Supporting Information.

FACTORS CORRELATING WITH SIGNIFICANT DIFFERENCES BETWEEN X-RAY STRUCTURES OF SPERM WHALE MYOGLOBIN.

Alexander A. Rashin, Marcin J. Domagalski, Michael T. Zimmermann, Wladek Minor, Maksymilian Chruszcz, and Robert L. Jernigan

SECTION S1

Table S1. LIST OF MYOGLOBIN STRUCTURES (set myo291)

101M,102M,**1**03M,**1**04M,**1**05M,**1**06M,**1**07M,**1**08M,**1**09M,**1**10M
111M,**1**12M,1A6G,1A6K,1A6M,1A6N,1ABS,1AJG,1AJH,**1AZI**
1BJE,**1**BVC,**1**BVD,1BZ6,1BZP,1BZR,1CH1,1CH2,1CH3,1CH5
1CH7,1CH9,1CIK,1CIO,1CO8,1CO9,1CP0,1CP5,1CPW,1CQ2
1DO1,1DO3,1DO4,1DO7,1DTI,**1**DTM,1DUK,**1**DUO,**1DWR**,**1DWS**
1DWT,1DXC,1DXD,**1**EBC,1F63,1F65,1F6H,1FCS,1**GJN**,**1**H1X
1HJT,**1HRM**,**1HSY**,**1**IOP,**1**IRC,1J52,**1**JDO,1JP6,1JP8,1JP9
1JPB,1JW8,1L2K,1LTW,1LUE,*1M6C*,*1M6M*,1MBC,1MBD,**1**MBI
1MBN,1MBO,1MCY,*1MDN*,1MGN,1MLF,1MLG,1MLH,1MLJ,1MLK
1MLL,1MLM,1MLN,1MLO,1MLQ,1MLR,1MLS,1MLU,*1MNH*,*1MNI*
1MNJ,*1MNK*,*1MNO*,1MOA,1MOB,1MOC,1MOD,1MTI,1MTJ,1MTK
1MWC,*1MWD*,1MYF,*1MYG*,*1MYH*,*1MYI*,*1MYJ*,1MYM,1MYZ,1MZ0
1N9F,**1**N9H,**1**N9I,**1**N9X,**1**NAZ,1**NPF**,**1NPG**,**1NZ2**,**1NZ3**,**1NZ4**
1NZ5,1O16,1OBM,1OFJ,1OFK,*1PMB*,**1RSE**,1SPE,**1**SWM,1TES
1U7R,1U7S,1UFJ,1UFP,1V9Q,1VXA,1VXB,1VXC,1VXD,1VXE
1VXF,1VXG,1VXH,**1WLA**,1WVP,**1XCH**,*1YCA*,*1YCB*,**1YMA**,**1YMB**
1**YMC**,1YOG,1YOH,1YOI,2BLH,2BLI,2BLJ,2BW9,2BWH,**2**CMM
2D6C,2E2Y,2EKT,2EKU,2EVK,2EVP,**2FRF**,**2FRI**,2**FRJ**,2**FRK**
2G0R,2G0S,2G0V,2G0X,2G0Z,2G10,2G11,2G12,2G14,**2IN4**
2JHO,2MB5,2MBW,2MGA,2MGB,2MGC,2MGD,2MGE,2MGF,2MGG
2MGH,2MGI,2MGJ,2MGK,2MGL,2MGM,**2**MYA,**2**MYB,**2**MYC,**2**MYD
2MYE,**2NSR**,**2NSS**,**2O58**,**2O5B**,**2O5L**,**2O5M**,**2O5O**,2**O5Q**,**2O5S**
2O5T,2OH8,2OH9,2OHA,2OHB,2SPL,2SPM,2SPN,2SPO,2**V1E**
2**V1F**,2**V1G**,**2V1H**,**2V1I**,**2V1J**,**2V1K**,**2VLX**,**2VLY**,**2VLZ**,2**VM0**
2W6W,2W6X,2W6Y,2Z6S,2Z6T,2ZSN,2ZSO,2ZSP,2ZSQ,2ZSR
2ZSS,2ZST,2ZSX,2ZSY,2ZSZ,2ZT0,2ZT1,2ZT2,2ZT3,2ZT4
3A2G,3ASE,3**BA2**,3E4N,3E55,3E5I,3E5O,3ECL,3ECX,3ECZ
3ED9,3EDA,3EDB,3H57,3H58,**3HC9**,**3HEN**,**3HEO**,**3HEP**,3K9Z
3LR7,**3LR9**,3M38,3M39,3M3A,3M3B,**3**MN0,3NML,3OGB,4MBN
5MBN,

Horse, bold-face; pig, italic; whale, plain-weight; mutant, underline. If the PDB contains an unusual ligand, we change the weight of the first character (either bold or plain). Thus, **1**01m is a mutant whale structure containing an unusual ligand and **1azi** is a wild-type horse structure with a usual ligand. Here we consider unusual ligands to be any that are hydroxide, nitric oxide, or cyanide and imidazole derivatives. The list of ligands considered to be unusual (by PDB residue name) is: AZI, NBN, ENC, MNC, NPN, BLA, 4MZ, 1MZ, CYN, OH, NO, IMD

Table S2. HORSE MYOGLOBINS (wild type sequences)

#	PDB ID	Resol (Å)	pH	T (K)	Detail	Salt	State
1	1az1	1.45	7.5	100			azide
2	1dwr	1.45	7.5	100	intermediate		cmonoxy
3	1dws	1.45	7.5	295	1h at photol		cmonoxy
4	1dwt	1.40	7.5	88	Aft photol		cmonoxy
5	1gjn	1.35	5.2	100			hydroxyl
6	1npf	1.9	7.4	100			nitric oxide
7	1npg	1.7	7.4	100	Photo-dis		nitrosoethane
8	1wla	1.7					aquomet?deoxy
9	1ymb	1.9			GH dif whale		met
10	1ymc	2.0			Iron-chlor heme		cyanomet?
11	2frf	1.2	7.4	100	nitrite		nitric oxide
12	2fri	1.6	7.4	100	similar		nitric oxide
13	2frj	1.3	7.4	100	similar		nitric oxide
14	2frk	1.3	7.4	100	similar		nitric oxide
15	2nsr	1.9	7.4	100			nitromethane
16	2nss	2.0	7.4	100			nitrobenzene
17	2o58	1.65	7.4	100	manganese		nitro oxide
18	2o5b	2.0	7.4	100	manganese		?
19	2o5l	1.7	7.4	100	manganese		?
20	2o5m	1.65	7.4	100	manganese		?
21	2o5o	1.6	7.4	100	manganese		?
22	2o5q	1.9	7.4	100	manganese		?
23	2o5s	1.6	7.4	100	manganese		?
24	2o5t	1.6	7.4	100	manganese		?
25	2vih	1.3	5.2	110	radiation prod		met
26	2vm0	1.6	6.8	100	radiation prod		peroxy
27	2vlz	1.3	6.8	110	radiation prod		peroxy
28	2vly	1.6	6.8	100	radiation prod		peroxy
29	2vlz	1.5	6.8	100			peroxy
30	2v1e	1.3	6.8	110	radiation interm		
31	2v1f	1.2	6.8	110	radiation interm		
32	2v1g	1.35	5.2	100	radiation interm		
33	2v1i	1.2	6.8	100	radiation prod		met
34	2v1j	1.4	8.7	100	radiation prod		met
35	2v1k	1.25	6.8	110			deoxy
36	3lr7	1.6	7.4	100	radiation induced		nitrite
37	3lr9	1.55	7.4	100	radiation induced		nitrite
38	1in4	2.15	7.0	100	discharged heme		
39	3ba2	1.8	7.4	100	large heme modif		cyanide

Table S3. SPERM WHALE MYOGLOBINS (set myo46)

#	lab [+]	PDB ID	Resid. in str	Group	Resol (Å)	pH	T (K)	Detail	Salt	State	Refinement (start, method)
1	1	1a6g	151	$P2_1$	1.15	6	90	D122N ?		cmonoxy	1mbc SHELXL-97
2	1	1a6k	151	$P2_1$	1.1	7	100			aquomet	1mbc SHELXL-97
3	1	1a6m	151	$P2_1$	1.0	7	100			oxy	1mbc SHELXL-97
4	1	1a6n	151	$P2_1$	1.15	7	100			deoxy	1mbc SHELXL-97
5	1	1abs	154	P6	1.5	9	20	D122N		cmonoxy	2mgk XPLOR3.1
6	2	1ajg	153	$P2_1$	1.69	6	40	Photo-dis		cmonoxy	1mbo XPLOR
7	2	1ajh	153	$P2_1$	1.69	6	40	Photo-dis		cmonoxy	1krn? XPLOR
8	3	1bz6	153	$P2_1$	1.2	6	287			aquomet	4mbn SHELXL-96
9	3	1bzp	153	$P2_1$	1.17	6	287			deoxy	4mbn SHELXL-96
10	3	1bzr	153	$P2_1$	1.15	5.9	287			cmonoxy	4mbn SHELXL-96
11	4	1cq2	153	$P2_1$	2.0	6.5	298	neutron	NH_4 sulfate	?	X-PLOR 3.1
12	5	1ebc	153	$P2_1$	1.8	7.0	300			cyanide	TNT V. 5-E
13	6	1hjt	153	$P2_1$	1.7		295			nitric ox	1yoi SHELXL-96, X-PLOR 3.851
14	6’	1jp6	152	$P2_1$	2.3	6.0	295			met	1vxh CNS 1.0
15	6’	1jp8	152	$P2_1$	2.3	6.0	295			met	same
16	6’	1jp9	151	$P2_1$	1.7	6.0	95	100Mpa		met	same
17	6’	1jpb	151	$P2_1$	1.7	6.0	77	200Mpa	NH_4 sulfate	met	same
18	6[x]	1jw8	154	P6	1.3	9.0	100	D122N	NH_4 sulfate Tris	cmonoxy	SHELXL-97
19	7	1l2k	151	$P2_1$	1.5	6.8	298	neutron		met	CNS1.0 XPLOR3.1
20	8	1mbc	153	$P2_1$	1.5		260			cmonoxy	PROLSQ
21	4’	1mbd	153	$P2_1$	1.4	8.4		neutron		deoxy?*	JACK-LEVITT?
22	5’	1mbi	153	$P2_1$	2.0					imidazol	TNT
23	9	1mbn	153	$P2_1$	2.0					hydroxide	
24	4*	1mbo	153	$P2_1$	1.6	8.4				oxy	JACK-LEVITT?
25	6*	1spe	153	$P2_1$	2.0	4.0	277			cmonoxy	X-PLOR 3.1
26	5’	1swm	153	$P2_1$	1.8					azyde?	TNT
27	6[x]	1u7r	153	$P2_12_12_1$	1.15	7.0	100			imidazol	SHELXL-97
28	6[x]	1u7s	153	P 6_1 22	1.4	4.5	100	low pH?	Citr. acetate, KCl	deoxy	SHELXL-97
29	6[†]	1vxa	153	$P2_1$	2.0	4.0	277			deoxy	X-PLOR 3.1
30	6[†]	1vxb	153	$P2_1$	2.0	4.0	277			met	X-PLOR 3.1
31	6[†]	1vxc	153	$P2_1$	1.7	5.0	277			cmonoxy	X-PLOR 3.1
32	6[†]	1vxd	153	$P2_1$	1.7	5.0	277			deoxy?	X-PLOR 3.1
33	6[†]	1vxe	153	$P2_1$	1.7	5.0	278			met	X-PLOR 3.1
34	6[†]	1vxf	153	$P2_1$	1.7	6.0	277			cmonoxy	X-PLOR 3.1
35	6[†]	1vxg	153	$P2_1$	1.7	6.0	277			deoxy	X-PLOR 3.1
36	6[†]	1vxh	153	$P2_1$	1.7	6.0	277			met	X-PLOR 3.1
37	6	1yog	153	$P2_1$	1.65		295			deoxy Co	X-PLOR 3.1
38	6	1yoh	153	$P2_1$	1.65		295			met Co	X-PLOR 3.1
39	6	1yoi	153	$P2_1$	1.65		295			oxy Co	X-PLOR 3.1
40	10	2jho	153	$P2_1$	1.4	6.5	100		NH_4 sulfate , phosphate	met cyan	REFMAC 5.2.0005
41	4	2mb5	153	$P2_1$	1.8			neutron	NH_4 sulfate	cmonoxy	PROLSQ
42	6	2mbw	154	P6	1.5	9.0	300	D122N		met	X-PLOR 3.1
43	11	2z6s	153	$P2_1$	1.25	6.0	100		NH_4 sulfate	oxy	SHELXL-97
44	11	2z6t	153	$P2_1$	1.2	6.0	100		NH_4 sulfate	peroxo	2Z6S SHELXL-97
45	12	4mbn	153	$P2_1$	2.0					deoxy	EREF
46	12	5mbn	153	$P2_1$	2.0					hydroxide	EREF

[+] List of labs (by main authors) where the structures were solved is given below

List of groups or their heads who solved each of 46 structures listed in Table S3.

