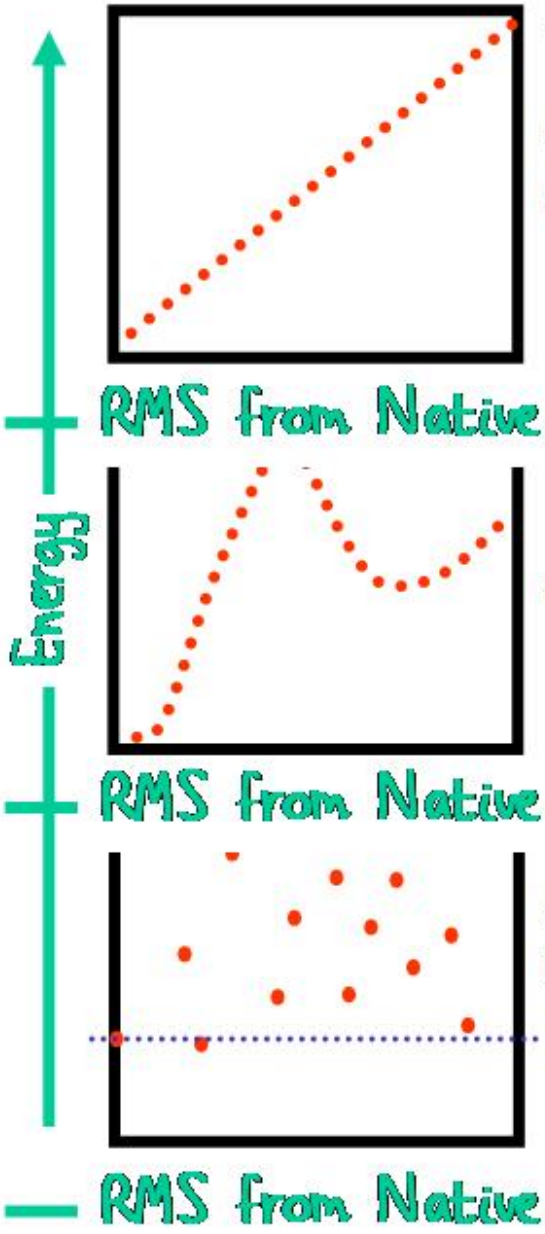
DISCRIMINATION

- Select the correct, native fold.

Need a good energy function

Need good decoy structures.

ENERGY vs. RMS CARTOONS



Energy increases linearly with distance from native. (Idealist's view)

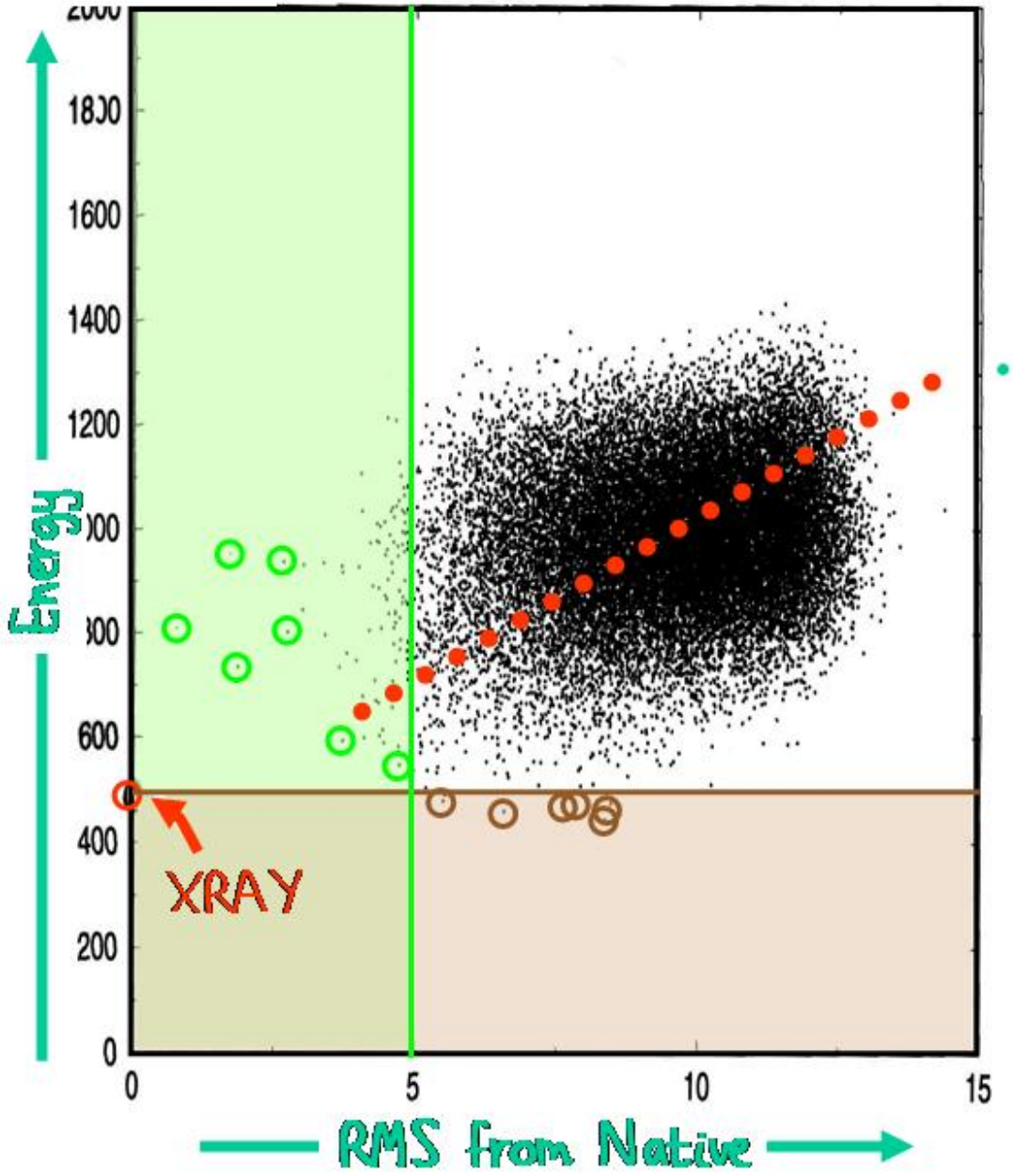
Energy varies smoothly as deviation increases. (Experimentalist's view)

Energy varies wildly with RMS deviation. (Realist's view)

Native is still low, but not lowest.

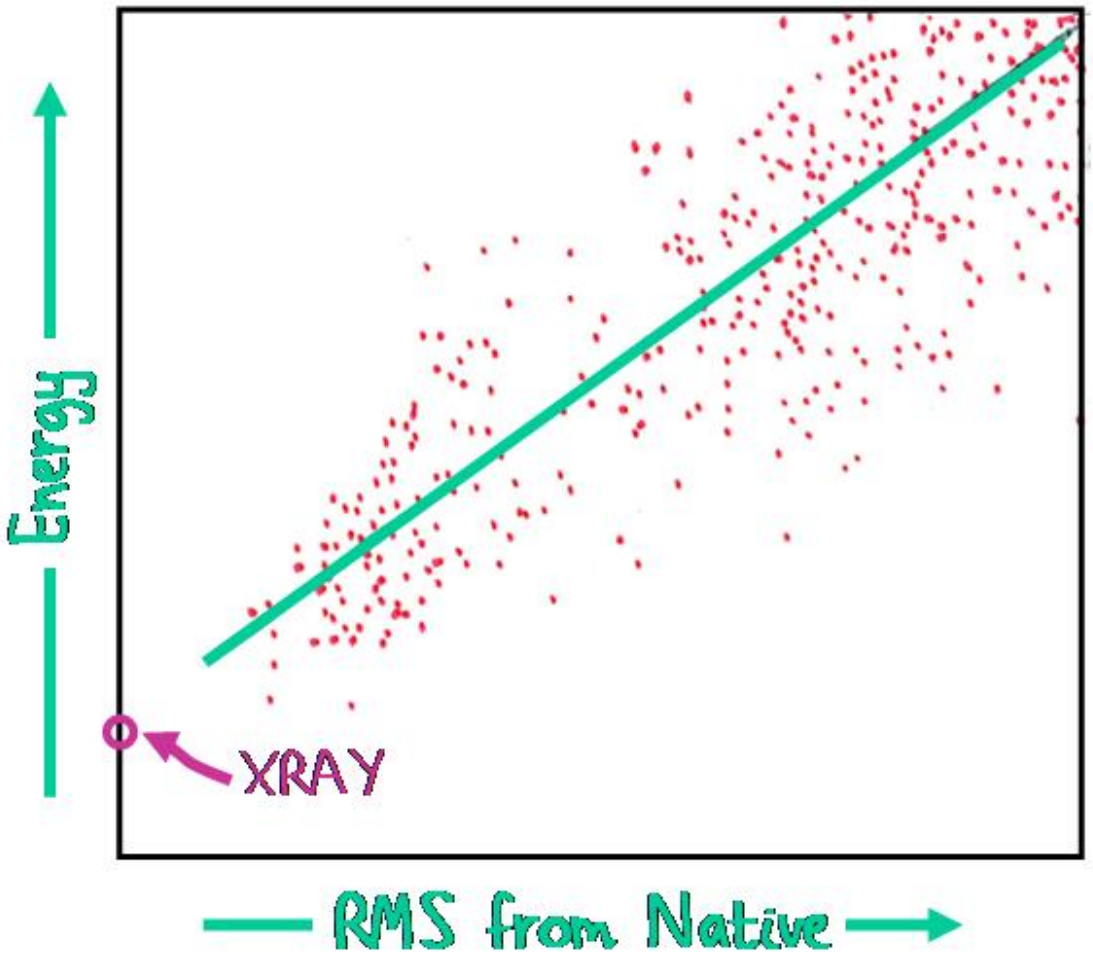
- Can easily calculate the energy of every decoy structure.
- Want to find the decoys that are closest to the real structure (low RMS).
- Are they also low energy?

ENERGY VS RMS REALITY



- There are low energy structures with low RMS (○).
- The are high RMS structures with lower energy than low RMS structures (○).
- There is some weak correlation between energy and RMS.

