WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments

Dmitry A. Kondrashov,[1]* Wei Zhang,[2]* Roman Aranda IV,[3] Boguslaw Stec,[4] and George N. Phillips Jr.[1]†

[1] Department of Biochemistry, University of Wisconsin – Madison, Madison, Wisconsin 53706

[2] Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030

[3] Department of Biomolecular Chemistry, University of Wisconsin – Madison, Madison, Wisconsin 53706

[4] Burnham Institute for Medical Research, La Jolla, California 92037

ABSTRACT

Proteins sample multiple conformational substates in their native environment, but the process of crystallization selects the conformers that allow for close packing. The population of conformers can be shifted by varying the environment through a range of crystallization conditions, often resulting in different space groups and changes in the packing arrangements. Three high resolution structures of myoglobin (Mb) in different crystal space groups are presented, including one in a new space group $P6_122$ and two structures in space groups $P2_12_12_1$ and P6. We compare coordinates and anisotropic displacement parameters (ADPs) from these three structures plus an existing structure in space group $P2_1$. While the overall changes are small, there is substantial variation in several external regions with varying patterns of crystal contacts across the space group packing arrangements. The structural ensemble containing four different crystal forms displays greater conformational variance (Cα rmsd of 0.54–0.79 Å) in comparison to a collection of four Mb structures with different ligands and mutations in the same crystal form (Cα rmsd values of 0.28–0.37 Å). The high resolution of the data enables comparison of both the magnitudes and directions of ADPs, which are found to be suppressed by crystal contacts. A composite dynamic profile of Mb structural variation from the four structures was compared with an independent structural ensemble developed from NMR refinement. Despite the limitations and biases of each method, the ADPs of the crystallographic ensemble closely match the positional variance from the solution NMR ensemble with linear correlation of 0.8. This suggests that crystal packing selects conformers representative of the solution ensemble, and several different crystal forms give a more complete view of the plasticity of a protein structure.

Proteins 2008; 70:353–362.

***Key words:** crystal forms; conformational substates; myoglobin; protein dynamics; X-ray diffraction; atomic displacement parameters; NMR ensemble.*

INTRODUCTION

The concept of the native state of a protein has evolved from a static, unique conformation with minimal energy to the model of a frustrated, hierarchical energy landscape.[1,2] In pioneering experiments on ligand rebinding in myoglobin (Mb), conformational heterogeneity was inferred from a nonexponential distribution of rebinding times following flash photolysis,[3] and further investigations using IR spectroscopy found evidence of multiple functionally significant states.[4] Computational studies using molecular dynamics simulations have contributed to the understanding of the conformational transitions between substates of the native state.[5–7] Evidence has accumulated that conformational flexibility of protein structures plays an important role in a diverse set of functions. Experiments have demonstrated that protein dynamics control substrate binding and release in enzymes, e.g., adenylate kinase,[8,9] and that catalytic activity can be regulated by dynamic coupling of the active site with distant residues in dihydrofolate reductase[10] and many others. The study of protein dynamics has emerged as central to understanding the mechanism of many biochemical processes.

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Library of Medicine; Grant number: NLM 5T15LM007359; Grant sponsor: National Institute of Health; Grant number: NIH 5 T32 GM08349; Grant sponsor: W.M. Keck Center for Computational Biology; Grant sponsor: Robert A. Welch Foundation; Grant number: C-1142; Grant sponsor: Wisconsin Alumni Research Foundation.

This work was supported by National Institutes of Health/National Institute for General Medical Sciences Grants P50 GM64598 and U54 GM074901 (J. L. Markley, PI).

*D. A. K. and W. Z. contributed equally to this work.

†Correspondence to: George N. Phillips Jr., Department of Biochemistry, University of Wisconsin, 433 Babcock Drive, Madison, Wisconsin 53706.
E-mail: phillips@biochem.wisc.edu

Received 12 November 2006; Revised 12 February 2007; Accepted 5 March 2007
Published online 6 August 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21499

X-ray crystallographic techniques have been used extensively to determine static structures of proteins, and more recently time-resolved crystallographic experiments have revealed protein dynamic information at the atomic level.[11,12] Crystallography has also provided information about conformational variation from structures determined under different conditions[13] and through the analysis of atomic displacement parameters (ADPs) that provide a measure of conformational flexibility of proteins in the crystal.[14,15] A recent study has shown that the reported coordinates in protein structures may contain substantial errors, being linked to conformational heterogeneity of proteins in the crystal.[16]

During the crystallization process the conformational ensemble of a protein is perturbed to varying degrees by packing interactions. While there is ample evidence that the gross structure of a protein in the crystal does not differ dramatically from the solution ensemble, a systematic analysis of the structural deviations has been missing. Specifically, it is not clear whether the conformational substates present in the crystal are part of the native ensemble in solution, or if the crystallographic contacts lead to significant local perturbations of the energy landscape, although some studies have found support for the former.[13,17] Intermediate conformations along a fairly well established coordinate of conformational change have been observed in several crystal structures.[18–20] Comparison of structures of the same protein crystallized in different space groups can help quantify the effect of crystal packing and varying solvent content on the observed structure and reveal new members of the ensemble as well. Previous studies of different crystal forms of T4 lysozyme[21] and cutinase[22] found substantial variation among the conformers matching the expected flexibility of the protein structures.

NMR spectroscopy provides both structural and dynamic information.[23] Refinement of NMR structures is performed via molecular dynamics simulations of a set of independent structures, with distance constraints from NOEs. The set of structures which converges to a neighborhood of a folded conformation is reported as the NMR structural model, after some refinement. The resultant conformational ensemble is influenced both by the computational force field and the experimental restraints. In macromolecular regions with few NOEs the spread in the ensemble is generated by the dynamics simulation rather than by experimental data. Thus, NMR models provide an independent view of the conformational ensemble in solution, with a different set of limitations from crystallography.

