Discovery of new ligand binding pathways in myoglobin by random mutagenesis

Xiaohua Huang and Steven G. Boxer

A random library of single amino acid mutants of myoglobin was generated using a highly efficient, single-base-misincorporation random mutagenesis method to discover new ligand-binding pathways in myoglobin. A surprisingly large fraction of the library exhibits ligand-binding kinetics that are substantially different from the wild-type protein. In addition to residues 45, 64 and 68, which comprise the best studied ligand-binding pathway single mutations of several other clusters of residues far away from that pathway are discovered which profoundly affect the ligand-binding kinetics. These results provide a new approach to explore the relationship between the fluctuations in protein structure and function.

Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA

Correspondence should be addressed to S.G.B.

The binding of diatomic ligands to myoglobin (Mb) is one of the best characterised examples of the coupling of protein dynamics and function. Examination of the X-ray structures of many Mbs shows that there are no obvious pathways from the protein surface to the haem-iron binding site[1–5]. Thus fluctuations in the structure are required for physiological function[3]. This dilemma has been addressed by examining X-ray structures with small[2,6,7] and bulky[8–10] ligands bound to the haem iron, temperature-dependent X-ray diffraction[11] and molecular dynamics simulations[12–15]. These studies suggest that distal residues His 64, Val 68 and Arg 45 (Lys 45 in many species) comprise a primary pathway, and this has been extensively explored with single and multiple mutations of these residues[16–24]. The question remains whether this traditional pathway is a dominant or even major pathway or whether, as some recent calculations[13,25] and experiments[18,26] suggest, many pathways compete.

Although a systematic approach to uncovering novel ligand-binding pathways, using site-specific mutagenesis guided by computer modeling and molecular dynamics simulations, can be very useful and has been pursued by many groups[16–24], an alternative is to prepare a random library of single-amino-acid mutants. This strategy is not biased by the history of prior efforts and predictions. There are many recent examples of the use of random libraries where some form of phenotypic selection is used to reduce the pool of interesting mutants to a manageable number for further study[27–29]. This approach is not useful in the present case because *Escherichia coli*, in which the Mb gene is expressed, does not depend on Mb for survival. Thus we have been forced to adopt the brute force approach of discovering interesting mutants by examining the function of a large number of mutant proteins. This represents one of a few cases to date where a library is screened for a specific function using a series of precise physical measurements, rather than by classical genetic or biochemical selection.

A random library of myoglobin mutants

A new method was used to generate a random library of single base mutations of the sperm whale myoglobin gene (see Methods section and Fig. 1). Time-serial digestion of the DNA by exonuclease III produced a uniform and random distribution of 3' ends over the whole gene. Single base mutations were then efficiently introduced into the 3' ends, and the DNA with the mis-paired base was properly extended. This was confirmed by standard DNA gel electrophoresis. A number of colonies were picked randomly from three out of eight time-serial pools of the mutants and their DNA was isolated and sequenced. Most of the colonies contain single base mutations, and these are evenly and randomly distributed over the expected regions of the Mb gene. The plasmids carrying the mutated genes were transformed into *E. coli*, and mutant Mbs were expressed[30]. A primary selection is made based on whether the *E. coli* colonies turn reddish-brown, indicative of the presence of Mb. Some colonies remain white, most likely because the genes contain a stop codon, deletions/insertions, or code for sequences which are not stable or do not fold correctly. The possible misfolding mutants are also potentially very interesting and will be discussed elsewhere.

