Protein Engineering

- **Protein Prediction vs. Protein Folding**
  - Thermodynamic state function vs. kinetics
  - Final state independent of path
  - Optimized density vs. molten globule
  - Global Minimum vs. Dynamic Ensemble of Conformers

Search Problem: How to Locate Global Minimum for a Given Amino Acid Sequence?

Protein Engineering

- **Levinthal Paradox** – So many degrees of freedom, so little time!

For example, average protein of 100 residues has 198 backbone degrees of freedom plus approximately 250 side-chain degrees of freedom. Only exploring trans isomers = 3^{448} combinations to be explored.

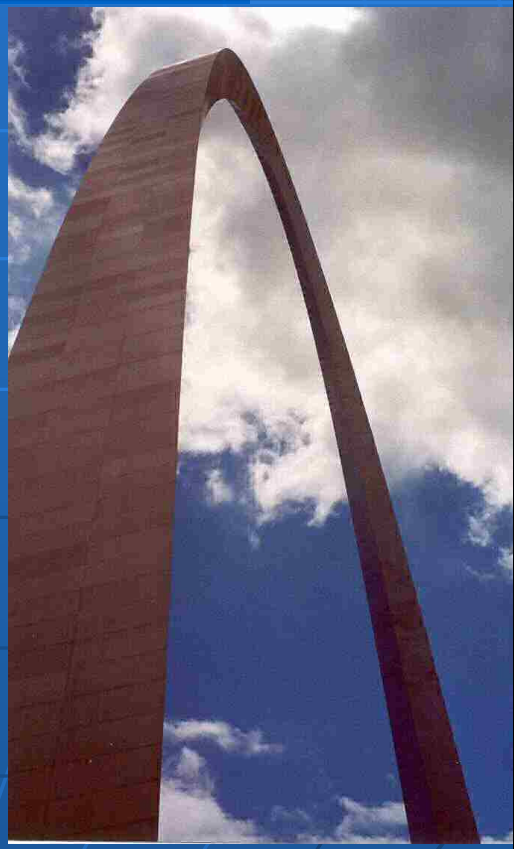
NEVERTHELESS, PROTEINS FOLD!

Proteins have standardized subunits as well - helices (blue), sheets (green) and reverse turns (orange); but any particular sequence is not constrained to only adopt only one of the substructures!

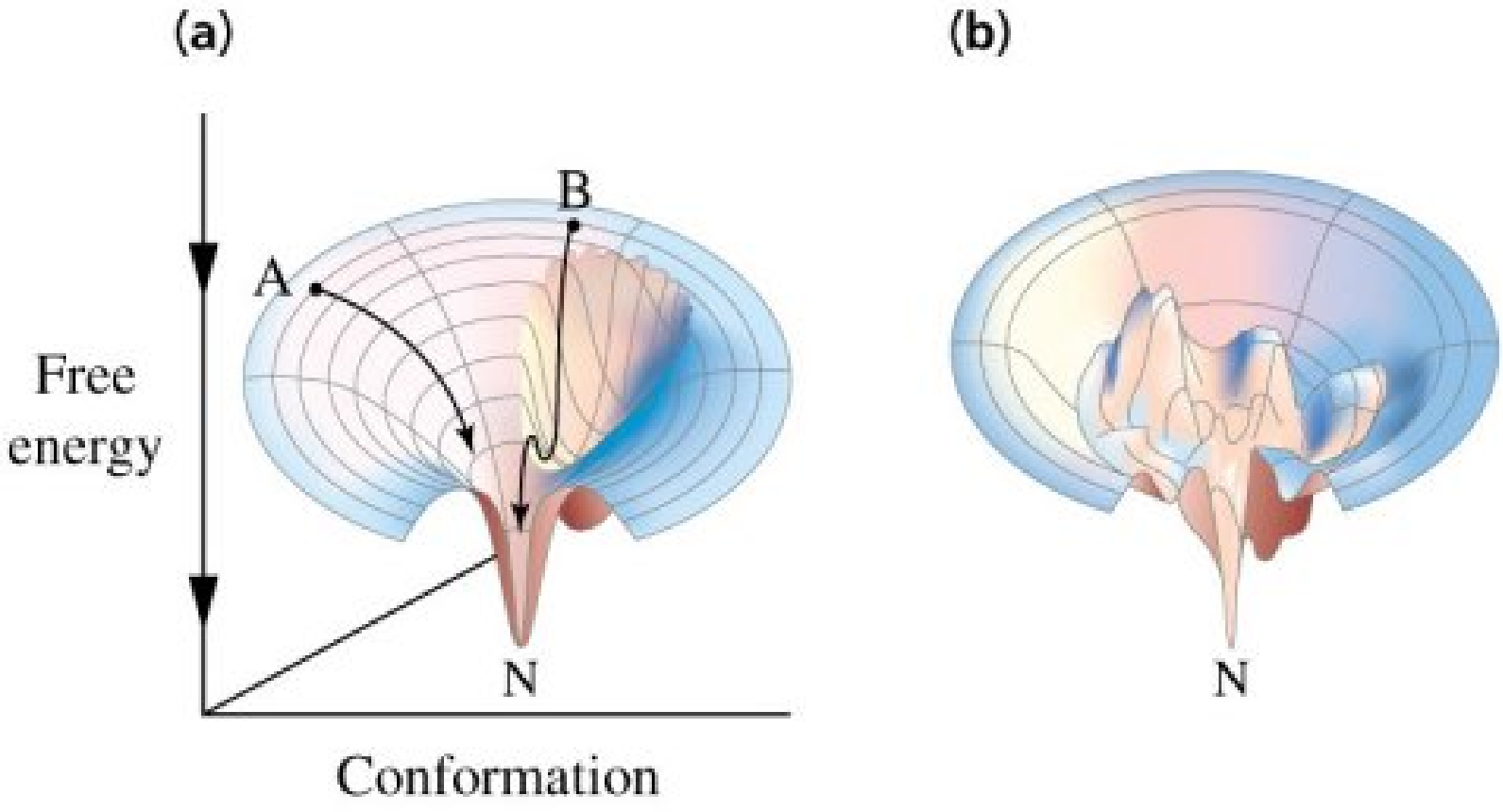
How do we design a building, such as the Gateway Arch?

Standardized materials with predefined geometry and structural properties assembled according to a plan.

Gateway Arch, Eero Saarinen



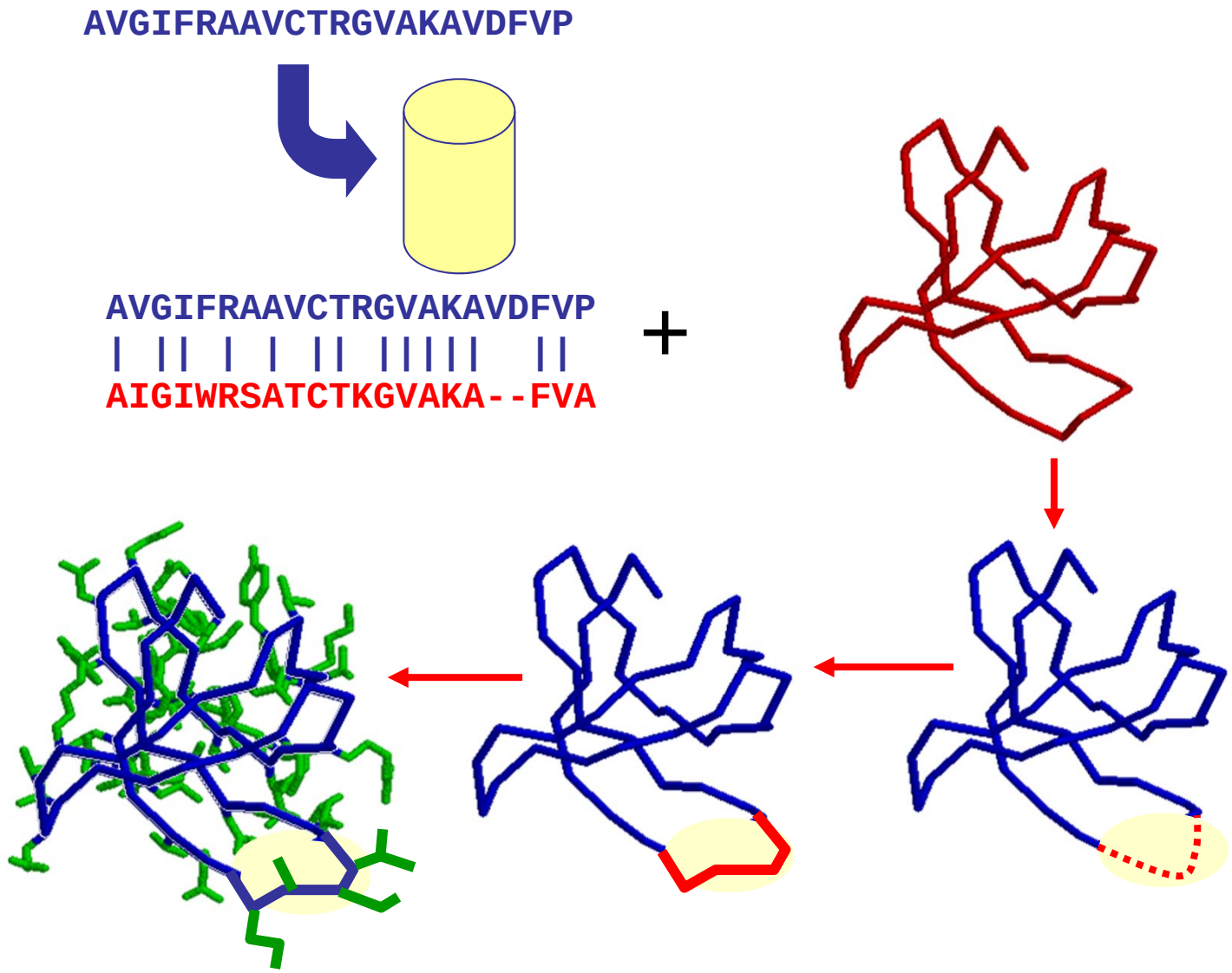
The free energy surface of a protein



Protein Structure Prediction & Design

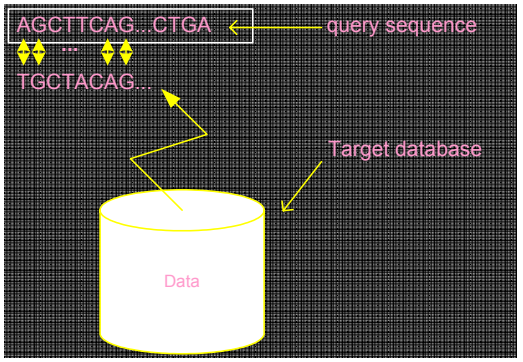
- Full Protein Structure from Sequence
  - High Alignment (comparative modeling)
  - Low Alignment (fold recognition)
  - No Alignment (ab Initio)
- Protein Engineering: Scaffold Design
- Protein Engineering: Function Design

Comparative modeling



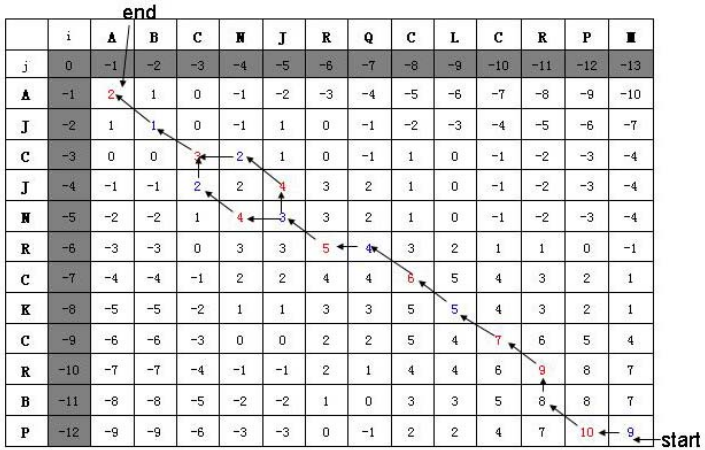
Comparative Homology Modeling

Proteins evolve gradually → 3D structures and functions are often **strongly conserved** during this process. Strong sequence similarity often indicates strong structure similarity
→ Finding family members or similar sequences



Scoring Matrix → Global -or- Local Alignment

	A	R	N	D	C	Q	E	G	H	I	L	K
A	5	-2	-1	-1	-2	0	-1	1	-2	-1	-2	-1
R	-2	9	0	-1	-3	2	-1	-3	0	-3	-2	3
N	-1	0	8	2	-2	1	-1	0	1	-2	-3	0
D	-1	-1	2	9	-2	-1	2	-2	0	-4	-3	0
C	-2	-3	-2	-2	16	-4	-2	-3	-4	-4	-2	-3
Q	0	2	1	-1	-4	8	2	-2	0	-3	-2	1
E	-1	-1	-1	2	-2	2	7	-3	0	-4	-2	1
G	1	-3	0	-2	-3	-2	-3	8	-2	-4	-4	-2
H	-2	0	1	0	-4	0	0	-2	13	-3	-2	-1
I	-1	-3	-2	-4	-4	-3	-4	-4	-3	6	2	-3
L	-2	-2	-3	-3	-2	-2	-2	-4	-2	2	6	-2
K	-1	3	0	0	-3	1	1	-2	-1	-3	-2	6



		H	E	A	G	A	W	G	H	E	E
	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	5	0	5	0	0	0	0	0
W	0	0	0	0	2	0	20	12	4	0	0
H	0	10	2	0	0	0	12	18	22	14	6
E	0	2	16	8	0	0	4	10	18	28	20
A	0	0	8	21	13	5	0	4	10	20	27
E	0	0	6	13	18	12	4	0	4	16	26

y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9 y_{10}

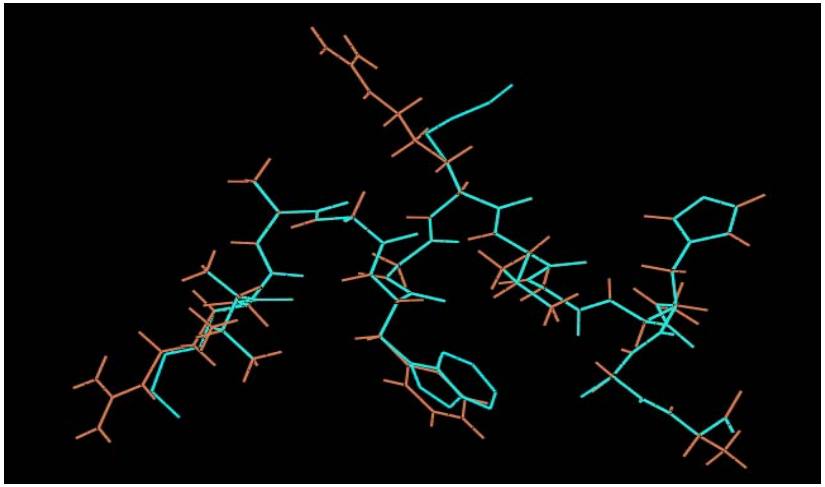
x_1 x_2 x_3 x_4 x_5 x_6 x_7

From Alignment to Structure

Copy aligned backbone from template

Retain conserved side chains

Predict new side chains and loops



- No current comparative modeling method can recover from an incorrect alignment
- Use multiple sequence alignments as initial guide.
- Once a suitable template is found, it is a good idea to do a literature search (PubMed) on the relevant fold to determine what biological role(s) it plays
- Hydrophobic residues exposed/Buried polar without charges balance
- Very large RMSD among the templates
- Different models give very different answers

3D-1D method (Inverse folding)

Instead of aligning a sequence to a sequence, we align a sequence by means of a string of descriptors that describe the 3D environment of the target structure. For each residue position in the structure, we determine:

how buried it is (buried, partly buried or exposed)
the fraction of surrounding environment that is polar (polar or apolar)
the local secondary structure (-helix, -sheet or other)

There are 6 classes of environments to each position in the structure. These environments (E, P1, P2, B1, B2 and B3) depend on the number of surrounding polar residues and how buried the position is. Since there are 3 possible secondary structures for each of these, we have a total of 6 x 3 = 18 environment classes.

$$\text{score}_{ij} = \ln\left(\frac{Pr(\text{residue } j \text{ in enviroment } i)}{Pr(\text{residue } j \text{ in any enviroment})}\right)$$

The denominator is obtained from amino acid frequencies present in the PDB

New score matrix based on structure but not evolution!!!

Position in fold	Environment class	Amino acid type A	C	D	E	F	G	...	R	S	T	V	W	Y	Gap penalty Opn	Ext
1	E	12	-46	22	3	-190	113	...	-32	32	12	-91	-214	-94	2	0.02
2	B_2	-66	-5	-128	-135	105	-166	...	-60	-117	-76	60	102	112	2	0.02
3	E α	46	-44	44	59	-220	68	...	-34	15	-17	-110	-135	-210	200	200
4	P_2 α	6	-93	28	56	-143	-50	...	50	-18	-5	-48	-114	-79	200	200
5	E α	46	-44	44	59	-220	68	...	-34	15	-17	-110	-135	-210	200	200
6	P_2 α	6	-93	28	56	-143	-50	...	50	-18	-5	-48	-114	-79	200	200
7	B_2 α	-69	-10	-162	-71	90	-149	...	6	-147	-150	68	50	85	200	200
8	E α	46	-44	44	59	-220	68	...	-34	15	-17	-110	-135	-210	200	200
9	P_2 α	6	-93	28	56	-143	-50	...	50	-18	-5	-48	-114	-79	200	200
10	B_1 α	-66	-73	-197	-174	132	-253	...	-167	-273	-129	66	100	18	200	200
.	.	.	.	.	.	.	.		.	.	.		.	.	.	.
.	.	.	.	.	.	.	.		.	.	.		.	.	.	.
.	.	.	.	.	.	.	.		.	.	.		.	.	.	.

