

Functional Significance of Hierarchical Tiers in Carbonmonoxy Myoglobin: Conformational Substates and Transitions Studied by Conformational Flooding Simulations

Brita G. Schulze,[†] Helmut Grubmüller,[‡] and Jeffrey D. Evanseck^{*,†,§}

Contribution from the Department of Chemistry, University of Miami, 1301 Memorial Drive, Coral Gables, Florida 33146-0431, and Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany

Received October 22, 1999. Revised Manuscript Received April 28, 2000

Abstract: The functional importance of large-scale motions and transitions of carbonmonoxy myoglobin (MbCO) conformational substates (CSs) has been studied by molecular dynamics (MD) and conformational flooding (CF) simulations. A flooding potential was constructed from an 800 ps MD trajectory of solvated MbCO to accelerate slower protein motions beyond the time scale of contemporary simulations. Two conformational transitions (tier-1 substates) resulting from seven principal molecular motions were assigned to the spectroscopic A₀ state (tier-0 substate) of MbCO, where His64 is solvated and not within the hydrophobic pocket binding site. The first computed conformational transition involves a distal pocket gate defined by the C and D helices and the interconnecting CD loop (residues 40–55). The gate-like motion is interpreted to regulate ligand access from the distal side of the heme. Simultaneously, a proximal pocket lever involving the F helix and surrounding EF and FG loops (residues 82–105) is found to shuttle the heme deep into the protein matrix (heme rmsd of 3.9 Å) as the distal pocket gate opened. The lever's effect on the heme motion is assumed to attract ligands into the heme pocket. The second major transition involves the compression and expansion of the cavity formed by the EF loop (residues 77–84) and the GH loop and H helix (residues 122–138). The motion is interpreted to modulate the hydrophobic pocket volume and regulate the ligand access from the proximal side of the heme. A third computed conformational transition was found to be a combination of the previous motions. For the first time, CF was applied in a series of room temperature simulations to accelerate molecular motions of the MbCO native fold and define the lower tier hierarchy of substate structure. The computed CSs and associated transitions coincide with previously suggested putative ligand escape pathways, and support a hierarchical description of protein dynamics and structure. A unified model that utilizes both mechanisms of distal His64 modulation (tier-0) and protein equilibrium fluctuations (tier-1) is presented to explain ligand diffusion in the MbCO dissociation reaction.

Introduction

The current understanding of protein structure, dynamics, and function has largely evolved from the systematic structural studies on myoglobin (Mb) and carbonmonoxy myoglobin (MbCO),^{1–7} and associated ligand dissociation reactions.^{8,9} MbCO is a compact globular protein composed of eight

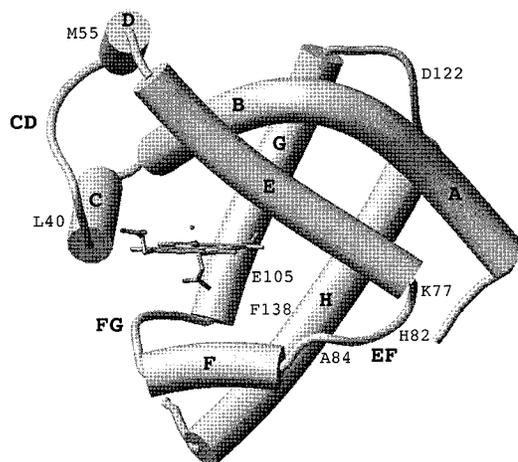


Figure 1. Schematic plot of MbCO based on the PDB structure.⁴ The helices are labeled A–H. Interconnecting loops and key residues which mark the ranges of significant structural variation are given. Figure was prepared using WebLab Viewer Light.¹⁵⁴

α-helices (labeled A–H), heme prosthetic group, and carbon monoxide ligand, as illustrated in Figure 1. The biomechanics associated with the dissociation reaction, in terms of the

[†] University of Miami.

[‡] Max-Planck-Institut für Biophysikalische Chemie.

[§] Current address: Department of Chemistry and Biochemistry, Duquesne University, 600 Forbes Ave., Pittsburgh, PA 15282.

(1) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, 1st ed.; John Wiley and Sons: New York, 1988; Vol. LXXI.

(2) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*, 1st ed.; Cambridge University Press: Cambridge, 1988.

(3) Cheng, X.; Schoenborn, B. P. *Acta Crystallogr.* **1990**, *B46*, 195.

(4) Cheng, X.; Schoenborn, B. P. *J. Mol. Biol.* **1991**, *220*, 381.

(5) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. *J. Mol. Biol.* **1986**, *220*, 381.

(6) Osapay, K.; Theriault, Y.; Wright, P. E.; Case, D. A. *J. Mol. Biol.* **1994**, *244*, 183.

(7) Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips, G. N., Jr. *J. Mol. Biol.* **1993**, *234*, 140.

(8) Sage, J. T.; Champion, P. M. *Small Substrate Recognition in Heme Proteins*; Elsevier Science Inc.: New York, 1996; Vol. 5, p 171.

(9) Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, G. N., Jr. *Chem. Rev.* **1994**, *94*, 699.

complementary structural relaxation of the protein and the diffusion of ligands between the heme and solvent, underscore the extensive range of structural and dynamic variability of the native-fold. The broad spectrum of amplitudes (0.01–100 Å), energies (0.1–100 kcal/mol), and time scales (10^{-15} – 10^3 s) associated with protein equilibrium and relaxation phenomena creates formidable scientific and technological challenges. Comprehensive reviews covering protein dynamics,^{1,2,10–14} and the specialized techniques used to probe specific time domains have been reported.^{15–21} Despite recent technological developments, only a partial and controversial assessment of the connection between protein structure, dynamics and function has been achieved.

A generalized picture of the energy landscape by Frauenfelder and co-workers has been described as a hierarchical model of structural and energetic changes.^{22–33} The strongest evidence in support of the hierarchical model comes from flash photolysis experiments and the monitoring of ligand rebinding over wide ranges of temperature and time.^{34–64} From the experiments, it

has been determined that Mb exists in a continuum of nearly isoenergetic local minima with slightly different conformations, which are called the conformational substates (CSs).^{22–33} In the case of MbCO, the top tier (tier-0) has three taxonomic substates bound by large intervening activation barriers.^{65–67} The three tier-0 taxonomic substates were first identified by discrete ligand IR absorption bands centered at 1965, 1949–1942, and 1932 cm^{-1} and labeled as the spectroscopic A states (A_{0-3}).^{68–73} Each A state has characteristic energetics, kinetics, and molecular properties, and their relative populations have been shown to depend on external conditions,²⁹ such as temperature,^{69–73} pressure,^{74–77} pH,^{68,78} and solvent viscosity.^{79–83} The structural and electrostatic determinants that differentiate between the three tier-0 taxonomic A states have been studied by pH and site-