Labs

1. Vojtechovsky, J., Chu, K., Berendzen, J., Sweet, R.M., Schlichting, I.
2. Teng, T.Y., Srajer, V., Moffat, K.
3. Popov, A.N., Kachalova, G.S., Bartunik, H.D.
4. Shu, F., Ramakrishnan, V., Schoenborn, B.P.

4’ Phillips, S.E., Schoenborn, B.P.

4* Phillips, S.E.

5. Rosano, C., Ascenzi, P., Rizzi, M., Losso, R., Bolognesi, M.

5’ Rizzi, M., Ascenzi, P., Coda, A., Brunori, M., Bolognesi, M.

6. Brucker, E.A., Phillips Jr., G.N.;

6’ Urayama, P., Gruner, S.M., Phillips Jr., G.N.

6^{x} Zhang, W., Phillips Jr., G.N.

6* Yang, F., Phillips Jr., G.N.

7. Ostermann, A., Tanaka, I., Engler, N., Niimura, N., Parak, F.G.
8. Kuriyan, J., Petsko, G.A.
9. Watson, H.C., Kendrew, J.C.
10. Arcovito, A., Benfatto, M., Cianci, M., Hasnain, S.S., Nienhaus, K., Nienhaus, G.U., Savino, C., Strange, R.W., Vallone, B., Della Longa, S.
11. Unno, M., Kusama, S., Chen, H., Shaik, S., Ikeda-Saito, M.
12. Takano, T.

SECTION S2. *Crystallographic validation*

Table S2. Myoglobin R-factors validation.

PDB id	R value	R-free value	Structure Factors available	% of Rama-chandran plot outliers	Clashscore percentile	% of rotamer outliers	%of bad bonds	%of bad angles
1a6n	0.123	0.145	Y	0.00	26	0.00	0.00	0.00
1ajh	0.171 (obs.)	0.234	Y	0.00	62	3.20	0.00	0.00
1ebc	N/A	N/A	Y	1.99	87	4.80	0.00	1.96
1jp6	N/A	0.218	Y	0.00	99	0.80	0.00	0.00
1jw8	0.135	0.157	Y	0.00	84	1.59	0.00	0.00
1u7s	0.153 (obs.)	0.179	Y	0.00	81	4.07	0.00	2.61
1yoh	0.164 (obs.)	N/A	N	0.00	34	5.60	0.00	0.65
2mb5	N/A	N/A	Y	0.66	34	14.40	0.00	1.31
2z6s	0.187	0.187	Y	0.00	38	2.44	0.00	1.32
1a6m	0.127	0.159	Y	0.00	37	1.61	0.00	0.66
1abs	0.207 (obs.)	N/A	Y	0.00	11	2.38	0.00	0.00
1ajg	0.171 (obs.)	0.238	Y	0.66	68	4.00	0.00	0.00
1mbc	0.171 (obs.)	N/A	N	0.00	0	5.60	10.46	14.38
1spe	0.173 (obs.)	N/A	Y	0.66	5	8.80	0.00	0.00
1u7r	0.138	0.165	Y	0.00	49	1.65	0.00	1.31
1vxb	0.200 (obs.)	N/A	Y	4.64	0	25.60	0.65	1.31
1vxd	0.153 (obs.)	N/A	Y	0.66	54	3.20	0.00	0.00
1vxe	0.156 (obs.)	N/A	Y	0.66	22	7.20	0.00	0.00
1yog	0.181 (obs.)	N/A	N	0.66	30	3.20	0.00	0.00
1yoi	0.162 (obs.)	N/A	N	0.66	43	4.80	0.00	0.65
2mbw	0.179 (obs.)	N/A	Y	0.00	35	3.97	0.00	0.00
2z6t	0.131	0.185	Y	0.00	86	0.81	0.00	1.32
4mbn	N/A	N/A	N	0.66	33	5.60	0.00	0.65
5mbn	N/A	N/A	Y	0.66	40	5.60	0.65	1.31
1a6k	0.132	0.153	Y	0.00	62	0.81	0.00	1.32
1bz6	0.091 (obs.)	N/A	Y	0.00	73	1.60	0.00	0.00
1hjt	0.166	0.252	Y	0.00	66	5.60	0.00	0.00
1l2k	N/A	0.238	Y	0.00	84	0.00	0.00	0.00
1mbi	0.148 (obs.)	N/A	Y	0.66	68	3.20	0.00	0.00
1swm	0.149 (obs.)	N/A	Y	0.66	35	4.00	0.00	0.00
1vxa	0.177 (obs.)	N/A	Y	1.99	7	20.00	0.00	0.00
1a6g	0.128	0.16	Y	0.00	66	1.61	0.00	1.32
1bzp	0.114 (obs.)	N/A	Y	0.00	81	3.20	0.00	0.00
1bzr	0.124	N/A	Y	0.00	78	1.60	0.00	0.65
1cq2	0.16	0.25	Y	0.00	100	0.80	0.00	0.00
1jp8	N/A	0.236	Y	0.00	98	1.60	0.00	0.66
1jp9	N/A	0.247	Y	0.00	82	0.00	0.00	0.00
1jpb	N/A	0.237	Y	0.00	72	0.00	0.00	0.00
1mbd	N/A	N/A	N	0.66	8	3.20	15.69	24.18
1mbn	N/A	N/A	N	0.66	0	15.20	25.49	20.92
1mbo	N/A	N/A	N	0.66	7	4.00	0.00	1.31
1vxc	0.155 (obs.)	N/A	Y	0.00	41	3.20	0.00	0.00
1vxf	0.146 (obs.)	N/A	Y	0.00	34	5.60	0.00	0.00
1vxg	0.157 (obs.)	N/A	Y	0.66	37	4.80	0.00	0.00
1vxh	0.14 (obs.)	N/A	Y	0.66	54	2.40	0.00	0.00
2jho	0.181 (obs.)	0.234	N	0.00	18	1.60	0.65	0.65

SPECIFIC WARNINGS.

The following crystal structures had specific warnings in crystallographic validation.

105m: poor R-free; error in structure factor file; poor clash-score;

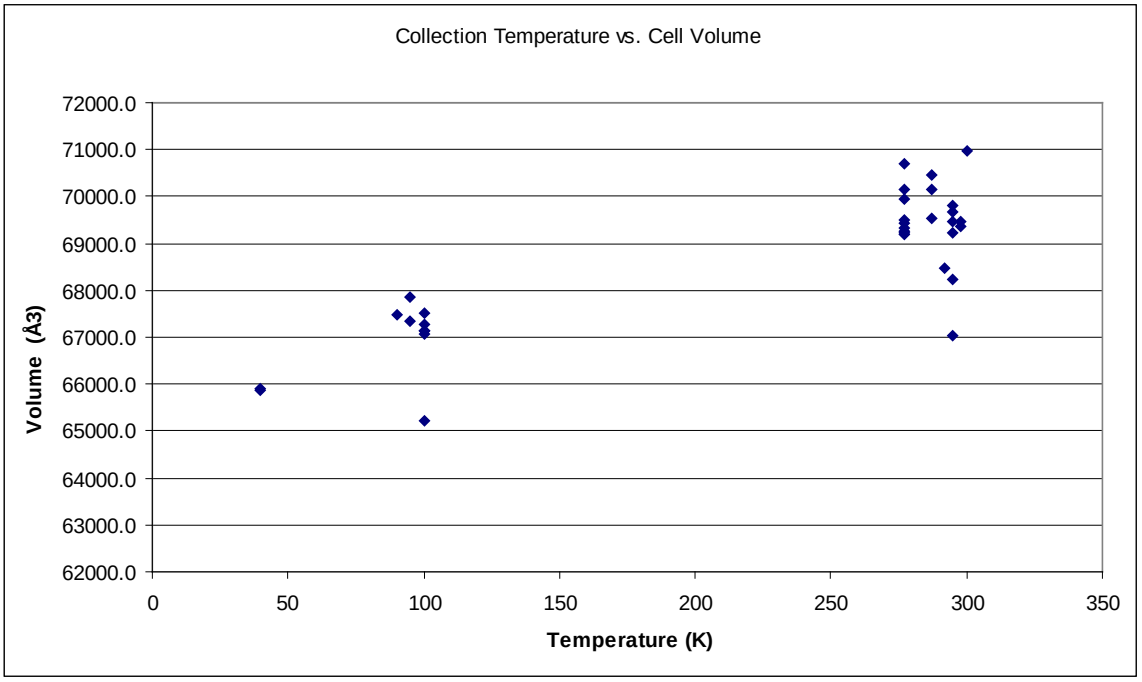
1ebc: R-factors not reported; Ramachandran outliers and weak density in 121-122 region;

1hjt: big difference between R-factors; many poor rotamers;

1abs: no R-free; poor clash-score; problem with structure factors;

1mbn: no detailed structural info; no refinement details; very bad stereochemistry; included below as the first myoglobin structure; because of bad geometry of this historical structure we did not include it in a detailed bonding analysis.

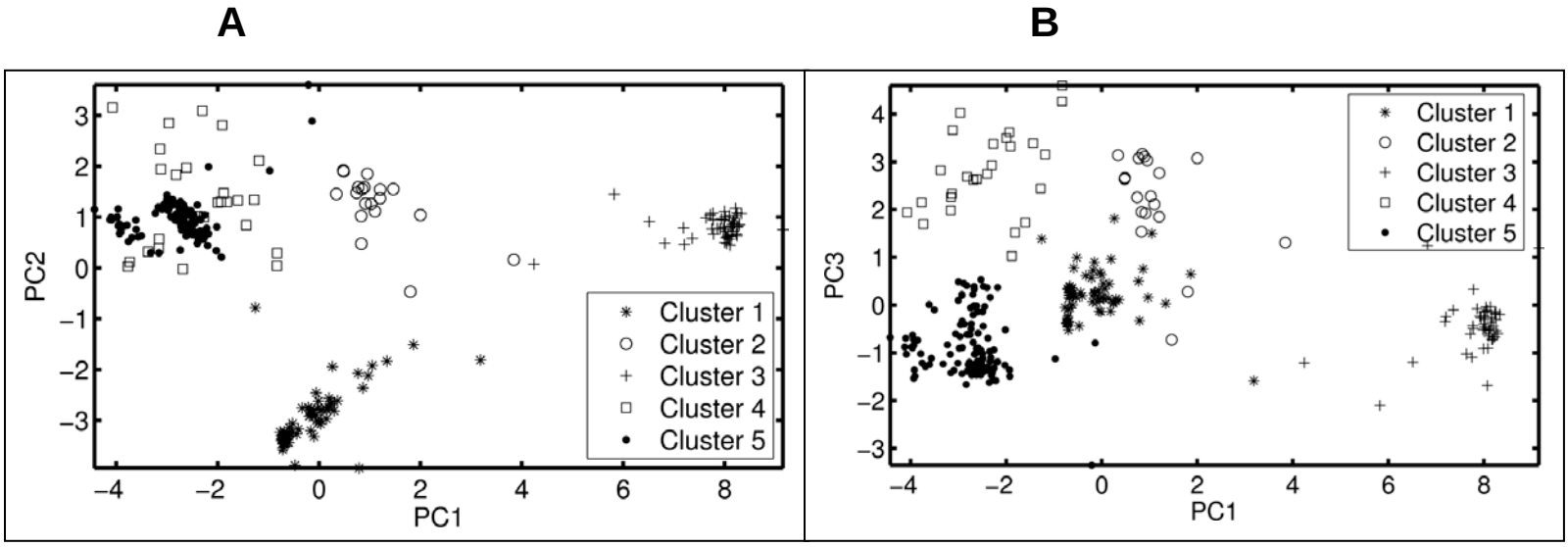
Figure S1. Temperature dependence of characteristics of myoglobin crystals.



