UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR.961 - 19

WORKSHOP ON:
PROTEINS, MEMBRANES and their INTERACTIONS

22 JULY - 2 AUGUST 1996

"Protein structure prediction using optimized energy functions"

PART II

Zaida LUTHEY-SCHULTEN
University of Illinois
School of Chemical Science
505 South Mathews Avenue
IL 61801 Urbana
U.S.A.

How to "Learn" an Energy Function

E. Look for Features

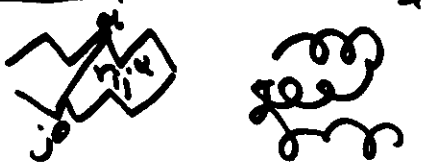
The "Associative Memory Energy"

$$E = \sum_{\text{pairs } \langle i,j \rangle} \sum_{\text{examples}} \gamma(\text{Target Sequence, Example}) \times \Theta\left(r_{ij} - r_{ij}^{(\alpha)}\right)$$

How much do sequence in target and example agree?

$\neq 0$ If distance matches example distance

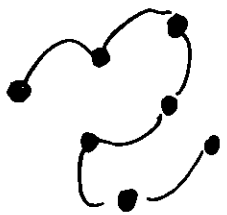
Examples



I. Assume simple physics

$$E = \sum_{\text{pairs}} \gamma(A_i, A_j)\, \Theta(r_{ij} - d)$$

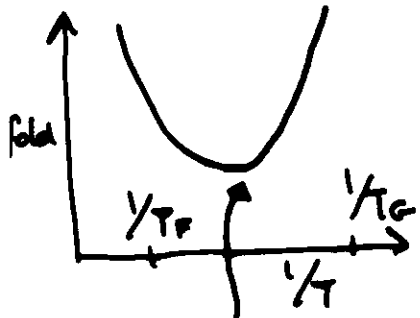
$\neq 0$ if pairs touch



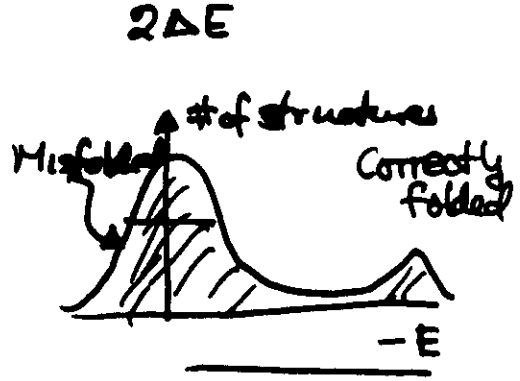
Use Minimum Frustration Principle to find γ's from known "worked" examples!

Optimization of Energy Functions

R. Goldstein
Z. Luthey-Schulten
P. Wolynes
PNAS 92



Best simulated annealing if T_F/T_G is a maximum



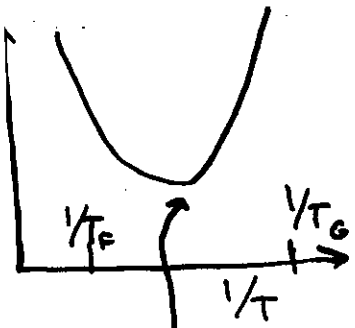
Best discrimination if $\delta E_S/\Delta E$ is a maximum

Same Variational Problem!

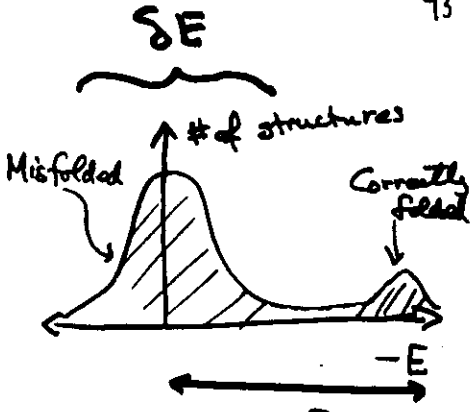
$$\text{Max } R = \frac{\delta E}{\Delta E} = \frac{(\underset{\sim}{A}\cdot\underset{\sim}{\gamma})}{\sqrt{\underset{\sim}{\gamma}\cdot\underset{=}{B}\cdot\underset{\sim}{\gamma}}} \Rightarrow \underset{\sim}{\gamma} = \underset{=}{B}^{-1}\cdot\underset{\sim}{A}$$

Energy Landscaping and the Minimal Frustration Principle

R. Goldstein
Z. Luthey-Schulten
PGW
PNAS '92
'92
'93



Best simulated annealing if T_F/T_G is a maximum.



Best discrimination if $E_F/\delta E$ is a maximum

Same Variational Problem!
within Random Energy Approximation

Minimize

$$r = \frac{E_F}{\delta E} = \frac{(\underset{\sim}{A}\cdot\underset{\sim}{\gamma})}{\sqrt{\underset{\sim}{\gamma}\cdot\underset{\sim}{B}\cdot\underset{\sim}{\gamma}}} \qquad \underset{\sim}{\gamma} = \underset{\sim}{B}^{-1}\cdot\underset{\sim}{A}$$

Heteropolymers and Protein Folding

What are the Problems?

1. Chemistry should tell us

$$E(\{q_i\}|\{r_i\})$$

Biology gives us $\{q_i\}$

Molecular Dynamics, simulated annealing $\{r_i\}$

2. Evolution's problem:
Given a $\{r_i\}$ find a $\{q_i\}$

3. Code breakers problem:
Given $\{r_i\}^{\mu}$ and $\{q_i\}^{\mu}$ find

$$E'(\{q_i\}|\{r_i\})$$

such that

$$E'(\{q_i\}^{new}|\{r_i\}) \Rightarrow \{r_i\}^{New}$$

Codebreakers need to understand both evolution and physics!

Protein Folding – What are the Problems?

1. Chemistry and physics should tell us

$E(\{q_i\}|\{r_i\})$

Biology gives us $\{q_i\}$

Molecular Dynamics simulated annealing

$\{r_i\}$

2. Evolution's problem: (Design)
Given a $\{r_i\}$ find a $\{q_i\}$

3. Code breaker's problem: $\{r_i\}^{\mu}$ ← database and $\{q_i\}^{\mu}$ find

$E'(\{q_i\}|\{r_i\})$

such that

$E'(\{q_i\}^{new}|\{r_i\}) \Rightarrow \{r_i\}^{New}$

Optimized Energy Functions

I. self-Consistently Optimized $\mathcal{H}_{local}$:

Threading Function

$\mathcal{H}_{local} = E_{profile} + E_{contact} + E_{hydrogen\ bonds\ (backbone\ \alpha,\beta)} + E_{gaps} + E_{exp\ constr}$

$E_{profile} = \sum_{i=1}^{N} \gamma^{P}(A_i, SS_i, SA_i)$

$E_{contact} = \sum_{i<j} \sum_{k=1}^{z} \gamma_k(A_i, A_j)\, u(r_k - r_{ij}) + E_{\beta_{bar}}$

r_{ij} — multiply counted

E_{gaps} ⋮

II Associative Memory Hamiltonian $\mathcal{H}_{AM}$

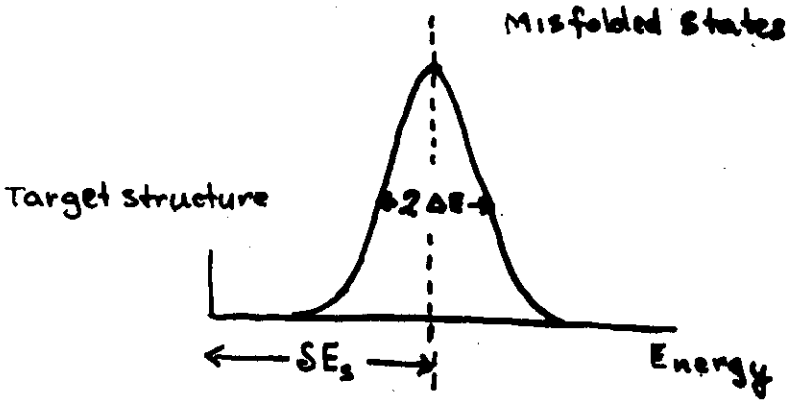
Molecular Dynamics

$\mathcal{H}_{AM}^{(r_{ij})} = -\sum_{\mu\ memories\ proteins} \sum_{i<j} \gamma_{ij}^{\mu}\, \theta_{gaussian}(r_{ij} - \hat{r}_{ij}^{\mu}) + \mathcal{H}_0$ (backbone rigidity)

pattern recognition

Ref. Friedrichs and Wolynes, Intn. Computer Method. (1990)
Goldstein, Luthey-Schulten, Wolynes PNAS (1992)

Optimization of γ



Target Configuration: $E_T = \Sigma_i \lambda_i^T \gamma_i$
Misfolded State k : $E_k = \Sigma_i \lambda_{ki} \gamma_i$

$$\delta E_s = \mathbf{A}\gamma$$
$$\Delta E^2 = \gamma \mathbf{B} \gamma$$

where

$$A_i = \lambda_i^T - \langle \lambda_{ki} \rangle$$
$$B_{ij} = \langle \lambda_{ki}\lambda_{kj} \rangle - \langle \lambda_{ki} \rangle \langle \lambda_{kj} \rangle$$

Optimal γ :$\Longrightarrow$ maximise $R = \frac{\delta E_s}{\Delta E}$ with respect to γ:

$$\gamma = \mathbf{B}^{-1}\mathbf{A} \qquad (1)$$

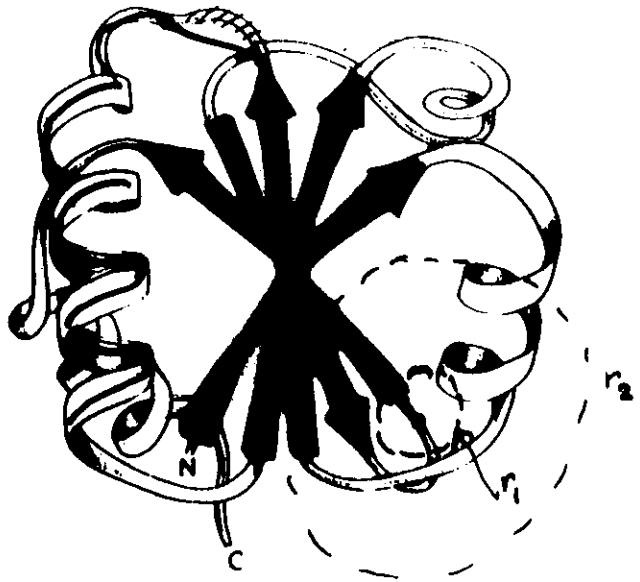
Mean-Field Sequence-structure Alignment

$$E_{local} = E_{profile} + E_{contact} + E_{\substack{hydrogen \\ bonds}} + E_{gaps} + E_{\substack{exp. \\ constraints}}$$

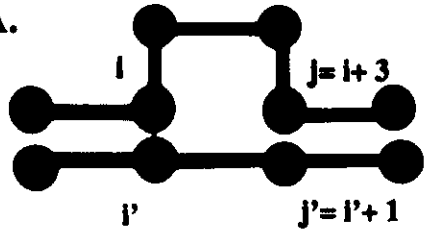
$$E_{profile} = \sum_{i=1}^{N} \gamma^P[A_i, \text{secondary structure}(i), \text{surf. acces}(i)]$$

$$E_{contact} = \sum_{i<j} \sum_{k=1}^{2} \gamma_k(A_i, A_j)\, u(r_k - r_{ij}) + E_{\substack{multiple \\ bonds}}$$

(Ref. Finkelstein & Reva, Skolnick et al, Thornton, et. al)



A.

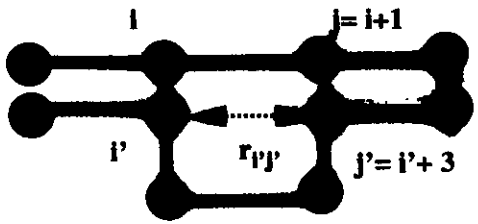


target

scaffold

$$E_g(A) = \gamma_1 + \gamma_2(j-1)$$

B.

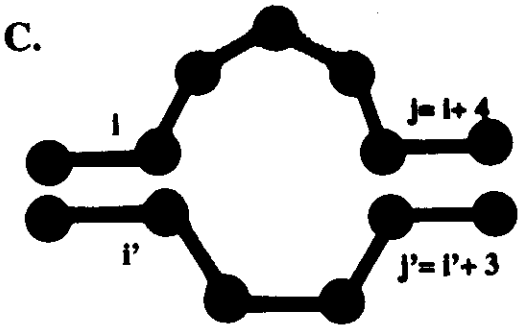


target

scaffold

$$E_g(B) = \gamma_3 + \gamma_4 r_{i',j'} + \gamma_5 r^2_{i',j'}$$

C.



target

scaffold

$$E_g(B) = \gamma_6 + \gamma_7(j-1) + \gamma_8 r^2_{i',j'}/(j-1)$$

R. Goldstein, E. Schulten, P. Wolynes in "New Developments in Theoretical Studies of Proteins" (World, 19[illegible]

Statistical Mechanical Basis of Sequence - Structure Alignment Algorithms

Conventional: (Smith-Waterman Algorithm)

$A = (a_1, a_2 \ldots\ldots a_m)$ 1-D information

$B = (b_1, b_2 \ldots\ldots b_n)$

2 Metrics: ① $S(A,B) = \{s(a_i, b_j) \ldots\ldots\}_{m \times n}$

$s(a_i, b_j)$: Dayhoff Evol. Scale

(indels) penalty ② $\omega = \alpha + \beta \log k$ ← length of indel

↑ non-uniform (Levitt, Chothia)

SM: Sequence - Structure Alignment (Wolynes et. al.)