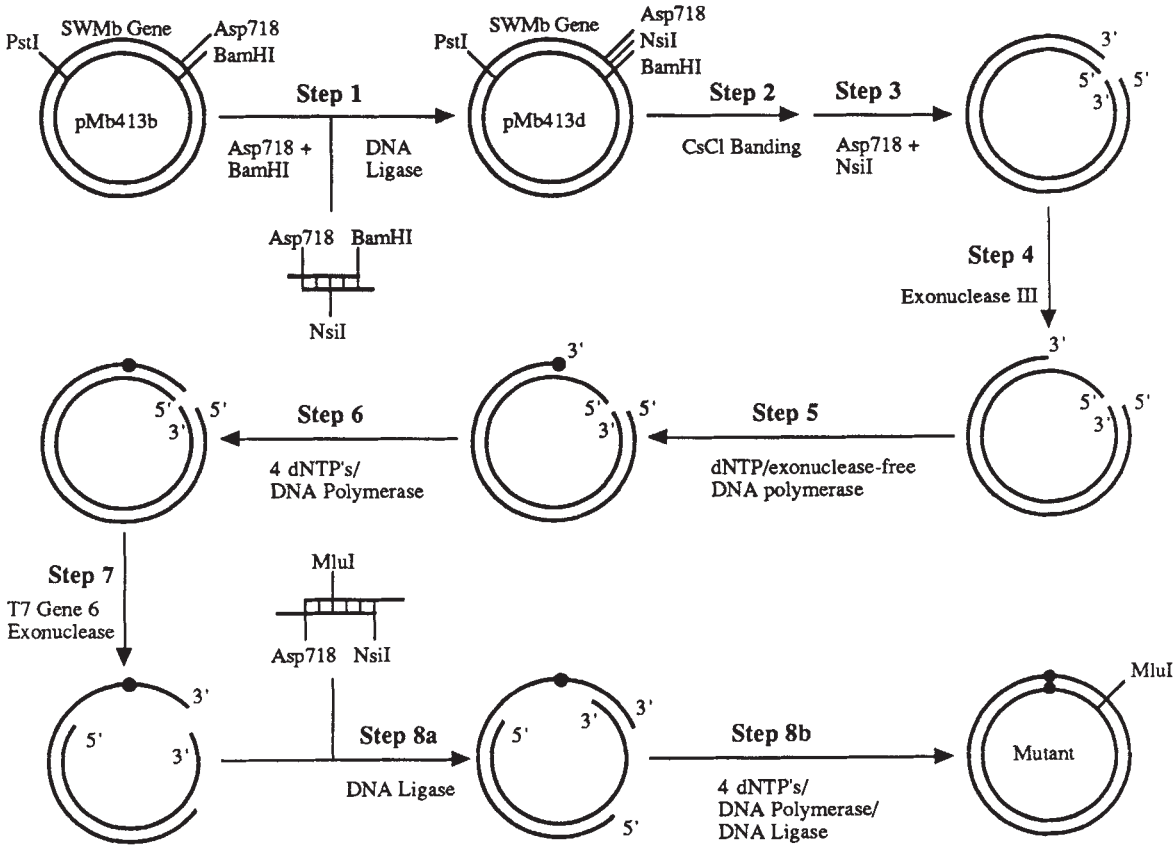


Fig. 1 A new random mutagenesis method used to prepare the random library of Mb mutants.

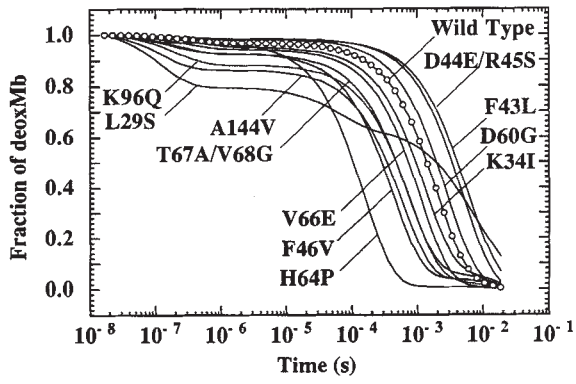


Fig. 2 Recombination of CO following photo-dissociation with a 9 ns laser flash for a number of mutants discovered by screening the random library (Mbs in crude lysate of *E. coli* in 50 mM Tris. Cl and 1 mM EDTA, pH = 8.0, saturated with 1 atm of CO at 23 °C). The data are plotted as normalized fraction of deoxy Mb vs. time on a logarithmic scale to include both geminate rebinding (times shorter than microseconds) and bimolecular recombination (the longest phase varying from about 10 μs to 10 ms).

