Determination of Ligand Pathways in Globins

APOLAR TUNNELS VERSUS POLAR GATES*‡

Mallory D. Salter1, George C. Blouin1,2, Jayashree Soman2, Eileen W. Singleton3, Sylvia Dewilde1, Luc Moens3, Alessandra Pesce2, Marco Nardini1, Martino Bolognesi4, and John S. Olson1,3

From the 1Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005-1892, the 2Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium, the 3Department of Physics, University of Genova, Via Dodecaneso 33, 16146 Genova, Italy, and the 4Dipartimento di Bioscienze, Università di Milano, Via Celoria 26, 20133 Milano, Italy

Background: O2 pathways in animal hemoglobin and myoglobin are controversial.

Results: Ligands enter and exit sperm whale Mb and Cerebratulus lacteae Hb by completely different pathways.

Conclusion: Rational mutagenesis mapping can identify ligand migration pathways and provides experimental benchmarks for testing molecular dynamics simulations.

Significance: Globins can use either a polar gate or an apolar tunnel for ligand entry.

Although molecular dynamics simulations suggest multiple interior pathways for O2 entry into and exit from globins, most experiments indicate well defined single pathways. In 2001, we highlighted the effects of large-to-small amino acid replacements on rates for ligand entry and exit onto the three-dimensional structure of sperm whale myoglobin. The resultant map argued strongly for ligand movement through a short channel from the heme iron to solvent that is gated by the distal histidine (His-64(E7)) near the solvent edge of the porphyrin ring. In this work, we have applied the same mutagenesis mapping strategy to the neuronal mini-hemoglobin from Cerebratulus lacteae (CerHb), which has a large internal tunnel from the heme iron to the C-terminal ends of the E and H helices, a direction that is 180° opposite to the E7 channel. Detailed comparisons of the new CerHb map with expanded results for Mb show unambiguously that the dominant (>90%) ligand pathway in CerHb is through the internal tunnel, and the major (>75%) ligand pathway in Mb is through the E7 gate. These results demonstrate that: 1) mutagenesis mapping can identify internal pathways when they exist; 2) molecular dynamics simulations need to be refined to address discrepancies with experimental observations; and 3) alternative pathways have evolved in globins to meet specific physiological demands.

Most animal myoglobins (Mbs) and red cell hemoglobins (Hbs) bind O2 reversibly for storage or transport. The ligand covalently coordinates to the heme iron atom and is fully enclosed by amino acid side chains to prevent exposure to solvent and oxidation. In crystal structures, there is often no obvious pathway for ligand migration into the distal portion of the heme pocket shown in Fig. 1A. In 1966 after publication of the first high resolution globin structure, Perutz and Matthews (1) suggested that O2 enters Hb and Mb by a short channel gated by the distal histidine at the E7 helical position, and this idea was later supported by outward movement of the distal histidine when large ligands were bound to Mb (2, 3). However, in 1969, internal cavities were discovered in sperm whale myoglobin (SwMb) capable of binding xenon at moderate pressures (≥10 atm (4–9)). This observation led Petsko and co-workers (5) and others in the 1980s to suggest that instead of passing through the His(E7) gate, diatomic gases may enter and exit globins through apolar pathways between helices and the loops connecting them.

The current evidence for movement of photodissociated ligands out of the distal pocket and into the Xe1 and Xe4 sites in SwMb is clear and comes from time-resolved spectroscopic methods (10–13) and from direct observations by time-resolved x-ray crystallography (14–20). However, these measure-
Mapping Ligand Pathways in Globins

ments cannot detect whether ligands enter and exit the protein from these transient positions. Ligand movements out of the protein involve overcoming larger activation barriers, and the resultant unstable transition states cannot be seen directly. Consequently, entry and exit pathways have to be inferred either theoretically from molecular dynamics (MD) simulations or experimentally from the effects of mutagenesis on kinetic parameters for ligand entry and escape.

In 1990, Elber and Karplus (21) carried out locally enhanced simulations of ligand movement in SwMb and observed multiple trajectories for ligand migration out of the protein. In 2010, Elber (22) reviewed the most recent computational literature, which still indicates that there are multiple ligand pathways in animal myoglobins and hemoglobins and that the channel gated by the distal histidine is only a minor route for entry and exit (23–31).

In 1994, Huang and Boxer (32) used an elegant random mutagenesis protocol to create and screen a library of ~1,500 single point mutants. Mutations were identified from colonies that produced lysates containing Ms with kinetic parameters significantly different (>-50%) from those of wild-type (WT) controls. Roughly 10% of the mutants examined showed differences, and a structural model of Mb was constructed to highlight the positions of the mutants showing altered kinetic parameters. In addition to positions near the bound ligand, other regions were highlighted, including interior positions near the Xe cavities and the C-terminal portion of the H helix. These results were interpreted to mean that multiple pathways for CO and O2 entry and exit exist, supporting the conclusions from the initial MD simulations (32).

Building on Huang and Boxer’s strategy, our group explored the E7 gate and Xe cavity pathways in SwMb by a rational mutagenesis mapping strategy in which small (Gly or Ala) and large (Phe or Trp) mutations were constructed in ligand-accessible cavities, along the major trajectories specified by the MD simulations and at some of the positions found in the random mutagenesis screen (13). Each of 90 different mutants was produced in milligram quantities; gerninate and bimolecular rate parameters for ligand binding were measured with pure proteins; and in many cases, crystal structures were determined (33–38).