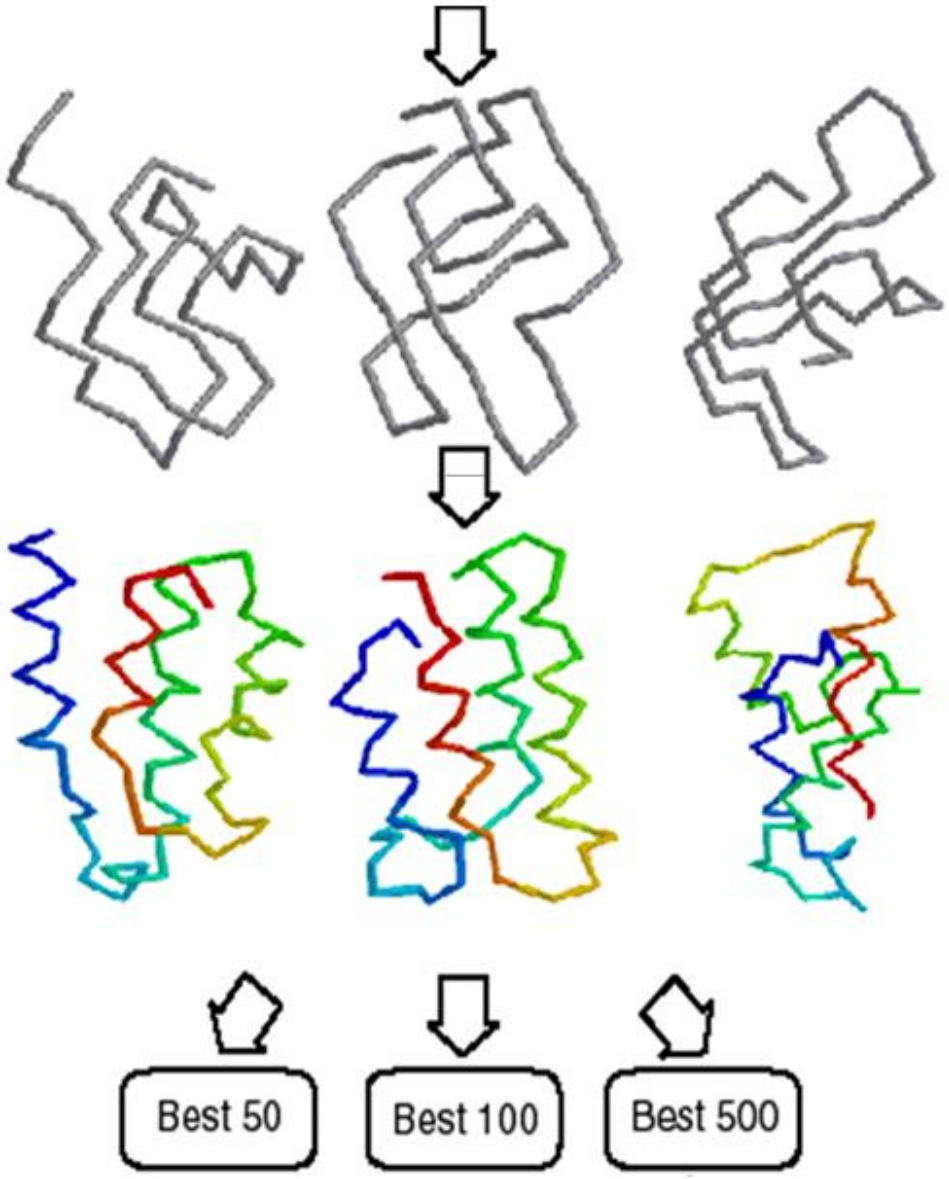
ENERGY VS RMS



- This is a particularly favorable case.
- The all-atom knowledge based energy of Samudrala and Moult (rapdf) sometimes has a high correlation with RMS.

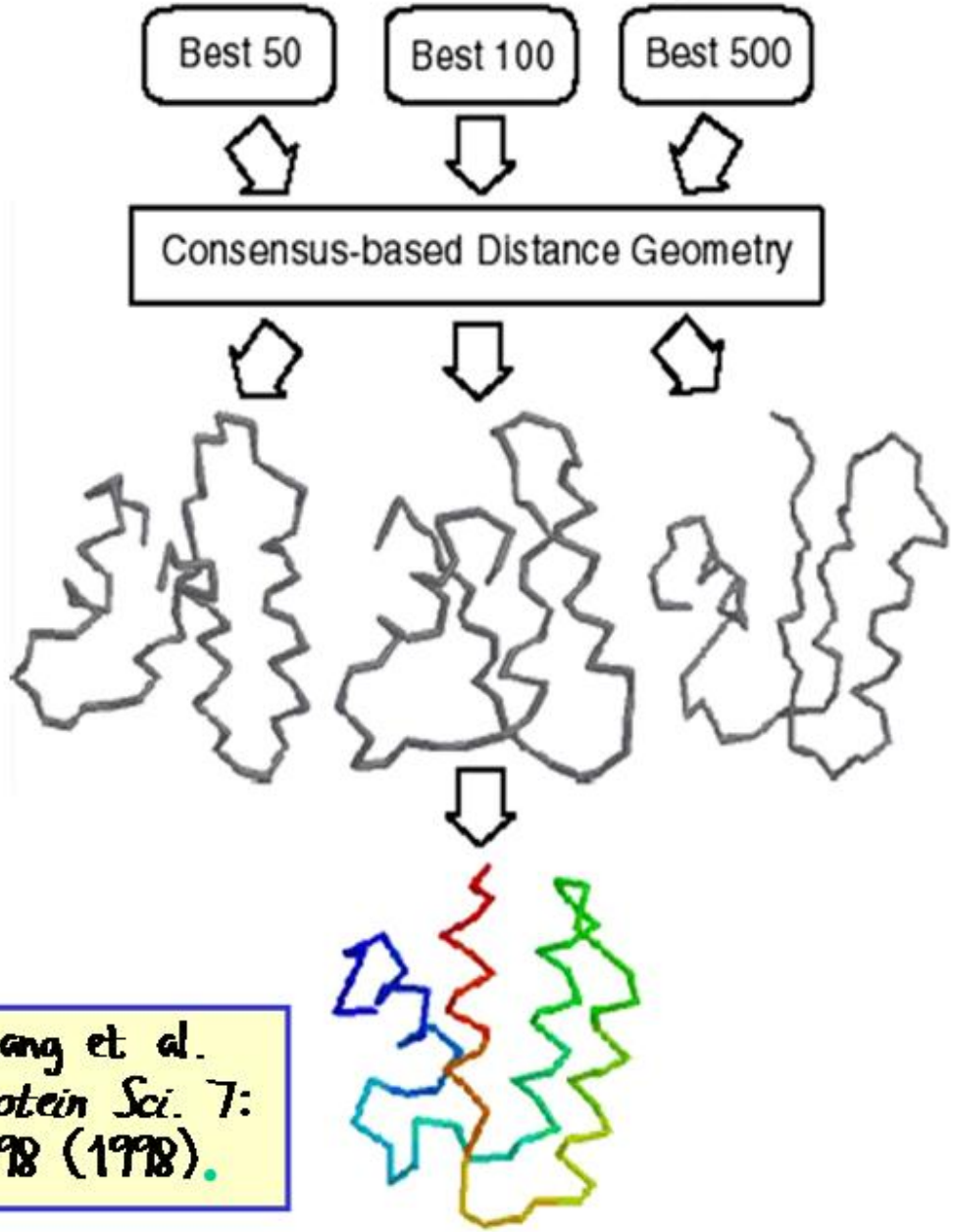
HIERARCHICAL STRUCTURE PREDICTION

ILYLDRVLGIMAPTRWFAEALSKYNSTILML



- Amino acid sequence.
- Simple lattice models give all possible low-resolution shapes.
- Set predicted secondary structure and add all-atom detail.
- Score with knowledge-based energy function.

GETTING AN AVERAGE MODEL



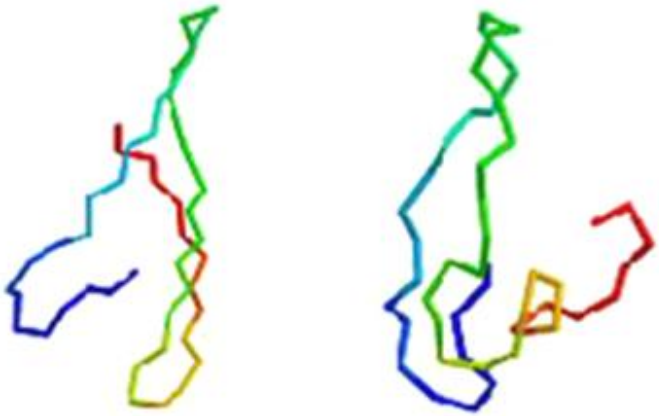
Huang et al. *Protein Sci.* 7: 1998 (1998).

- Get pair-wise CA...CA distances in each structure.
- Average CA...CA distances for each pair.
- Use distance geometry to get a consensus model that best fits the average distances.

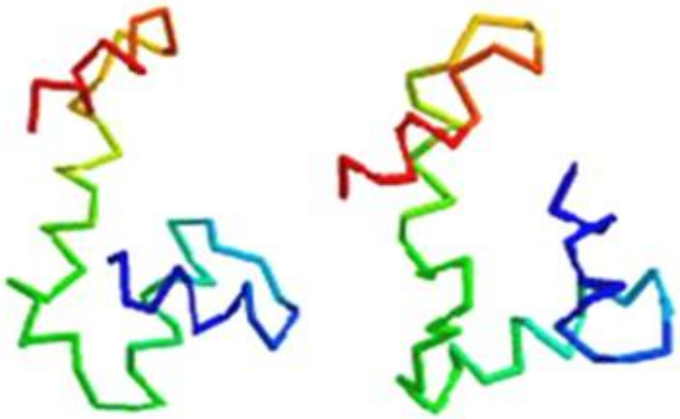
HIERARCHICAL PREDICTION DOES WELL

Does well at CASP3 (1998).

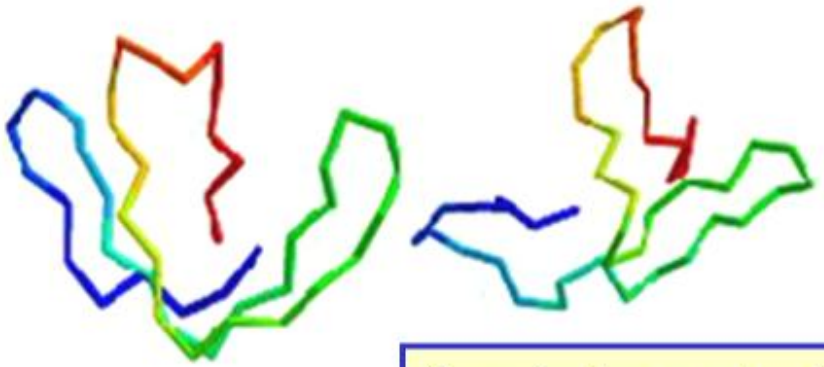
T46/adg 7.5 Å (49 residues; 66113)



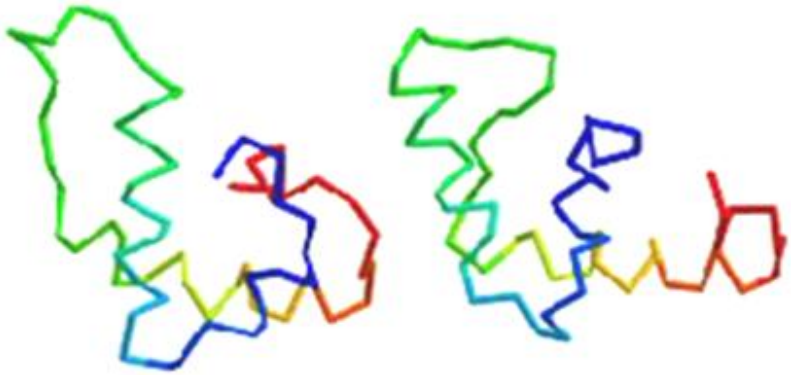
* T56/dnab 6.8 Å (60 residues; 67126)



** T59/smd3 6.7 A (46 residues; 3075)



** T61/hdea 7.4 Å (66 residues; 974)



Samudrala et al. *Proteins*, 37 (3S): 194 (1999).