$F = E + \sum n_\ell \mu_\ell$ (free-energy)

where μ_ℓ is a gap penalty $\mu_\ell(\ell = \underline{i-j}, r^{ij}_{i'j'})$

↑ gap length ↑ distance

$$\mu_\ell = \frac{\Gamma_r^2}{\Delta E_{rc}} \log \left\{ \frac{P_{r_\ell} P_{c_0}^{\lambda_\ell + 1}}{P_{r_0}^{\lambda_\ell + 1} P_{c_\ell}} \right\}$$

$P_{c_\ell}, P_{r_\ell} \ldots$ = Prob. that gap of type ℓ appears in correct { random alignments.

① Hawaii Conference, ② R. Elber, editor "New Developments ..."

Smith-Waterman (1981)

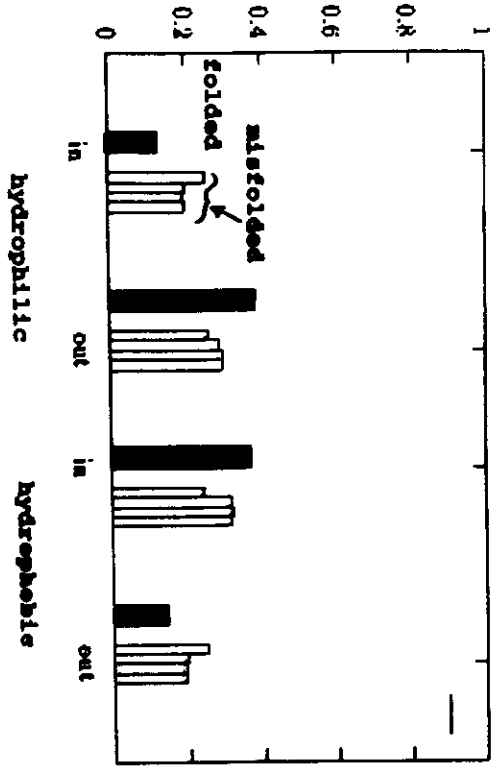
Local & global alignments based on similarity score H where

$$H_{ij} = \max \begin{cases} H_{i-1\,j-1} + \Delta(a(i), b(j)), \\ \max\limits_{k \geq 1} H_{i,j-k} - W_k, \\ \max\limits_{l \geq 1} H_{i-l,j} - W_l, \; 0 \end{cases}$$

Discrimination Scores

Training Proteins

PDB	NRES	Name	Zeroth	First	Second	$\frac{\delta E_n}{\Delta E_n}$
		α proteins				
2cro	63	434 Cro Protein	3.70	3.60	4.40	1.19
1wrp(R)	102	TRP Repressor (Trigonal Form)	2.41	3.80	4.44	1.84
1ccr	111	Cytochrome C	4.43	4.62	5.02	1.13
1hmq(A)	113	Hemerythrin (MET)	3.80	4.22	3.48	0.92
1bp2	123	Phospholipase	6.35	6.35	7.63	1.20
2hhb(A)	141	Hemoglobin (Deoxy)	2.37	2.90	3.36	1.42
3cln	143	Calmodulin	5.37	5.95	6.99	1.30
1fdh(G)	146	Hemoglobin (Deoxy, Human Fetal)	3.31	3.63	4.30	1.33
5mbn	153	Sperm Whale Myoglobin (Deoxy)	3.79	4.62	5.49	1.45
		β proteins				
5pcy	99	Plastocyanin	5.35	6.28	6.82	1.28
1rei(A)	107	Bence-*Jones Immunoglobulin	3.93	4.15	3.59	0.91
2pas	123	Pseudoazurin (Cupredoxin)	6.55	8.69	8.85	1.35
2il1b	153	Interleukin-1*β	6.14	7.62	9.46	1.54
1f19(H)	215	R19.9 (IG*G2B=K=) Fab Fragment	2.14	2.79	2.55	1.20
3hfm(H)	215	IG*G1 Fab Fragment	4.71	5.40	6.62	1.41
2plv(3)	235	Poliovirus (TYPE 1, Mahoney Strain)	2.58	3.79	5.16	2.00
1r1a(2)	236	Rhinovirus Serotype 1 (HRV1) Coat Protein	2.53	3.18	3.91	1.55
1cms	263	Chymosin B	3.36	4.19	5.56	1.65
		$\alpha+\beta$ or α/β proteins				
1fdx	54	Ferredoxin	1.97	2.53	2.21	1.12
1alc	122	α-*Lactalbumin	6.63	7.51	7.78	1.17
1rbb(A)	124	Ribonuclease B	5.68	6.49	6.91	1.22
1snc	136	Staphylococcal Nuclease	6.31	8.15	9.61	1.52
2dhf(A)	162	Dihydrofolate Reductase	3.36	4.28	5.01	1.49
2act	218	Actinidin (Sulfhydryl Proteinase)	5.06	6.59	7.90	1.50
2prk	279	Proteinase K	3.76	6.33	6.89	1.83
1pfk(A)	319	Phosphofructokinase (R-State)	3.85	5.56	7.10	1.84
2ldx	331	Apo-Lactate Dehydrogenase	2.92	3.92	5.03	1.72
3gpd(G)	334	D-Glyceraldehyde-3-Phosphate Dehydrogenase	3.44	4.72	7.24	2.10
2liv	344	Leucine/Isoleucine/Valine-Binding Protein	5.34	6.90	8.83	1.65
		mean				1.44



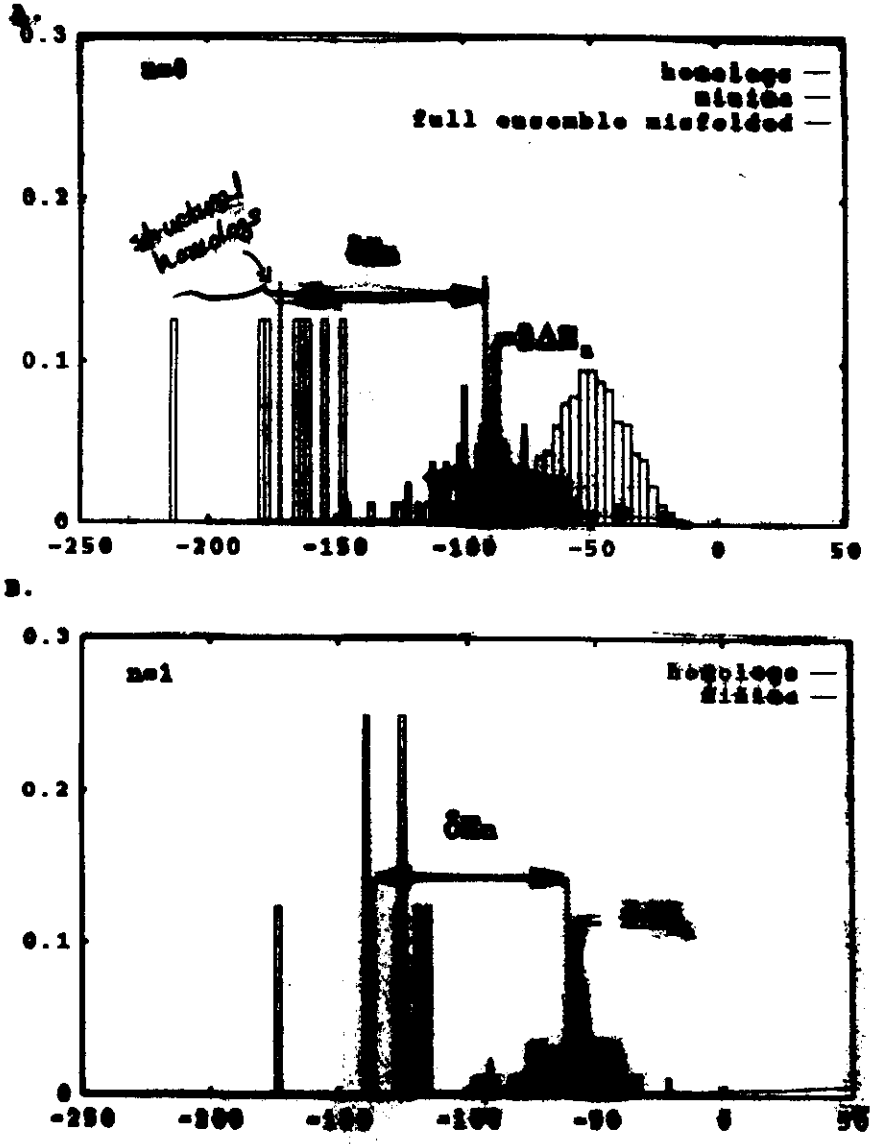
Partial Ordering

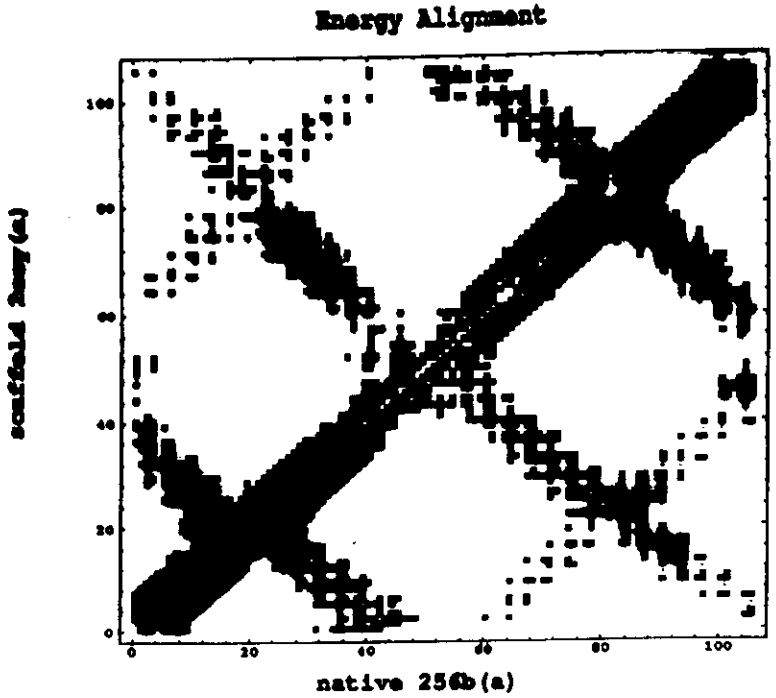
Comparision of Energetic and Genetic Alignments

PDB	Self-comparison Target	q	PDB	Predicted Structure Analog	RMS	RMS FW	q	q FW	q FRW	% Ident.
	α-proteins									
2MK(A)	Cytochrome MMF (Oxidized)	1.00	3cyt(A)	Cytochrome C'	1.68	11.08	0.41	0.08	0.40	27.3
5cab	Calcium binding protein	1.00	4tnc	Troponin C	2.30	3.97	0.82	0.51	0.51	20.3
1lhv	Crab Hemoglobin I	1.00	2lhb	Hemoglobin V	2.74	4.39	0.64	0.48	0.48	22.5
2lh4	Leghemoglobin (Deoxy)	1.00	5mbn	Sperm Whale Myoglobin	2.82	4.01	0.87	0.55	0.59	18.3
1mbn	Sea Hare Myoglobin (MET)	1.00	5mbn	Sperm Whale Myoglobin	2.35	5.39	0.68	0.48	0.48	25.0
1gta	Glutathione S-transferase	1.00	1gta	Glutathione S-transferase	4.20	3.67	0.48	0.65	0.65	21.6
1cI0	434 Repressor (Amino-Terminal)	1.00	1hcd(C)	λ Repressor-Operator	3.37	2.07	0.85	0.61	0.61	28.6
	β-proteins									
1cnj	Azurin	1.00	4pcy	Plastocyanin	3.30	4.36	0.35	0.40	0.42	25.5
2fn	Rous-Sarcoma Protein	1.00	2fdn(B)	IG*G1 Fab Fragment	4.66	4.66	0.46	0.43	0.47	27.2
1cd3	CD3 (T-cell CD-complex)	1.00	2fb4(B)	Immunoglobin Fab Fragment	5.09	9.36	0.40	0.30	0.36	21.3
2fdn(L)	IG*G1 Fab Fragment	0.94	2fdn(B)	IG*G1 Fab Fragment	7.04	7.09	0.31	0.30	0.32	26.2
2lb	Intestinal Fatty Acid binding protein	1.00	1crb	Cellular retinol binding protein	2.02	2.09	0.64	0.63	0.67	20.53
	α+β and α/β proteins									
1fc1	Flavodoxin	1.00	3fxn	Flavodoxin	2.32	2.69	0.51	0.63	0.65	30.4
3dfr	Dihydrofolate Reductase	1.00	8dfr	Dihydrofolate Reductase	4.35	3.30	0.51	0.56	0.60	20.6
6ldh	Apo-Lactate Dehydrogenase	0.98	4mdh(A)	Malate Dehydrogenase	3.98	9.39	0.47	0.36	0.40	20.2
2gbp	Galactose/Glucose Binding Protein	0.98	1abp	Arabinose-Binding Protein	6.94	5.38	0.53	0.47	0.47	23.5