- (10) Kurzynski, M. *Prog. Biophys. Mol. Biol.* **1998**, *69*, 23.
- (11) Frauenfelder, H. In *Physics of Biological Systems from Molecules to Species*; Flyvbjerg, H., Hertz, J., Jensen, M. H., Mouritsen, O. G., Sneppen, K., Eds.; Springer: New York, 1997; p 29.
- (12) Jardetzky, O. *Prog. Biophys. Mol. Biol.* **1996**, *65*, 171.
- (13) Nienhaus, G. U.; Young, R. D. *Encycl. Appl. Phys.* **1996**, *15*, 163.
- (14) Parak, F.; Frauenfelder, H. *Physica A* **1993**, *201*, 332.
- (15) Hochstrasser, R. M. *J. Chem. Educ.* **1998**, *75*, 559.
- (16) Réat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterheld, D.; Zaccari, G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4970.
- (17) Chen, E.; Goldbeck, R. A.; Klinger, D. S. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 327.
- (18) Schlick, T.; Barth, E.; Mandziuk, M. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 181.
- (19) Ben-Nun, M.; Cao, J.; Wilson, K. R. *J. Phys. Chem. A* **1997**, *101*, 8743.
- (20) Srajer, V.; Teng, T.-Y.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S.-I.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. *Science* **1996**, *274*, 1726.
- (21) Doster, W.; Cusack, S.; Petry, W. *Nature* **1989**, *337*, 754.
- (22) Frauenfelder, H.; Wolynes, P. G.; Austin, R., H. *Rev. Mod. Phys.* **1999**, *71*, S419.
- (23) Nienhaus, G. U.; Müller, J. D.; McMahon, B. H.; Frauenfelder, H. *Physica D* **1997**, *107*, 297.
- (24) Thorn Leeson, D.; Wiersma, D. A.; Fritsch, K.; Friedrich, J. J. *Phys. Chem. B* **1997**, *101*, 6331.
- (25) Frauenfelder, H. *Nat. Struct. Biol.* **1995**, *2*, 821.
- (26) Frauenfelder, H.; Wolynes, P. G. *Phys. Today* **1994**, 58.
- (27) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598.
- (28) Young, R. D.; Frauenfelder, H.; Johnson, J. B.; Lamb, D. C.; Nienhaus, G. U.; Philipp, R.; Scholl, R. *Chem. Phys.* **1991**, *158*, 315.
- (29) Hong, M. K.; Braunstein, D.; Cowen, B. R.; Frauenfelder, H.; Iben, I. E. T.; Mourant, J. R.; Ormos, P.; Scholl, R.; Schulte, A.; Steinbach, P. J.; Xie, A.-H.; Young, R. D. *Biophys. J.* **1990**, *58*, 429.
- (30) Frauenfelder, H. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 451.
- (31) Ansari, A.; Berendsen, J.; Braunstein, D.; Cowen, B. R.; Frauenfelder, H.; Hong, M. K.; Iben, I. E. T.; Johnson, J. B.; Ormos, P.; Sauke, T. B.; Scholl, R.; Schulte, A.; Steinbach, P. J.; Vittitow, J.; Young, R. D. *Biophys. Chem.* **1987**, *26*, 337.
- (32) Frauenfelder, H.; Wolynes, P. G. *Science* **1985**, *229*, 337.
- (33) Ansari, A.; Berendsen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5000.
- (34) Mizutani, Y.; Kitagawa, T. *Science* **1997**, *278*, 443.
- (35) Lim, M.; Jackson, T. A.; Anfinrud, P. A. *Nat. Struct. Biol.* **1997**, *4*, 209.
- (36) Olson, J. S.; Phillips, G. N., Jr. *J. Biol. Chem.* **1996**, *271*, 17593.
- (37) Shibata, Y.; Kutita, A.; Kushida, T. *J. Chem. Phys.* **1996**, *104*, 4396.
- (38) Agmon, N.; Sastry, M. *Chem. Phys.* **1996**, *212*, 207.
- (39) Lim, M.; Jackson, T. A.; Anfinrud, P. A. *Science* **1995**, *269*, 962.
- (40) Lim, M.; Jackson, T. A.; Anfinrud, P. A. *J. Chem. Phys.* **1995**, *102*, 4355.
- (41) Ansari, A.; Jones, C., M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Biochemistry* **1994**, *33*, 5128.
- (42) Walda, K., N.; Liu, X. Y.; Sharma, V., S.; Magde, D. *Biochemistry* **1994**, *33*, 2198.
- (43) Jackson, T. A.; Lim, M.; Anfinrud, P. A. *Chem. Phys.* **1994**, *180*, 131.
- (44) Post, F.; Doster, W.; Karvounis, G.; Settles, M. *Biophys. J.* **1993**, *64*, 1833.
- (45) Mourant, J. R.; Braunstein, D.; Chu, K.; Frauenfelder, H.; Nienhaus, G. U.; Ormos, P.; Young, R., D. *Biophys. J.* **1993**, *65*, 1496.
- (46) Tian, W. D.; Sage, J. T.; Srajer, V.; Champion, P. M. *Phys. Rev. Lett.* **1992**, *68*, 408.
- (47) Steinbach, P. J.; Chu, K.; Frauenfelder, H.; Jonson, J. B.; Lamb, D. C.; Nienhaus, G. U.; Sauke, T. B.; Young, R., D. *Biophys. J.* **1992**, *61*, 235.
- (48) Steinbach, P. J.; Ansari, A.; Berendzen, J.; Braunstein, D.; Chu, K.; Cowen, B. R.; Ehrenstein, D.; Frauenfelder, H.; Johnson, J. B.; Lamb, D. C.; Luck, S.; Mourant, J. R.; Nienhaus, G. U.; Ormos, P.; Philipp, R.; Xie, A.; Young, R. D. *Biochemistry* **1991**, *30*, 3988.
- (49) Genberg, L.; Richard, L.; McLendon, G.; Miller, R. J. D. *Science* **1991**, *251*, 1051.
- (50) Srajer, V.; Champion, P. M. *Biochemistry* **1991**, *30*, 7390.
- (51) Ahmed, A. M.; Campbell, B. F.; Caruso, D.; Chance, M. R.; Chavez, M. D.; Courtney, S. H.; Friedman, J. M.; Iben, I. E. T.; Ondrias, M. R.; Yang, M. *Chem. Phys.* **1991**, *158*, 329.
- (52) Lambright, D. G.; Balasubramanian, S.; Boxer, S. G. *Chem. Phys.* **1991**, *158*, 249.
- (53) Chance, M. R.; Courtney, S. H.; Chavez, M. D.; Ondrias, M. R.; Friedman, J., M. *Biochemistry* **1990**, *29*, 5537.
- (54) Petrich, J. W.; Poyart, C.; Martin, J. L. *Biochemistry* **1988**, *27*, 4049.
- (55) Jongeward, K. A.; Magde, D.; Taube, D. J.; Marsters, J., C.; Traylor, T. G.; Sharma, V., S. *J. Am. Chem. Soc.* **1988**, *110*, 380.
- (56) Gibson, Q. H.; Olson, J. S.; McKinnie, R. E.; Rohlfis, R. J. *J. Biol. Chem.* **1986**, *261*, 10228.
- (57) Young, R. D.; Bowne, S. F. *J. Chem. Phys.* **1984**, *81*, 3730.
- (58) Henry, E. R.; Sommer, J. H.; Hofrichter, J.; Eaton, W. A. *J. Mol. Biol.* **1983**, *166*, 443.
- (59) Agmon, N.; Hopfield, J. J. *J. Chem. Phys.* **1983**, *79*, 2042.
- (60) Doster, W.; Beece, D.; Bowne, S. F.; DiIorio, E. E.; Eisenstein, L.; Frauenfelder, H.; Reinisch, L.; Shyamsunder, E.; Winterhalter, K. H.; Yue, K. T. *Biochemistry* **1982**, *21*, 4831.
- (61) Cornelius, P. A.; Steele, A. W.; Chernoff, D. A.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 7526.
- (62) Alberding, N.; Austin, R. H.; Chan, S. S.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C.; Nordlund, T. M. *Chem. Phys.* **1976**, *65*, 4701.
- (63) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355.
- (64) Austin, R. H.; Beeson, K.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C.; Marshall, V. P. *Science* **1973**, *18*, 541.
- (65) Johnson, J. B.; Lamb, D. C.; Frauenfelder, H.; Müller, J. D.; McMahon, B.; Nienhaus, G. U.; Young, R., D. *Biophys. J.* **1996**, *71*, 1563.
- (66) Tian, W. D.; Sage, J. T.; Champion, P. M.; Chien, E.; Sligar, S. G. *Biochemistry* **1996**, *35*, 3487.
- (67) Huang, J.; Ridsdale, A.; Wang, J.; Friedman, J. M. *Biochemistry* **1997**, *36*, 14353.
- (68) Shimada, H.; Caughey, W. S. *J. Biol. Chem.* **1982**, *257*, 11893.
- (69) Alben, J. O.; Beece, D.; Bowne, S. F.; Doster, W.; Eisenstein, L.; Frauenfelder, H.; Good, D.; McDonald, J. D.; Marden, M. C.; Moh, P. P.; Reinisch, L.; Reynolds, A. H.; Shyamsunder, E.; Yue, K. T. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3744.
- (70) Caughey, W. S.; Shimada, H.; Choc, M. G.; Tucker, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2903.
- (71) Makinen, M. W.; Houtchens, R. A.; Caughey, W. S. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 6042.
- (72) Caughey, W., S.; Alben, J. O.; McCoy, S.; Boyer, S. H.; Charache, S.; Hathaway, P. *Biochemistry* **1969**, *8*, 59.
- (73) Alben, J. O. *Biochemistry* **1968**, *7*, 175.
- (74) Galkin, O.; Buchter, S.; Tabirian, A.; Schulte, A. *Biophys. J.* **1997**, *73*, 2752.

directed mutagenesis,^{84–88} X-ray crystallography,⁸⁹ single-crystal Raman spectroscopy,⁹⁰ MD simulations,^{91–94} and time-resolved infrared spectroscopy.^{95,96} The experiments indicate that the geometric positioning of His64 with respect to the hydrophobic pocket determines the A₀ (open) and A_{1–3} (closed) spectroscopic states. In addition, a hydrogen bond between His64 and the ligand is thought to induce the A_{1,2} state,^{97–99} and the A₃ state is postulated to result from an additional and cooperative electrostatic interaction with Arg45.⁹¹

There has been less success in defining the hierarchical model of the lower tiers associated with ligand diffusion in the MbCO dissociation reaction.⁷⁵ Two proposed mechanisms for ligand diffusion involve either distal His64 control by moving in and out of the heme pocket,^{86,89,99–102} or structural equilibrium fluctuations which generate longer hydrophobic pathways involving both the distal and proximal sides of the heme.^{100,103–108} The first mechanism is related to the A states, since ligand binding rates are increased when His64 is forced out of the heme

pocket consistent with the A₀→A_{1–3} transition.^{86,89,99} The second mechanism is supported by crystallography with a myoglobin–xenon complex,¹⁰³ ligand binding studies with various mutants,^{104,109,110} MD simulations,^{100,105–108,111} and ultrafast spectroscopy of the photodissociated CO in Mb.³⁵ The main problem lies in assigning the correct motions of the fluctuating protein matrix to the lower tiers of the hierarchical model. Within the tier-1 substates, the protein can assume more than 10⁶ substates separated by small barriers, as compared to the taxonomic states.^{11,25,31} It has been previously suggested that protein conformations based upon heme iron displacements in the unbound state may be important in defining the tier-1 substates.^{112,113} It has also been reported that localized distal pocket configurations, rather than changes in the heme and proximal side, define the tier-1 CSs.⁶⁶ Despite considerable work, the tier-1 CSs structures have not been unambiguously defined, and their functional significance is far from being understood. As a consequence, the description of ligand diffusion through Mb remains as a controversial issue.

Computational methods have indirectly determined the ligand diffusion pathways through the protein.^{100,103–108} The indirect nature of this approach lies in the time scale of the dynamics, where ligand motion is significantly faster than the important lower frequency motions of the protein. During MD simulations, ligand dynamics are relatively equilibrated, whereas protein dynamics are far from equilibrium. With current computing powers, it is impossible to carry out the long-time MD simulations necessary to equilibrate protein dynamics.^{114,115} This limitation has prompted the development of several different MD techniques to improve the sampling of conformational space, such as simulated annealing,^{116,117} simulation of multiple short-time trajectories,^{91,108,118,119} locally enhanced sampling,¹⁰⁶ and conformational flooding (CF).¹²⁰ Alternatively, information about large-scale fluctuations in a protein can be obtained by normal-mode analysis,^{121–123} quasiharmonic analysis,^{124–128} or essential dynamics.¹²⁹

(75) Frauenfelder, H.; Alberding, N. A.; Ansari, A.; Braunstein, D.; Cowen, B. R.; Hong, M. K.; Iben, I. E. T.; Johnson, J. B.; Luck, S.; Marden, M. C.; Mourant, J. R.; Ormos, P.; Reinisch, L.; Scholl, R.; Schulte, A.; Shyamsunder, E.; Sorensen, L. B.; Steinbach, P. J.; Xie, A.; Young, R. D.; Kwok, T. Y. *J. Phys. Chem.* **1990**, *94*, 1024.

(76) Gekko, K.; Hasegawa, Y. *Biochemistry* **1986**, *25*, 6563.

(77) Iben, I. E. T.; Braunstein, D.; Doster, W.; Frauenfelder, H.; Hong, M. K.; Johnson, J. B.; Luck, S.; Ormos, P.; Schulte, A.; Steinbach, P. J.; Xie, A. H.; Young, R. D. *Phys. Rev. Lett.* **1989**, *62*, 1916.

(78) Uno, T.; Nishimura, Y.; Tsuboi, M.; Makino, R.; Iizuka, T.; Ishimura, Y. *J. Biol. Chem.* **1997**, *262*, 4549.