Mb is a small globular protein that functions as an oxygen transport, storage, and mobile buffer protein in muscle.[24] Mb folds as a compact monomer and does not undergo large conformational changes in comparison with multidomain proteins, but localized changes are observed[17] as it reversibly binds various ligands,[11,25] and with pH variation.[26] Conformational substates of Mb have been identified from vibrational spectra of the bound carbon monoxide (CO) ligand, corresponding to configurations of the neighboring residues in the heme binding pocket.[27] Conformational relaxation in Mb has been investigated using vibrational spectroscopy,[4] time-resolved crystallography,[11,12] and molecular dynamics simulations.[28,29] The accumulated evidence suggests that the conformational ensemble of Mb is shifted by external conditions, modulating functional properties such as ligand-binding propensity and accessibility of internal cavities.[30,31] The subtlety of the conformational changes necessitates a detailed description of the conformational ensemble of the protein.

To date, nearly all of the Mb protein crystals have been grown from high concentrations of ammonium sulfate or phosphate as precipitants with different buffers, pH levels, and additives.[32–35] We crystallized Mb in various conditions to yield high-resolution $P2_12_12_1$, P6 and $P6_122$ space group structures. The $P6_122$ space group is a new protein packing symmetry for Mb, with a similar solvent content to the $P2_12_12_1$ Mb space group crystal. A previously published high-resolution Mb crystallized in the $P2_1$ space group (PDB 1BZR[36]) was used for comparison with the three reported structures, and analyzed for changes in the conformational substates. We analyze the variance in the resulting structural ensemble of the four structures and compare the results with ensembles of Mb structures from the same space group and with an NMR-determined ensemble.

MATERIALS AND METHODS

Crystallization and data collection

Native sperm whale skeleton Mb protein was purchased from Sigma prior to the restriction of its sale, and recombinant (D122N mutant) protein was generously provided by John S. Olson. The protein concentration for crystallization of the native protein was 50 mg/mL. The $P2_12_12_1$ form was crystallized using 50 m*M* imidazole (Im) at pH 7.0 and 12–16% polyethylene glycol (PEG) 3550; the $P6_122$ form was crystallized using 50 m*M* citric acid at pH 4.5, 100 m*M* KCl, and 12–20% PEG 1550. The P6 form was crystallized from the recombinant protein in 20 m*M* Tris HCl at pH 9.0, EDTA 1 m*M*, and 2.4–2.6*M* ammonium sulfate.[33] For detailed comparisons, the crystallization conditions are listed in Table I.

Diffraction data were collected at 100 K, at the APS BioCARS beamline at Argonne National Lab on a ADSC Quantum-4 detector using a 1.00 Å wavelength source ($P2_12_12_1$ and $P6_122$ crystals), and at NLSL at Brookhaven National Lab using a 0.91 Å wavelength radiation (P6 crystal). The $P2_12_12_1$ crystal diffracted to 1.40 Å, $P6_122$ to 1.15 Å, and P6 to 1.30 Å. Data sets were reduced and processed on site with DENZO/HKL program,[38] and Friedel pairs were not merged in the structure factor files. The summary of the data collection statistics is given in Table II.

DOI 10.1002/prot

Table I
Different Crystallization Conditions for Sperm Whale Myoglobin

PDB ID	Space group	Buffer	Precipitant	Salt	Reference
1U7S	$P6_122$	50 m*M* Citric acid, pH 4.5	12–20% PEG 1550	100 m*M* KCl	This work
1U7R	$P2_12_12_1$	50 m*M* Imidazole, pH 7.0	12–16% PEG 3550	—	This work
1JW8	P6	20 m*M* Tris.HCl, 1 m*M* EDTA, pH 9.0	2.4–2.6*M* Ammonium sulfate	—	48
1BZR	$P2_1$	100 m*M* Potassium phosphate, pH 6.0	2.3–2.8*M* Ammonium sulfate	—	36

Structure determination and refinement

A high resolution model CO-bound Mb (PDB 1BZR,[36]) was used as the starting model for molecular replacement. The rotation and translation search was carried out for the three space-group structures using Crystallography and NMR System (CNS).[39] The search resolved the enantiomorphic ambiguity between the space groups $P6_122$ and $P6_522$ for the $P6_122$ crystal structure. Structures were further improved through a rigid-body refinement and simulated annealing in CNS and with the restrained least squares procedure in SHELXL97.[40] The final round of refinement was carried out with anisotropic temperature factors for all atoms with the observed reflection data. The alternative conformations of residue side chains were identified from difference electron density or sigma-a weighted $2F_o-F_c$ maps. Electron density with clear peaks in the difference (F_o-F_c) maps for alternative conformations were modeled as such.

The refinement was considered to be converged when both the R and R_{free} factors were stabilized and no apparent peaks at the 3σ level in the final F_o-F_c difference map were detected. Manual fitting was carried out using the computer graphics program XtalView[41] and graphic illustrations were created with the programs MOLMOL[42] and PyMOL (www.pymol.org). Structures were validated through Ramachandran plots and PROCHECK.[43]

Quantitative analysis

PyMOL was used to overlay the structures from different space group with minimal root-mean squared deviation (rmsd). The rms command was used to measure the average distance between different subsets of atoms. To measure similarity in the atomic displacement parameters, we separated the magnitudes and directional components. The former is equivalent to the trace of the ADP matrix, and to compare the latter, each matrix was divided by its trace. Directionality comparison was restricted to ADPs with nonspherical distributions, specifically, those with the ratio of the lowest to highest eigenvalue (known as the anisotropy[44]) less than 0.5. Two measures were used to compare directionality of ADPs: dot product of the largest major axis (the eigenvector corresponding to the largest eigenvalue) and the overlap fraction, for comparing two electron density distributions, first computed by Merritt.[45] We used a modified formula for the overlap fraction, where a score of 1 indicates perfect alignment, and 0, perfect misalignment. The modified overlap score between two normalized variance matrices U and V is computed as follows, where rho is the Merritt overlap measure, and V^* is the matrix V rotated with the major axis coinciding with the minor axis of U, and vice versa.

