J|A|C|S
ARTICLES
Published on Web 04/01/2006

Modeling of Ligation-Induced Helix/Loop Displacements in Myoglobin: Toward an Understanding of Hemoglobin Allostery

Victor Guallar,*,† Andrzej A. Jarzecki,‡,# Richard A. Friesner,§ and Thomas G. Spiro*,‡

Contribution from the Department of Biochemistry, Washington University School of Medicine, St, Louis, Missouri 63108, Department of Chemistry and Center for Biomolecular Simulations, Columbia University, New York, New York 0027, and Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received October 26, 2005; E-mail: spiro@Princeton.edu; guallarv@biochem.wustl.edu

Abstract: Combining quantum and molecular mechanics (QM/MM) methods and protein structure prediction algorithms, helix and loop movements are computed along the pathway of CO dissociation from myoglobin (Mb). The results are compared with high-resolution crystallographic data using sequence-displacement graphs. These graphs provide an unbiased method for evaluating main-chain segmental motions; they resolve an apparent disagreement between two sets of high-resolution crystal structures for MbCO and deoxyMb. The QM/MM modeling of the CO deligation reproduces the experimentally observed spin states and photodissociated crystal structure. The principal effect of CO dissociation is shown to be a concerted rotation of the E and F helices, which hold the heme like a clamshell. The rotation is a response to deligation forces, which impel the F helix away from the heme because of the Fe spin conversion, and which allow the E helix to collapse toward the heme as nonbonded contacts on the distal side are relieved. Additional helix and loop displacements stem from these primary events. In particular, the CD loop is found to be repositioned as a result of steric interactions with the water molecule that becomes H-bonded to the distal histidine in deoxyMb. A similar EF rotation and CD loop displacement are proposed to be the first steps along the allosteric pathway from the R to the T state in hemoglobin.

1. Introduction

Hemoglobin (Hb), the paradigmatic allosteric protein, provides a laboratory for understanding the mechanisms whereby proteins change their shape in response to the binding of small molecules, a phenomenon fundamental to a wide rage of biological processes. The alternative quaternary structures of Hb, R and T, are known in atomic detail,[1−7] but the pathway by which one is converted to the other continues to challenge both experimental and theoretical investigations.

Time-resolved vibrational spectroscopy has yielded insight into the structures of kinetic intermediates along the allosteric pathway.[8,9] When the CO ligand is photodissociated from HbCO, the protein, initially in the R structure, undergoes a series of kinetically defined motions leading to the T structure, before the CO recombines from solution. The first of these motions breaks interhelical H-bonds involving the E and F helices, which hold the heme like a clamshell (Figure 1). These H-bonds are then re-formed via further helix motions, which induce the subunit rotation that establishes the T-state quaternary contacts. The initial motion is proposed to involve rotation of the EF helix clamshell in response to deligation, in line with the original stereoelectronic hypothesis of Perutz.[10] When the CO dissociates from the heme, the Fe^{2+} ion converts from a low- to a high-spin state and moves out of the heme plane toward the proximal histidine ligand (H87 in Figure 1). The proximal histidine is on the F helix, and the electronic change therefore creates a downward force on the F helix, away from the heme. At the same time, relief of steric interactions between the bound CO and E helix residues on the distal side of the heme permits the E helix to collapse toward the heme as the CO departs the binding site. These combined effects can lead to a concerted rotation of the EF clamshell.[9,11]

† Washington University.
‡ Princeton University.
§ Columbia University.
# Current address: Department of Chemistry, Brooklyn College and the Graduate School of the City University of New York, Brooklyn, NY 11230.

(1) Baldwin, J.; Chothia, C. *J. Mol. Biol.* **1979**, *129*, 175−200.
(2) Park, S.-Y.; Tame, J. R. H. RCSB Protein Data Bank, 2001, PDB code 1IRD.
(3) Safo, M. K.; Burnett, J. C.; Musayev, F. N.; Nokuri, S.; Abraham, D. J. *Acta Crystallogr. D, Biol. Crystallogr.* **2002**, *58*, 2031−2037.
(4) Silva, M. M.; Rogers, P. H.; Arnone, A. *J. Biol. Chem.* **1992**, *267*, 17248−17256.
(5) Lukin, J. A.; Kontaxis, G.; Simplaceanu, V.; Yuan, Y.; Bax, A.; Ho, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 517−520.
(6) Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R. *J. Mol. Biol.* **1984**, *175*, 159−174.
(7) Tame, J. R. H.; Vallone, B. RCSB Protein Data Bank, 1998, PDB code 1A3N.
(8) Balakrishnan, G.; Case, M. A.; Pevsner, A.; Zhao, X. J.; Tengroth, C.; McLendon, G. L.; Spiro, T. G. *J. Mol. Biol.* **2004**, *340*, 843−856.
(9) Balakrishnan, G.; Tsai, C. H.; Wu, Q.; Case, M. A.; Pevsner, A.; McLendon, G. L.; Ho, C.; Spiro, T. G. *J. Mol. Biol.* **2004**, *340*, 857−868.
(10) Perutz, M. F.; Wilkinson, A. J.; Paoli, M.; Dodson, G. G. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 1−34.

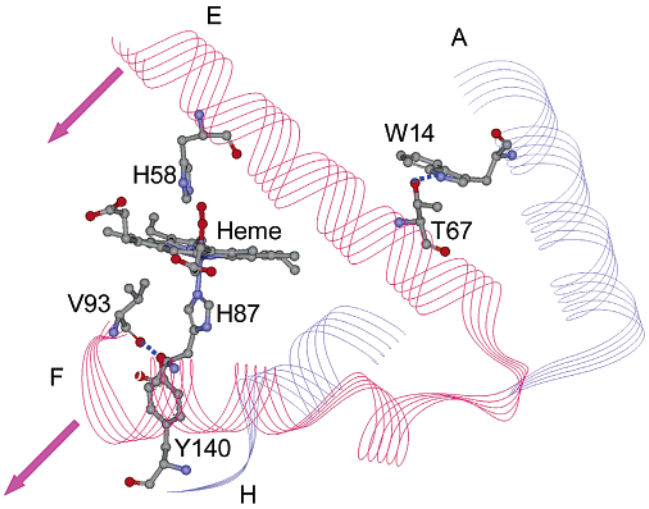


Figure 1. Rotation of the EF clamshell, with breaking of mechanical H-bonds, is proposed to be the first step in the R−T allosteric pathway of Hb. Residue numbering for the α chain (PDB code 1HHO).