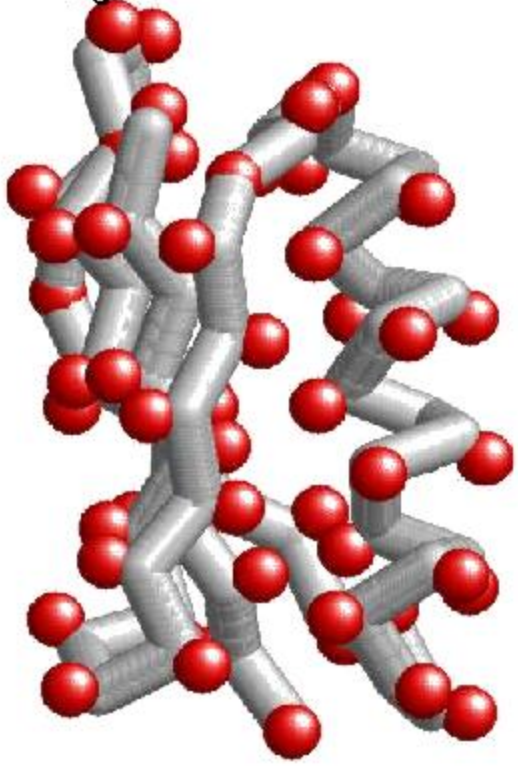
Now you align your target sequence with many **know structure** proteins in a database using this matrix score, obtaining templates, as in comparative modeling…

WHAT IS FOLD RECOGNITION?

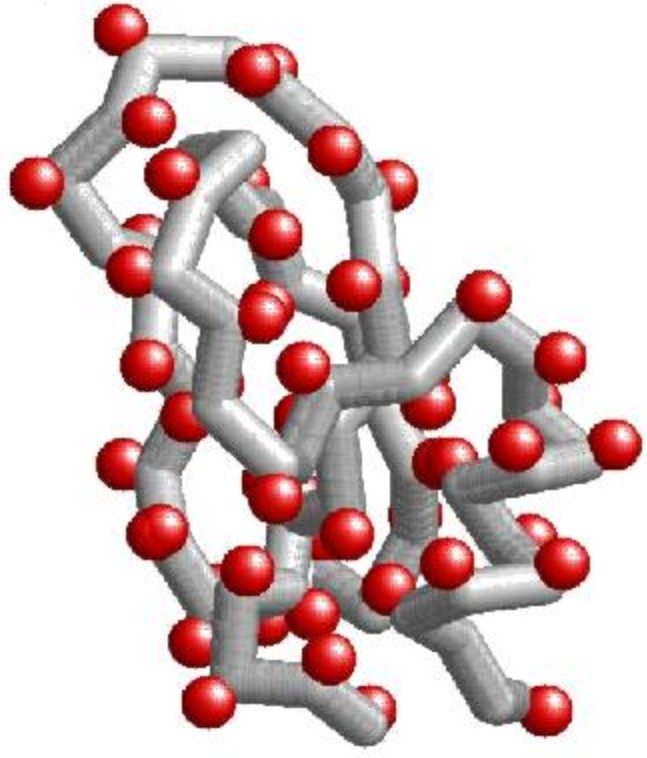
Find the fold that best fits the query sequence.

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

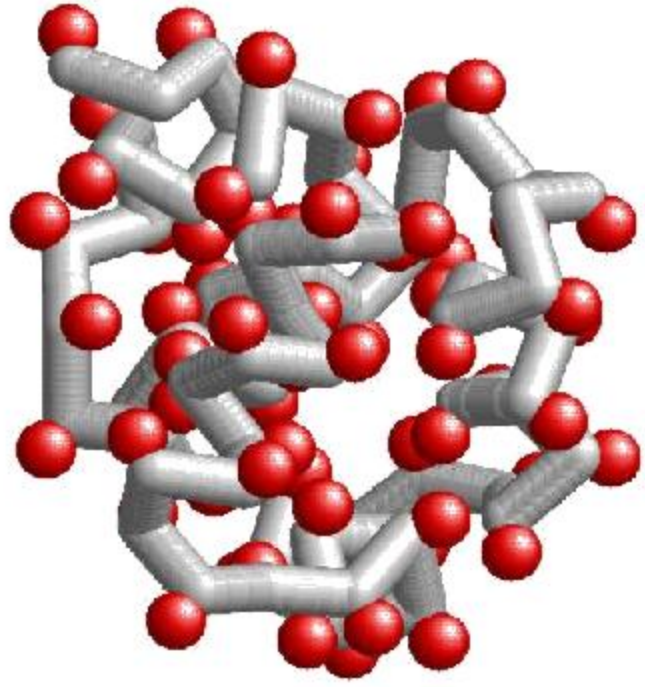
1pgb.pdb



5pti.pdb



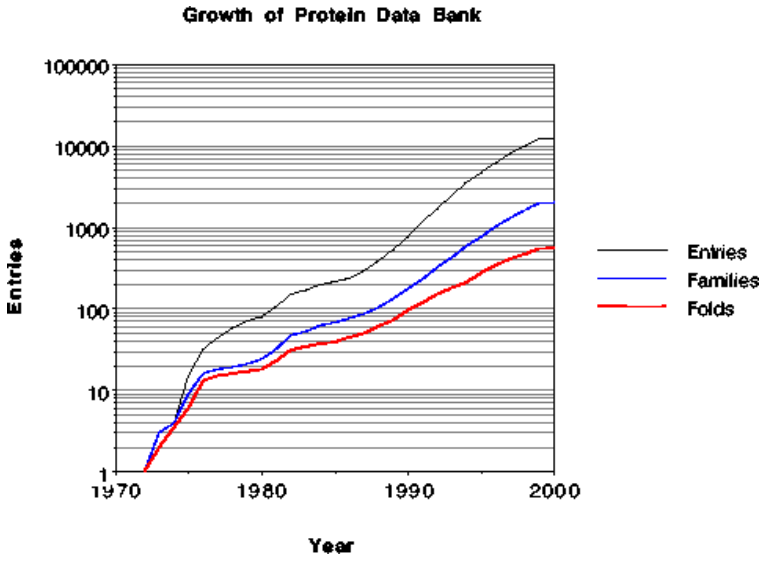
2cro.pdb

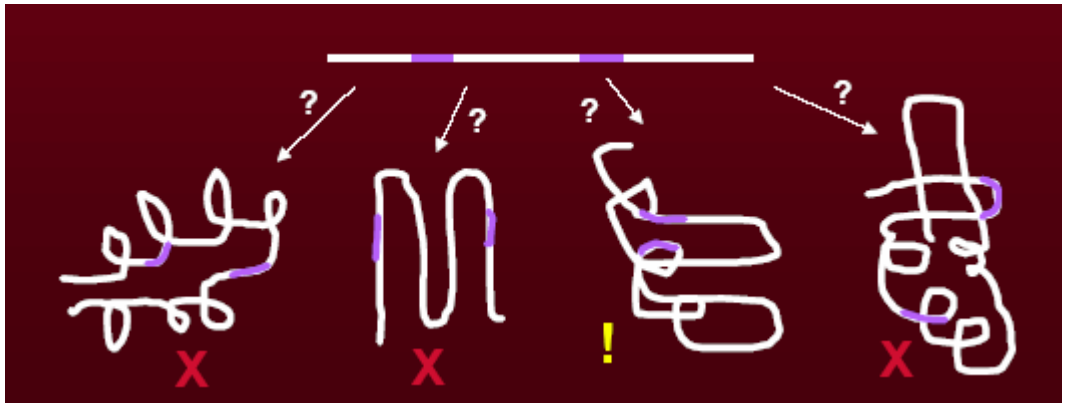


Plus 1000 more folds

Fold Recognition Alignment Method (also called "Threading")

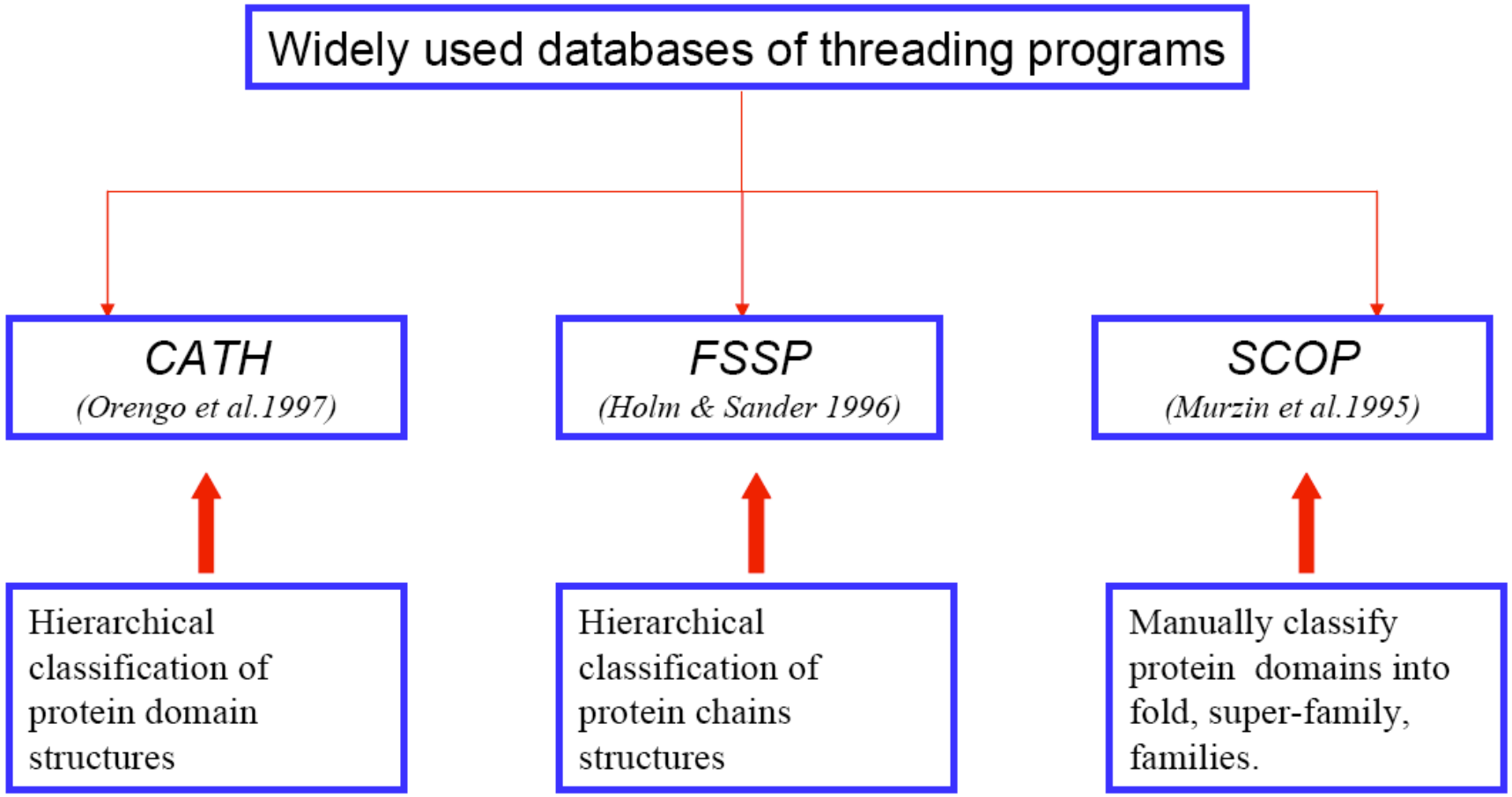
Applied when we have < 20% alignment
Used with a limited # of fold types





→ **Score the folds**

- Pair-wise potential function
- Fold alignment using structural scoring matrix


Widely used databases of threading programs
CATH
(Orengo et al.1997)
FSSP
(Holm & Sander 1996)
SCOP
(Murzin et al.1995)
Hierarchical classification of protein domain structures
Hierarchical classification of protein chains structures
Manually classify protein domains into fold, super-family, families.

SCOP-Structural Classification of Proteins

http://scop.berkeley.edu/

- A database that describes **structural** and **evolutionary** relationships between proteins of known structure.
- Many levels exist in the hierarchy; the principal levels are **family**, **superfamily** and **fold**
- Created mainly by manual inspection.

Family: *Clear evolutionarily relationship*

• Proteins clustered together into families are clearly evolutionarily related. Generally, this means that pairwise residue identities between the proteins are 30% and greater.

• In some cases similar functions and structures provide definitive evidence of common descent in the absence of high sequence identity; for example, many globins form a family though some members have sequence identities of only 15%.

Fold: *Major structural similarity*

• A common fold - the same major secondary structures in the same arrangement and with the same topological connections.

• Different proteins with the same fold often have peripheral elements of secondary structure and turn regions that differ in size and conformation.

• Proteins placed together in the same fold category may not have a common evolutionary origin

Superfamily: *Probable common evolutionary origin*

• Proteins that have low sequence identities, but whose structural and functional features suggest that a common evolutionary origin is probable are placed together in superfamilies. For example, actin, the ATPase domain of the heat shock protein, and hexakinase together form a superfamily.

SCOP: Structural Classification of Proteins. **1.61** release
17406 PDB Entries (1 September 2002). 44327 Domains. 28 Literature References
(excluding nucleic acids and theoretical models)

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	151	257	409
All beta proteins	111	213	362
Alpha and beta proteins (a/b)	117	190	467
Alpha and beta proteins (a+b)	212	308	488
Multi-domain proteins	39	39	52
Membrane and cell surface proteins	12	19	34
Small proteins	59	84	128
Total	701	1110	1940

SCOP: Structural Classification of Proteins. **1.65** release
20619 PDB Entries (1 August 2003). 54745 Domains. 1 Literature Reference
(excluding nucleic acids and theoretical models)

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	179	299	480
All beta proteins	126	248	462
Alpha and beta proteins (a/b)	121	199	542
Alpha and beta proteins (a+b)	234	349	567
Multi-domain proteins	38	38	53
Membrane and cell surface proteins	36	66	73
Small proteins	66	95	150
Total	800	1294	2327

Energy Function

$$E_{total} = E_{mutate} + E_{single} + E_{pair} + E_{gap}$$

E_{mutate} *the* sum of the compatibility measurements $e_{mutate}(a_1; a_2)$ for substituting template amino acid a_1 by target amino acid a_2, PROSPECT use PAM250 matrix for e_{mutate}.

E_{single} the sum of the preferences $e_{single}(a; s; t)$ for aligning amino acid a of the target sequence onto a template position with a structural environment defined by secondary structure s and solvent accessibility t

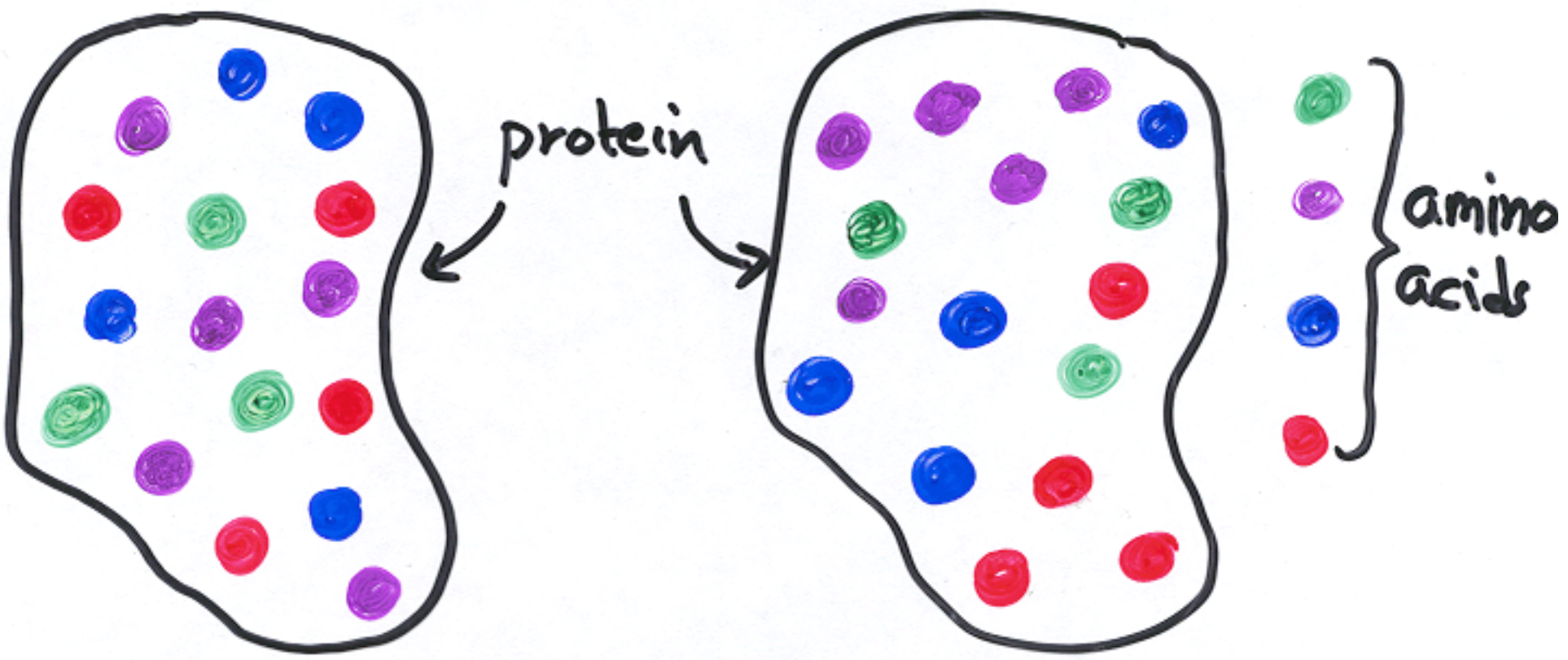
E_{pair} the sum of pair-contact potentials $e_{pair}(a_1; a_2)$ between amino acids a_1 and a_2 of the target sequence when they are aligned to template positions that are spatially close