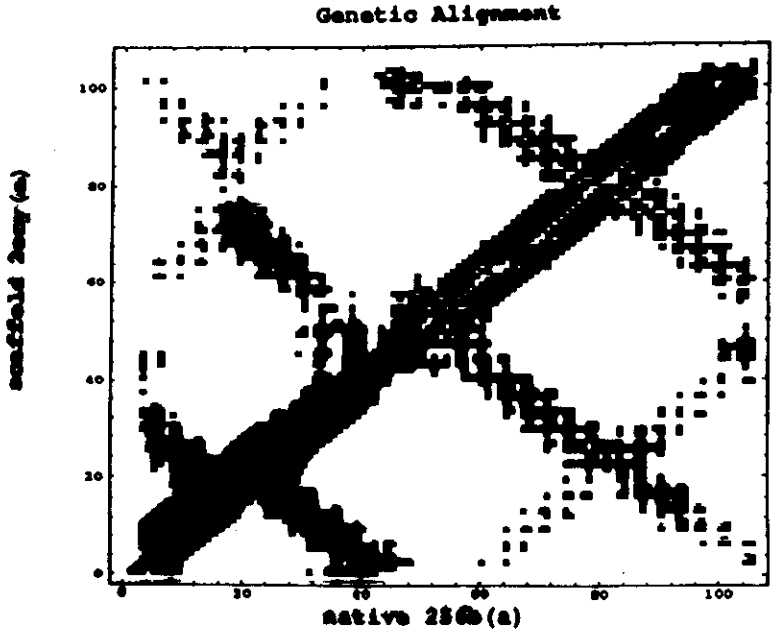
$$q = \frac{1}{N(N-1)} \sum_{i<j} \theta\left(r_{c_a}^{x\text{-ray}}(i,j) - r_{c_a}^{align}(i,j)\right)$$

Self - Consistent Optimization





95% Sec. Str.

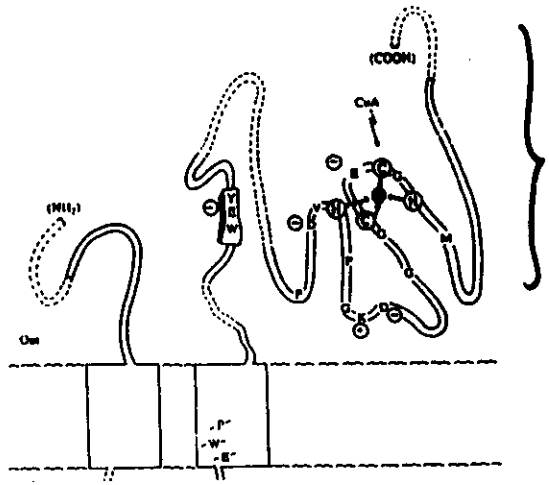


65% Sec. Str.

Partial Unknowns

Prosite Signatures

	Target Name		Predicted Homolog	
	Name	PDB	Name	% Ident.
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	Bovin β-2 μ-globuline	[illegible]	IG*G1 Fab Fragment	20.2
B0MG_PONPY	Orangutan β-2 μ-globuline	3fdm(H)	IG*G1 Fab Fragment	19.6
B0MG_RAT	Rat β-2 μ-globuline	3fdm(H)	IG*G1 Fab Fragment	18.6
HA2B_MOUSE	Mouse H-2 Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	16.9
HA2P_HUMAN	Human H-2 Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	26.3
HA2Q_MOUSE	Mouse H-2 Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	19.3
HA2R_MOUSE	Mouse H-2 Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	18.5
HA2S_HUMAN	Human HLA Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	19.4
HB2S_HUMAN	Human HLA Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	20.0
HB2B_HUMAN	Human HLA Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	21.3
HB2G_HUMAN	Human HLA Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	19.7
HB2S_HUMAN	Human HLA Histocomp. Antigen	3hla(A)	Human Class I Histocompatibility	26.3
KACA_RAT	Rat IG κ Chain	3hfm(H)	IG*G1 Fab Fragment	23.6
KACB_RAT	Rat IG κ Chain	3hfm(H)	IG*G1 Fab Fragment	21.7
KAC_MOUSE	Mouse IG λ Chain	3hfm(H)	IG*G1 Fab Fragment	18.6
LAC1_MOUSE	Mouse IG λ Chain	3hla(A)	Human Class I Histocompatibility	24.3
LAC3_MOUSE	Mouse IG λ Chain	3hfm(H)	IG*G1 Fab Fragment	27.3
LAC6_MUSSP	Mouse IG λ Chain	3hla(A)	Human Class I Histocompatibility	24.3
LAC_HUMAN	Human IG λ Chain	3hla(A)	Human Class I Histocompatibility	23.3
avidoxin signature	[illegible]			
FLAV_ANASP	A. sp. Flavodoxin	1fx1	Flavodoxin	25.5
FLAV_AZOVI	A. vinelandii Flavodoxin	1fx1	Flavodoxin	26.2
FLAV_CHOCR	C. crispus Flavodoxin	1fx1	Flavodoxin	29.3
FLAV_CLOAB	C. acetobutylicum Flavodoxin	1fx1	Flavodoxin	23.1
FLAV_DESVH	D. vulgaris Flavodoxin	3fxn	Flavodoxin	20.4
FLAV_KLEPN	K. pneumoniae Flavodoxin	1fx1	Flavodoxin	23.1
FLAV_RHOCA	R. capsulatus Flavodoxin	1fg1	Flavodoxin	26.3
DHF signature	[illegible]			
DYRE_ECOLI	E. coli DHFR	4dfr(B)	Dihydrofolate Reductase	29.7
DYRC_ECOLI	E. coli DHFR	4dfr(B)	Dihydrofolate Reductase	24.2
DYR_BPT4	Bacteriophage T4 DHFR	4dfr(B)	Dihydrofolate Reductase	25.8
DYR_ENTFC	E. faecium DHFR	4dfr(B)	Dihydrofolate Reductase	23.3
DYR_LACCA	L. casei DHFR	4dfr(B)	Dihydrofolate Reductase	27.7
ioredoxin signature	[illegible]			
THIL_YEAST	S. cerevisiae Thioredoxin I	2trx(A)	E. coli Thioredoxin	22.7
THIO_CHICK	Chick Thioredoxin	2trx(A)	E. coli Thioredoxin	24.8
THIO_HUMAN	Human Thioredoxin	2trx(A)	E. coli Thioredoxin	22.1
THIO_MOUSE	Mouse Thioredoxin	2trx(A)	E. coli Thioredoxin	22.1
THIO_RABIT	Rabit Thioredoxin	2trx(A)	E. coli Thioredoxin	22.1
THIO_RAT	Rat Thioredoxin	2trx(A)	E. coli Thioredoxin	22.1



Cu A Domain

Cytochrome c oxidase

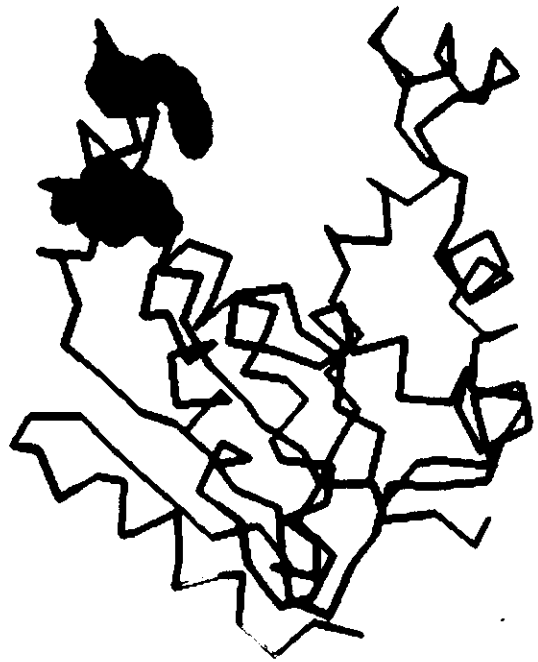
Redox Reaction: $O_2 \longrightarrow H_2O$

!fold Secondary structure assignment

```
      131
AV:    tttBBBBBtt  tHHHHHHHtt  ttttttttt  tttBBBtttt  BBBBBBBttt
/:     ---BBBBBBB  TTBBBBBBT-  -----CBBT-  ---BBBBCT  TBBBBBBBBT
E:     ---BBBBBBB  BCCCTTCCC-  -----CBBT-  ---BBBBCT  TBBBBBBBTT
E:     ---BBBBBBB  BTTTTTT---  -----BBBBB  BBBBBBCTTC  BBBBBBBB--
      181
       ttBBBBBttt  tBBBBBtttt  tBBBBBtttt  BBBBBBBBBt  ttttttttBB
       CBBBBBB-CC  CCTTCCCCTT  BBBBBBBBCT  CBBBBBBBB-  TT---TTTBB
       TTCCBBBCTT  TTCCCBBBCT  TCBBBBBBCT  CBBBBBBB-T  TT---TTTBB
       TTCCBBBTTC  TCBBBBBCTT  CBBBBB-CCT  CBBBBBBC-T  TT---TTTCC
      231
       BBBBBBHHHH  HHHHHHHHHH  H
       BBBBBB----  ----------  -     (75% secondary structure)
       BBBBCTTHHH  HHHTCCCHHH  H     (69% secondary structure)
       BBBBBBTTTH  HHHHHHHHCH  H     (73% secondary structure)
```

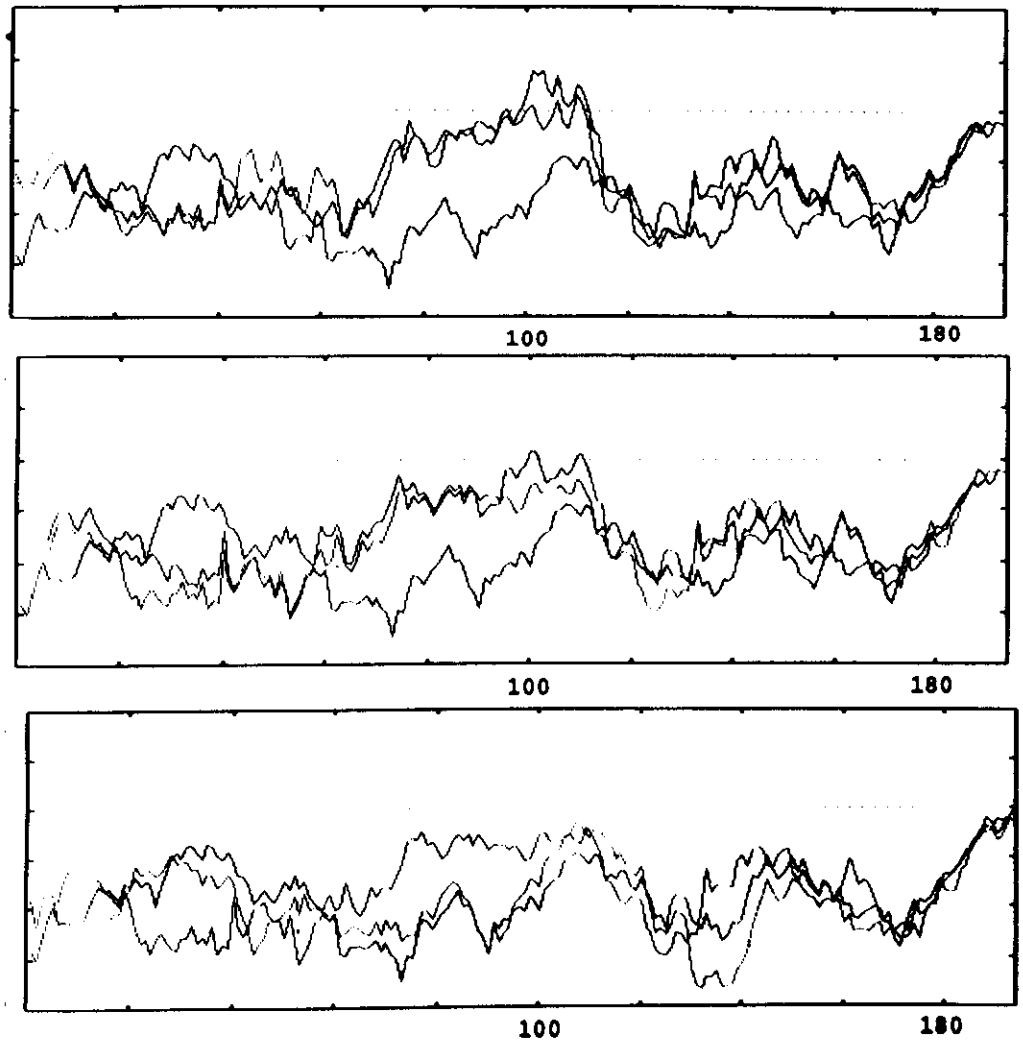
Threading as good as the scaffold!