This view of the initial response of the Hb molecule to deligation received striking support from high-resolution crystal structures of myoglobin (Mb) in its unligated and CO-bound forms. Mb is a monomeric protein but has the same helix architecture as Hb. Kachalova et al.[12] (referred to as KPB) superimposed the MbCO and deoxyMb structures and reported E and F helix displacements corresponding to the proposed clamshell rotation. However, an independent high-resolution crystallographic study by Vojtechovsky et al.[13] (referred to as VCBSS) did not support rotation of the helices between the two structures.

To understand the apparent discrepancy, we have reexamined the structural data, using sequence-displacement graphs as a tool for comparison. These graphs confirm the EF helix rotation in both sets of structures, although there are quantitative differences between the graphs. We have also examined the mechanism of ligation-induced helix motion computationally. Because the mechanism links electronic change at the heme with protein motion at a distance, the computation requires a combination of quantum and molecular mechanics (QM/MM) methods with protein structure sampling tools. The helix displacements are small, <1 Å, putting severe demands on the quality of the computations. We are encouraged, however, by the success of recent calculations combining QM/MM methods with protein structure prediction algorithms.[14,15] Our results indicate that such a combination allows for an accurate characterization of protein conformational changes coupled to a quantum mechanical description of the active site.

2. Methods

2.1. QM/MM Protocol. QM/MM and QM Models. The QM/MM model has been derived from the experimental structure (PDB code 1AJG). The full protein plus all crystallographic waters were placed in an equilibrated box containing 6753 molecules of SPC waters, defining a 15 Å solvation layer of waters around the protein. Upon inspection of the local environment, histidines 64, 82, 97, and 119 had their N-bound proton placed at the ϵ N (neutral histidines). The waters and all the hydrogen atoms of the protein were further equilibrated by heating the system to 300 K in five steps of 1 ps each, followed by 10 ps of molecular dynamics at 300 K and another five steps of 1 ps during which we cooled the system. Periodic boundary conditions were used through the annealing process. The system was then reduced by deleting all water beyond 10 Å from any protein atom. The final system consisted of 7900 atoms. The QM region for all QM/MM calculations included the iron atom, the CO (or water) ligand, the protoporphyrin ring, and the proximal histidine (His93). Geometry optimizations were carried out using the B3LYP functional in combination with the 6-311G* (LACVP* effective core potential for the iron). CO dissociation was modeled by constraining the Fe−C distance while optimizing all other parameters, with the exception of the last solvation layer. After the deligation event, two more structures were modeled, first by removing the CO ligand and second by inserting a water molecule (not bound to the heme group). This latter model aims to reproduce the deoxy crystallographic structures, where a loosely bound water molecule is seen in the active site. For all geometry optimizations, the coordinates of the oxygen in the last layer of waters (all oxygens from 8 Å to 10 Å) were held fix. The QM/MM calculations were carried out with the QSite program,[16] which was developed via a close coupling of Jaguar and the IMPACT[17] protein modeling program of Levy and co-workers. The QM/MM methodology and protocol, as well as extensive validation studies demonstrating excellent agreement with fully QM calculations for peptides and protein active sites, have been described elsewhere.[18,19] Recent studies on enzymatic reactions have demonstrated the capabilities of this methodology when applied to large biological systems.[14,15,20,21]

We have also performed the deligation study using a gas-phase active site model. This reduced QM model consisted of (ImH)FeP(CO), the CO adduct of Fe(II) porphine, with imidazole as the sixth ligand. This complex has been used in previous computations.[22−25] Comparison of both QM and QM/MM models allows us to address the role of the protein in the dissociation process. All gas-phase electronic ab inito calculations were performed with the Jaguar[26] program and with the same basis set and functional used in the QM/MM study.

2.2. Protein Structure Prediction and Refinement. We have developed methods for high-resolution protein structural prediction and refinement based on the use of an accurate protein molecular mechanics force field and a surface-generalized Born (SGB) continuum solvent model.[27] For the molecular mechanics potential energy function, we used a recently updated version of the OPLS-AA parametrization (OPLS2001), which has been fit to high-level ab initio quantum chemical data for more than 200 rotamer states of the various amino acid dipeptides. The new force field was demonstrated[28] to provide significant improvements in side-chain prediction as the fit to quantum chemical data in experimentally relevant regions of phase space was improved. We employ specialized sampling algorithms for both loop

(11) Rodgers, K. R.; Spiro, T. G. *Science* **1994**, *265*, 1697−1699.
(12) Kachalova, G. S.; Popov, A. N.; Bartunik, H. D. *Science* **1999**, *284*, 473−476.
(13) Vojtechovsky, J.; Chu, K.; Berendzen, J.; Sweet, R. M.; Schlichting, I. *Biophys. J.* **1999**, *77*, 2153-2174.
(14) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *J. Mol. Biol.* **2004**, *337*, 227−239.
(15) Guallar, V.; Borrelli, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3954−3959.
(16) *QSite*; Schrödinger, Inc., Portland, OR, 2001.
(17) *IMPACT*; Schrödinger, Inc., Portland, OR, 2000.
(18) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442−1457.
(19) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468−1494.
(20) Guallar, V.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 8501−8508.
(21) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. A. *J. Comput. Chem.* **2005**, *26*, 915−931.
(22) Harvey, J. N. *J. Am. Chem. Soc.* **2000**, *122*, 12401−12402.
(23) Ugalde, J. M.; Dunietz, B.; Dreuw, A.; Head-Gordon, M.; Boyd, R. J. *J. Phys. Chem. A* **2004**, *108*, 4653−4657.
(24) McMahon, B. H.; Stojkovic, B. P.; Hay, P. J.; Martin, R. L.; Garcia, A. E. *J. Chem. Phys.* **2000**, *113*, 6831−6850.
(25) Spiro, T. G.; Kozlowski, P. M. *J. Am. Chem. Soc.* **1998**, *120*, 4524−4525.
(26) *Jaguar 4.1*; Schrödinger, Inc., Portland, OR, 2000.
(27) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.
(28) Jacobson, M. P.; Friesner, R. A.; Xiang, Z. X.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597−608.

