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

H4.SMR/916 - 40

SEVENTH COLLEGE ON BIOPHYSICS:

Structure and Function of Biopolymers: Experimental and Theoretical Techniques.

4 - 29 March 1996

Protein Folding Recognition and Dynamics in the Space of Contact Maps

E. Domany

Weismann Institute
Dept. of Biophysics
Rehovot
Israel

PROTEIN FOLDING

1. THE PROBLEM:

   PREDICTING STRUCTURE FROM SEQUENCE

2. REDUCED REPRESENTATION: CONTACT MAPS

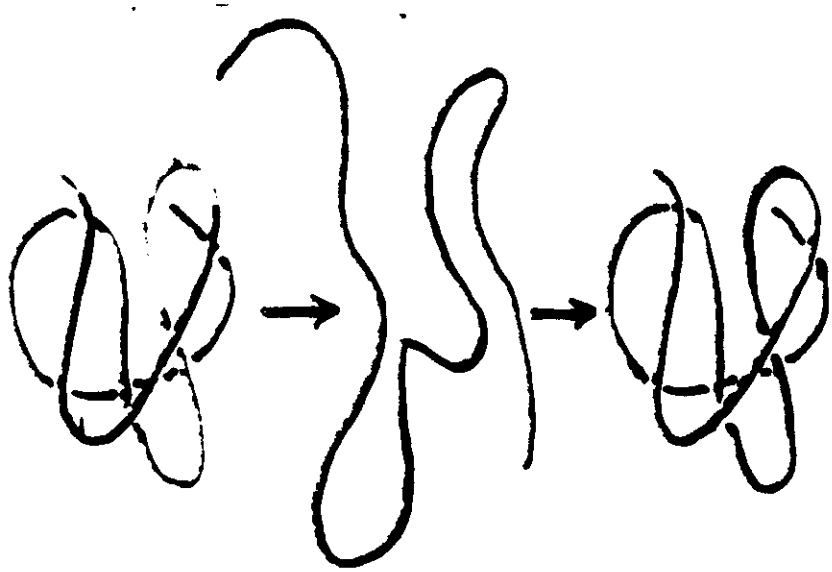
   - "ENERGY FUNCTION FOR CONTACT MAPS
   - DYNAMIC RULES

3. SELECTED RESULTS

DYNAMICS IN THE SPACE OF CONTACT MAPS - A VERY PROMISING TECHNIQUE

THE CENTRAL PROBLEMS:

PROTEINS FOLD (REVERSIBLY) INTO THEIR FUNCTIONALLY EFFECTIVE NATIVE STATE.



FOR A GIVEN SEQUENCE:

1. WHAT IS THE NATIVE FOLD?

2. HOW DOES THE PROTEIN "FIND" IT?

PARAMETRIZATION OF THE STRUCTURE

1. POSITION OF EVERY ATOM
2. $\phi - \psi$ ANGLES (N – C_α – C')
3. TRACE (COORDINATES OF C_α)
4. N14 (NO. NEIGHBORS IN SPHERE)
5. DISTANCE MAP
6. CONTACT MAP

CONTACT MAP OF A PROTEIN

DENOTE BY A THE AMINO-ACID SEQUENCE:

$$\text{A} = (a_1 . a_2 . a_3 ..... a_N)$$

(EXAMPLE: BPT: N=58, a_1=R, a_2=P, etc)

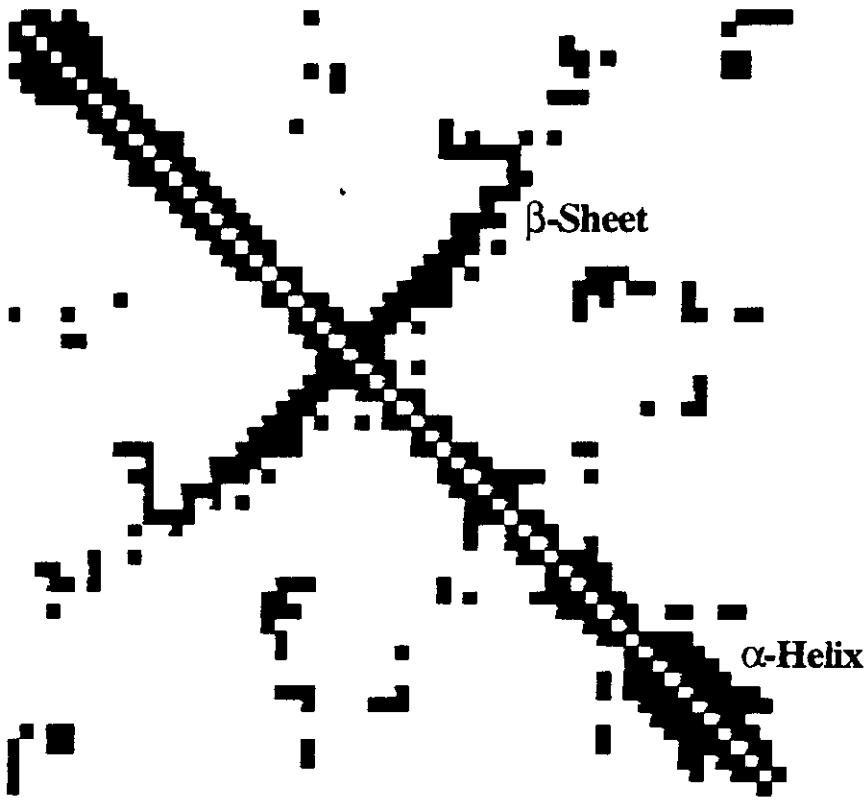
CONTACT MAP: N $\times$ N MATRIX S

$$S_{ij} = \begin{cases} 1 & \text{if } a_i \text{ and } a_j \text{ in CONTACT} \\ 0 & \text{otherwise} \end{cases}$$

CONTACT: IF THERE ARE TWO HEAVY ATOMS,

ONE IN a_i AND ONE IN a_j

WHOSE DISTANCE D$\leq 4.5\mathring{A}$



Contact map of the native BPTI structure.
Secondary structure elements (α-helices and β- sheets) are represented by typical patterns on a contact map.

THE CONTACT MAP S IS A

REDUCED REPRESENTATION

OF THE STRUCTURE: MANY MICROSCOPIC

CONFIGURATIONS PRODUCE THE SAME S.

$\mathcal{H}$ (S,A)= FREE ENERGY OF S AND A

$$\text{Prob(S)} \propto e^{-\mathcal{H}\text{(S,A)}} = \sum_{Config} e^{-\frac{1}{kT}E(Config)} \Delta(Config, S)$$

$$\Delta(Config, S) = \begin{cases} 1 & \text{if S consistent with Conf} \\ 0 & \text{otherwise} \end{cases}$$

MINIMAL $\mathcal{H}$ HYPOTHESIS

FOR ANY FIXED SEQUENCE A, THE NATIVE MAP

S_0 MINIMIZES $\mathcal{H}$ (S,A).

PROBLEM: HOW TO FIND $\mathcal{H}$?

$$\begin{aligned} \mathcal{H}(S,A) &= \mathcal{H}^{pair}(S,A) \\ &+ \mathcal{H}^{hydro}(S,A) \\ &+ \mathcal{H}^{constraint}(S,A) \end{aligned}$$

PARAMETRIZATION OF $\mathcal{H}^{pair}$ (S,A)

$$\mathcal{H}^{pair}(\mathrm{S,A}) = \sum_{i<j}^{N} \mathrm{S}_{ij} \Xi_{a_i,a_j}$$

THE 210 PARAMETERS $\Xi_{a,a'}$ ARE DETERMINED BY A STATISTICAL ANALYSIS OF 60 PROTEINS OF KNOWN STRUCTURE

$e^{-\Xi_{a,a'}} \propto$ Prob. contact of a and a'

$$\Xi_{a,a'} = -\log\left(\frac{\sum_P N_P w^P_{a,a'}}{\sum_P N_P}\right)$$

$$w^P_{a,a'} = \left[\frac{N_{a,a'}/M_{a,a'}}{N_{cont}/M_{cont}}\right]_P$$

Table I: The contact energies $E_{\mu\nu}$ of amino acids μ and ν, as obtained, using eq. (18), from 54 known protein structures.