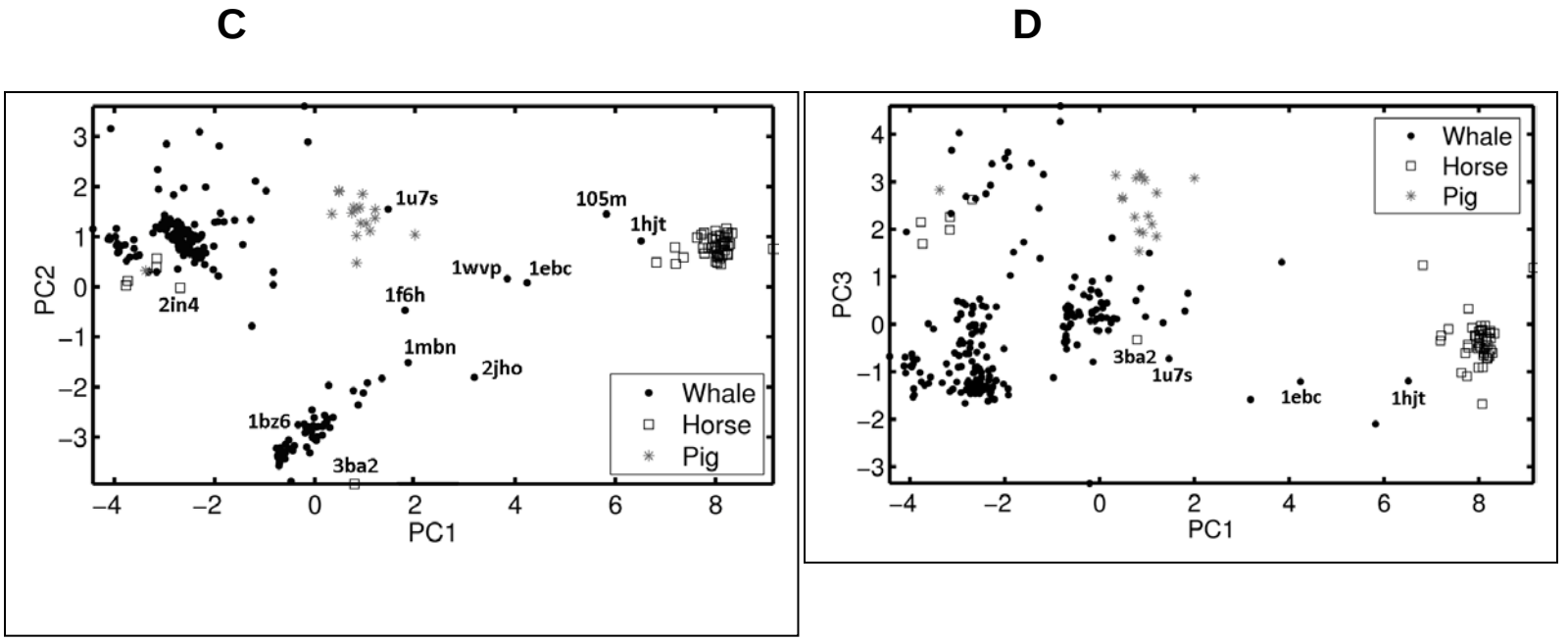
It might be worth noting that CCP4i rather often lists short contacts of the types: E85OE2-D126OD2 (2.9Å, 1jpb), E136OE2-Q152O (2.6Å, 2jho), N122OD1-E18OE2 (3.0Å, 1jw8). There are no alternative conformations in PDB files for these atoms and they cannot form hydrogen bonds because they have no attached hydrogens. The last case can be explained by the erroneous assignment of OD1 and ND2 in N side chain. Two other cases evade this simple explanation, and large changes in pKs of the involved side chains of Glu and Asp seem unlikely. In the case of alternative conformations, *Contact* from the CCP4 package might need a modification to avoid confusion. All of these cases ask for simple structure validation tests.

SECTION S3 *PCA and clustering in myo sets.*

Figure S2. PCA Clusters.



PCA clustering of myo291 set in three Principal Component dimensions.
A: projection on PC1/PC2 plane; B: projection on PC1/PC3 plane.



C and D are the same as A and B but structures are marked according to species.
Some of the outliers have their PDB names shown.

PCA and clustering of a large set of 291 myoglobin structures reveals distinct conformational clusters with high conformational consistency within each cluster. Representation of the results with five clusters (see Fig. S2A) was found to be the most effective. Clusters 1, 3, and 5 contain the majority of structures. The clusters have a high correspondence with species-specific variations. Comparison of Fig. S2A and Fig1 shows that all but one pig myoglobins are in cluster 2, and horse myoglobins dominate cluster 3. However, there are three sperm whale structures in cluster 2, three sperm whale structures in cluster 3, and five horse myoglobin structures in cluster 4. Sperm whale is represented by three clusters (1, 4 and 5) two of which (1 and 5) are dominant and clearly separated in PCA space. Cluster 1 is comprised of mostly wild-type sperm whale myoglobin structures. Cluster 5 contains mostly mutant sperm whale myoglobins, while cluster 4 is again sperm whale myoglobins, but these mostly contain unusual ligands (see Methods, section 2.5).

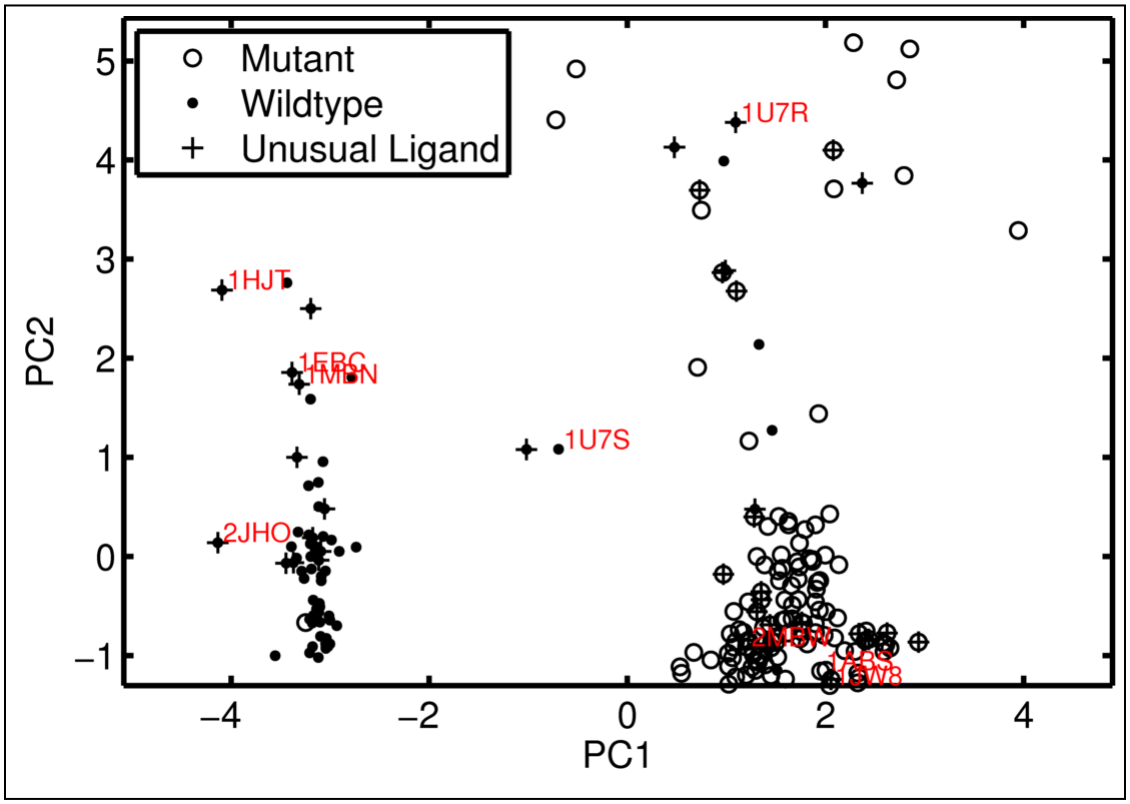
Due to clusters 4 and 5 overlapping in the projection on the PC1/PC2 plane, we also present clusters projection on the PC1/PC3 plane (Fig. S1B). In the PC1/PC3 plane these clusters 4 and 5 separate but clusters 1, 2, 4 and 5, while separated, are rather close to each other. To quantify the pair wise proximity between the clusters in PC space, we calculated ten unique pair wise distances (see Table S3) between the centroids of the five clusters. We give a descriptive identity to each cluster in Table S3, however some exceptions do exist. For instance, there is one mutant in the wild-type cluster of whale myoglobin. Not all of the ten pairs are distinguished by significantly different pair wise distances. In addition there are outliers of each cluster. The difference in the mutual orientation of the clusters in the P1/P2 plane versus the P1/P3 plane reflects the generally observed dependence of PCA results on changes in the set size and in particular on an the exclusion of outliers.

Table S3. Distances between myoglobin clusters in PCA space.

Cluster # and Identity	Cluster # and Identity	Centroid Distance
4 Whale Unusual Ligand	5 Whale mutant	3.7
4 Whale Unusual Ligand	2 Pig	3.75
1 Whale wild-type	5 Whale mutant	4.84
1 Whale wild-type	2 Pig	4.87
2 Pig	5 Whale mutant	5.05
1 Whale wild-type	4 Whale Unusual Ligand	5.49
2 Pig	3 Horse	7.05
1 Whale wild-type	3 Horse	8.91
3 Horse	5 Whale mutant	10.65
3 Horse	4 Whale Unusual Ligand	10.76

The centroid of each cluster is computed as the mean position (in PC-space using the first 3 PCs) of all members within the cluster. The distance between clusters is then the Euclidean distance (in PC-space) between cluster centroids.

Figure S3. Myo216 set in PC1/PC2 projection with PDB names of outliers from Myo46 shown in red. Comparison of Figs. 4a & S3 is addressed in the main text.



PCA clustering of 216 whale myoglobin structures (myo216 set) yields two distinct dominant clusters, with multiple structures occupying a diffuse region (see Fig. 1B). One cluster contains most of the wild-type sequences and only one 'mutant' (1a6g with D122N). (We suspect that this mutation, D122N, listed only in CO form of four myoglobin structures, obtained from the same aquomet-Mb, is an artifact of the refinement (Vojtechovsky *et al*., 1999) which used as a starting model a 1.5Å resolution MbCO structure (Kiriyan *et al*., 1986) with no density for terminal atoms of Asp122). The other cluster and most outliers are either mutants or contain unusual ligands. Structures with wild-type sequences that exist in the mutant cluster are 1bvc, 1bvd, 2ekt, 2d6c, 1iop, 1ufp, 1u7r, 2w6w (see end of 3.6) with the first five having modified or substituted hemes, and 1ufp is an apo-form with salophen substituting heme. Cryogenic u7r has is binding an umusual ligand – imidazolium. Two wild-type structures occupy the region of space between the major clusters; 2cmm (with a modified heme) and 1u7s (see Fig. 1B). All of which still evades a simple and unique interpretation. Further subdivision of the myo291 set into smaller subsets and re-computing of PCs on each may elucidate additional relationships.

SECTION S4.

Characterizing ligand binding sites by distance between Fe and NE2 of distal histidine in some whale myoglobins with usual and unusual ligands.

Thorough mutational, crystallographic, and kinetic analysis (Scott *et al*., 2001) have led to the conclusion that "by analogy with the baseball glove, myoglobin 'catches' and then 'holds' incoming ligand molecules long enough to allow for bond formation with the iron atom. Opening of the glove occurs by outward movements of the distal histidine (His64), and the ligands are trapped in the distal pocket … and either bind to the iron atom or escape through the His64 gate. … Net escape (entry) through the interior of wild-type myoglobin is <20-25%. " Thus, the distal histidine serves as, what protein functional movements studies refer to as (e.g. Krebs *et al*., 2003) "an active site lid". A distinction between usual and unusual ligands apparently correlates with structures within or beyond the "uncertainty threshold" (see Introduction, Methods 2.3 and Results 3.1). Therefore it is worthwhile to find a simple characteristic of the geometric difference in the position of the distal histidine in structures with different ligands as a possible "trigger" of significant conformational changes near or far from the ligand binding site. We choose the distance between the heme iron and the NE2 atom of the distal histidine (which can simultaneously bind to usual ligands) as such a characteristic (see Table S4). Table S4 shows that the minimal distance between these functional atoms in the aquomet structures and 1hjt with NO bound is around 4.3Å. In all other structures with usual ligands the distance between these functional atoms is between 4.4Å and 4.75Å. The upper limit in of 4.75Å might reflect a mixture of three side chain conformations (A,B,C) that are observed in the cryogenic carbonmonoxy 1a6g, but could not be separated in the room temperature structure 1bzr. The maximum distance of 8-8.5Å is seen in structures where the distal histidine is moved away and the ligand binding site cavity is open (1spe and 1a6g C). The largest distance between functional atoms (about 5Å) is seen in the cyanide liganded structures, 1ebc and 2jho, where the binding cavity is closed by the distal histidine. However, the position of the C^{α} atom of the distal histidine in its open

1a6g(A-B) and closed 1a6gC conformations is the same, while the maximum shift of this C^{α} atom is about 0.5Å relative to its position in 1bz6 and it is found only in cyanide-liganded 1ebc (in DDM 1bz61ebc). Contrary to what is typically observed in active site lid movements of most proteins (Rashin *et al*., 2009, 2010), myoglobin lid movements of the distal histidine are limited to its side-chain and do not involve any large repositioning of its main chain. Due to this limitation, functional movement, when considering C^{α} solely, is undetectable by any method, including PCA.