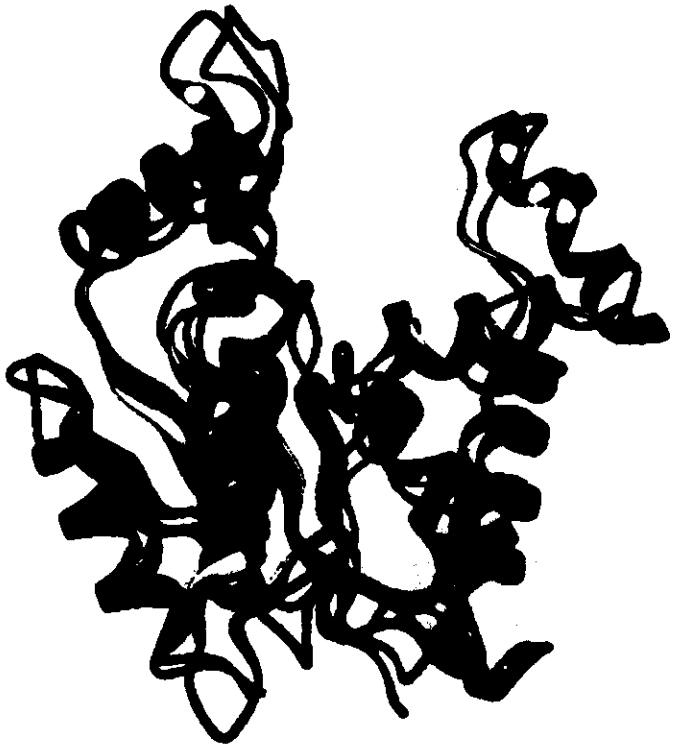
Methanococcal adenylate kinase



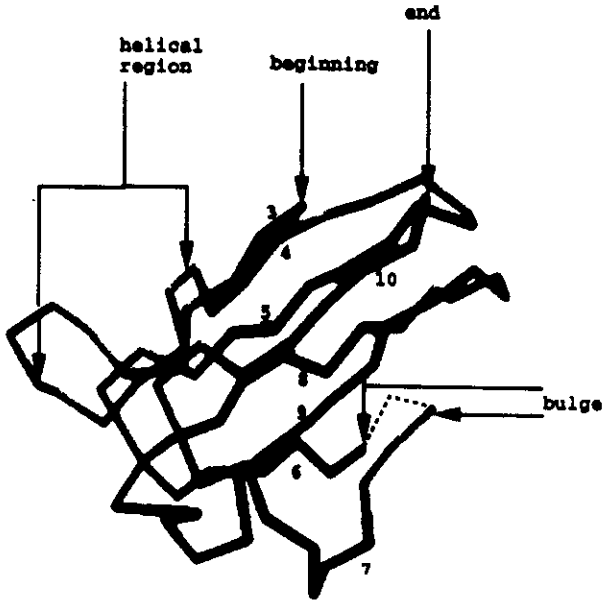
```
     MKNKV VVIVGVPGVG STTVTNKAIE ELKKEGIEYK IVNFGTVMFE
MEEKLKKSKI IFVVGGPGSG KGTQCEKIVQ K-----YGYT HLSTGDLLRA

IAK-----EE GLVEHRDQLR KLPPEEQKRI QKLAGKKIAE MAKEFNIVVD
EVSSGSARQK MLSEIMEKGQ LVP--LETVL DMLRDAMVAK VDTSKGFLID

PHSTIKTPKG YLPGLPAWVL EELNPDIIVL VEAE---NDE ILMRRLKDET
---GYPREVK QGEEPERKIG QPT---LLLY VDAGPETMTK RLLKRGETSG

RQRDFESTED IGEHIFMNRC AAMTYAVLTG ATVKIIKNRD F--LLKDAVQ
RVDDNEETIK KRLETYYKAT EPTIAFYEKR GIVRKVNAEG SVDDVFSQVC

ELIEVLK
PHLDTLK
```

Koretke, Haney, Konisky, BS, PW

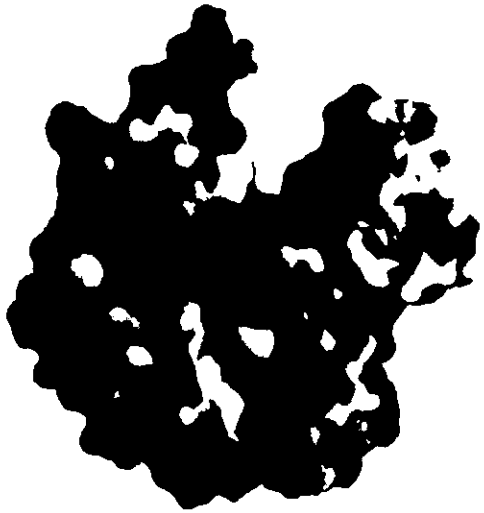

100
180
100
180
100
180



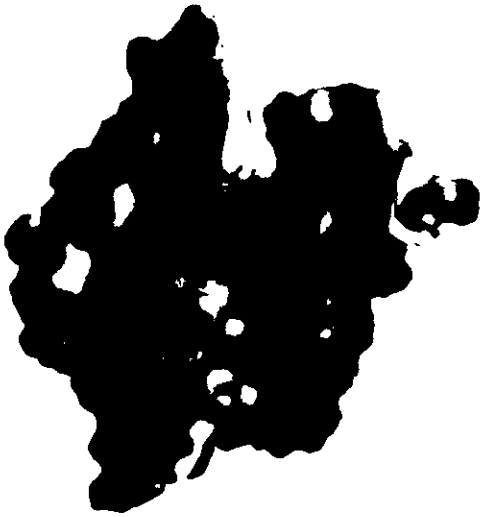
Prediction of Copper A Structure



A.



B.

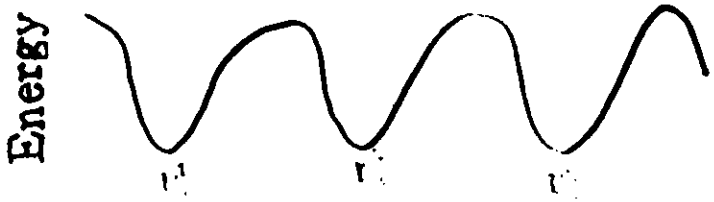


1D - Associative Memories Hamiltonian

Multiple Memories $\{\mu\}$

$$\mathcal{H}_{AM}(\mathbf{r}_{ij}) = -\gamma \sum_{\mu} \sum_{<ij>} \theta(r_{ij} - r^{\mu}_{ij}) + \mathcal{H}_o$$

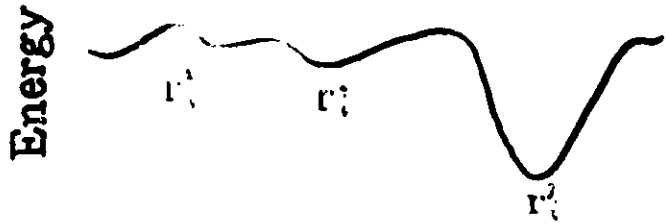
pre-aligned; gaussian; Backbone



Including correlations between target and memory sequences

$$\mathcal{H}_{AM}(\mathbf{r}_{ij}) = -\sum_{\mu} \sum_{<ij>} \gamma^{\mu}_{ij}\, \theta(r_{ij} - r^{\mu}_{ij}) + \mathcal{H}_o$$

$$\gamma^{\mu}_{ij} = \gamma(S^T_i, S^T_j, S^{\mu}_i, S^{\mu}_j)$$



Associative Memory Hamiltonian

Multiple Memories $\{\mu\}$

$$\mathcal{H}_{AM}(r_{ij}) = -\sum_{\mu} \sum_{<ij>} \theta(r^{c_\alpha,c_\beta}_{ij} - r^{c_\alpha,c_\beta}_{ij,\mu}) + \mathcal{H}_o$$

gaussian; memories pre-aligned

Including correlations between target and memory sequences

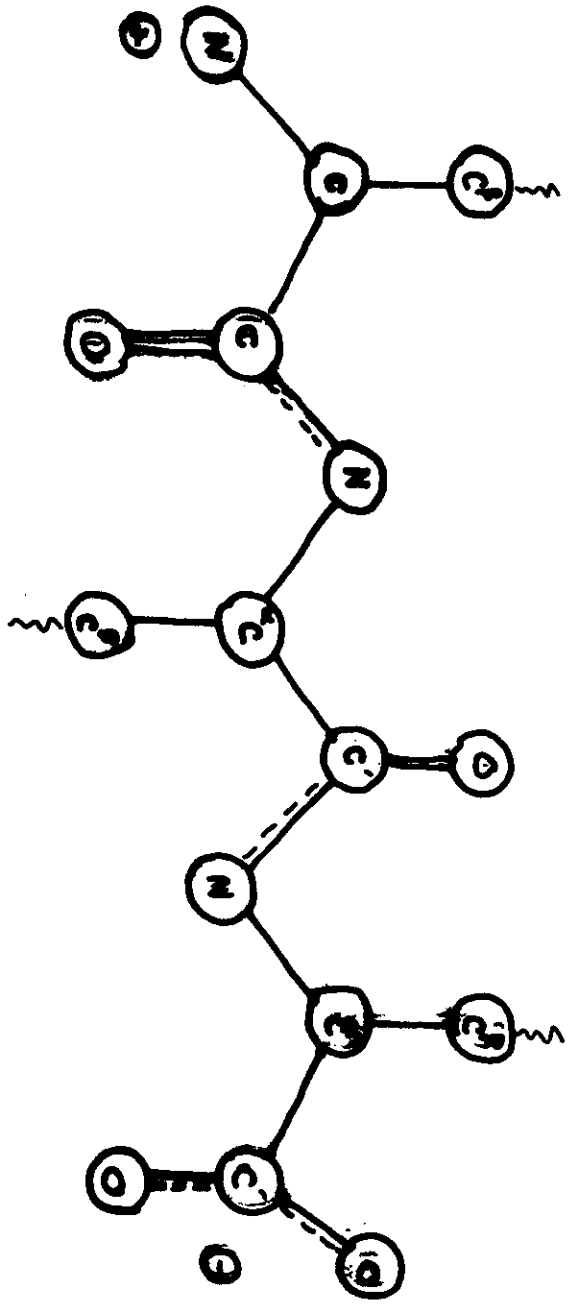
$$\mathcal{H}_{AM}(r_{ij}) = -\sum_{\mu} \sum_{<ij>} \gamma^{\mu}_{ij}\, \theta(r^{c_\alpha,c_\beta}_{ij} - r^{c_\alpha,c_\beta}_{ij,\mu}) + \mathcal{H}_o$$

gaussian

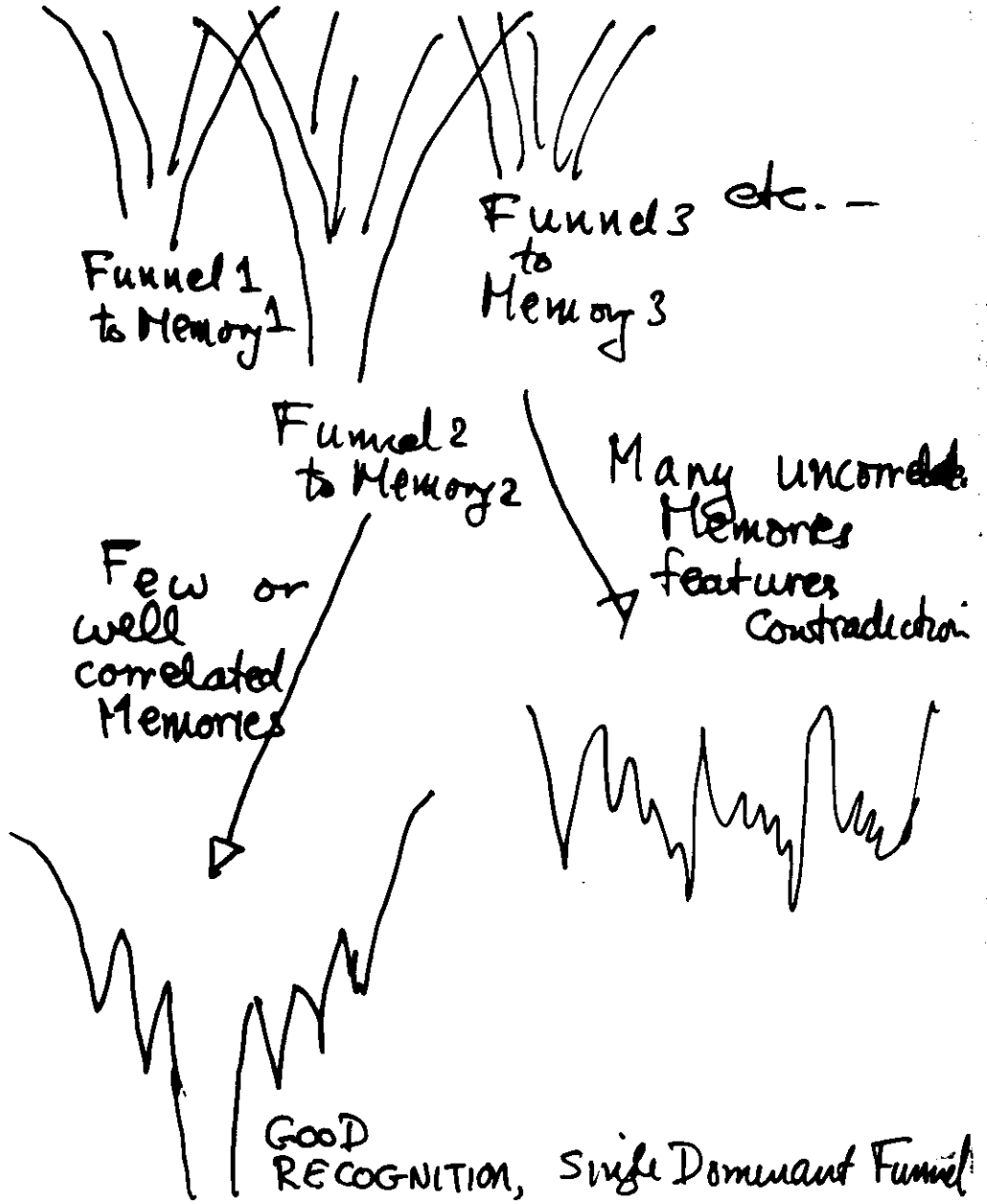
γ^{μ}_{ij} can encode:

- Hydrophobicity
- Probability of mutation
- Predicted, Actual, or Instantaneous secondary structure
- Location of the residue in the protein
- Properties of nearby residues

48 gammas: hydrophobicity × 3 categories ($j-i<5$), ($j-i>5$, $n_{ij}<9$)



Energy Landscape of AM Energy Function



Full $\mathcal{H}_{AMH}$ and Protein Backbone Geometry

$$\mathcal{H}(\{r_i^{C_\alpha\beta}\}|\{s_i^\tau\}) = \sum_{\mu \text{ memories}} \sum_{i<j} \sum_{k,l=\alpha,\beta} \gamma^{\mu}_{s_i s_j} \theta(r^{C_\alpha\beta}_{k_i l_j} - r^{C_\alpha\beta\,\mu}_{k_i l_j})$$

H_0 (reduced representation of backbone)

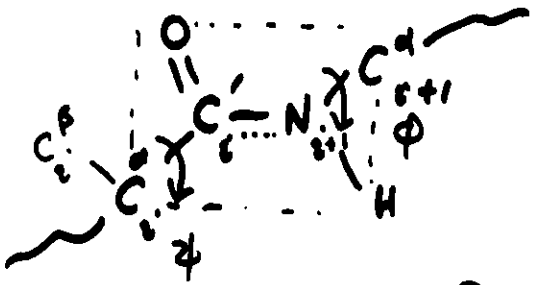
$$+ \mathcal{H}_{ev} : \Delta_{ev} \sum_{i<j} \begin{cases} (r_{ij} - 3A^0)^2 & r_{ij} < 3 \\ 0 & r_{ij} > 3 \end{cases}$$

$$+ \mathcal{H}_{chirality} : T\left[\vec{r}_{C_i' C_i^\alpha} \times \vec{r}_{C_i^\alpha N_i}\right] \cdot \vec{r}_{C_i' C_i^\beta}$$

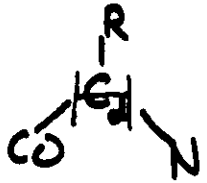
$$+ T\mathcal{H}_\phi + TH\psi$$

$$+ T\mathcal{H}_{harmonic}(r_{C_i' C_i^\beta}, r_{N_i C_i'}, r_{N_i C_i^\beta})$$

Planar Peptide Bond Geometry



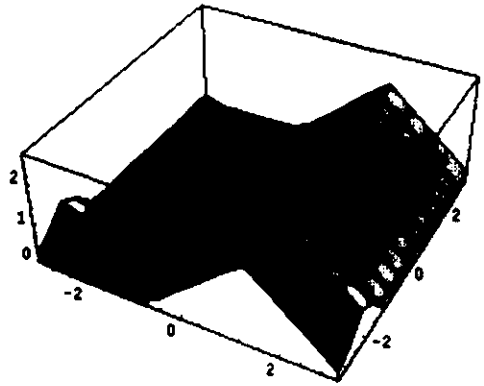
L-amino acid



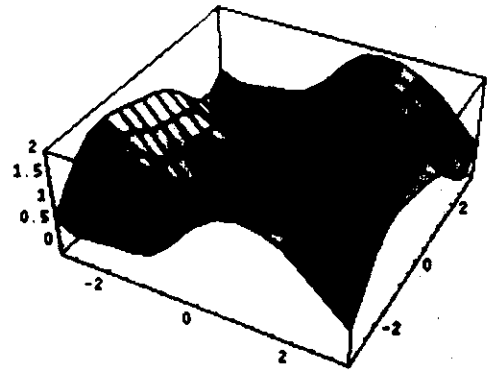
R: 

, H, CH_3, ...