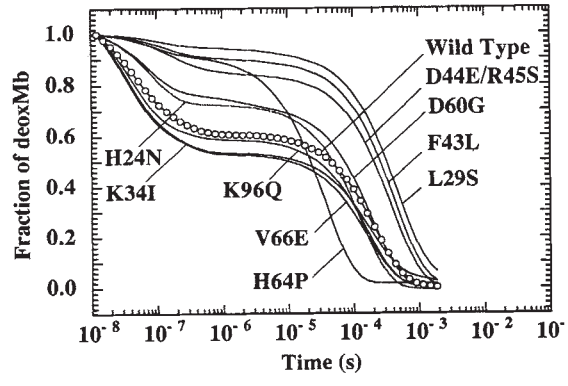
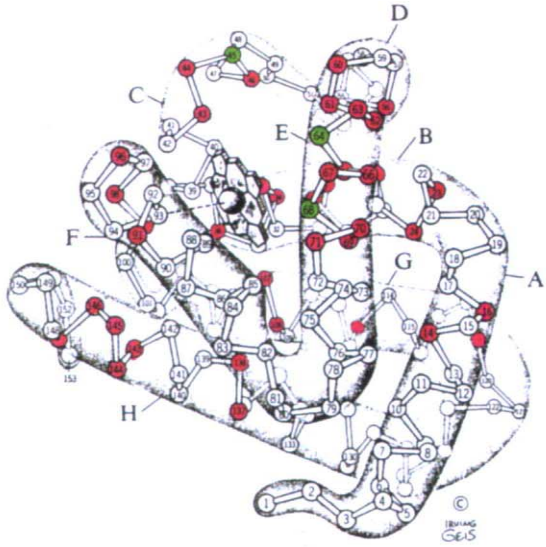


Fig. 3 Recombination of O_2 following photo-dissociation with a 9 ns laser flash for a number of mutants discovered by screening the random library (Mbs in crude lysate of *E. coli* in 50 mM Tris. C1 mM EDTA, pH = 8.0, saturated with 1 atm of air at 23 °C). The data are plotted as normalized fraction of deoxy Mb vs. time on a logarithmic scale to include both geminate rebinding (times shorter than microseconds) and bimolecular recombination (the longest phase varying from about 10 μs to 1 ms). The full timecourse of the geminate process is not resolved because the earliest geminate events are faster than the pulsewidth of the laser.

Ligand-binding kinetics of the mutants

O_2 and CO rebinding kinetics on the nanosecond to millisecond timescale were measured on mutant Mbs in crude lysates of *E. coli* using standard flash photolysis methods[31]. These kinetics are highly sensitive to changes in protein structure, and can be measured accurately on relatively crude fractions of mutant proteins. All aspects of this process were automated using a robot of local design. The following criteria were used to distinguish significant from insignificant changes in ligand binding kinetics compared to the wild-type Mb: for the bimolecular recombination rate constant, an increase greater than 50% for CO and 40% for O_2, or a decrease greater than 25% for CO and 15% for O_2; for the relative geminate recombination yield, an increase or decrease greater than 50% for CO and 20% for O_2. Although arbitrary, these represent large changes in ligand binding kinetics and lead to a manageable number of interesting mutants. The genes coding for a large number of the mutants which met these criteria were then sequenced.

To date, the kinetics of approximately 1500 mutants have been measured. A large fraction (about 10%) was found to exhibit major differences compared with the wildtype. Some examples of the large variations in CO and O_2 recombination kinetics are shown in Figs 2 and 3, respectively. Fig. 4 shows three-dimensional representations of Mb with positions of high sensitivity to mutations highlighted in red. It is immediately evident from these representations that regions of the structure far removed from the classical pathway consisting of residues 64, 68 and 45 (shaded green, and which were also re-discovered in the course of the search), can profoundly affect the ligand binding kinetics. Although a substantial fraction of the library has been examined, it is likely that more pathways will be discovered by further study. Several interesting trends emerge. The regions of sensitivity to amino acid changes tend to occur as clusters in the three-dimensional structure. Residues adjacent to those comprising the traditional pathway are found to be as sensitive as those on the pathway itself. This includes many residues on the E-helix (in the vicinity of residues 64 and 68), and also residues adjacent to Arg 45. The ligand binding kinetics of Mb are insensitive to amino acid changes in large regions of the primary sequence. A substantial fraction of the mutants exhibiting the largest changes involve replacements by proline, for example His 24, Phe 46, Leu 61, His 64, Gln 91 and Lys 145 to proline. Insertions of Pro in the middle of α helices might be expected to produce unstable proteins; however, in these cases the proteins are still stable and the function is radically altered.