Amino Acid	ALA	GLU	GLN	ASP	ASN	LEU	GLY	LYS	SER	VAL	ARG	THR	PRO	ILE	MET	PHE	TYR	CYS	TRP	HIS
ALA	0.175	0.206	0.150	0.366	0.242	-0.168	0.355	0.216	0.303	-0.151	0.154	0.272	0.479	-0.279	-0.162	-0.200	-0.156	0.350	-0.345	0.015
GLU	0.206	-0.173	0.153	0.336	-0.074	0.099	0.522	-0.175	0.286	0.188	-0.372	0.148	0.298	-0.097	-0.053	-0.184	0.085	0.445	-0.253	-0.320
GLN	0.150	0.153	-0.036	0.231	-0.128	0.034	0.291	0.124	0.234	-0.017	-0.349	0.087	-0.072	0.098	0.200	-0.197	-0.141	-0.389	-0.251	-0.130
ASP	0.366	0.336	0.231	0.285	0.120	0.217	0.358	-0.143	0.226	0.332	-0.442	0.184	0.539	0.157	0.080	0.036	-0.230	0.510	-0.149	-0.103
ASN	0.242	-0.074	-0.128	0.120	0.264	0.045	0.267	0.081	0.339	0.122	-0.164	0.216	0.466	0.111	-0.382	0.166	-0.333	0.039	0.039	-0.057
LEU	-0.168	0.099	0.034	0.217	0.045	-0.685	0.102	0.067	0.212	-0.591	-0.353	-0.020	0.093	-0.781	-0.393	-0.742	-0.529	-0.380	-0.654	-0.169
GLY	0.355	0.522	0.291	0.358	0.267	0.102	0.327	0.223	0.550	0.211	0.184	0.327	0.414	0.296	0.296	0.312	0.043	0.019	-0.173	0.275
LYS	0.216	-0.175	0.124	-0.143	0.081	0.067	0.223	0.397	0.253	-0.004	0.238	0.092	0.218	-0.040	-0.021	-0.052	-0.279	0.212	-0.185	0.187
SER	0.303	0.286	0.234	0.226	0.339	0.212	0.550	0.253	0.250	0.207	0.047	0.215	0.232	0.132	0.140	-0.021	0.034	0.119	-0.120	0.250
VAL	-0.151	0.188	-0.017	0.332	0.122	-0.591	0.211	-0.004	0.207	-0.451	0.019	0.033	0.280	-0.588	-0.326	-0.600	-0.261	-0.205	-0.598	0.054
ARG	0.154	-0.372	-0.349	-0.442	-0.164	-0.353	0.184	0.238	0.047	0.019	-0.017	-0.130	0.052	-0.433	-0.397	-0.466	-0.544	-0.223	-0.335	0.043
THR	0.272	0.148	0.087	0.184	0.216	-0.020	0.327	0.092	0.215	0.033	-0.130	0.044	0.254	-0.178	-0.044	-0.189	-0.131	0.110	-0.286	0.000
PRO	0.479	0.298	-0.072	0.539	0.466	0.093	0.414	0.218	0.232	0.280	0.052	0.254	0.341	0.403	-0.242	0.082	-0.289	0.216	-0.345	0.309
ILE	-0.279	-0.097	0.098	0.157	0.111	-0.781	0.296	-0.040	0.132	-0.588	-0.433	-0.178	0.403	-0.716	-0.727	-0.748	-0.547	0.042	-0.849	0.016
MET	-0.162	-0.053	0.200	0.080	-0.382	-0.393	0.296	-0.021	0.140	-0.326	-0.397	-0.044	-0.242	-0.727	-0.869	-0.664	-0.705	-0.632	-1.281	-0.563
PHE	-0.200	-0.184	-0.197	0.036	0.166	-0.742	0.312	-0.052	-0.021	-0.600	-0.466	-0.189	0.082	-0.748	-0.664	-0.953	-0.501	-0.650	-1.114	-0.367
TYR	-0.156	0.085	-0.141	-0.230	-0.333	-0.529	0.043	-0.279	0.034	-0.261	-0.544	-0.131	-0.289	-0.547	-0.705	-0.501	-0.258	-0.263	-0.619	-0.609
CYS	0.350	0.445	-0.389	0.510	0.039	-0.380	0.019	0.212	0.119	-0.205	-0.223	0.110	0.216	0.042	-0.632	-0.650	-0.263	-1.843	-0.147	-0.814
TRP	-0.345	-0.253	-0.251	-0.149	0.039	-0.654	-0.173	-0.185	-0.120	-0.598	-0.335	-0.286	-0.345	-0.849	-1.281	-1.114	-0.619	-0.147	-0.826	-0.682
HIS	0.015	-0.320	-0.130	-0.103	-0.057	-0.169	0.275	0.187	0.250	0.054	0.043	0.000	0.309	0.016	-0.563	-0.367	-0.609	-0.814	-0.682	-0.568

PARAMETRIZATION OF $\mathcal{H}^{hydro}$ (S,A)

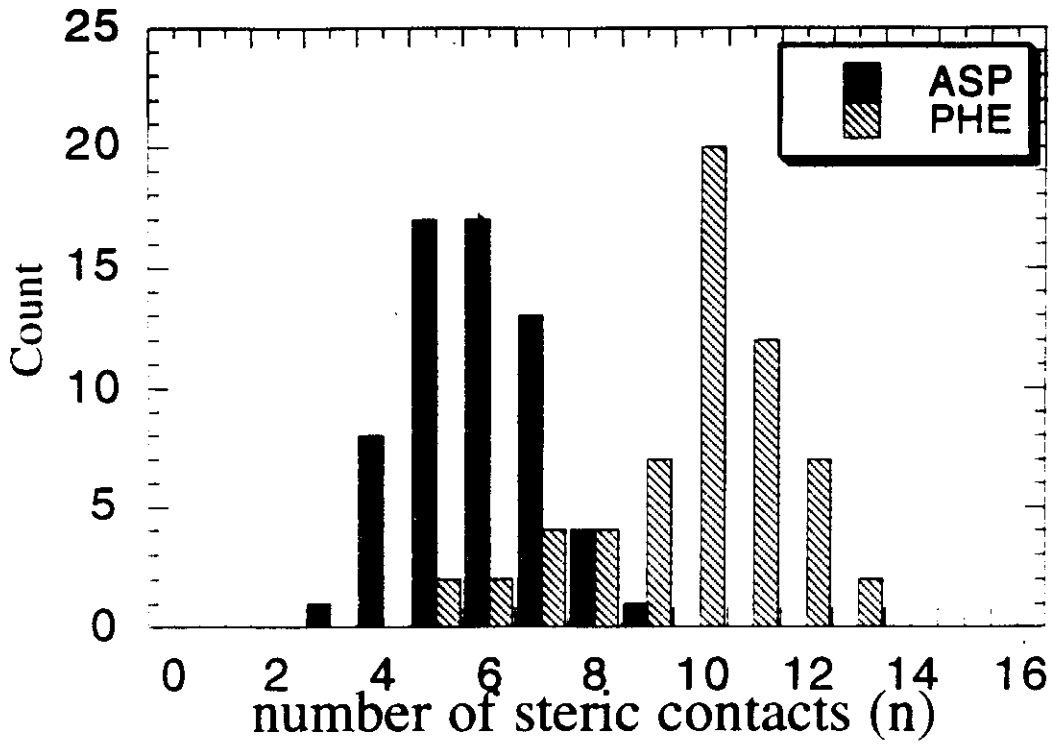
$$\mathcal{H}^{hydro}(\mathrm{S,A}) = \sum_{i}^{N} a_i \left(a_i - \sum_{k}^{N} \mathrm{S}_{ik} \right)^2$$

IMPOSES A TENDENCY FOR AMINO ACID a TO HAVE n_a CONTACTS.

FOR HYDROPHOBIC a - LARGE n_a

FOR HYDROPHILIC a - SMALL n_a

n_a, β OBTAINED FROM STATISTICAL ANALYSIS OF 60 KNOWN CONTACT MAPS



amino acid	ALA	GLU	GLN	ASP	ASN	LEU	GLY	LYS	SER	VAL
n_u	6.7	6.4	7.0	6.1	6.4	9.3	5.6	6.5	5.8	8.7
σ_u	2.4	2.4	2.7	2.6	2.6	2.6	2.3	2.4	2.6	2.4

ARG	THR	PRO	ILE	MET	PHE	TYR	CYS	TRP	HIS
8.0	6.7	5.8	9.5	9.4	10.4	9.7	8.1	11.4	8.0
3.2	2.6	2.3	2.4	2.8	2.3	2.8	2.2	2.8	2.7