FOLDING ENERGY FUNCTIONS

- Minimize all-atom energy with respect to all torsion angles.

- Augment the normal potential energy function with:
  - Cooperative long-range hydrophobic interactions.
  - Cooperative long-range hydrogen bonds.
  - Forced exposure of charges.

Keasar and Levitt, A Novel Approach to Decoy Set Generation: Designing a Physical Energy Function Having Local Minima with Native Structure Characteristics. *J. Mol. Biol.* 329, 159 (2003).

POTENTIAL ENERGY IN TORSION SPACE

$$U = \sum K_\phi [1 - \cos(n\phi + \delta)]$$

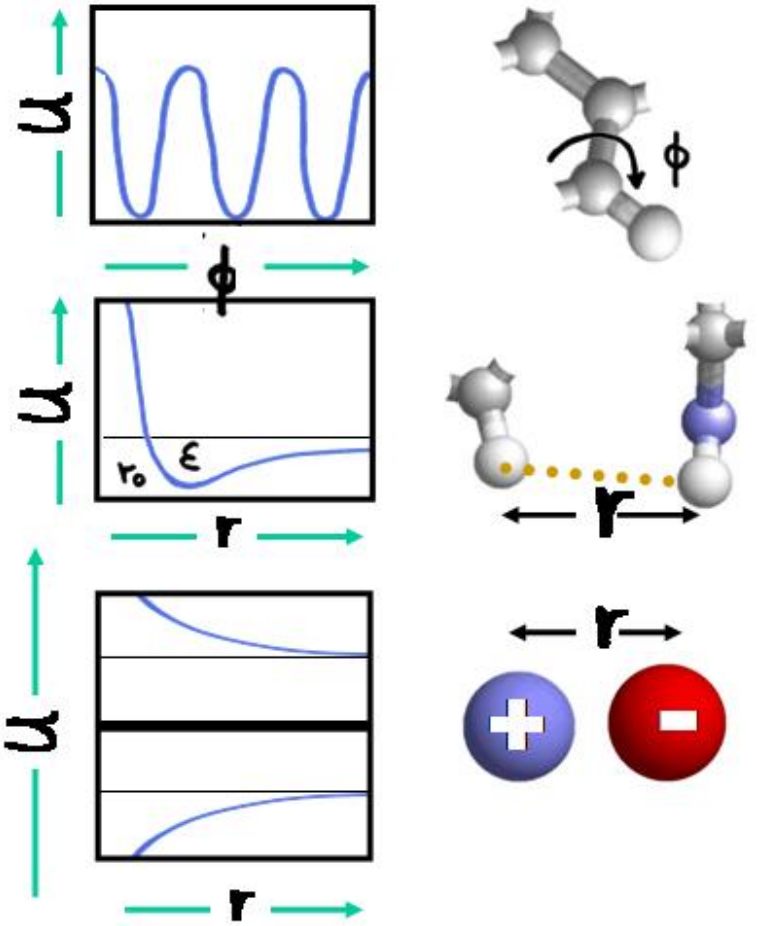
All Torsion Angles

$$+ \sum \varepsilon \left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^{6} \right]$$

All nonbonded pairs

$$+ \sum 332 q_i q_j / r$$

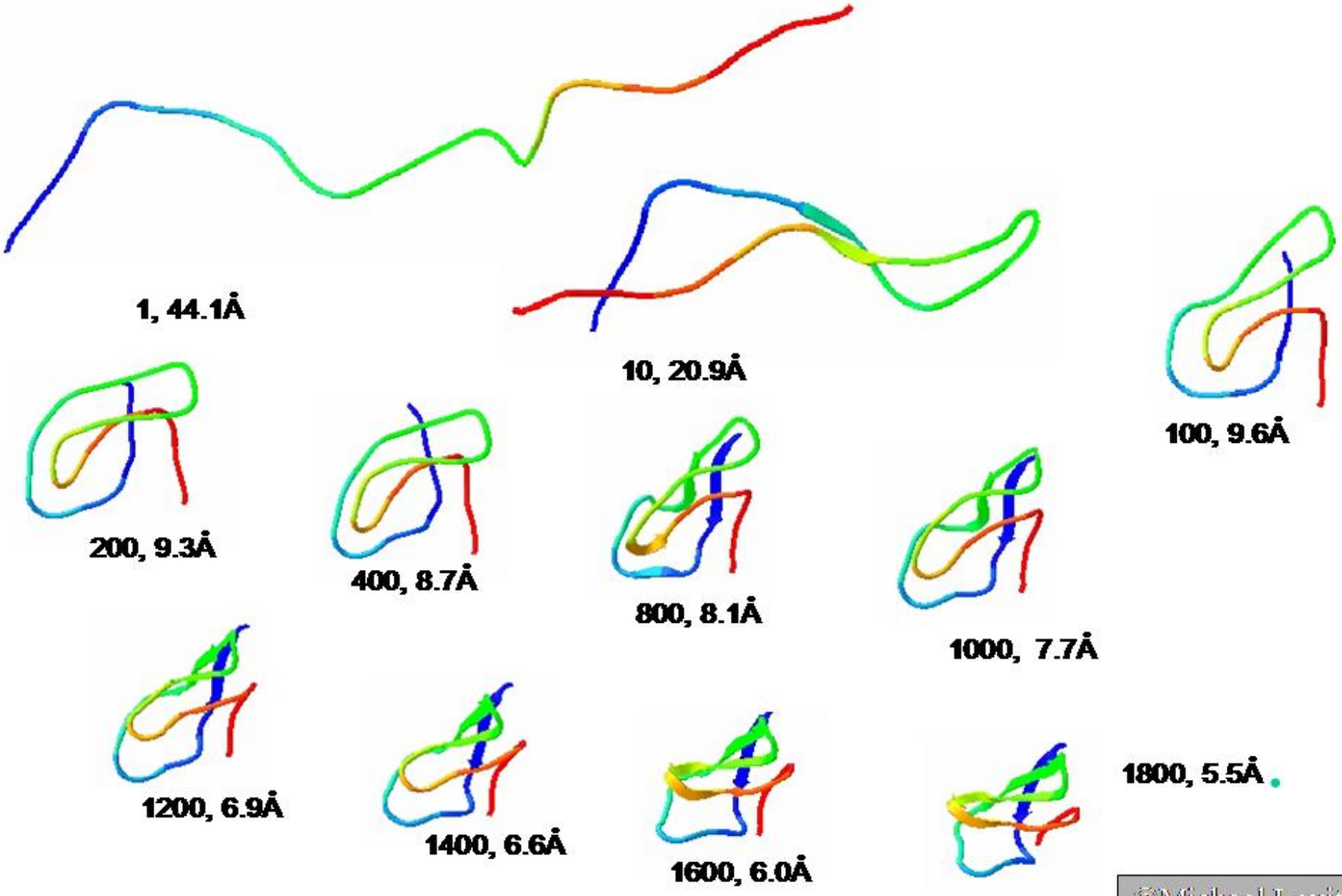
All partial charges



+ Clever cooperative terms for
  - Hydrophobic contacts
  - Hydrogen bonds
  - Buried charges

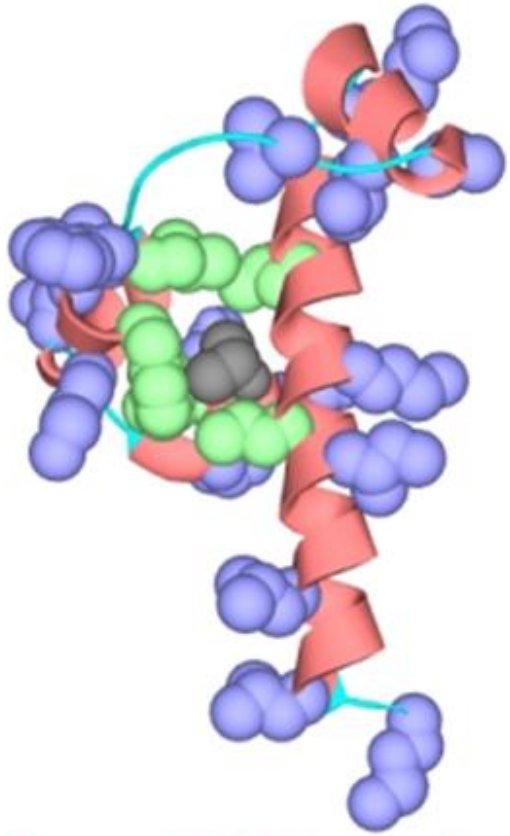
- A protein with N residues has about 4N (φ, ψ, χ) single bond torsion angles. The same protein has about 50N Cartesian coordinates (x, y, z).

ENERGY MINIMIZATION FOLDS THE CHAIN

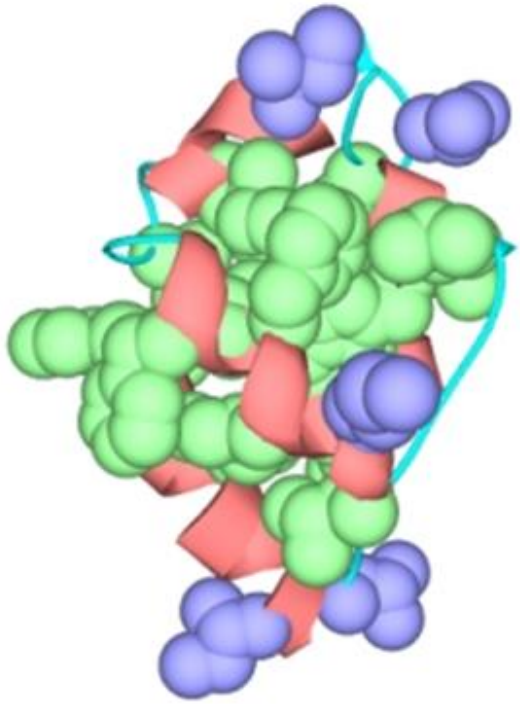


COOPERATIVE HYDROPHOBIC PACKING

- Cooperative hydrophobic compaction makes a good core.