$$\rho(U,V) = \frac{\rho(U,V) - \rho(U^*,V^*)}{1 - \rho(U^*,V^*)}$$

Table II
Data Collection and Refinement Statistics, the Highest Resolution Shell Statistics are Show in Parentheses

PDB ID	1U7S	1U7R	1JW8
Space group	$P6_122$	$P2_12_12_1$	P6
Unit cell parameters Å	$a = b = 73.74$, $c = 96.70$	$a = 39.95$, $b = 47.97$, $c = 78.32$	$a = 90.38$, $b = 90.38$, $c = 45.34$
Solvent content (%)	44	44	59
Resolution (Å)	50–1.40 (1.45–1.40)	30–1.15 (1.19–1.15)	30–1.30 (1.35–1.30)
Total reflections	55,715	100,212	269,537
Unique reflections	29,988	53,703	49,946
Completeness of data (%)	96.0 (82.3)	98.4 (92.8)	94.6 (83.0)
R merge (%)	4.1 (37.2)	3.9 (23.5)	6.7 (27.2)
Refinement statistics			
R_{cryst} (%)	14.3	13.8	13.3
R_{free} (%)	17.9	16.5	15.7
Model statistics			
Protein atoms	1221	1218	1272
Solvent atoms	203	183	260
rmsd bond lengths (Å)	0.016	0.019	0.013
rmsd angle distances (°)	0.035	0.035	0.027

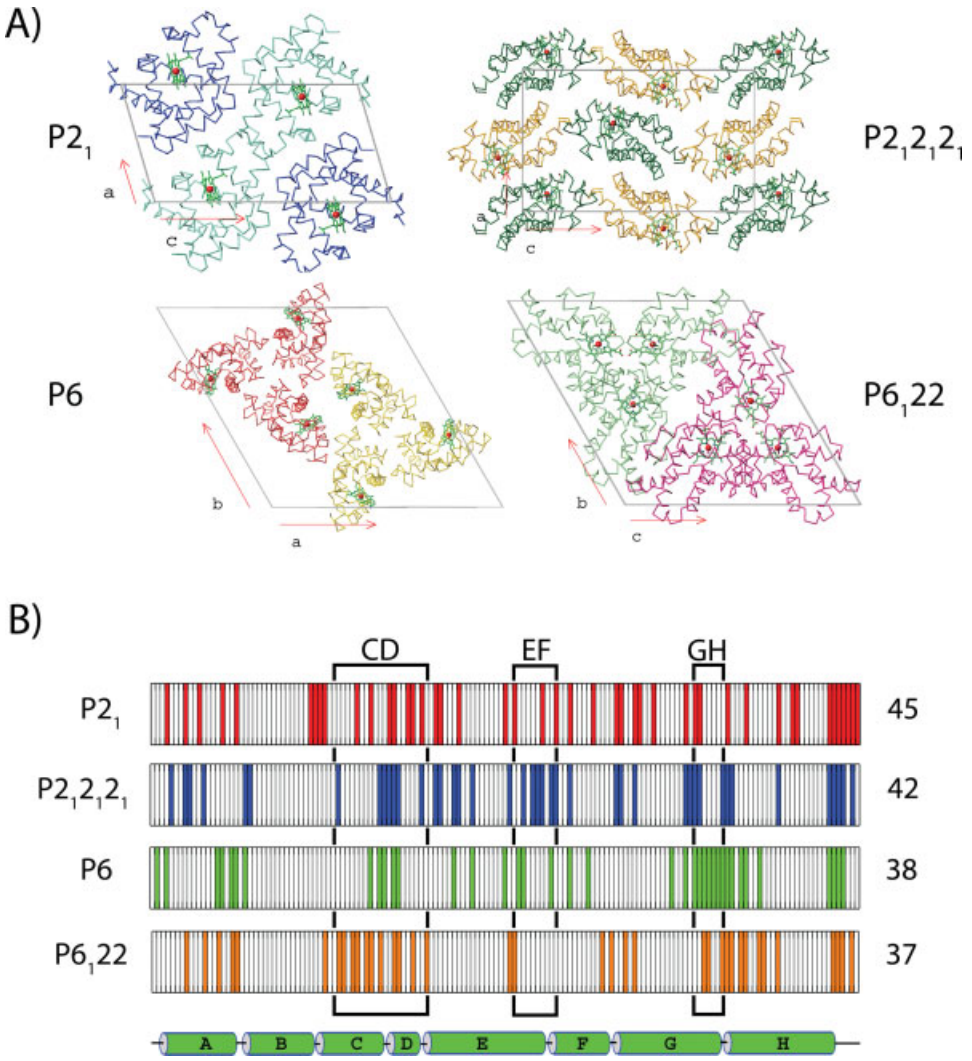

Figure 1

*Crystal packing properties of Mb structures in four different crystalline forms. (**A**) Unit cell packing of the Mb in four space groups. (**B**) Residues with atoms within 4 Å of an atom in the symmetry molecules are highlighted for each space group. Regions of particular interest, the CD, EF, and GH, are identified with boxes and labeled accordingly. The schematic representation of secondary structure of Mb is shown below.*

To compare the structural variance from crystallographic structure with the NMR ensemble, a composite variation was computed from both coordinates and ADPs. In crystal structures, atomic coordinates are the means of Gaussian distributions with variance described by ADPs. The variance of a combination of several Gaussian distributions was computed as follows:

$$\frac{1}{N}\sum_i \int (x-\bar{x})^2 p(x)dx = \frac{1}{N}\sum_i \mathrm{Var}(x_i) + \frac{1}{N}\sum_i (\bar{x}-\bar{x}_i)^2$$

where N is the number of structures, x_i is the atomic position variable in structure i, $p(x)$ is a Gaussian distribution of positions, $\bar{x}_i$ denotes the mean coordinate as reported in crystal structure i, $\mathrm{Var}(x_i)$ is the positional variance as measured by the ADP, and $\bar{x}$ is the average of the set of coordinates from N different structures. The expression holds because the Gaussian distribution is symmetric. Thus, the combined variance in a set of crystal structures is obtained as a sum of the average ADPs (variances) and mean square coordinate deviations from the average position.