(79) Shibata, Y.; Kurita, A.; Kushida, T. *Biochemistry* **1999**, *38*, 1789.

(80) Kleinert, T.; Doster, W.; Leyser, H.; Petry, W.; Schwarz, V.; Settles, M. *Biochemistry* **1998**, *37*, 717.

(81) Sastry, G. M.; Agmon, N. *Biochemistry* **1997**, *36*, 7097.

(82) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Science* **1992**, *256*, 1796.

(83) Beece, D.; Eisenstein, L.; Frauenfelder, H.; Good, D.; Marden, M. C.; Reinisch, L.; Reynolds, A. H.; Sorensen, L. B.; Yue, K. T. *Biochemistry* **1980**, *19*, 5147.

(84) Morikis, D.; Champion, P. M.; Springer, B. A.; Sligar, S. G. *Biochemistry* **1989**, *28*, 4791.

(85) Li, T.; Quillin, M. L.; Phillips, G. N.; Olson, J. S. *Biochemistry* **1994**, *33*, 1433.

(86) Tian, W. D.; Sage, J. T.; Champion, P. M. *J. Mol. Biol.* **1993**, *233*, 155.

(87) Ray, G. B.; Li, X.-Y.; Ibers, J. A.; Sessler, J. L.; Spiro, T. G. *J. Am. Chem. Soc.* **1994**, *116*, 162.

(88) Braunstein, D. P.; Chu, K.; Egeberg, K. D.; Frauenfelder, H.; Mourant, J. R.; Nienhaus, G. U.; Ormos, P.; Sligar, S. G.; Springer, B. A.; Young, R. D. *Biophys. J.* **1993**, *65*, 2447.

(89) Yang, F.; Phillips, G. N., Jr. *J. Mol. Biol.* **1996**, *256*, 762.

(90) Zhu, L.; Sage, J. T.; Rigos, A. A.; Morikis, D.; Champion, P. M. *J. Mol. Biol.* **1992**, *224*, 207.

(91) Schulze, B. G.; Evanseck, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 6444.

(92) Jewsbury, P.; Kitagawa, T. *Biophys. J.* **1994**, *67*, 2236.

(93) Spiro, T.; Kozlowski, P. M. *J. Am. Chem. Soc.* **1998**, *120*, 4524.

(94) Kuczera, K.; Kuriyan, J.; Karplus, M. *J. Mol. Biol.* **1990**, *213*, 351.

(95) Moore, J. N.; Hansen, P. A.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5062.

(96) Ormos, P.; Braunstein, D.; Frauenfelder, H.; Hong, M. K.; Lin, S.-L.; Sauke, T. B.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8492.

(97) Unno, M.; Christian, J. F.; Olson, J. S.; Sage, J. T.; Champion, P. M. *J. Am. Chem. Soc.* **1998**, *120*, 2670.

(98) Sigdriss, E.; Ryde, U. *J. Mol. Inorg. Chem.* **1999**, *4*, 99.

(99) Lai, H. H.; Li, T.; Lyons, D.; Phillips, G. N., Jr.; Gibson, Q. H. *Protein Struct., Funct., Genet.* **1995**, *22*.

(100) Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979**, *132*, 343.

(101) Perutz, M. J. *Mol. Biol.* **1965**, *13*, 646.

(102) Perutz, M. F.; Mathews, F. A. *J. Mol. Biol.* **1966**, *21*, 199.

(103) Tilton, R. F., Jr.; Kuntz, I. D., Jr.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849.

(104) Huang, X.; Boxer, S. G. *Nat. Struct. Biol.* **1994**, *1*, 226.

(105) Tilton, R. F., Jr.; Singh, U. C.; Kuntz, I. D., Jr.; Kollman, P. A. *J. Mol. Biol.* **1988**, *199*.

(106) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161.

(107) Sassaroli, M.; Rousseau, D. L. *J. Biol. Chem.* **1986**, *261*, 16292.

(108) Carlson, M. L.; Regan, R. M.; Gibson, Q. H. *Biochemistry* **1996**, *35*, 1125.

(109) Uchida, T.; Unno, M.; Ishimoro, J.; Morishima, I. *Biochemistry* **1997**, *36*, 324.

(110) Scott, E. E.; Gibson, Q. H. *Biochemistry* **1997**, *36*, 11909.

(111) Tilton, R. F., Jr.; Singh, U. C.; Weiner, S. J.; Connolly, M. L.; Kuntz, I. D., Jr.; Kollman, P. A.; Max, N.; Case, D. A. *J. Mol. Biol.* **1986**, *192*, 443.

(112) Srajer, V.; Schomacher, K. T.; Champion, P. M. *Phys. Rev. Lett.* **1986**, *57*, 1267.

(113) Srajer, V.; Reinisch, L.; Champion, P. M. *J. Am. Chem. Soc.* **1988**, *110*, 6656.

(114) Grubmüller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997.

(115) Heymann, B.; Grubmüller, H. *Chem. Phys. Lett.* **1999**, *1*.

(116) Kirpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.

(117) Brünger, A. T. In *Crystallographic Refinement by Simulated Annealing*; Isaacs, N. W., Taylor, M. R., Eds.; Clarendon Press: Oxford, 1988.

(118) Elofsson, A.; Nilsson, L. *J. Mol. Biol.* **1993**, *233*, 766.

(119) Caves, L. S. D.; Evanseck, J. D.; Karplus, M. *Protein Sci.* **1998**, *7*, 649.

(120) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893.

(121) Levitt, M.; Sander, C.; Stern, P. S. *Int. J. Quantum Chem.* **1983**, *10*, 181.

(122) Gō, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696.

(123) Brooks, B.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(124) Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 809.

(125) Hayward, S.; Kitao, A.; Hirata, F.; Go, N. *J. Mol. Biol.* **1993**, *234*, 1207.

(126) Teeter, M. M.; Case, D. A. *J. Phys. Chem.* **1990**, *94*, 8091.

(127) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696.

(128) Kitao, A.; Hirata, F.; Go, N. *Chem. Phys.* **1991**, *158*, 447.

(129) Amadei, A.; Linssen, F.; Go, N. *Chem. Phys.* **1991**, *158*, 447.

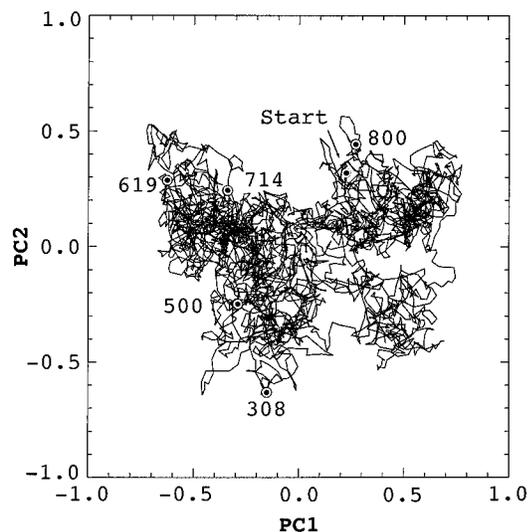


Figure 2. Projection of the 800 ps MD trajectory of MbCO onto the two largest eigenvectors from PCA. The indicated times correspond to the starting points (308, 500, 619, 714, and 800 ps) for the five flooding simulations. Units for the PC1 and PC2 axes are in Å.

Due to the intrinsic importance of forging a general connection between protein structure, dynamics, and function, it is both timely and appropriate to report on the computed hierarchical structure of conformational substates and corresponding transitions that define a portion of the MbCO energy landscape. For the first time, conformational flooding is used in a series of room-temperature simulations to accelerate molecular motions of the MbCO native-fold, define the controversial lower tier hierarchy of substate structure, and explain ligand diffusion in the dissociation reaction of MbCO.

Methods

MD Simulations. An 800 ps MD simulation was started from the neutron diffraction structure (PDB entry 2MB5) at 1.8 Å resolution.⁴ The protein and 89 crystal water molecules were embedded within a 37 Å sphere of 6198 equilibrated TIP3P water molecules¹³⁰ using the SOLVATE_1.0 program.¹³¹ MD simulations were carried out with the EGO_VIII software, which employs a combination of structure-adapted multipole method¹³² and multiple time step algorithm¹³³ to calculate efficiently long-range electrostatic forces to avoid the use of nonbond cutoffs. The all-atom CHARMM force field was used.¹³⁴ The surface region of the water sphere was subjected to an SBOUND potential¹³⁵ to confine water molecules within the sphere and to counterbalance surface tension. The system was relaxed in several steps with a gradient minimizer and then equilibrated for 30 ps at 300 K. During the equilibration phase, the system was coupled to a heat bath with a time constant of 10^{-13} s. For the remaining MD simulation (770 ps), a weaker heat bath coupling with a time constant of 10^{-11} s was employed.

(130) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(131) Grubmüller, H. *Solvate: A Program to Create Atomic Solvent Models*; 1.0 ed., 1996; Electronic access: www.mpibpc.gwdg.de/abteilungen/071/solvate.html.

(132) Eichinger, M.; Grubmüller, H.; Heller, H.; Tavan, P. *J. Comput. Chem.* **1997**, *18*, 1729.

(133) Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. *Mol. Simul.* **1991**, *6*, 121.

(134) MacKerell, A.; Bashford, D.; Bellott, D.; Dunbrack, R.; Evanseck, J. D.; Field, M.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586. Note added in proof: The Urey–Bradley (NC2 C NC2) term for arginine had the wrong sign. This has been determined to have an insignificant effect on the computed results.

(135) Brünger, A.; Brooks, C. L.; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495.

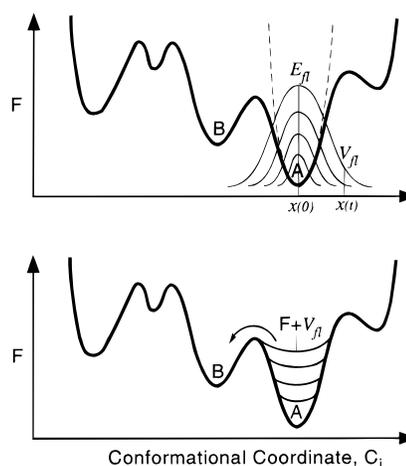


Figure 3. The method of conformational flooding. Shown are plots of the free energy landscape F as a function of high-dimensional conformational coordinates c_i . See text for detailed explanation.