TESTING $\mathcal{H}$: 1.SEQUENCE SPECIFITY

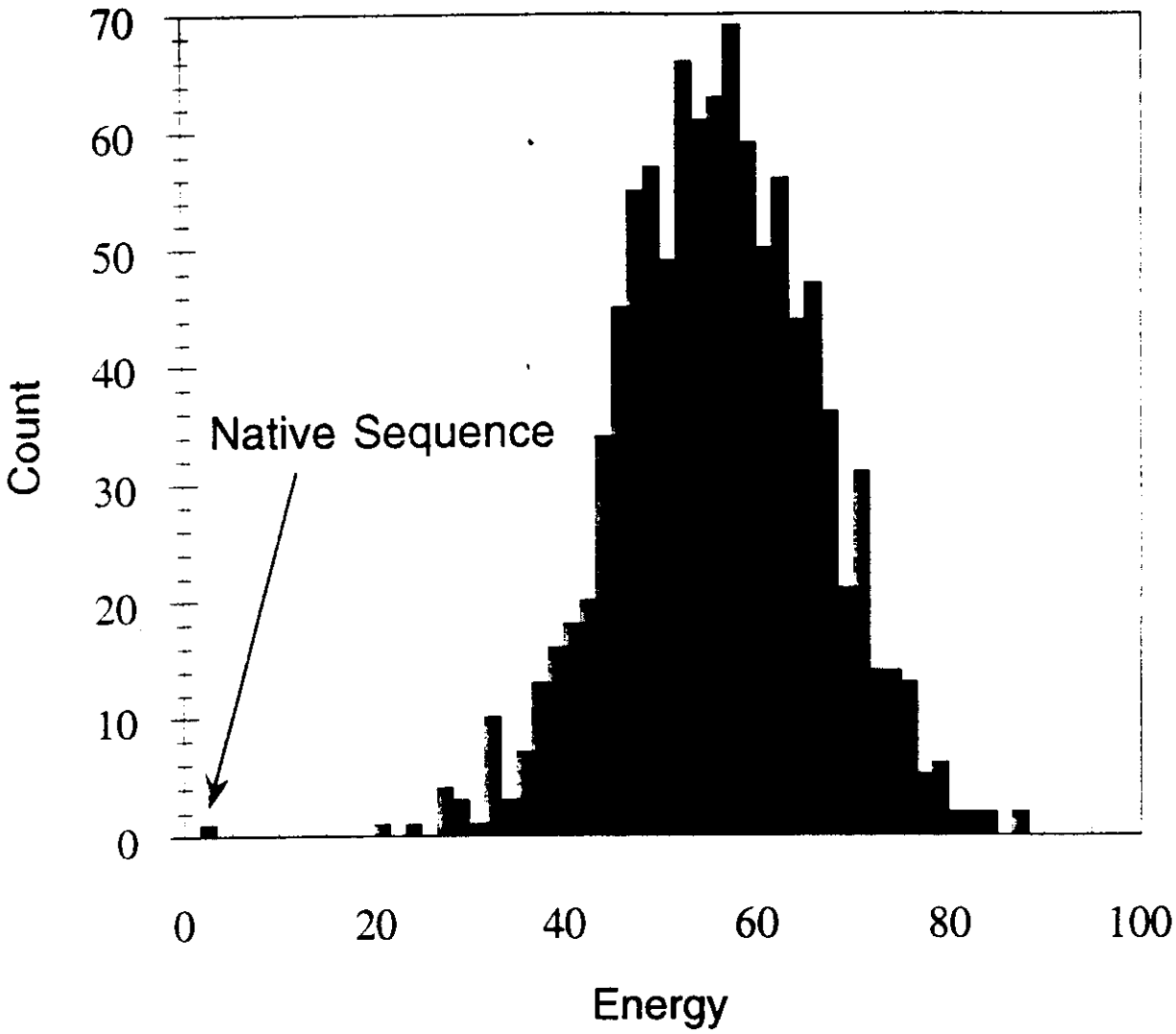
FIX MAP AND VARY SEQUENCE: PICK A PROTEIN WITH KNOWN NATIVE MAP S AND SEQUENCE A:

SCRAMBLE THE SEQUENCE RANDOMLY:

$$A \rightarrow A'$$

WE CALCULATED $\mathcal{H}(S,A')$ FOR MANY A'

THE ENERGY OF THE CORRECT SEQUENCE A IS 5σ BELOW THE AVERAGE OF THE SCRAMBLED ONES A': OUR ENERGY FUNCTION RECOGNIZES CORRECTLY THE AMINO ACIDS



Histogram of energies obtained by randomly shuffled sequences folded into the BPTI native deviates from the average energy value by 5 standard deviations.

TESTING $\mathcal{H}$: 2.FOLD RECOGNITION

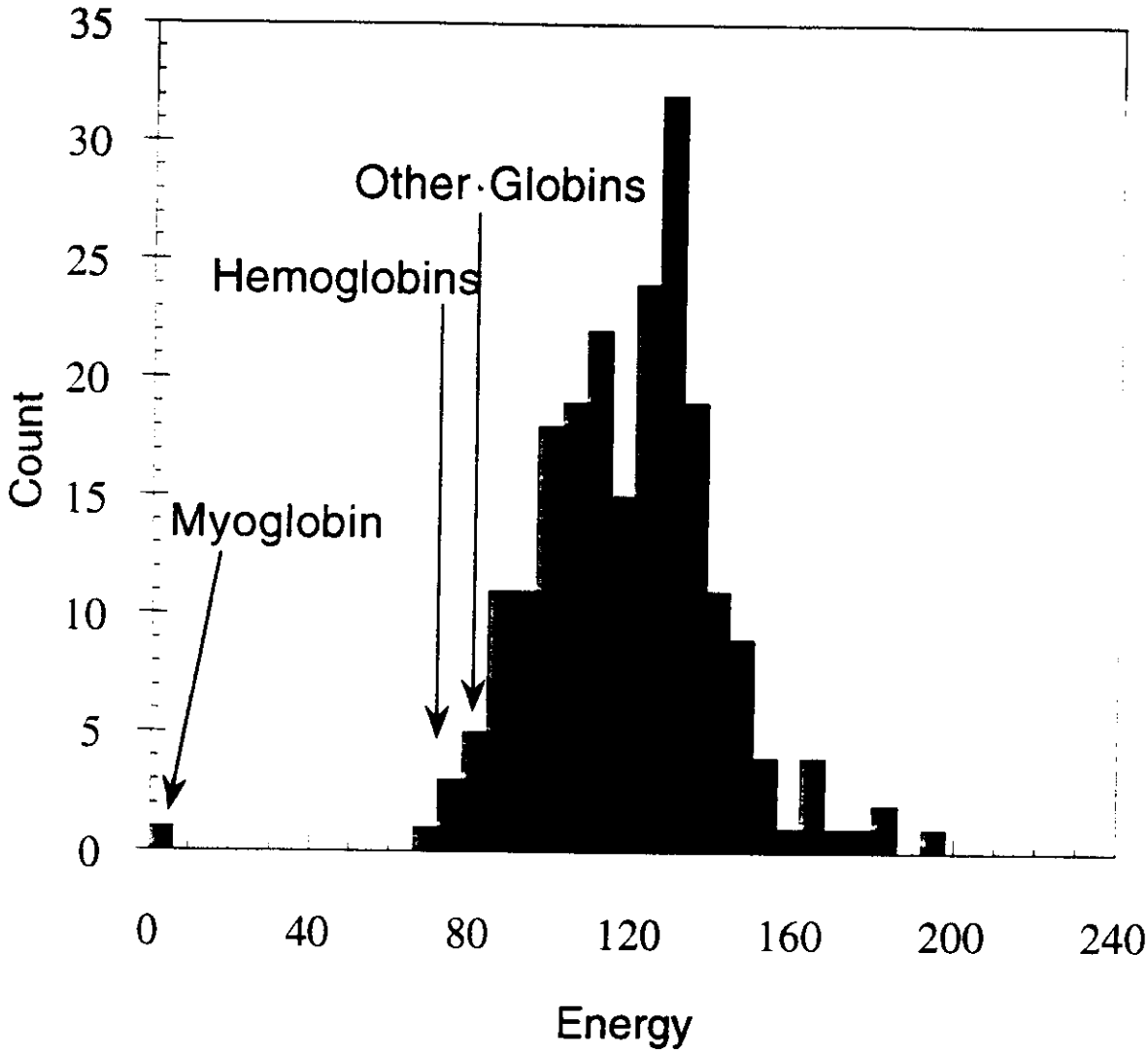
FIX SEQUENCE AND VARY MAP:

PICK A PROTEIN WITH KNOWN NATIVE MAP S AND SEQUENCE A: USE BANK OF KNOWN STRUCTURES TO EXTRACT ALL POSSIBLE N $\times$ N CONTACT MAPS S'. "THREAD" A THROUGH ALL MAPS AND CALCULATE $\mathcal{H}$(S',A).

1. FOR 238 OUT OF 249 PROTEINS TESTED, THE NATIVE MAP HAD THE LOWEST ENERGY.

2. THE GAP BETWEEN LOWEST ENERGY AND MEAN ENERGY OF ALL MAPS WAS 5σ

3. RECOGNIZED BPTI1,2; 3cro; spectrin



Fold recognition test. Histogram of energies obtained by threading the myoglobin sequence through the different conformations of a protein set. The native structure has the lowest energy and other members of the globin family come next.