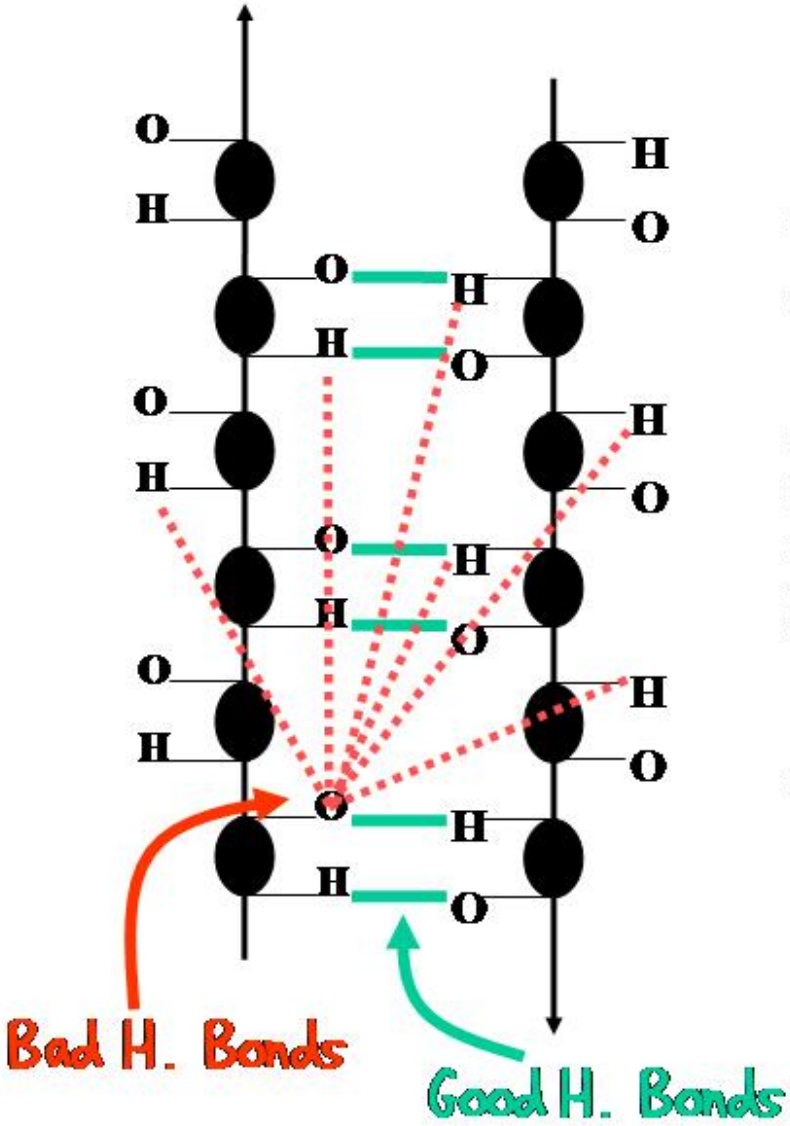


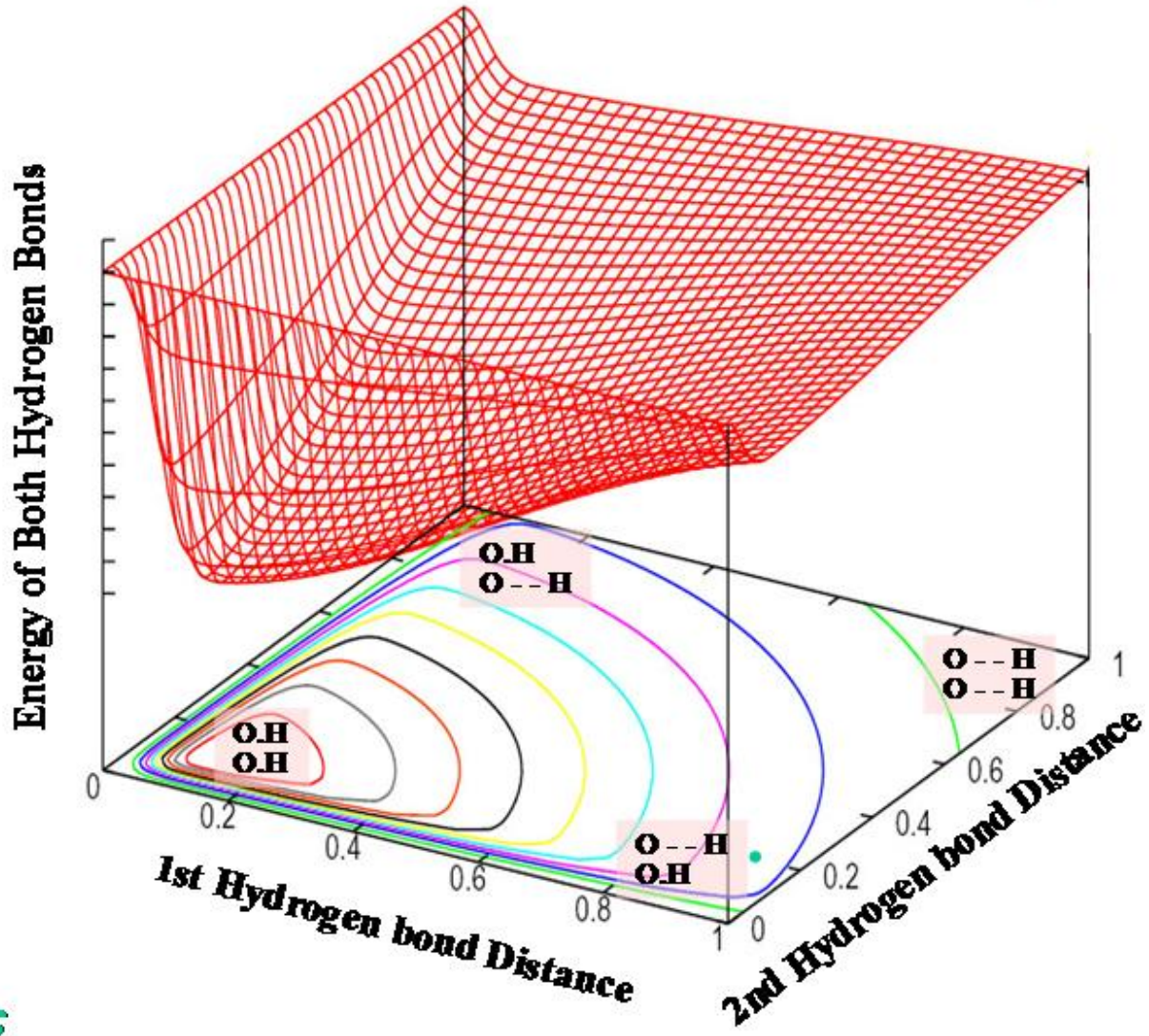
Original Potential



Modified Potential.

COOPERATIVE HYDROGEN BONDS

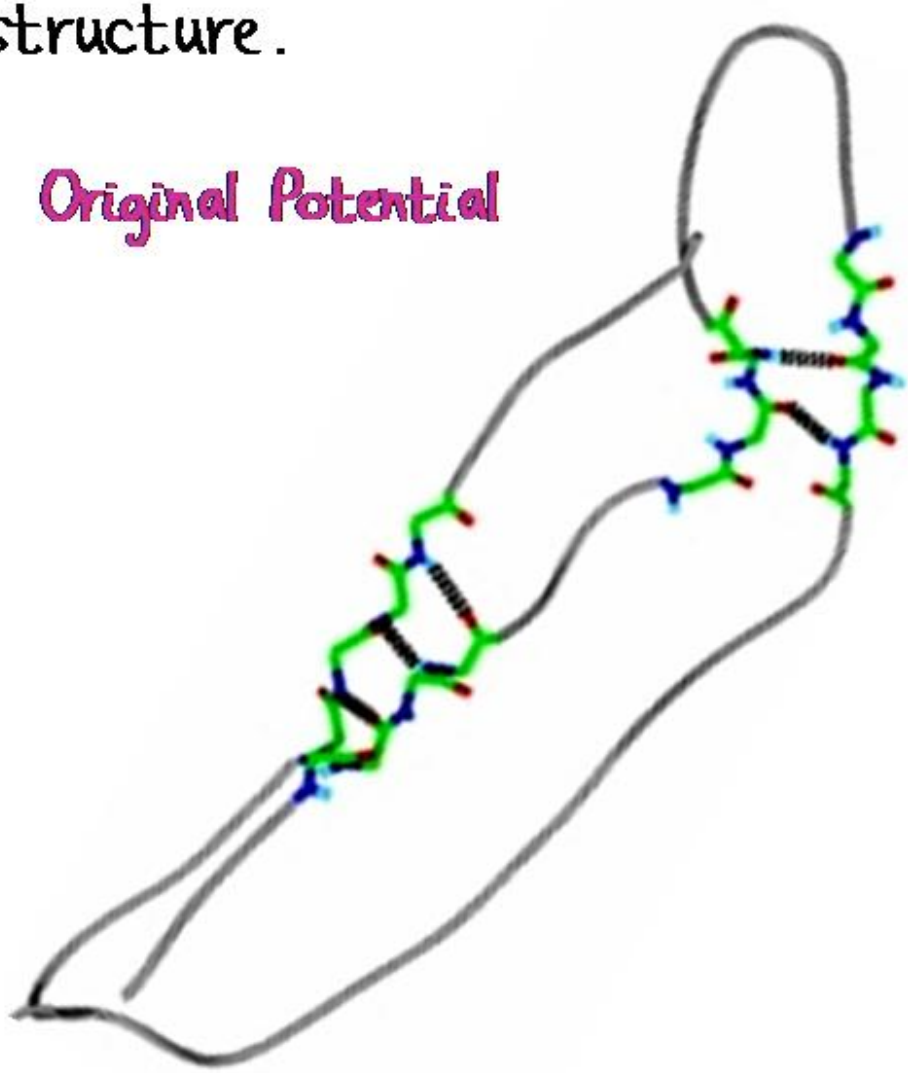




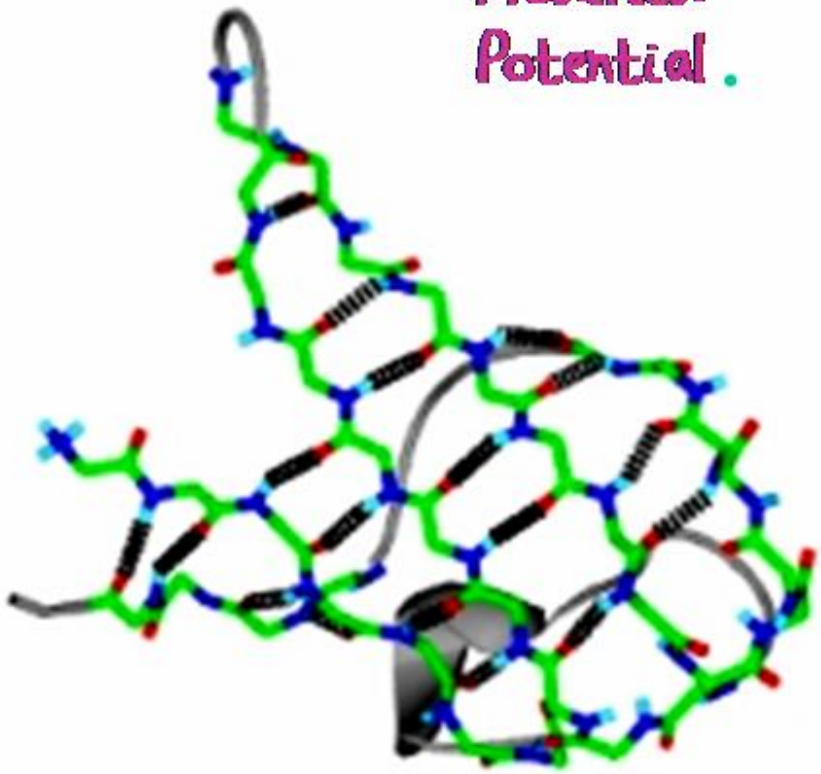
COOPERATIVE HYDROGEN BONDS

- Cooperative hydrogen bonds give good secondary structure.