RESULTS

High resolution structures in different space groups

The Mb structure in the new space group P6$_1$22 was obtained in a citric acid buffer and solved to a resolution of 1.40 Å. This type of crystal arrangement is not repre-

DOI 10.1002/prot

sented in over 200 crystal structures of different Mbs deposited in the PDB. Two more structures, in space groups P6 and $P2_12_12_1$, are presented here at very high resolution, 1.30 and 1.15 Å, respectively. The resolution surpasses all published Mb structures in these space groups with the exception of a P6 structure of the YQR Mb mutant at 1.04 Å (PDB 1NAZ[46]). The high resolution allowed us to detect alternate conformers and refine anisotropic displacement parameters (ADPs), which were not included in the 1NAZ model. The residues near the heme pocket of Mb are in similar conformations with the exception of the $P2_12_12_1$ Im-bound structure, where the presence of Im forces the distal histidine (His64) to rotate ~90° about its χ_1 dihedral angle towards the solvent in comparison with the other space group structures. The rotation also displaces Phe46 1.3 Å away from His64.

We compare four Mb structures: a previously published 1.15 Å CO-bound Mb structure in space group $P2_1$ (PDB 1BZR), and three new structures presented here: CO-bound Mb in space group P6 (1JW8), Im-bound Mb in space group $P2_12_12_1$ (1U7R), and aqua Mb in space group $P6_122$ (1U7S). The arrangement of protein molecules in the four unit cells and the resulting intermolecular crystal contacts are shown in Figure 1. Counting as contacts atomic pairs within 4 Å, there are 45, 42, 38, and 37 residues involved in crystal contacts in space groups $P2_1$, $P2_12_12_1$, P6, and $P6_122$, respectively. The differences in packing are also reflected in the solvent content of the unit cell, shown in Table II, which is highest in P6 at 59%, and is significantly lower in the space groups $P2_12_12_1$, $P6_122$, which both consist of 44% solvent. The $P2_1$ space group structure has the lowest solvent content at 34%. The differences in crystal contacts and solvent environment result in conformational variation between the different space groups.

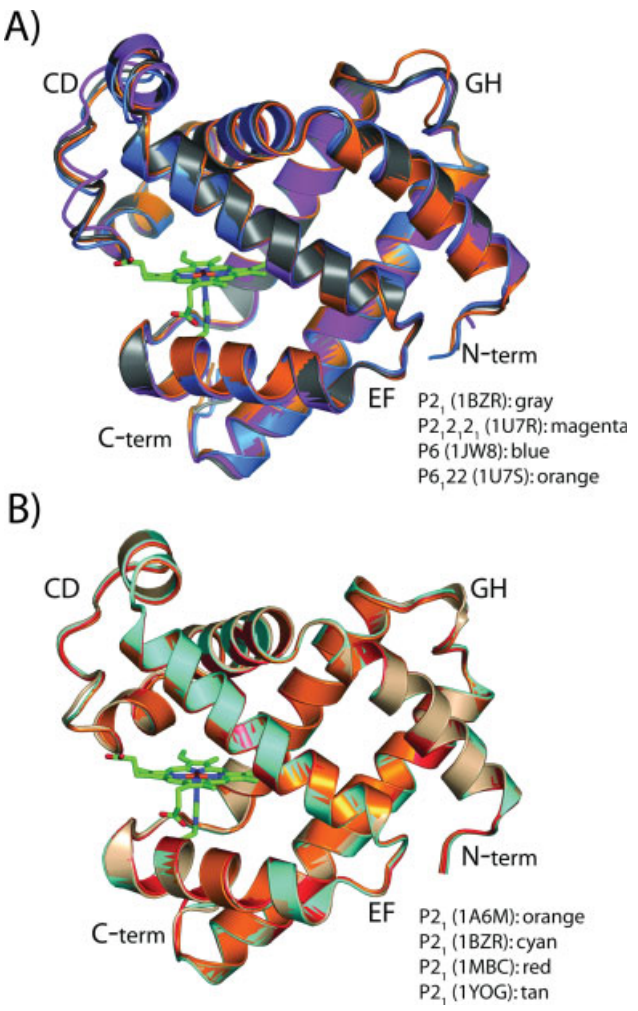

Figure 2

*Superposition of Mb structures from (**A**) $P2_1$, $P2_12_12_1$, P6, $P6_122$ space groups, contrasted with (**B**) an ensemble of four $P2_1$ Mb structures with diverse conditions and mutations. The CD, EF, GH loops and termini are labeled for each ensemble.*

Structural variation and crystal contacts

Crystal structures in different space groups show localized but significant structural variation. The structures were aligned by rmsd fitting over carbon alpha (Cα) atoms from the five largest helices (A, B, E, G, H), and the resulting ensemble is shown in Figure 2(A). The variation in structure is most pronounced in four regions: the N-terminus, short helices C and D (residues 40–60, including the CD corner), the loop between helices G and H (residues 118–125, GH loop), and the C-terminus. Quantitative comparison of the structures is presented in Table III. The Cα rmsd values over the entire structure vary between 0.54 and 0.79 Å, while the rmsd values for all atoms vary between 1.0 and 1.3 Å. If comparison is restricted to the five large helices, the variation is much smaller, but still measurable: the Cα rmsd values range between 0.28 and 0.37 Å, while the all atom rmsd values are between 0.87 and 1.10 Å. To ascertain whether this variation is a result of differences in crystal packing we performed the same analysis on sets of high-resolution structures in space groups $P2_1$ (PDBs 1A6M, 1BZR, 1MBC, 1YOG), shown in Figure 3(B), and P6 (1DO1, 1DXC, 1JW8, 1NAZ), all solved by different authors, and with varying ligands and mutations. The variation in these ensembles is significantly smaller than in the ensemble with different space groups (Figure 2). The Cα rmsd values for all residues are between 0.13 and 0.29 Å for the $P2_1$ and P6 sets of structures, and the all atom rmsd ranges from 0.59 and 0.95 Å. Also, the variation in structure is not as localized, with the core residues (helices A, B, E, G, H) showing only slightly smaller rmsd values than the entire structure.