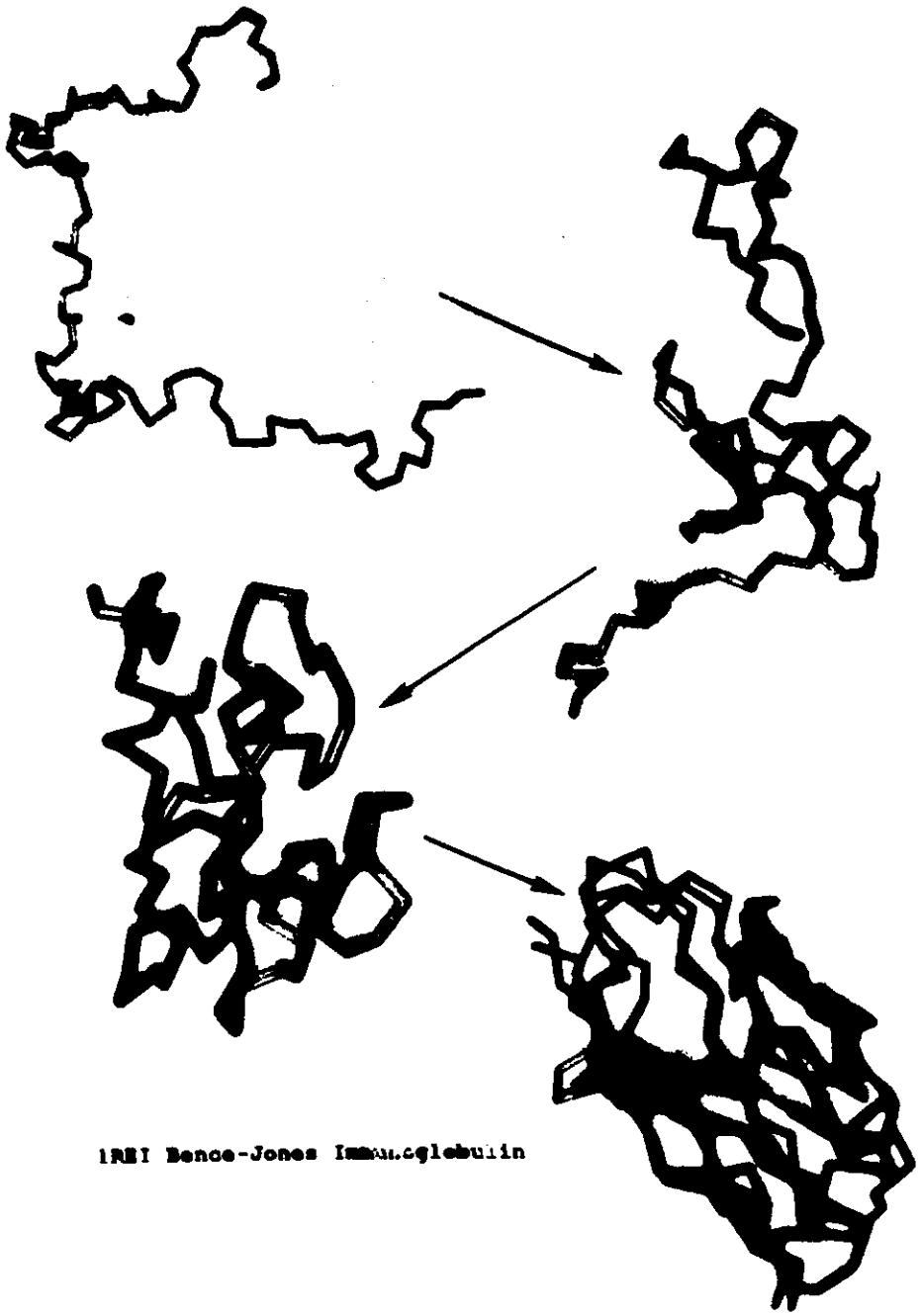
AMH Ramachandran Potentials



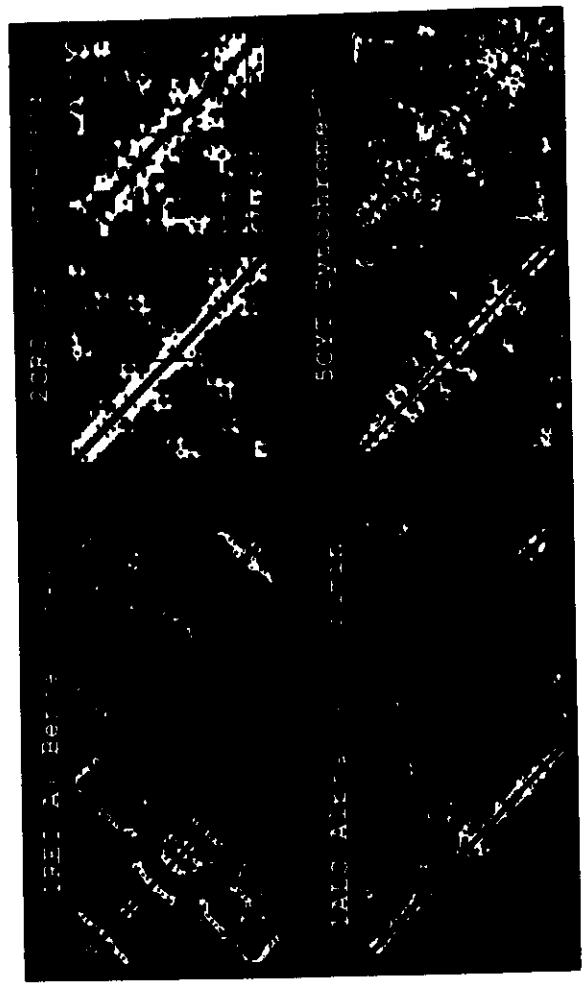
original potential



current potential



1REI Bence-Jones Immunoglobulin

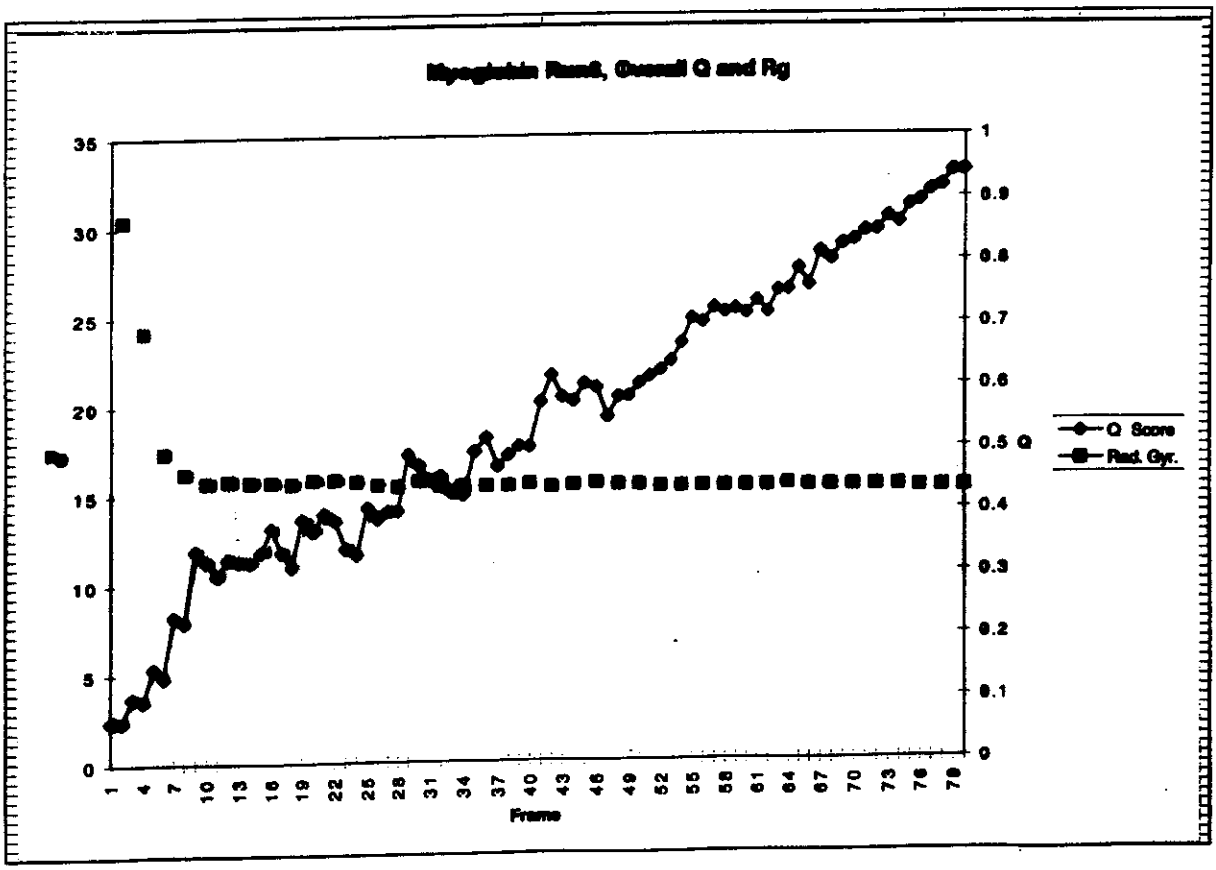


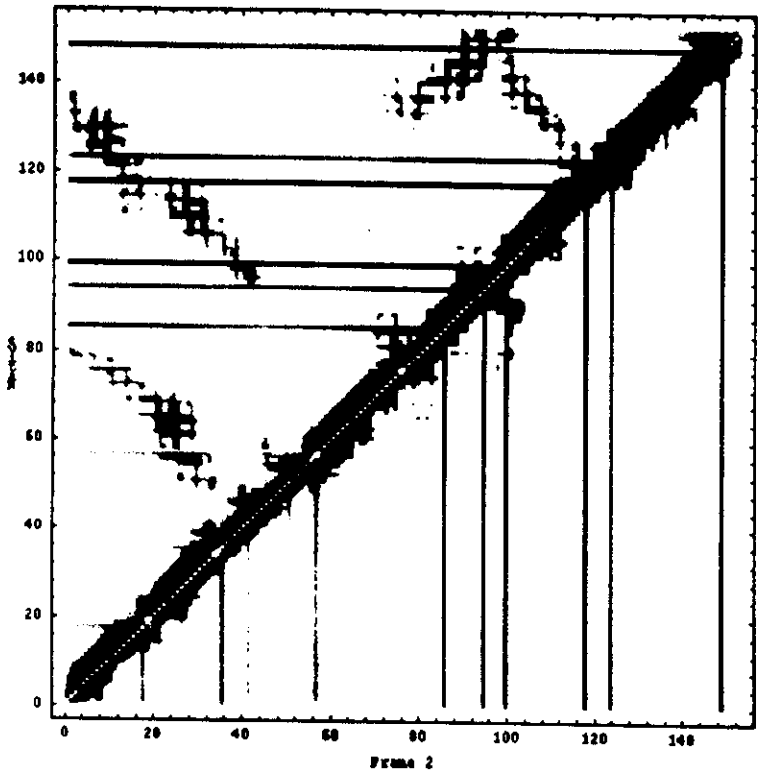
2 D Methods
(1 homolog)

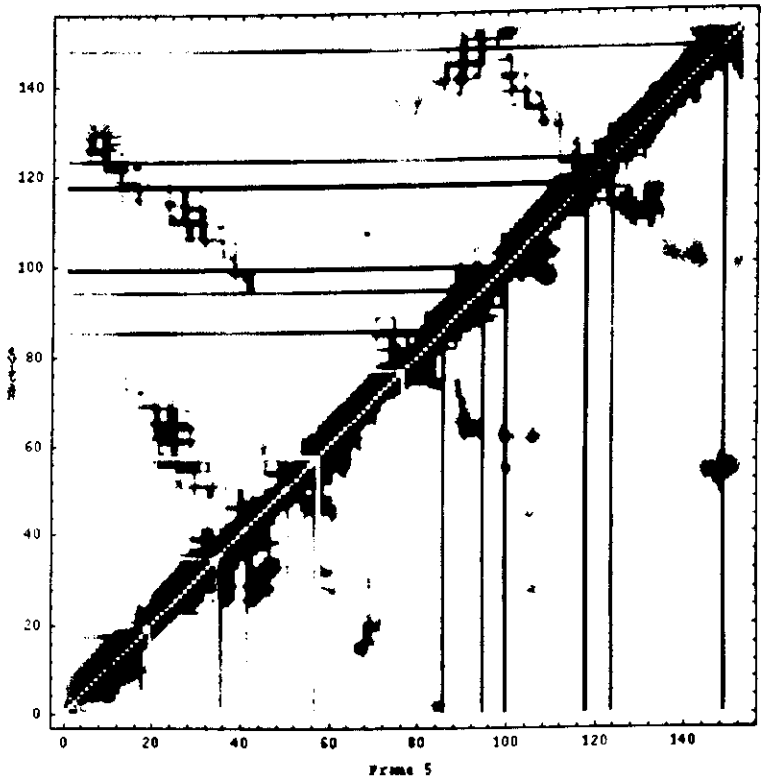
Molecular Dynamics Results

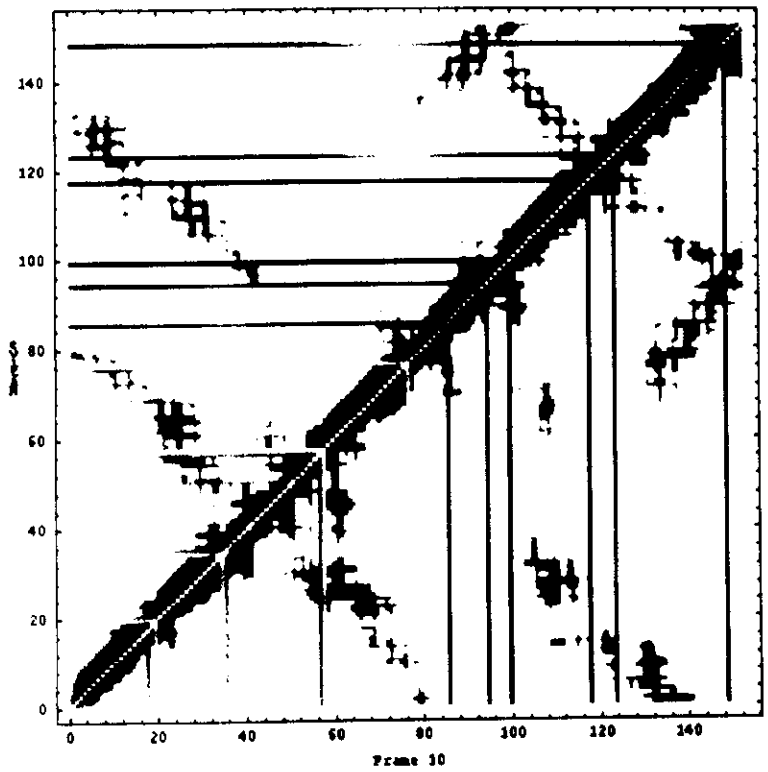
Target		Homolog					Predicted structure	
PDB	Name	PDB	Name	% I	q	RMS	q	RMS
Alpha								
2CRO	434 Cro	1LRD	Lambda Repressor	16.4	0.46	4.21	0.32 0.32	6.06 6.35
2MC	*P. aeruginosa* Cytochrome C551	3C2C	*R. rubrum* Cytochrome C2	20.0	0.51	3.48	0.36	5.41
3CLN	Bovine Ca-Binding	3CLN	Rat Calmodulin	38.3	0.49	3.15	0.40	5.15
3CYT	Tuna Cytochrome C	3C2C	*R. rubrum* Cytochrome C2	40.0	0.70	2.91	0.23 0.24	9.95 10.14
Beta								
1REI	Human Bence-Jones (Variable)	3HFM	Mouse Fab (H)	30.2	0.60	2.85	0.51 0.60	2.68 3.10
2HLA	Human Histocompatibility (β-2)	1MCP	Mouse Fab (L)	26.5	0.66	2.80	0.50	5.15
5PCY	Poplar Plastocyanin	1PAZ	*A. faecalis* Pseudoazurin	25.8	0.55	3.78	0.27 0.31	8.81 7.34
Other								
1ALC	Baboon Alpha-lactalbumin	2LYZ	Hen Egg White Lysozyme	38.7	0.83	1.96	0.47 0.47	7.58 7.81
2SNI	Barley Chymotrypsin Inhibitor II	2SEC	Leech Eglin-C	37.1	0.80	1.64	0.40 0.41 0.34	5.54 6.43 8.72