Our mutagenesis mapping results strongly indicate the following: 1) although O2 can move from the binding pocket into the Xe cavities, it must return to the distal pocket to exit the protein, and 2) the major pathway (>75%) for O2 entry and exit is the short channel that opens by outward movement of the distal histidine side chain. The discrepancy between our mutagenesis map and that of Huang and Boxer (32) is not as great as it first appears. The largest changes in their random mutagenesis map and that of Huang and Boxer (32) is not as rational mutagenesis mapping strategy or the models and interactions used in the simulations cannot accurately define ligand movements into solvent.

In 2002, Pesce et al. (39) reported the structure of the mini-hemoglobin from the sea worm Cerebratulus lacteus (CerHb). CerHb is a neuronal O2 storage protein, which surrounds the axons and brain of the Nemertean worm, and it rapidly releases O2 when the worm burrows into sand and becomes anaerobic while searching for prey (40). The analysis of the CerHb crystal structure revealed the presence of an apolar tunnel leading from the distal pocket to solvent, with an exit point between the E and H helices created by the loss of the N-terminal A helix (Fig. 1B). This tunnel was suggested to be the major route of entry and exit for O2 binding, and later work indicated that this “open” channel accounts for the very large association rate constants for O2 and NO binding to CerHb (200–300 μM⁻¹ s⁻¹), which are close to the values expected for diffusion-controlled processes (41).

Our initial mutagenesis and Xe binding studies of CerHb supported the original idea of Pesce et al. (39) that the internal tunnel is the major ligand pathway (39, 42, 43). Thus, we have used CerHb as a counter example to SwMb where the histidine gate mechanism for O2 binding appears to apply. However, Orlowski and Nowak (44) carried out simulations with CerHb and, as with SwMb, obtained multiple trajectories with the apolar tunnel being only one of them.

To resolve this discrepancy and determine whether the experimental mutagenesis mapping strategy can identify internal pathways, we constructed and examined the O2, CO, and NO ligand binding properties of 120 different mutants of CerHb. Rates of ligand entry and exit for this library of CerHb variants were compared with results for an expanded library of SwMb mutants, which extends the results of Scott et al. (13) to regions near the ends of the E and H helices.

Our goals were: 1) to verify the mutagenesis mapping strategy; 2) to define experimentally the ligand pathways in the two proteins; 3) to compare the results to MD simulations; and 4) to provide benchmarks for future theoretical work.

**EXPERIMENTAL PROCEDURES**

**Hb and Mb Sample Preparations**—WT and mutant recombinant CerHbs were expressed and purified as described previously using a synthetic gene with codon usage optimized for expression in Escherichia coli (39, 41). The WT SwMb gene and expression plasmid (pMb413) were constructed by Springer and Sligar (45) to optimize codon usage in *E. coli*. All new mutants were constructed using this gene in a pUC19-based expression vector and the Stratagene PCR-based QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). Purification was achieved as described previously by Springer and Sligar (45, 46) and modified by Carver et al. (47). The WT control and the mutant Ms have an extra N-terminal Met residue and a D122N substitution for crystallization in the P6 space group, neither of which has an effect on ligand binding (48). All experiments to determine rate constants were carried out in 0.1 M phosphate buffer, pH 7.0, 1.0 mM EDTA, 20 °C.

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**

**Measurement of Rate Constants for Ligand Association (kΔ)**
SwMbCO and CerHbCO samples containing various [CO] under pseudo first order conditions as described previously (42). The association rate constants for NO binding to deoxy-SwMb and deoxy-CerHb were measured using a flow-flash multilinear apparatus and a 500-nm dye laser (42, 43, 49). In these experiments, SwMbCO or CerHbCO samples in deoxygenated buffer were mixed with various concentrations of NO and then photolyzed ~10–20 ms after mixing, allowing measurement of time courses for NO binding to the newly formed deoxy-SwMb or deoxy-CerHb without the limitations of the low quantum yields of CerHbNO and SwMbNO complexes. O$_2$ association and dissociation time courses were measured after complete laser photolysis of SwMbCO and CerHbCO samples in O$_2$/CO mixtures using a 500-nm dye laser as described previously (42, 49, 50). Experimental procedures and fitting routines were obtained for $k_{O2}$ and $k_{O2}$ ($k_{O2}$ indicates oxygen association rate constant; and $k_{O2}$ indicates oxygen dissociation rate constant) from analyses of the fast bimolecular binding and slow O$_2$ displacement phases have been described in previous work (42, 49, 50).

**Measurement of Geminate Recombination**—Time courses for internal geminate CO and O$_2$ recombination within CerHb and SwMb were measured at 436 nm after excitation with a 7-ns excitation pulse from a Lumonics YAG-laser system, using a Tektronix TDS3052 digitizing oscilloscope with nanosecond time resolution. Experimental procedures and fitting routines were followed as described previously (13, 42). Sample time courses are shown in Figs. 5–7. The results were analyzed and interpreted in terms of the previously published mechanisms for ligand binding to SwMb (13) and CerHb (42, 43), as described below.

**Analysis of Bimolecular and Geminate Rate Parameters**—In Scott et al. (13), bimolecular and geminate O$_2$ bindings to SwMb were analyzed in terms of the three-step, side-path model shown in Fig. 1A. The rate constants for intramolecular rebinding, escape, and movement through the protein between sites B and C were obtained by fitting sets of time courses to this model using numerical integration algorithms described by Scott and Gibson (12). The geminate rebinding time courses for WT SwMb and many of the mutants can also be fitted well by the sum of two exponential processes and an offset reflecting the amount of escape to solvent. In this analysis, the rate of the fast phase equals $k_{bond} + k_{bc}$ + $k_{escape}$, the total fraction of geminate recombination, $F_{gem,total}$ equals $k_{bond}/(k_{bond} + k_{escape})$, and the fraction of the slow phase, $F_{gem,slow}$ approximately equals $k_{bc}/(k_{bond} + k_{bc} + k_{escape})$, allowing an estimation of all three rate parameters (13). It should be noted that a simple two-step reaction analysis, using the correct $F_{gem,total}$ provides the same value of $k_{entry}$ and an only slightly larger $k_{escape}$ value.