E_{gap} the sum of the penalties $e_{gap}(g) = 10.8+0.6^* (g - 1)$ for an alignment gap of length g

D. Xu and Ying Xu, "Computational Studies of Protein Structure and Function using Threading Program PROSPECT"

KNOWLEDGE - BASED ENERGIES

(Potentials of mean force, statistical energy)



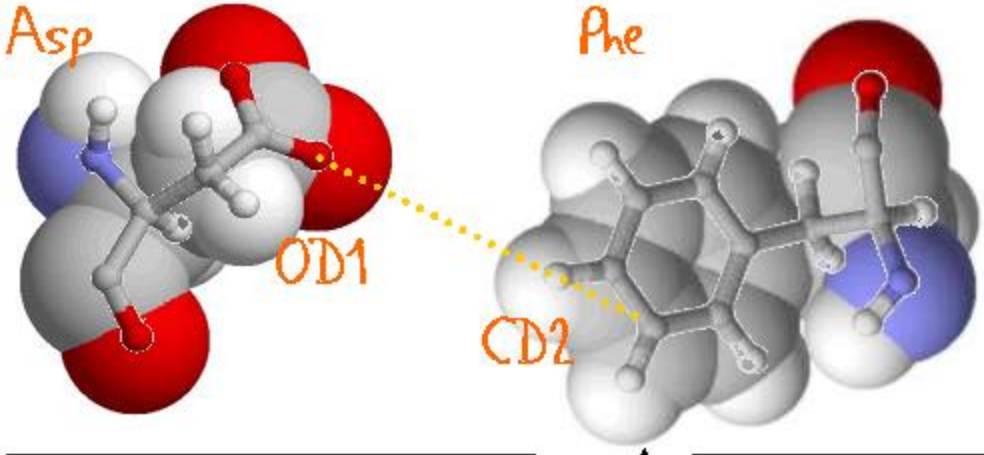
Random State
(Shuffle labels)

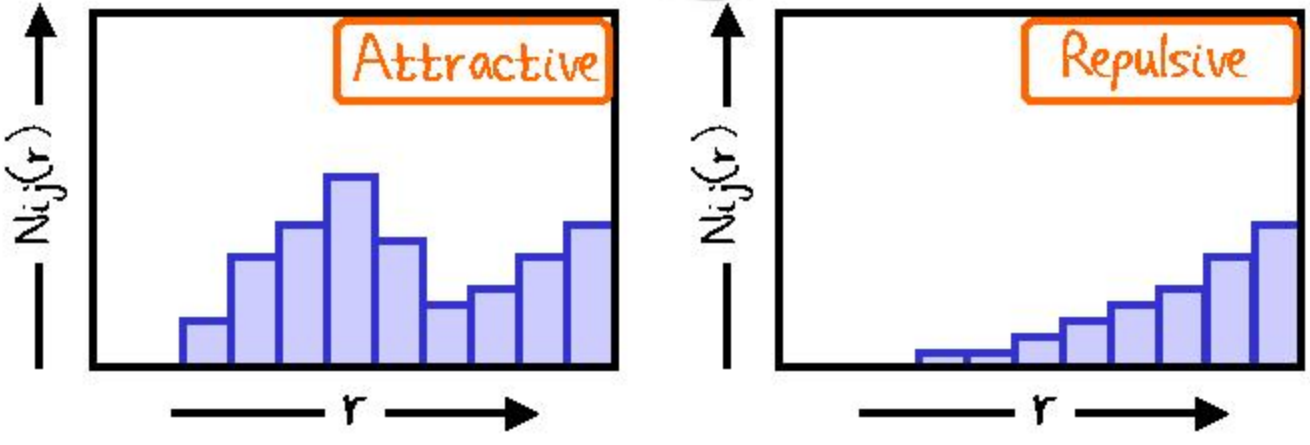
Native Protein

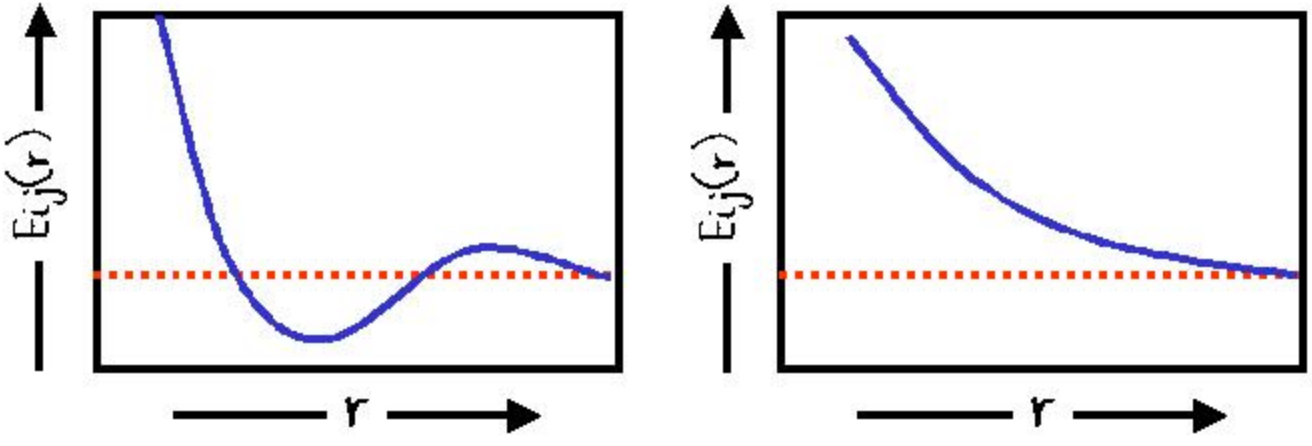
- Count pairs of each residue or atom type at different separations $n_{ij}(r)$

- Normalize by expected count $n_{ij}^{o}(r)$
- Convert to additive "Energy" $\log_e (n_{ij}(r) / n_{ij}^{o}(r))$

KNOWLEDGE-BASED ENERGIES







- Get distribution of distances between pairs of atom centers of a particular type, e.g. D–OD1...F–CD2.

- Normalize and take log to get Energy score:

$$E_{ij}(r) = \log[N_{ij}(r)/M_{ij}(r)]$$

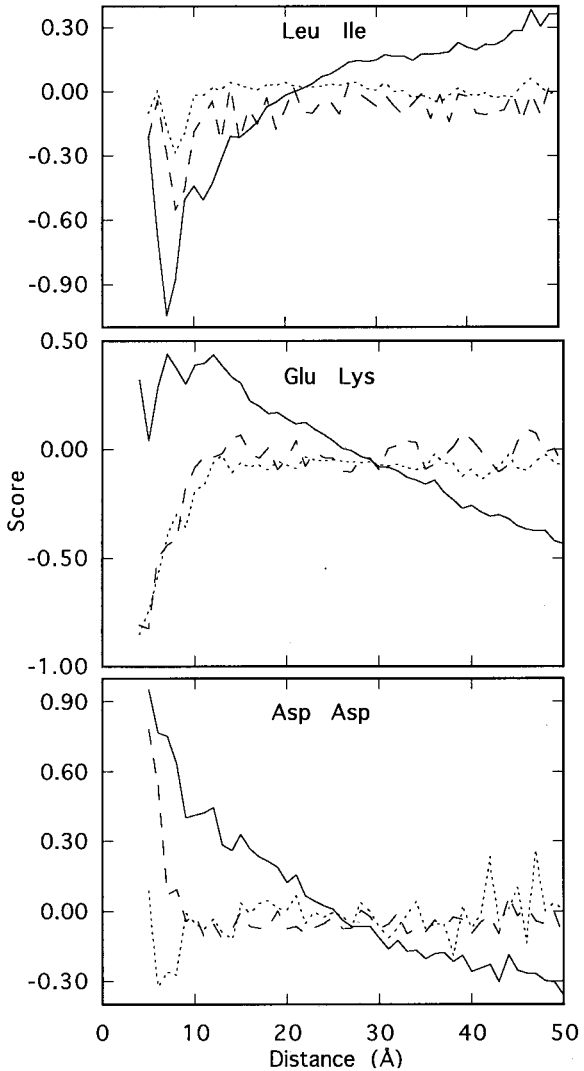


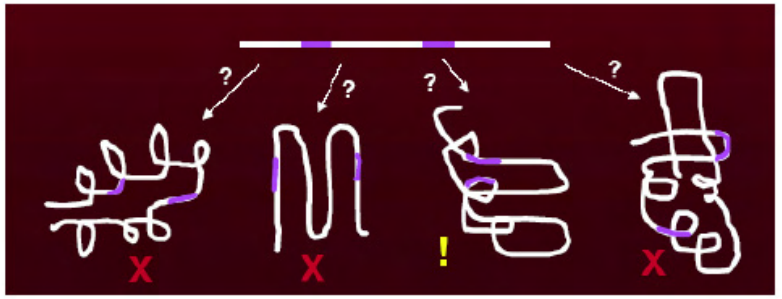
Figure 4. Comparison of the negative logarithms of equation (5) and the residue pair specific second term in equation (8) for sequence separations greater than ten. Residues with greater than 16 neighbors were considered buried. Continuous lines, equation (5); dotted lines, equation (8) both residues buried; broken line, equation (8) both residues exposed.

Threading methods:

Given: a structure P with positions $p_1, p_2, \ldots, p_n$, and a sequence $s_1, \ldots, s_m$.
Find: $t_1, t_2, \ldots, t_n$ (where $1 \leq t_1 < t_2 < \cdots < t_n \leq m$ and t_i indicates the index of the amino acid from s that occupies p_i) such that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \text{score}\,(i, j, s_{t_i}, s_{t_j})$$

is maximized.



Instead of modeling energies from first physical principles, simplify the problem by positioning only amino acids, and compute empirical energies from the observed associations of amino acids

E(interaction) = -KT ln[frequency of interaction]

where *K* is constant, *T* is temperature (constant), frequency of interaction measured in database of known structures.

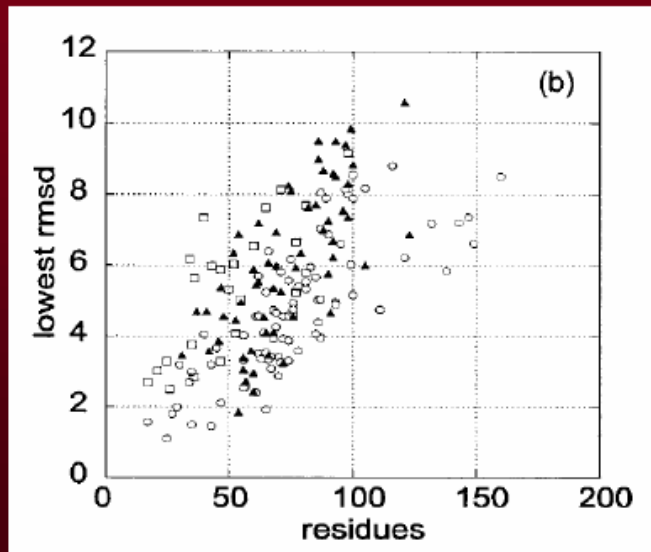
More frequent → more favorable.

Rosetta Method for *ab initio Modeling*

Creative ways to memorize sequence: structure correlations in short segments from the PDB, and use these to model new structures. ROSETTA Method.

1. Break target into fragments of 9 (25) and 3 (200) amino acids → Using MSA profile and predicted secondary structure
2. (from totally extent) Insert 9 mers from database. Metropolis Monte Carlo for 2000 steps (steric clashes criteria)
3. 2000 more steps with residue-residue scores : **hydrophobic burial** and **specific pair interactions** and secondary structure packing scores.
4. 10 iterations 2000 steps during which the local strand-pairing score is cycled on and off to promote formation of nonlocal -strand pairing over local strand kinetic traps, whereas the local atom density is pushed toward that of native protein structures
5. 4000 3-mer fragment insertions; a term linear in the radius of gyration is added to help condense the model and a higher resolution model of strand pairing is used

Performance of Rosetta Method



The final decoy is stored only if it passes several filters. Between 10,000 and 400,000 independent simulations starting from different random number seeds

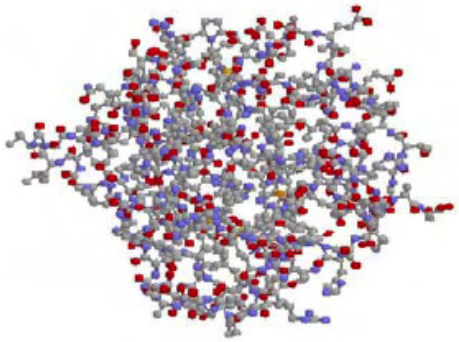
Ab initio prediction {

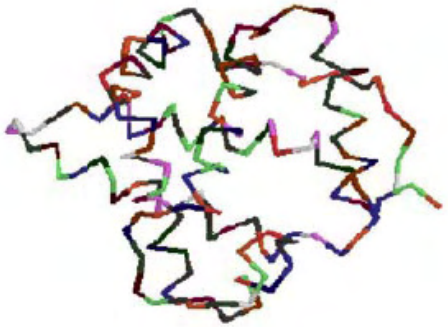
an accurate potential function that permits calculation of the free energy given a structure

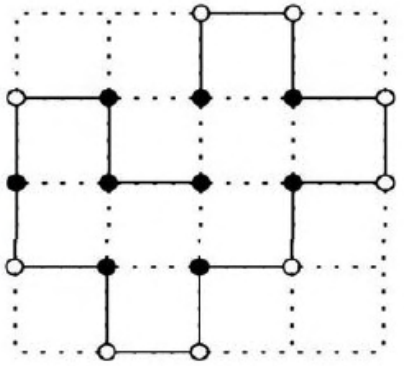
an efficient method for searching for energy minima

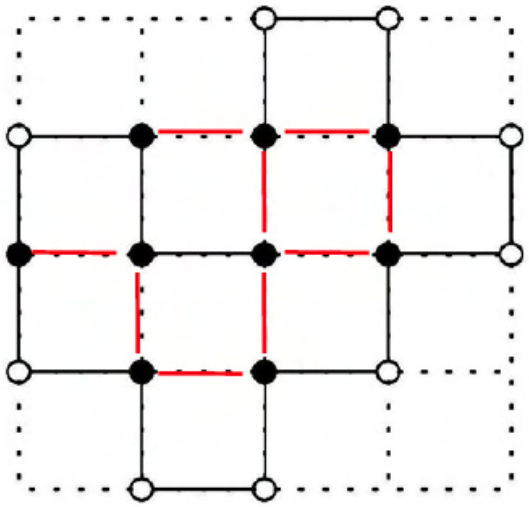
Simplified potential model: Lattice models

- Models are used to reduce the search space (simplify the computation)
  Three kinds of models
  - Lattice models
  - Discrete state off-lattice models