Principal Component Analysis (PCA). Principal component analysis (PCA)^{136–138} was used in this work as a basis for the CF simulations and as a tool to analyze and visualize the MD trajectories in conformational space. The basic principles and computational steps relevant for its application have been described recently.⁹¹ In the present 800 ps trajectory, the first three PCs describe 41.9% (24.9, 9.0 and 8.0%, respectively) of the total structural variation. In Figure 2, the PCs from the largest eigenvector ψ_1 are plotted on the x -axis and those from the second largest eigenvector ψ_2 are plotted on the y -axis. To enable comparison with rmsd values, all PCs have been divided by the square root of the number of atoms considered. In this particular case, the number of α -carbons or the square root of 153 is used. The PCs of smaller eigenvalues show diffusely scattered data points, which indicates nonspecific atomic fluctuations.

Conformational Flooding (CF) Theory. A rigorous derivation of the statistical mechanics on which CF is based has been presented before,¹²⁰ and only a brief description, as outlined in Figure 3, is given here. The upper part of the figure sketches a cross-section of the high-dimensional multi-minima free energy landscape of a protein as a function of a number of conformational degrees of freedom c_i . Each of the minima in that landscape defines a CS. We assume that the barriers separating an initial substate from its neighboring substates are so large that they are not overcome with conventional MD simulations, whereas CF allows the investigation of transitions at longer time scales. The transition from CS A to B is illustrated in the figure. As a first step, a conventional MD simulation (typically 500–1000 ps) is carried out, yielding an ensemble $\{\bar{x}\}$ of structures within the initial CS. From the covariance matrix C of that ensemble

$$C = \langle (\bar{x} - \langle \bar{x} \rangle)(\bar{x} - \langle \bar{x} \rangle)^T \rangle \quad (1)$$

(where $\langle \dots \rangle$ denotes an ensemble average), an approximation of the configuration space density within the initial substate is derived in terms of a multivariate Gaussian function

$$\rho(\bar{x}) = Z^{-1} \exp \left[-\frac{1}{2} (\bar{x} - \langle \bar{x} \rangle)^T \mathbf{A} (\bar{x} - \langle \bar{x} \rangle) \right] \quad (2)$$

where $\mathbf{A}^{-1} = C$, and Z is the configurational partition function

$$Z = \int d^{3N} \bar{x} \exp \left[-\frac{1}{2} (\bar{x} - \langle \bar{x} \rangle)^T \mathbf{A} (\bar{x} - \langle \bar{x} \rangle) \right] \quad (3)$$

The vector \bar{x} contains the $3N$ Cartesian coordinates of the molecule. From the Gaussian approximation of the configuration space density,

(136) Pearson, K. *Philos. Mag.* **1901**, *2*, 557.

(137) Hotelling, H. *J. Educ. Psychol.* **1933**, *24*, 417.

(138) Jolliffe, I. T. *Principal Component Analysis*; Springer-Verlag: New York, 1986.

the quasiharmonic free energy approximation

$$F(\vec{x}) = -k_B T \ln \rho(\vec{x}) \quad (4)$$

is obtained. Here, k_B is the Boltzmann constant, and T is the temperature. The free energy model (dashed line in upper part of Figure 3) is valid in the vicinity of the initial CS and is multi-harmonic. On the basis of eqs 2–4, a flooding potential is constructed to “fill up” the initial free energy minimum (Figure 3, bottom, thin solid lines in the A substate)

$$V_{\bar{n}}(\vec{x}) = E_{\bar{n}} \exp\left[-\frac{1}{2} k_B T (\vec{x} - \langle \vec{x} \rangle)^T \mathbf{A} (\vec{x} - \langle \vec{x} \rangle) / E_{\bar{n}}\right] \quad (5)$$

Thus, the free energy barrier “seen” by the protein is lowered, and the conformational transitions of interest are accelerated dramatically, for example, from microseconds to picoseconds. In eq 5, $E_{\bar{n}}$ denotes the flooding strength, which regulates the acceleration.

On the basis of the assumption that conformational transitions can indeed be described by relatively few conformational coordinates,^{124,127,129} the flooding potential is modified to act only on those p conformational coordinates which exhibit the largest fluctuations, as characterized by their PCA eigenvalues. Since the proper value of the flooding strength is not known a priori, flooding potentials of varying magnitude have to be tested. In this work, we consider the flooding potential, $V_{\bar{n}}$, to be of suitable strength if it triggers conformational transitions at a time scale of 100 ps or less, yet is weak enough to avoid artificial deformation of the structure. This is symbolically depicted with the four curves for flooding potentials in Figure 3.

Flooding Simulations. The atomic coordinates of all α -carbon atoms from 300 to 800 ps of the MD trajectory were used as the configuration vectors from which the flooding matrix was calculated according to eqs 1–5. Thus, the flooding potential (eq 5) was based on the specific structural and dynamic properties of the original MD trajectory between 300 and 800 ps. When applied with the appropriate strength in the flooding simulation, it was expected to drive the structure from the original conformation into a new CS previously unexplored by the MD simulations.

Two observables, $D_{\bar{n}}$ and $V_{\bar{n}}$, were monitored during the flooding simulations to identify conformational transitions. Figure 3 (upper part) illustrates that the flooding energy $V_{\bar{n}}(\vec{x})$ decreases as the structure $\vec{x}(t)$ moves away from its original position $\vec{x}(0)$. An additional criterion for a conformational transition was the flooding distance $D_{\bar{n}}(\vec{x})$, defined as

$$D_{\bar{n}}(\vec{x}) = \sum_{n=1}^{153} [\bar{x}_n(0) - \bar{x}_n(t)]^2 \quad (6)$$

The sum is over the α -carbon atom coordinates, $\bar{x}_n(0)$ is the n th coordinate of the starting conformation, and $\bar{x}_n(t)$ is the n th coordinate at time t . During the flooding simulation, $V_{\bar{n}}$ typically decreased smoothly as the flooding potential became less effective, and $D_{\bar{n}}$ increased steadily since the structure moved away from its original conformation. A steep drop in $V_{\bar{n}}$ and a concurrent sudden rise in $D_{\bar{n}}$ was assumed to indicate the crossing of an energy barrier, followed by the relaxation into a new CS.

Subsequent MD Simulations. To check if the generated structures were stable, additional MD simulations of 300–400 ps were carried out after each CF run. They are called MD2 to distinguish them from the original 800 ps trajectory. Snapshots before flooding ($t = t_0$), after flooding ($t = t_n$, typically 100–150 ps after t_0) and after MD2 ($t = t_{MD2}$, 300–400 ps after t_n) were compared by their root-mean-square deviation (rmsd). If $[\text{rmsd}(t_{MD2}) - \text{rmsd}(t_n)] < [\text{rmsd}(t_{MD2}) - \text{rmsd}(t_0)]$, the new conformation was assumed to be stable, since the structure remained in the new conformation after MD2 (without the flooding potential). If $[\text{rmsd}(t_{MD2}) - \text{rmsd}(t_n)] > [\text{rmsd}(t_{MD2}) - \text{rmsd}(t_0)]$, the structure was evaluated to check if it had moved back to its original conformation. If so, then the new conformation was rejected as unstable.

Analysis. By using a probe radius of 1.6 Å the solvent-accessible surface area was computed with the algorithm by Lee and Richards as implemented in the CHARMM program.¹²³ Structures were averaged over a time period of 10 ps to calculate the solvent-accessible surface

Table 1. Description of the Flooding and Subsequent MD2 Trajectories

CF trajectory	$E_{\bar{n}}$ (kcal/mol)	affected residues	secondary structure elements	radius of gyration	rmsd in Å before CF and after MD2
308–436 ps (CF 1)	54	40–55 82–88	C, CD, D EF, F	15.68 (308 ps) 16.19 (436 ps)	2.30 2.05
436–840 ps (MD2)		95–112 heme, CO	FG, G	15.95 (840 ps)	1.95 3.87
500–600 ps (CF 2)	60	77–84 122–138	EF GH, H	15.67 (500 ps) 15.86 (600 ps)	3.03 4.11
600–900 ps (MD2)		43–63 149–153	CD, D, E C-terminus	15.74 (900 ps)	2.25 3.40
619–750 ps (CF 3)	54–66				
714–836 ps (CF 4)	66	40–47 75–84	C, CD EF	15.54 (714 ps) 15.68 (836 ps)	3.07 4.63
836–1200 ps (MD2)		135–140 152–153	H C-terminus	15.76 (1200 ps)	1.89
800–950 ps (CF 5)	54–66				

area. The program MOLMOL¹³⁹ was used to determine the secondary structure elements in snapshots from flooding and MD. The α -helices were identified on the basis of 1,4-backbone hydrogen bond formation.

Results and Discussion

The starting structures for the flooding simulations were chosen as the structures at 308, 500, 619, 714, and 800 ps from the original MD trajectory (Figure 2). The entire 800 ps trajectory sampled the A_0 taxonomic state, since His64 was always removed from the hydrophobic ligand binding site. This is in agreement with previously reported computational^{91,140,141} and experimental^{84–90,95,96,142} studies in the literature. To determine the appropriate flooding potentials, which promote conformational transitions belonging to the A_0 hierarchy, a series of CF simulations at room temperature with $E_{\bar{n}}$ values between 54 and 66 kcal/mol were carried out from each starting structure (refer to Table 1).

Flooding Trajectory 1 (CF 1). The first flooding trajectory was started from the MD transient structure at 308 ps. Test simulations with various values for $E_{\bar{n}}$ showed that a flooding strength of 54 kcal/mol was necessary to induce a conformational transition, which was then observed to occur at a flooding energy of approximately 35–40 kcal/mol as reflected in the changes of $V_{\bar{n}}$ and $D_{\bar{n}}$ (see Figure 4). From 317 to 323 ps, the flooding energy, $V_{\bar{n}}$, dropped from 40 to 20 kcal/mol (Figure 4a), while the flooding distance, $D_{\bar{n}}$, increased by approximately 170 Å² (Figure 4b). The sudden changes in $V_{\bar{n}}$ and $D_{\bar{n}}$ can be qualitatively understood by examining the energy surface (thick line) and the flooding potential (thin line), as shown in Figure 4c. At the beginning of the flooding simulation ($t = 308$ ps), the flooding potential raised the energy of the conformation, as indicated by the dashed line. The system was destabilized sufficiently, such that at 317 ps it underwent the necessary conformational changes to overcome the original barrier (curved arrow) and move toward a new CS. Since the protein moved abruptly away from the region in conformational space where $E_{\bar{n}}$ was effective, a steep drop in $V_{\bar{n}}$ and concurrent increase in $D_{\bar{n}}$ until 323 ps is indicative of an energy barrier crossing followed by relaxation into a new CS.