Table S4.
Distances between heme iron and NE2 atom of distal His in whale myoglobins.

PDB name[+]	State/ligand	T/K	Resolution/Å	Distance Fe-H64NE2
1mbo	oxy	not shown	1.6	4.52
1bzp	deoxy	287	1.17	4.61
1a6m A-B	oxy	100	1.0	4.58 - 4.66
1a6n A-B	deoxy	100	1.15	4.40 – 4.49
1z6t A-B	peroxo	100	1.2	4.54 – 4.44
1bzr	cmonoxy	287	1.15	4.77
1a6g A-B	cmonoxy	90	1.15	4.42 - 4.48
1ebc	CN-	300	1.8	5.05
2jho	CN- met	100	1.4	4.93
1a6g C	cmonoxy	90	1.15	8.00
1spe (pH=4)	cmonoxy	277	2.0	8.47
1vxb (pH=4)	met	277	2.0	4.65
1bz6	aquemet	287	1.2	4.38
1a6k	aquemet	100	1.1	4.30
1hjt	NO	295	1.7	4.34

[+] A, B and C after a PDB name refers to alternative conformations found in high resolution structures. Note that for DDMs we used only conformation A, because coordinate differences between alternative conformations are usually in the side chains (only with a few minor exceptions) and not C^{α} atoms used for DDMs.

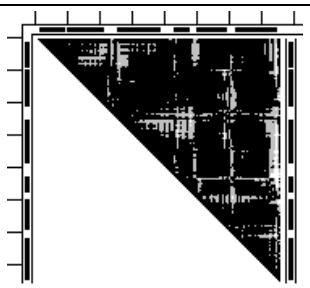
SECTION S5. *Estimation of positional uncertainties.*

Table S5. Whale myoglobin pairs with RMSDD over 0.45 for 151 residues and RMSDD(-2N) after excluding from calculation two N-terminal residues.

PAIR	RMSDD (Å)	RMSDD(-2N) (Å)	PAIR	RMSDD (Å)	RMSDD(-2N) (Å)
1A6G1EBC!!	0.53		1HJT1JP8!!	0.50	
1A6G1HJT!!	0.55		1HJT1JP9!!	0.57	
1A6G1U7R!!	0.46	<0.45	1HJT1JPB!!	0.59	
1A6K1EBC!!	0.51		1HJT1JW8!!	0.63	
1A6K1HJT!!	0.54		1HJT1MBC!!	0.49	
1A6K1U7R!!	0.46	<0.45	1HJT1MBD!!	0.46	
1A6M1EBC!!	0.53		1HJT1MBN!!	0.47	
1A6M1HJT!!	0.55		1HJT1SPE!!	0.50	
1A6M1U7R!!	0.46	<0.45	1HJT1U7R!!	0.64	
1A6N1EBC!!	0.55		1HJT1VXA!!	0.51	
1A6N1HJT!!	0.55		1HJT1VXB!!	0.59	
1A6N1U7R!!	0.46	<0.45	1HJT1VXC!!	0.48	
1ABS1EBC!!	0.59		1HJT1VXD!!	0.49	
1ABS1HJT!!	0.61		1HJT1VXE!!	0.47	
1ABS1MBN!!	0.49		1HJT1VXF!!	0.47	
1ABS1SPE!!	0.47		1HJT1VXG!!	0.48	
1ABS1U7R!!	0.46	<0.45	1HJT1VXH!!	0.47	
1ABS1VXB!!	0.50		1HJT1YOG!!	0.47	
1AJG1EBC!!	0.61		1HJT2MBW!!	0.52	
1AJG1HJT!!	0.60		1HJT2Z6S!!	0.55	
1AJG1SPE!!	0.46	<0.45	1HJT2Z6T!!	0.55	
1AJG1U7R!!	0.51		1JP91U7R!!	0.47	<0.45
1AJG1VXB!!	0.48		1JPB1U7R!!	0.48	<0.45
1AJH1EBC!!	0.60		1JW81MBN!!	0.50	
1AJH1HJT!!	0.60		1JW81SPE!!	0.49	
1AJH1U7R!!	0.51		1JW81U7R!!	0.48	<0.45
1AJH1VXB!!	0.48		1JW81VXA!!	0.46	<0.45
1BZ61HJT!!	0.47		1JW81VXB!!	0.52	
1BZP1HJT!!	0.47		1MBD1U7S!!	0.46	<0.45
1BZP1U7S!!	0.50		1MBN1U7R!!	0.54	
1BZR1U7S!!	0.46	<0.45	1MBN1VXB!	0.50	
1CQ21HJT!!	0.46	<0.45	1SPE1U7R!!	0.47	<0.45
1EBC1JP8!!	0.46		1SPE1U7S!!	0.49	
1EBC1JP9!!	0.53		1SPE2JHO!!	0.47	
1EBC1JPB!!	0.57		1U7R1U7S!!	0.56	
1EBC1JW8!!	0.61		1U7R1VXB!!	0.56	
1EBC1U7R!!	0.61		1U7R2JHO!!	0.59	
1EBC1U7S!!	0.48		1U7R2MB5!!	0.50	
1EBC1VXB!!	0.53		1U7R2Z6S!!	0.46	<0.45
1EBC2MBW!	0.47		1U7R2Z6T!!	0.46	<0.45
1EBC2Z6S!!	0.54		1U7S1VXA!!	0.46	<0.45
1EBC2Z6T!!	0.53		1U7S1VXB!!	0.55	
			1VXA2JHO!!	0.46	<0.45

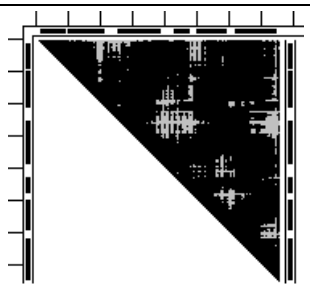
Figure S4. DDMs of myoglobin pairs. Black bars on top/sides – helices; ticks – every 20 residues; black – DDs<0.5Å, gray - <1Å, white - >1Å. 1ebc, 1hjt, 1u7s, 1mbn, 2mb5 and 2jho exhibit bright white L-shaped (Rashin et al., 2009) broken strips corresponding to shifts of the end of G-helix and GH loop by over 1Å. They vary in width with some including only GH loop shifts (1mbn, 2mb5). 1u7r exhibits a similar L-shaped strip corresponding to shifts in the CD region and beginning of E-helix.

1a



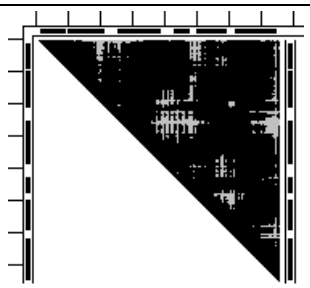
1abs1bz6

1b



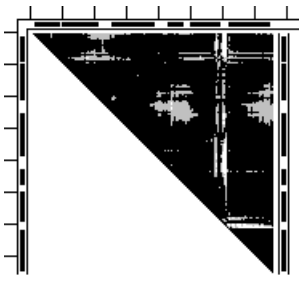
1ajg1bz6

1c



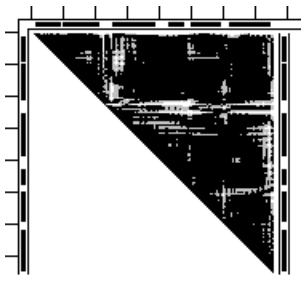
1ajh1bz6

1d



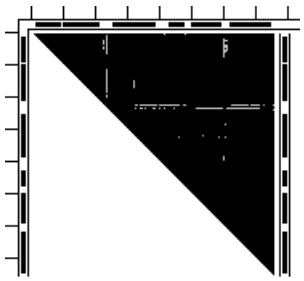
1bz62jho

1e



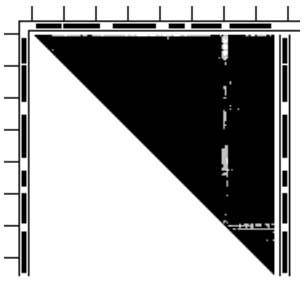
1bz61u7r

1f



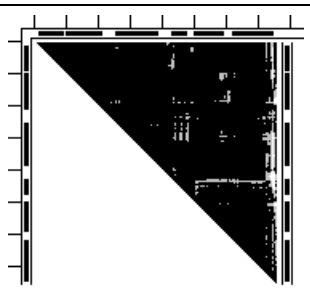
1bz61mbi

1g



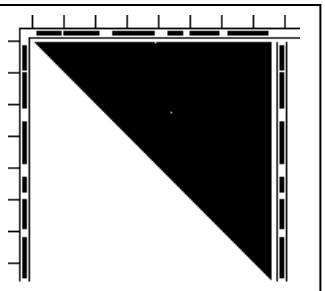
1bz62mb5

1h



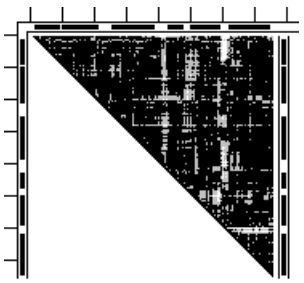
1bz62mbw

1i



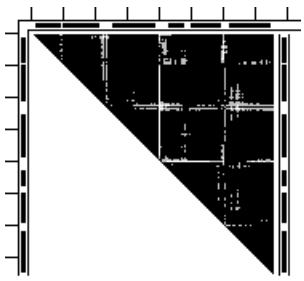
1bz61bzp

1j



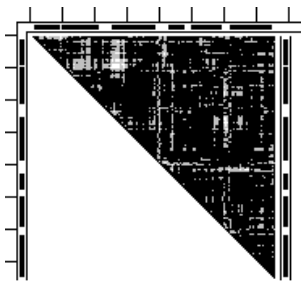
1bz61mbn

1k



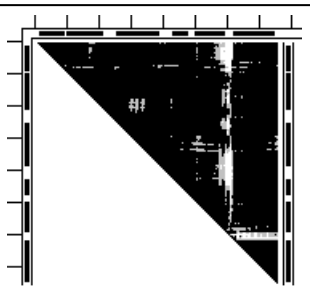
1bz61spe

1l



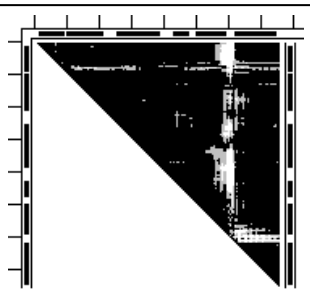
1bz61vxb

1m



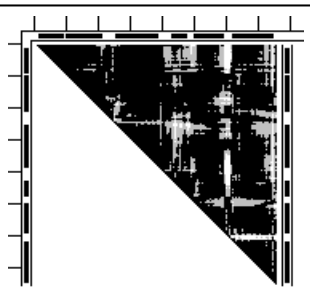
1bz61ebc

1n



1bz61hjt

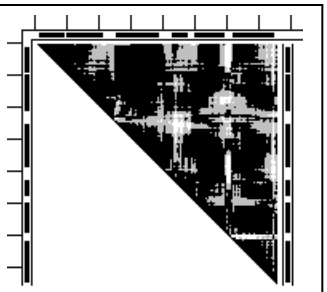
1o



1bz61u7s

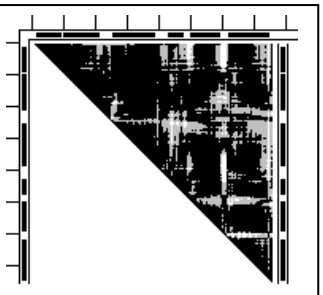
Fig. S5. See Section 3.4. Notations are the same as for Fig. S4.