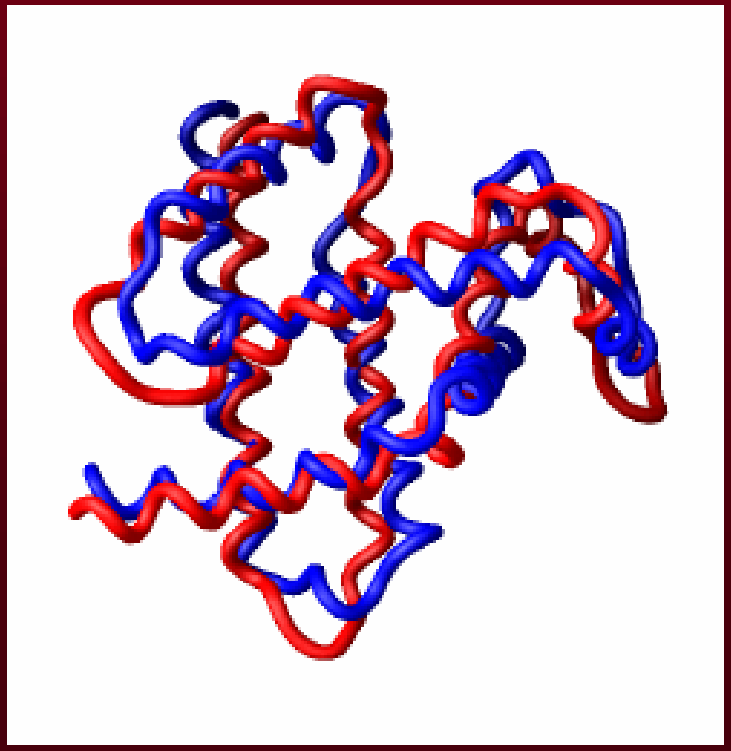




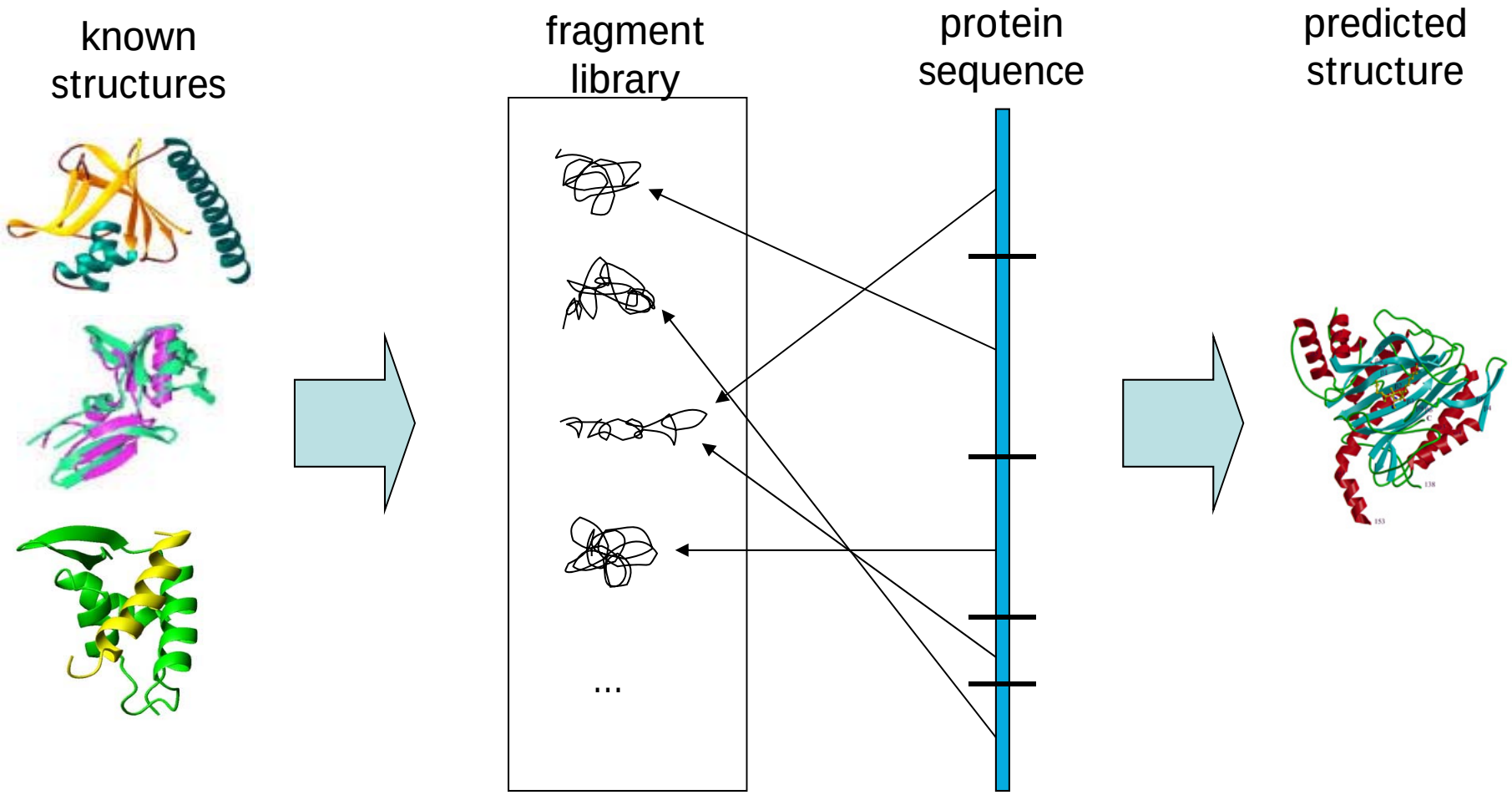
- ● hydrophobic amino acid
- ○ hydrophilic amino acid
- — Covalent bond
- — H-H contact

- Goal: maximize the number of H-H contacts

Comparison of calculated (red) and experimental (blue) structures for the protein myoglobin using the refined potential function. The calculated structure is the lowest energy structure obtained from 3 different jobs with clustering and energy selection. The total simulation time on a 16 node partition CM-5 massively parallel computer was 60 hours, in which about 5 billion structures were generated. The RMS deviation of the two structures is 6.2 Å.



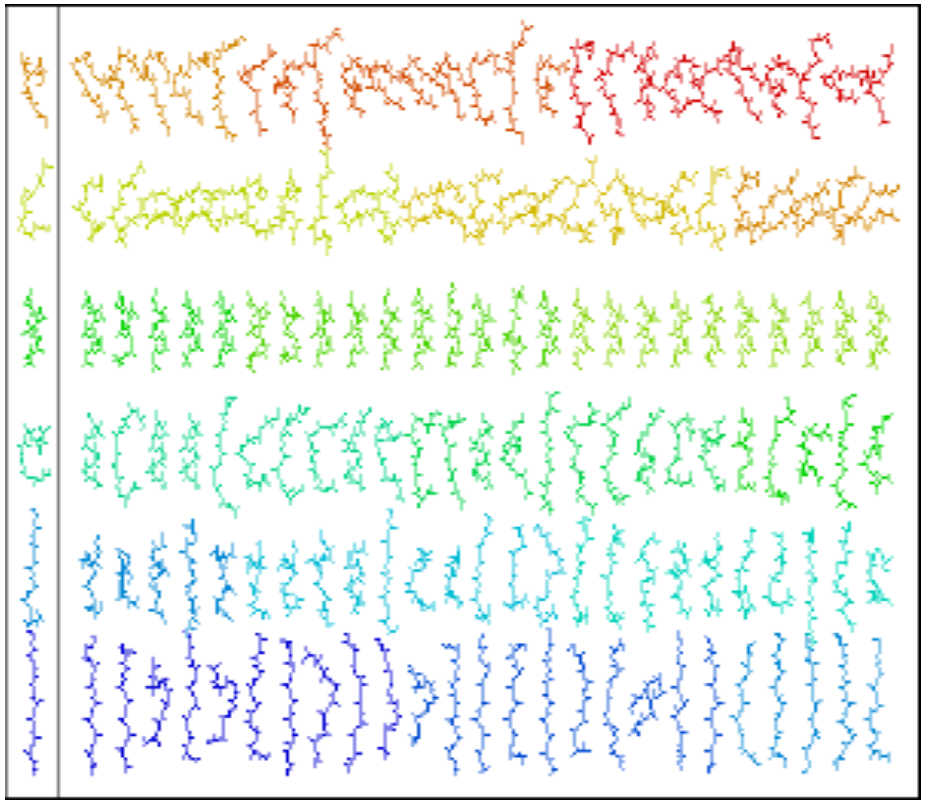
Fragment assembly



Rosetta/Robetta

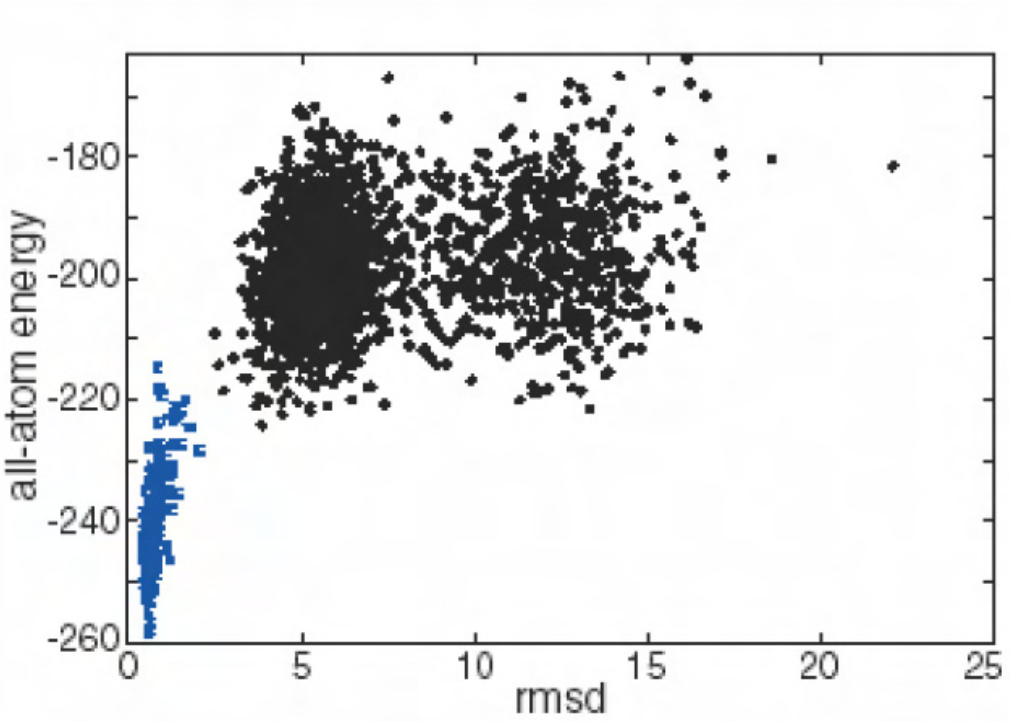
- Decoys are assembled from fragments
- Lowest energy model from a set of generated decoys is selected as the prediction
- Monte Carlo simulated annealing
- Physical energy function with elements of a statistical potential

Fragment library



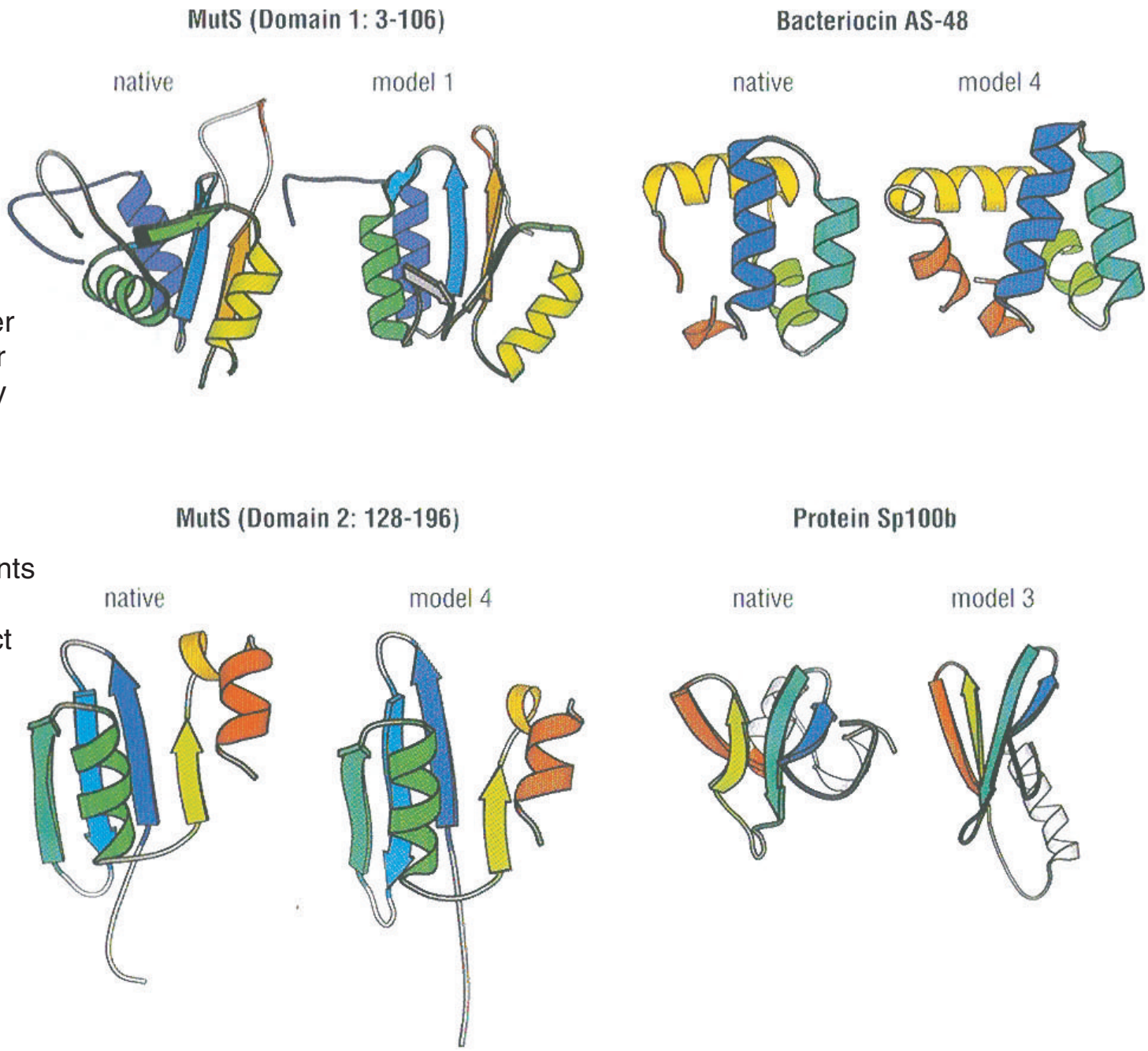
http://robetta.bakerlab.org/

Conformational sampling challenge



Rosetta Uses a Fragment Library + Monte Carlo Search

Examples of the best-center cluster found by *Rosetta* for some test proteins. In many cases the overall fold is predicted well enough to be recognizable. However, relative positions of the secondary structure elements are almost always shifted somewhat from their correct values.



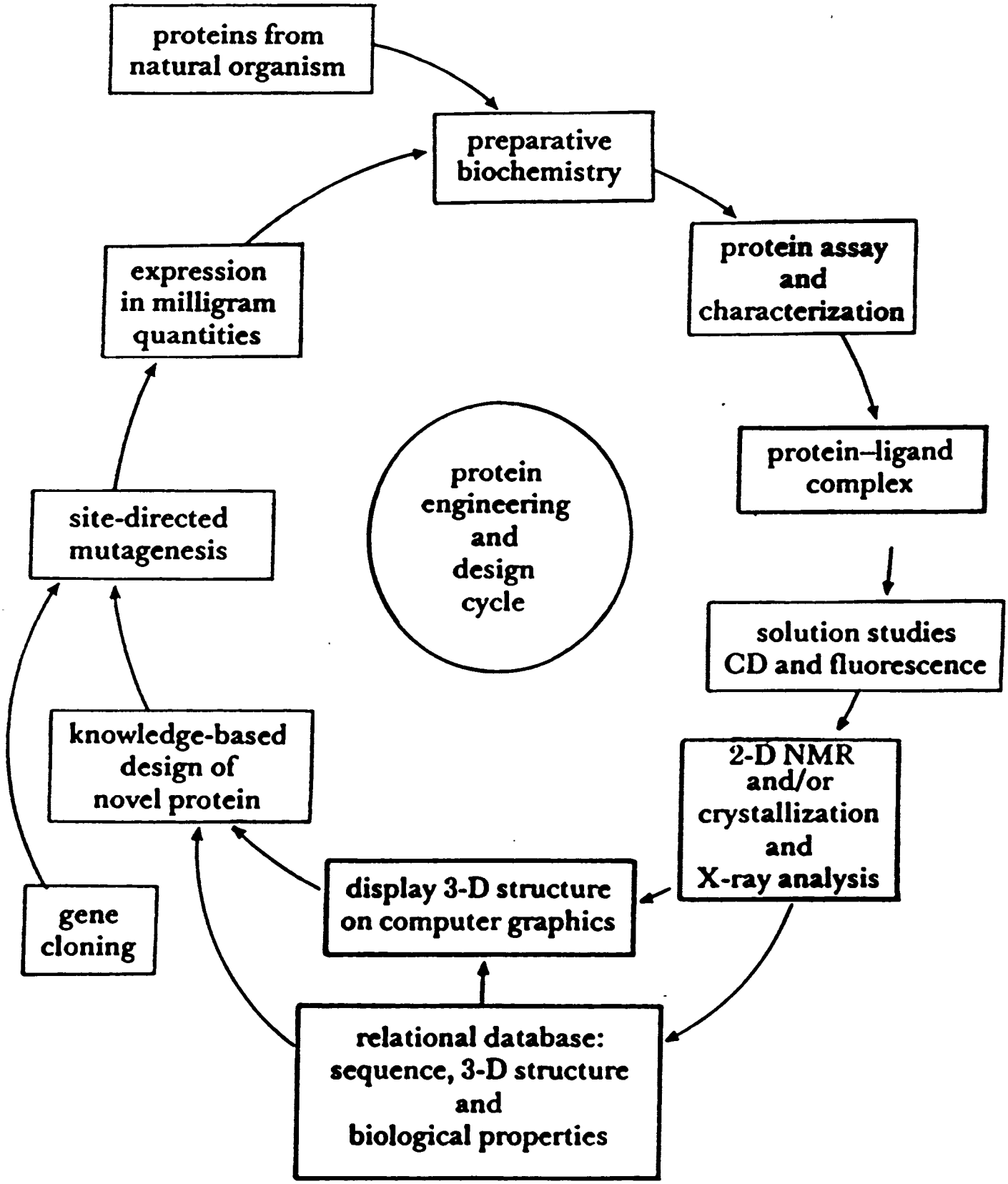


FIGURE 1. The protein engineering and design cycle.