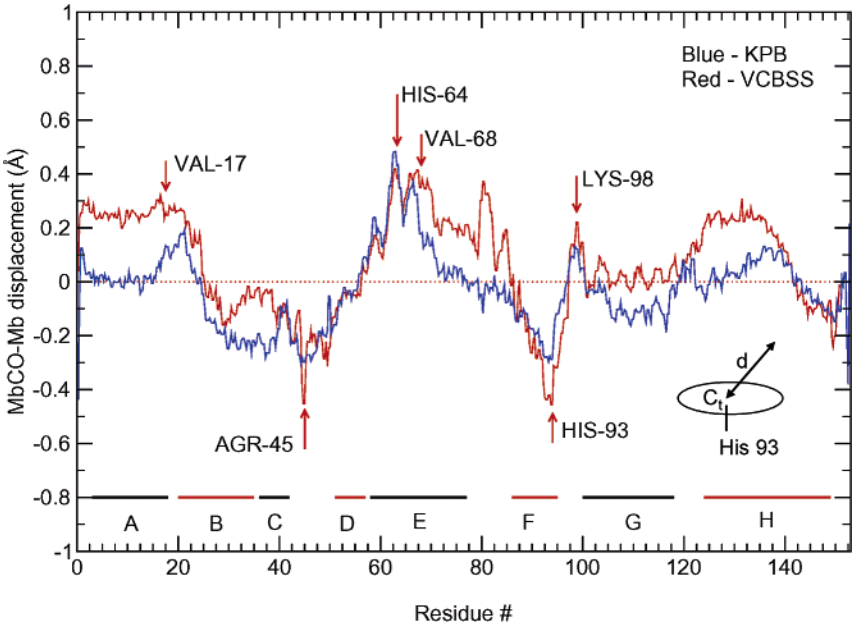


Figure 2. Sequence-displacement graph (d plot) for deoxyMb relative to MbCO, comparing the two sets of high-resolution crystal structures: KPB[12] (blue) and VCBSS[13] (red). The data are differences in distance from the center of the porphyrin N atoms to successive main-chain atoms, N, C_α, and C(O). Helix units (A−H) are indicated near the bottom.

and side-chain prediction.[29,30] These algorithms include the use of highly detailed rotamer-state libraries for side-chain conformational searching, hierarchical screening methods based on steric overlap and approximate electrostatics to rapidly eliminate obviously incorrect conformations, and a multiscale minimization algorithm that is 1−2 orders of magnitude more efficient than conventional approaches based, for example, on conjugate gradient and uniform evaluation of the energy at every minimization step.

To project conformational changes along the deligation process, we adopted the following sampling protocol. Starting from the fully optimized CO-bound structure, we perturbed the system by inserting the quantum mechanical region geometry and charges from a QM/MM intermediate following the Fe−CO unbinding reaction coordinate. The quantum mechanical region includes the heme group, the CO ligand, and His93. We then proceeded with an iterative procedure where we sampled all side chains and performed a full minimization, excluding the quantum mechanical region in both the side-chain sampling and the minimization. The process iterated until two consecutive sampled structures did not introduce any further modification, using the criterion of 0.01 Å rmsd for α carbons. This final structure was then perturbed again, introducing the geometry of another intermediate point along the CO deligation coordinate.

3 Results and Discussion

3.1. Sequence-Displacement Analysis of Crystallographic Data. To evaluate helix and loop displacements between MbCO and deoxyMb, we calculated the distances from the center of the four N atoms of the porphyrin ring to three main-chain atoms (N, C_α, C(O)) on each residue. The differences in these distances between deoxyMb and MbCO were then plotted against sequence for the high-resolution KPB[12] and VCBSS[13] structures. These plots (Figure 2) show excursions of up to 0.5 Å, although the displacements are mostly <0.2 Å. Importantly, a periodic pattern is seen, indicative of concerted motions of backbone segments. Particularly significant is the prominent downward-sloping segment encompassing the E and F helices. The largest positive displacement is for His 64 (the distal histidine residue), near the start of the E helix, while the largest negative displacement is for His 93 (the proximal histidine) at the end of the F helix; between these points, the displacements descend monotonically. This is precisely the pattern expected for the postulated EF helix rotation; when the CO is lost, the E helix moves toward the heme, producing positive displacements, while the F helix moves away from the heme, producing negative displacements. The displacements are close to zero near the EF corner, which acts as a pivot for the rotation. *Both* sets of structures show this pattern, although there is a quantitative difference in the middle of the E helix. Thus, the EF rotation is a reproducible feature of the high-resolution crystal structures.

The sequence-displacement plots reveal additional backbone motions. The EF rotation appears to be accompanied by a counter-rotation of the D helix and the first part of the E helix, on one end, and of the FG corner on the other. The two sets of data are coincident in these regions, both showing a negative peak (Arg 45) at the loop connecting the C and D helices and a positive peak (Lys 98) at the FG corner. They both show an additional positive peak at the AB corner. However, they differ somewhat at the A, B, and H helices, the VCBSS displacements lying some 0.2 Å higher than the KPB displacements in these regions. These quantitative deviations may be connected with the proximity of the A and H regions to the C- and N-termini of the protein chain, which may be sensitive to the different experimental conditions and crystal packing.

Figure 2 (the d plot) does not capture the actual directions of the displacements, but only their sign and magnitude with respect to a common point. To examine the motions three-dimensionally, we decomposed the displacement vectors into *x*, *y*, and *z* components (Figure 3), relative to the heme plane. *z* is perpendicular to the plane, while the *x* and *y* directions were chosen to bisect the A and C pyrrole rings on the one hand, and the B and D rings on the other. Figure 4 is a ribbon

(29) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.
(30) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Tyler, J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins* **2004**, *55*, 351−367.

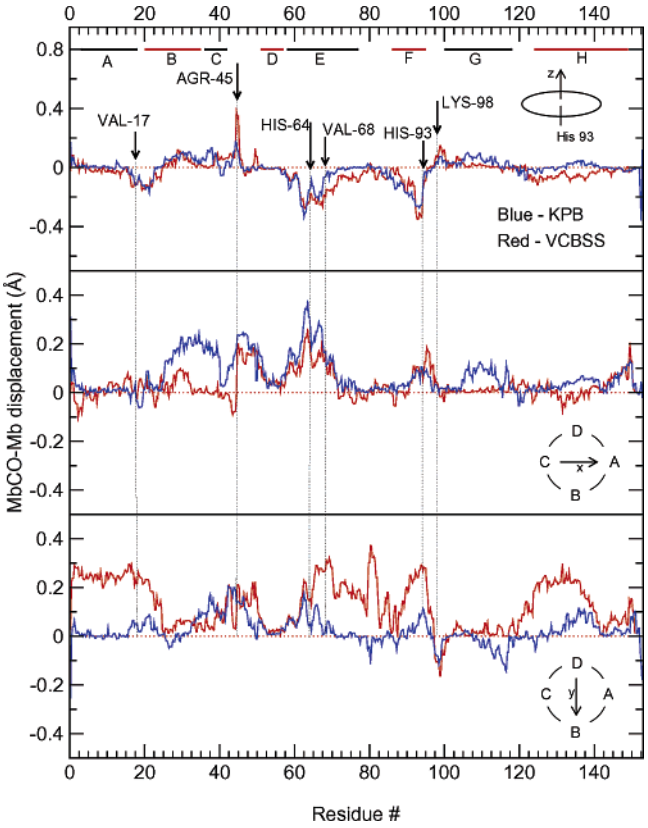


Figure 3. As Figure 2, but showing the *z* component of the displacement perpendicular to the heme plane (top), the *x* component of the displacement along A and C ring N atoms (middle), and the *y* component of the displacement along B and D ring N atoms (bottom).

diagram of MbCO, projected on the *zx* and *xy* planes; the color shades (see legend) indicate the key backbone motions.