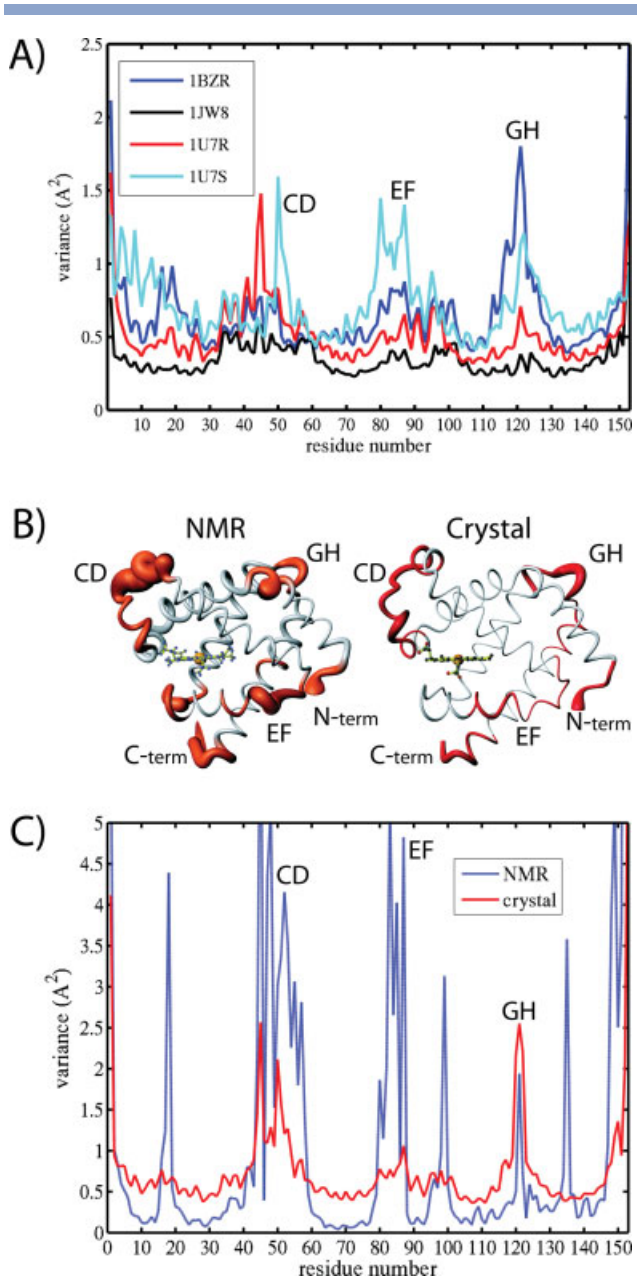

Figure 3

*Variation in the crystallographic and NMR Mb structure ensembles. The CD, EF, and GH regions are labeled in each panel. (**A**) ADP magnitudes are plotted for the structures in four space groups. (**B**) The NMR ensemble of 12 models and crystallographic ensemble of four space group structures shown in cartoon format with ribbon thickness proportional to positional variance. For NMR cartoon, fragments of the RMSD difference of Cα greater than 1.0 Å are highlighted in orange, and for crystal cartoon, regions with variance greater than the mean are colored in red. (**C**) Variance obtained from combination of ADPs and coordinate variance in crystallographic ensemble plotted together with positional variation from NMR ensemble.*

Conformational changes between Mb structures in different crystal forms are primarily restricted to external residues, specifically the terminal regions, small helices C and D, and the GH loop. The terminal regions are expected to be mobile in most protein structures, thus we investigate the cause of the variation in the other two locations. Figure 1(B) shows large differences in the number of crystal contacts in the CD corner and the GH loop between the Mb crystal forms. The CD corner (residues 20–40) is well packed in $P6_122$ space group with 10 contacts, while there are only five contacts in the P6 packing. On the other hand, the GH loop (residues 118–125) is best packed in P6 space group, with each of the eight residues involved in contacts, in contrast to the packing in space group $P2_1$ with only three contacts. However, the structures with the largest structural deviations in the CD corner and the GH loop are from the $P2_12_12_1$ and $P6_122$ space groups, respectively [see Fig. 2(A)], which do not contain the highest or lowest number of crystal contacts. Further, the E-F region (residues 79–88) also shows substantial differences in crystal contacts, with six contacts in space group $P2_12_12_1$ and only one in space group $P6_122$, but the differences induce little structural variation. The evidence suggests that greater numbers of crystal contacts do not necessarily induce conformational changes, but that specific interactions from a smaller number of contacts may stabilize one conformation at the expense of others.

Dynamics in crystal structures and solution NMR ensemble

In addition to atomic coordinates, crystal structures also include atomic displacement parameters (ADPs) to model uncertainty in atomic location. Because of the high resolution of the structures presented here, the contribution of model error and noise to ADPs is minimized. Figure 3(A) presents isotropic ADPs from the four Mb structures in different crystal forms. All four structures show higher variance in loop regions compared with core helices. Three regions with the greatest differences in ADPs are the ones with the highest disparity in crystal contacts: CD corner, E-F loop, and GH loop. In each region, the ADPs are suppressed for structures with a high number of crystal contacts, for instance, in the GH loop region, the P6 space group structure (1JW8) has eight crystal contacts and the lowest variance, while the $P2_1$ structure (1BZR) shows the highest variance with only three crystal contacts. Similar patterns are seen in other variable regions, the E-F loop and the CD corner.

