SUPPLEMENTARY MATERIAL

PROTEIN FLEXIBILITY: COORDINATE UNCERTAINTIES AND INTERPRETATION OF STRUCTURAL DIFFERENCES

by

Alexander A. Rashin, Abraham H. L. Rashin and Robert L. Jernigan

SM1. An example of CDDM (contact distance difference matrix)

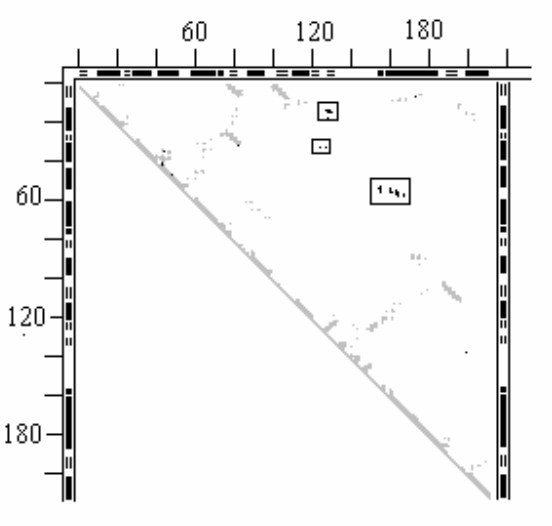

Fig. S1. Bitmap of the CDDM for adenylate kinase (4ake–1ank). Boxed are dark spots indicating formation of new contacts between C^{α}–atoms that moved by over 5 Å. Gray spots indicate contact distance changes of less than 3 Å. Short thick bars or segments of thin double lines along the tops and sides of the triangular matrix denote positions of helices or β-strands (taken from the PDB file). Distances between neighboring ticks on the top and left are at intervals of 20 residues.

SM2. RMSDs for 32 fragments calculated with *SUPERPOSE* and our algorithm

	Fragmetns										
Pair 6LDH-1LDM	3-95	96-104	105-108	109-121	122-215	216-218	219-223	224-305	306-324	233-236	
Our RMSD	0.37	0.99	0.35	0.60	0.57	0.92	0.91	0.61	0.56	0.53	
SuperPose	0.35	1.01	0.32	0.47	0.53	2.10	0.82	0.56	0.46	0.44	
Pair 1AKZ-1SSP	3-7	8-35	36-48	49-63	64-76	77-121	122-132	133-148	149-170	171-173	174-223
Our RMSD	0.50	0.30	0.34	0.31	0.27	0.50	0.36	0.38	0.44	0.03	0.38
SuperPose	0.50	0.26	0.22	0.29	0.26	0.40	0.31	0.35	0.37	0.43	0.33
Pair 1LFH-1LFG	5-86	87-92	93-138	139-142	143-250	251-329	330-337	338-417	418-420	421-423	424-691
Our RMSD	0.45	0.85	0.57	0.74	0.39	0.50	0.41	0.36	0.19	0.37	0.56
SuperPose	0.37	0.80	0.46	1.07	0.39	0.46	0.36	0.34	1.50	0.81	0.54

Better than 0.05 Å agreement is marked in black; between 0.07 and 0.1 Å is marked in blue; that between 0.1 and 0.13 Å is marked in brown; all failures of *SUPERPOSE* larger than 0.32 Å are in red.

SM3. List of the structures used in determining coordinate uncertainty thresholds

The list of RNase monomers is comprised from the PDB file names with identifiers of individual chains from the same asymmetric unit cell, which were treated as individual molecules, and the resolution shown in parenthesis after the PDB name or chains identifiers.

For the myoglobins only the resolution is shown in parenthesis after the PDB name because all myoglobins in this study had one monomer per unit cell.

RNases A:

1AFK (A,B; 1.7Å),. 1AFL (A,B; 1.7Å), 1AFU (A,B; 2.0 Å), 1AQP (2.0Å), 1BEL (1.6Å), 1EOS (A,B; 2.0Å), 1EOW(2.0Å), 1FS3 (1.4Å),. 1JVT (A,B; 2.05Å), 1JVU (A,B; 1.78Å), 1JVV (A,B; 2.2Å), 1QHC (A,B; 1.7Å), 1RBW (1.69Å), 1RBX (1.69Å), 1RCA (1.9Å), 1RCN (2.32Å), 1RNC (1.5Å), 1RND (1.5Å), 1RNM (2.0Å), 1RNN (1.8Å), 1RNQ (2.0Å), 1RNW (1.8Å), 1RNX (1.9Å), 1RNY (2.0Å), 1RNZ (1.9Å), 1ROB (1.6Å), 1RUV (1.25Å), 1XPT (A,B; 1.9Å), 3RN3 (1.45Å), 5RSA (2.0Å), 6RSA (2.0Å), 7RSA (1.26Å), 9RAT (1.5Å).