The *z* displacements (top plot in Figure 3) strikingly confirm the EF rotation hypothesis. Two prominent negative peaks are seen, near the start of the E helix and at the end of the F helix; the peaks are negative because the lips of the clamshell both move along $-z$ in the concerted rotation (Figure 1). Near the EF pivot, the displacement is close to zero, as expected. The *z* displacement reveals a smaller positive peak at the FG corner, accounting for part of the FG counter-rotation in the d plot (Figure 2). There are also positive *z* displacements at the CD loop and along the C and B helices, and a negative peak at the AB corner. The two sets of crystal structures give essentially coincident *z* displacements.

Likewise, the two sets of crystal structures show essentially the same *x* displacements (middle plot in Figure 3), with the significant exception of the B and C helices; these show a positive displacement for the KPB but not for the VCBSS structures. This disparity can be traced to the different location of two sulfate ions in the two sets of crystallographic structures largely affecting the side-chain positions of the residues in the vicinity of the BC corner.Both sets of structures show a $+x$ peak near the beginning of the E helix, of a magnitude similar to that of the $-z$ peak. This bisection of $+x$ and $-z$ reflects the ~45° angle of the E helix relative to the heme (Figure 4). The F helix, which is parallel to the heme plane, shows little *x* displacement.

The greatest difference between the two sets of structures is revealed in the *y* displacement plot (Figure 3). The VCBSS structures display ~0.2 Å displacements along the A, E, F, and H helices, while the KPB structures do not. This systematic difference is equivalent to a $+y$ displacement of the heme group in the frame of these helices (Figure 4). This displacement might reflect differential interactions of the heme propionate groups with water and/or ions in the crystals.

3.2. Computed Protein Deligation and Displacements. QM/MM Deligation. Initial calculations involved optimizing the MbCO structure starting with crystal coordinates (PDB code 1AJG) and using the QM/MM protocol. This crystal structure[31] represents a different set of MbCO coordinates than that of KPB and VCBSS, reducing any possible initial bias. QM/MM optimization produced a rmsd of 0.4 Å for the α carbon atoms and 0.55 Å for all heavy atoms. The only large deviations, >0.4 Å, occur at the C- and N-terminal sites of the protein. The computed electronic energies and relevant geometries along the CO dissociation are shown in Table 1. In both the QM and QM/MM models, the singlet state is the ground state at the equilibrium Fe−CO distance and is steadily destabilized as the Fe−CO distance is increased. Importantly, the computation yields the same small (~8°) deviation from FeCO collinearity as seen crystalographically. In agreement with experiment and with some of the previous computational studies,[22,32−34] the deoxy ground state corresponds to the quintet state. In this state, the Fe−N distances expand and the Fe is displaced from the porphyrin plane, as electrons occupy the antibonding iron orbitals ($d_{x^2-y^2}$ and d_{z^2}).[23] However, the iron is also displaced, though to a lesser extent, in the singlet and triplet states.

Removing the Fe−C constraint at the QM/MM 3.25 Å point relaxed this distance only slightly, to 3.21 Å, and the energy diminished by only 0.2 kcal/mol. This deligated QM/MM model gave a remarkably accurate (0.34Å heavy atom rmsd) superposition with the low-temperature crystal structure of photodissociated MbCO,[31] with the CO lying parallel and adjacent to the heme (see Figure S1, Supporting Information). Both orientations were tested, with the C or the O atom closest to the Fe, and found to produce negligible rearrangement of the rest of the structure (0.04 all atom rmsd), with an energy difference of only 0.1 kcal/mol between them. The two orientations have been recently interpreted as the origin of the Stark splitting in the infrared transition of the dissociated CO.[35]

The dissociation energy profiles are similar for the QM and QM/MM models. The quintet state becomes the ground state just before the Fe−C separation reaches 2.6 Å distance, giving an estimate for the recombination barrier of ~2 kcal/mol for both models. This value is in good agreement with previous calculations (2.4 kcal/mol, Harvey[22]) and with experiment.[36] However, the triplet and quintet states are nearly degenerate at the crossing with the singlet. Moreover, vertical excitation from the singlet structure is lower in energy for the triplet than for the quintet state at all Fe−C separations. Thus, the triplet state no doubt plays an important role in the singlet-to-quintet transition (spontaneous or induced dissociation) or in the reverse process (geminate recombination); a superexchange mechanism

(31) Teng, T. Y.; Srajer, V.; Moffat, K. *Nat. Struct. Biol.* **1994**, *1*, 701−705.
(32) Dunietz, B. D.; Dreuw, A.; Head-Gordon, M. *J. Phys. Chem. B* **2003**, *107*, 5623−5629.
(33) Harvey, J. N. *Faraday Discuss.* **2004**, *127*, 165−177.
(34) Sigfridsson, E.; Ryde, U. *J. Inorg. Biochem.* **2002**, *91*, 101−115.
(35) Nienhaus, K.; Olson, J. S.; Franzen, S.; Nienhaus, G. U. *J. Am. Chem. Soc.* **2005**, *127*, 40−41.
(36) Steinbach, P. J.; et al. *Biochemistry* **1991**, *30*, 3988−4001.