Original Potential



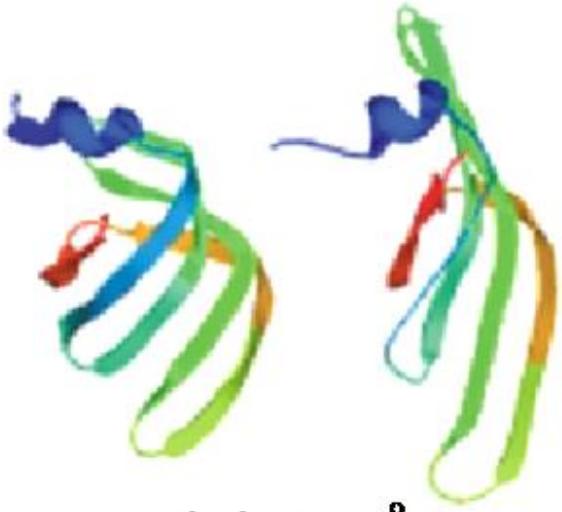
Modified Potential.



MINIMIZATION GIVES GOOD FOLDS

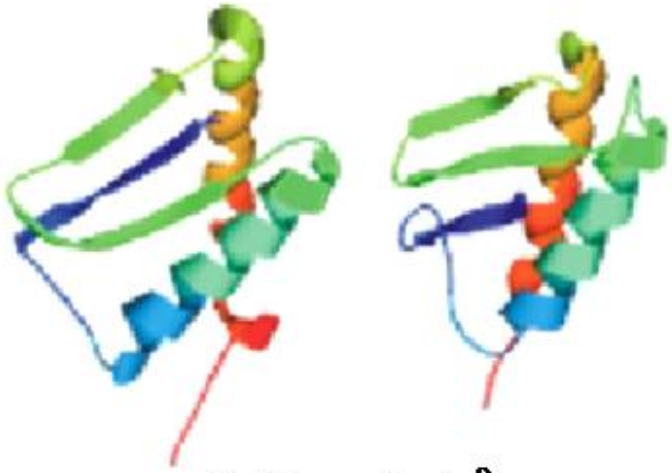
Native structure on left

1e68 3.7Å

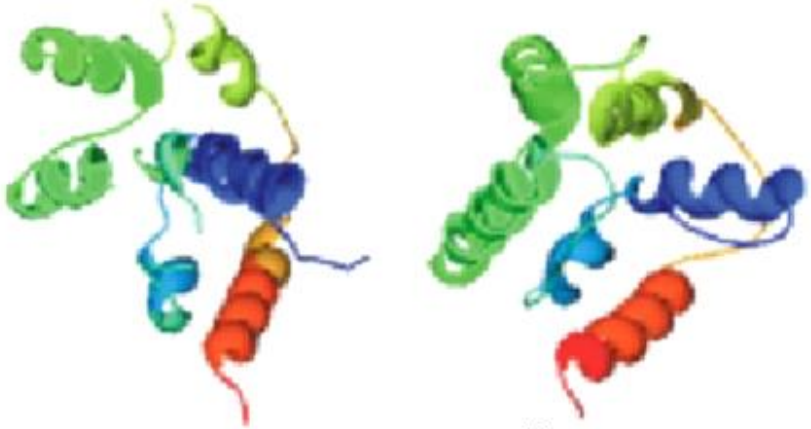


1d3b 5.5Å

1ubi 4.1Å



1f0a 4.4Å



1jwe 5.6Å

Winning NF Methods at CASP
Concept 9.6

CASP4 NEW FOLD WINNERS

RANK	GROUP	N	ZSCORE
1	Baker	15	39.40
2	Friesner	6	13.51
3	Murzin	5	12.81
4	Skolnick-Kolinski	6	12.03
5	I-Sites-Server	7	10.15
6	GNM-Fr	4	10.06
7	Jones-Ab	5	10.02
8	Lomize-Andrel	4	9.88
9	Kollman-Baker	3	8.78
10	Levitt	4	8.66
11	SAM-T2k	5	8.42
12	Ram-Samudrala	4	7.80

- Baker, Skolnick, I-Sites & Jones also do well at CASP5

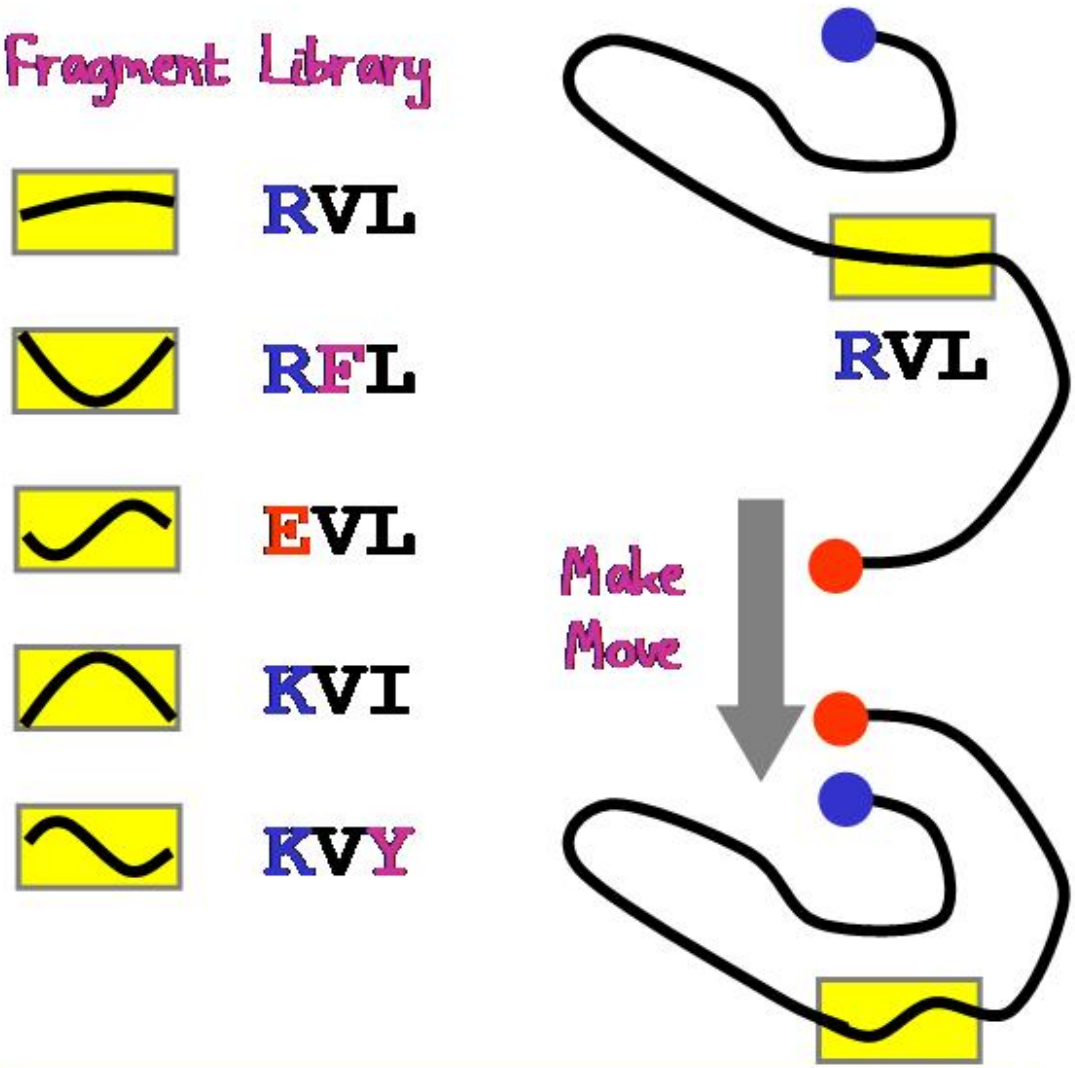
WINNING NF METHODS AT CASP4

- Segment Monte Carlo. Baker.
- I-Sites method. Bystroff.
- Lattice Monte Carlo. Skolnick & Kolinski.
- Threading & Monte Carlo. Friesner.
- Segment assembly. Jones.
- Segment Folding. Samudrala & Levitt.
- All-Atom Energy Minimization. Keasar & Levitt.

SEGMENT EXCHANGE MONTE CARLO

- Rich Bonneau, Jerry Tsai, Charlie Strauss and David Baker using their program, Rosetta.
- Fragment swap Monte Carlo.
- Specially designed energy function.
- Cluster resulting folds to select the best ones.
- I-Sites also uses Rosetta.

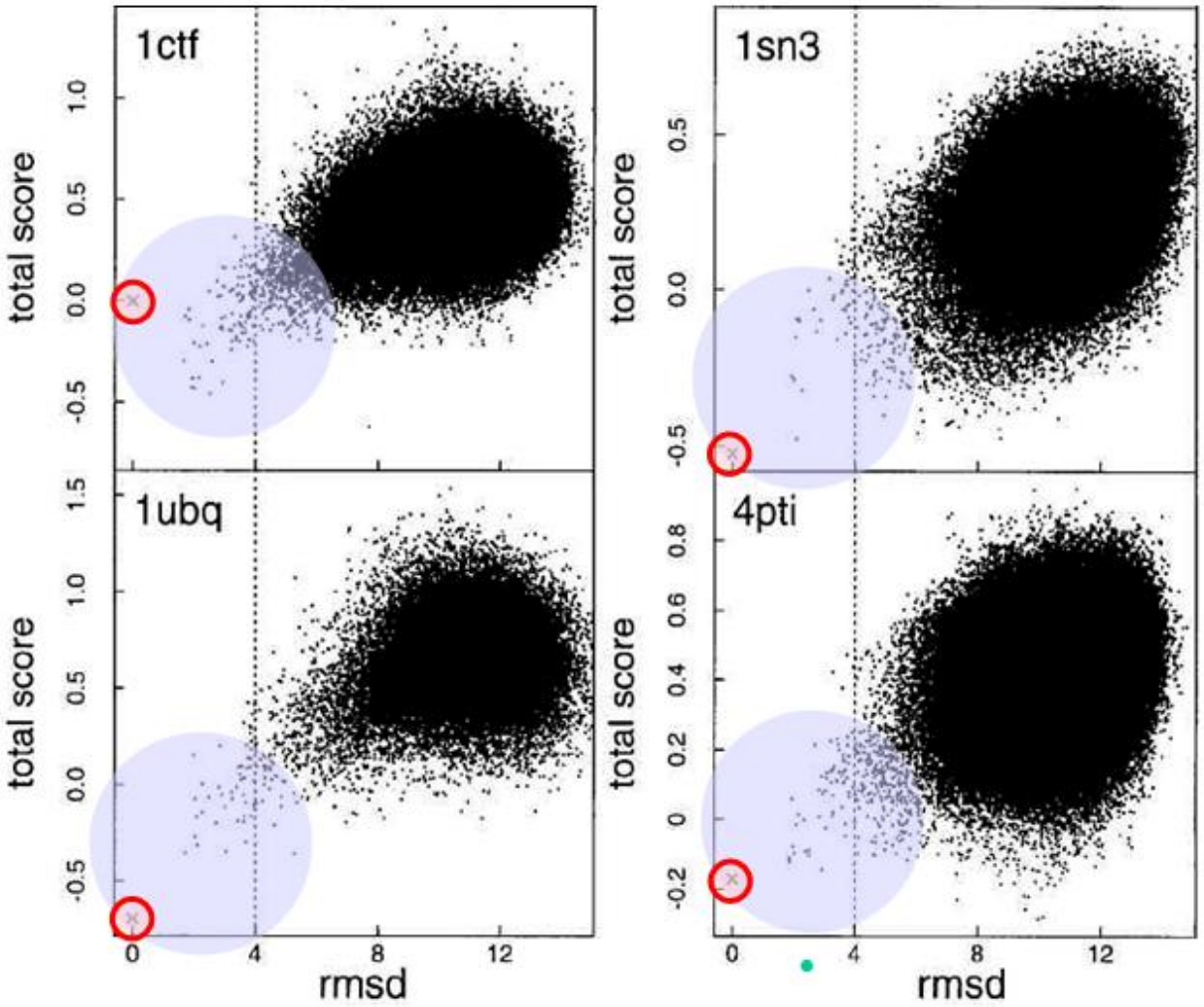
FRAGMENT INSERTION MONTE CARLO



- Use a library of small fragments of similar sequence.
- Swap in a new fragment by setting six (ϕ, ψ) torsion angles.
- Accept move by Monte Carlo and anneal.