The side-path mechanism was chosen for analysis of SwMb data because both Xe binding and mutations that fill the Xe1 and Xe4 cavities collapse the time courses to a single rapid phase with little or no change in the total fraction of geminate rebinding (12, 13). Similar results were obtained for geminate ligand rebinding to WT Scapharca inaequalvis Hbl, where again, Xe binding and Phe and Trp replacements in internal cavities caused little change in the total fraction of geminate recombination (51). In the case of HbA subunits, single geminate phases for CO rebinding are observed at room temperature on nanosecond time scales and indicate that there are no side path C states, and again, the E7 channel appears to be the path for ligand entry and escape (49, 52). Thus, the scheme shown in Fig. 1A appears to apply to most animal hemoglobins with a tertiary and distal pocket structure similar to that of SwMb.

The fraction of photodissociated ligands that geminately rebind after photolysis is defined solely by $k_{bond}/(k_{bond} + k_{escape})$ in the SwMb scheme in Fig. 1A. There is no dependence on the rates of interconversion of the C and B sites, which only affect the rates and amplitudes of the slower geminate phases. Thus, regardless of the complexity of the internal movements, the overall bimolecular rate constant for ligand binding to SwMb under physiological conditions ([C], [B], d[C]/d$t$, and d[B]/d$t = 0$) is given by (13) Equation 1,

$$k_X = k_{entry}F_{geminate}; F_{geminate} = \frac{k_{bond}}{k_{bond} + k_{escape}} \quad (Eq. 1)$$

where $X$ stands for O$_2$, CO, or NO. In the scheme shown in Fig. 1A, $k_{entry}$ represents the second order rate constant for formation of state B, starting with ligand in the solvent phase. Values of $k_{entry}$ for O$_2$ entry in SwMb were calculated from $k_{O2}/F_{geminate}$ and are listed in Table 3 for new mutants and in Scott et al. (13) for the remaining 90 variants.

It is important to note that the experimental values of $F_{gem}$ are model-independent and depend only on an accurate definition of the absorbance change for the total amount of geminate recombination. Thus, Equation 1 can also be used empirically to estimate the rate of entry into any globin, including CerHb, where in Fig. 1B binding involves a linear sequential set of ligand movements. As discussed by Scott et al. (13), the steady-state expression for $k_X$ using a linear scheme is still $k_X = k_{entry}$ $F_{gem}$, but in this case $k_{entry}$ represents the bimolecular rate of entry into the first discrete site in the protein, which, in the case of CerHb, would be the apolar tunnel. However, if the barriers along the tunnel and from the tunnel into the distal pocket are small compared with bond formation, the mechanism reduces to a two-step process with a single geminate intermediate. Under these conditions, $F_{gem}$ can again be approximated as $k_{bond}/(k_{bond} + k_{escape})$.

In the case of CO binding to CerHb, there is evidence of a barrier between the tunnel and the distal pocket implying a linear three-step scheme. However, this extra complexity is only evident at low temperatures (53) or at room temperature when restrictions in the tunnel are completely removed as is observed in the L86A mutant (43).

When movement into and through the tunnel is restricted by mutagenesis, the geminate time courses become single exponentials and are readily analyzed as a simple two-step scheme (Equation 2),

$$k_{gem} = k_{bond} + k_{escape}; F_{geminate} = \frac{k_{bond}}{k_{bond} + k_{escape}}$$

$$k_{bond} = k_{gem}F_{geminate}; k_{escape} = k_{gem}(1 - F_{geminate}) \quad (Eq. 2)$$
We used the expressions in Equations 1 and 2 to estimate values of $k_{\text{entry}}$, $k_{\text{escape}}$, and $k_{\text{bond}}$ for CO binding to all of the mutants of CerHb examined (Tables 1 and 4 and supplemental Tables S1–S4). These estimations, although empirical, are the simplest and closest to the observed experimental data, making the maps shown in Fig. 4 experimentally defined and not model-dependent.

**NO Association and the Validity of Using** $k_{\text{entry}}$ and $k_{\text{escape}}$ **—** In principle, the calculated values of $k_{\text{entry}}$ and $k_{\text{escape}}$ should be used to map the ligand pathways quantitatively because these parameters factor out the differences in reactivity ($k_{\text{bond}}$ values) between the O$_2$, CO, and NO. The values of these entry and exit parameters should be the same for all three diatomic gases because of their similar sizes and diffusion constants at room temperature. In the case of CO binding to WT CerHb, the relative errors in these calculated parameters are very large because the extent of CO geminate recombination is very small, $\approx 0.05$, and it has an absolute error of approximately $\pm 0.03$, which is roughly the same for all values of $F_{\text{gem}}$ measured. To reassure ourselves that the calculated $k_{\text{entry}}$ values are valid for the CerHb mutants, we examined the correlations between $k_{\text{entry}}$, $k_{\text{bond}}$, and $k_{\text{CO}}$ for some of the key variants (Table 1).