1



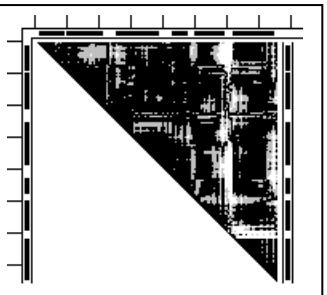
1bzp1u7s

2



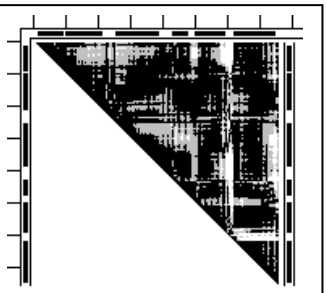
1bzr1u7s

3



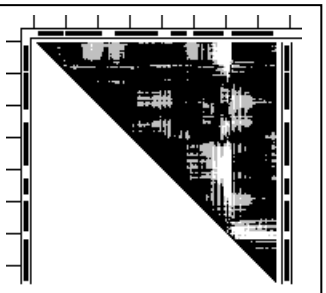
1abs1hjt

4



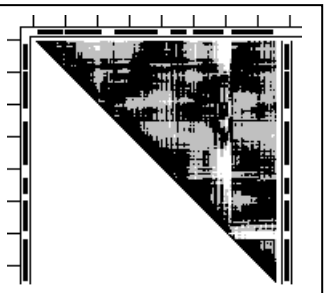
1abs1ebc

5



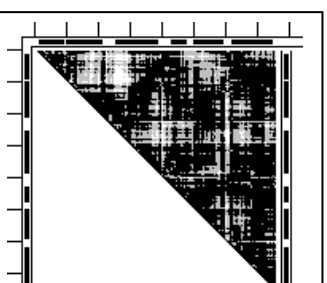
1ajg1hjt

6



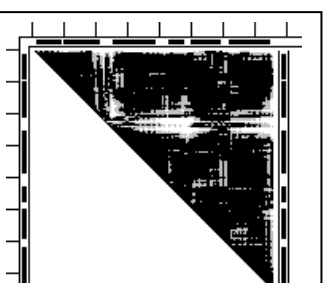
1ajg1ebc

7



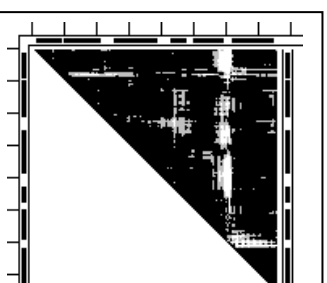
1ajg1vxb

8



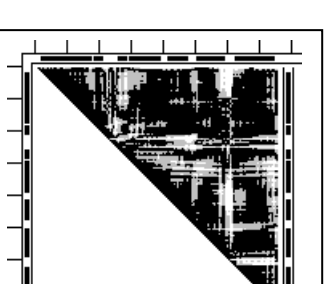
1ajg1u7r

9



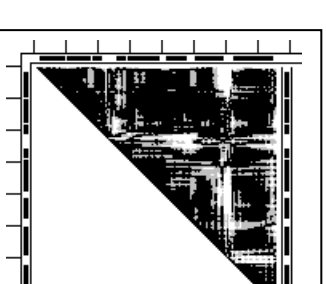
1bzp1hjt

10



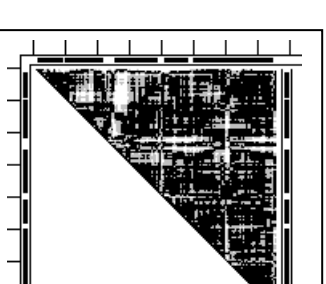
1ebc1u7r

11



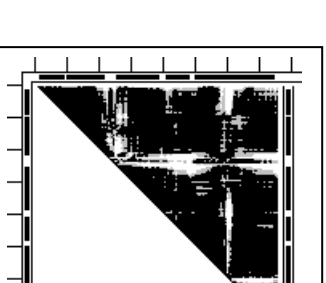
1hjt1u7r

12



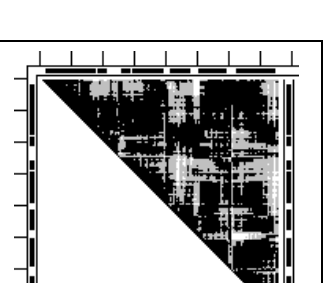
1u7r1vxb

13



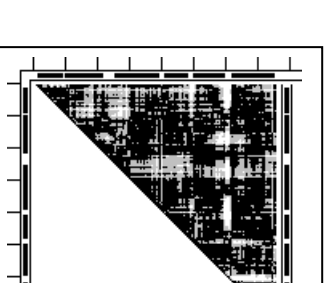
1u7r1u7s

14



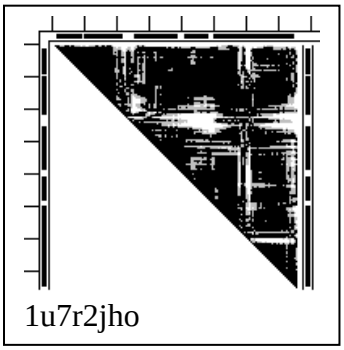
1ebc1u7s

15



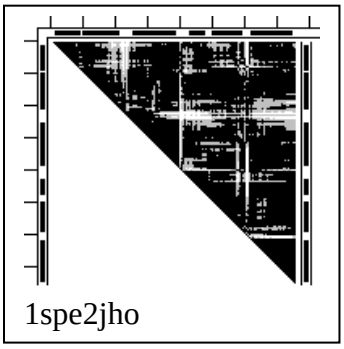
1u7s1vxb

16



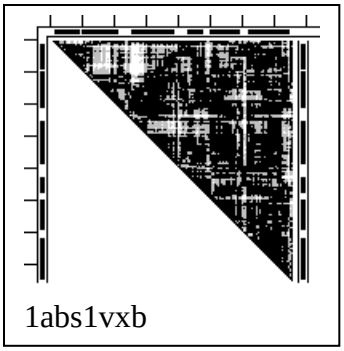
1u7r2jho

17



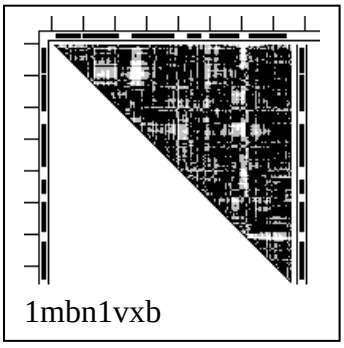
1spe2jho

18



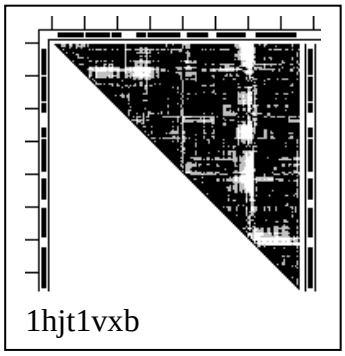
1abs1vxb

19



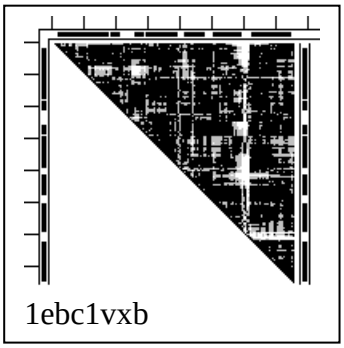
1mbn1vxb

20



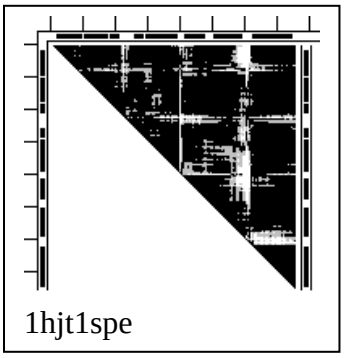
1hjt1vxb

21



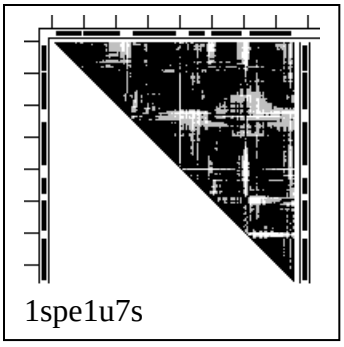
1ebc1vxb

22



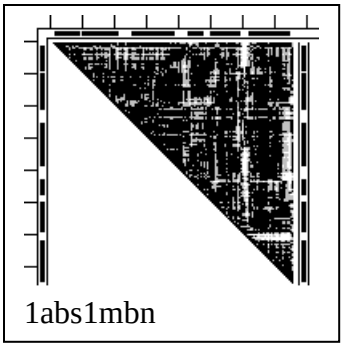
1hjt1spe

23



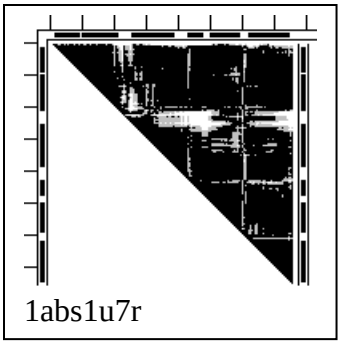
1spe1u7s

24



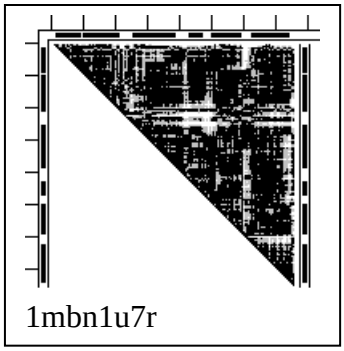
1abs1mbn

25



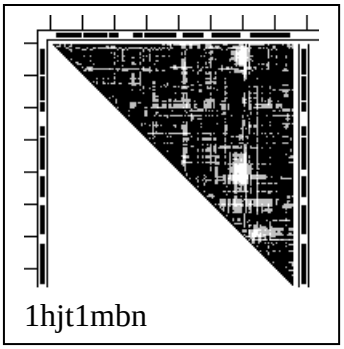
1abs1u7r

26



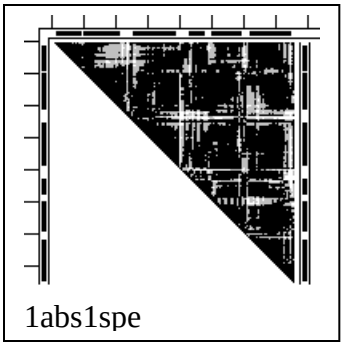
1mbn1u7r

27



1hjt1mbn

28



1abs1spe

Pairs with 1ajh are not shown, as they are identical to pairs with 1ajg (see Fig. S4 1b-c); the same is done with pairs including 1jw8, which are identical to the corresponding pairs including 1abs.

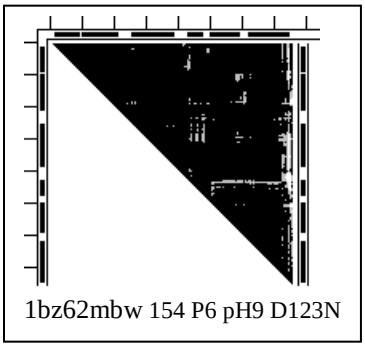
Note that the DDM 1bz61bzp in Fig. S4 1i shows no significant conformational difference. 1bz61bzr and 1bzr1bzp look the same and therefore are not shown. However, first two DDMs in Fig. S5 show that small differences between 1bzp and 1bzr lead to significant differences between 1bzp1u7s and 1bzr1u7s (Fig. S5 1-2). First of these two DDMs has RMSDD above the uncertainty threshold while the second one has it's RMSDD below the threshold (see also Table 2 and Table S5).

A careful observation of Figs. S4 and S5 reveals repeating patterns. All pairs including 1ebc, 1hjt, 1u7s, 1mbn, 2mb5 and 2jho exhibit bright white L-shaped (Rashin *et al*., 2009) broken strips corresponding to shifts of the end of G-helix and GH loop by over 1Å. They vary in width with some including only GH loop shifts (1mbn, 2mb5). 1u7r exhibits a similar L-shaped strip corresponding to shifts in the CD region and beginning of E-helix. 1spe seems to add a spatter of thin broken lines and gray spots, while 1vxb differs from 1spe by less lines, larger spatter of small spots, and a larger white spot near C-D + beginning of E-helix.

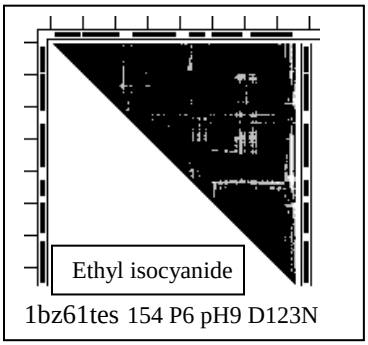
Comparison of all these structures paired in DDMs with the same reference structure 1bz6 (one of the most accurate room-temperature structures) supports the impression that all but a few DDMs in Fig. S5 are mostly superpositions of the reference DDMs (Fig. S4) of structures paired in the DDMs of Fig. S5 (e.g., 16 from Fig. S5 is superposition of d and e from Fig. S4). There is an exception from this "superposition" observation in DDMs 1spe2jho and 1spe1u7s (Fig. S5 17,23). It seems that 1spe weakens the strong GH L-shaped strips of 2jho and 1u7s and adds strong C-D strips to DDMs 1spe2jho and 1spe1u7s, while the reference DDM 1bz61spe itself does not exhibit a strong C-D strip. (This might be due to the fact that DDMs used here represent absolute values of DDs, and thus in superposing reference DDMs the contribution of 1bz6 to reference DDMs is not necessarily cancelled in DDMs of some pairs, e.g. 1spe2jho and 1spe1u7s.)