Simons et al. *J. Mol. Biol.* 268: 209 (1997)

NEW POTENTIAL IMPROVES RECOGNITION



- Simons et al. derive a new potential that discriminates decoys.
- Add a term to enforce the hydrogen bonding of segments of secondary structure.

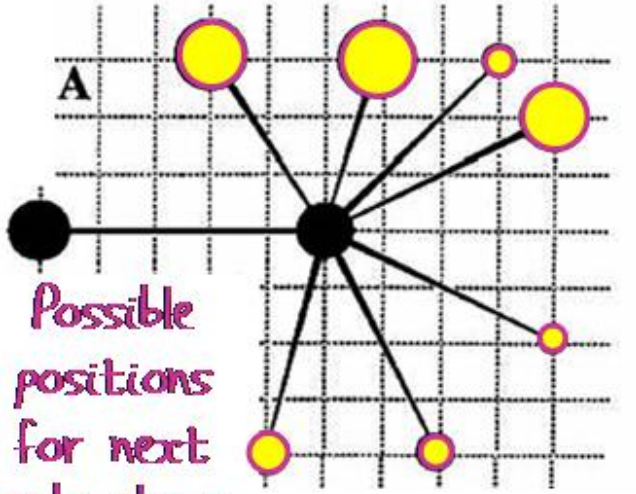
Kim, Simons, Ruczinski, Kooperberg, Fox, Bystroff & Baker. Improved Recognition of Native-Like Protein Structures Using a Combination of Sequence-Dependent and Sequence-Independent Features of Proteins. *Proteins*, 34: 82-95 (1999).

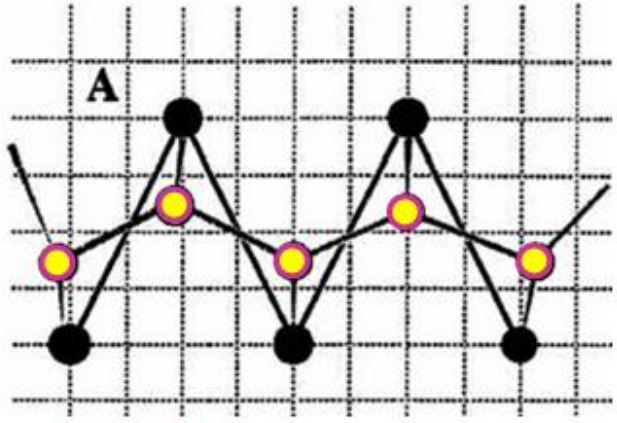
LATTICE MONTE CARLO

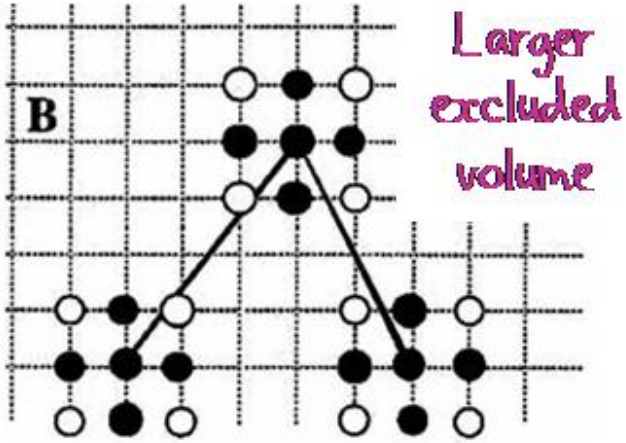
- Jeff Skolnick & Andrzej Kolinski.
- Lots of experience with lattice models and Monte Carlo.
- Use a complicated lattice model for the side chain centroids.

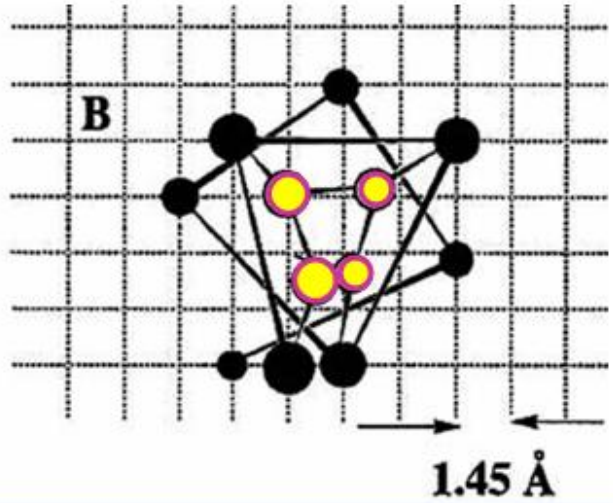
Ortiz, Kolinski & Skolnick. Nativelike Topology Assembly of Small Proteins Using Predicted Restraints in Monte Carlo Folding Simulations. *PNAS*, 95: 1020 (1998).

COMPLICATED LATTICE







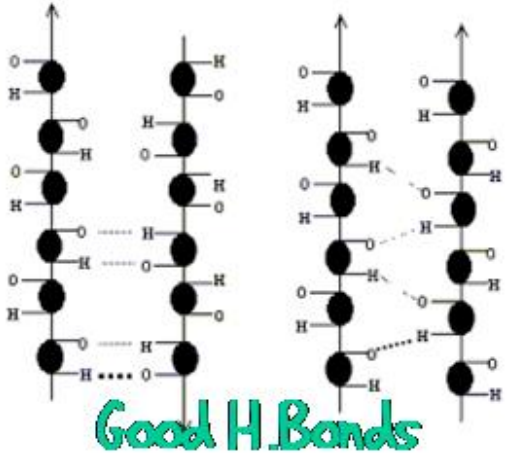


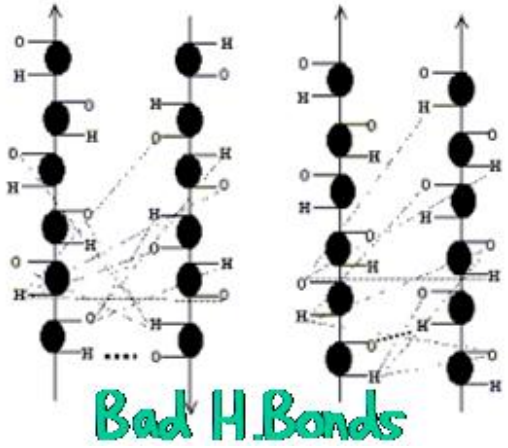
- Skolnick & Kolinsky.
- Use a complicated lattice with many choices for chain extension.
- Use Multi-Replica Monte Carlo.

Kolinski & Skolnick. Assembly of Protein Structure From Sparse Experimental Data: An Efficient Monte Carlo Model. Proteins, 32: 475 (1998).

STRUCTURE PREDICTION BY MINIMIZATION

- Minimize special energy function with respect to torsion angles (phi, psi and chi).
- Add energy terms for cooperative hydrogen bonds and hydrophobic compaction.

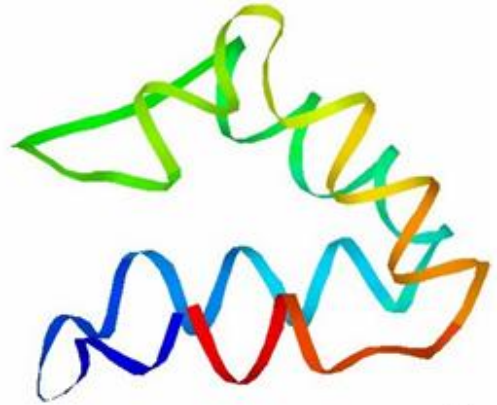




Does well at CASP4.

Structure of T102

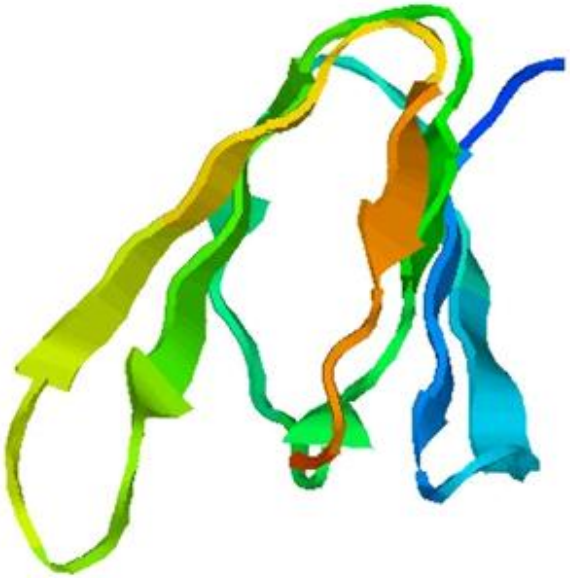
Best T102 3.3Å



T102 Submitted 5.0Å

ALL-BETA PREDICTION SUCCESS

- All-beta proteins are the hardest to predict.
- Torsion minimization does well on T114, an all beta-protein.