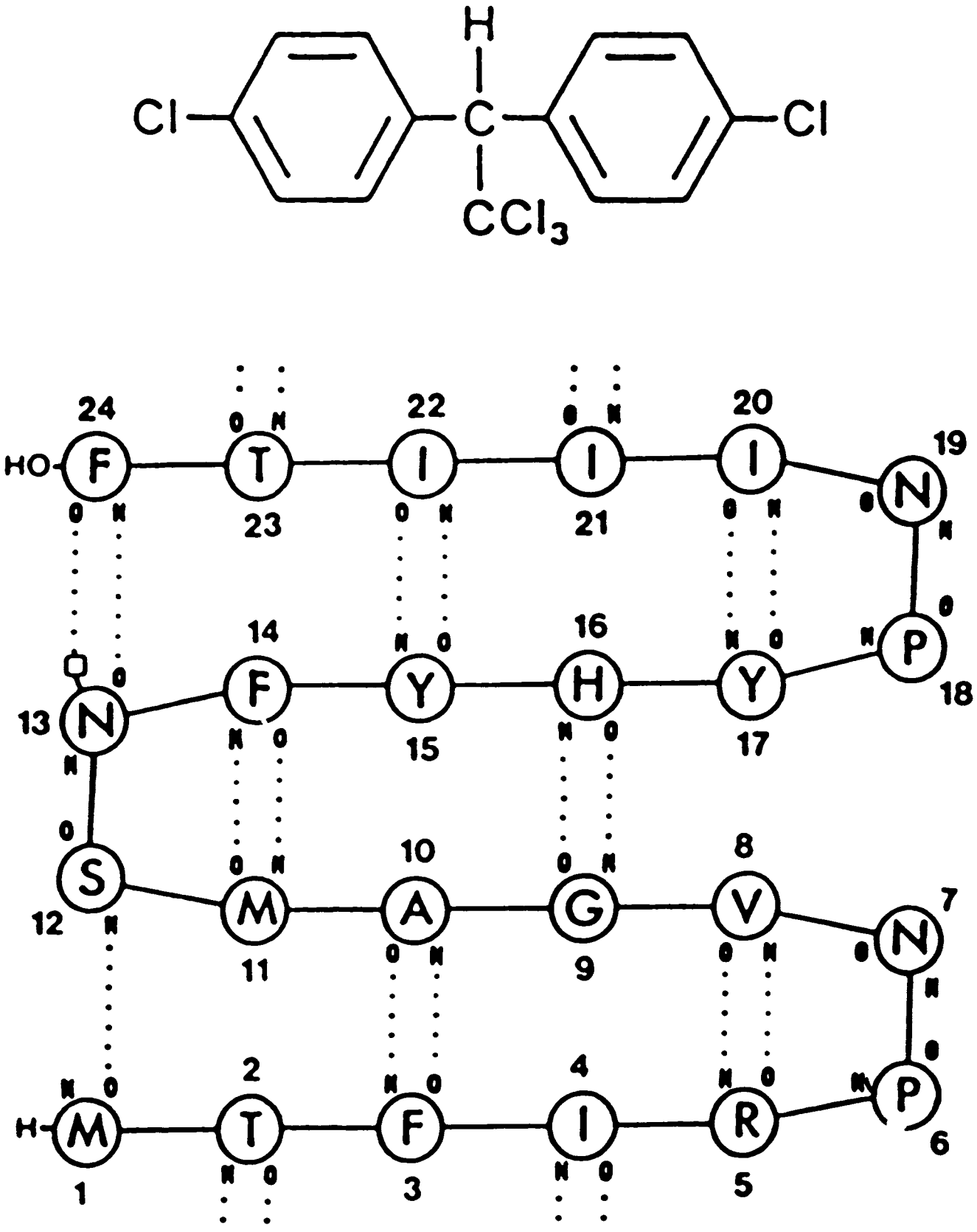


Fig. 1. Structural formula of 4,4'-DDT [1,1,1-trichloro-2,2-bis(4-chlorophenyl)-ethane] (top) and sequence and proposed secondary structure of the designed 24-residue DDT-binding polypeptide (bottom). Dotted lines indicate hydrogen bonds between NH and CO groups of the backbone and side chains.

Table 1. Dissociation constants of the complexes of artificial DDT-binding peptides with DDT and DDT analogues; dioxin, hemin, 2'-CMP, and tyrosine were used as controls

DBP analogue	DBP[a,b]	Val^{14}-	Trp^{16}-	Thr^{16}-	Thr^{15}, Thr^{17}-	Ile^{3}, Phe^{4}-	random[c]
K_D(M) of DDT-peptide complex	9×10^{-7}	3×10^{-5}	2×10^{-6}	7×10^{-5}	4×10^{-6}	2×10^{-6}	6×10^{-4}

Ligand	DDD	DDA	dioxin	hemin	2'-CMP	tyrosine
K_D(M) of ligand-DBP complex	1.5×10^{-6}	4.5×10^{-5}	$\gg10^{-3}$	5.0×10^{-4}	$>10^{-3}$	$>10^{-3}$

[a]Original DBP (Figure 1). [b]Using UV difference spectroscopy, the K_D of the DDT-DBP complex was 5×10^{-7} M. [c]Randomized DBP sequence: Ser-Arg-Pro-Thr-Ile-His-Asn-Asn-Ile-Thr-Tyr-Phe-Val-Pro-Gly-Phe-Ala-Met-Tyr-Met-Asn-Ile-Ile-Phe.

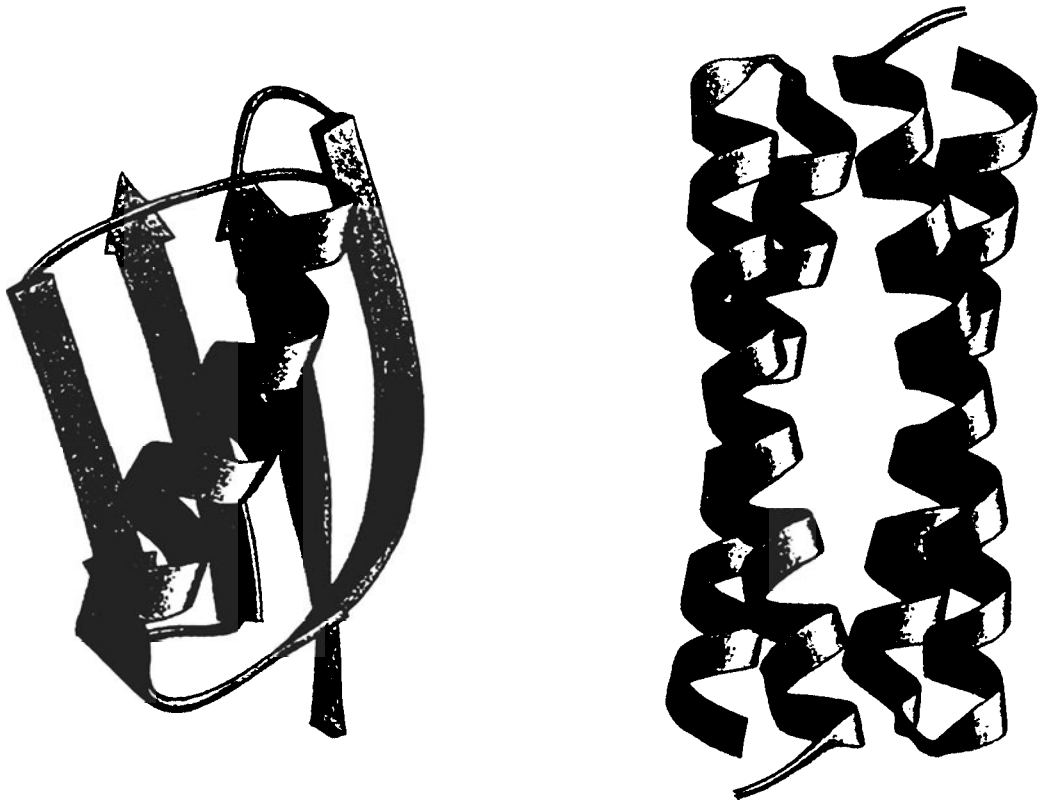


Figure 17.16 Ribbon diagram representations of the structures of domain B1 from protein G (blue) and the dimer of Rop (red). The fold of B1 has been converted to an α-helical protein like Rop by changing 50% of its amino acids sequence. (Adapted from S. Dalal et al., *Nature Struct. Biol.* 4: 548–552, 1997.)

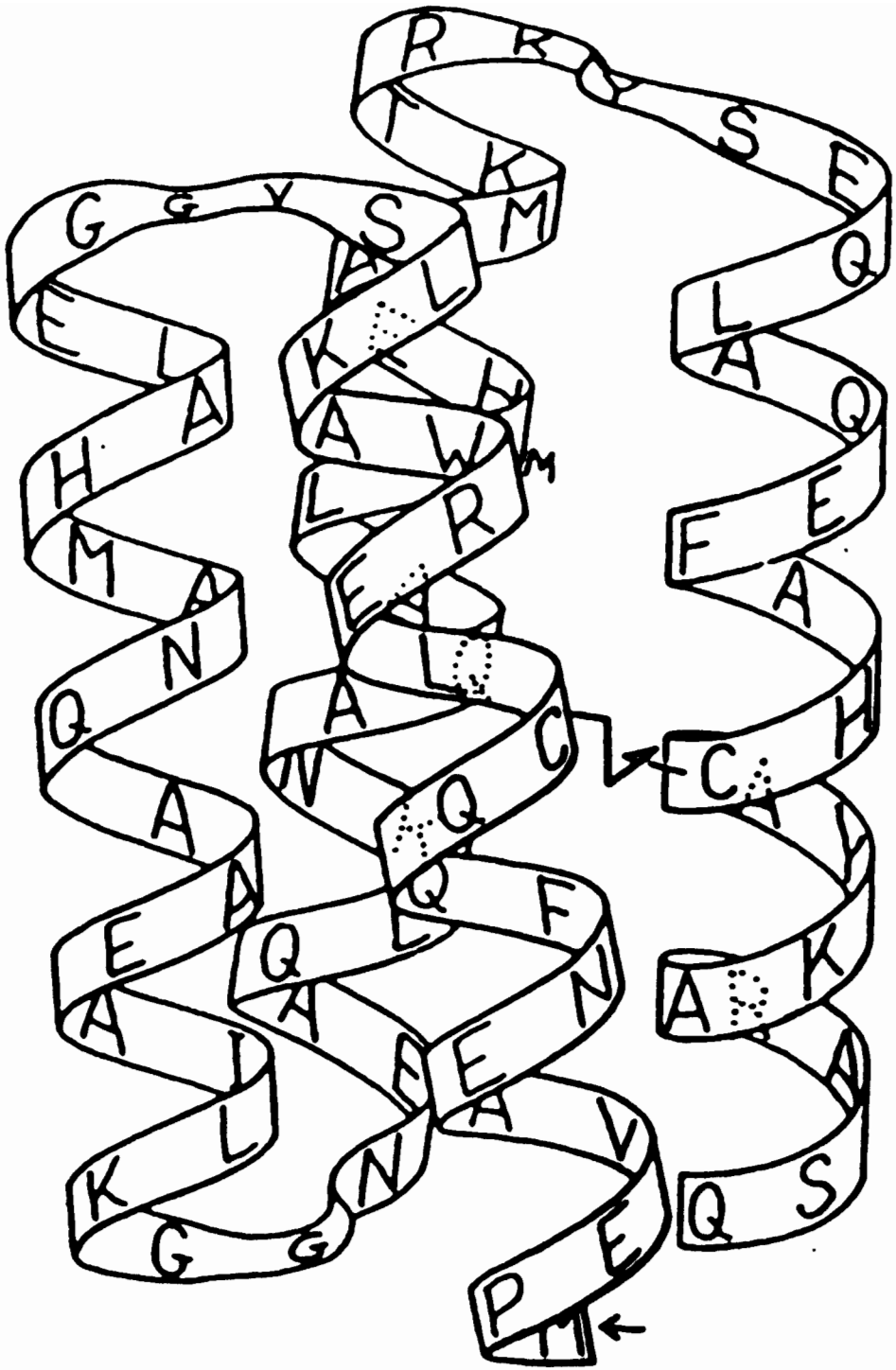


Fig. 1. Ribbon drawing of the sequence and proposed three-dimensional structure of Felix, with the disulfide indicated. The chain turns left at the top of the first helix, making this a "left-turning", up-and-down, four-helix bundle. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

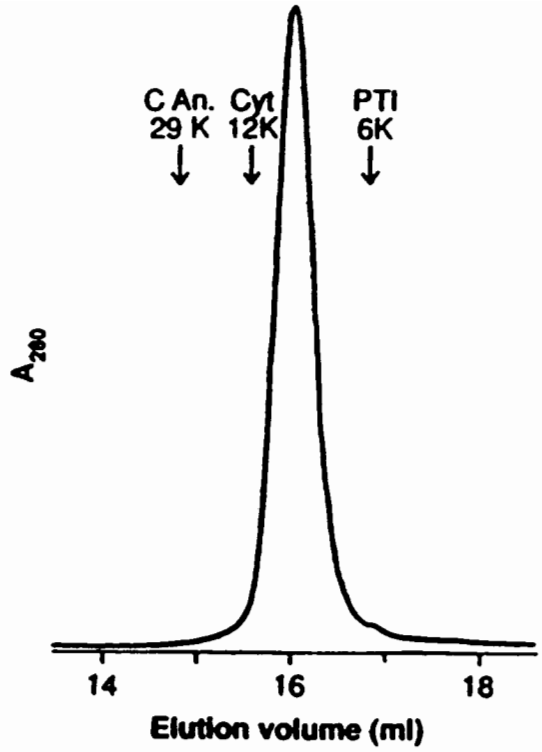


Fig. 8. Size exclusion chromatography indicating that Felix is a monomer under nondenaturing conditions. Felix S-S, at an initial concentration of 0.2 mg/ml, was chromatographed on a prepacked Superose-12 gel filtration column (Pharmacia) in 50 mM sodium phosphate, pH 7.6, 100 mM NaCl. Molecular size markers are carbonic anhydrase (C An), cytochrome c (Cyt), and bovine pancreatic trypsin inhibitor (PTI).

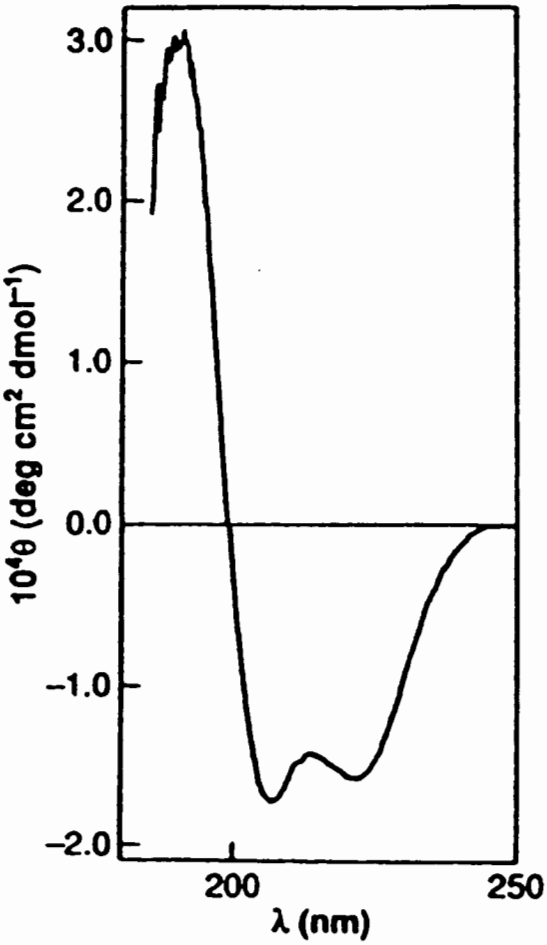


Fig. 9. Circular dichroism spectrum of Felix S-S. Felix S-S was dissolved at 0.09 mg/ml (~10 μM) in 50 mM sodium phosphate, pH 7.6 in a cuvette with a 1-mm path length. Spectra were measured at 2°C with a Jobin Yvon Mark 5 spectropolarimeter interfaced with an Apple computer. The data were collected in triplicate from 260 nm to 185 nm with a step size of 0.2 nm, and a response time of 2 seconds.

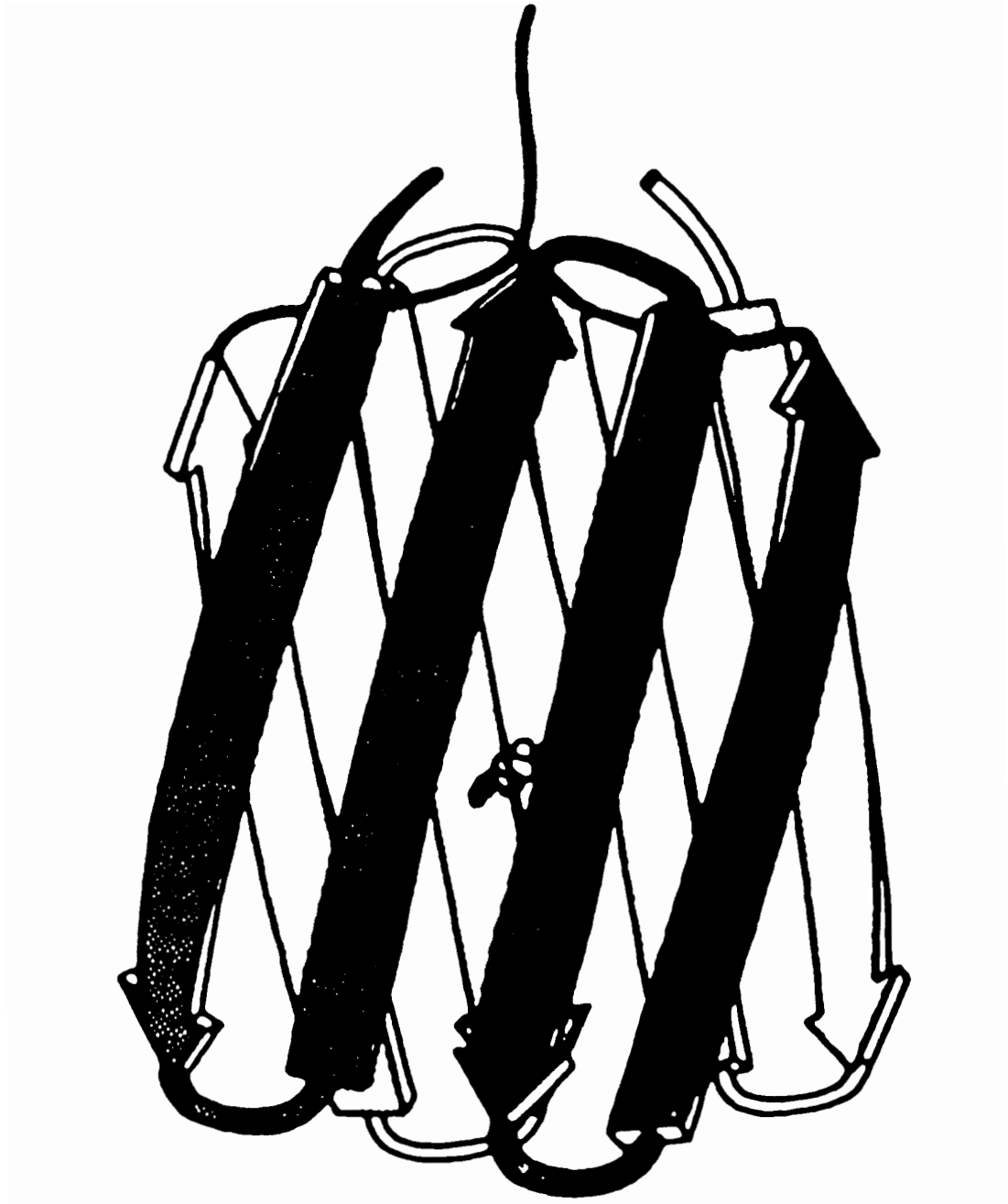


Fig. 3. Schematic ribbon drawing of betabellin[19], with the two-armed crosslinker in black. The two β-sheets have identical 32-residue sequences and are joined by a disulfide bond. In recent versions of betabellin, the crosslinker is omitted.