(139) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(140) Kushkuley, B.; Stavrov, S. S. *Biophys. J.* **1996**, *70*, 1214.

(141) Kushkuley, B.; Stavrov, S. S. *Biophys. J.* **1997**, *72*, 899.

(142) Oldfield, E.; Guo, K.; Augspurger, J. D.; Dykstra, C. E. *J. Am. Chem. Soc.* **1991**, *113*, 7537.

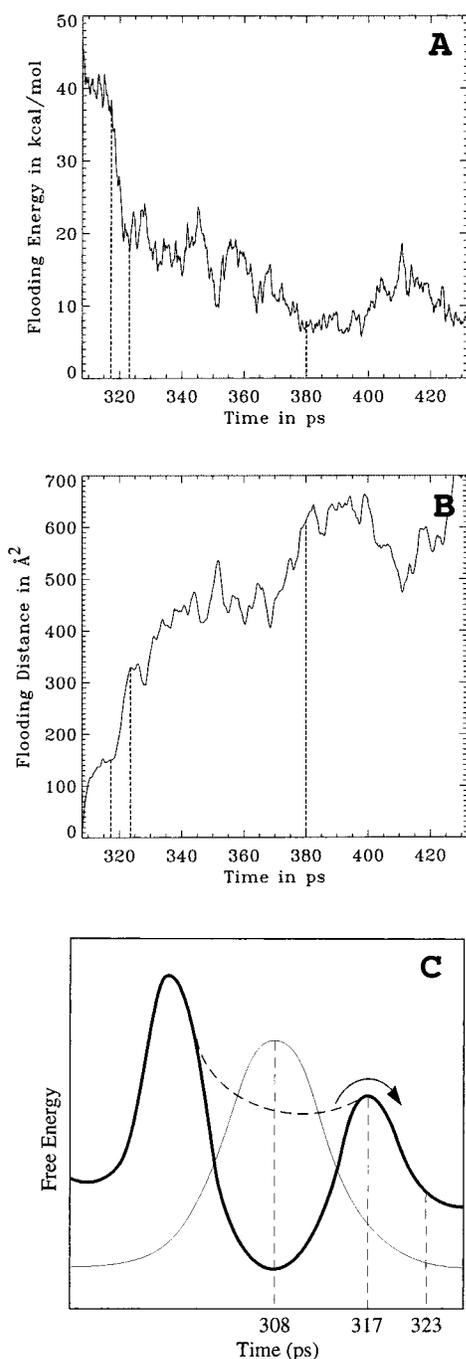


Figure 4. Development of flooding simulation CF 1. (a) Flooding energy V_f . (b) Flooding distance D_f . (c) Schematic sketch of the energy surface indicating the start of floodings (308 ps), barrier crossing (317 ps), and arrival in a new substate (323 ps).

The flooding simulation was stopped at 436 ps. In the subsequent MD2 simulation (436–840 ps), the protein could either stabilize in the new conformation, move back toward the original structure, or convert into another CS. After the start of MD2, the structure underwent a number of conformational changes, as visualized in Figure 5. This plot was constructed by applying PCA to the Cartesian coordinates of the structures from the original MD trajectory (0–308 ps), CF 1 (308–436 ps), and MD2 trajectory (436–840 ps). The conformational changes were mainly between 308 and 500 ps. At 500 ps, the conformation entered an energetically favorable conformational region and did not move back to the original substate. The figure demonstrates that a larger amount of conformational space was

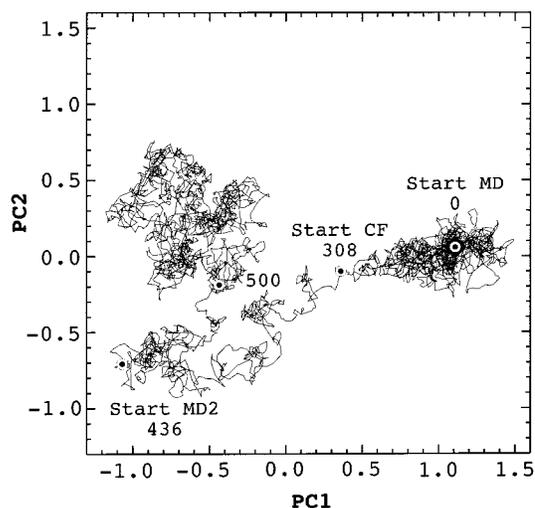


Figure 5. Projected trajectory for CF 1: original MD (0 to 308 ps), CF 1 (308 to 436 ps), and subsequent MD2 simulation (436 to 840 ps). Units for the PC1 and PC2 axes are in Å.

sampled with CF as compared to MD, and the protein assumed a new and stable conformation significantly different from the original CS. Note that the trajectories in Figures 2 and 5 appear different between 0 and 308 ps, even though the same spatial coordinates are used. The difference originates since flooding trajectory coordinates are also included in the matrix **A** (see eqs 1–3) from which the eigenvectors and eigenvalues are computed. Since the flooding period caused relatively large changes in the structural coordinates, it dominates the eigenvectors and results in the apparent distortion of the first 308 ps.

To understand the structural change, residue-based rmsd plots between structures at 308, 436, and 840 ps were constructed. We examined the rmsd's between structures before and after flooding (Figure 6a), before flooding and after MD2 (Figure 6b), and after flooding and after MD2 (Figure 6c). The principal structural deviation for the first conformational transition was concentrated in two regions of primary sequence (40–55 and 82–105), throughout the different stages of computational treatment. The rmsd comparison of the structures show that the repositioning of the C, D, F, and G helices are significant in defining the newly located tier-1 CS. After MD2, four additional residues outside the region of structural change, produced by flooding, were found to have rmsd values greater than 3 Å (Figure 6b). The first three residues, Gln8, Asp20, and Lys56, formed new hydrogen bonds, while His113 moved due to changes in the G helix. Each isolated residue change was determined to have little impact on the newly determined tier-1 CS but is thought to define lower-level tier CS substates. The rmsd values between structures after flooding and after MD2 (Figure 6c) show an overall reduction in the two primary regions of change. The interpretation is that the CF method coarsely located the new tier-1 CS, while the following MD2 simulation relaxed the protein into one of the many possible lower-order tier conformations. In summary, residues 40–55 (C, CD, and D), 82–105 (EF, F, FG, and G), and the heme were identified as the regions which changed the most (Table 1).

While the rmsd plots in Figure 6 suggest that there are three major regions of structural change (including the heme), Figure 7 illustrates the three-dimensional structures of these motions in overlaid snapshots before and after conformational flooding. In Figure 7a, the protein backbone with the strongest structural variation during CF 1 (C, D, F, and G helices and connecting loops) is drawn with thick lines. The lower plot in Figure 7a is

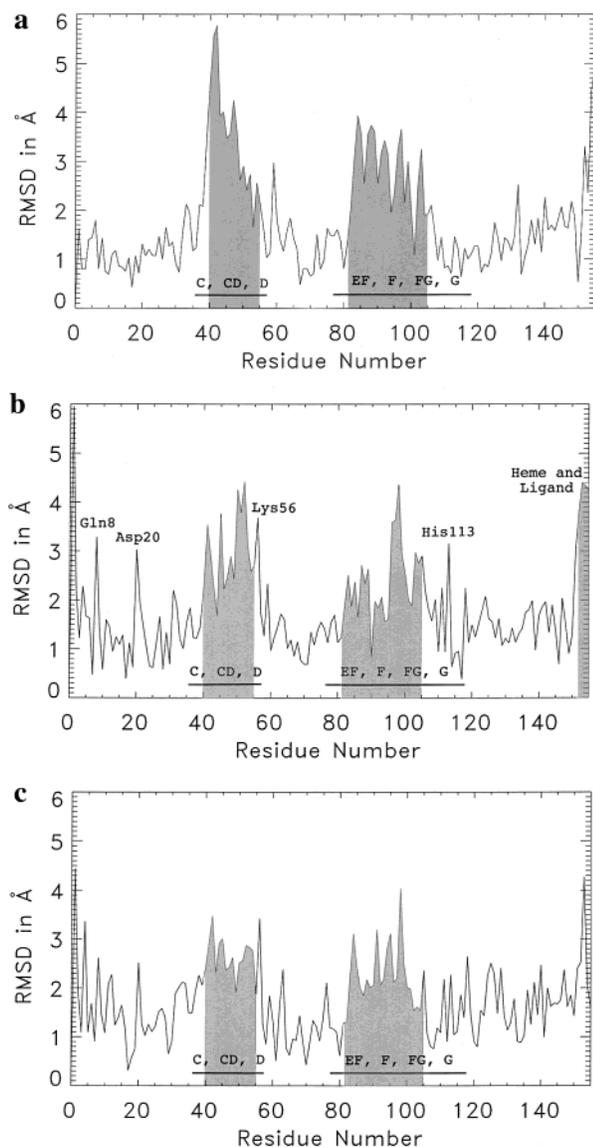


Figure 6. Residue-based rmsds for CF 1 (residues 154 and 155 correspond to the heme and CO, respectively). (a) rmsd between structures before (308 ps) and after flooding (436 ps). (b) rmsd between structures before flooding (308 ps) and after MD2 (840 ps). (c) rmsd between structures after flooding (436 ps) and after MD2 (840 ps).

rotated 90° to allow a full evaluation of distortion. The gray lines of Figure 7a refer to the structure before flooding (308 ps), and only the regions of major structural change (residues 40–55 and 82–105) are indicated for clarity. The solid black lines represent the protein after flooding (436 ps) for all backbone atoms. The conformation achieved after MD2 (840 ps) is not shown, since it primarily differentiates between lower-order tier conformations which are difficult to visualize. The seven regions of molecular motion are indicated by number in Figure 7a and 7b. The overlaid structures from 308 and 436 ps show that the most significant structural change appears on both the distal and proximal sides of the heme pocket involving the C, D, F, and G helices and connecting loops. First, residues 40 to 55 move (motions 1, 4, and 5) to uncover the distal side of the heme pocket, as shown in Figure 7a. Second, residues 82 through 105 move (motions 2 and 3) to translate the F helix. Both motions are the main contributors to the increased radius of gyration listed in Table 1. The biomechanics of the first five combined motions provide a distal pocket gate, defined by the

C and D helices and interconnecting CD loop (residues 40–55), and a proximal pocket lever, created by the EF and FG loops that cause a significant translation of the F helix (residues 82–105). Due to the heme covalent bond with His93(F), the heme is translated with the F helix, and a large rmsd of 3.9 Å results between the heme coordinates of the 308 and 840 ps configurations, as shown in Table 1. Thus, in addition to the suspected role of His64 in the taxonomic states, the computed gate-like motion and heme displacement is interpreted to regulate ligand access from the distal side of the heme. After the flooding potential was turned off (436 ps), it took the protein approximately 60 ps (as shown in Figure 5) to complete the conformational transition.

