SUPPORTING INFORMATION.

Diversity of function-related conformational changes in proteins: coordinate uncertainty, fragment rigidity and stability

Alexander A. Rashin,[*‡‖] Abraham H.L. Rashin,[‡§] and Robert L. Jernigan[‖]

*Corresponding author: Tel: 201-836-7960. E-mail: Alexander_Rashin@hotmail.com

[‡]BioChemComp Inc, 543 Sagamore Ave, Teaneck, NJ 07666

[§]Rutgers, The State University of New Jersey

22371 BPO WAY Piscataway, NJ 08854-8123

[‖]LH Baker Center for Bioinformatics and Department of Biochemistry, Biophysics and Molecular Biology, 112 Office and Lab Bldg, Iowa State University, Ames, IA 50011-3020

1) Parametrization of the stability calculations in this work.

An estimate of the free energy ΔG_D^N corresponding to the stability of a protein fragment of molecular weight M can be obtained by using the following expression (42):

$$-\Delta G_D^N = \Delta G_B - T\Delta S_{conf0} + T\Delta S_{S\text{-}S} \quad (1)$$

where ΔG_D^N is protein stability; ΔG_B – is assumed to be proportional to the surface area, B, buried upon fragment's folding; $T\Delta S_{conf0}$ is the loss of conformational entropy of the chain without disulphide crosslinks upon folding and is a function of its molecular weight; $T\Delta S_{S\text{-}S}$ is the decrease in the absolute value of this entropy loss caused by disulphide crosslinks. (Note, that expression 1 should be looked upon just as an empirical formula allowing a reasonable degree of success but leaving aside controversies regarding the mechanism of protein folding (*44,60,61*))

$$\Delta G_B = 0.022 \text{ kcal/(mol•Å}^2\text{)} \times B \quad (2)$$

$$T\Delta S_{S\text{-}S} = 0.9N_{s\text{-}s}(\ln \tilde{n} + 3) \text{ kcal/mol} \quad (3)$$

$$T\Delta S_{conf0} \text{ (kcal/mol)} = \begin{cases} 0.01612M, & M<2320 \\ -16.94+0.02341M, & 2320<M<10370 \\ -60.64+0.02762M, & M>10370 \end{cases} \quad (4)$$

where: M is the molecular weight; $N_{s\text{-}s}$ is the number of disulphide cross-links in a protein; $\tilde{n}=Ns/2N_{s\text{-}s}$, and Ns is the total number of residues between the crosslinks. An N×N table of stabilities of all contiguous fragments of a protein forms a "stability matrix" (SM).

All but two proteins in the set studied here do not contain disulphide cross-links. Thus all calculations of the buried area were initially performed presuming that Cys residue can bury up to its full accessible surface of 144 $Å^2$ (*42*). However, Cys forming S-S bond can bury only 78 $Å^2$ of surface area. Therefore, Cys-Cys bond buries 132 $Å^2$ of surface area less than two fully buried but non-cross-linked Cys (2×144 $Å^2$-2×78 $Å^2$). Thus a destabilizing correction of 2.9 kcal/mol per disulphide crosslink should be added to the results of initial calculations together with the stabilizing contribution from eq. 3.

The estimated standard error σ in $T\Delta S_{conf0}$ is ~1kcal/mol for molecular weights M<2600 and increases to 3-5 kcal/mol for higher M. If the calculated $\Delta G_D^N = -a\sigma$, then the assumption of a Gaussian distribution of the error in $T\Delta S_{conf0}$ leads to the probability, P, that $\Delta G_D^N > 0$ (i.e., that the fragment is actually unstable in its studied conformation):

$$P = 1 - 1/(2\pi)^{1/2} \int_{-a}^{\infty} \exp(-x^2/2)dx \qquad (5)$$

Eq. 5 implies that if a=1 then P is about 16%, and if a=2 then P is about 2%.

2) An example (2tbv) of the process of sequential rigid-body fittings is shown below:

Input1 Input2 Output Residues to fit Reference residues Residues to move

Move #1

2tbvA.pdb 2tbvC.pdb 2tbvA_2tbvC _1.pdb 174-286 174,279 1-286

Reference center of mass: -20.834 47.466 150.334

ToMove center of mass: -8.884 4.850 152.490

difference of above: -11.949 42.616 -2.156

Angle of rotation: 130.265 deg

Axis of rotation: 0.1082 -0.1322 -0.9852

Initial RMSD of regions 174-286: 47.8048733335172

Final RMSD of regions 174-286: 0.0011711880361099

(Note that our simplified method produces almost perfect fit superimposing segments 174-286)

Next move (#2):

Input1 Input2 Output Residues to fit Reference residues Residues to move

2tbvA.pdb 2tbvA_2tbvC_1.pdb 2tbvA_2tbvC_2.pdb 2-164 60,102 1-164

Reference center of mass: -3.603 48.836 128.771

ToMove center of mass: -2.486 45.929 129.943

difference of above: -1.116 2.906 -1.172

Angle of rotation: 21.600 deg

Axis of rotation: 0.4076 0.7247 0.5554

Initial RMSD of regions 2-164: 5.002

Final RMSD of regions 2-164: 0.376

Final RMSDD and Δ are calculated by another program from above file 2tbvA_2tbvC_2.pdb as an input.