betabellin 9

HTLTASIPDLTYSIDPNTATCKVPDΦTLSIGB

A

```
position  1     5     10      15      20      25      32 positio:
    12    HTLTASIpdLTYSINpdTATCKVpdFTLSIGA    12
 common   |  |||||   ||      ||||  |   |  |    commo
    14    HSLTASIkaLTIHVQakTATCQVkaYTVHISE    14
 pattern  epnpnpnttnpnpnpnrrpnpnpnttnpnpnpne pattern
chirality LLLLLLLDDLLLLLLDDLLLLLLDDLLLLLLL chirali
```

B

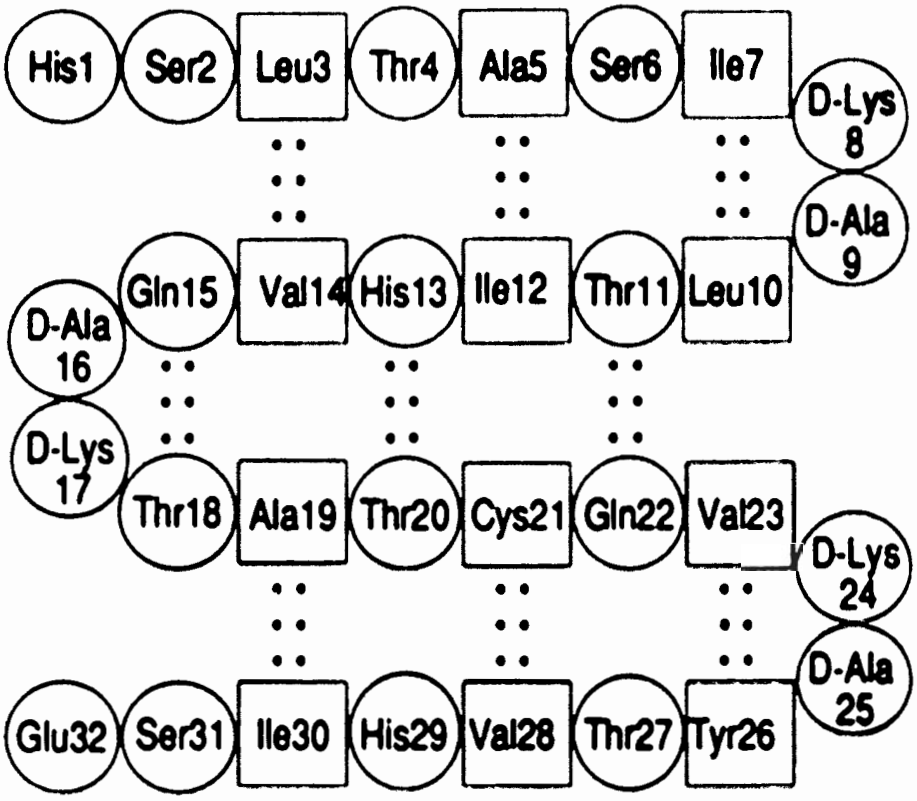


Fig. 1. Amino acid sequence of the betabellin-14 chain. **A:** Palindrom pattern of polar (p), nonpolar (n), end (e), and turn (t, r) residues. Ea half contains the same 14-residue palindromic pattern (underlined). Tl chiral pattern of L-amino acid (L) and D-amino acid (D) residues and rc idues common (|) to the betabellin-12 chain are indicated. **B:** Betab lin target structure with 18 β-sheet hydrogen bonds ($\cdots$) between 3 pai of polar residues (circled, side chains back) and 6 pairs of nonpolar rc idues (boxed, side chains forward).

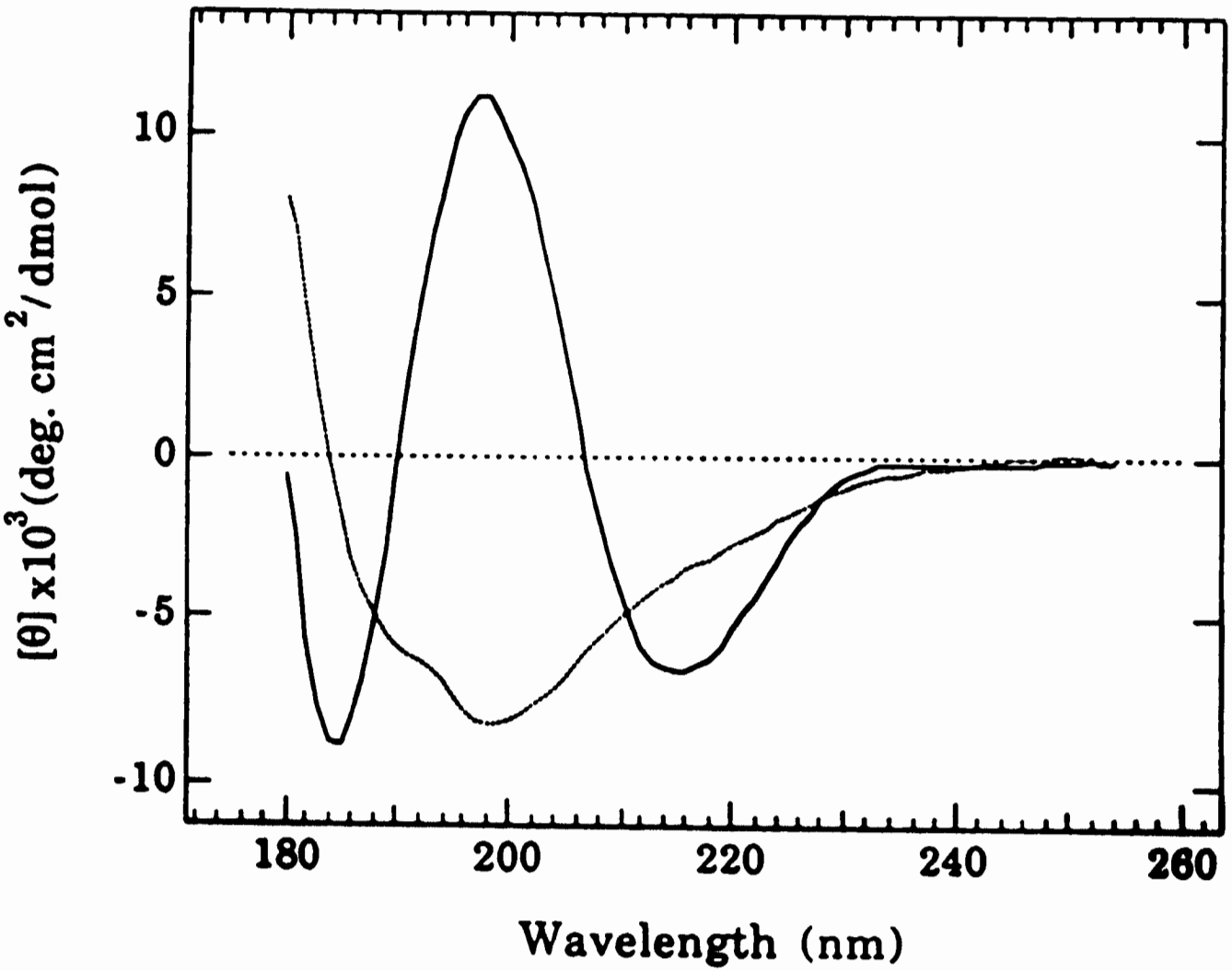


Fig. 2. CD spectra of betabellins 14S (------) and 14D (———) in water at pH 6.5 (Yan & Erickson, 1994).

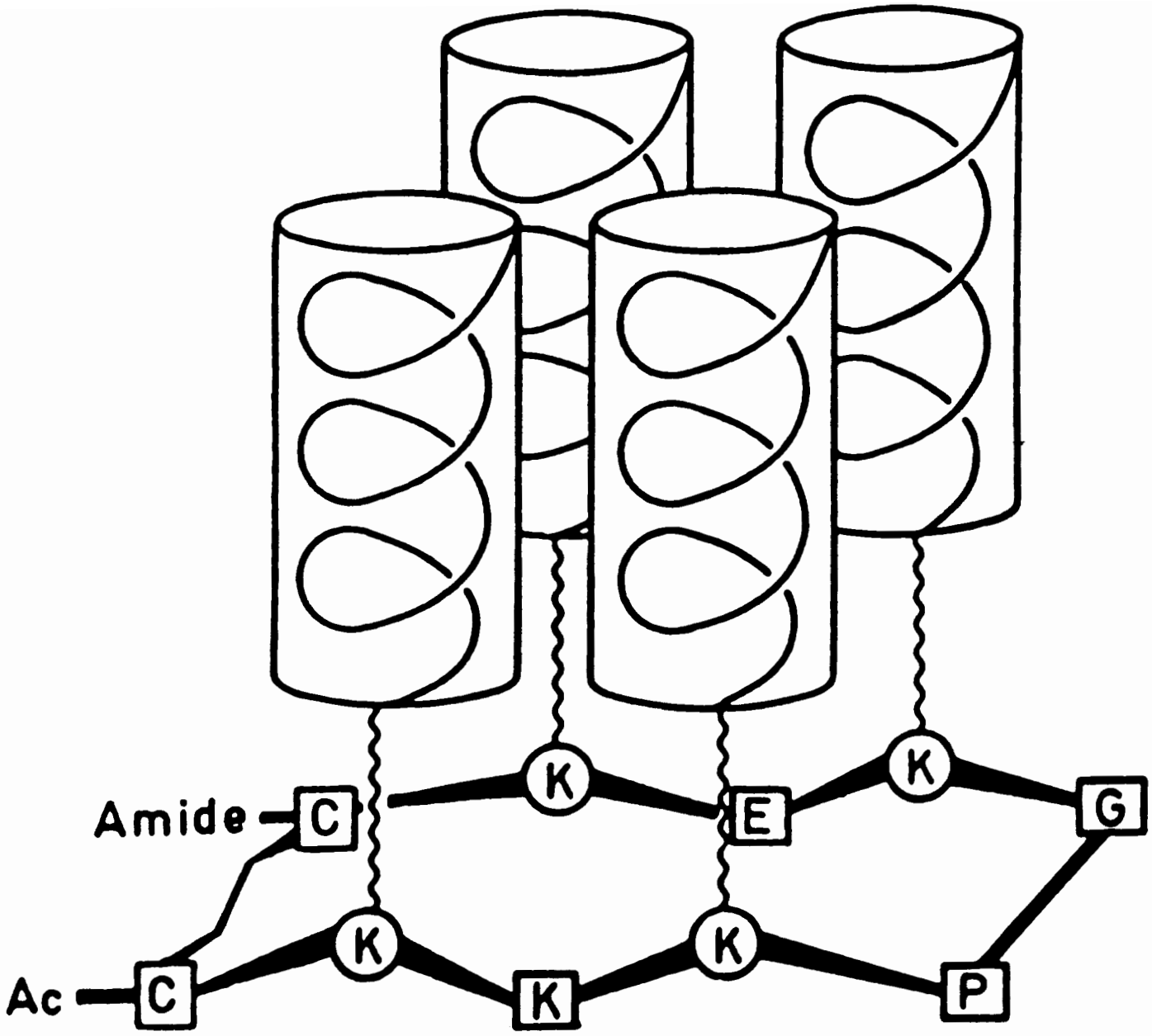


Fig. 8. Design of a new TASP molecule using a cyclic peptide (Ac–Cys–Lys–Ala–Lys–Pro–Gly–Lys–Ala–Lys–Cys–NH_2) as the template. The four helical blocks are identical and have the sequence X–Glu–Ala–Leu–Glu–Lys–Ala–Leu–Lys–Glu–Ala–Leu–Ala–Lys–Leu–Gly (X = H, TASP; X = Ac, TASP′; see text).

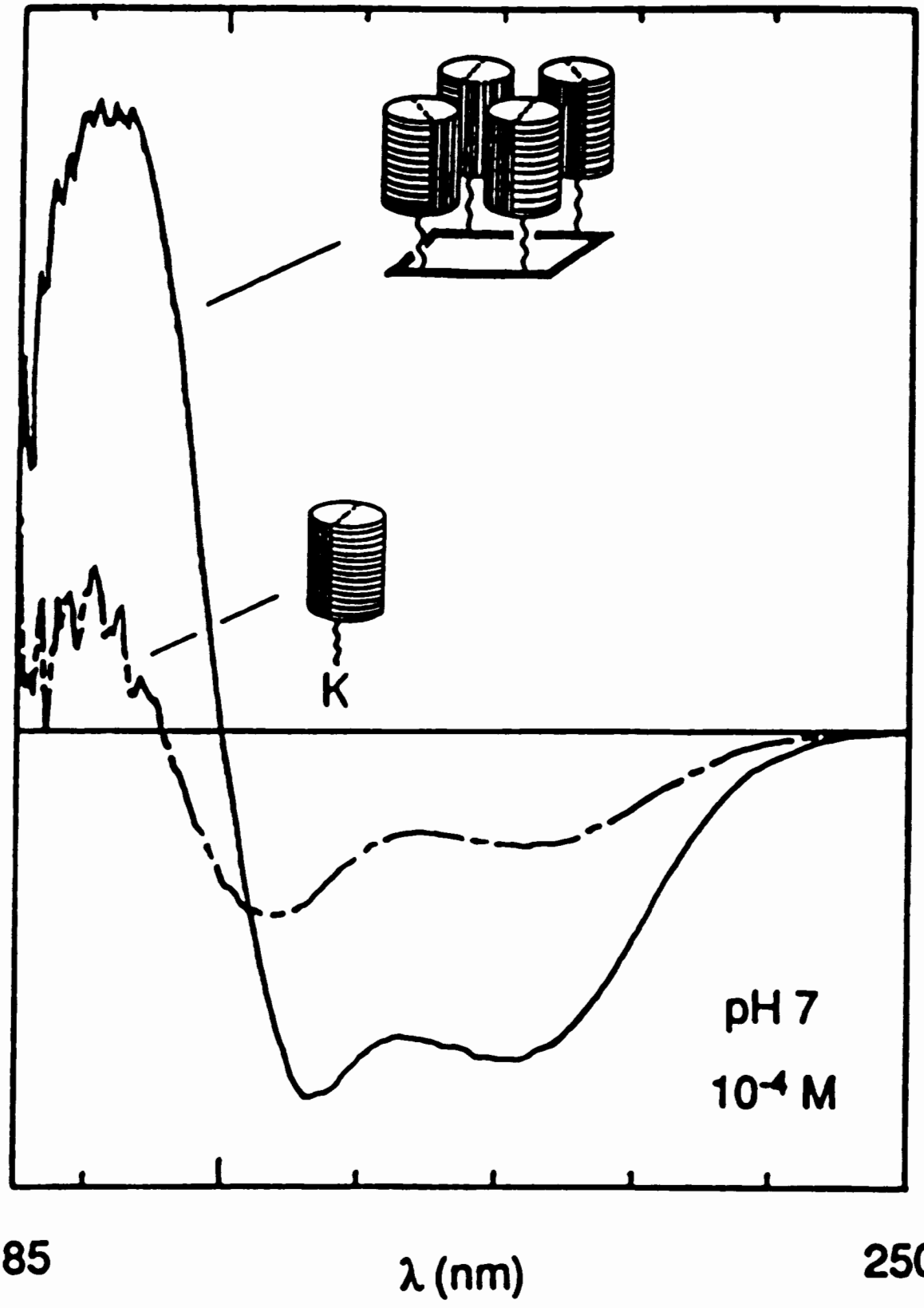


Fig. 9. CD spectra of TASP (X = H, *cf* Fig. 8) (———) and the corresponding helix block (attached to the ϵ-amino group of Ac–Lys–NH_2) (– – –) in water at neutral pH.