In contrast to CO, the fraction of internal NO recombination after photolysis is very large, $\approx 0.99$ due to its high reactivity with iron. Once inside the protein, NO reacts immediately with the iron atom, with almost every internal collision between them being productive. As a result, the rate-limiting step for NO binding from solvent is the bimolecular rate of ligand entry into the protein, and the observed association rate constant, $k_{\text{NO}}$, should equal the calculated value of $k_{\text{entry}}$ (Equation 1, $k_{\text{X}} \approx k_{\text{entry}}$ when $F_{\text{gem}} \to 1.0$).

As shown in Fig. 2A, there is a linear correlation between $k_{\text{entry}}$, calculated using $k_{\text{CO}}$ and $F_{\text{gem,CO}}$, and the experimentally measured value of $k_{\text{NO}}$ for a series of Val-7(B6), Gln-44(E7), Thr-48(E11), Ala-55(E18), and Leu-86(G12) CerHb mutants (Table 1). The error in $k_{\text{entry}}$ is very large for those mutants with large $k_{\text{CO}}$ and very small $F_{\text{gem,CO}}$ values, making it hard to determine the slope of the correlation between $k_{\text{entry}}$ and $k_{\text{CO}}$. In addition, there appears to be a systematic overestimation of $k_{\text{entry}}$ for the CerHb variants with $F_{\text{gem,CO}} < 0.1$ (Fig. 2A). However, the correlation is strong for mutants where $F_{\text{gem}}$ is larger due to partial blockage of the tunnel. Similar linear correlations between $k_{\text{entry}}$ and $k_{\text{NO}}$ were observed for all of the SwMb mutants reported in Scott et al. (13), verifying the $k_{\text{entry}}$ values for myoglobin.

As shown in Fig. 2B, there is a 1:1 correlation between $k_{\text{NO}}$ and $k_{\text{O}_2}$ for the CerHb variants in Table 1 implying that O$_2$ binding is also primarily limited by the bimolecular rate of entry into the protein. The only exceptions are the T48V mutant, where the loss of $\gamma$-OH at the E11 position releases the phenol side chain of Tyr(B10), allowing it to sterically inhibit iron-ligand bond formation, and the Q44F mutation, where the larger Phe(E7) side chain directly inhibits O$_2$ binding at the iron atom. For both these mutants, overall O$_2$ binding is slowed because bond formation is markedly inhibited. As a result, $k_{\text{bond}}$ becomes similar to $k_{\text{escape}}$ causing a decrease in $k_{\text{O}_2}$ due to a decrease in $F_{\text{gem}}$ and not the rate of entry (Equations 1 and 2). In the case of NO, its high reactivity with the iron atom is still much greater than that of ligand escape, keeping ligand entry as the rate-limiting step even for the T48V and Q44F CerHb mutants.

Fig. 2C shows the correspondence between the calculated ligand entry rate using overall and geminate CO binding data with the bimolecular rate constant for O$_2$ binding. In this case, there is an $\approx 1:1$ correspondence with all the mutants examined in this work, even though the scatter is fairly large. All of the results in Fig. 2 for CerHb and the correlations between $k_{\text{NO}}$ and $k_{\text{entry}}$ values for the SwMb mutants (13) support the use of calculated $k_{\text{entry}}$ and $k_{\text{escape}}$ values to map ligand pathways and demonstrate that similar conclusions would be obtained if $k_{\text{NO}}$ or $k_{\text{O}_2}$ values were used directly in the mapping analysis instead of the calculated $k_{\text{entry}}$ values.

**Crystalization, Data Collection, and Refinement of** H100F, H100W, and Y11F/Q44L/T48V/A55W CerHb Variants—The O$_2$, CO, and aquomet forms of H100F and H100W CerHb were crystalized using the sitting drop vapor diffusion method at a protein concentration of 25 mg/ml. The best crystals grew from solutions containing 2.3 M ammonium sulfate, 100 mM sodium citrate, pH 5.4, at 277 K. The aquomet form of the quadruple mutant with an apolar distal pocket was crystalized from solutions containing 200 mM sodium thiocyanate and 2.2 M ammonium sulfate. The O$_2$ and CO forms of this multiple mutant were obtained by reduction of the ferric crystals in the presence of either air or 1 atm CO. These manipulations often caused high mosaicsities and reversion to the ferric or completely deoxygenated reduced form. As a result, crystal manipulations and data collection were repeated several times.

In all cases crystals were transferred to appropriate storage solutions containing 2.8 M ammonium sulfate and 25% sucrose.
(w/v) immediately before data collection at 100 K. For CO complexes, the mounting solution was saturated with 1 atm of pure carbon monoxide.

A cryo-cooling N2 system was used to maintain low temperature (100 K) in the environment of the mounted crystals to reduce radiation damage. Complete x-ray diffraction data sets for each of the CerHb variants were obtained from single crystals using CuKα radiation (λ = 1.5418 Å) from a Rigaku RUH3R rotating anode x-ray generator operated at 50 kV and 90 mA and a Rigaku R-AXIS IV++ image plate detector (Rigaku Americas Co.). Data were collected, scaled, and reduced using d*TREK software (54).