SECTION S6. ***Factors correlating with crystallization in P6 symmetry.***

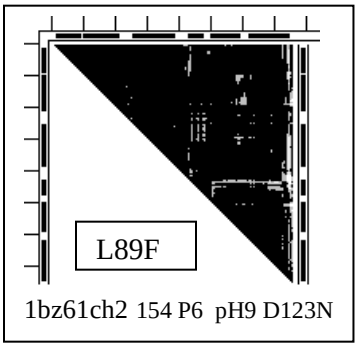
FIGURE S6



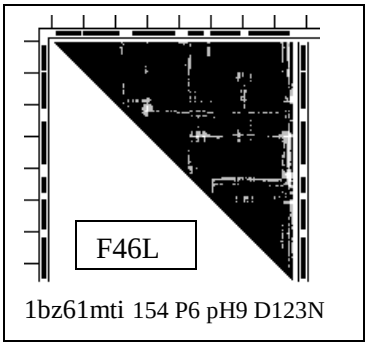
1bz62mbw 154 P6 pH9 D123N



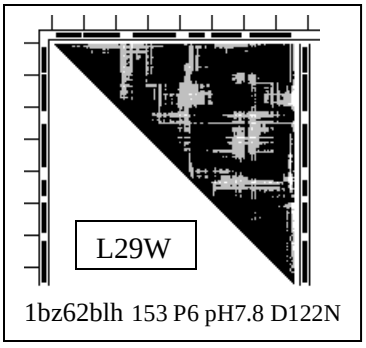
1bz61tes 154 P6 pH9 D123N



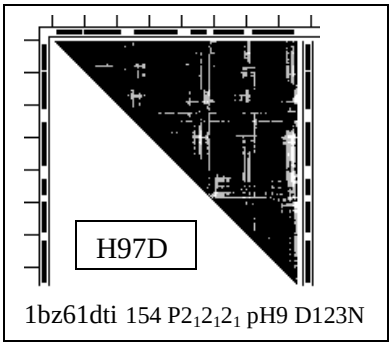
1bz61ch2 154 P6 pH9 D123N



1bz61mti 154 P6 pH9 D123N



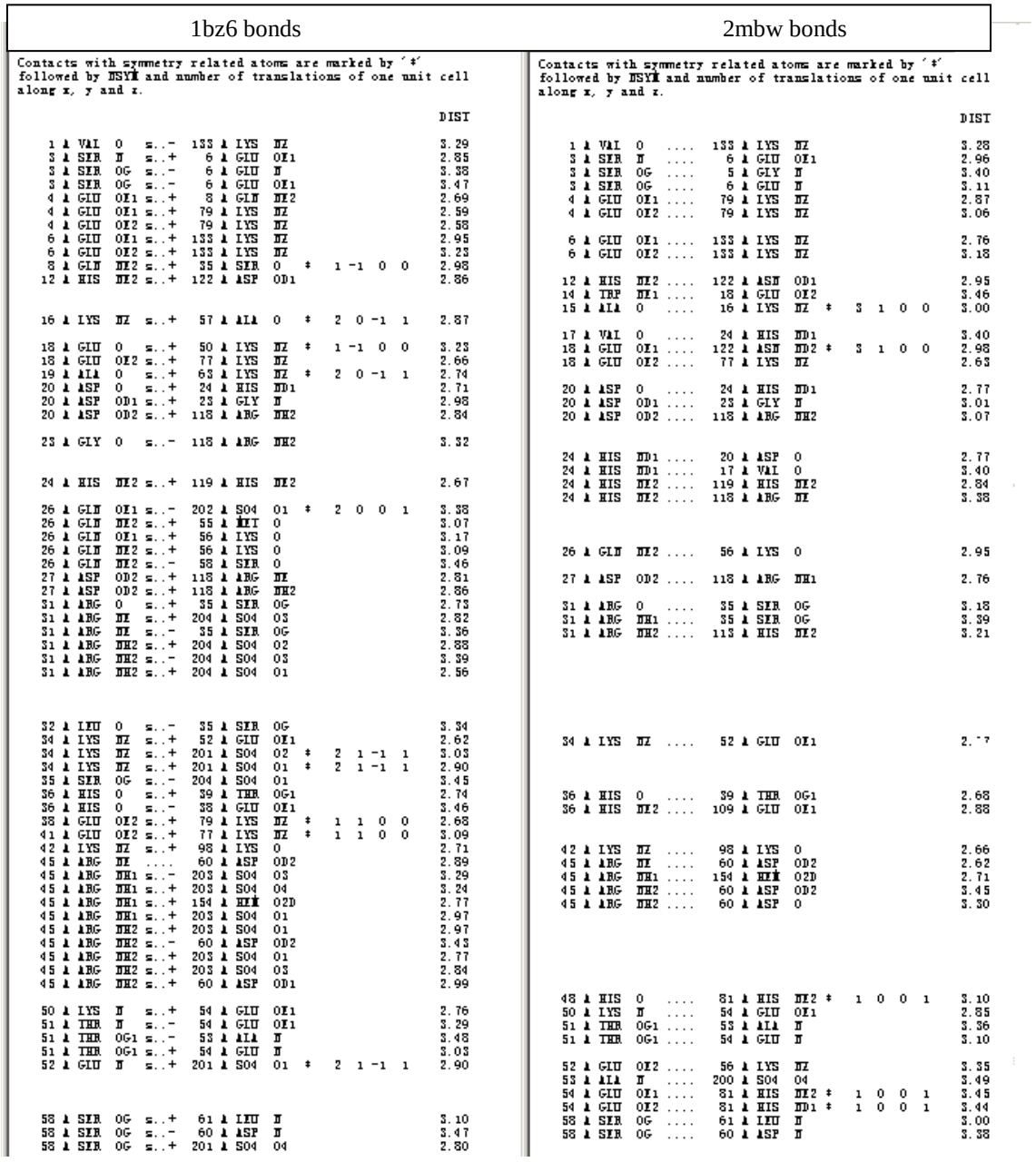
1bz62blh 153 P6 pH7.8 D122N



1bz61dti 154 $P2_12_12_1$ pH9 D123N

SECTION S7. <u>On power of intermolecular crystal bonds to predict main chain conformational differences.</u>

FIGURE S7



1bz6 bonds

Contacts with symmetry related atoms are marked by '*' followed by [SYM] and number of translations of one unit cell along x, y and z.

										DIST
1	A	VAL	O	s..-	133	A	LYS	NZ		3.29
3	A	SER	N	s..+	6	A	GLU	OE1		2.85
3	A	SER	OG	s..-	6	A	GLU	N		3.38
3	A	SER	OG	s..-	6	A	GLU	OE1		3.47
4	A	GLU	OE1	s..+	8	A	GLN	NE2		2.69
4	A	GLU	OE1	s..+	79	A	LYS	NZ		2.59
4	A	GLU	OE2	s..+	79	A	LYS	NZ		2.58
6	A	GLU	OE1	s..+	133	A	LYS	NZ		2.95
6	A	GLU	OE2	s..+	133	A	LYS	NZ		3.23
8	A	GLN	NE2	s..+	35	A	SER	O	* 1 -1 0 0	2.98
12	A	HIS	NE2	s..+	122	A	ASP	OD1		2.86
16	A	LYS	NZ	s..+	57	A	ALA	O	* 2 0 -1 1	2.87
18	A	GLU	O	s..+	50	A	LYS	NZ	* 1 -1 0 0	3.23
18	A	GLU	OE2	s..+	77	A	LYS	NZ		2.66
19	A	ALA	O	s..+	63	A	LYS	NZ	* 2 0 -1 1	2.74
20	A	ASP	O	s..+	24	A	HIS	ND1		2.71
20	A	ASP	OD1	s..+	23	A	GLY	N		2.98
20	A	ASP	OD2	s..+	118	A	ARG	NH2		2.84
23	A	GLY	O	s..-	118	A	ARG	NH2		3.32
24	A	HIS	NE2	s..+	119	A	HIS	NE2		2.67
26	A	GLN	OE1	s..-	202	A	SO4	O1	* 2 0 0 1	3.38
26	A	GLN	NE2	s..+	55	A	[MET]	O		3.07
26	A	GLN	OE1	s..+	56	A	LYS	O		3.17
26	A	GLN	NE2	s..+	56	A	LYS	O		3.09
26	A	GLN	NE2	s..-	58	A	SER	O		3.46
27	A	ASP	OD2	s..+	118	A	ARG	NE		2.81
27	A	ASP	OD2	s..+	118	A	ARG	NH2		2.86
31	A	ARG	O	s..+	35	A	SER	OG		2.73
31	A	ARG	NE	s..+	204	A	SO4	O3		2.82
31	A	ARG	NE	s..-	35	A	SER	OG		3.36
31	A	ARG	NH2	s..+	204	A	SO4	O2		2.88
31	A	ARG	NH2	s..-	204	A	SO4	O3		3.39
31	A	ARG	NH2	s..+	204	A	SO4	O1		2.56
32	A	LEU	O	s..-	35	A	SER	OG		3.34
34	A	LYS	NZ	s..+	52	A	GLU	OE1		2.62
34	A	LYS	NZ	s..+	201	A	SO4	O2	* 2 1 -1 1	3.03
34	A	LYS	NZ	s..+	201	A	SO4	O1	* 2 1 -1 1	2.90
35	A	SER	OG	s..-	204	A	SO4	O1		3.45
36	A	HIS	O	s..+	39	A	THR	OG1		2.74
36	A	HIS	O	s..-	38	A	GLU	OE1		3.46
38	A	GLU	OE2	s..+	79	A	LYS	NZ	* 1 1 0 0	2.68
41	A	GLU	OE2	s..+	77	A	LYS	NZ	* 1 1 0 0	3.09
42	A	LYS	NZ	s..+	98	A	LYS	O		2.71
45	A	ARG	NE	....	60	A	ASP	OD2		2.89
45	A	ARG	NH1	s..-	203	A	SO4	O3		3.29
45	A	ARG	NH1	s..+	203	A	SO4	O4		3.24
45	A	ARG	NH1	s..+	154	A	[HIX]	O2D		2.77
45	A	ARG	NH1	s..+	203	A	SO4	O1		2.97
45	A	ARG	NH2	s..+	203	A	SO4	O1		2.97
45	A	ARG	NH2	s..-	60	A	ASP	OD2		3.43
45	A	ARG	NH2	s..+	203	A	SO4	O1		2.77
45	A	ARG	NH2	s..+	203	A	SO4	O3		2.84
45	A	ARG	NH2	s..+	60	A	ASP	OD1		2.99
50	A	LYS	N	s..+	54	A	GLU	OE1		2.76
51	A	THR	N	s..-	54	A	GLU	OE1		3.29
51	A	THR	OG1	s..-	53	A	ALA	N		3.48
51	A	THR	OG1	s..+	54	A	GLU	N		3.03
52	A	GLU	N	s..+	201	A	SO4	O1	* 2 1 -1 1	2.90
58	A	SER	OG	s..+	61	A	LEU	N		3.10
58	A	SER	OG	s..-	60	A	ASP	N		3.47
58	A	SER	OG	s..+	201	A	SO4	O4		2.80

2mbw bonds

Contacts with symmetry related atoms are marked by '*' followed by [SYM] and number of translations of one unit cell along x, y and z.