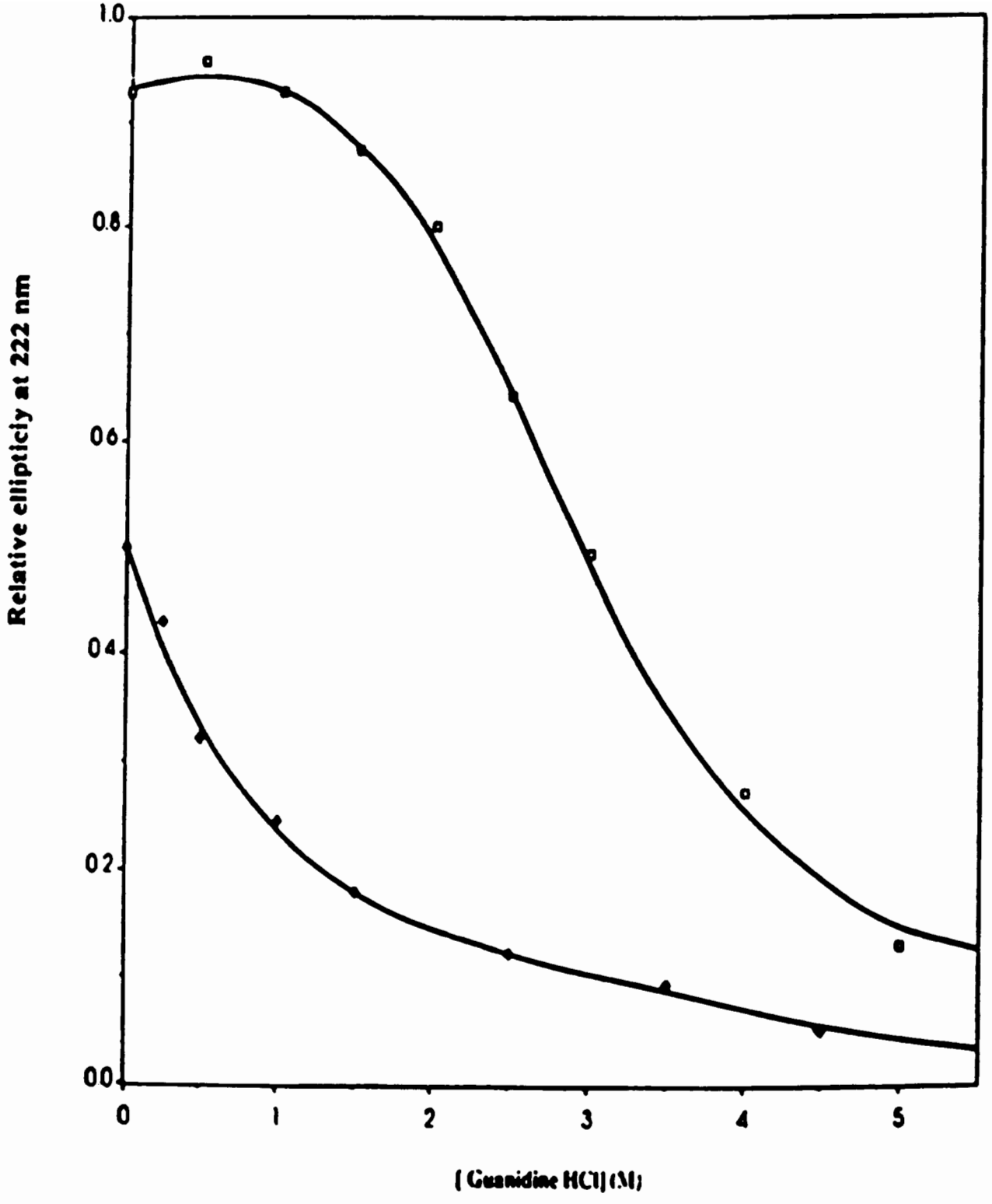


Figure 9. Denaturation of T_4-($4\alpha_{15}$-H) (□) (concentration = 1 mg of peptide/mL) by increasing concentrations of guanidine hydrochloride (guanidine HCl). The solutions were prepared in 50 mM phosphate and adjusted to pH 7 by addition of 4 N NaOH. The curve for the single helix block, Ac-Lys(α_{15}-H)-NH_2 (c = 1 mg of peptide/mL) is shown for comparison (◆).

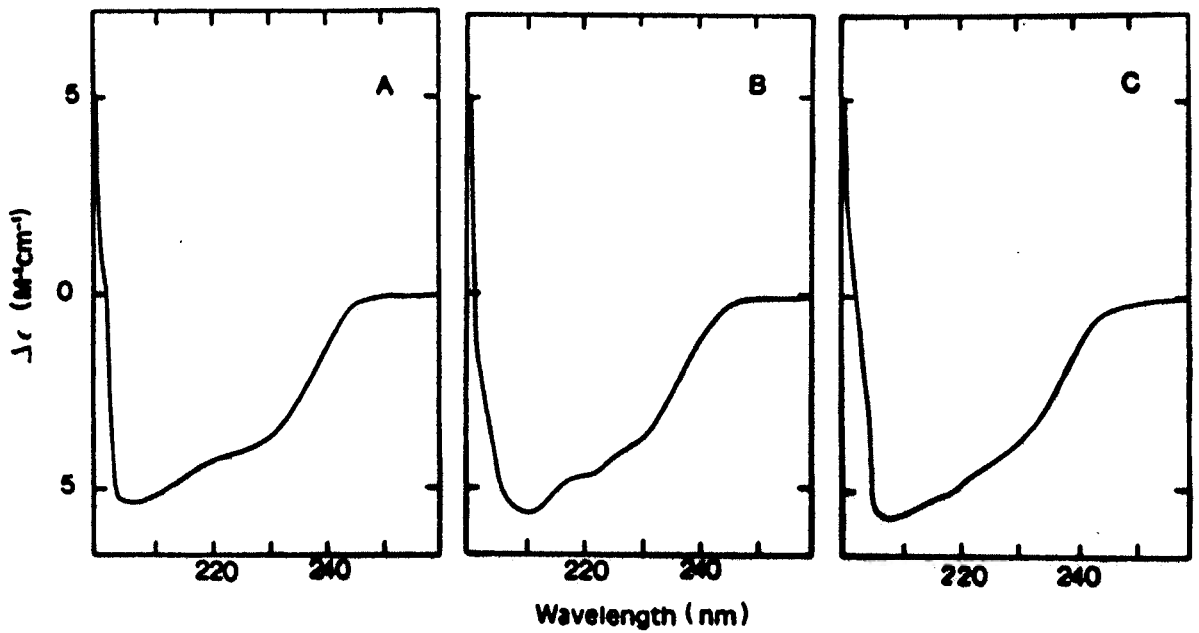


Fig. 9. CD spectra of octarellin, heptarellin and nonarellin in 60 mM ethanolamine; 10% (w/v) sucrose (pH 9.5). (A) Octarellin at an approximate concentration of 0.2×10^{-3} M. (B) heptarellin and (C) nonarellin at an approximate concentration of 0.43×10^{-3} M. The [illegible] were recorded in a cell with a 1-mm path length.

Figure 17.15 Schematic diagrams of the main-chain conformations of the second zinc finger domain of Zif 268 (red) and the designed peptide FSD-1 (blue). The zinc finger domain is stabilized by a zinc atom whereas FSD-1 is stabilized by hydrophobic interactions between the β strands and the α helix. (Adapted from B.I. Dahiyat and S.L. Mayo, *Science* 278: 82–87, 1997.)

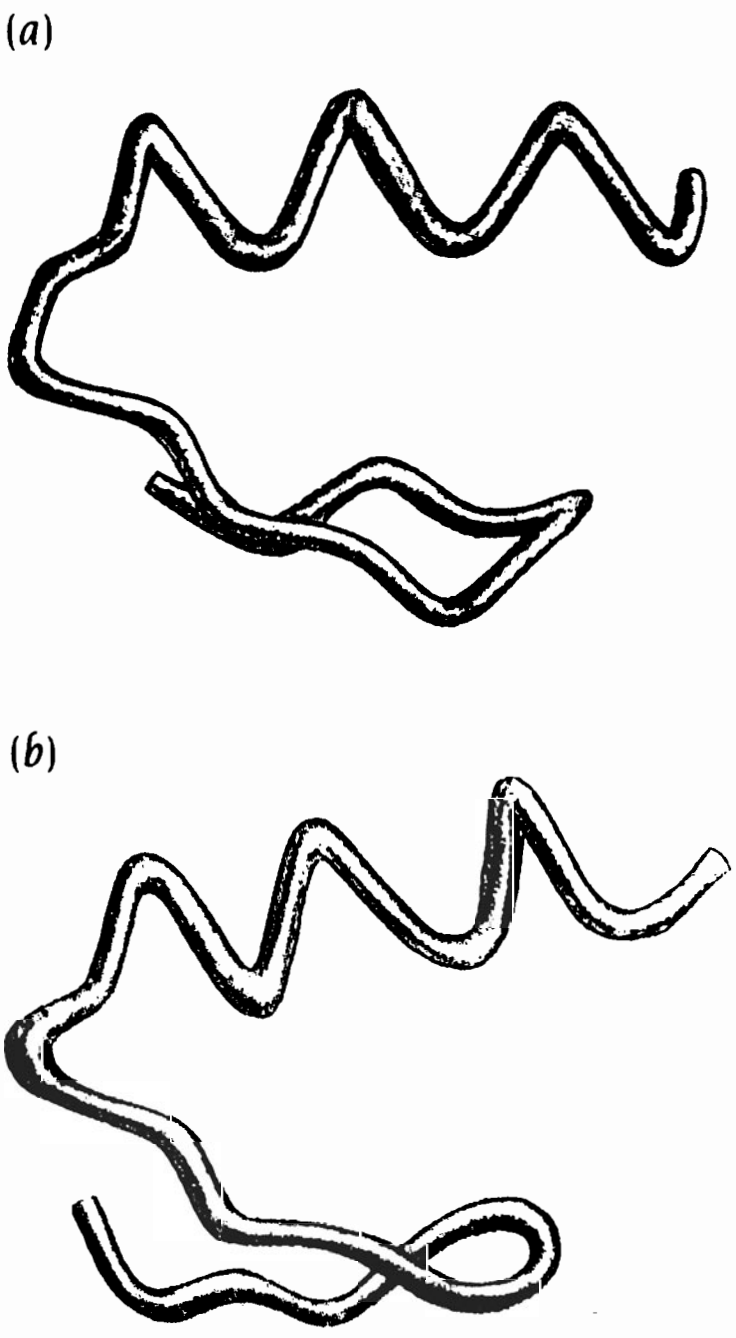


Table 17.2 Amino acid sequences of the second zinc finger of Zif 268 and the designed peptide FSD-1

	1										11										21									28
FSD-1	Q	-	Q	-	**Y**	-	T	-	A	-	K	-	**I**	-	K	-	G	-	R	-	T	-	**F**	-	R	-	N	-	E	-	K	-	E	-	**L**	-	R	-	D	-	**F**	-	**I**	-	E	-	K	-	F	-	K	-	G	-	R
Zif 268	K	-	P	-	F	-	Q	-	C	-	R	-	I	-	C	-	M	-	R	-	N	-	F	-	S	-	R	-	S	-	D	-	H	-	L	-	T	-	T	-	H	-	I	-	R	-	T	-	H	-	T	-	G	-	E

Residues in the hydrophobic core of FSD-1 are green.

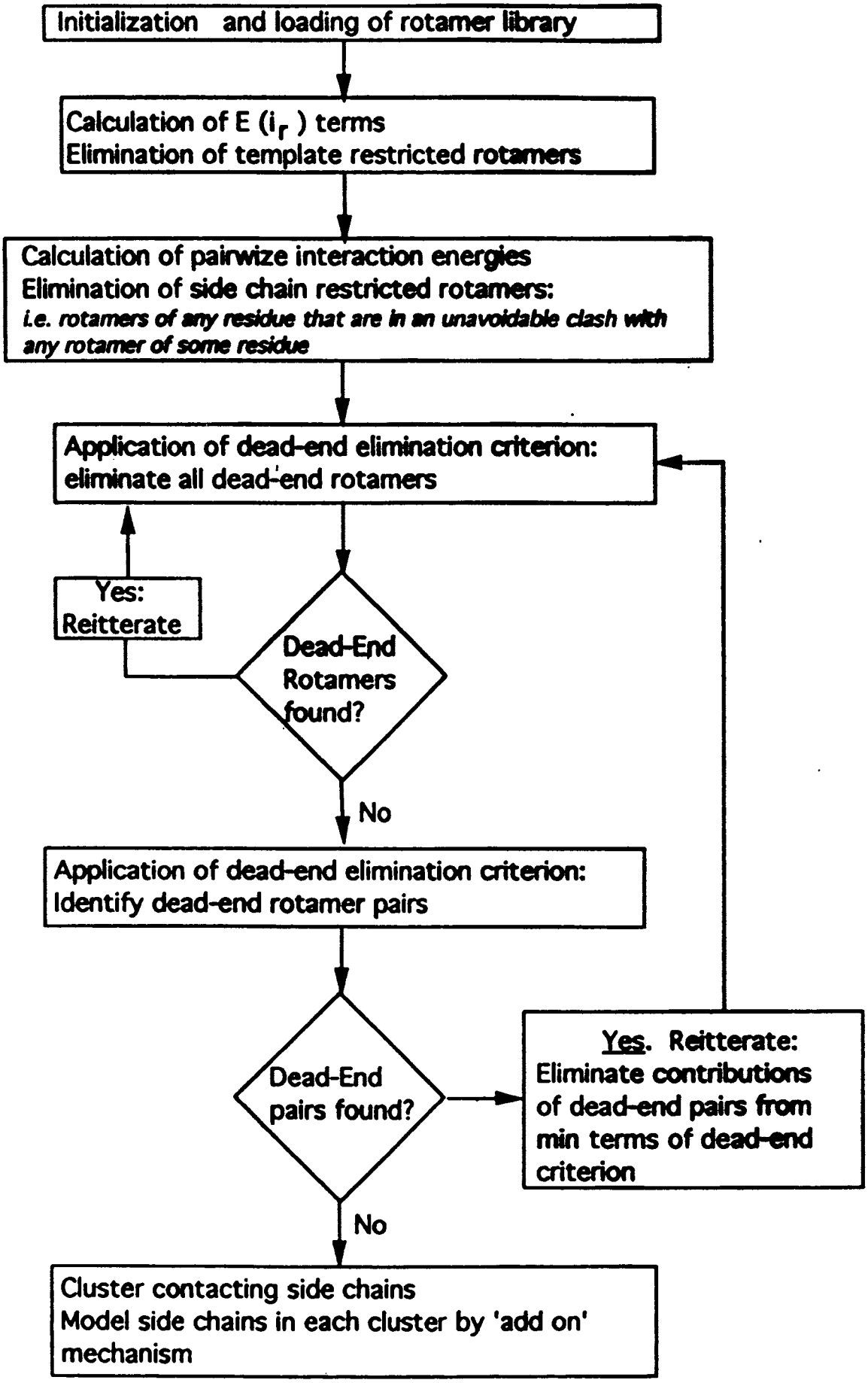
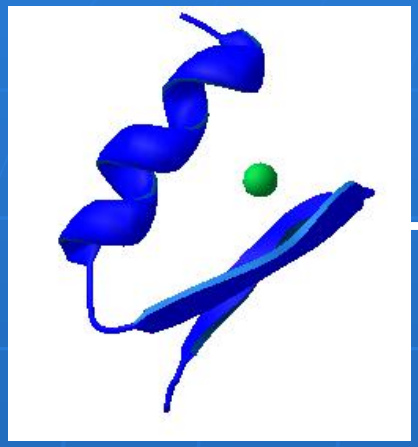


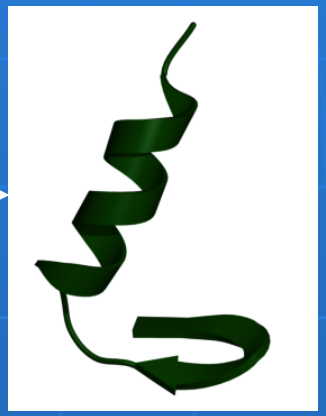
Fig. 2. General flow of our side chain placement algorithm

Zif268 Transformed to Heterodimer



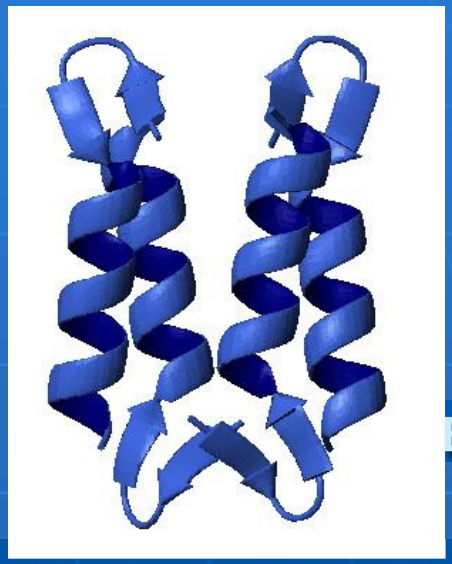
Zif268

Pavletich, et al. *Science*, 252:809 (1991).



BBA5

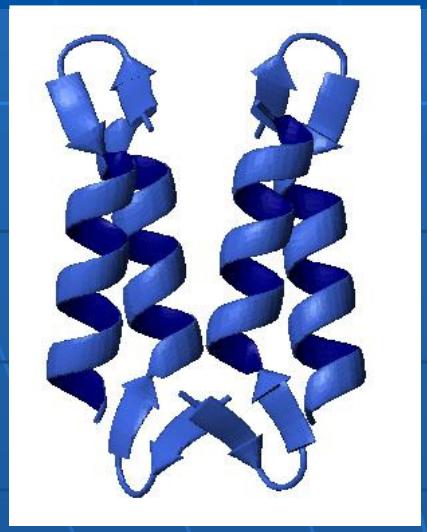
Struthers, et al. *Science*, 271:342 (1996); *Folding and Design*, 3:95(1998).



BBAT2

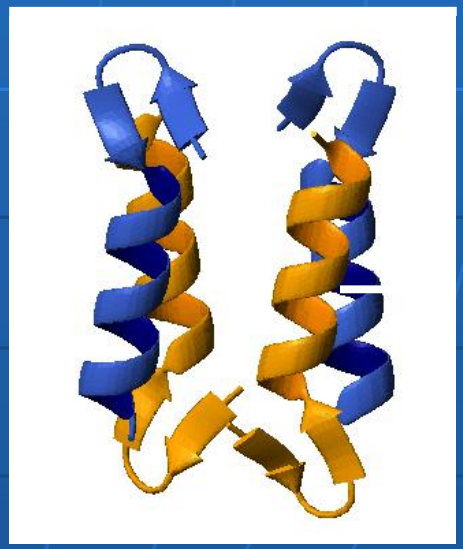
Ali, M, Peisach, E., Allen, K., and Imperiali, B, *PNAS*, 101:12183 (2004).