The program PHENIX (55) was used for both structure solution and refinement. The structures of the H100F and H100W CerHb derivatives (PDB codes 4F6B, 4F6D 4F6F, 4F6G, 4F6I, and 4F6I) were solved by difference Fourier syntheses using the structure of wild-type CerHb (PDB code 1KR7) with the ligand and solvent molecules omitted. The structures of the Y11F/Q44L/T48V/A55W CerHb mutants (PDB codes 4F68 and 4F69) were also solved by difference Fourier syntheses using the structure of the A55W CerHb mutant (PDB code 2YY). An initial round of simulated annealing was used to calculate unbiased electron density maps to confirm the correct placement of the mutated residues. This process was followed by several refinement macrocycles with maximum likelihood as the target of refinement and included bulk solvent correction and anisotropic scaling of the data, individual coordinate refinement with minimization, and individual isotropic ADP refinement interspersed with manual fitting. The ligands were modeled into the electron density maps at this point, and refinement was continued with restraints imposed on the expected Fe2+CO, Fe2+O2, or Fe3+-H2O geometries. Solvent molecules were then included, and in a few cases, combined TLS and individual ADP refinements were carried out in the final stages of refinement.

Map fitting and other manipulations with molecular models were performed using the graphic software COOT (56). The accession codes for the models, crystal parameters, and statistics of x-ray data collection and refinement are provided in Table 2. Figs. 1, 4, 5, and 6 were prepared using the graphic software COOT (56). The structures of x-ray data collection and refinement are provided in Table 2. Figs. 1, 4, 5, and 6 were prepared using the graphic software COOT (56). The overall rate constants for O2 and CO binding were measured using rapid mixing, laser photolysis methods with O2/CO mixtures, and flow-flash techniques as described in Salter et al. (42) and Pesce et al. (43). Internal CO recombination within each mutant was also measured using a 7-nm YAG laser. The rate, kgem, and the fraction, Fgem, of geminate recombination and the overall bimolecular rate constant for CO binding, kCO, were used to estimate the first order rate constant for ligand escape (kescape = kgem,CO(1 − Fgem,CO)) following its photodissociation from the heme iron and the bimolecular rate of ligand re-entry from solvent (kentry = kCO/Fgem,CO) (42, 43).

In key cases, the calculated value of kentry was verified by measuring the bimolecular rate constant for NO association as described under “Experimental Procedures” (Table 1 and Fig. 2). Complete sets of O2 and CO kinetic parameters for all the CerHb variants examined are listed in the supplemental Tables S1–S4.

Underlying Assumptions—The first premise of our mutagenesis approach is that larger and smaller apolar amino acid side chains should impose larger and smaller kinetic barriers, respectively, for both ligand entry and escape if they are located along the primary pathway for ligand binding (13, 42). Thus, the overall bimolecular rate constant of ligand entry for a mutant with a small amino acid, kentry,small, should be significantly larger than the entry rate constant for a mutant with a large amino acid, kentry,large, at positions located along the migration pathway that leads to bond formation. Similar relationships should occur for the ligand escape rate constants for the same set of mutants. To quantify these effects, we computed the overall rate enhancement effect, Renhance, of a large-to-small mutation as shown in Equation 3 (13),

\[
R_{\text{enhance}} = \log\left(\frac{k_{\text{entry,small}}}{k_{\text{entry,large}}}\right) + \log\left(\frac{k_{\text{escape,small}}}{k_{\text{escape,large}}}\right)
\]  

(Eq. 3)

Large-to-small replacements at positions on the major ligand pathway will increase both kentry and kescape, which will lead to positive values for both logarithmic terms in Equation 3 and a large value of Renhance when added together. In contrast, mutations at positions distant from the ligand trajectory should have little or no effect on entry and escape.

The second premise of our mapping approach is that ligand binding and single point mutations have only minor effects on overall globin structure. Ligand binding to Mb does produce structural changes involving replacement of distal pocket water, small outward movements of the distal histidine, and alterations of the proximal pocket and minor tilting of the heme to accommodate in-plane movement of the iron atom. However, none of these changes have a significant effect on the overall tertiary structure and the internal Xe cavities observed in Fig. 1A (14, 19, 36, 37). Recently, Germani et al. (57) have shown that CO and O2 binding to deoxy-CerHb causes even smaller changes in the heme cavity and has no effect on the size or shape of the apolar tunnel.
### TABLE 2