TESTING H: 3. RECOGNITION OF SIMILAR STRUCTURES

S

Globins :

Hemoglobin - Colicin A
Hemoglobin - Phycocyanin-C
Cytochromes - Hemerythrin
Hemerythrin - Apolipoprotein E2

TIM barrel :

Adenosine deaminase - D-xylose isomerase
Glycolate oxidase - D-xylose isomerase
Aldose reductase - Adenosine deaminase
Triosephosphate isomerase - D-xylose isomerase

Dehydrogenases :

Lactate dehydrogenase - Malate dehydrogenase

Hematopoetic cytokins:

Interleukin-2 - Granulocyte-colony stimulating factor
Interleukin-4 - Granulocyte-colony stimulating factor

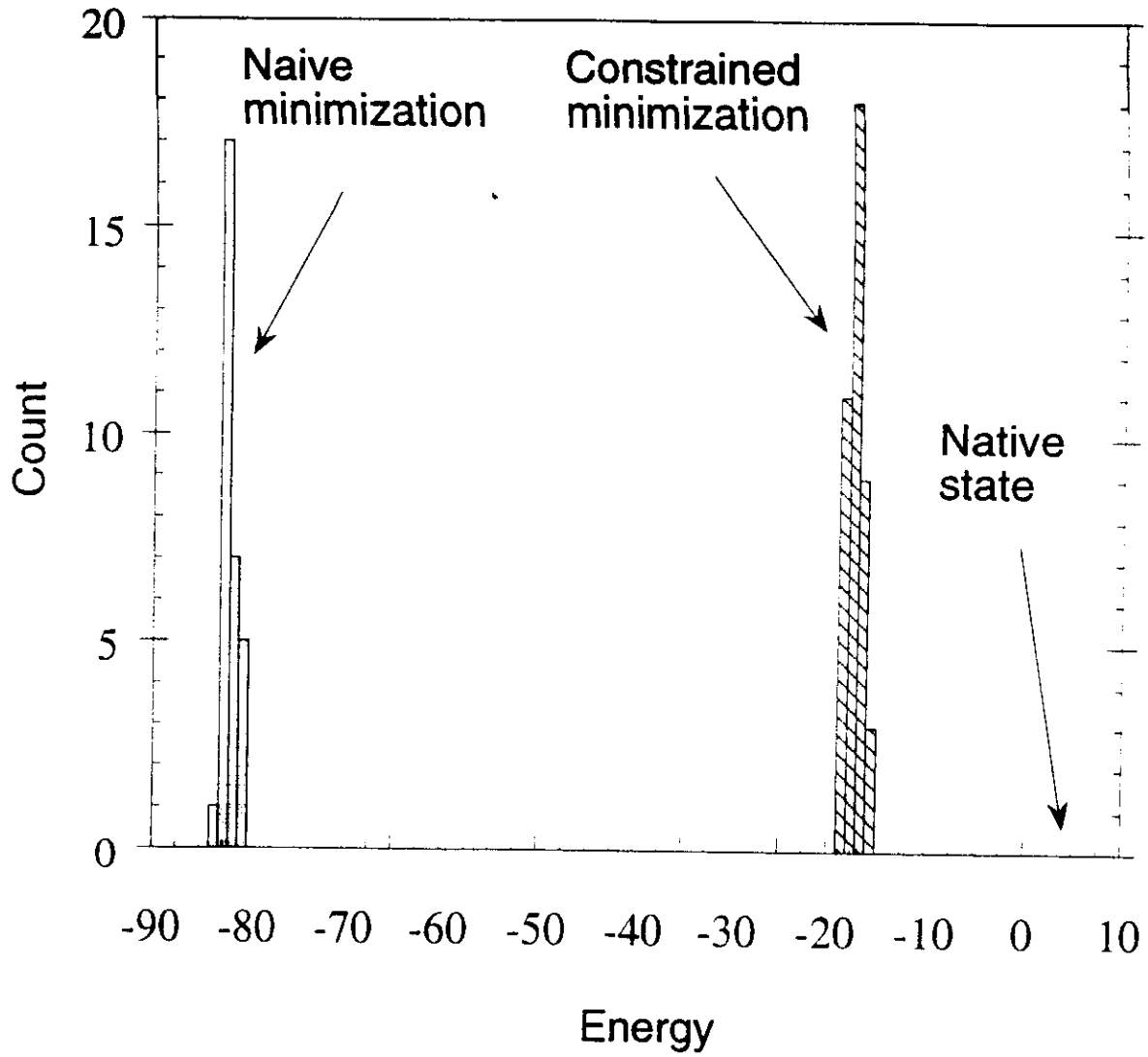
Others :

Ferritin-H - Ribonucleotide reductase R2
Chloramphenicol acetyltransferase - Dihydrolipoamide acetyltransferase
Class I MHC - Immunoglobin
Tumor necrosis factor - Viral coat and capsid protein (SBMW)

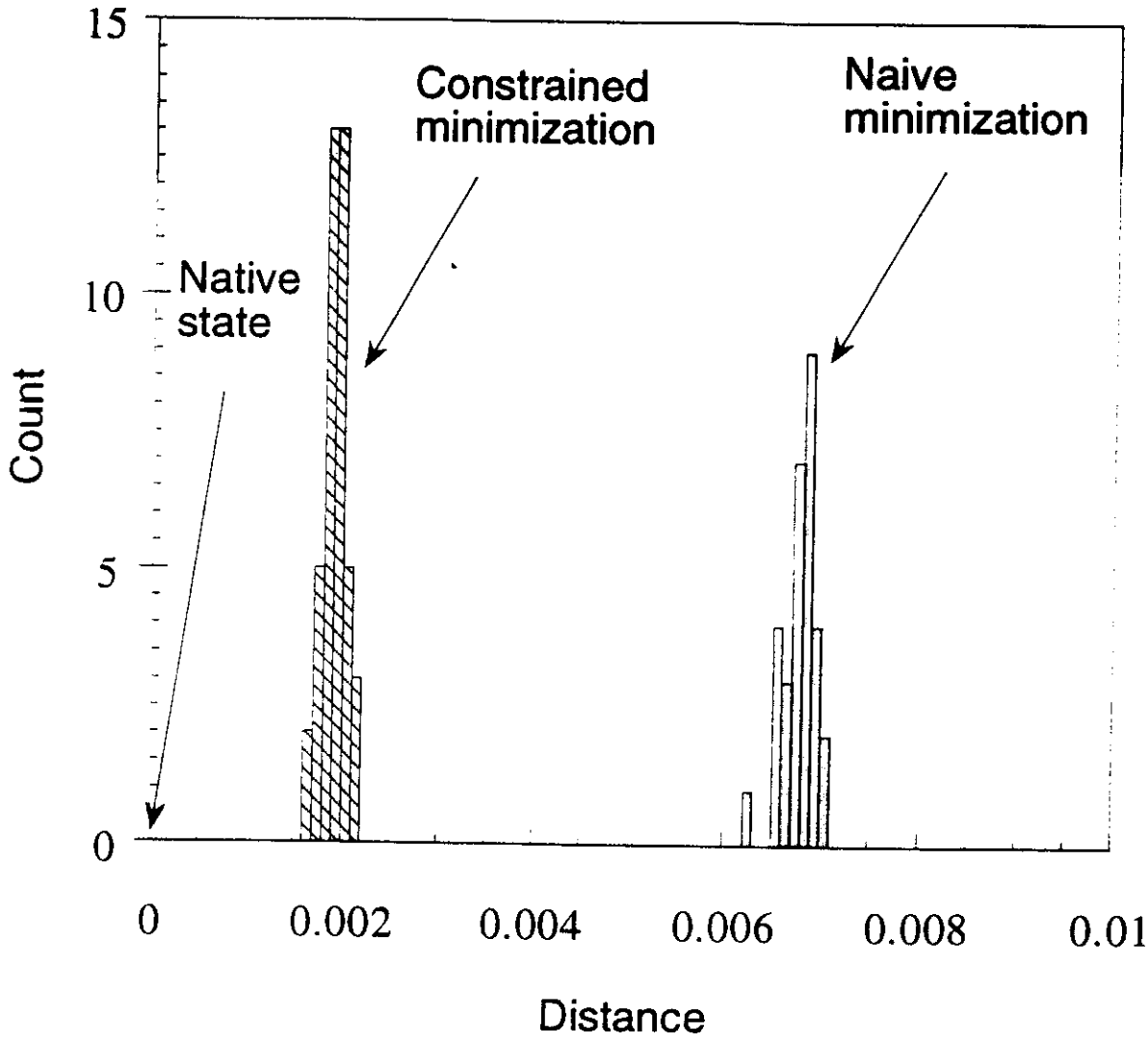
DYNAMICS IN THE SPACE OF CONTACT MAPS

MONTE CARLO PROCEDURE: FOR A GIVEN FIXED A START AT NATIVE S: ADD/REMOVE ONE CONTACT AT A TIME, ACCEPT NEW MAP S' IF $\mathcal{H}$ WENT DOWN.