Myoglobins:

1BZ6 (1.2Å), 1BZP (1.15Å), 1BZR (1.15 Å), 1CQ2 (2.0Å), 1JP6 (2.3Å), 1L2K (1.5 Å), 1MBC (1.5Å), 1MBD (1.4Å), 1MBO (1.6Å), 1SPE (2.0Å), 1VXB (2.0Å), 1VXD (1.7Å), 1VXG (1.7Å), 1YOG (1.65Å), 2MB5 (1.8 Å), 2MYE (1.68Å), 4MBN (2.0Å), 5MBN (2.0Å).

SM4. B-factors, average positional errors and DDs

Fig. S2(*a*)

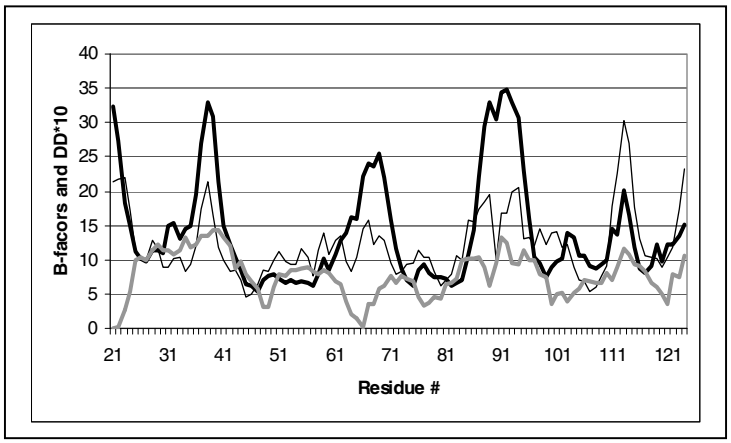

Fig. S2(*b*)

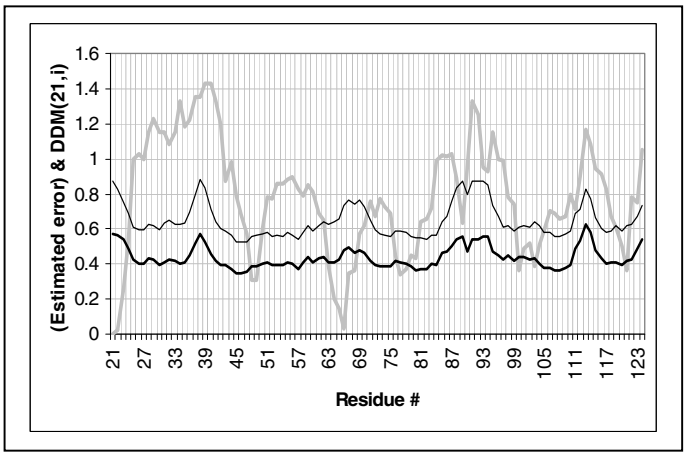

Fig. S2(*c*)

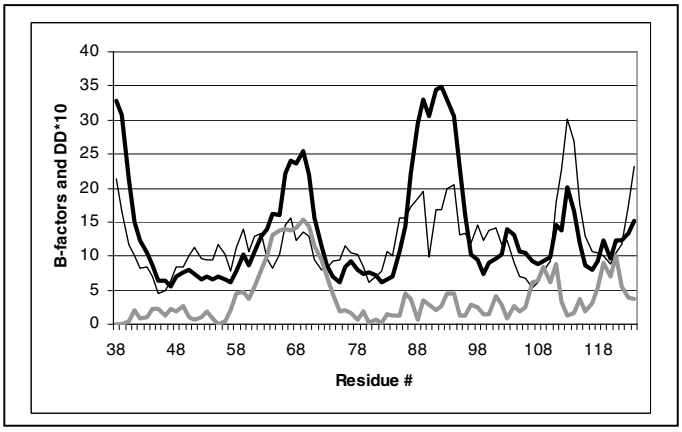

Fig. S2(*d*)