										DIST
1	A	VAL	O	....	133	A	LYS	NZ		3.28
3	A	SER	N	....	6	A	GLU	OE1		2.96
3	A	SER	OG	....	5	A	GLY	N		3.40
3	A	SER	OG	....	6	A	GLU	N		3.11
4	A	GLU	OE1	....	79	A	LYS	NZ		2.87
4	A	GLU	OE2	....	79	A	LYS	NZ		3.06
6	A	GLU	OE1	....	133	A	LYS	NZ		2.76
6	A	GLU	OE2	....	133	A	LYS	NZ		3.18
12	A	HIS	NE2	....	122	A	ASN	OD1		2.95
14	A	TRP	NE1	....	18	A	GLU	OE2		3.46
15	A	ALA	O	....	16	A	LYS	NZ	* 3 1 0 0	3.00
17	A	VAL	O	....	24	A	HIS	ND1		3.40
18	A	GLU	OE1	....	122	A	ASN	ND2	* 3 1 0 0	2.98
18	A	GLU	OE2	....	77	A	LYS	NZ		2.63
20	A	ASP	O	....	24	A	HIS	ND1		2.77
20	A	ASP	OD1	....	23	A	GLY	N		3.01
20	A	ASP	OD2	....	118	A	ARG	NH2		3.07
24	A	HIS	ND1	....	20	A	ASP	O		2.77
24	A	HIS	ND1	....	17	A	VAL	O		3.40
24	A	HIS	NE2	....	119	A	HIS	NE2		2.84
24	A	HIS	NE2	....	118	A	ARG	NE		3.38
26	A	GLN	NE2	....	56	A	LYS	O		2.95
27	A	ASP	OD2	....	118	A	ARG	NH1		2.76
31	A	ARG	O	....	35	A	SER	OG		3.18
31	A	ARG	NH1	....	35	A	SER	OG		3.39
31	A	ARG	NH2	....	113	A	HIS	NE2		3.21
34	A	LYS	NZ	....	52	A	GLU	OE1		2.77
36	A	HIS	O	....	39	A	THR	OG1		2.68
36	A	HIS	NE2	....	109	A	GLU	OE1		2.88
42	A	LYS	NZ	....	98	A	LYS	O		2.66
45	A	ARG	NE	....	60	A	ASP	OD2		2.62
45	A	ARG	NH1	....	154	A	[HIX]	O2D		2.71
45	A	ARG	NH2	....	60	A	ASP	OD2		3.45
45	A	ARG	NH2	....	60	A	ASP	O		3.30
48	A	HIS	O	....	81	A	HIS	NE2	* 1 0 0 1	3.10
50	A	LYS	N	....	54	A	GLU	OE1		2.85
51	A	THR	OG1	....	53	A	ALA	N		3.36
51	A	THR	OG1	....	54	A	GLU	N		3.10
52	A	GLU	OE2	....	56	A	LYS	NZ		3.35
53	A	ALA	N	....	200	A	SO4	O4		3.49
54	A	GLU	OE1	....	81	A	HIS	NE2	* 1 0 0 1	3.45
54	A	GLU	OE2	....	81	A	HIS	ND1	* 1 0 0 1	3.44
58	A	SER	OG	....	61	A	LEU	N		3.00
58	A	SER	OG	....	60	A	ASP	N		3.38

62	A	LYS	NZ	s..	+	118	A	ARG	O	#	2	0	0	1	3.20
62	A	LYS	NZ	s..	+	202	A	SO4	O1	#	2	0	0	1	2.11
62	A	LYS	NZ	s..	+	202	A	SO4	O4	#	2	0	0	1	3.12
64	A	HIS	ND1	s..	+	203	A	SO4	O1						2.78
64	A	HIS	ND1	s..	+	203	A	SO4	O4						3.08
66	A	VAL	O	s..	+	70	A	THR	OG1						2.77
67	A	THR	OG1	s..	+	203	A	SO4	O1						3.16
67	A	THR	OG1	s...		203	A	SO4	O2						2.39
78	A	LYS	O	s..	-	81	A	HIS	ND1						3.40
78	A	LYS	NZ	s..	+	85	A	GLU	OE1						3.00
78	A	LYS	NZ	s..	+	85	A	GLU	OE2						2.56
79	A	LYS	O	s..	-	81	A	HIS	ND1						3.49
81	A	HIS	O	s..	-	83	A	GLU	OE1						3.38
82	A	HIS	O	s..	-	85	A	GLU	OE1						3.49
82	A	HIS	NE2	s..	+	141	A	ASP	OD2						2.74
89	A	LEU	O	s..	-	92	A	SER	OG						3.28
89	A	LEU	O	s..	+	93	A	HIS	ND1						3.00
91	A	GLN	O	s..	+	95	A	THR	OG1						2.82
91	A	GLN	OE1	s..	+	152	A	GLN	NE2	#	2	0	0	0	2.83
91	A	GLN	OE1	s..	-	153	A	GLY	N	#	2	0	0	0	3.39
91	A	GLN	NE2	s.?	+	152	A	GLN	NE2	#	2	0	0	0	1.71
91	A	GLN	NE2	s..	+	153	A	GLY	N	#	2	0	0	0	2.57
92	A	SER	OG	s..	+	154	A	HIS	O1A						2.65
93	A	HIS	NE2	s..	+	154	A	HIS	NA						2.97
93	A	HIS	NE2	s..	+	154	A	HIS	NB						3.02
93	A	HIS	NE2	s..	+	154	A	HIS	FE						2.16
93	A	HIS	NE2	s..	+	154	A	HIS	NC						3.04
93	A	HIS	NE2	s..	+	154	A	HIS	ND						3.00
96	A	LYS	NZ	s..	-	204	A	SO4	O4	#	1	0	1	0	3.33
97	A	HIS	NE2	s..	+	154	A	HIS	O2A						2.83
99	A	ILE	O	s..	+	146	A	TYR	OH						2.68
102	A	LYS	NZ	s..	+	105	A	GLU	OE1						2.72
104	A	LEU	O	s..	+	108	A	SER	OG						2.85
108	A	SER	OG	s..	+	139	A	ARG	NH1						3.23
108	A	SER	OG	s..	+	139	A	ARG	NH2						3.16
109	A	GLU	OE1	s..	-	204	A	SO4	O4						3.45
109	A	GLU	OE2	s..	+	147	A	LYS	NZ	#	2	0	-1	0	3.02
113	A	HIS	O	s..	-	117	A	SER	OG						3.26
116	A	HIS	NE2	....		203	A	SO4	O4	#	1	0	-1	0	3.21
116	A	HIS	NE2	....		203	A	SO4	O2	#	1	0	-1	0	3.06
122	A	ASP	OD2	s..	+	202	A	SO4	O3						2.88
123	A	PHE	O	s..	+	128	A	GLN	NE2						2.91
124	A	GLY	O	s..	-	128	A	GLN	NE2						3.43
128	A	GLN	O	s..	+	132	A	ASN	ND2						3.19
132	A	ASN	ND2	s..	-	153	A	GLY	OXT	#	2	0	-1	0	3.46
132	A	ASN	ND2	s..	-	136	A	GLU	OE2						3.42
136	A	GLU	OE1	s..	-	153	A	GLY	OXT	#	2	0	-1	0	3.37
136	A	GLU	OE1	s..	+	153	A	GLY	N	#	2	0	-1	0	2.97
136	A	GLU	OE2	s..	-	153	A	GLY	N	#	2	0	-1	0	3.31
139	A	ARG	NE	s..	+	147	A	LYS	O	#	2	0	-1	0	3.20
139	A	ARG	NE	s..	+	148	A	GLU	O	#	2	0	-1	0	2.72
139	A	ARG	NH1	s..	+	147	A	LYS	O	#	2	0	-1	0	2.86
139	A	ARG	NH2	s..	+	147	A	LYS	O	#	2	0	-1	0	2.64
139	A	ARG	NH2	s..	+	108	A	SER	OG						3.16
143	A	ALA	O	s..	+	152	A	GLN	OE1						3.06
145	A	LYS	NZ	s..	+	148	A	GLU	OE1						2.78

59	A	GLU	OE2	....		62	A	LYS	NZ						3.37
63	A	LYS	O	....		67	A	THR	OG1						3.38
66	A	VAL	O	....		70	A	THR	OG1						2.79
77	A	LYS	NZ	....		122	A	ASN	OD1	#	3	1	0	0	3.39
78	A	LYS	O	....		81	A	HIS	ND1						3.37
78	A	LYS	NZ	....		81	A	HIS	O						3.09
78	A	LYS	NZ	....		85	A	GLU	OE2						3.28
81	A	HIS	O	....		83	A	GLU	OE2						3.48
82	A	HIS	O	....		85	A	GLU	OE1						3.43
82	A	HIS	NE2	....		141	A	ASP	OD2						2.75
87	A	LYS	NZ	....		126	A	ASP	OD1	#	6	0	0	0	2.99
87	A	LYS	NZ	....		126	A	ASP	OD2	#	6	0	0	0	2.86
88	A	PRO	O	....		92	A	SER	OG						3.38
89	A	LEU	O	....		93	A	HIS	ND1						3.02
89	A	LEU	O	....		92	A	SER	OG						3.32
91	A	GLN	O	....		95	A	THR	OG1						3.07
91	A	GLN	OE1	....		125	A	ALA	N	#	6	0	0	0	3.49
91	A	GLN	NE2	....		128	A	GLN	OE1	#	6	0	0	0	2.91
91	A	GLN	NE2	....		123	A	PHE	O	#	6	0	0	0	2.82
91	A	GLN	NE2	....		124	A	GLY	O	#	6	0	0	0	3.32
92	A	SER	OG	....		154	A	HIS	O2A						2.61
93	A	HIS	NE2	....		154	A	HIS	NB						3.08
93	A	HIS	NE2	....		154	A	HIS	FE						2.19
93	A	HIS	NE2	....		154	A	HIS	NC						3.13
93	A	HIS	NE2	....		154	A	HIS	ND						3.08
93	A	HIS	NE2	....		154	A	HIS	NA						2.88
95	A	THR	O	....		98	A	LYS	NZ						2.59
95	A	THR	OG1	....		128	A	GLN	OE1	#	6	0	0	0	3.32
95	A	THR	OG1	....		128	A	GLN	NE2	#	6	0	0	0	2.89
96	A	LYS	O	....		98	A	LYS	NZ						3.14
97	A	HIS	NE2	....		154	A	HIS	O1A						2.77
98	A	LYS	O	....		151	A	TYR	OH						3.49
99	A	ILE	O	....		146	A	TYR	OH						2.59
100	A	PRO	N	....		151	A	TYR	OH						3.30
101	A	ILE	N	....		153	A	GLY	OXT						3.38
104	A	LEU	O	....		108	A	SER	OG						2.83
108	A	SER	OG	....		139	A	ARG	NH1						2.98
108	A	SER	OG	....		104	A	LEU	O						2.83
113	A	HIS	O	....		117	A	SER	OG						3.17
128	A	GLN	NE2	....		149	A	LEU	O	#	2	0	0	0	3.00
132	A	ASN	ND2	....		136	A	GLU	OE2						3.00
132	A	ASN	ND2	....		148	A	GLU	O	#	2	0	0	0	3.11
143	A	ALA	O	....		152	A	GLN	OE1						2.78
145	A	LYS	NZ	....		148	A	GLU	OE1						2.78

Table S6. "Standard" long-range hydrogen/ionic bonds involving side chains in myoglobins and their variations in some structures.

	1bzp	*1bzr*	*1bz6*	*1spe*	*1vxb*	*1hjt*	*1ebc*	*2jho*	*1u7s*	*1jw8*	*1u7r*	*2z6s*	*1blh*	*1iop*	*2cmm*
K34NZ-E52OE1	+	+	+							*OE2*		+		-	
E38OE2-K79NZ	*	*	*	*	*	*	*	*				*			
S35O-H12NE2		*													
E41OE2-K77NZ	*		*									*OE1*			
K42NZ-K98O	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
R45NE-D60OD2	+	+	+			+	+	+	+			+			
R45NH1-HEMO2D	+	+	+			+	+	+	*NH1*	*NH2*		+	*NH2*	+	
R45NH1,2-D60OD1,2	+	+	+	+	+	+	+	+	+			+	+		
K50NZ-E18O			*			*									
M55O-Q26NE2		+	+			+									
K56O-Q26NE2	+	+	+	+		+	+	-	+	-	+	*KNZ*	-	-	-
A57O-K16NZ	*	*	*	*	*							*			
K62NZ-R118O	*-	*-	*		*			*				*			
K63NZ-A19O	*	*	*				*					*			
K77NZ-E18OE2,1	+	+	+	+	+	+	+	+	+	*OE1*	+	+	+	+	+
K79NZ-E4OE1,2	+	+	+	+	+	+	+	+	+	*OE1*	+	+	+	+	
H82NE2-D141OD2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Q91NE2-G153O	*	*	*	*				*							
S92OG-HEMO1A	+	+	+			*O2A*	+		+	*O2A*	+	+	*O2A*		
H93NE2-HEMFe	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
H97NE2-HEMO2A	+	+	+	*O1A*	*O1A*	*O1A*	+	+	+	*O1A*	+	+	*O1A*		
R118NE-D27OD2	+	+	+	+	+	+				+		+			
R118NH2-D27OD2	+	+	+			+	*NH1-*	*NH1*	*NH1*	*NH1*	*OD1*	+	*NH1*	+	*NH1*
R118NH2-D20OD2	+	*NH1*	+		-	*NH1*	+	+	+			*NH1*	-	+	+
H119NE2-H24NE2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
D122OD1-H12NE2	+	+	+	+	+						+	+	+		+
K133NZ-V1O	-	+	+					+	+		+	+	+		+
K133NZ-E6OE1,2	+	+	+	+		+		+	+	+	+	+		+	+
S35O-Q8NE2	*	*	*	*-	*	*									
I99O-Y146OH	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
S108OG-R139NH2	*NH1*	*NH1*	*NH1*		+	-	*NH1*	*NH1*	*NH1*	+	*NH1*	*NH!*	*NH1*	*NH1*	*NH1*
E109OE2-K147NZ	*	*	*				*	*				*-			
K98O-Y151OH	+								+			+		-	
I99N-Y151OH	-											-		-	
R31NH1/2-H113NE2				+	+	*ND1*			-	-		-			-
L32O-T39OG1					+										
R45NE-D60OD2												+			
R45NE-H64NE2				-											
H64NE2-HEMO1D				+											
E85OE2-D126OD1,2				*	*							*			
Q91NE2-Q152OE1				*											
S92OG-HEMO2A				+	+										
E136OE1-Q152N				*	*										
R139NH1-K147O				*	*	*NH2*	**NH2*	**NH2*				**NH2*			
R139NE-E148O						*	*-	*-				*			
R118NH1-H24NE2					+			*NH2-*	*RNE-*	*NH2-*					
E38OE1-Y103OH					-										