1. NATIVE FOLD - UNSTABLE
2. FINAL STABLE MAP HAS MUCH LOWER ENERGY
3. NATIVE STRUCTURE WAS DESTROYED (LARGE FINAL DISTANCE FROM NATIVE)

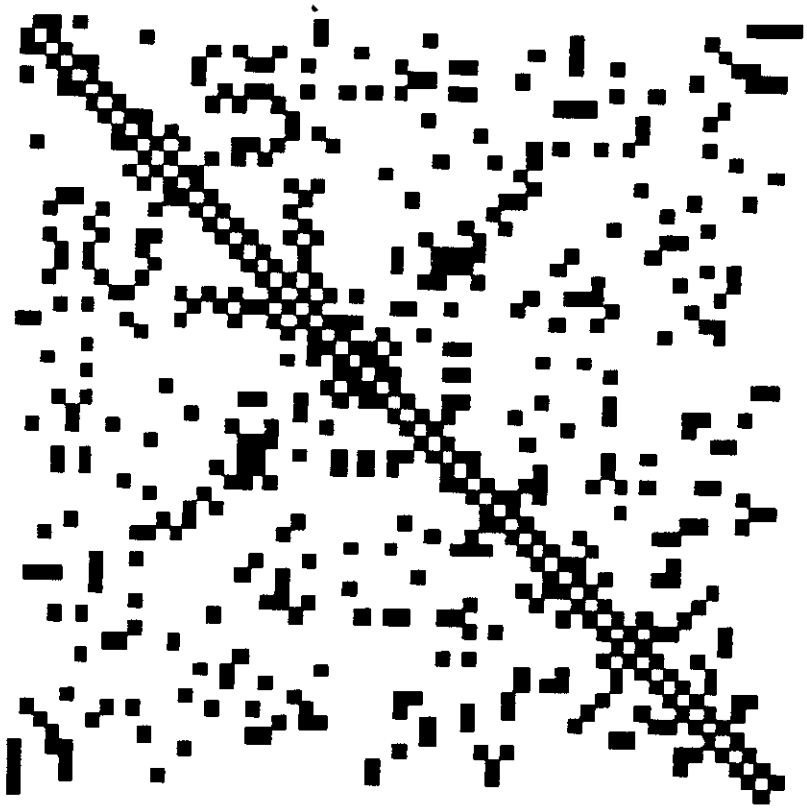


Histogram of energies of contact maps obtained by constrained (dashed bars) and 'naive' (solid bars) energy minimizations, starting from the native state.



Histogram of distances between the native contact map and the contact maps obtained by constrained (dashed bars) and 'naive' (solid bars) energy minimizations, starting from the native state.

$$\text{DISTANCE}(S, S') = \frac{\sum_{i<j} |S_{ij} - S'_{ij}|}{N_{\text{CONT.}}(S) \cdot N_{\text{CONT.}}(S')}$$

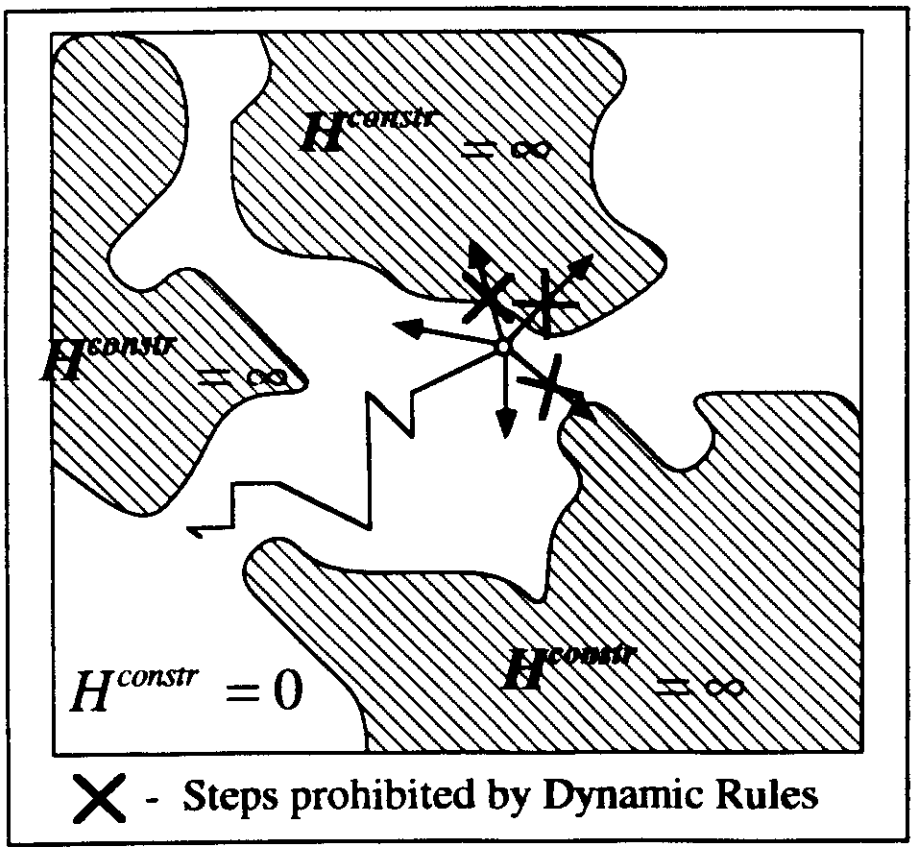


Final contact map corresponding to local minimum reached by "naive" Monte-Carlo energy minimization that ignored $H^{constraints}$. The minimization started from the native contact map (see Fig 1).

NEED FOR $\mathcal{H}^{constraint}$ (S,A)

"DYNAMICS" GENERATED NON-PHYSICAL MAPS

$$\mathcal{H}^{constraint}(\text{S,A}) = \begin{cases} 0 & \text{if S PHYSICAL} \\ \infty & \text{otherwise} \end{cases}$$



WE DO NOT KNOW WHICH S ARE PHYSICAL:

PRESCRIPTION:

- ALWAYS START FROM A PHYSICAL MAP
- ONLY THOSE "MOVES" THAT SATISFY SOME (HEURISTIC) DYNAMIC RULES ARE ALLOWED.

TEST: STABILITY OF NATIVE MAPS

APPLICATIONS:

1. EXPLORE ENERGY LANDSCAPE (vs DISTANCE)
2. HEATING/COOLING DYNAMICS
3. IMPROVE ON FOLD RECOGNITION

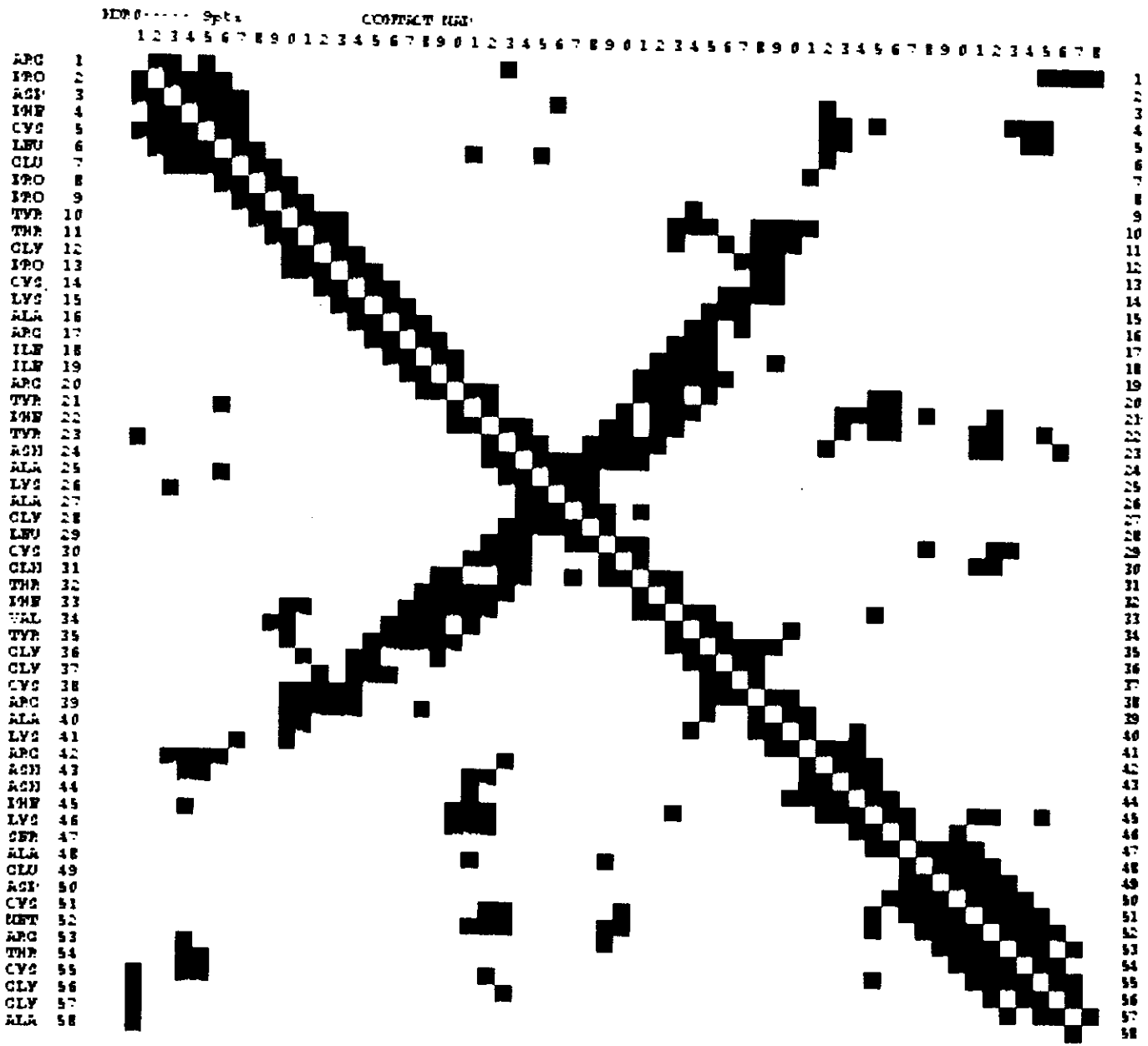
MAP VE
PLOT I

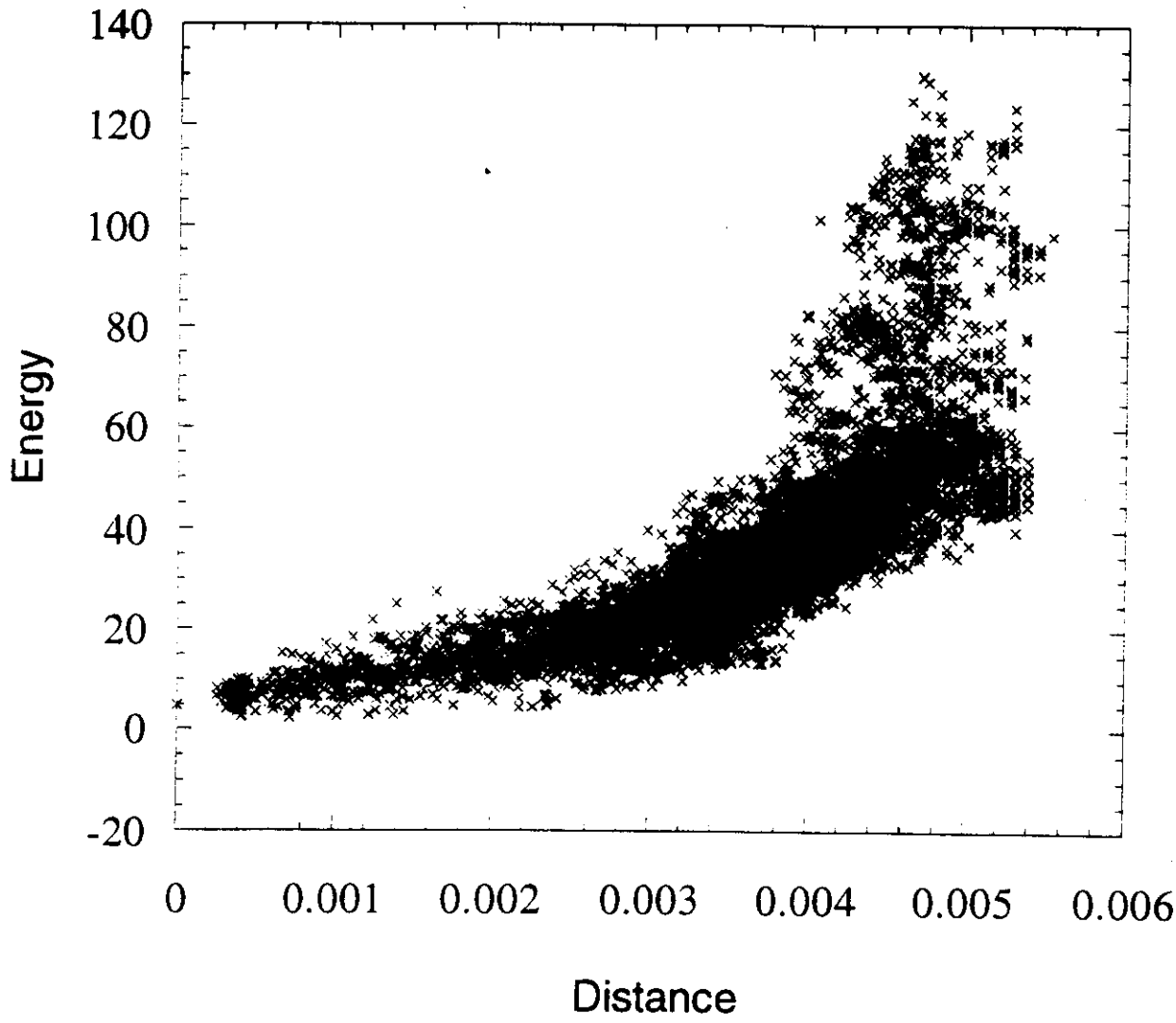
BPTI

MAP AFTER

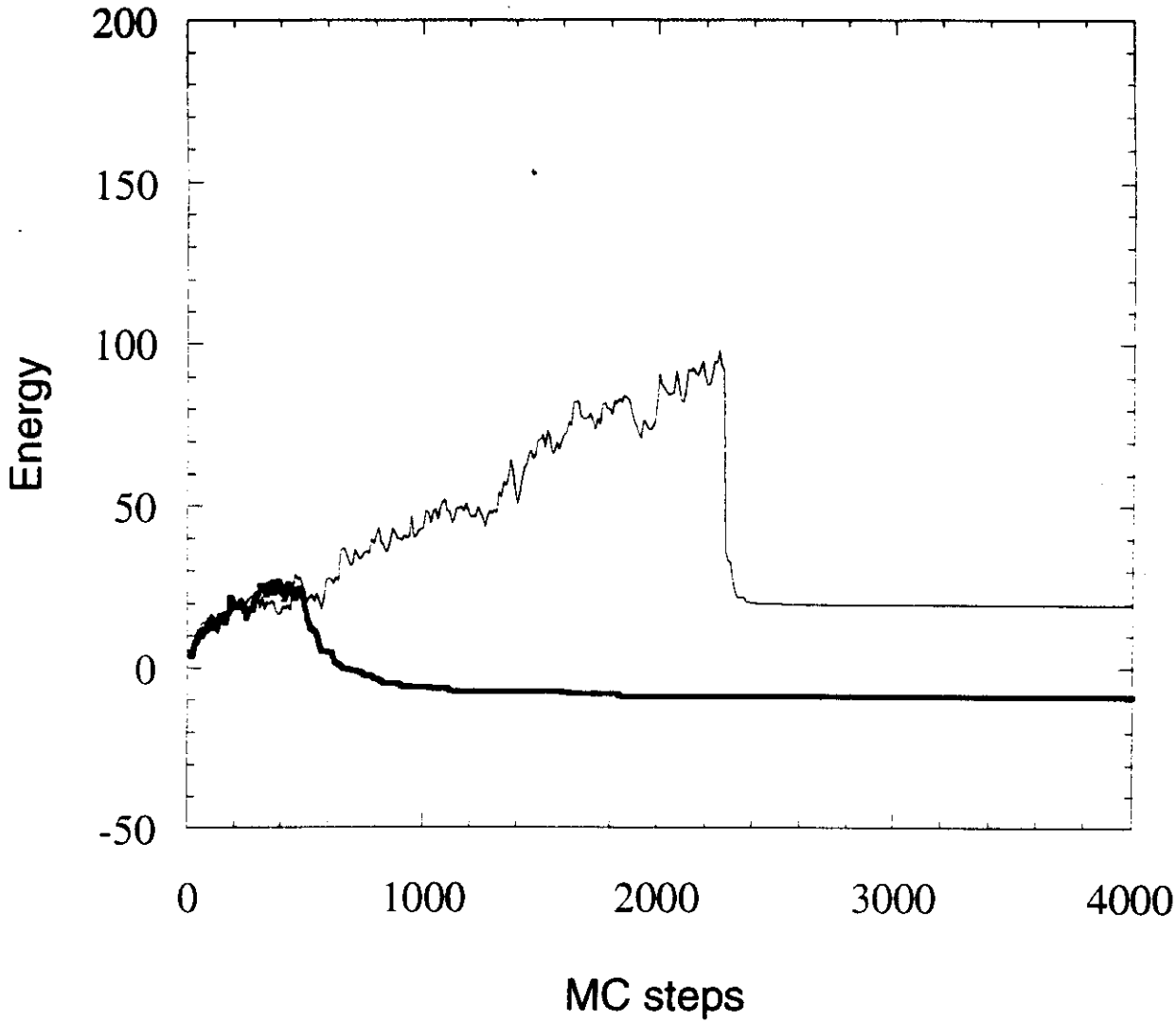
CONSTRAINED

MINIMIZATION

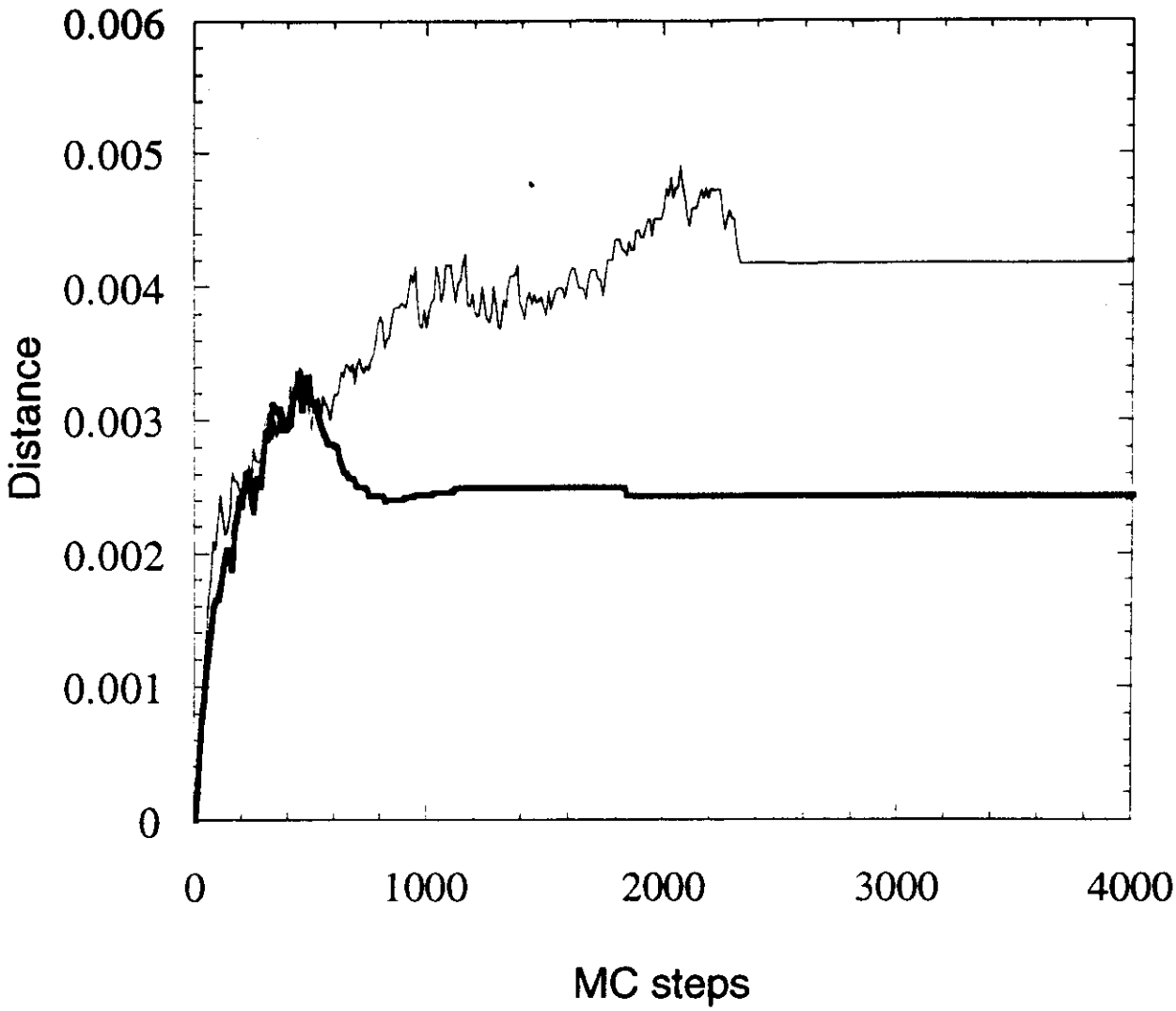




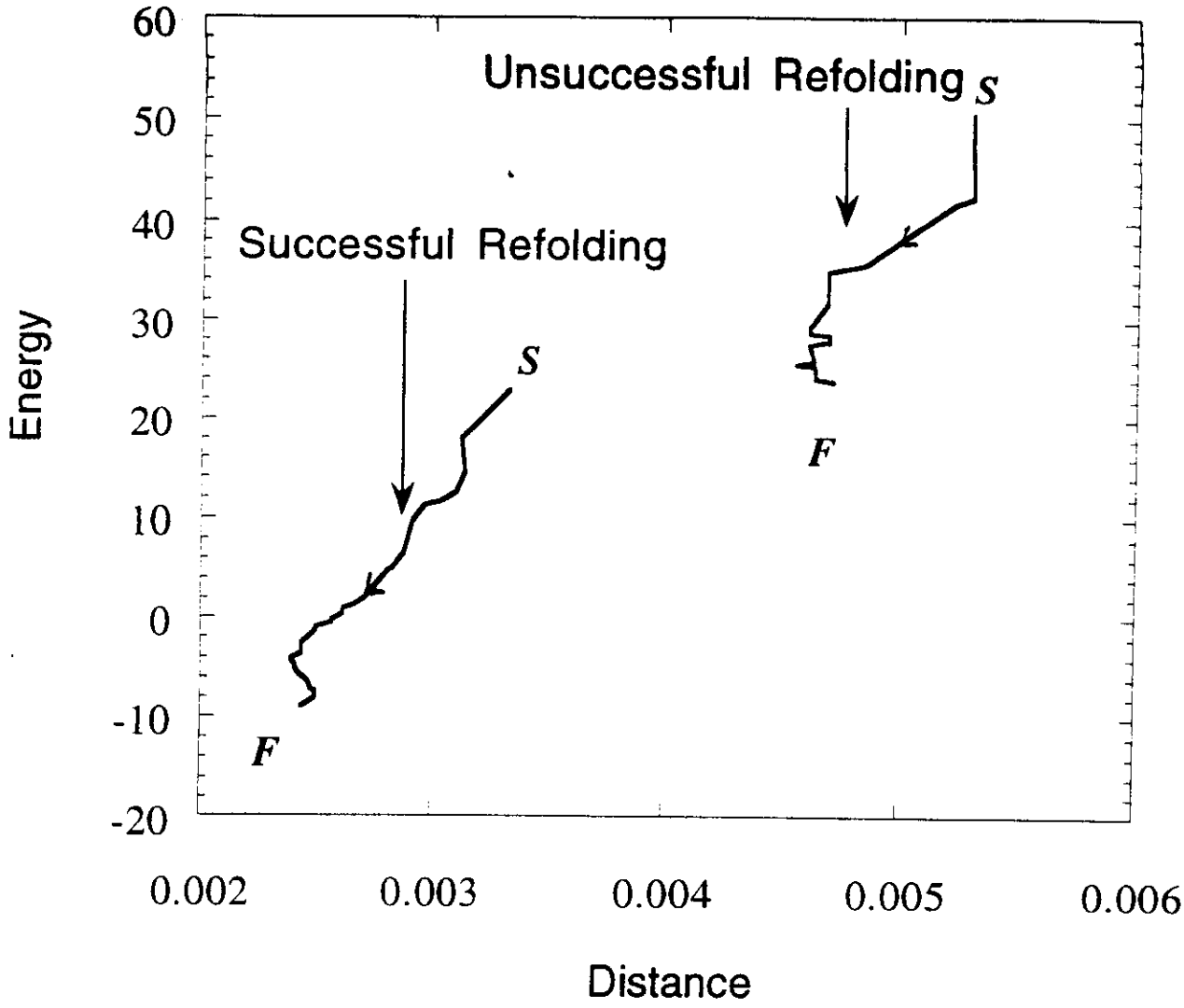
Energy landscape of a protein in the vicinity of the native state. Energy is plotted as a function of distance from the native state when the system is heated in several Monte-Carlo runs.



Refolding of the BPTI contact map after heating at infinite temperature and then slowly cooling. Energy of the contact map during heating and slow cooling. A slightly heated protein was able to refold and return to initial energy values (thick line), whereas cooling of a significantly heated protein did not return the energy to the initial values.



Refolding of the BPTI contact map, after heating at infinite temperature and then slowly cooling. Distance from the native contact map as a function of time during heating and cooling. For small number of heating steps the protein successfully refolded (thick line), whereas it failed to refold for large number of heating steps (thin line). A Monte Carlo step is defined as an attempt to change the state of one contact on a map.