Data collection and refinement statistics for CerHb mutants

<table>
<thead>
<tr>
<th></th>
<th>Y11F/Q44L/ T48V/A55W CerHbO₂</th>
<th>Y11F/Q44L/ T48V/A55W CerHbCO</th>
<th>H100F met-CerHb</th>
<th>H100F CerHbO₂</th>
<th>H100F CerHbCO</th>
<th>H100W met-CerHb</th>
<th>H100W CerHbO₂</th>
<th>H100W CerHbCO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PDB entry code</strong></td>
<td>4F68</td>
<td>4F69</td>
<td>4F6D</td>
<td>4F6F</td>
<td>4F6G</td>
<td>4F6I</td>
<td>4F6J</td>
<td></td>
</tr>
<tr>
<td><strong>Crystal data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range</td>
<td>24.66–1.80 Å</td>
<td>27.0–1.60 Å</td>
<td>24.55–1.65 Å</td>
<td>27.0–1.80 Å</td>
<td>27.04–1.56 Å</td>
<td>24.55–1.64 Å</td>
<td>27.0–1.56 Å</td>
<td>30.24–1.45 Å</td>
</tr>
<tr>
<td>Space group</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
<td>P 2,2,2</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td>a = 42.62, b = 43.55, c = 59.75 Å</td>
<td>a = 42.62, b = 43.55, c = 59.5 Å</td>
<td>a = 42.53, b = 43.47, c = 59.50 Å</td>
<td>a = 42.37, b = 43.55, c = 59.41 Å</td>
<td>a = 42.68, b = 43.48, c = 59.48 Å</td>
<td>a = 42.15, b = 43.40, c = 59.70 Å</td>
<td>a = 42.10, b = 43.51, c = 59.49</td>
<td></td>
</tr>
<tr>
<td><strong>Reflections</strong></td>
<td>25,517/10,643</td>
<td>38,297/14,641</td>
<td>45,856/13,828</td>
<td>23,325/10,039</td>
<td>47,165/14,607</td>
<td>46,250/13,789</td>
<td>37,156/12,733</td>
<td>39,047/16,823</td>
</tr>
<tr>
<td>(measured/unique)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completeness</td>
<td>98.8% (100%)*</td>
<td>97.0% (89.7%)</td>
<td>98.0% (98.1%)</td>
<td>93.7% (94.0%)</td>
<td>90.0% (62.3%)</td>
<td>98.0% (98.8%)</td>
<td>95.0% (68.6%)</td>
<td>83.8% (46.8%)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>2.4 (2.3)</td>
<td>2.6 (2.1)</td>
<td>3.3 (3.4)</td>
<td>4.9 (4.3)</td>
<td>4.6 (4.1)</td>
<td>4.9 (3.9)</td>
<td>2.4 (1.9)</td>
<td>2.3 (1.2)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.8% (17.7%)</td>
<td>4.8% (24.0%)</td>
<td>4.9% (11.4%)</td>
<td>4.6% (10.1%)</td>
<td>4.6% (16.0%)</td>
<td>4.9% (26.0%)</td>
<td>4.4% (33.0%)</td>
<td>4.0% (27.9%)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range</td>
<td>24.66–1.80 Å</td>
<td>27.0–1.60 Å</td>
<td>24.55–1.65 Å</td>
<td>27.0–1.80 Å</td>
<td>27.04–1.56 Å</td>
<td>24.55–1.64 Å</td>
<td>27.0–1.56 Å</td>
<td>30.24–1.45 Å</td>
</tr>
<tr>
<td>R-factor</td>
<td>20.3%</td>
<td>20.6%</td>
<td>18.0%</td>
<td>16.0%</td>
<td>19.3%</td>
<td>18.4%</td>
<td>19.8%</td>
<td>19.8%</td>
</tr>
<tr>
<td>Rfree</td>
<td>22.1%</td>
<td>23.3%</td>
<td>20.2%</td>
<td>19.7%</td>
<td>21.7%</td>
<td>20.6%</td>
<td>21.4%</td>
<td>21.9%</td>
</tr>
<tr>
<td><strong>Root mean square deviations from ideal values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond length</td>
<td>0.009 Å</td>
<td>0.008 Å</td>
<td>0.007 Å</td>
<td>0.025 Å</td>
<td>0.014 Å</td>
<td>0.007 Å</td>
<td>0.010 Å</td>
<td>0.010 Å</td>
</tr>
<tr>
<td>Bond angles</td>
<td>0.956°</td>
<td>0.973°</td>
<td>0.943°</td>
<td>0.922°</td>
<td>1.004°</td>
<td>0.966°</td>
<td>0.902°</td>
<td>0.900°</td>
</tr>
<tr>
<td>Residues in most favorable region</td>
<td>95.7%</td>
<td>96.8%</td>
<td>94.7%</td>
<td>94.7%</td>
<td>96.8%</td>
<td>94.7%</td>
<td>95.7%</td>
<td>94.7%</td>
</tr>
<tr>
<td>Residues in additional allowed region</td>
<td>4.3%</td>
<td>3.2%</td>
<td>5.3%</td>
<td>5.3%</td>
<td>4.3%</td>
<td>5.3%</td>
<td>4.3%</td>
<td>5.3%</td>
</tr>
</tbody>
</table>

* Parameters in parentheses are for the outer resolution shell.
FIGURE 1. Side views of WT CerHbO2 and SwMbO2. In both globins the molecules are oriented so that the planar heme group is viewed edge-on, its propionates are facing left and the "distal pocket" is above the heme. Portions of the protein matrix are removed so the interior view is clear. Xe atoms are depicted as teal spheres; key amino acids and the heme prosthetic groups are indicated by sticks with carbon in green, oxygen in red, and nitrogen in blue; the remainder of the protein structure is represented as gray spheres. A, upper panel, SwMbO2 with most of the E-helix removed (PDB code 2MGM) and Xe1 and Xe4 inserted into positions obtained from the metMb structure determined at high Xe pressures (PDB code 2W6W). Key distal pocket residues include Leu-29(B10), His-64(E7), Val-68(E11), and Phe-43(CD1). Additional mutants constructed in this study include mutations made at residues Arg-45(CD3), Leu-61(E4), Leu-69(E12), Ile-75(E18), and Leu-135(H12). Middle panel, mechanism for ligand binding to Mb based on Scott et al. (13, 38). Lower panel, cavity surfaces (gray mesh) in SwMbO2 (PDB code 2MGM), obtained using the surface function in PyMOL, are compared with the Xe atom positions found in metMb under high Xe pressures (PDB code 2W6W). As described in Pesce et al. (39), no apolar tunnels were found in SwMb using Surfinet (64). We also found no tunnel in SwMb using MOLE online 2.0 (65) using the default settings. The cavities found using Surfinet in SwMbO2 are not connected. Xe1–Xe4 atoms were inserted into the structure based on their positions in 2W6W. B, upper panel, WT CerHbO2 (PDB code 1KR7) with most of the E helix removed and bound Xe atoms inserted at their locations in L86A CerHbO2 equilibrated with Xe gas (PDB code 2XKH). Key distal pocket and apolar tunnel residues include Val-7(B6), Tyr-11(B10), Gln-44(E7), Thr-48(E11), Ala-55(E18), Ala-82(G8), Leu-86(G12), Leu-98(H9), and His-100(H11) in WT CerHbO2. Middle panel, mechanism for ligand binding to CerHb taken from Salter et al. (42) and Pesce et al. (43). Lower panel, ribbons drawing for WT CerHbO2 (1KR7) with an apolar tunnel (red mesh) computed using MOLE online 2.0 (65) with a probe radius of 3 and 1.25 Å interior threshold (default settings). Xe atoms were inserted based on their positions in PDB code 2XKH.