3) Supporting figures.

Legends:

Fig. S1. DDM for 2gd11gd1 (glyceraldehyde-3-phosphate dehydrogenase)

Notations: White space means that the absolute value of the distance difference (DD) between the corresponding pair of C^{α}-s in the two structures (e.g., PDB entries) is more than 1Å; black areas means that DD is below 0.5Å, and gray areas - that it is between 0.5Å and 1Å. The white spots indicate that the corresponding DDs are larger than 1Å.

Short thick bars or segments of thin double lines along the tops and sides of the triangular matrices denote positions of helices or β-strands (taken from PDB file).

Neighboring ticks on the top and left are at intervals of twenty residues.

Fig. S2. DDM and CDDM for calmodulin pair 1cll1ctr.

a) DDM 1cll1ctr (for notations see Fig. S1). DDM **1cll1ctr** resembles DDM of 2tbv (Fig. 1a).

b) CDDM 1cll1ctr shows a formation of new long-range contacts (enclosed in little rectangles),

Fig. S3. DDM 1akz1ssp (DNA-uracil glycosylase). For notations see legend to Fig. S1.

Fig. S4. DDM 8adh6adh (alcohol dehydrogenase). For notations see legend to Fig. S1.

Fig. S5. Phospho-glycerate kinase (PGK)

a) DDM 16pk13pk (for notations see legend to Fig. 1)

b) SM 13pk (for notations see legend to Fig. 3c).