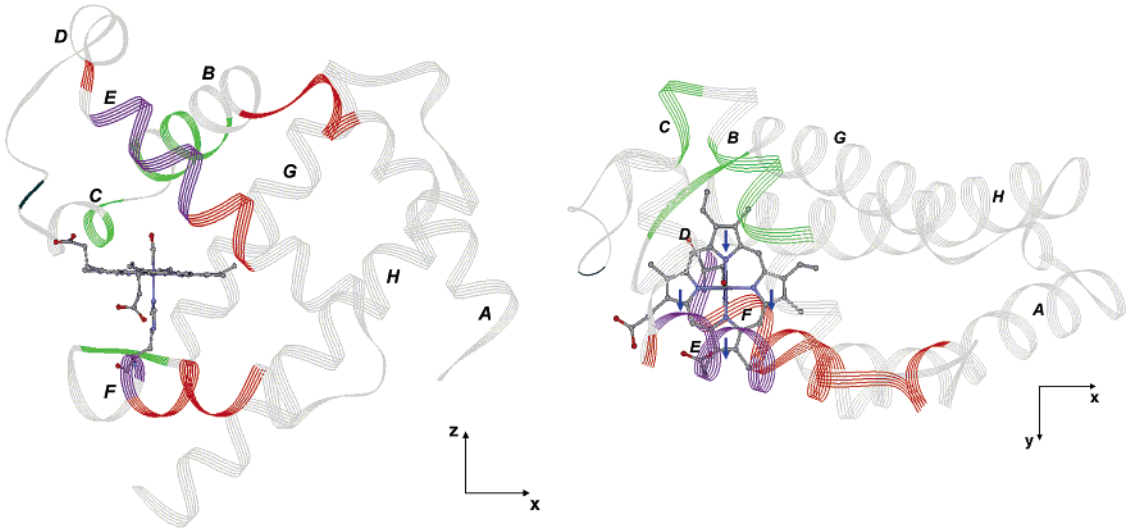
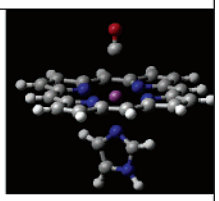
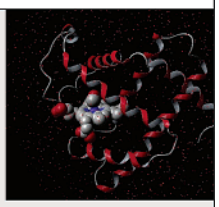


Figure 4. Ribbon diagram of MbCO structure projected on the *xz* (left) and *xy* (right) plane. Shades of color indicate extent of displacement along the *z* axis for residues showing the greatest deligation-induced changes. Purple and red indicate large and small $-z$ displacements, while dark and light green indicate large and small $+z$ displacements. The blue arrows indicate the direction of the ligation-induced heme shift in the VCBSS structures.

Table 1. QM and QM/MM (Shaded) Energies and Relevant Distances for the Single, Triplet, and Quintet Spin States along the Deligation Coordinate[a]

	Fe-C: Min:1.8	Fe-C:2.3	Fe-C:2.6	Fe-C:3.25	Model
S_0	0.0 2.09/2.03-0.03	13.1 2.00/2.02/0.05	19.5 1.99/2.02/0.10	22.2 1.98/2.02/0.13	
T_0	31.8/27.3 2.08/2.09/-0.03	28.4/20.9 2.33/2.03/0.04	27.0/19.8 2.31/2.03/0.08	26.2/19.1 2.29/2.03/0.12	
Q_0	51.0/38.0 2.45/2.10/-0.15	33.1/20.8 2.30/2.09/0.10	30.0/18.9 2.26/2.09/0.17	31.1/17.0 2.23/2.10/0.23	
S_0	0.0 2.13/2.04/-0.02	11.6 2.04/2.03/0.06	17.8 2.02/2.03/0.11	24.0 2.02/2.03/0.14	
T_0	30.1/26.3 2.11/2.10/-0.01	26.9/16.5 2.34/2.03/0.06	24.8/16.3 2.34/2.03/0.09	24.1/15.9 2.33/2.03/0.12	
Q_0	40.4/36.6 2.33/2.11/-0.06	29.3/17.2 2.29/2.10/0.12	27.6/16.2 2.27/2.10/0.15	28.8/14.3 2.19/2.11/0.27	

[a] The upper numbers (blue) are the computed vertical/relaxed energies at the indicated Fe−C separation (kcal/mol, relative to the singlet-state minimum for the unconstrained CO adduct). The vertical energies were computed at the singlet minimized geometries, while the relaxed energies were obtained after geometry optimization. The lower numbers (black) are the optimized distances (Å) from the iron to the imidazole N atom, to the porphyrin N atoms, and to the center of the porphyrin ring (Fe−N_I/Fe−N_P/Fe−Ct). The sign of Fe−Ct indicates displacement toward the imidazole (+) or the CO (−) ligand.

may not be applicable. However, the choice of basis set influences the computed energy differences of the heme spin states.[22,23,37,38] Further studies employing a larger basis and higher-level correlation methods should address this point.

The character of the lowest triplet state changes with increasing Fe−C separation. At Fe−C = 1.8 Å, the iron−imidazole distance (Fe−N_I) is the same for the singlet and triplet states, but the iron−porphyrin (Fe−N_P) distance expands by 0.06 Å for the triplet, reflecting electron occupation of the $d_{x^2-y^2}$ orbital; the d_{z^2} triplet appears ∼7 kcal/mol higher in energy. However, at Fe−C = 2.3 Å, the Fe−N_P distance shrinks back to the singlet value, while the Fe−N_I distance expands, by a full 0.25 Å, as the antibonding electron switches from the $d_{x^2-y^2}$ to the d_{z^2} orbital. Thus, there is a diabatic crossing of triplet states in the initial phase of Fe−CO dissociation.

A key issue in the dissociation mechanism is the impulse toward a change in spin state. Clearly the quintet (and also the $d_{x^2-y^2}$ triplet) is stabilized by displacement of the Fe atom from the heme plane, which reduces the overlap of the porphyrin N orbitals with the $d_{x^2-y^2}$ orbital. What induces the Fe displacement? The standard view[5,39,40] is that steric repulsion between the imidazole and the porphyrin ring is responsible. However, our computations indicate a negligible change in the imidazole−porphyrin separation (sum of Fe−N_I and Fe−C_t distances).

(37) Rovira, C.; Ballone, P.; Parrinello, M. *Chem. Phys. Lett.* **1997**, *271*, 247−250.
(38) Green, M. T. *J. Am. Chem. Soc.* **1998**, *120*, 10772−10773.
(39) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1789−1793.
(40) Olafson, B. D. G. W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1315−1319.