FIGURE 2. Correlation of calculated rates of ligand entry, $k_{entry}$, and $k_{O2}$ in CerHb. A, $k_{entry}$ versus $k_{NO}$ for Val-7(B6)Ala, -Phe, and -Trp; Thr-48(E11)Val; Ala-55(E18)Val, -Leu, -Phe, and -Trp; Leu-86(G12)Ala, -Phe, and -Trp; Glu-44(E7)Trp; and Val-7(B6)Phe/Ala-55(E18)Phe/Leu-84(G12)Phe CerHb mutants (supplemental Table S2). B, $k_{O2}$ versus $k_{NO}$, open circles, all single mutants are located in and along the apolar tunnel (supplemental Table S1); closed circles, multiple mutants are located in the apolar tunnel (supplemental Table S4).
Our previous work with distal pocket mutants of CerHb also showed very little change in overall structure (41), and crystallographic studies of the mutations at the 55(E18) (42) and 100(H11) positions (this report) near the solvent entrance to CerHb have little effect on overall globin structure or the tunnel. In contrast, Xe addition to WT CerHbO2 crystals causes loss of diffraction, and we could only determine the structure of Xe bound to L86A CerHb where the tunnel cavity near the internal G12 position is enlarged. Similarly, we have been unable to crystallize V7F, V7W, L86F, and L86W variants previously suggested by previous MD simulations (Fig. 3) and do not significantly alter the rates for entry and escape in SwMb were measured using O2 instead of CO, as described in Scott et al. (13). The iron atom in Mb is less reactive due to proximal constraints, and as a result, very little internal CO recombination is observed at room temperature in ordinary buffers. In contrast, WT Mb shows ~50% geminate O2 recombination, and the rate is slow enough ($k_{gem} \approx 10^{-20}$ $\mu s^{-1}$, $t_{1/2} \approx 30–60$ ns) to be measured readily using a 7-ns YAG laser excitation pulse. In the case of CerHb, O2 geminate recombination is too fast to measure accurately with WT and most mutant proteins ($k_{gem} \geq 130 \mu s^{-1}$, $t_{1/2} \leq 5$ ns) (42).

**Rate Enhancement Maps**—Bar graphs of the rate enhancements caused by large-to-small mutations, $R_{enhance}$, versus sequence position are shown in Fig. 3 for both SwMb and CerHb. In both cases, significant effects are observed at the active site positions where the amino acid side chains interact directly with bound ligands (see B10, CD1, E7 and E11 bars labeled with blue diamonds in Fig. 3). However, beyond those regions the differences between Mb and CerHb are dramatic.

In the case of Mb, large rate enhancements are only observed at or near the E7 channel or in the back of the distal pocket where ligands are initially captured before binding to the heme iron atom, i.e. positions 29(B10), 32(B13), 43(CD1), 46(CD4), 64(E7), and 68(E11). No significant enhancements of the rates of entry and escape are observed at positions around the Xe1 cavity (i.e. positions 89(F4), 99(FG5), 104(G5), and 138(H15)), in regions equivalent to the apolar tunnel in CerHb (i.e. between the E, G, and H helices), or at other remote locations suggested by previous MD simulations (Fig. 3A). In contrast, large rate enhancements for entry into and escape from CerHb are observed for replacements lining the apolar tunnel shown in Fig. 1A, including positions near the C-terminal ends of the E and H helices and at position Val-7(B6) where the distal pocket
connects to the apolar tunnel (bars labeled with green asterisks, Fig. 3).

The magnitudes of the $R_{\text{enhance}}$ parameters are mapped onto the three-dimensional crystal structures of WT SwMb and CerHb in Fig. 4 (13). The order of $R_{\text{enhance}}$ effects is red/orange/yellow spheres at native amino acid positions (see legend to Fig. 4). The white or gray sticks in Fig. 4 indicate positions where little or no effect is observed on $k_{\text{entry}}$ and $k_{\text{escape}}$ for large-to-small amino acid substitutions.

The results are quite clear for both proteins. Only positions near the E7 gate or in the distal heme pocket are highlighted in SwMb, and no rate enhancements are observed along alternative pathways, including the Xe pockets or other interior apolar regions (Fig. 4, A–C). In contrast, mutagenesis mapping clearly highlights the internal tunnel in CerHb (Fig. 4, D–F), proving experimentally that ligands enter and exit through this pathway. This conclusion is particularly evident in Fig. 4F where the tunnel between the E and H helices is clearly seen and circumscribed by red/orange/yellow side chains.

Identifying the Solvent Exit of the Apolar Tunnel—Three further tests of the apolar pathway in CerHb were performed. First, we explored the solvent-exposed exit aperture of the apolar tunnel at the ends of the E and H helices; second, we examined the influence of polar interactions at the E7 gate; and third, we filled the apolar channel with aromatic amino acid side chains to completely block this pathway.