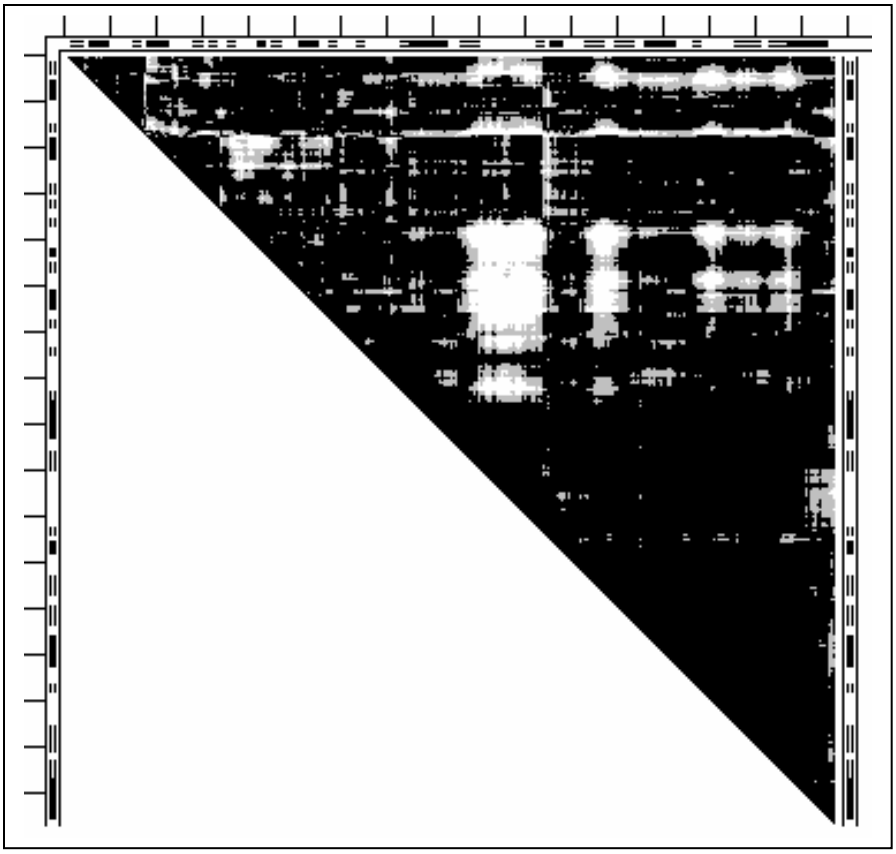


Fig. S1

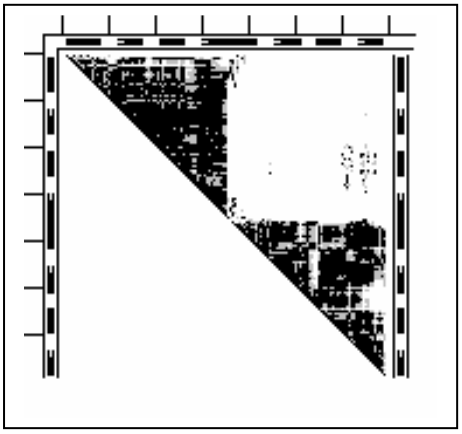


Fig. S2a

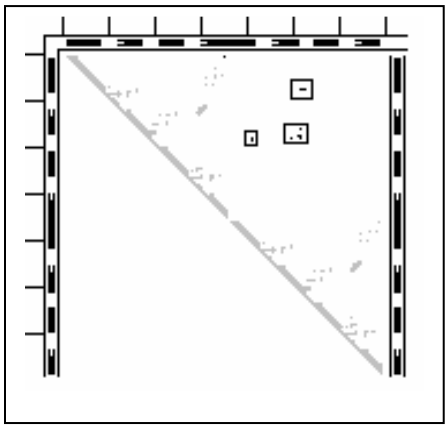


Fig. S2b

Fig. S3.

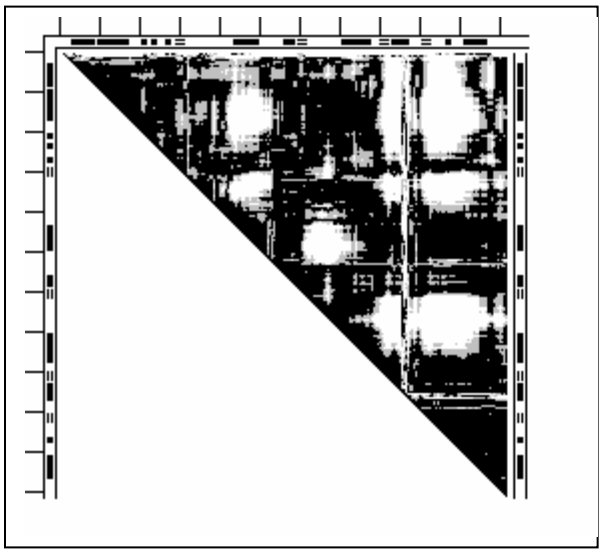
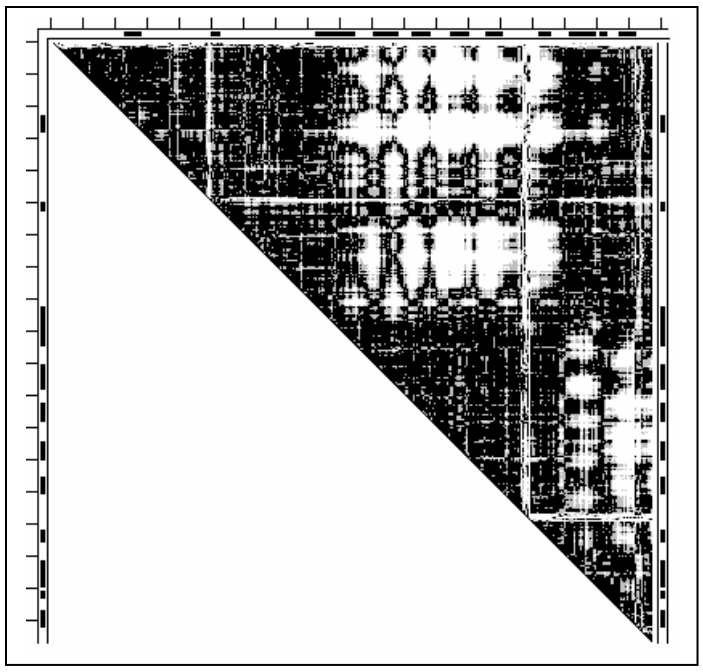