Energy and distance from the native state, shown at a sequence of (Monte Carlo) times for successful and unsuccessful refolding of a protein, starting from two partially unfolded states. The starting state for refolding is denoted by the letter S and the final state - by F.

SUMMARY AND PLANS

1. CONTACT MAPS ARE A USEFUL REDUCED REPRESENTATION OF PROTEIN STRUCTURE.

2. WE FOUND A GOOD ENERGY FUNCTION FOR ALLOWED, PHYSICAL MAPS.

3. HEURISTIC DYNAMIC RULES LEAVE US IN THE SUBSPACE OF PHYSICAL MAPS

4. PROMISING RESULTS

PLANS: 1. PREDICT UNKNOWN CONTACT MAPS

2. BETTER DYNAMIC RULES

3. STUDY DYNAMICS OF FOLDING

REFERENCES

1. Creighton, T.E. Protein Folding. New York: W.H. Freeman and Company. 1993.

2. Levitt, M.: Protein folding. Curr. Opinion Struct. Biol. 1:224-229,1991.

3. Dill K.A.: Folding proteins: finding a needle in a haystack. Curr. Opinion in Struc. Biology 3:99-103,1993.

4. Frauenfelder H., Wolynes P.G. Biomolecules: where the Physics of complexity and simplicity meet. Physics Today, 47: 58-64,1994.