In addition to examining the magnitudes of ADPs, we also compare the directions of thermal ellipsoids generated by the anisotropic ADPs. We use only those which are sufficiently anisotropic for the directional analysis, specifically, with the ratio of smallest to largest principal axis less than 0.5. We compare ADPs from the corresponding atoms in different structures by computing the dot product between the largest principal axes of the ellipsoids, and by computing the overlap score for the

DOI 10.1002/prot

Table III
RMSD Between Myoglobin Structures for Cα and All Atoms

	1U7S	1U7R	1BZR	1JW8
1U7S	—	1.320	1.220	1.006
1U7R	0.784	—	1.348	1.262
1BZR	0.582	0.657	—	1.274
1JW8	0.535	0.651	0.523	—

*Upper right half of table represents all atoms rmsd and lower left half of table shows Cα rmsd (residues 1–151).

	1U7S	1U7R	1BZR	1JW8
1U7S	—	0.868	0.955	0.959
1U7R	0.308	—	1.025	1.075
1BZR	0.353	0.371	—	1.055
1JW8	0.315	0.279	0.370	—

*Upper right half of table represents all atoms rmsd and lower left half of table shows Cα rmsd (residues 6–17,22–34,60–77,102–117,126–148).

ADPs (see Materials and Methods). The results show close agreement in directionality of position distribution of atoms in different crystal forms (Table IV). The dot products and the overlap score are both in the range of 0.6–0.8 in pairwise comparisons of ADPs from both Cα and all atoms.

To test the hypothesis that Mb structures in different crystalline forms are conformers selected out of the solution state, we compared the set of four crystal structures with an NMR Mb ensemble. We use the NMR structure of a CO-bound Mb (PDB 1MYF[47]), consisting of 12 models that represent a converging set of structures derived from molecular dynamics simulations constrained by NOE distance and other sequential constraints. The variation observed in the crystal structures is very similar to the NMR ensemble, as shown in Figure 3(B). The NMR solution ensemble of Mb and ensemble of structures from different crystal forms are shown with the regions of highest variation highlighted. These include the termini, the CD corner, the E-F turn, and the GH loop. It is evident that the NMR ensemble shows greater positional deviation in most regions, as represented by the thickness of the ribbon. To enable a quantitative description of the dynamics observed by crystallography, we combine the positional variance from the four structures with the magnitude of the ADPs. Figure 3(C) overlays the plots of NMR positional variance and the crystallographic variation. The agreement is close, with a correlation coefficient of 0.80 (0.56 if three residues at each terminus are omitted). In the mobile regions mentioned above, both methods show significantly elevated variation. There are only two isolated locations where NMR ensemble shows an elevated variation that is not matched by a peak in the crystallographic variance: the loop between helices A and B, and a one-residue spike in the middle of helix H. Overall, the crystallographic variation apparently overestimates the variation in the core regions, due to rigid body motions in the lattice and noise contributions to the ADPs, and shows much lower variance in the mobile regions, while the NMR ensemble shows a more extreme contrast between the different parts of the protein, perhaps due to

Table IV
Dot Product and Overlap Fraction for ADPs for Cα and all Atoms

	1U7S	1U7R	1BZR	1JW8
1U7S	—	0.727/0.651	0.818/0.747	0.702/0.600
1U7R	0.735/0.658	—	0.819/0.765	0.758/0.690
1BZR	0.840/0.795	0.813/0.784	—	0.805/0.703
1JW8	0.738/0.623	0.776/0.724	0.769/0.687	—

*Values reported are mean dot product / mean overlap fraction. The comparison is for all residues, with upper right half of table showing results for all atoms and lower left half of table shows Cα.

	1U7S	1U7R	1BZR	1JW8
1U7S	—	0.725/0.658	0.818/0.745	0.691/0.598
1U7R	0.734/0.688	—	0.838/0.789	0.766/0.707
1BZR	0.811/0.777	0.867/0.820	—	0.815/0.721
1JW8	0.752/0.639	0.803/0.743	0.853/0.711	—

*Values reported are mean dot product / mean overlap fraction. The comparison is for residues not involved in crystal contacts, with upper right half of table showing results for all atoms and lower left half of table shows Cα.

variations in density of NOE constraints. Further, comparison of anisotropic variance in both ensembles resulted in median dot product of 0.65 and median overlap fraction of 0.46. The results indicate very similar dynamic profiles, indicating that the crystallographic ensemble is consistent with the solution NMR ensemble.

DISCUSSION

The extent of flexibility of protein structures varies greatly depending on location. Simulations have suggested that residues in the core have dynamic properties of a solid, while those in external residues show characteristics of a liquid.[48] The collection of Mb structures from four crystal space groups presented here shows structural variation, primarily restricted to external residues, of greater magnitude than in structures in the same space group with varying ligands and mutations. The basic question is whether this conformational heterogeneity is inherent in the native state solution ensemble, or largely induced by the crystalline environment. Our results strongly indicate the former mechanism.

A detailed examination of conformational changes in the CD corner and the GH loop shows no clear relationship between the number of crystal contacts in a region and its conformational deviation from other structures. The conformational variation appears to be the result of specific interactions involving flexible residues, such as the intermolecular main chain hydrogen bond formed in loop GH in space group $P6_122$, or the conformational change induced by imidazole binding in the CD corner, which is not constrained by crystal contacts in space group $P2_12_12_1$. As previously reported,[37] we observe that a greater number of crystal contacts suppresses the ADPs in participating residues, suggesting that the intermolecular interaction energy is primarily[17] stabilizing existing conformers. This is in conceptual agreement with the experimental evidence that crystallization is driven primarily by entropy gain of released solvent from protein–protein interfaces,[49] rather than by formation of strong interactions across crystal contacts.