New ligand-binding pathways in Mb

It is possible that an amino acid substitution leads to the introduction of new ligand binding pathways or that the observed effect is the result of more global structural changes. The observation that the substitutions that affect ligand binding kinetics occur in clusters from the

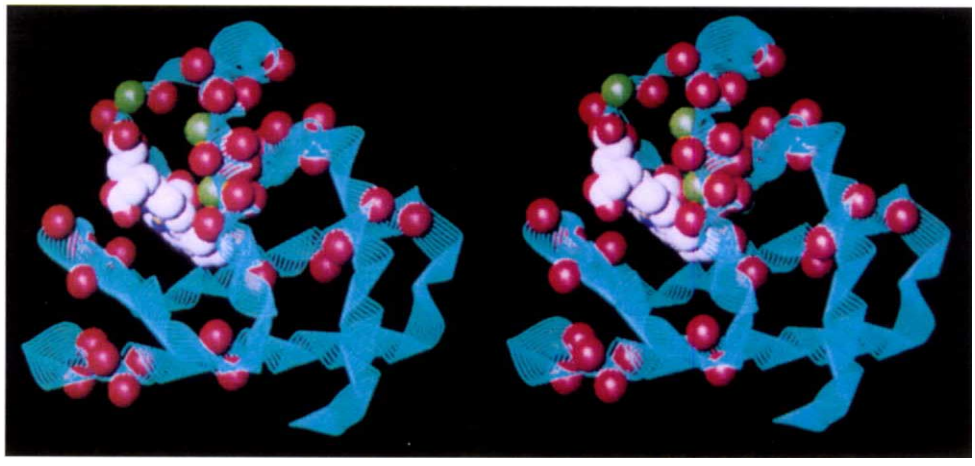


Fig. 4 *a* (top), three-dimensional structure of sperm whale Mb (adopted from Geis[33]) illustrating in red the amino acids where mutations lead to large changes in ligand rebinding kinetics, and, in green, the traditional ligand binding pathway involving residues 45, 64 and 68. Examples of specific changes affecting CO and O_2 rebinding kinetics are shown in Figs 2, 3. *b* (left), Stereo view of the myoglobin structure. Cα's of other residues where mutations lead to large change in ligand binding kinetics are highlighted in red. The α-helices are represented by ribbons, and the haem is represented by CPK model (Coordinates of sperm whale myoglobin are from Protein Data Bank[7]. The graphics were generated using O (ref. 34).

interior to the exterior of the protein, rather than in isolated residues, suggests that genuine, native pathways are being revealed. Nine clusters of residues where amino acid changes lead to large changes in ligand binding kinetics have been identified: Arg 45, His 64, Thr 67 and Val 68 between the CD loop, E helix and the haem (these are the residues forming the classical pathway); Phe 43, Asp 44, Arg 45 and Phe 46 between the CD loop and the haem; Trp 14 and Leu 69 between the A and E helices; Gln 26, Leu 29, Ile 30, Leu 61 and Lys 56 between the B, E and D helices; Phe 33, Lys 34, His 36, Leu 107 and Ser 108 between the B and G helices; Phe 33, Phe 43, Phe 46 between the B and C helices and the CD loop; His 24, Lys 16 and His 119 between the A-B and G-H helix corners; Gln 91, Ile 99, Ala 144 to Lys 145 and Tyr 146 between the F-G helix corner, and the F and H helices; and finally Ala 71, Leu 137 and Phe 138 between the E and H helices. Some of these residues (Leu 29, Val 68, Ser 108, Leu 137 and Phe 138) are associated with interior cavities revealed in the X-ray structure of the Mb-xenon complex[32]. A substantial number of residues, including Trp 14, His 24, Gln 26, Leu 29, Ile 30, Phe 33, Phe 43, Arg 45, Phe 46, Ile 61, His 64, Thr 67, Val 68, Leu 69, Gln 91, Ile 99, Ile 107, Ser 108, Leu 137 and Phe 138, coincide with key residues suggested by molecular dynamics simulation[12,13,25], demonstrating the utility of these simulations.