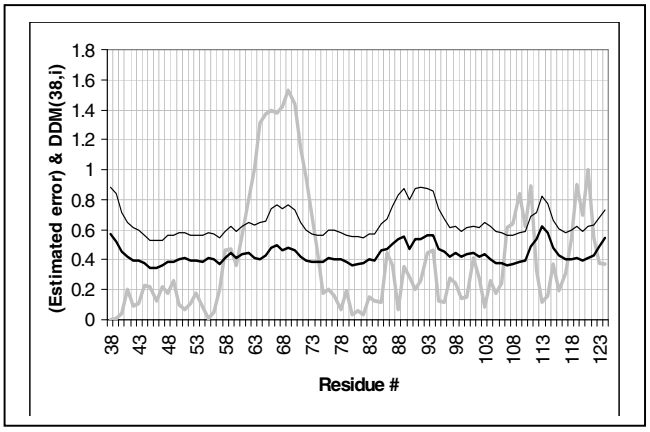

Fig. S2. Comparisons of the absolute values of DDs and B-factors (*a*, *c*) or σ (*b*, *d*). (*a*) Absolute values of DDs from C^{α} of residue 21 to C^{α}-s of all following residues increased 10-fold for comparison (gray); B-factors of C^{α}-s of 1FS3 (thick black) and of 1XPT (thin black). (*b*) – absolute values of DDs from C^{α} of residue 21 to C^{α}-s of all following residues (thick gray); error, σ, estimated from the B-factors according to Eq. 3-4 (thin black using σ_{ave} from Luzzati plots, and thick black using for 1FS3 $\sigma_{ave\ from\ SIGMAA}$). (*c*) and (*d*) – same as a-b but for residue 38 instead of 21.

Inspection of Figs. S2(*a*) and S2(*b*) shows a consistent discrepancy between DD(21-i) and B-factor based characteristics: the deepest minimum in DD(21-i) is at residue i=67 while B-factors for both compared structures 1fs3 and 1xpt have significant maxima at $C^{\alpha 67}$. However, the four other major peaks in DDs agree generally with the areas of higher B-factors. Much more dramatic

discrepancies are apparent between DD(38-i) and B-factor based characteristics (Fig. 2Sc-d): out of three major peaks in DD only one corresponds to the lowest peak of B-factor characteristics.

The PDB mainly provides Luzzati $\sigma_{ave}(r)$ if any at all. It is widely accepted that "the use of Luzzati plots to estimate final errors in protein structures is often badly flawed. The Luzzati method, based preferably on Rfree, can be applied to the low-B atoms in such structures. As the number of observations increases and the resolution improves, the Luzzati $\sigma(r)$ increasingly overestimates the true $\sigma(r)$ of the low-B atoms" (Cruickshank, 1999).

The diffraction-component precision index (DPI), suggested as an improvement to Luzzati's $\sigma_{ave}(r)$ "is not to be regarded as having absolute validity. It is a quick and rough guide" (Cruickshank, 1999). We did not find any DPI $\sigma_{ave}(r)$ values in the PDB entries used in this work.

Another $\sigma_{ave}(r)$ values rarely provided by the PDB is SIGMAA. "This (SIGMAA) is based on the same theory as the Luzzati plot, but with fewer assumptions. Such overall measures of coordinate error should be taken as a rough guide only, because the remaining necessary assumptions are poorly-founded. (R.J. Read, 2005).

Thus there seems to be no really reliable way to determine $\sigma_{ave}(r)$ for the use in eq. 3-4.

However, this might matter only in particular cases. Equations (3) and (4) can be rewritten as

$$\begin{aligned}\sigma(DD_{ij}^{ab}) &= \left[(\sigma_i^a)^2 + (\sigma_j^a)^2 + (\sigma_i^b)^2 + (\sigma_j^b)^2\right]^{1/2} \\ &= \left\{(\sigma_{ave}^a)^2 \times \left[(B_i^a / B_{ave}^a)^2 + (B_j^a / B_{ave}^a)^2\right] + (\sigma_{ave}^b)^2 \times \left[(B_i^b / B_{ave}^b)^2 + (B_j^b / B_{ave}^b)^2\right]\right\}^{1/2}\end{aligned} \qquad \text{(S1)}$$

Thus, (S1) is the square root of a sum of positive terms and no cancellation of one term by another is possible. Let us assume that, as in our Fig. S2, *i* is fixed. Therefore if the expressions in the square brackets have maxima along *j* for both molecules *a* and *b*, then changing σ_{ave} for any one or both molecules can change only the magnitude (and details of the shape) but not the

positions of maxima in $\sigma(DD_{ij}^{ab})$. Different choices of σ_{ave} might change positions of maxima in $\sigma(DD_{ij}^{ab})$if, for example, some maxima along *j* are absent in the square brackets corresponding to one of the molecules in (S1).

This is not the case for data in Fig. S2. The thick black line in Figs. S2(*b*) and S2(*d*) just shifts down from the thin black line when we use low SIGMAA value (0.08Å) for 1FS3 and keep using the Luzzati value for 1XPT (SIGMAA is not available). Different choices of σ_{ave} would not change our conclusions.