Flooding Trajectory 2 (CF 2). The second flooding trajectory with a flooding strength of 60 kcal/mol was started from the 500 ps MD structure (Figure 2), and a transition was observed after approximately 30 ps. After 100 ps ($t = 600$ ps), the flooding simulation was terminated, and MD2 was continued for an additional 300 ps ($t = 900$ ps). After the start of MD2, the structure underwent additional conformational change, as visualized in Figure 8. This plot was constructed in the same fashion as Figure 5 by applying PCA to the Cartesian coordinates of the structures from the original MD trajectory (0–500 ps), CF 2 (500–600 ps), and MD2 trajectory (600–900 ps). The conformational changes were mainly between 500 and 600 ps, again demonstrating that the CF method drove the protein into a new region of phase space. After 640 ps, the conformation remained in the new substate. The figure demonstrates that the protein assumed a new and stable conformation, significantly different from the CF 2 starting point at 500 ps and original MD structures. Note how the original MD trajectory (shown from 0 to 500 ps) appears different from the trajectory in Figure 2. As described earlier, the structural change induced by the flooding period influences the computed eigenvectors used for plotting.

The magnitude of structural change between the starting point (before flooding) and final CS (after MD2) was similar to that found for CF 1, yet occurred for different residues. The regions which changed most significantly involved residues 77–84 (EF loop) and 122–138 (GH loop and first part of H helix), as summarized in Table 1. The corresponding overlaid three-dimensional structures are shown in Figure 7b, where the protein backbone residues 77–84 and 122–138 are drawn with thick lines to emphasize the regions of largest structural change. The gray lines correspond to the most significant structural change before flooding (500 ps) with the other atoms removed for clarity. The solid black lines indicate the structure after flooding (600 ps) for all backbone atoms. The most significant conformational change is seen in Figure 7b (lower) where the EF loop bends outward (motion 6) and the GH loop and H helix bend away from the protein (motion 7). This produces a considerable widening of the ligand binding cavity. Consequently, the GH loop includes more residues at 900 ps, and the H helix has been effectively shortened. This motion is interpreted to modulate the hydrophobic pocket volume and regulate ligand access from the proximal side of the heme.

Flooding Trajectories 3 and 5. Additional flooding trajectories with values for E_{fl} between 54 and 66 kcal/mol were started from the 619 (CF 3) and 800 ps (CF 5) structures of the original MD trajectory, respectively. No stable conformational transitions were computed. In the case of CF 3, a clear transition was computed with E_{fl} of 54 kcal/mol, but the new conformation was not stable, and the protein moved back into its original CS within 200 ps of MD2. In the case of CF 5, no clear

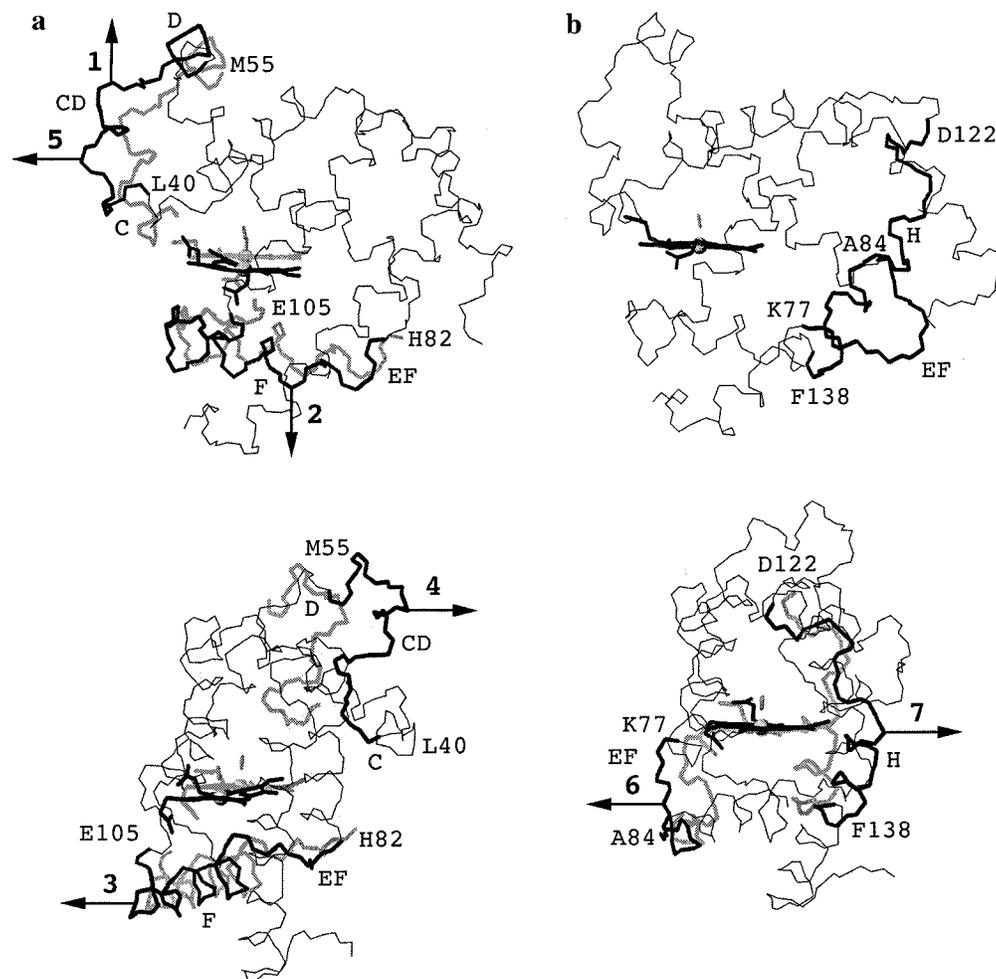


Figure 7. Superimposed protein backbone plots for the two unique CF 1 and CF 2 conformational transitions. (a) Side and top views involve the overlap between “before CF 1” (308 ps, gray) and “after CF 1” (436 ps, black). (b) Side and top views involve the overlap between “before CF 2” (500 ps, gray) and “after CF 2” (600 ps, black). The five molecular motions are illustrated using arrows. See text for explanation.

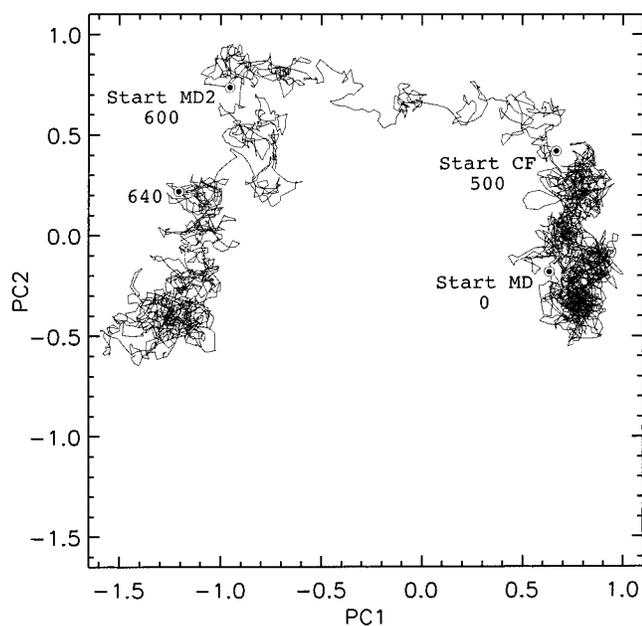


Figure 8. Projected trajectory for CF 2: original MD (0–500 ps), CF 2 (500–600 ps) and subsequent MD2 simulation (600–900 ps). Units for the PC1 and PC2 axes are in Å.

conformational changes were observed, and only nonspecific deformations occurred. Figure 2 indicates that during the last

10 ps of the original MD simulation the protein moved to a conformation similar to the initial PDB structure (upper right part of figure). Although the flooding matrix which defines the overall shape of the flooding potential has been computed from all structures between 300 and 800 ps, the main contribution comes from the structures of the left-hand cluster in Figure 2. Thus, the flooding potential is specific for the large substate. It does not sufficiently represent the structural characteristics of the conformation at 800 ps and cannot induce a conformational transition in CF 5. This illustrates that if the protein is not in the CS for which V_{fl} is designed (e.g., already flooded out of this valley), then its structure and dynamics are no longer affected by the flooding potential.

Flooding Simulation 4 (CF 4). The third complete conformational transition was observed during CF 4, which started from the 714 ps transient structure of the original MD simulation (Figure 2) and used a flooding strength of 67 kcal/mol. A transition was observed after approximately 5 ps. The flooding simulation was terminated at 863 ps, and MD2 was continued until 1200 ps. After the start of MD2, the structure underwent more conformational change, as visualized in Figure 9 (analogous to Figure 5). After 880 ps, the conformation stabilized and did not return to the original substate. Thus, a new and stable conformation was located.