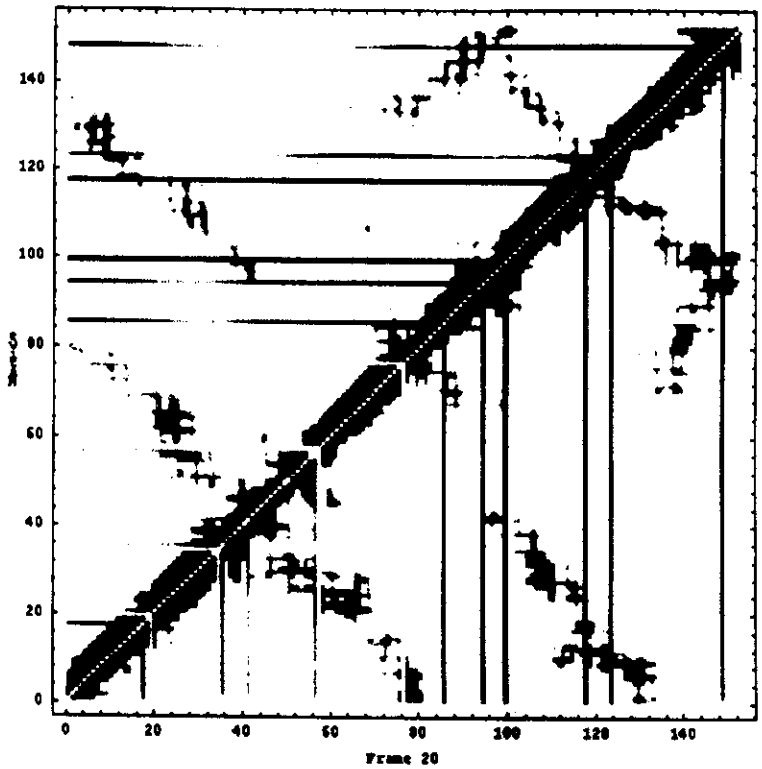
Q_table_myo3



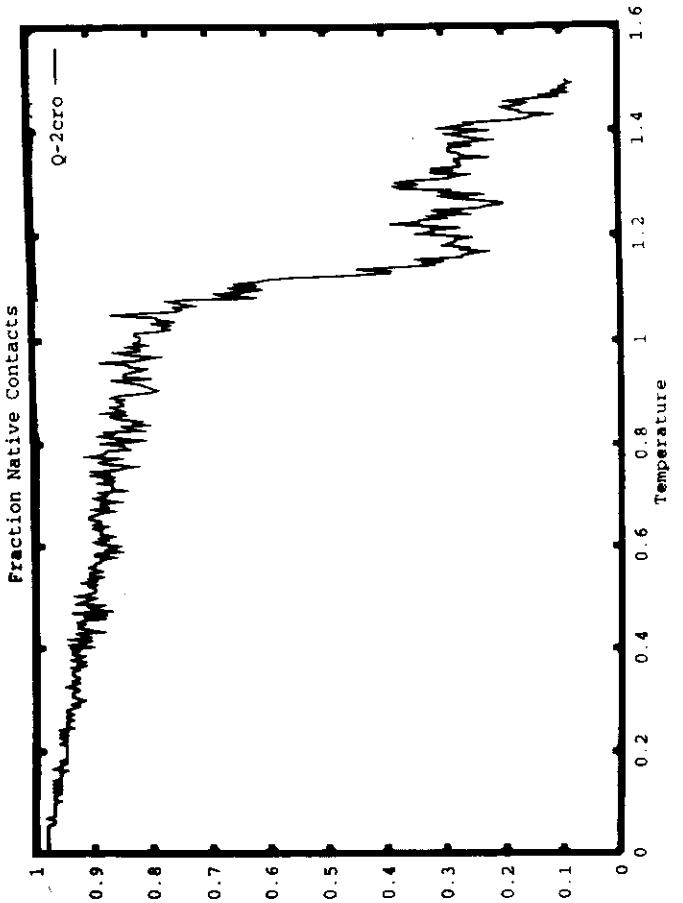

0
20
40
60
80
100
120
140
Frame 2


0
20
40
60
80
100
120
140
Frame 5

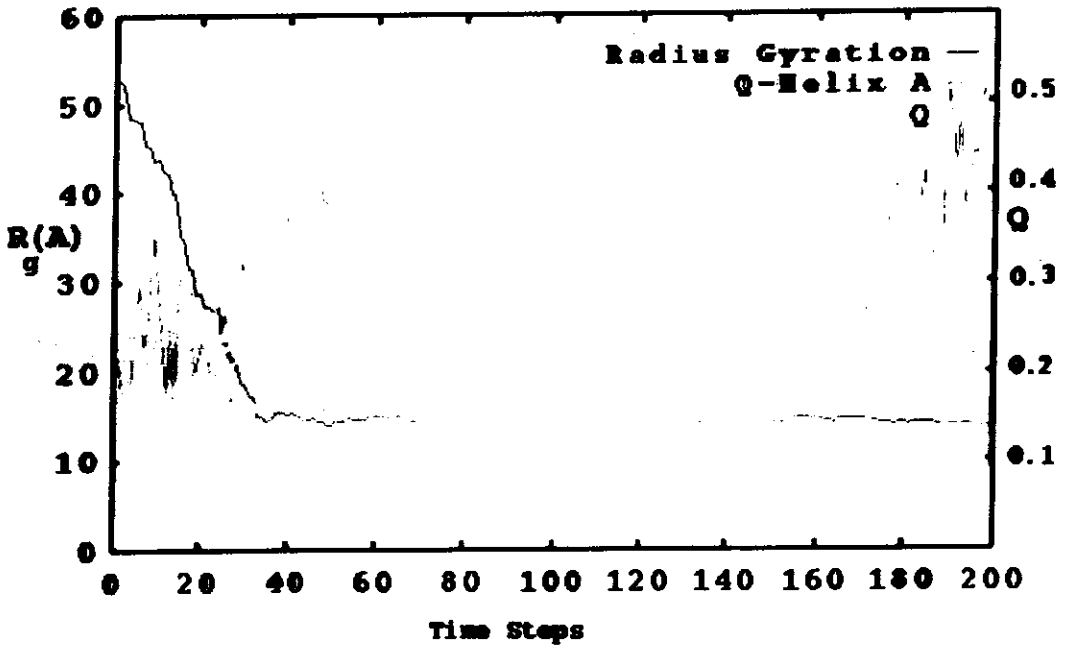

0
20
40
60
80
100
120
140
Frame 10


0
20
40
60
80
100
120
140
Frame 20

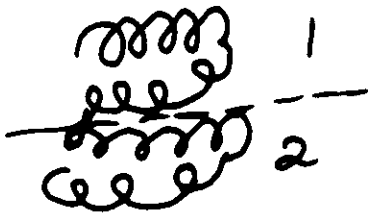
New Ψ, ϕ Potential



Multiple Memories H_{AMH}



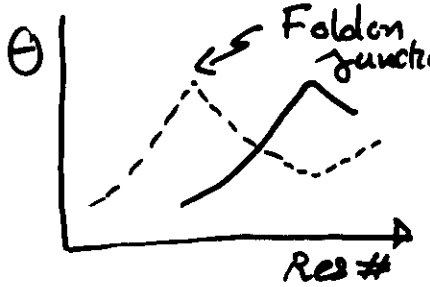
Foldons and Funnels



Is there a (sub-)domain structure in Folding Kinetics?

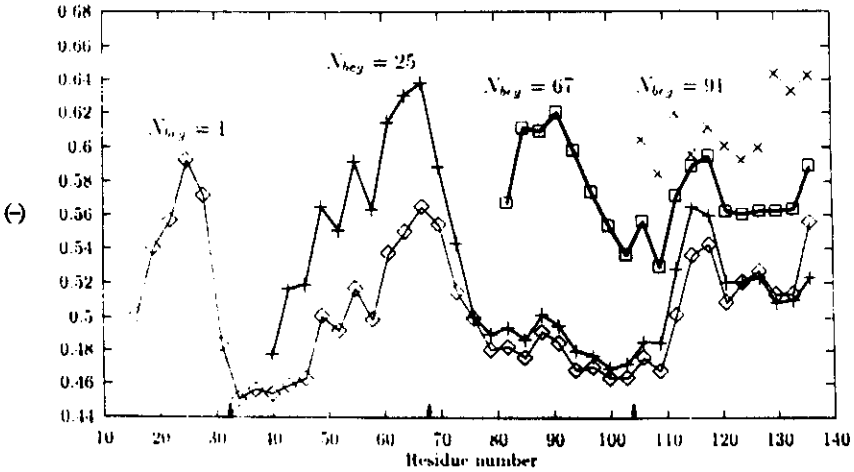
Maximize over cuts

$$\Theta = \frac{\delta E_s^{(1)}}{\sqrt{N}\,\Delta E^{(1)}} + \frac{\delta E_s^{(2)}}{\sqrt{N}\,\Delta E^{(2)}}$$

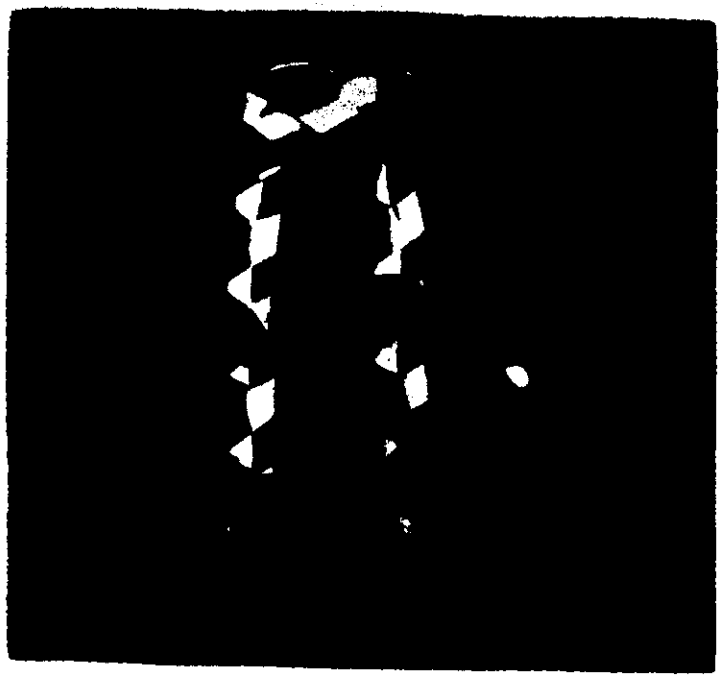


⇓

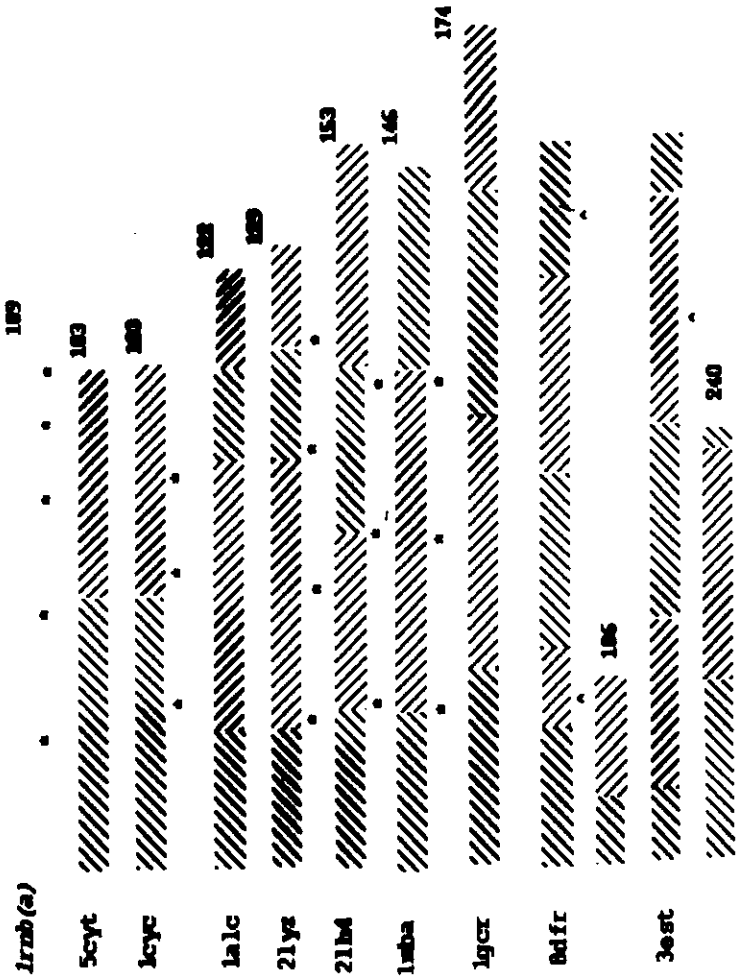
Rough Division into "Fast" Folding Contiguous Segments = Foldons



Comparing "Foldons" and Exons



Regions of maximum $\left(\frac{\delta E_s}{\Delta E}\right)_{local}$



Panchenko, Z. Schulten, P. Wolynes, PNAS (1995)

Table 1.

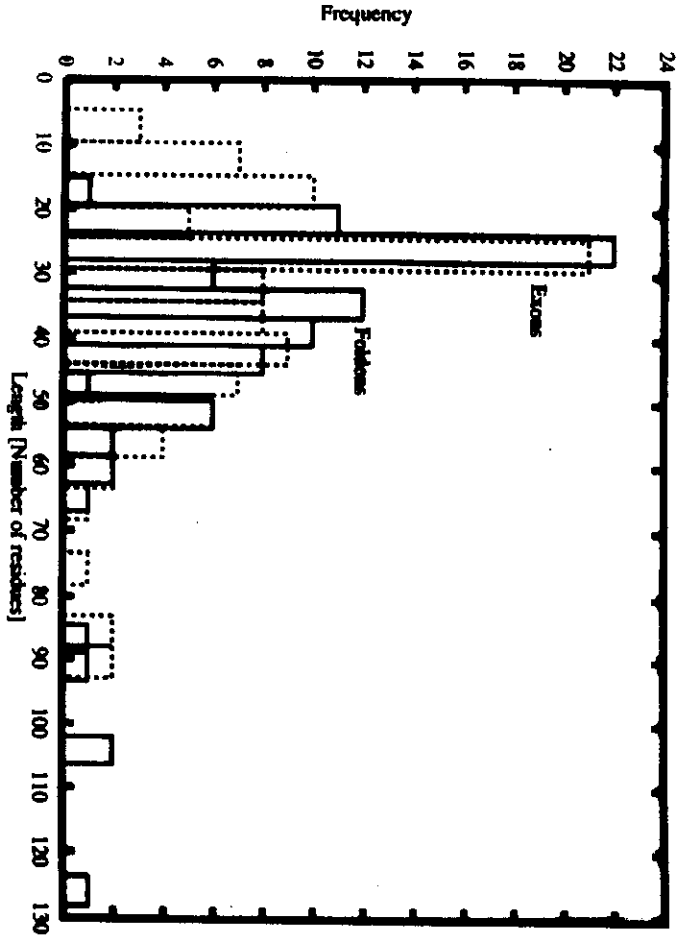
Foldons, exons and structural domain boundaries for some represantative proteins: myoglobin (MBA), triosephosphate isomerase (TIM), lysozyme (LYZ), dihydrofolate reductase (DFR) and carboxypeptidase (CPA). Foldon positions are defined according to BPTI criterion, ones that satisfied the CI-2 criterion are indicated by asterisk.