Native Structure

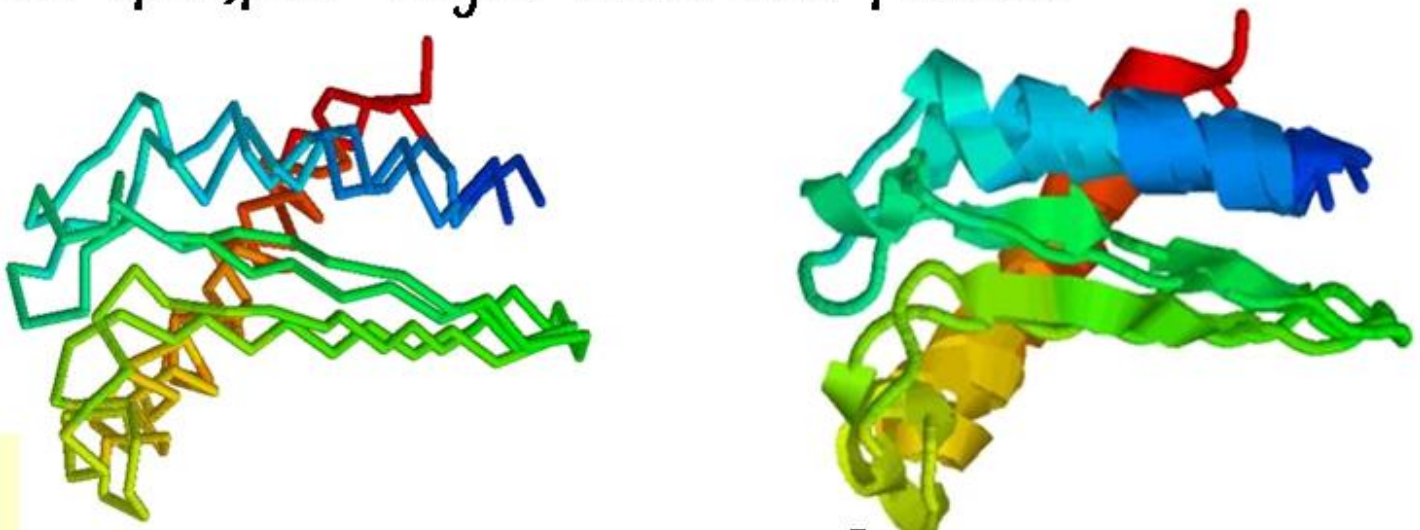
Prediction is somewhat similar

SEGMENT FOLDING MONTE CARLO

- Ram Samudrala & Michael Levitt.
- Use all-atom knowledge-based energy function (rapdf) with a compaction term.
- Make moves by swapping the torsion angles of a 3-residue segment.
- Reduce temperature during Monte Carlo run (simulated annealing).

SEGMENT FOLDING PREDICTION

- Do Monte Carlo moves with respect to (phi, psi) torsion angles. Simulated annealing.
- Use all-atom Knowledge-Based energy function. Add terms to enforce compaction.
- Get reasonable (phi,psi) angles from real protein fragments.



Good at CASP4.

Fit 80 residues to 4.0Å

CASP5 NEW FOLD WINNERS

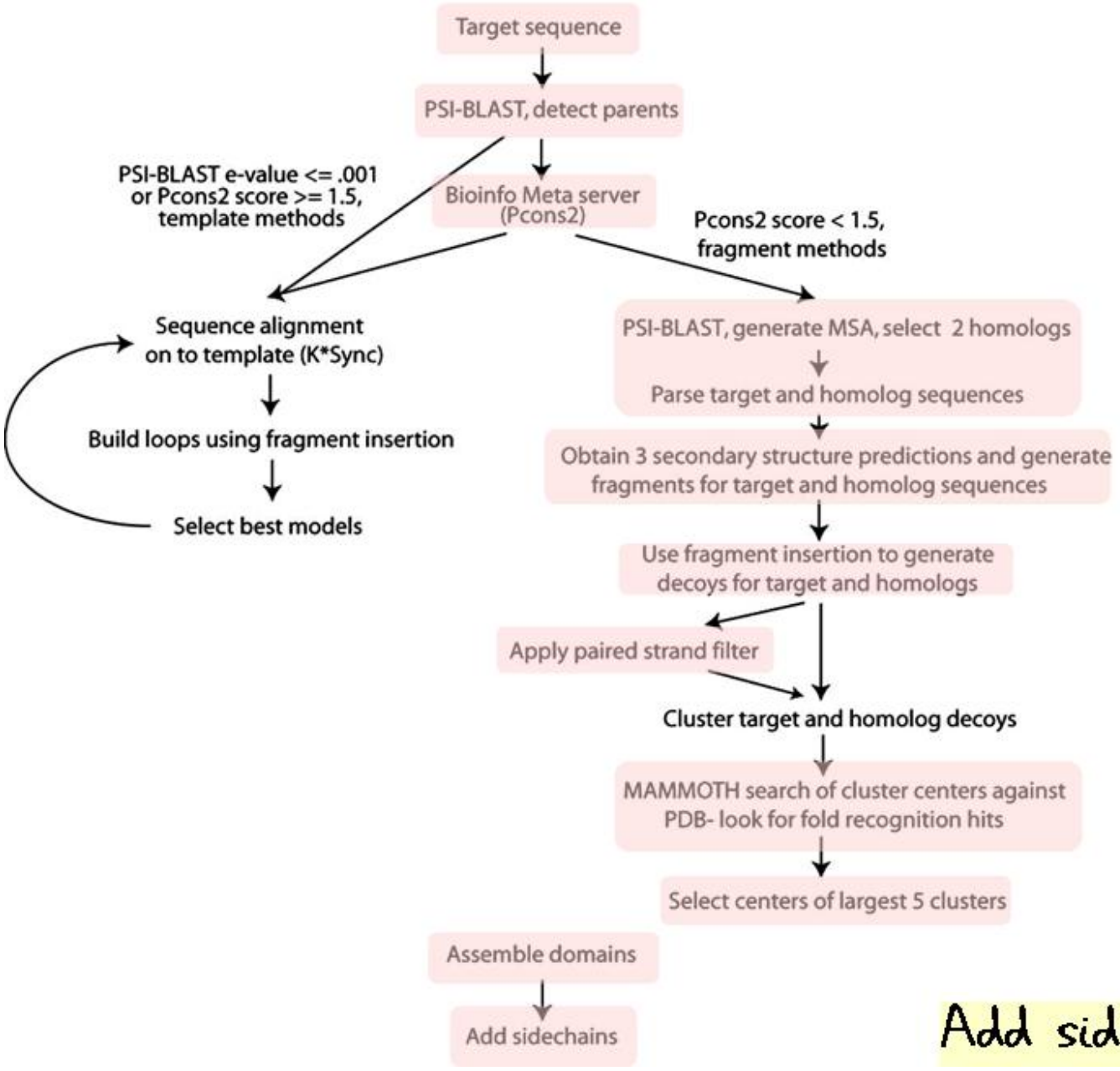
RANK	GROUP	N	ZSCORE
1	Baker	9	25.7
2	Shortle	8	17.5
3	I-sites/Bystroff	5	13.4
4	Skolnick-Kolinski	5	11.8
5	Sam-T02-human	5	11.3
6	Levitt	6	10.5
7	Jones-NewFold	5	10.4
8	Chimera	4	9.1
9	Tome	4	7.6
10	BAKER-ROBETTA	4	7.6
11	PMODEL3	4	7.6
12	SAMUDRALA-NF	3	7.3

- Methods used by top 7 groups.
- Meta-servers do OK.
- One server did well.

WINNING NF METHODS AT CASP5

- Segment Monte Carlo. Baker.
- Shortle.
- I-Sites method. Bystroff.
- Lattice Monte Carlo. Skolnick & Kolinski.
- Segment Assembly. Jones.
- Samudrala Server.

BAKER AT CASP5



Sequence

PSI-BLAST

Bioinfo.pl meta-server

2 homologues

3 predicted secondary structures

Fragment insertion

Strand filter

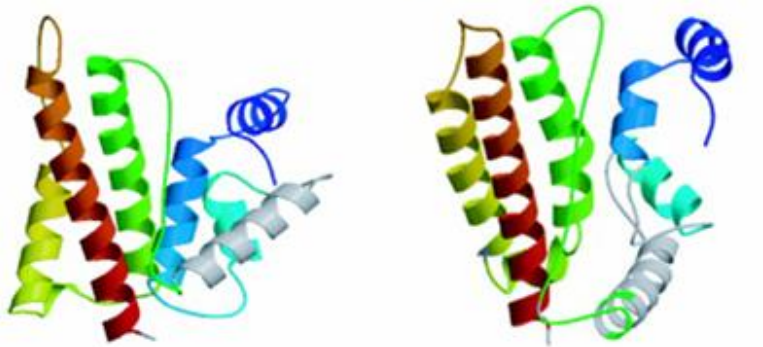
Structure match cluster centers to PDB

Select from 5 clusters

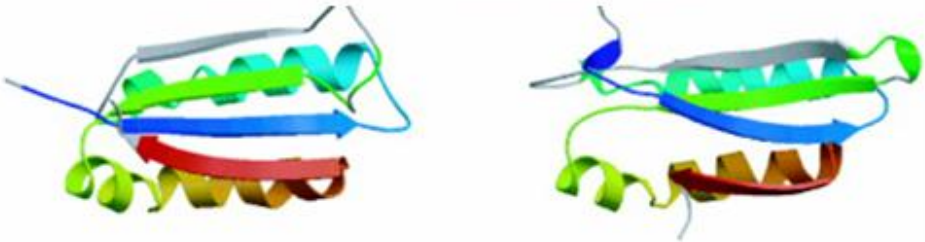
Assemble domains

Add sidechains.

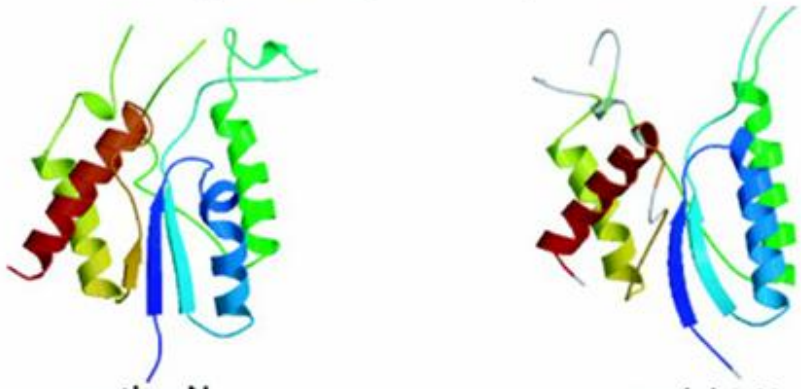
BAKER AT CASP5



T129: HI0817 (full chain 1-182)