5. Karplus M., Shakhnovich E. Protein Folding: Theoretical Studies of Thermodynamics and Dynamics, in Protein Folding, Creighton, T.E. ed. W.H.Freeman and Co: New York. 1992.

6. Protein Data Bank 1993 Release 67, Brookhaven National Laboratory

7. Kendrew, J.C., Bodo,G., Dintzis,H.M., Parrish, R.G., Wycoff,H. Phillips, D.C. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature, 181: p. 662-666, 1958.

8. Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H. Will, G., North, A.C.T. Structure of haemoglobin. A three-dimentional Fourier synthesis at 5.5 A resolution obtained by X-ray analysis. Nature,185: 416,422,1960.

9. Richardson, J.S. The Anatomy and Taxonomy of Protein Structures. Adv. Prot. Chem., 34: 167-339,1981;

10. Orengo, C. Classification of protein folds. Curr. Opinion in Struc. Biology 4:429-440,1994.

11. Goodsell, D.S., Olson, A.J. Soluble proteins: size, shape and function. TIBS 18: 65-69,1993.

12. Rose, G.D., Wolfenden R. Hydrogen Bonding, Hydrophobicity, Packing and Protein Folding. Ann. Rev. Biomol. Struct. 22: p. 381-415,1993.

13. Barksdale, A.D., Stuehr, J.E. Kinetics of helix-coil transition in aqueous poly(L-glumtamic acid) J.Amer. Chem. Society 94: 3334-3338,1972.

14. Scheraga, H. Calculation of stable conformations of polypeptides, proteins, and protein complexes. Chemica Scripta 29A: 3-13,1989;

15. Sali, A., Shakhnovich, E., Karplus, M. Kinetics of Protein Folding. A lattice model study of the requirements for folding to the native state. J.Mol.Biol. 235: 1614-1636,1994.

16. Anfinsen, C.B. Principles that govern the folding of protein chains. Science 181:223-230,1973.

17. Kim, P., Baldwin R. Intermediates in the folding reactions of small proteins and the mechanism of protein folding. Ann. Rev. Biochem. 51: 459- 489,1990.

18. Karplus, M., Weaver, D.L. Protein folding dynamics: The diffusion- collsion model and experimental data. Protein Science 3: 650-668,1994.

19. Levinthal, C. Are there pathways for protein folding? Chim. Phys. 65: 44-45,1968.

20. Sali, A., Shakhnovich, E. I., Karplus, M. How does a protein fold? Nature 369: 248-251,1994.

21. Brooks, C.L.I., Karplus, M., Pettit, B.M. Proteins: A theoretical Perespective of Dynamics, Structure and Thermodynamics. New York: John Wiley and Sons,1988.

22. Bouzida, D., Kumar, S. and Swendsen, R. H. Efficient Monte Carlo methods for computer simulation of biological moleculePhys.Rev.A45:8894 -8901, 1992.

23. Kolinski, A., Skolnick, J. Monte Carlo Simulations of Protein Folding. I. Lattice Model and Interaction Scheme. Proteins 18:338-352, 1994.

24. Wodak S.J., Rooman M.J. Generating and testing protein folds, Curr. Opinion in Struc. Biology 3:247-259, 1993.

25. Rooman, M.J., Kocher, J.P., Wodak, S.J. Prediction of protein backbone conformation based on 7 structure assignments: influence of local interactions. J.Mol.Biol. 221:961-979, 1991.

26. Wilson, C., Doniach, S. A Computer Model to Dynamically Simulate Protein Folding: Studies with Crambin. Proteins 6: 193-209, 1989.

27. Levitt, M., A Simplified representation of protein conformations for rapid simulations of protein folding. J. Mol. Biol. 104: 59-107,1976.

29. Bowie, J.U., Luthy, R., Eisenberg, D. A. Method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164-170,1991.

30. Lifson, S., Sander, C. Antiparallel and parallel beta-strands differ in amino acid residue preferences. Nature, 282: 109-111,1979.

31. Yee, D.P., Dill, K.A. Families and the structural relatedness among globular proteins. Protein Science 2: 884-899,1993.

32. Godzik, A., Skolnick, J. Sequence-Structure matching in globular proteins: application to supersecondary and thretary structure determination PNAS 89: 12098-12102,1992.

33. Holm, L. Sander, C. Protein Structure Comparison by Alignment of Distance Matrices, J.Mol.Biol. 233: 123-132,1993;

34. Ouzounis, C., Sander, C., Scharf, M., Schneider R. Prediction of Protein Structure by evaluation of sequence-structure fitness. J.Mol.Biol. 232:805-825,1993.

35. Hinds, D. A., Levitt, M. A Lattice Model for Protein Structure Prediction at Low Resolution. Proc. Natl. Acad. Sci. U.S.A. 89, 2536-40, 1992; Hinds, D. A., Levitt, M. Exploring Conformational Space with a Simple Lattice Model for Protein Structure. J. Mol. Biol. 243, 668-82, 1994.

36. Sippl, M.J. Calculation of conformational ensembles from potentials of mean force. J.Mol.Biol. 213:859-883,1990.

37. Chotia, C. Evolution of proteins formed by beta-sheets. J.Mol.Biol. 160: 309- 303, 1982.

38. Havel, T. F., Crippen, G. M., Kuntz, I. D. Effects of Distance Constraints on Macromolecular Conformation II. Simulation of Experimental Results and Theoretical Predictions. Biopolymers 18: 73-81, 1979.

39. Crippen, G. M. A Novel Approach to Calculation of Conformation: Distance Geometry. J. Comput. Phys. 24: 96-100, 1977.

40. Saitoh, S., Nakai, T., Nishikawa, K. A Geometrical Constraint Approach for Reproducing the Native Backbone Conformation of a Protein.

41. Bohr, J., Bohr, H., Brunak, S., Cotterill, R. M., Fredholm, H., Lautrup, B., Petersen, S. B. Protein Structures from Distance Inequalities. J. Mol. Biol. 231: 861-869, 1993.

42. Casari, G., Sippl, M.J. Structure derived hydrophobic potential. J.Mol.Biol. 224:725-732,1992.

43. Shakhnovich, E. I., Gutin, A.M. A new approach to the design of stable proteins. Protein Engineering 6:793-800,1993.

44. Holm, L., Sander, C. The FSSP Database of Structurally Aligned Protein Fold Families. Nucleic Acids Res. 22, 3600-3609, 1994.

45. Orengo, C. A., Flores, T. P., Jones, D. T., Taylor, W. R., Thornton, J. M. Recurring Structural Motifs in Proteins with Different Functions. Current Biology 3, 131-139, 1993.

46. Subbiah S., Laurents D.V., Levitt M. Structural similarity of DNA-binding domains of bacteriophage repressor and the globin core. Curr. Biology 3:141- 148,1993

47. Jaritz, M., Sippl, M. J., Unpublished, Reported in Ref. 48.

48. Sippl, M. J. Knowledge-based Potenetials for Proteins. Cur. Op. Struct. Biol. 5: 229-235, 1995.

49. Huang, E. S., Subbiah, S., Levitt, M. Recognizing Native Folds by Arrangement of Hydrophobic and Polar Residues. J. of Mol. Biol. 252: 709-720, 1995.

50. Keasar, C., Elber, R. Homology as a Tool in Optimization Problems: Structure Determination of 2D Heteropolymers. J. Phys. Chem. 99: 11550-11555, 1995.

51. Sippl, M.J., Weitckus, S. Prediction of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. Proteins 13:258-271,1992.

52. Crippen, G.M. Prediction of protein folding from amino acid sequence over discrete conformational space. Biochemistry 30:4232-4237,1991.

53. Monge, A., Lathrop, E. J., Gunn, J. R., Shenkin, P. S., Friesner, R. A. Computer Modeling of Protein Folding: Conformational and Energetic Analysis of Reduced and Detailed Protein Models. J. Mol. Biol. 247, 995-1012, 1995.

54. Miyazawa, S., Jernigan, R. Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-chemical Approximation. Macromolecules 18, 534-552, 1985.

55. Elber, R., Karplus, M. Multiple conformational states of proteins: molecular dynamics of myoglobin. Science: 235, 318-321,1987.

56. Horovitz, A., Fersht, A.R. Co-operative interactions during protein folding. J.Mol.Biol. 224:733-740,1992.

57. Dill, K.A., Feibig, K.M., Chan, H.S. Cooperativity in protein-folding kinetics. PNAS 90: 1942-1946,1993.

58. Daggett, V., Levitt, M. A model of the molten globule state from molecular dynamics simulation. PNAS 89: 5142-5246,1992.

59. Feller, W. An Introduction to Probability Theory 1968 Wiley, New-York