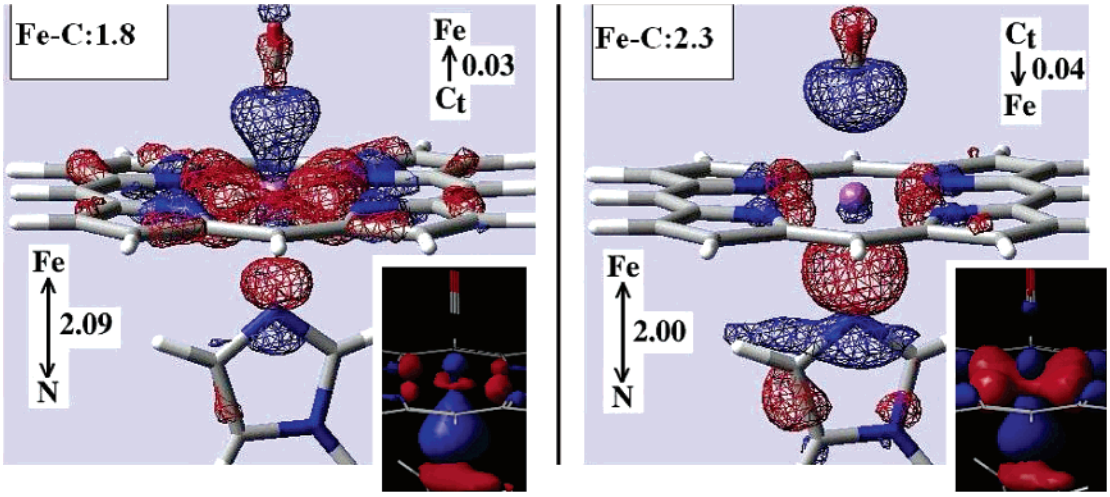


Figure 5. Molecular orbitals describing the evolution of σ bonding changes along the CO deligation coordinate in the singlet spin state.

Instead, there is a ~0.1 Å *decrease* in the Fe−N_I distance by the time the Fe−CO bond is stretched by 0.5 Å. A smaller increase, 0.05 Å, is seen only at the dissociated geometry. Thus, the initial out-of-plane displacement is induced by Fe−N_I bond shortening, rather than ligand steric repulsion. The reason for the abrupt shortening of Fe−N_I can be traced to σ competition between the CO and ImH lone-pair orbitals for the Fe d_{z^2} orbital, which is slightly mixed with the ligand bonding orbitals. Figure 5 shows the evolution of key occupied molecular orbitals as the Fe−CO bond is stretched. As bonding overlap is lost for Fe−CO (main figure, orbital 1), it increases for Fe−N_I (inset, orbital 2). Thus, the Fe−N_I bond strengthens as σ competition with the departing CO diminishes, and the Fe moves out of the heme plane to shorten the bond. (Steric factors do play a role, of course, or else the imidazole would have approached the heme even more closely.) The out-of-plane motion in turn stabilizes the higher spin states.

Finally, we note that the surrounding protein has little effect on the energetics or the geometry of the heme. The computed energies, relative to the equilibrium singlet state, are ~1−2 kcal/mol lower for the QM/MM than the QM model. The state crossing points and the estimated recombination barrier are essentially the same. The computed bond distances and Fe displacements are also very similar. The only notable difference is that the Fe−N_I distance of the starting structure is 0.05 Å *longer* in the QM/MM than in the QM model, suggesting that the protein is predisposed to stabilize the high-spin structure. However, the energy involved in this initial bond stretching is small. When Fe−N_I was constrained to the initial QM model distance, the QM/MM energy rose by only 0.5 kcal/mol.

Protein Response. We examined progressive changes in the protein backbone along the deligation coordinate using the same sequence-displacement plots discussed above. Starting from the QM/MM structures, the protein response to the CO deligation was further sampled using protein structure prediction algorithms as described in the Methods section. Figure 6 shows the maps for three different models along the oxy-to-deoxy transition: Fe−CO distance of 2.3 Å, deletion of CO (no CO), and insertion of a water molecule. As the Fe−CO bond is stretched, the protein evolves in the direction seen in the MbCO/deoxyMb crystal structure comparisons. In the final simulation, with water present in the heme pocket, the displacement plot fits the

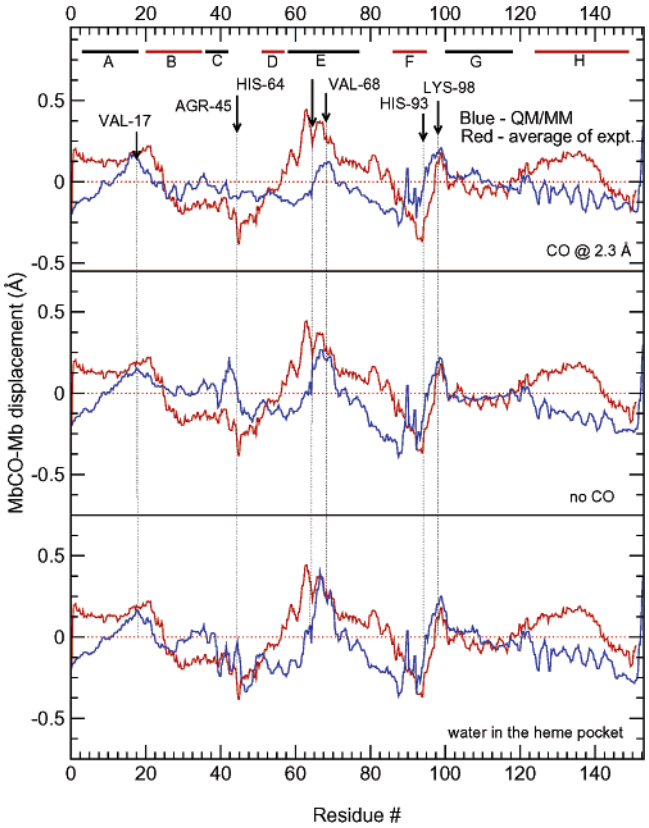


Figure 6. Sequence-displacement graphs (d plots) from the QM/MM computations at successive stages on the CO dissociation pathway. The average for the two crystallographic data sets (shown in red) is shown for comparison.

experimental data quite well. The EF rotation, the D and G helix counter-rotation, and the displacements of the CD loop and the AB corner are all present in the simulation and occur to the observed extent. Most of the deviations are in the same regions where the two crystallographic data sets show inconsistencies (Figure 2).

However, there is significant deviation from both sets of crystal structures in the region around residue 60 (ED corner), including the distal His64; this residue is strongly displaced in the crystal structures but not in the simulation. The reason for this difference is that the adjacent pair of residues, Lys63 and

Lys62, have positively charged side chains that extend from the protein surface and are associated with a sulfate ion in the crystals.[12,13] This feature was, of course, not included in the computation and may not be relevant to physiological conditions. We infer that the ionic interaction influences the displacement of the ED corner and of His64.