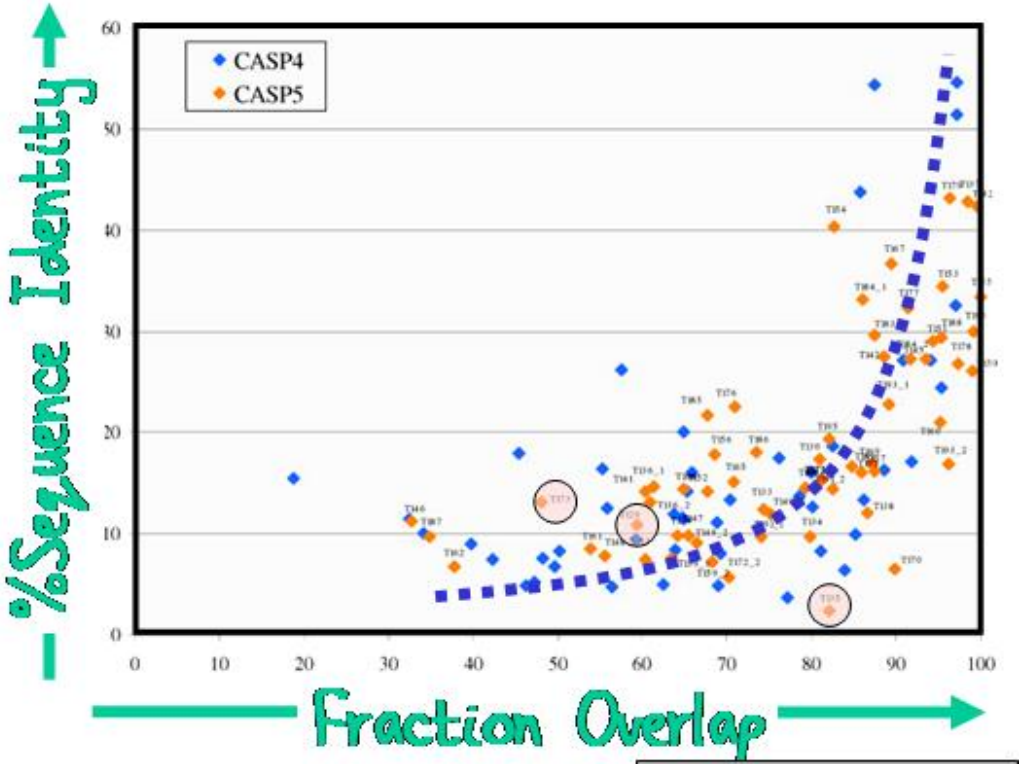


T135: Boiling stable protein (full chain 1-108)

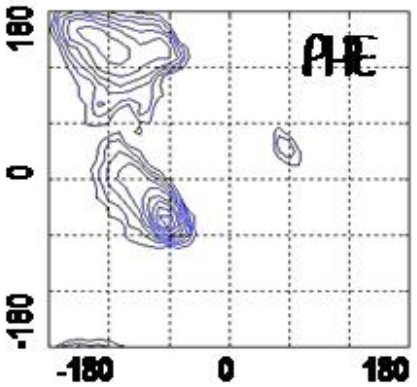


T173: Rv1170 (N-terminal region, 1-127)

- Rosetta does well on some very difficult targets.

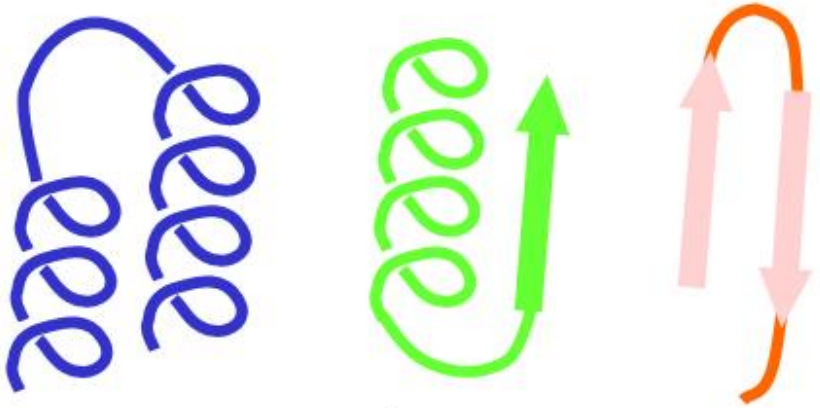


SHORTLE AT CASP5

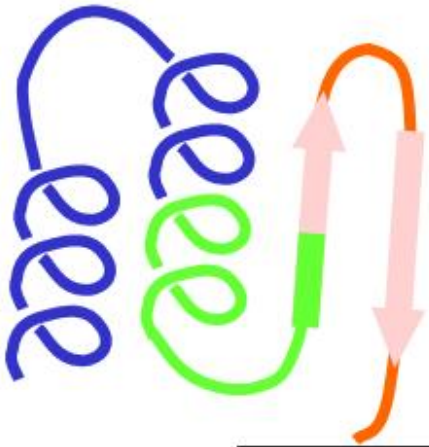


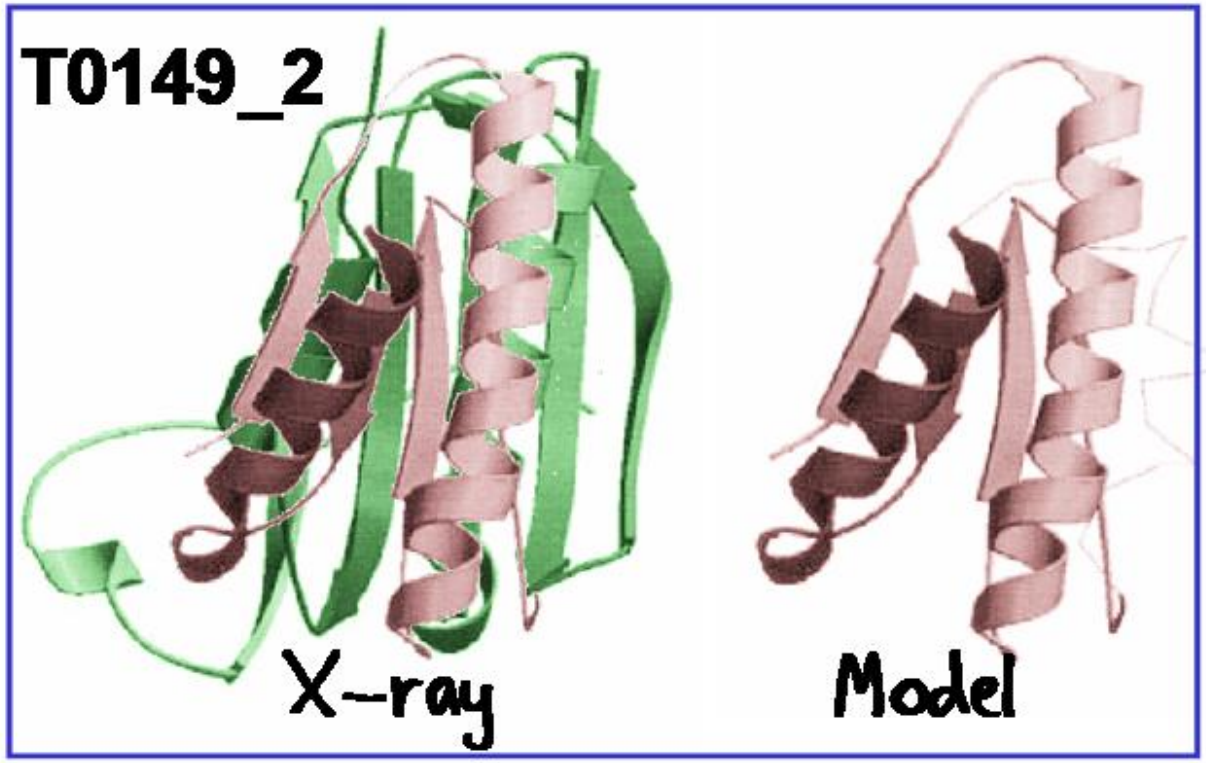
Use (phi,psi) maps to predict extended loops common to homologues.

Novel Ideas.



Assemble by joining in middle of overlapping secondary structure segments.



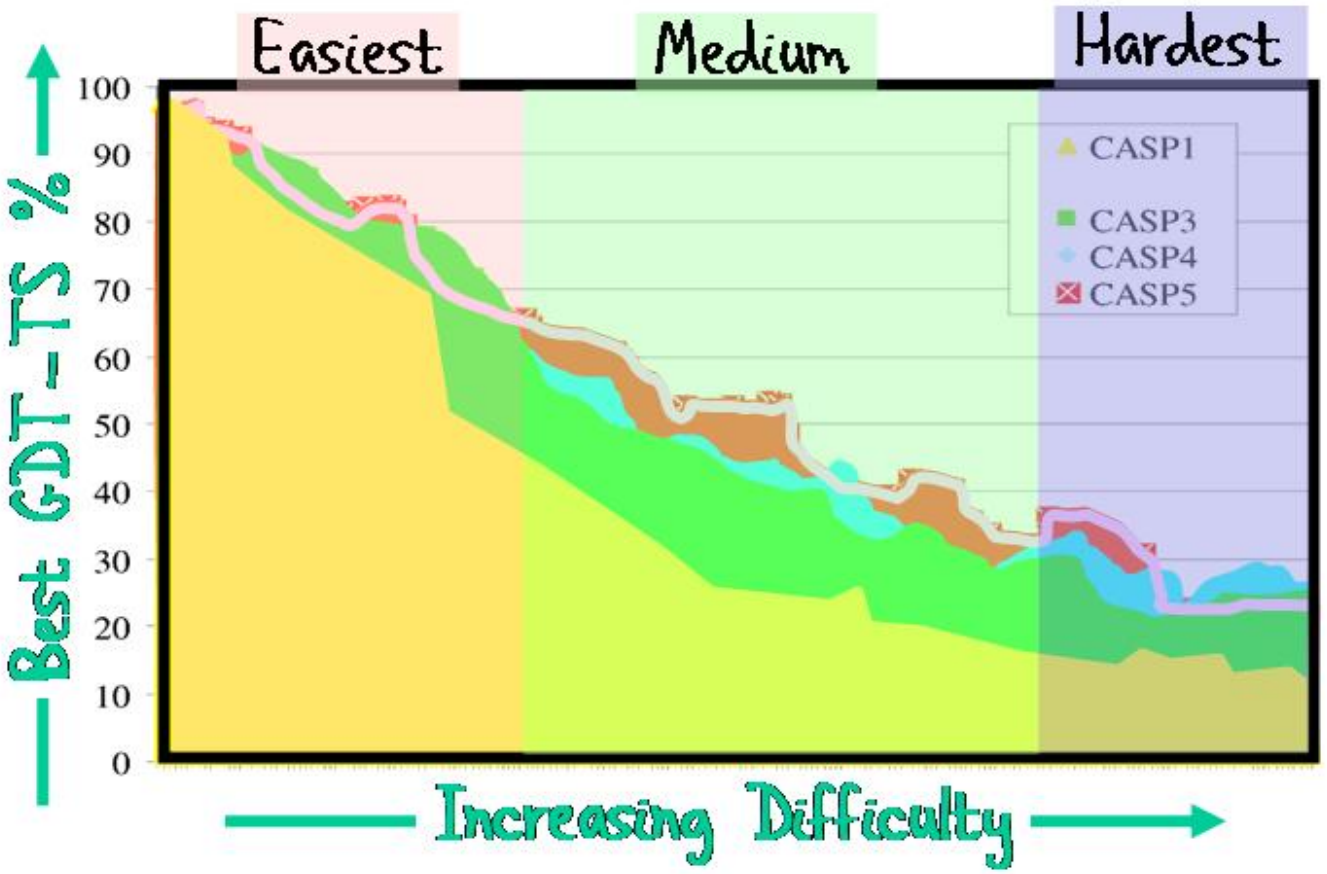


CASP5 LESSONS

- Medium targets improve steadily.

%GDT_TS doubled

Servers and meta-servers



- Easiest targets best at CASP3.

Hand modeling takes time

- Hardest targets best at CASP4.

No new methods

CASP SUMMARY

Did CASP lead to improved methods?

How can Meta-Servers be used widely?

Does CASP inhibit new methods?

THE END

of Lecture 9