Fig. S4



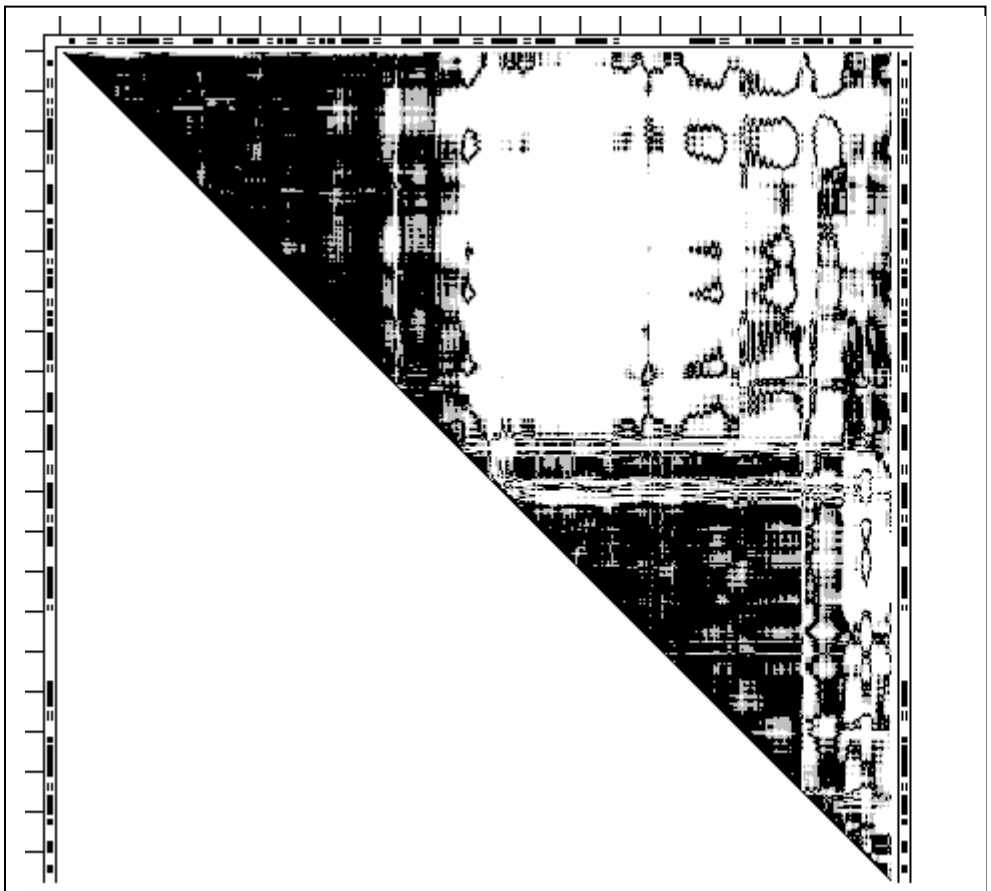


Fig. S5a

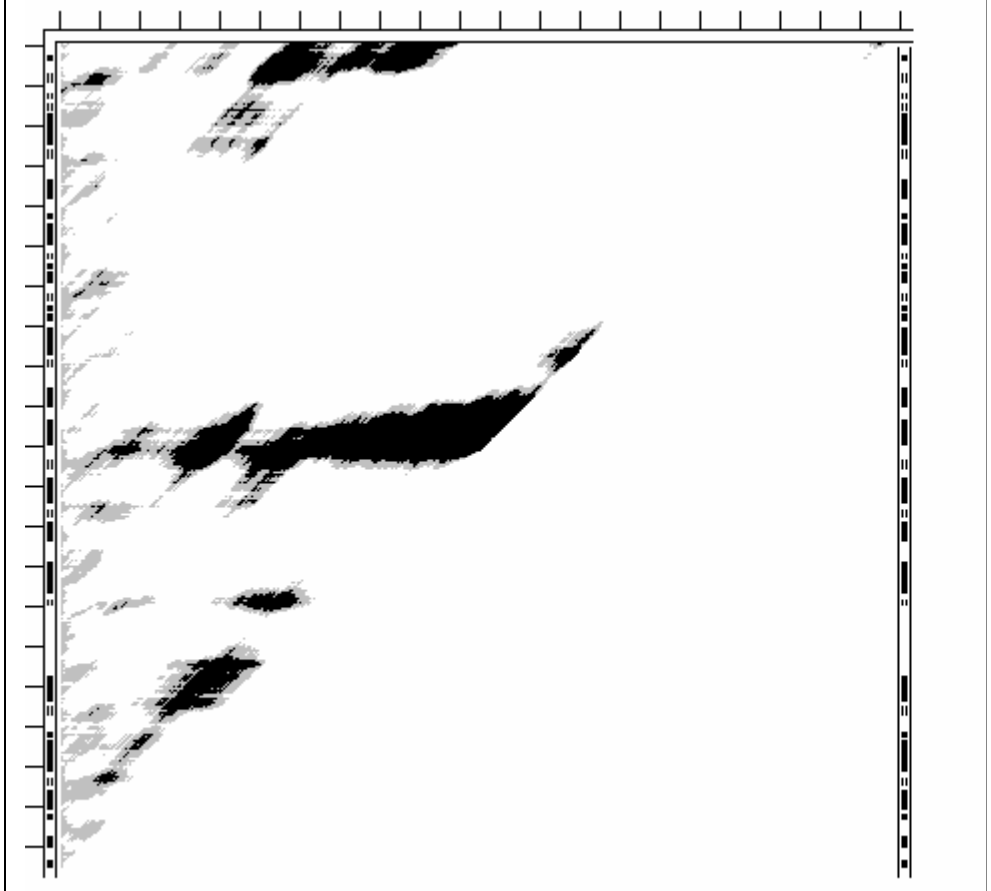


Fig. S5b