Methods

The method used to generate the random single mutation library of sperm whale myoglobin is summarized in Fig. 1 and will be presented in detail elsewhere. Briefly, two unique restriction sites are inserted into a plasmid carrying the Mb gene, and plasmid DNA is linearized with these two unique restriction enzymes. The linearized DNA is subjected to time serial 3' to 5' digestion with Exonuclease III to produce a random distribution of 3' ends in the gene. Misincorporation of a single base is effected by exonuclease-free DNA polymerases, exonuclease-free Klenow enzyme and Sequenase version 2.0 (from United States Biochemicals, U.S.A.) in the presence of a single deoxynucleoside triphosphate (dNTP). The mutated target DNA is repaired to blunt ends by the same exonuclease-free DNA polymerases in the presence of four deoxynucleoside triphosphates (dNTPs: dATP, dCTP, dGTP and dTTP) and the unmutated strand is removed by 5' to 3' digestion with T7 Gene 6 exonuclease. Finally, the DNA is repaired to a double-stranded closed circular form by oligonucleotide-directed recircularization ligation and DNA repair synthesis.

Received 10 January; accepted 22 February 1994.

Acknowledgements
We thank D.G. Lambright for the design and setup of the apparatus for flash photolysis, for his help with the automation of data collection and with data analysis. We also thank S.G. Sligar for making the synthetic sperm whale Mb gene available and R.L. Baldwin for providing the plasmid, pMb413b, in which the codon for residue Asn 122 has been corrected to the codon for Asp 122 in the Mb gene. This work is supported in part by a grant from the National Institutes of Health.