To understand the structural change, residue based rmsd plots between structures at 714 ps (before flooding) and 1200 ps (after MD2) were constructed (not shown), as summarized in Table

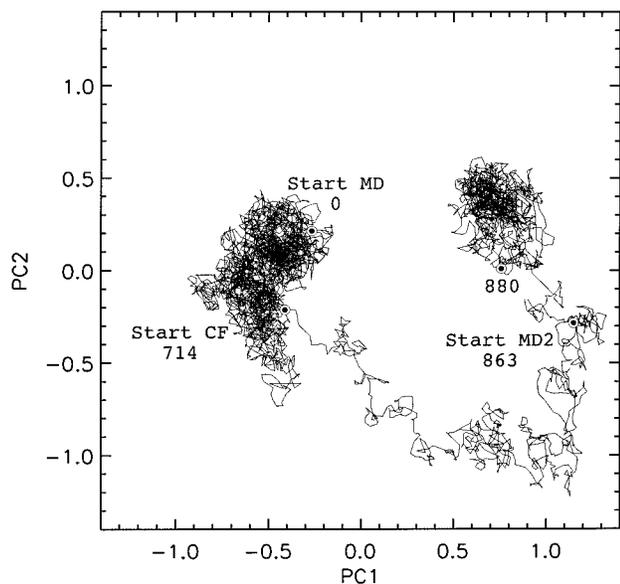


Figure 9. Projected trajectory for CF 4: original MD (0–714 ps), CF 4 (714–863 ps) and subsequent MD2 simulation (863–1200 ps). Units for the PC1 and PC2 axes are in Å.

1. As found with the other flooding trajectories, principal structural changes occurred in several regions. A rmsd comparison showed that the repositioning of the C, D, F, and G helices are significant in defining the newly located CS. Interestingly, the conformational transition resulted from a combination of seven molecular motions that had been observed in CF 1 and CF 2. Overall, the reproduction of features from CF 1 and 2 suggests that only few tier-1 conformational transitions are allowed energetically, which assists in defining the hierarchy of conformational substates.

Solvent Accessibility of the Heme. The structural changes in the CD region indicate that conformational transitions in MbCO can affect the immediate environment of the heme prosthetic group and the ligand binding process. Therefore, the solvent-accessible surface area of the heme was studied in all flooding trajectories. The free heme has a surface area of 900 Å². Figure 10a shows the heme's water-accessible surface area in the original MD trajectory, which underwent rather large fluctuations, from 110 to 240 Å². At the starting point of CF 1 (308 ps), the computed solvent-accessible surface area was rather large, around 195 Å². Application of the flooding potential at this time point resulted in a further opening of the cleft into which the heme is embedded. Therefore, the heme's solvent-accessible surface area increased dramatically. Whereas in the original trajectory, only the heme propionates interacted with water, CF 1 and the subsequent MD trajectory resulted in a larger area of the heme exposed to water. The number of water molecules within 3 Å of the heme increased by 50% for snapshots before, after CF 1, and after MD2. The starting points of the other two flooding trajectories (CF 2 and CF 4) coincided with conformations where less of the heme was exposed to solvent (160 and 140 Å², respectively), and consequently the flooding potential had less effect on the heme environment. Comparison of the original 800 ps MD trajectory with a set of four other MD simulations of the solvated protein (Figure 10b) indicates that the heme's water-accessible surface area varies between trajectories as much as within a single trajectory.⁹¹ The surface area of these four trajectories (starting from different PDB structures) fluctuated between 80 and 160 Å². The computed differences could be partially due to the starting

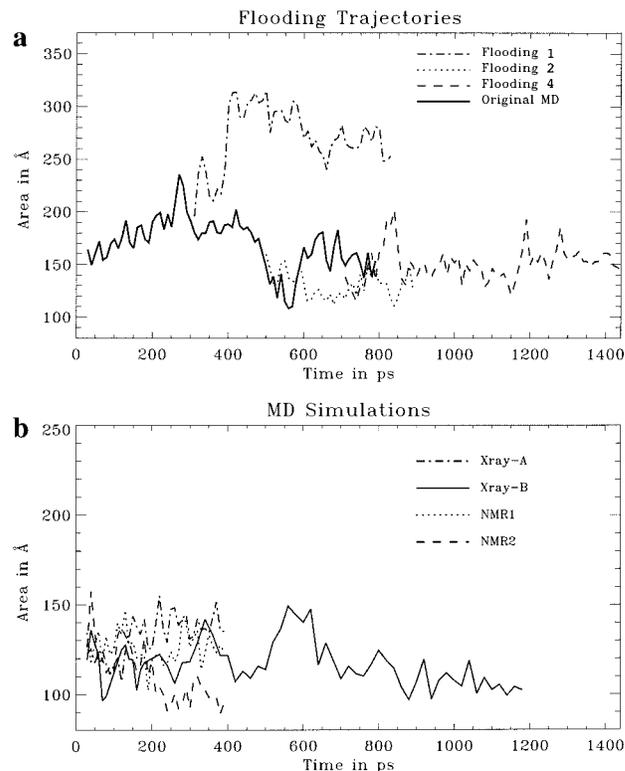


Figure 10. Solvent-accessible surface area of the heme group. (a) Analysis of the 800 ps MD trajectory and the three CF simulations described in the text (each composed of flooding and MD2). (b) Analysis of four regular MD simulations with MbCO simulated analogously to the 800 ps trajectory of the Neutron diffraction PDB structure. X-ray-A (400 ps) and X-ray-B (1200 ps) start from two conformations of an X-ray structure, NMR1 (400 ps) and NMR2 (400 ps) start from conformations determined by NMR spectroscopy.

conformation where the neutron diffraction structure has the greatest solvent-exposed heme group.

Helical Reorganization. The effect of the flooding potential on the helical arrangement during the three flooding trajectories was studied. Helices A–H were determined for each structure using MOLMOL.¹⁴³ Figure 11 summarizes the results for the PDB structure and for snapshots from the three flooding simulations before and after CF and after MD2. Overall, the changes are rather small, and all large helices remained relatively undisturbed. Most change involved helices C, F, and H. Helix C was diminished during CF 1 and in the original MD simulation before CF 4 had started. During CF 4 it was reformed. CF 2 did not affect helix C, but the length of helix F changed significantly. During CF 1 it increased by more than 50% from 8 to 13 residues, during CF 4 there was a decrease of the same magnitude. In most structures, the GH loop rearranged to form a small helix. Only in CF 2, this helix transformed into a loop, and helix H was shortened by 3 residues, giving rise to the increased radius of the GH loop in Figure 7b. Overall, the changes appear small compared to the large rmsd values reported for helices C, D, F, and H in Table 1. The data confirms that the helical structures, besides the changes noted for the C, D, F, and H helices, are quite stable and remain largely unaffected by the conformational transitions. Instead, structural deformations occur almost exclusively in the loops, which is in agreement with findings by Elber and Karplus that motions in loop regions initiate conformational transitions.¹⁴⁴

(143) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51.

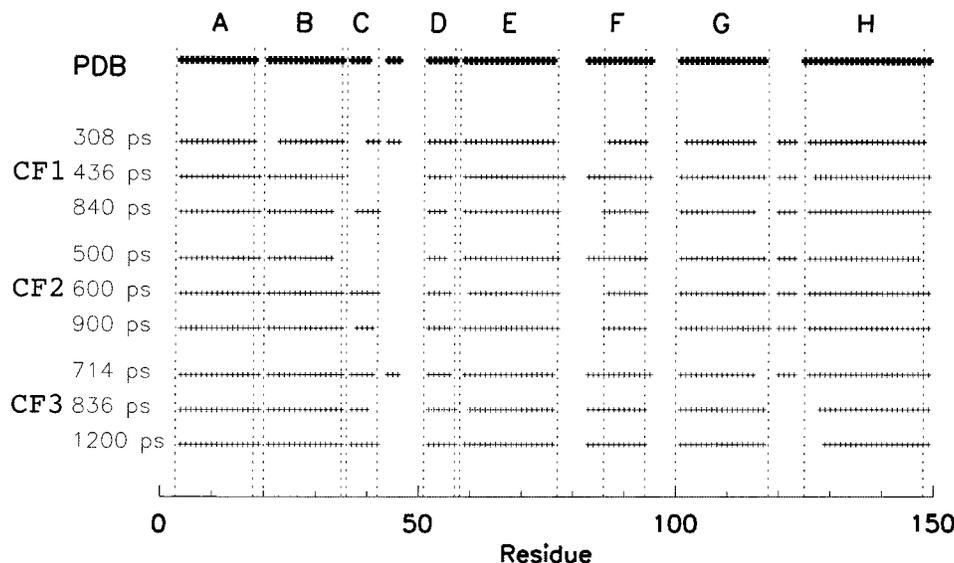


Figure 11. Helical arrangement for PDB structures and snapshots from flooding simulations. The vertical dashed lines indicate PDB helices A through H as defined in the text. A small cross symbol indicates that the respective residue is part of a α -helix.

Table 2. Correlation between CF Conformational Transitions and Ligand Escape Routes

flooding simulation	transitions during flooding simulation	correlating ligand escape pathway
1	40–55 82–88 95–112 heme,CO	CD loop ¹⁰⁶ overlap of EF loop and N-terminus ¹⁰⁶ proximal heme pocket ^{103,106,110}
2	43–63 77–84 122–138 149–153	CD loop ¹⁰⁶ overlap of EF loop and N-terminus ¹⁰⁶ overlap of G helix and AB loop ^{106,147}
4	40–47 75–84 135–140 152–153	CD loop ¹⁰⁶ overlap of EF loop and N-terminus ¹⁰⁶ overlap of G helix and AB loop ^{106,147}

Correlation of Experimentally Determined Flexible Regions in Mb. Previous publications of a Mb–Xe complex,¹¹¹ Mb mutants,^{104,109,110} ultrafast phase grating studies,¹⁴⁵ analysis of mean-square displacements in an X-ray diffraction investigation,¹⁴⁶ and MD simulations of Mb–ligand complexes^{105,111} have shown that the dissociated ligand can take several pathways through the protein matrix to move from the heme to the exterior and vice versa. As a result, a number of regions have been identified in the protein where the ligand is most likely to move. The correlation between strong motions in the flooding simulations and previously observed ligand escape routes is summarized in Table 2. The data shows that the flooding transitions correspond especially well to ligand pathways through the CD region (distal pocket),¹⁰⁶ the proximal pocket,^{103,106,110} through the region around the EF loop,^{106,109,146} and through the GH loop.^{106,146,147} In two simulations (CF 2 and to a lesser degree CF 4) large translation in the H helix occurred. Distortion of MbCO in the specific direction would support ligand escape via the EF and GH loop areas. The motion corresponds to ligand

dissociation which occurs parallel to the heme plane, consistent with translation of the F helix.¹⁴⁵

Correlation to Results of Normal-Mode Analysis. Normal-mode analysis allows the study of the dynamic properties of a protein, for example, atomic fluctuations and their correlation to each other, characterization of low modes, and identification of rigid and softer structural units at a local minimum of its energy surface. This method was applied by Seno and Gō to study the heme–globin interactions¹⁴⁸ and conformational changes upon oxygenation.¹⁴⁹ The authors conclude that the protein can be divided into two large and rigid units (B to E helix and F to H helix) connected by the EF loop, as the most flexible unit in the protein. They identified 55 residues within 6 Å near the heme with atomic fluctuations closely correlated to heme fluctuations. Upon oxygenation, large atomic displacements were found in the region from the F helix to the beginning of the G helix (residues 85–105). The analysis did not identify low-frequency fluctuations in the distal region around the heme (helices C and D and loop CD), which one would expect to be functionally important for ligand diffusion through the protein. Overall, our results agree with the findings on the significance of the EF, F, and FG regions. However, our flooding simulations identified more flexible regions on the distal side of the heme (especially residues 40–55). Therefore, greater insight in terms of an increased solvent accessibility of the heme pocket and changes in helical structures were made with the CF simulations.