The observed Mb dynamics are limited in scope and are not generated by distinct conformational states, such as those frequently associated with enzymatic reactions or motor proteins. The dynamic profile of the ensemble of X-ray crystal structures from four space groups agrees closely with the variance extracted from a 12-model solution NMR ensemble, with 0.8 correlation between variances of Cα atoms. The results are consistent with a study that found significant overlap between principal components from an NMR ensemble and the conformational differences in two crystal forms of photoactive yellow protein.[13] In the present study, the larger number of crystal structures allowed us to directly compare the ensembles, and to include the positional variances from ADPs which were not considered in the earlier work. The close agreement indicates that the conformational variation observed in the crystal forms is consistent with fluctuations observed in solution, and adds to the evidence that crystalline environment does not induce conformational substates significantly deviating from the range of allowed conformers in solution. However, the results caution against overinterpretation of crystal structures as representing the definitive and unique native conformation. Many external protein residues are inherently mobile and need to be modeled as a collection of conformers.

Some systematic differences were observed between the NMR- and crystallography-based positional variance, which suggests that both methods introduce errors and biases into the models. Crystallographic displacement parameters incorporate experimental noise and model error, leading to an overestimation of variance in the ordered core residues. Further, the Gaussian model of atomic positional variation is ill-suited for the positional distribution of atoms undergoing large anharmonic fluctuations,[50] such as those in the external residues, leading to underestimation of the flexibility, which is partially ameliorated by the inclusion of conformers from different crystal forms. On the other hand, the NMR positional ensemble is determined from NOE distance constraints, and in mobile regions with poor data quality the coordinates of the model are underconstrained and likely overestimate the actual positional variation. This is consistent with a recent study of sets of crystal structures of proteins with high sequence identity, which found that the crystal ensembles reproduce the NMR order parameters better than the NMR coordinate ensembles.[51]

Protein flexibility is essential for Mb to bind and release ligands and the differences between the structures from four space groups reveal information about the plasticity of Mb. The variability in the CD loop of the Mb crystallographic ensemble may provide room for the His64 to swing out towards the solvent and allow access for ligand entry as proposed by Johnson, *et al.*[52] This is most evident in the Im-bound $P2_12_12_1$ structure where His64 is rotated out towards the solvent resulting partly in the large deviation between the four space group structures in the CD loop. His64, in a $P2_1$ ethyl isocyanide-bound Mb structure,[52] is also rotated out towards the solvent, however, the CD loop does not deviate from other ligand bound $P2_1$ structures suggesting that the lack of packing contacts in the $P2_12_12_1$ structure in comparison to the other space group structures may allow for more flexibility in the CD loop region. The variation in the loops, such as the GH corner in the four structures, may allow for helical adjustments while the protein binds and releases its ligand. Comparative analysis of multiple structures from different crystalline environments enables a fuller representation of functionally significant conformational changes.

DOI 10.1002/prot

Conformational plasticity in proteins spans a variety of scales. In some, functional dynamics involves large scale transitions between distinct energy minima, usually referred to as "conformational changes," while in Mb small fluctuations around the distal pocket govern ligand binding and release. Thus, functionally relevant conformational fluctuations in this protein are roughly described by Gaussian distributions from equilibrium statistical mechanics, such as the ADPs. However, we have shown that the influence of contacts in crystal packing confines the system to a substate of an energy well, and that consideration of multiple substates can flesh out the native conformational ensemble of a protein structure.

CONCLUSION

Four high resolution structures of Mb in different crystal space groups were compared using structural and dynamic metrics. The differences in crystal contacts were found to induce conformational perturbations that were within the range observed in NMR models. Combined variance from ADPs and coordinate deviations is in good quantitative agreement with variance derived from NMR data. This indicates that a crystal structure represents a subset of conformational substates of the protein native state ensemble. While the coordinates for the immobile core residues are generally highly accurate, the flexible external regions are not adequately described by a single set of coordinates.

ACKNOWLEDGMENTS

We thank the staff at the Advanced Photo Source (APS) BioCARS beamline for assistance in collecting the diffraction data and Joel Berendzen and Ilme Schlicting for collection of the P6 data at NSLS. We also acknowledge John S. Olson for providing the recombinant protein and Tod Romo for help with refining the P6 data.

REFERENCES

1. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
2. Garcia AE, Blumenfeld R, Hummer G, Krumhansl JA. Multi-basin dynamics of a protein in a crystal environment. Physica D 1997; 107:225–239.
3. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC. Dynamics of ligand binding to myoglobin. Biochemistry 1975;14: 5355–5373.
4. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA. Conformational relaxation and ligand binding in myoglobin. Biochemistry 1994;33:5128–5145.
5. Elber R, Karplus M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. Science 1987;235:318–321.
6. Andrews BK, Romo T, Clarage JB, Pettitt BM, Phillips GN, Jr. Characterizing global substates of myoglobin. Structure 1998;6:587–594.
7. Garcia AE, Hummer G. Conformational dynamics of cytochrome c: Correlation to hydrogen exchange. Proteins: Struct Funct Genet 1999;36:175–191.
8. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, Kern D. Linkage between dynamics and catalysis in a thermophilic–mesophilic enzyme pair. Nat Struct Mol Biol 2004; 11:945–949.
9. Bae E, Phillips GN, Jr. Roles of static and dynamic domains in stability and catalysis of adenylate kinase. Proc Natl Acad Sci USA 2006;103:2132–2137.
10. Wong KF, Selzer T, Benkovic SJ, Hammes-Schiffer S. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. Proc Natl Acad Sci USA 2005;102:6807–6812.
11. Aranda R, Levin EJ, Schotte F, Anfinrud PA, Phillips GN, Jr. Time-dependent atomic coordinates for the dissociation of carbon monoxide from myoglobin. Acta Crystallogr D Biol Crystallogr 2006;62 (Part 7):776–783.
12. Schotte F, Lim M, Jackson TA, Smirnov AV, Soman J, Olson JS, Phillips GN, Jr, Wulff M, Anfinrud PA. Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. Science 2003;300:1944–1947.
13. van Aalten DM, Crielaard W, Hellingwerf KJ, Joshua-Tor L. Conformational substates in different crystal forms of the photoactive yellow protein—correlation with theoretical and experimental flexibility. Protein Sci 2000;9:64–72.
14. Frauenfelder H, Petsko GA, Tsernoglou D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. Nature 1979;280:558–563.
15. Wilson MA, Brunger AT. The 1.0 A crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol 2000;301:1237–1256.
16. DePristo MA, de Bakker PI, Blundell TL. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure 2004;12:831–838.
17. de Groot BL, Hayward S, van Aalten DM, Amadei A, Berendsen HJ. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. Proteins 1998;31:116–127.
18. Hall BE, Bar-Sagi D, Nassar N. The structural basis for the transition from Ras-GTP to Ras-GDP. Proc Natl Acad Sci USA 2002;99:12138–12142.
19. Kavanaugh JS, Rogers PH, Arnone A. Crystallographic Evidence for a new ensemble of ligand-induced allosteric transitions in hemoglobin: the T-to-T(High) quaternary transitions. Biochemistry 2005;44: 6101–6121.
20. Berry MB, Phillips GN, Jr. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap5A, Mg^{2+} Ap5A, and Mn^{2+} Ap5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg^{2+} and Mn^{2+}. Proteins 1998; 32:276–288.
21. Zhang XJ, Wozniak JA, Matthews BW. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J Mol Biol 1995;250:527–552.
22. Jelsch C, Longhi S, Cambillau C. Packing forces in nine crystal forms of cutinase. Proteins: Struct Funct Bioinformatics 1998;31:320–333.
23. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature 2005;433:128–132.
24. Radding W, Phillips GN, Jr. Kinetic proofreading by the cavity system of myoglobin: protection from poisoning. Bioessays 2004;26: 422–433.
25. Silva MM, Rogers PH, Arnone A. A third quaternary structure of human hemoglobin A at 1.7-A resolution. J Biol Chem 1992;267: 17248–17256.
26. Yang F, Phillips GN, Jr. Crystal structures of CO-, deoxy- and metmyoglobins at various pH values. J Mol Biol 1996;256:762–774.

