

Tables

A)

TERM	Definition or URL Location
Macromolecular Motions Database	http://bioinfo.mbb.yale.edu/MolMovDB Used for classification and annotation of motions in outlier database
SCOP Database	http://scop.mrc-lmb.cam.ac.uk/scop/ Used for classification and annotation of motions via SCOP extension technique.
Wilson et al. set	As shown in Figure 1, a set of 30,000 of SCOP identifier pairs was constructed for Wilson CA, Kreychman J, and Gerstein M (2000), J Mol Biol 297: 233-49. This was then separated into two sets: the 30,000 pair “Wilson et al.” set used in that paper, and the “Full Outlier Set” (described immediately below), which we use in this text. See the caption to Figure 1 for more information.
Full Outlier Set	Text file http://bioinfo.mbb.yale.edu/molmovdb/datasets/outliers.txt Pairs of proteins (SCOP domains) whose structural similarity score was more than two standard deviations above the mean structural similarity for their sequence similarity. See the caption to Figure 1 for more information on the construction of this set.
Workable Outlier Set	This is the subset of the full outlier set on which both morph server processing and normal mode analysis were successful. It consists of 3,814 motion pairs.
Manual Training Set	This is the training set that was produced by examining the SCOP domains in the outlier set for matches against PDB IDs in the set of manually classified motions in the Database of Macromolecular (Gerstein and Krebs (1998) Nuc. Acid. Res., 26(18):4280). Matches received the same classification as in the database, which were determined by manual examination of the scientific literature. Thus, confidence in the accuracy of these classifications is high.
Extended Training Set	The outlier set was searched for pairs that shared the same SCOP fold family as pairs classified in the Manual Training Set; these then received an identical classification. We found empirically that, because proteins which share the same SCOP fold often share similar mechanisms, proteins with the same SCOP fold have a high probability of undergoing similar conformation change and, hence, sharing the same motion size classification. Consequently, these classifications should be accurate but are less reliable than the classifications in the Manual Training Set.
Classified Set	This is simply the entire workable outlier set (minus those already classified in the extended training set) run through the automatic classifier defined by the decision tree which we produced when we analyzed the extended training set.

B)

TERM	Definition
Mode Concentration	This is discussed extensively in the text. It is a simple measure of how much the protein's motion is concentrated into any single low-frequency normal mode.
#CAatoms	Number of C-alpha atoms in the protein
Residuals	This is the Euclidean length of the residual difference between the atomic displacements between protein pairs and the SVD fit of the normal modes to the atomic displacements (in Angstroms)
Norm0	Maximum Value of the SVD displacement vector (unitless)
Norm1	Mean of the SVD displacement vector (unitless)
Norm2	Root-mean-square of the SVD displacement vector (unitless)
Frequency	The frequency in relative units of the normal mode with the highest SVD coefficient.
Ranking Overlap	Rank of the normal mode with the largest overlap (unitless). Overlap is defined in the caption to Figure 2.
Maximum Overlap	Value of the largest overlap (unitless quantity). Overlap is defined in the caption to Figure 2.
Size of 2 nd Core	This is the number of residues in the 2 nd core (the 2ndCoreCAs key in the database). This is typically related to the size of the protein, although in poorly matches protein pairs the number can be less.
Trimmed RMS	This is the trimmed RMS score, as defined in Wilson CA, Kreychman J, and Gerstein M (2000), <i>J Mol Biol</i> 297: 233-49 and Gerstein and Krebs (1998) <i>Nuc. Acid. Res.</i> , 26(18):4280-4290.
Maximum CA Movement	This is the largest movement (in Angstroms) of any residue during the course of the motion, as computed by the Morph Server.
Number of Atoms	This is the number of atoms in the protein as computed by the Morph Server. (Atoms in non-standard amino acids are excluded.) This is a measure of the size of the protein.
Energy of Frames	The Morph Server computes energies for the various intermediate structures. These show a strong relationship to the sequence similarity between the two structures, and are indicator of how "good" a given morph is. The relationship of intermediate energies (energy of 4 th frame, for example) with endpoint frames (energy of 8 th frame, for example) can sometimes provide a rough sense of activation energies.
Translation	In hinge motions, the approximate translation (in Angstroms) the moving domains undergoes in the course of the motion, as automatically computed by the morph server. (This number is also computed for non-hinge motions, where it is less meaningful.)
Hinge Rotation	In hinge motions, the rotation (in degrees) of the moving domain around the screw axis in the course of the motion, as automatically computed by the morph server. (This number tends to be small in non-hinge motions.)
Number of Hinges	The number of putative hinges, or flexible linkages involved in the motion, as determined by the Morph Server
Traditional RMS	This is simply the traditional RMS score between the domains.
Rank of Norm0 Mode	This is a software index that identifies the normal mode contributing the most to the motion as computed within our SVD framework. (The same normal mode that sets norm 0.)

key	#CAatoms	Residuals	Norm1	Norm2	Frequency	Ranking Overlap	Maximum Overlap
mean	220	480	-0.001	540	3.1	2.7	0.0031
std. dev.	110	660	0.051	360	0.89	3.6	0.005
minimum	39	0.23	-0.14	15	4.2E-08	0	4.7E-5
maximum	1000	8800	0.15	2700	8.6	19	0.11
median	210	330	0.00093	520	3.1	1	0.0017

Motion Size	Predicted		Observed			
	Classified Set		Extended set		Manual Set	
	Count	Percent	Count	Percent	Count	Percent
Domain	2165	95%	1549	93%	180	73%
Fragment	94	4%	107	6%	50	20%
Subunit	14	1%	14	1%	15	6%
Totals	2273	100%	1670	100%	245	100%

Database Statistic	Depth in Tree Built upon Extended Set	Depth in Tree Built upon Manual Set
Size of 2 nd Core	1	1
Trimmed RMS	3	2
Maximum CA Movement	5	2
Number of Atoms	4	3
Mode Concentration	6	4
Energy of 2 nd frame	6	4
Translation	4	5
Hinge Rotation (Degrees)	4	6
Number of Hinges		6
Energy of 3 rd frame		6
Norm0 (maximum value)	5	9
Energy of 9 th frame	3	
Number of Residues	5	
Frequency	5	
Residuals	6	
Norm1 (average norm)	6	
Rank of Norm0 Mode	7	
Traditional RMS	8	
Norm2 (Euclidean norm)	8	
Energy of 4 th frame	9	
Energy of 9 th frame	9	
Energy of 8 th frame	13	