Although His64 is in contact with the bound CO in MbCO, analysis of the effects of His64 mutations on binding affinity and kinetics indicates that the interaction is energetically unimportant.[41,42] Consistent with this inference, the MbCO crystal structures show that the His64 side chain is not tightly packed and experiences large-amplitude motions.[12,13] It is therefore not surprising that the main chain can be distorted at this position by environmental interactions.

However, Val68, whose side chain is also in contact with bound CO, is rigidly held, and it moves toward the Fe in the transition to deoxyMb. Kachalova et al.[12] proposed that this motion might account for Mb's ability to discriminate against CO and in favor of O_2 binding. However, replacing Val68 with the sterically less demanding residue glycine does not alter the CO/O_2 binding ratio.[41] The CO/O_2 discrimination is attributable instead to stronger H-bonding of the His64 side chain to bound O_2 than to bound CO. Nevertheless, the Val68Gly substitution does increase the affinity for both ligands,[41] showing the nonbonded contact to be energetically significant. In addition, a fully QM computation of the MbCO binding pocket has shown that Val68, and not His64, is responsible for the observed ~8° deviation from linearity of the FeCO unit.[43] Thus, relief of the Val68···CO contact is the likely mechanism for the E helix motion toward the heme as the CO dissociates. The simulation is in exact agreement with crystallography with respect to the Val68 displacement (Figure 6, bottom panel).

Likewise, the observed displacement of His93, at the end of the F helix, is exactly reproduced, as is that of the FG corner residue Lys98. His93 is the proximal histidine that is ligated to the heme Fe. It is impelled downward when the Fe lengthens its bonds and is displaced from the heme plane as a result of its singlet/quintet conversion upon loss of CO. The downward motion applies a torque to the FG corner, impelling Lys98 upward.

The EF clamshell rotation already appears in the simulation at Fe···C = 2.3 Å (top panel in Figure 6); the displacements are small, but one can clearly see movements in the expected direction for the E and F helices, and for the FG corner. We were surprised to see, at this early stage in the CO deligation reaction coordinate, additional displacement in the expected direction at the distant (25 Å from the heme Fe) AB corner (Val17). Analysis of the intervening residues, however, revealed a direct steric connection between Val 17 and Val68 via the side chain of Leu69 (Figure 7). Thus, relief of the CO···Val68 nonbonded contact is tightly coupled to displacement of the AB corner, as well as the E helix.

When the CO is removed from the binding pocket (Figure 6, middle panel), the EF helix rotation reaches nearly its full extent. (Essentially the same displacements are seen when Fe····C is increased to 3.25 Å.) However, the observed downward

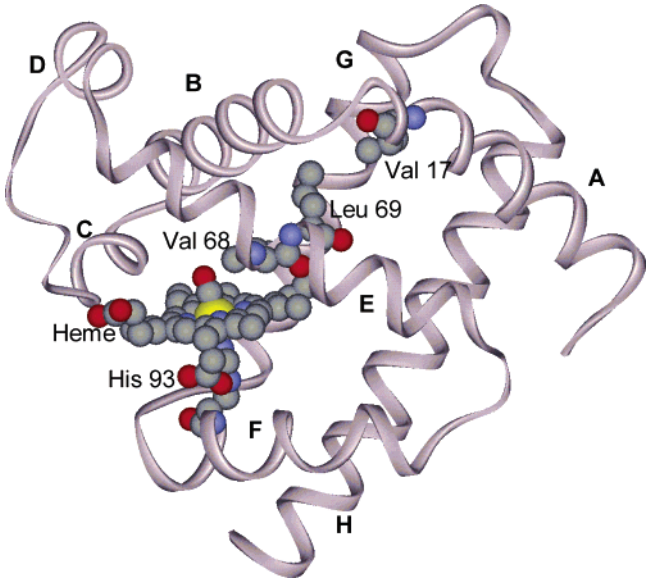


Figure 7. MbCO structure showing juxtaposition of Val 68 on the bound CO, and the steric interactions path through Leu69 to Val17, at the AB corner.

displacement of the CD loop is not reproduced in this simulation. Inclusion of the His64-bound water molecule remedies this deficiency, and the CD loop is now correctly displaced (Figure 6, bottom panel). The physical mechanism involves nonbonded contact of the water molecule with the side chains of the CD loop residues Phe43 and Phe46, which move out of the way when the H_2O is introduced (see Figure S2, Supporting Information). This plasticity of the CD loop is consistent with mutational studies showing that rates of ligand entry and escape are sensitive only to the nature of the CD loop residues adjacent to His64.[44] Furthermore, the importance of the CD region in the globin heme-binding family has been recently shown by Fernandez-Alberti et al.[45] Using a Gaussian network model analysis, the authors showed the existence of conserved low-frequency modes involving the C and D helixes and the loop connecting them. We note also that the E helix is fully displaced only when the H_2O is introduced (Val68 displacement of 0.30 Å vs 0.25 Å without H_2O), because the excursion of the CD loop away from the porphyrin center allows the E helix to collapse further toward the heme. Thus, the displacement of the CD loop locks the E helix into place.

There are quantitative differences between computed and experimental displacements, particularly toward the highly flexible chain termini (A and H helices), where end effects may be imposed by the crystal lattice; these effects are absent from the computation.

The studies reported above provide a detailed atomic-level picture of the structural and energetic factors responsible for the conformational transition in myoglobin upon ligand binding. The level of quantitative agreement with experiment that is achieved, both for the heme itself and for the protein structure, is highly satisfactory and increases confidence in the QM/MM energy models and sampling algorithms that we have developed. First, the modeling of the changes in the potential energy surface

(41) Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, G. N. J. *Chem. Rev.* **1994**, *94*, 699−714.
(42) Spiro, T. G.; Kozlowski, P. M. *Acc. Chem. Res.* **2001**, *34*, 137−144.
(43) De Angelis, F.; Jarzecki, A. A.; Car, R.; Spiro, T. G. *J. Phys. Chem. B* **2005**, *109*, 3065−3070.
(44) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177−5188.
(45) Maguid, S.; Fernandez-Alberti, S.; Ferrelli, L.; Echave, J. *Biophys. J.* **2005**, *89*, 3−13.

associated with switching of the heme electronic state from singlet to quintet as the ligand is dissociated is reasonably accurate; the difference in these surfaces is what is ultimately responsible for the conformational transition. Second, the molecular mechanics potential function and associated continuum solvation model deliver adequate performance in tracking the initial protein reorganization. Simple minimization of the structure is inadequate to manifest the conformational transition; when this was tried, the system quickly fell into local minima, the changes being absorbed locally (within ~5 Å of the active site) instead of being propagated to remote areas, as is observed experimentally. Our results show that this propagation requires iterative sampling of the side chains as well as the backbone. It is an interesting question whether a molecular dynamics approach would succeed with a similar level of computational effort. Finally, the deficiencies of the continuum solvent model are manifested in the final phase of the calculation, where a specific explicit water molecule is required to trigger shifting of the CD loop. The use of a small number of explicit water molecules in conjunction with continuum approaches is likely to be necessary in many situations in modeling important biological phenomena, if accuracy in structural and energetic predictions is to be achieved.