Foldon junctions	Intron positions	Domain junctions	Foldon junctions	Intron positions	Domain junctions
Myoglobin			Lysozyme		
28	31	30†,33§	28*	28	30†, 38#, 41§
64*		66†, 67§	55		55†,61§
91*				82	84†,86§,88#
	105	102§,101†	103*	108	109†
Triose-phosphate isomerase			Dihydrofolate reductase		
			28	28	33#
22*	13	28†		45	
	37	37†		80	
61*	61	63†,#	121	122	134#
	78	78†	163	161	
88*		93†	Carboxypeptidase		
	107	109†	22*	17	29§
121		125#,127†	55*	59	50#
	132			84	73§,96#
	151	147†	112*	126	122§,125#
163	167	162†	163*	156	141#,144§
	180/183	179†	193		189§,196#
202	209	201†,210#	226*	215	211§,234§
	237/239	230†	271	260	271§

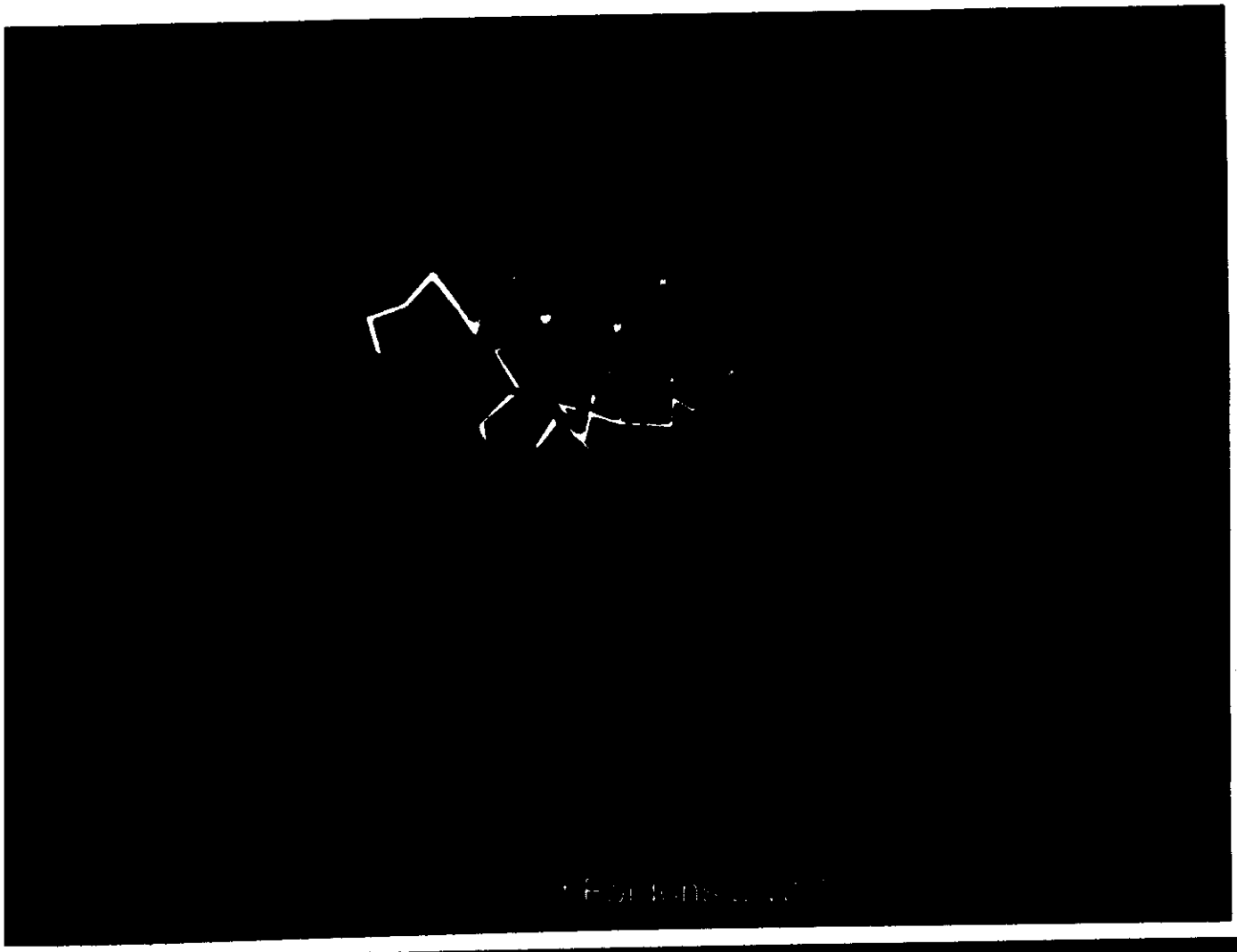
†, ‡ structural module boundaries calculated by C^{α} distance algorithm [8,10,33,31]

§ domain boundaries defined by geometrical algorithm of Rose [6]

domain boundaries found by surface accessibility algorithm [9]

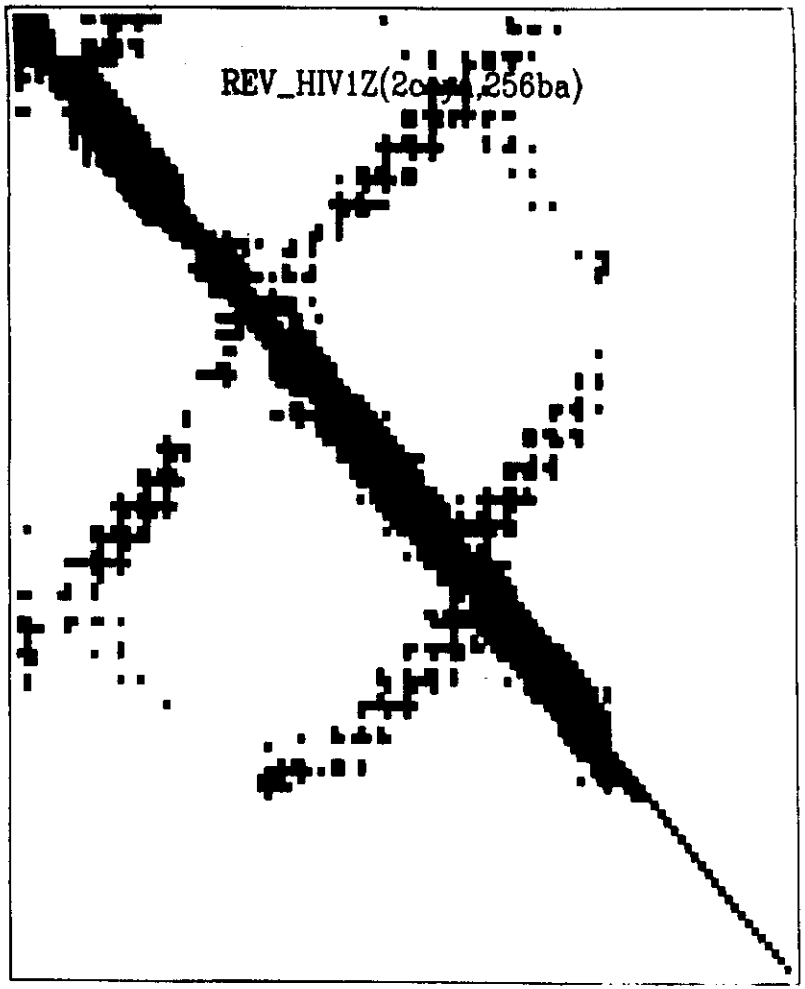


$$\frac{\theta_{ij}}{\theta_{NC}}$$



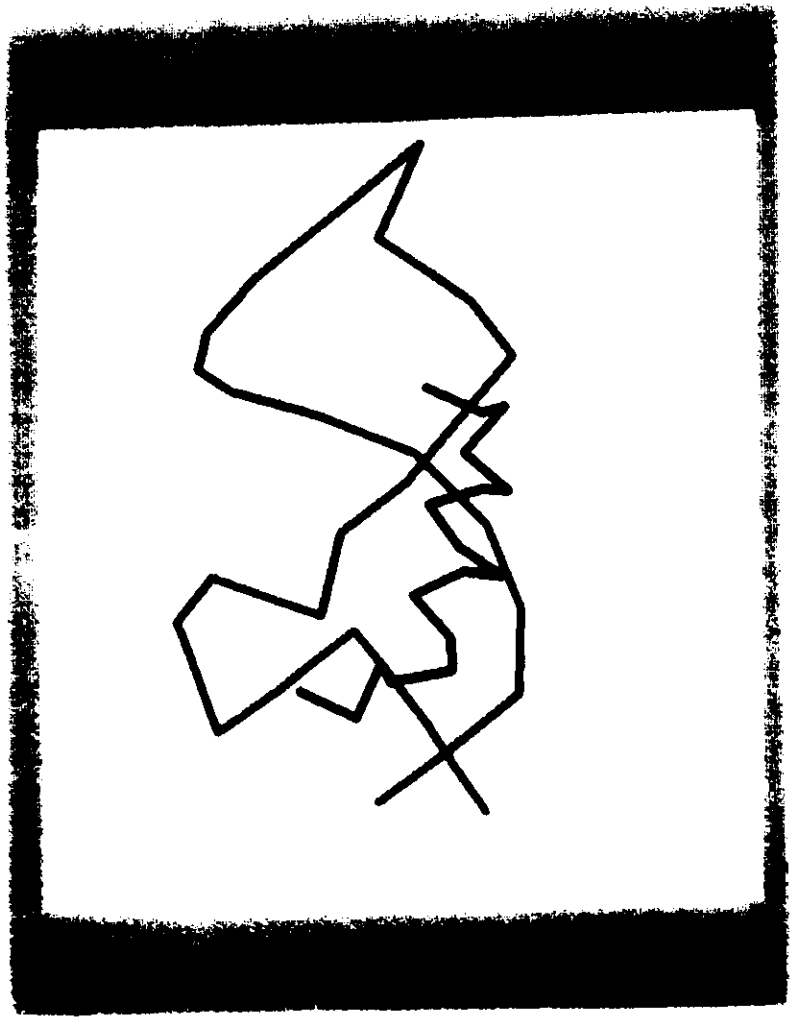

8atca(185-211), Q=0.34, RMS=3.5A

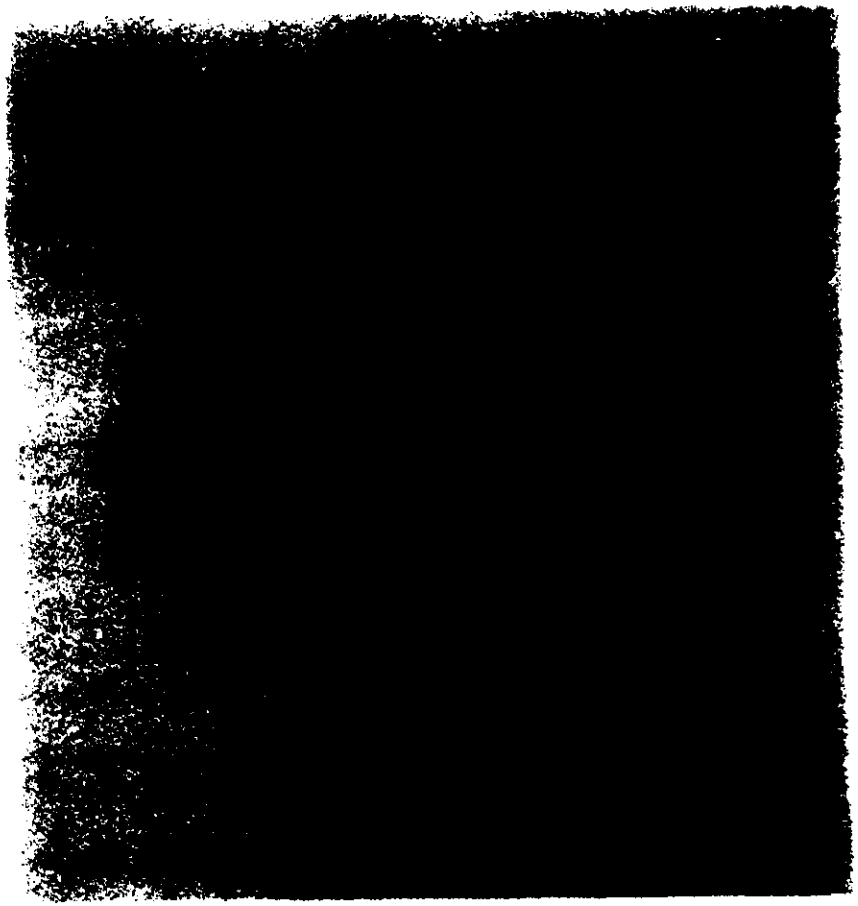
Local Threading



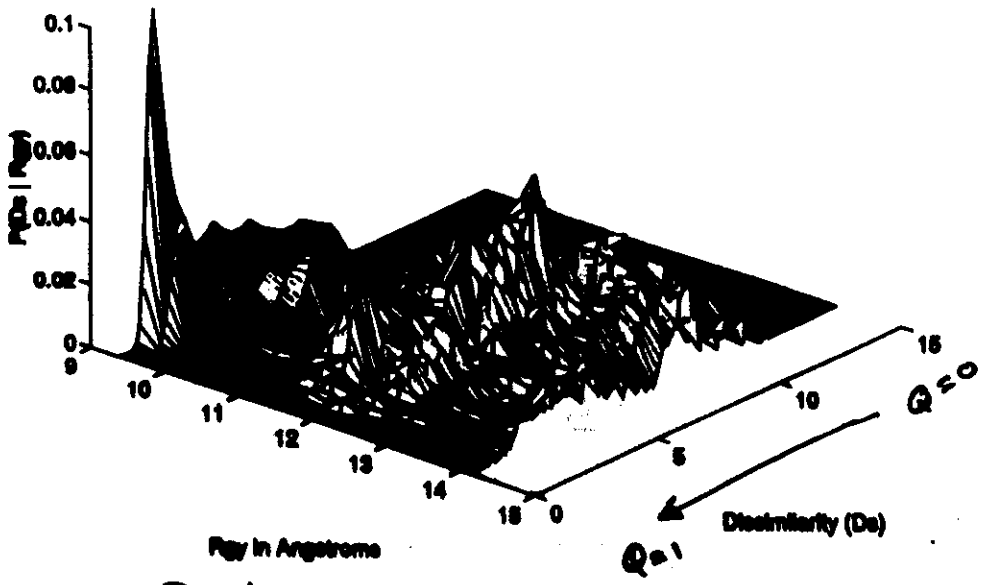
REV Regulatory Protein in HIV

Prediction: 3-helix Bundle





C. Brooks, et. al
(pre-print) 1995



Radius of Gyration