	1bzp	1bzr	1bz6	1spe	1vxb	1hjt	1ebc	2jho	1u7s	1jw8	1u7r	2z6s	1blh	1iop	2cmm
R45NH1-D60O					+										
K78NZ-E85OE2					+	+	-		-	OE1,2				+	OE1,2
A94N-Y146OH					+										
K140NZ-Y151OH					+	-									
A143O-Q152OE1					-	+		NE2-							
D122OD2-K16NZ						+	+ !		-						-
K16NZ-E59OE2						*-		*							
R118NH2-G23O						+	NH1	NH1-	NH1-	-	-	+		NH1-	+
H119ND1-K62NZ						*									
S117O-K63NZ						*									
H64NE2-NO(N)						+	CYNN	CYNN				O2		CYNN	CYNC
S92OG-HEMO2(1) A						+	O1A								
H97NE2-HEMO1(2)A						+	O2A								
Q128NE2-HEMO2A						*									
K34NZ-K50O							+								
K34NZ-K52N							+							-	
K79NZ-E38OE2							*	*							
H116NE2-Q128OE1							+	+				NE2-			
H116ND1-Q128NE2,OE1											NE2-				OE1-
D122OD2-K62NZ							*								
K96NZ-E109OE1							*	*-							
E83OE2-V1O								*							
R118NE-H24ND1,NE2								-		-					-
E83OE1-D141OD2											-	-			
E83OE2-K145NZ								-		+	OE1,2	+		OE1,2	
E83O-R145NZ											+	+			
Q91OE1-G153N								+							
H93O-Y146OH								-							
Q128NE2-HEMO1A								*-							
E136OE2-Q152O								*?				QN			
E136OE2-G153O								*?							
Q128NE2-E18OE1,2									*						
E41OE2-K42NZ									*						
D44OD2-K98NZ									*-						
K47O-Q152NE2									*						
H48NE2-H48NE2									*						
K63NZ-HEMO2D									*						
K79NZ-E136OE1									*						
P100N-Y151OH									+	+	-	-		+	
G121O-A125N									*						
A125N-G121O									*						
K79NZ-E4OE1,2							+			+	+				
K133NZ-E6OE1,2										+	+			+	
H24ND1-V17O										-	-			-	
H119ND1-E18O										*-					
N122ND2-E18OE2										*OD1?			*OE1		
N122OD1-K77NZ													*		
A15O-K16NZ													*		
R118NH1-D27OD2										+					
K98NZ-R31NH2										*					

	1bzp	*1bzr*	*1bz6*	*1spe*	*1vxb*	*1hjt*	*1ebc*	*2jho*	*1u7s*	*1jw8*	*1u7r*	*2z6s*	*1blh*	*1iop*	*2cmm*
E109OE2-H36NE2										+			*OE1*		
R45NE-HEMO1D										+					
H81NE2-E54OE1,2										*			*	*	
H64NE2-CMOO										+					
K77NZ-G121O										*					
K87NZ-D126OD1,2										*			*	*	
Q91NE2-F123O										*			*	*	
Q91NE2-G124O										*-				*	
Q91NE2-Q128NE2										*			*	*OE1?*	
T95O-H113ND1										*-					
T95OG1-Q128NE1/2										*			*	*	
Q128NE2-L149O										*			*	*	
I101N-G153OXT													*		
R45NH2-H64ND1														+	
N132ND2-K147O										*-					
N132OD1-E148O										*			*	*?	
E54OE1,2-H81NE2										*				*	
K56O-Q26OE1											+			*NE2-*	
K56NZ-Q26NE2,OE1											+	+			
K56NZ-D27OD2											+				
H81NE2-E41OE1											*-				
K42NZ-H97O											+			-	
E59OE1-H81O											*-				
K62NZ-E85OE2											*				
T67OG1-K87NZ											*-				
H64ND1-IMDN3											+				
E105OE2-A125N											*				
H116ND1-F123O											-				
Q152NE2-S117O,OG											*				
R118NH1-E148O											*				
E4N-S35O															*
A19O-E102N															*
V21N-G153O															*
V22N-G153OXT															*
R45NH1-G121O															*
R45NH2-P120O															*
H48ND1-D126OD2															*-
S58OG-D126OD1,2															*-
E59N-D126OD1															*
D60N—D126OD1															*
D60OD1-A125N															*
D60OD2-D126N															*
T67OG1-Q91NE2															*
K78NZ-Q85N,OE1															*
E85OE1,2-K98NZ															*
E85OE2-T95O															*-
K50NZ-G80O														*	*
K87NZ-K63O,T67OG1															*, *
K96NZ-K147O,E148O															*, *-

See 2.4 -2.5. * strong intermolecular hydrogen/ionic bond; *- weak; + and - strong & weak intramolecular bonds.

Section S8.

On the role of cryogenic temperature in myoglobin conformational changes.

After a seminal 1987 publication (Frauenfelder *et al*., 1987) it became an accepted belief that cryogenic temperatures induce significant conformational changes in myoglobin. The reality is, however, more complex. In this study (Section 3.3) we do not find any pair of native myoglobin structures (one at the cryogenic and one at room temperature at neutral pH) in deoxy-, oxy-, met- or CO-state with $P2_1$ symmetry and RMSDDs above the uncertainty threshold. Perhaps the extra contribution to the RMSDDs in 1987 study came from C-terminal residues 152 and 153, which were excluded from comparisons in our study. While RMSDDs for some pairs were only slightly below the uncertainty threshold, results of other RMSDD comparisons are confounding.

Table S7. Low RMSDD between cryogenic and room-temperature structures.

Structure 1 PDB code	Structure 2 PDB code	RMSDD Å	Temp 1 K	Temp 2 K	Resol 1 Å	Resol 2 Å	State of Struct. 1	State of Struct. 2
1bz6	1bzp	0.15	287	287	1.2	1.17	aquomet	deoxy
1bz6	2z6s	0.24	287	100	1.2	1.25	aquomet	oxy
1bz6	1a6k	0.21	287	100	1.2	1.1	aquomet	aquomet
1vxh	1a6k	0.19	277	100	1.7	1.1	met	aquomet
1mbc	1a6m	0.19	260	100	1.5	1.0	cmonoxy	oxy
1mbc	2z6s	0.18	260	100	1.5	1.25	cmonoxy	oxy
1jp8	1a6k	0.18	295	100	2.3	1.1	met	aquomet
1a6n	1ajh	0.19	100	40	1.15	1.69	deoxy	cmonoxy

Apparently RMSDD (and thus conformational differences) between two high resolution structures at room temperature, and between cryogenic and room temperature structures are rather similar and do not significantly depend on the resolution of the room temperature structures. In fact 1bz6 and 1bzp used as a starting structure for refinement 4mbn with resolution of 2Å (see Table S3). However, RMSDD 1bz64mbn is 0.17 Å which is very similar to RMSDD 1bz61bzp in Table S7 above. These data might invalidate the widely held beliefs about the role of cryogenic temperatures in proteins conformational differences. Some cryogenic structures do significantly differ from the

room-temperature structures of the same protein, and some do not. Process of validation includes, in our view, invalidation of opinions not based on wide range of observations.

Section S9.

Data set size and resolution influence on the value of the uncertainty threshold and a general applicability of the approach.

This section is largely related to our initial publication (Rashin *et al*., 2009). In that publication uncertainty threshold had been determined from RMSDD of 18 whale myoglobin structures and 41 bovine ribonucleaseA structures (two molecules in a unit cell were considered as independent). These subsets did not contain any structures with characteristics that had been claimed to sometimes produce significant conformational changes (e.g., mutations or low water content). Some or all of such structures have coordinate variations explained or "justified" and thus inappropriate for finding uncertainty thresholds. This kind of careful selection is necessary to determine uncertainty thresholds.

Reasonably reliable thresholds apparently require relatively large data sets. Fig. 4 in our initial paper (Rashin *et al*., 2009) shows that 41 RNaseA fill well the entire RMSDD distribution provided by two joined subsets. On the other hand, 18 myoglobin structures provide a discontinuous RMSDD distribution up to RMSDD of 0.40Å but with most values in RMSDD range below 0.24Å. This distribution has a rather long sparsely populated RMSDD distribution with the right-most bar with two RMSDDs at 0.38Å. However, addition of some cryogenic myoglobin structures studied in this work to the data set would probably add a few RMSDDs of 0.43-0.44Å. Large sets of independently determined structures of the same protein are rare. The next promising data set might be for egg-white lysozyme from birds. However, it might be much more difficult than the myoglobin set studied in this paper, because of reported large number of loops deformed between crystal forms and associated with intermolecular contacts. We started with the myoglobin often consider a like of "a hydrogen atom" for proteins.

We have not previously checked influence of structures resolution in data sets on uncertainty thresholds values. Low resolution is considered as worse than 2.5Å (Kleywegt, 1999). In our data sets (Rashin *et al*., 2009) there are no structures with such low resolution. High resolution used to be considered 2Å or better (Kuriyan *et al*., 1987). All but three protein crystals in our data sets of 2009 were studied at 2Å or better resolution. Five structures from these three crystals are: 1rcn (2.32 Å), 1jvv(A,B, 2.2 Å) and 1jvt(A,B, 2.05 Å). In the parenthesis are shown chains in the unit cell, considered as independently determined structures, and the resolution. Elimination of these five structures from 2009 data set eliminates in Fig. 4 (Rashin et al., 2009) three last highest RMSDD bars from 0.42Å to 0.44 Å, possibly reducing the uncertainty threshold to 0.41Å. However (see the previous paragraph), inclusion of cryogenic whale myoglobins into the set might restore the threshold back to 0.44Å. (We use "might" because the cryogenic structure of most interest in this contribution to the threshold are photodissociated cmonoxy structures 1ajg and 1ajh at 40K, which possibly have their particular reasons to be different). Exclusion of structures at 2 Å resolution from 2009 data set does not change the uncertainty threshold. Therefore, we do not see particular reasons to reconsider the uncertainty threshold value at this time,

Our approach can be currently fruitfully applied to: **a)** any set of protein structures from PDB which were claimed to be "almost identical" by the authors (we are working on one such set now); **b)** any pair of independently determined structures of a same protein, including pairs of molecules in crystals with more than one molecule in the unit cell (see Rashin *et al.*, 2009, 2010) (we are working on 1,000 structures with two monomers in unit cells.); **c)** presumed "functional" motions as described in detail in 2009 and 2010 (Rashin *et al*.) papers.

References.

Kleywegt, G.J. (1999). *Acta Cryst*. D**55**, 1878-1884.

Krebs, W.G., & Gerstein, M. (2000). *Nucleic Acids Res*. **28**,1665-1675.

Kuriyan, J., Wilz, S., Karplus, M. & Petsko, G.A. (1986). *J. Mol. Biol*. 192, 133-154.

Kuriyan, J., Karplus, M. & Petsko, G.A. (1987). *Proteins,* **2**, 1-12.

Vojtechovsky, J., Chu, K., Berendsen, J., Sweet, R.M. & Schlichting, I. (1999). *Biophys. J*. **77**, 2153-2174.