27. Ansari A, Berendzen J, Braunstein D, Cowen BR, Frauenfelder H, Hong MK, Iben IET, Johnson JB, Ormos P, Sauke TB, Scholl R, Schulte A, Steinbach PJ, Vittitow J, Young RD. Rebinding and relaxation in the myoglobin pocket. Biophys Chem 1987;26:337–355.
28. Hummer G, Schotte F, Anfinrud PA. Unveiling functional protein motions with picosecond X-ray crystallography and molecular dynamics simulations. Proc Natl Acad Sci USA 2004;101:15330–15334.
29. Bossa C, Anselmi M, Roccatano D, Amadei A, Vallone B, Brunori M, Di Nola A. Extended molecular dynamics simulation of the carbon monoxide migration in sperm whale myoglobin. Biophys J 2004;86:3855–3862.
30. Teeter MM. Myoglobin cavities provide interior ligand pathway. Protein Sci 2004;13:313–318.
31. Olson JS, Soman J, Phillips GN. Ligand pathways in myoglobin: a review of Trp Cavity Mutants. IUBMB Life, in press.
32. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC. Structure of myoglobin: a three-dimensional Fourier synthesis at 2 A resolution. Nature 1960;185:422–427.
33. Phillips GN, Jr, Arduini RM, Springer BA, Sligar SG. Crystal structure of myoglobin from a synthetic gene. Proteins 1990;7:358–365.
34. Brucker EA, Olson JS, Ikeda-Saito M, Phillips J, George N. Nitric oxide myoglobin: crystal structure and analysis of ligand geometry. Proteins: Struct Funct Genet 1998;30:352–356.
35. Cheng XD, Schoenborn BP. Neutron-diffraction study of carbonmonoxymyoglobin. J Mol Biol 1991;220:381–399.
36. Kachalova GS, Popov AN, Bartunik HD. A steric mechanism for inhibition of CO binding to heme proteins. Science 1999;284:473–476.
37. Phillips GN, Jr. Comparison of the dynamics of myoglobin in different crystal forms. Biophys J 1990;57:381–383.
38. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276 (Part B): 307–326.
39. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54 (Pt 5):905–921.
40. Sheldrick GM, Schneider TR. SHELXL: high-resolution refinement. Methods Enzymol 1997;277 (Part B):319–343.
41. McRee DE. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999; 125:156–165.
42. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
43. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
44. Trueblood KN, Burgi HB, Burzlaff H, Dunitz JD, Gramaccioli CM, Schulz HH, Shmueli U, Abrahams SC. Atomic displacement parameter nomenclature—report of a subcommittee on atomic displacement parameter nomenclature. Acta Crystallogr Sect A 1996;52: 770–781.
45. Merritt EA. Comparing anisotropic displacement parameters in protein structures. Acta Crystallogr 1999;D55:1997–2004.
46. Miele AE, Federici L, Sciara G, Draghi F, Brunori M, Vallone B. Analysis of the effect of microgravity on protein crystal quality: the case of a myoglobin triple mutant. Acta Crystallogr D Biol Crystallogr 2003;59 (Part 6):982–988.
47. Osapay K, Theriault Y, Wright PE, Case DA. Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. J Mol Biol 1994;244: 183–197.
48. Zhou Y, Vitkup D, Karplus M. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. J Mol Biol 1999;285:1371–1375.
49. Vekilov PG, Feeling-Taylor AR, Yau ST, Petsev D. Solvent entropy contribution to the free energy of protein crystallization. Acta Crystallogr D Biol Crystallogr 2002;58 (Pt 10 Pt 1):1611–1616.
50. Garcia AE, Krumhansl JA, Frauenfelder H. Variations on a theme by Debye and Waller: from simple crystals to proteins. Proteins: Struct Funct Genet 1997;29:153–160.
51. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between native ensembles and experimental structures of proteins. Proc Natl Acad Sci USA 2006;103:10901–10906.
52. Johnson KA, Olson JS, Phillips GN, Jr. Structure of myoglobin-ethyl isocyanide. Histidine as a swinging door for ligand entry. J Mol Biol 1989;207:459–463.

 DOI 10.1002/prot