BBAT2



Homodimer to

Heterodimer,
Taylor and Keating
MIT



RMSD between predicted minimized design structure and crystal structure:

All Atom: 1.76Å

Backbone: 0.71Å

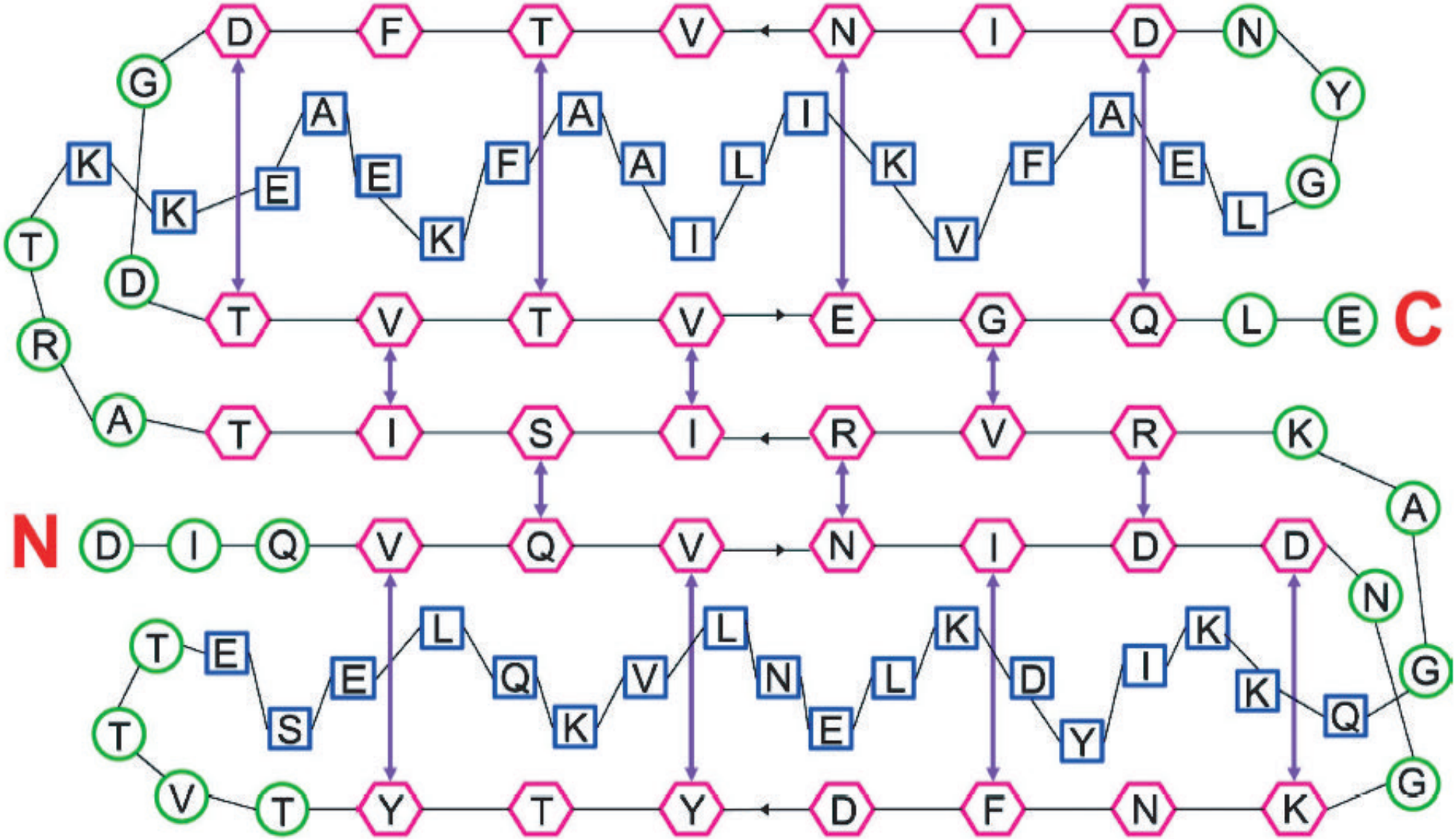


Fig. 1. A two-dimensional schematic of the target fold (hexagon, strand; square, helix; circle, other). Hydrogen bond partners are shown as purple arrows. The amino acids shown are those in the final designed (Top7) sequence.

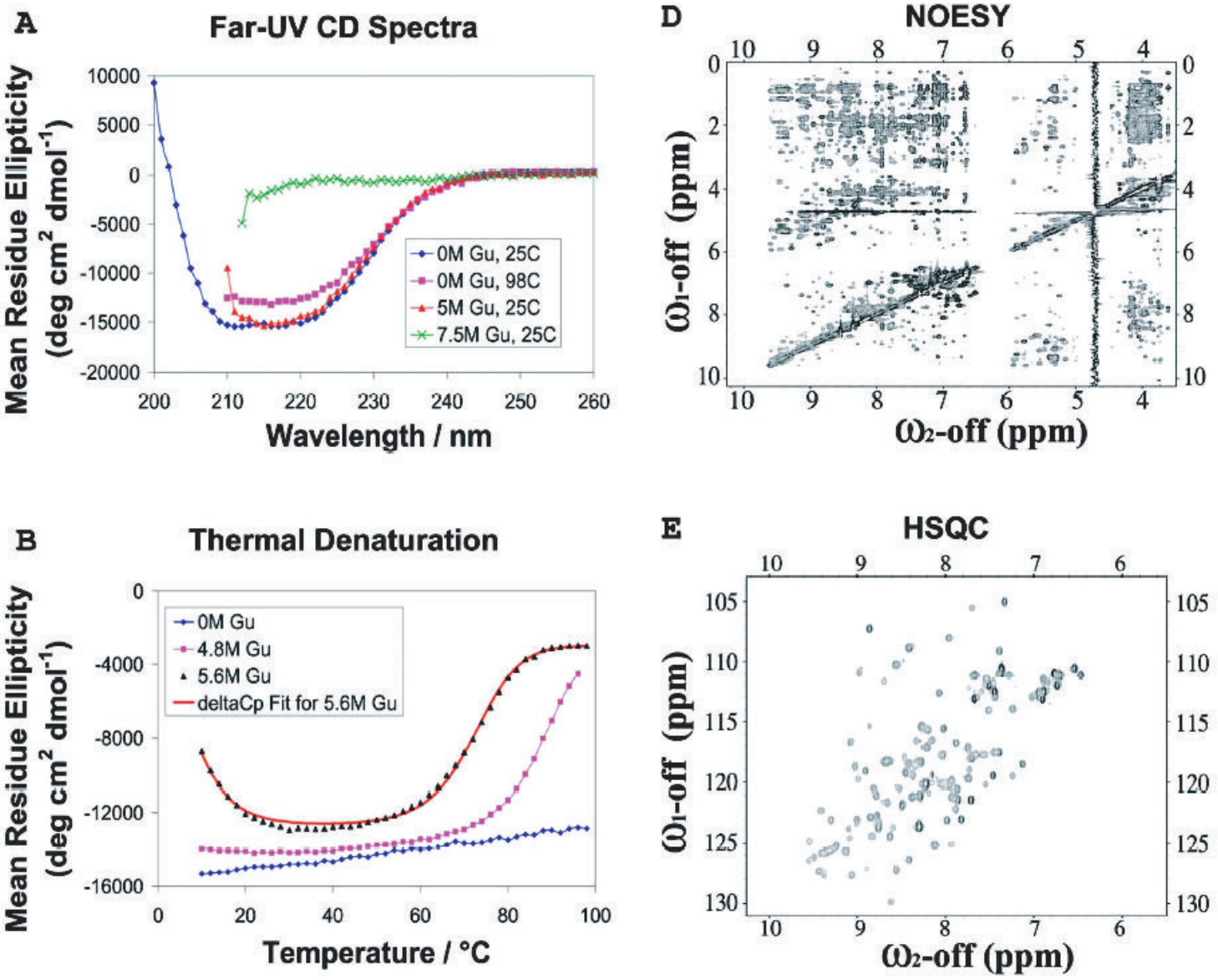
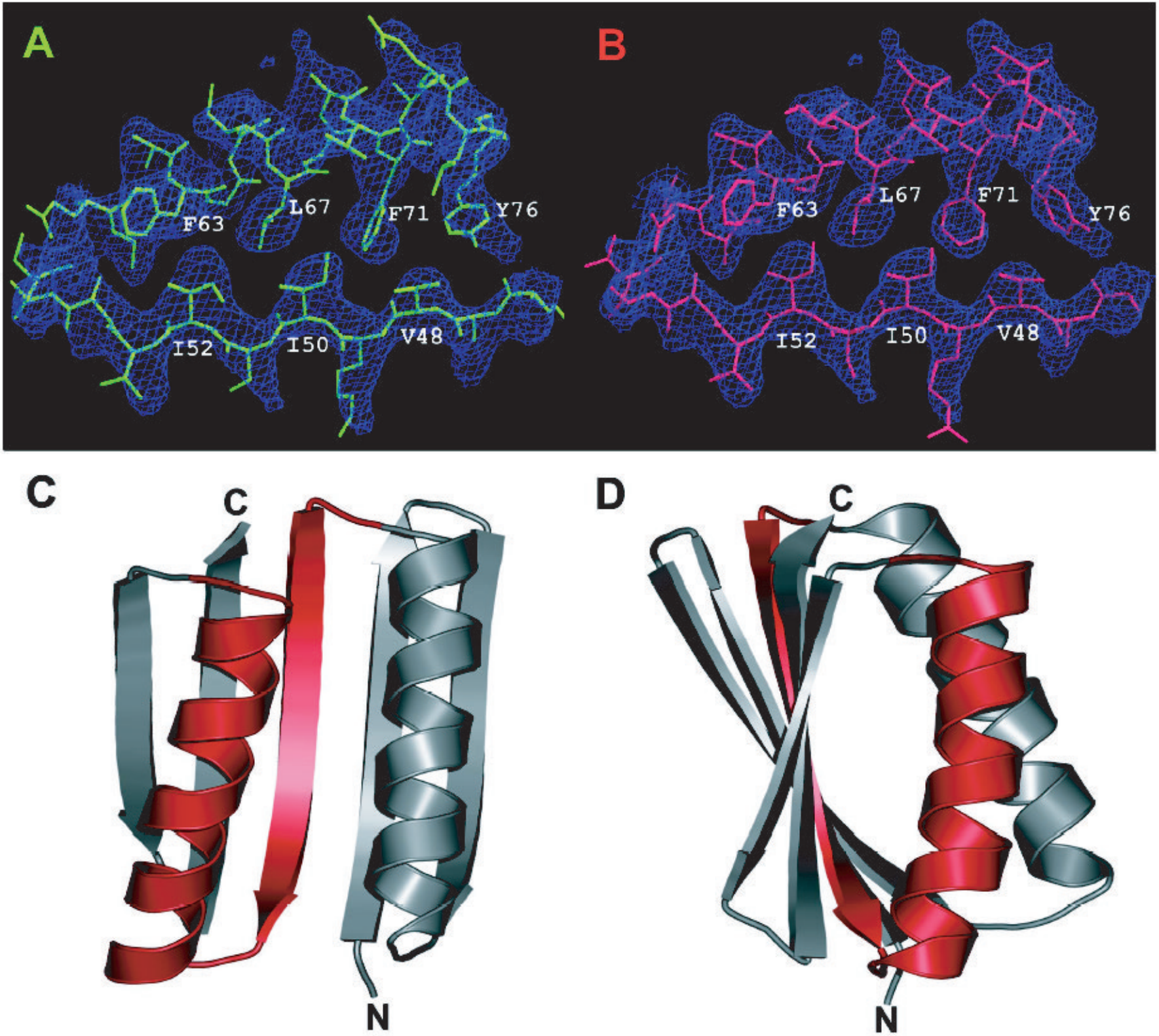


Fig. 2. Biophysical characterization of Top7. (**A**) The far-ultraviolet (UV) CD spectrum of 20 μM Top7 in 25 mM tris-HCl, 30 mM NaCl, and pH = 8.0 at varying temperatures and concentrations of GuHCl. (**B**) CD signal at 220 nm as a function of temperature and GuHCl for 8 μM TOP7 in 25 mM tris-HCl, 30 mM NaCl, pH = 8.0, in a 2-mm cuvette. (**C**) CD signal at 220 nm as a function of GuHCl concentration for 5 μM protein in 25 mM tris-HCl, 30 mM NaCl, pH = 8.0, at 25°C in a 1-cm cuvette. (**D**) The NOESY spectrum of ~1 mM Top7 at pH = 6.0 recorded at 298 K, 500 Mhz, and 200-ms mixing time with the use of Watergate suppression. ω, frequency. (**E**) The HSQC spectrum of ~1 mM 15N-Top7 at pH = 6.0 recorded at 298 K and 500 Mhz with the use of the fast HSQC scheme of Mori *et al.* (*43*).


A
F63
L67
F71
Y76
I52
I50
V48
B
F63
L67
F71
Y76
I52
I50
V48
C
C
N
D
C
N

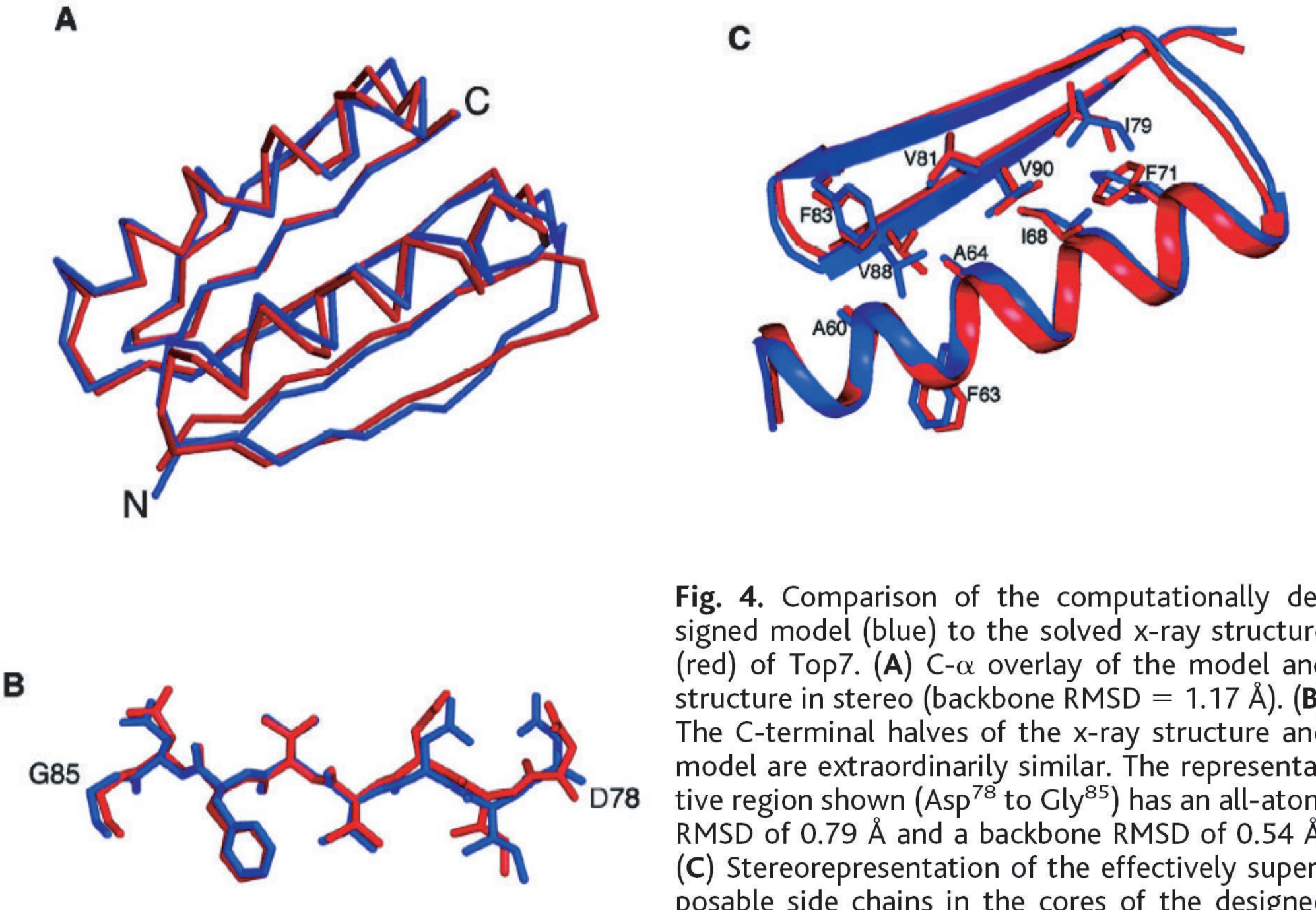


Fig. 4. Comparison of the computationally designed model (blue) to the solved x-ray structure (red) of Top7. (**A**) C-α overlay of the model and structure in stereo (backbone RMSD = 1.17 Å). (**B**) The C-terminal halves of the x-ray structure and model are extraordinarily similar. The representative region shown (Asp^{78} to Gly^{85}) has an all-atom RMSD of 0.79 Å and a backbone RMSD of 0.54 Å. (**C**) Stereorepresentation of the effectively superposable side chains in the cores of the designed model and the solved structure.