1. Kendrew, J.C. *et al.* Structure of myoglobin. A three-dimensional fourier synthesis at 2 Å resolution *Nature* **185**, 422–427 (1960).
2. Perutz, M.F. & Mathews, F.S. An x-ray study of azide methaemoglobin *J. molec. Biol.* **21**, 199–202 (1966).
3. Nobbs, C.L. in *Heme and Hemoproteins* (eds Chance, B., Estabrook, R. W. & Yonetani, T.) 143–147 (Academic Press, New York, 1966).
4. Takano, T. Structure of myoglobin refined at 2.0 Å resolution. I. Crystallographic refinement of metmyoglobin from sperm whale *J. molec. Biol.* **110**, 537–568 (1977).
5. Takano, T. Structure of myoglobin refined at 2.0 Å resolution. II. Structure of deoxymyoglobin from sperm whale *J. molec. Biol.* **110**, 569–584 (1977).
6. Phillips, S.E.V. Structure and refinement of oxymyoglobin at 1.6 Å resolution *J. molec. Biol.* **142**, 531–554 (1980)
7. Kuriyan, J., Wilz, S., Karplus, M. & Petsko, G.A. X-ray structure and refinement of carbon-monoxy (FeII)-myoglobin at 1.5 Å resolution *J. molec. Biol.* **192**, 133–154 (1986).
8. Bolognesi, M. *et al.* Reactivity of ferric *aplysia* and sperm whale myoglobins towards imidazole. X-ray and binding study *J. molec. Biol.* **158**, 305–315 (1982).
9. Ringe, D., Petsko, G.A., Kerr, D.E. & Ortiz de Montellano, P.R. Reaction of myoglobin with phenylhydrazine: A molecular doorstop *Biochemistry* **23**, 2–4 (1984).
10. Johnson, K.A., Olson, J.S. & Phillips, G.N., Jr. Structure of myoglobin-ethyl isocyanide. Histidine as a swinging door for ligand entry *J. molec. Biol.* **207**, 459–463 (1989).
11. Frauenfelder, H., Petsko, G.A. & Tsernoglou, D. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics *Nature* **280**, 558–563 (1979).
12. Case, D.A. & Karplus, M. Dynamics of ligand binding to heme proteins *J. molec. Biol.* **132,** 343–368 (1979).
13. Tilton, R.F. Jr *et al.* Computational studies of the interaction of myoglobin and xenon *J. molec. Biol.* **192,** 443–456 (1986).
14. Kottalam, J. & Case, D.A. Dynamics of ligand escape from the heme pocket of myoglobin *J. Am. chem. Soc.* **110**, 7690–7697 (1988).
15. Tilton, R.F. Jr, Singh, U.C., Kuntz, I.D. Jr & Kollman, P.A. Protein-ligand dynamics. A 96 picosecond simulation of a myoglobin-xenon complex *J. molec. Biol.* **199**, 195–211 (1988).
16. Lambright, D.G., Balasubramanian, S. & Boxer, S.G. Ligand and proton exchange dynamics in recombinant human myoglobin mutants *J. molec. Biol.* **207**, 289–299 (1989).
17. Balasubramanian, S., Lambright, D.G., Marden, M.C. & Boxer, S.G. CO recombination to human myoglobin mutants in glycerol-water solutions *Biochemistry* **32**, 2202–2212 (1993).
18. Lambright, D.G., Balasubramanian, S. & Boxer, S.G. Dynamics of protein relaxation in site-specific mutants of human myoglobin *Biochemistry* **32**, 10116–10124 (1993).
19. Rohlfs, R.J. *et al.* The effects of amino acid substitution at position E7 (residue 64) on the kinetics of ligand binding to sperm whale myoglobin *J. biol. Chem.* **265**, 3168–3176 (1990).
20. Egeberg, K.D. *et al.* The role of Val[68] (E11) in ligand binding to sperm whale myoglobin. Site-directed mutagenesis of a synthetic gene *J. biol. Chem.*. **265**, 11788–11795 (1990).
21. Carver, T.E. *et al.* Analysis of the kinetics barriers for the ligand binding to sperm whale myoglobin using site-directed mutagenesis and laser photolysis techniques *J. biol. Chem.* **265**, 20007–20020 (1990).
22. Smerdon, S.J. *et al.* Distal pocket polarity in ligand binding to myoglobin: structural and functional characterization of a threonine (E11) mutant *Biochemistry* **30**, 6252–6260 (1991).
23. Olson, J.S. *et al.* The role of the distal histidine in myoglobin and haemoglobin *Nature* **336**, 265–266 (1988).
24. Rizzi, M. *et al.* Crystal structure of a distal site double mutant of sperm whale myoglobin at 1.6 Å resolution *FEBS Letters* **320**, 13–16 (1993).
25. Elber, R. & Karplus, M. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglogin *J. Am. chem. Soc.* **112**, 9161–9175 (1990).
26. Chatfield, M.D., Walda, K.N. & Magde, D. Activation parameters for ligand escape from myoglobin proteins at room temperature *J. Am. chem. Soc.* **112**, 4680–4687 (1990).
27. Shortle, D. & Lin, B. Genetic analysis of staphylococcal nuclease: identification of three intragenic "global" suppressors of nuclease-minus mutations *Genetics* **110**, 539–555 (1985).
28. Oliphant, A.R. & Struhl, K. An efficient method for generating proteins with altered enzymatic properties: application to β-lactamase *Proc. natn. Acad. Sci. U.S.A.* **86**, 9094–9098 (1989).
29. Rennell, D., Bouvier, S.E., Hardy, L. & Poteete, A.R. Systematic mutation of bacteriophage T4 lysozyme *J. molec. Biol.* **222**, 67–87 (1991).
30. Springer, B.A. & Sligar, S.G. High-level expression of sperm whale myoglobin in *Escherichia Coli. Proc. natn. Acad. Sci. U.S.A.* **84**, 8961–8965 (1987).
31. Lambright, D.G., Balasubramanian, S. & Boxer, S.G. Protein relaxation dynamics in human myoglogin *Chem. Phys.* **158**, 249–260 (1991).
32. Tilton, R.F., Jr., Kuntz, I.D., Jr. & Petsko, G.A. Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9 Å *Biochemistry* **23**, 2849–2857 (1984).
33. Dickerson, R.E. & Geis, I. in *Hemoglobin* (The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, U.S.A., 1983).
34. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).