SM5. Comparing DDMs after the fitting with three and seven rigid-body motions in 9aat–1ama

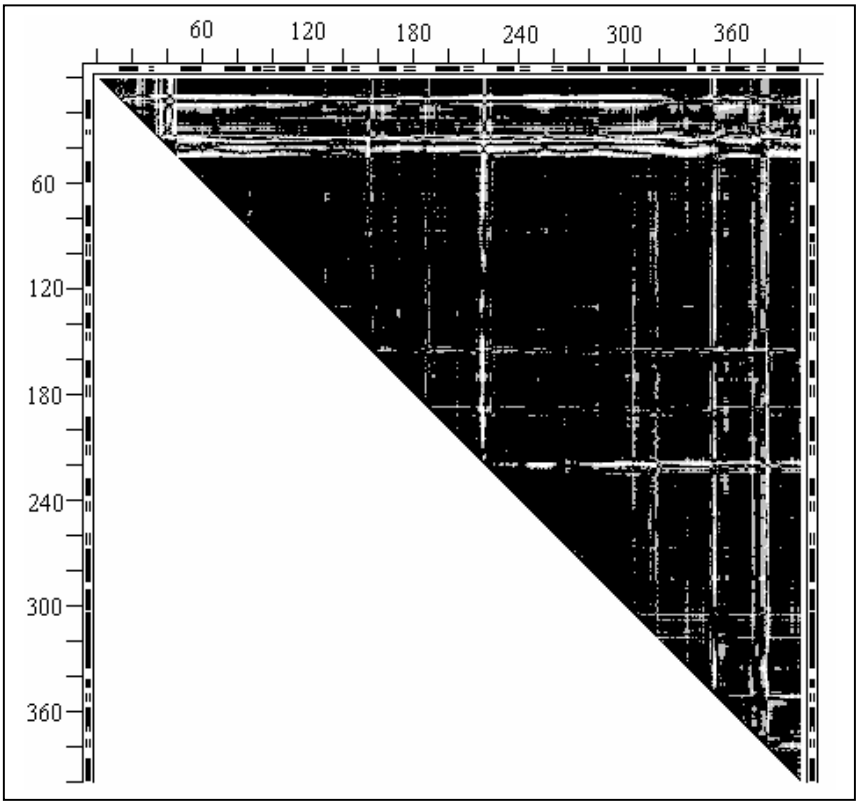

Fig. S3(*a*) Three moves leave much more white mess on the top of the DDM and heavier white lines on the right than seven moves.

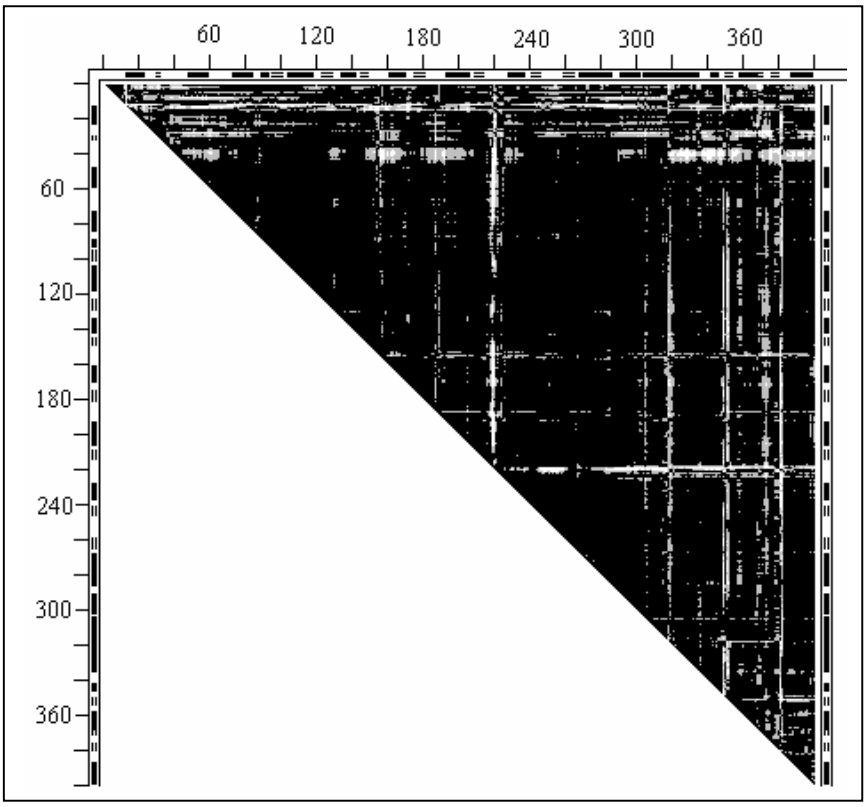

Fig. S3(*b*) – Seven moves lead to less white on the top and right side of the DDM.