Biomechanics and Energy Landscape. Transition rates between the taxonomic states in 75% glycerol solution have been previously reported using infrared spectroscopy in combination with flash photolysis,⁶⁵ pressure release,^{28,75,77} double-pulse kinetic selection,⁶⁶ spectral hole burning and photon echo,²⁴ and kinetic hole burning and filling techniques.⁶⁷ The existence of many more than 10⁶ substates within the tier-1 substates have been shown by non-exponential rebinding⁶³ and hole burning experiments¹⁵⁰ at low temperatures. The computed structures in this report support the hierarchical model of protein structural and energetic change by Frauenfelder and co-workers.^{22–33} The current results assist in piecing together the structural details to better understand the hierarchy of the energy

(144) Elber, R.; Karplus, M. *Science* **1987**, 235, 318.

(145) Deak, J.; Chiu, H.-L.; Lewis, C. M.; Miller, R. J. D. *J. Phys. Chem. B* **1998**, 102.

(146) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature* **1979**, 280, 558.

(147) Ulitsky, A.; Elber, R. *J. Phys. Chem.* **1994**, 98, 1034.

(148) Seno, Y.; Gō, N. *J. Mol. Biol.* **1990**, 216, 95.

(149) Seno, Y.; Gō, N. *J. Mol. Biol.* **1990**, 216, 111.

(150) Gafert, J.; Ober, C.; Orth, K.; Friedrich, J. *J. Phys. Chem.* **1995**, 99, 14561.

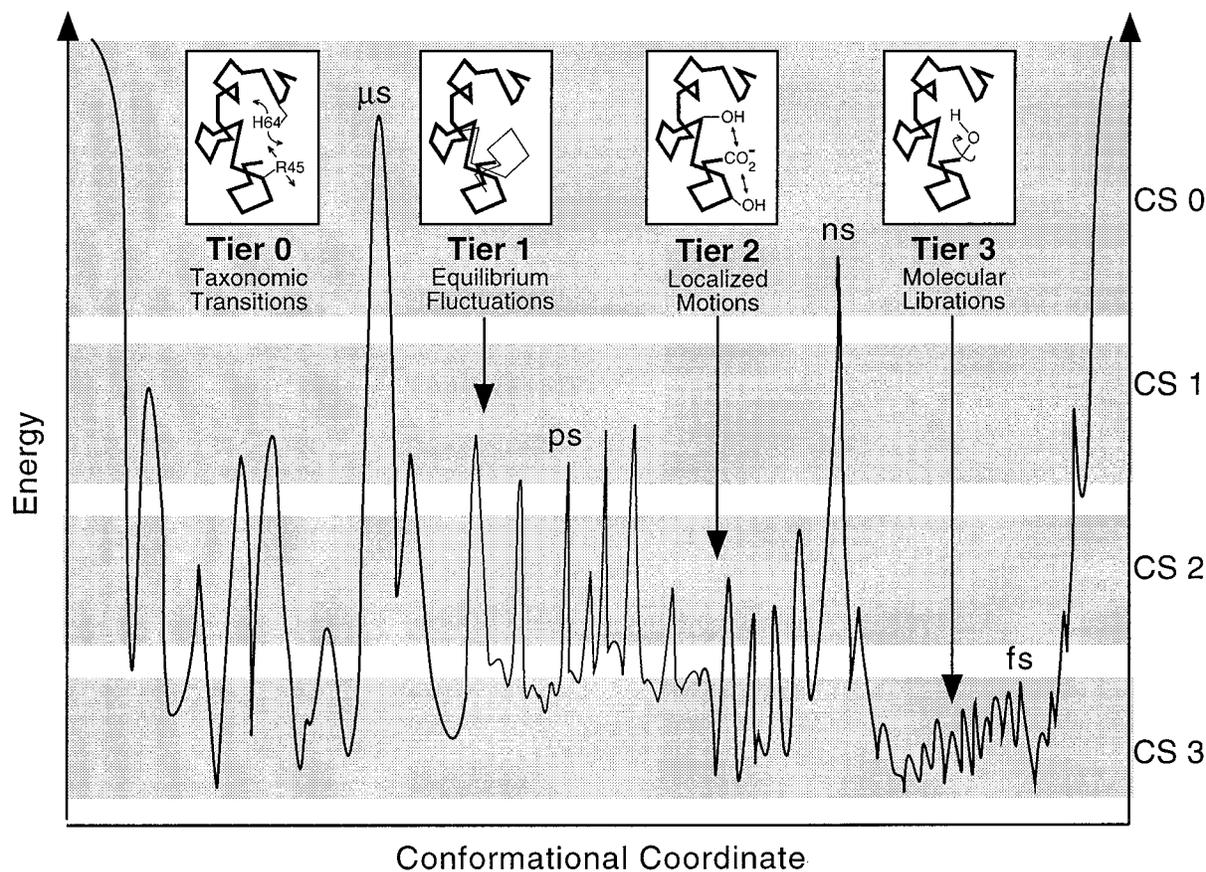


Figure 12. One-dimensional cross-section sketch of the free energy landscape of a protein. The heights of the energy barriers separating conformational substates determine the conformational transition times.

landscape, as shown in Figure 12. Seven principal motions in MbCO were identified repeatedly in three flooding simulations. Combinations of the seven motions produce the two computed conformational transitions. The combination of the first, fourth, and fifth molecular motion results in a *distal pocket gate* defined mainly by the C and D helices and interconnecting CD loop (residues 40–55). The computed gate-like motion is interpreted to regulate ligand access from the distal side of the heme. The second and third principal mode can be summarized as a *proximal pocket lever* involving the F helix and surrounding EF and FG loops (residues 82–105), which shuttled the heme deep into the protein matrix (heme rmsd of 3.9 Å) as the distal pocket gate simultaneously opened. The lever's effect on the heme motion was assumed to attract ligands into the heme. The second conformational transition is called the *hydrophobic pocket expansion* modulated by EF loop motion (sixth motion) and GH loop expansion (seventh motion). The computed motion is most likely related to the hydrophobic pocket's ability to accommodate CO ligands and regulate ligand access from the proximal side of the heme.

The technique of conformational flooding has produced conformational transitions beyond the time domain reached by conventional molecular dynamics. Despite the discovery of new tier-1 structural details and conformational transitions, there is no guarantee that all structures and higher-energy transitions have been located, although, it seems unlikely that a significant number of physically relevant structures have been missed by the combined MD and CF method. Therefore, the computed results suggest that there are a finite number of localized motions and conformational transitions for the tier-1 substate within the spectroscopic tier-0 A₀ state. This important result is consistent

with the experimental data of more than 10⁶ statistical substates beyond the taxonomic states, since it is plausible that a large number of transitions occur in the lower tiers when the higher-energy tier-1 barriers are frozen at low temperatures.¹¹ The rmsd structural comparisons, as previously discussed, show that several residues switch hydrogen bonding partners after MD2 simulations. We propose that these low-energy molecular liberations belong to the lower-tier substates and account for the inhomogeneous broadening of the individual stretch bands^{151–153} and the non-exponential time dependence of CO rebinding³¹ in each of the taxonomic states.

Conclusions

The first application of CF simulations to a protein has been presented. The strength of the flooding potential was adjusted to induce realistic conformational transitions in MbCO. Three flexible regions are identified to be responsible for conformational transitions on long time scales. The two computed transitions are composed of seven principal molecular motions. The localized motions involve (1) helices C and D and the CD loop, (2) helix F with the EF and FG loops and the heme, (3) GH loop and part of the H helix. The two computed conformational transitions are attributed to equilibrium fluctuations on the tier-1 hierarchical description of protein structure and dynamics. The computations indicate that changes in both the

(151) Agmon, N. *Biochemistry* **1988**, 27, 3507.

(152) Campbell, B. F.; Chance, M. R.; Friedman, J. M. *Science* **1987**, 238, 373.

(153) Ormos, P. A.; Ansari, A.; Brunstein, D.; Cowen, B. R.; Frauenfelder, H.; Hong, M. K.; Iben, I. E. T.; Sauke, T. B.; Steinbach, P. J.; Young, R. D. *Biophys. J.* **1990**, 57, 191.

(154) *WebLab ViewerLite*, ver. 3.2; Molecular Simulations Inc., 2000.

distal and proximal sides of the heme define the tier-1 CSs. All computed biomechanical motions correlate well to previously found ligand diffusion pathways through the protein. The observed motions rationalize the most significant ligand pathways through the protein established in experimental and theoretical studies and confirm results from normal-mode analysis. A unified model that utilizes both mechanisms of distal His64 modulation (tier-0) and protein equilibrium fluctuations (tier-1) explains ligand diffusion in the MbCO dissociation reaction.

Acknowledgment. We thank Professor Paul Tavan and Dr. Markus Eichinger from the Ludwig-Maximilians-Universität

Munich, Germany, for their support and for helpful discussions. We are grateful to the Summer Award in Natural Science and Engineering and the University Research Council from the University of Miami, the Petroleum Research Fund (PRF No. 30918-G4), the Department of Defense (DAAH04-96-1-0311 and DAAG55-98-1-0067), and the IBM Corporation for financial support to make this research possible. The MD trajectories were computed at the SP2 of the Cornell Theory Center and at the CRAY T3E at the Research Centre Jülich, Germany, and we acknowledge the computer time allocated for this work.

JA993788Y