3.3. Implications for Hemoglobin. The Hb tetramer contains two α and two β polypeptide chains, each of which has the basic Mb architecture. In the transition from the low-affinity T structure to the high-affinity R structure, one $\alpha\beta$ dimer rotates against the other by ~15°, producing extensive changes in intersubunit contacts, especially at the $\alpha_1\beta_2$ interface.[1] The question of how the elementary act of ligand binding or dissociation induces this large-scale protein motion poses a continuing challenge. The present results on Mb reveal a mechanism for transmission of ligation-induced displacements of chain segments throughout the structure. Analogous displacements could serve to impel specific quaternary shifts in Hb.

The EF clamshell rotation, now clearly established in Mb, provides the protein transducer for Perutz' stereochemical trigger[10] involving heme deligation and the out-of-plane displacement of the Fe. In Hb, the same rotation was inferred from Raman spectroscopic evidence for the breaking of interhelical H-bonds at the initial phase of the R−T transition.[11] Comparison of the crystallographic structures of HbCO and deoxyHb via sequence-displacement maps (Figure 8) clearly reveals the same pattern of chain segment displacements as in Mb. There are three HbCO structures (revealing alternative R conformations of comparable energy), producing somewhat different displacement plots, especially for the α chains. The β chains, however, produce consistent plots, with the same shape as Mb, but with about double the displacement amplitudes. This chain also presents larger amplitudes than the α chain, most likely as a result of the presence of a longer CD loop area in the β chain. As shown by Fernandez-Alberti et al.,[45] this is the most flexible region, and its dynamics might limit overall conformational changes of the entire protein. We caution that these plots convolute changes in both tertiary and quaternary structure and obscure the fact that there are distinct intermediates along the R−T pathway.[8,9] Nevertheless, they are suggestive of the likely influence of the mechanisms traced for Mb on the quaternary reaction coordinate in Hb.

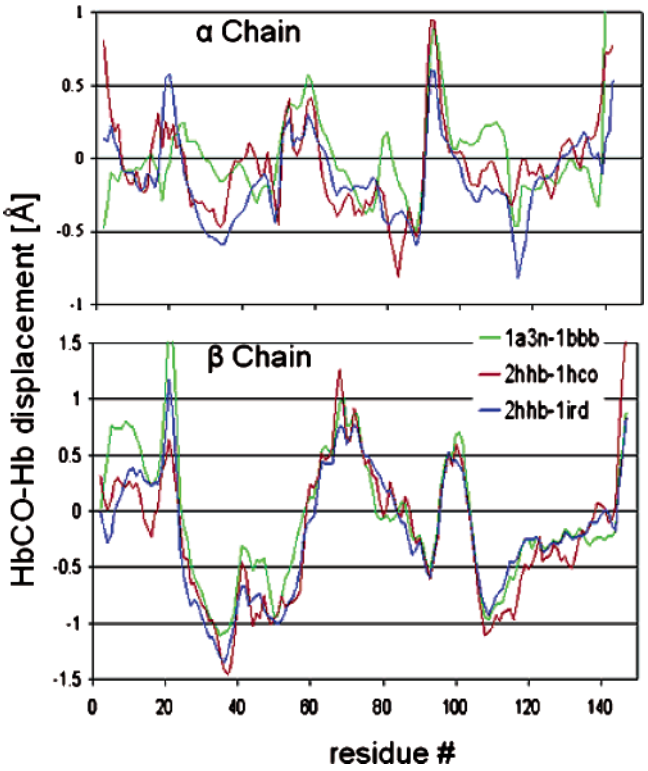


Figure 8. Sequence-displacement graph (d plot) for the α and β chains of deoxy Hb relative to HbCO, comparing three different sets of crystal structures: 1A3N−1BBB (green line), 2HHB−1HCO (red line), and 2HHB−1IRD (blue line).

For example, the AB corner displacement is prominent in the Hb plots, and the Val68—Leu69···Val117 steric pathway illustrated in Figure 7 for Mb finds correspondences in both β (Val67—Leu68···Val118) and α chains (Val62—Ala63···Val117) of Hb, aligned in the same way. The AB corners are near the $\alpha_1\beta_1$ interface and may be involved in intradimer coupling.[9] On the other hand, the CD loops and FG corners, both of which are strongly displaced in Hb, are both the locus of critical quaternary contacts at the $\alpha_1\beta_2$ interface[1] and are likely to play key roles in the later stages of the R−T transition.

4. Conclusions

The main-chain sequence-displacement graphs provide an unbiased method for evaluating helix and loop motions in the frame of the heme group as deligation of MbCO proceeds. The experimental plots make clear that the dominant effect is the EF clamshell rotation, brought about by deligation forces. They also reveal additional concerted motions, including counter-rotations adjacent to the E and F helices, producing well-defined displacements of the FG corner and of the CD loop; a further displacement of the AB corner is also seen.

These features are all present in the QM/MM-computed displacements, which show the basic pattern evolving continuously as the Fe−CO bond is stretched and the CO is removed. The computations also reveal the mechanisms for the displacements and show how deligation forces are transduced into well-defined protein motions.

These mechanisms are expected to be operative in Hb as well and may be involved in the allosteric pathway. Calculations on Hb are underway to assess this hypothesis.

Acknowledgment. This work was supported by NIH grant GM 33576 (to T.G.S.) and GM40526 (to R.A.F.). V.G. acknowledges Washington University for startup funds.

Supporting Information Available: Figure S1, illustrating the comparison between the QM/MM and crystallographic unligated structures; Figure S2, illustrating the displacements of Phe 43 and 46, and the connecting CD loop, induced in the simulation by the interaction of the His64-bound water molecule; complete ref 36. This material is available free of charge via the Internet at http://pubs.acs.org.

JA057318H