In WT CerHbO$_2$, the polar imidazole side chain of His-100 at the H11 helical position is extended out into solvent and does not appear to block entry into the protein (Figs. 1B and 5B) (42). As expected, removing the imidazole at this position by a H100A mutation has little effect on ligand binding; however, both Phe and Trp mutations at this position cause ~3-fold decreases in both $k_{\text{O}_2}$ and $k_{\text{O}_2}$, with little change in affinity (Table 4). These same aromatic amino acid mutations also cause large increases in the fraction of CO geminate recombination, from ~0.05 for WT CerHbCO to ~0.34 for H100W (Fig. 5A), and a marked decrease in $k_{\text{gem,CO}}$ due to an ~10-fold decrease in $k_{\text{escape}}$ (Table 4). All of these effects suggest that large Phe and Trp side chains at the H11 helical position in CerHb rotate back to the protein surface to partially cover the solvent exit of the ligand tunnel and restrict ligand escape.

The imidazole ring of the naturally occurring His-100 side chain is poorly defined in the WT CerHbO$_2$ structure, with only a portion of the expected electron density of the ring apparent in the calculated map (Fig. 5B). In the structures containing the A55W replacement, the His-100 chain is better defined but in a different orientation due the large indole ring at the E18 position, which restricts the motion at the H11 helical position (i.e., see Fig. 6C). Fig. 5C shows the conformation of His-100 in seven different CerHb structures in both reduced and oxidized forms. The side chain appears to show free rotation about the C-H bond with no specific energy minimum, presumably because of equally favorable interactions with solvent water molecules. As a result, the entrance into the apolar channel is kept open. When Phe or Trp is inserted into this position, the more apolar aromatic rings appear to be forced into the entrance of the channel by a hydrophobic effect and restrict ligand entry and escape (Table 4 and Fig. 5A).

To verify this interpretation, we crystallized and determined structures of the H100F and H100W CerHb mutants in

![FIGURE 3. Effects of Ala or Val to Phe or Trp mutations on entry and escape in CerHb and SwMb. Blue diamonds denote helical positions typically associated with the E7 gate pathway. Green asterisks are associated with positions along the apolar tunnel pathway. Data for SwMb were taken from this study and Ref. 13. The enhancement factor, $R_{\text{enhance}}$, on the y axis is defined in Equation 3, and the topological helical position is given along the x axis.](https://www.jbc.org/content/287/40/33171/F1.large.jpg)
reduced O₂- and CO-bound forms and in the oxidized state (Table 2). In both cases, the aromatic side chains rotated back onto the protein surface, partially entering and blocking the tunnel from the H11 helical position (Fig. 5D). The Trp-100 and Phe-100 side chain conformations have 100% occupancy, are well defined by electron density (Fig. 5D), and account for the decreases in rates of ligand entry and exit and the increase in geminate CO recombination observed for H100F and H100W mutants (Table 4 and Fig. 5A).

In WT CerHb, the small residue Ala-55 at E18 is located on the opposite side of the entrance to the apolar channel from His-100 at H11. Substitution of a Trp for Ala at this position also blocks ligand entry, decreases $k_{\text{entry}}$ and $k_{\text{escape}}$ and traps photodissociated ligands inside the protein, increasing $F_{\text{gem}}$ (Table 4 and Figs. 3B and 6C) (42). Similar large decreases in $k_{\text{entry}}$ and $k_{\text{escape}}$ and increases in $F_{\text{gem}}$ occur when Ala-101 at H12 is replaced with Trp (Fig. 3B and supplemental Table S1). Combined, these mutagenesis results unambiguously locate the entrance for ligand movement into CerHb between the ends of the E and H helices at positions E18, H11, and H12.

**Constructing an Apolar Distal Pocket**—Pesce et al. (39, 43) suggested that extensive polar interactions between Tyr-11(B10), Gln-44(E7), Lys-47(E10), and Thr-48(E11) limit opening of the E7 gate, causing ligands to use the apolar tunnel for entry and exit. To test this idea, we examined ligand binding to a triple CerHb mutant with a completely apolar active site co-
FIGURE 5. His-100 is at the entrance to the apolar tunnel in CerHb. A, time courses for geminate recombination in His-100(H11) mutants of CerHbCO at 20 °C, pH 7.0. The fraction of geminate recombination increases 6-fold upon mutating the polar His to large aromatic amino acids. B, 2Fo – Fc electron density maps for Ala-55(E18) cutoff at 1.5σ/Hb, and His-100(H11) cutoff at 0.7σ/Hb, in WT CerHbO2 (PDB code 1KR7). The polar imidazole side chain of His-100 at the H11 helical position is poorly defined, cannot be seen at the higher contour levels, and appears to extend out and interact with solvent molecules, keeping the channel open. C, positions of the His-100(H11) side chain in the structures for the O2 and aquomet complexes of WT CerHb (PDB codes 1KR7 and 2XKI), L86A CerHbO2/H11006 Xe (PDB codes 2XKG and 2XKH), T48V CerHbO2 (PDB code 1V07), A55W CerHbO2 (PDB code 2VYY), and A55F CerHbO2 (PDB code 2VYZ). D, 2Fo – Fc electron density map for the Trp-100(H11) cutoff at 1.5σ/Hb in H100W CerHbO2 (PDB code 4F6I). The indole side chain is well defined.
taining Y11F/Q44L/T48V mutations at the B10, E7, and E11 helical positions, respectively. As shown in Table 4, rates of O₂ binding to and dissociation from this triple mutant increase by only 50%, and there is little change in the computed values of $k_{\text{entry}}$ and $k_{\text{escape}}$ compared with WT CerHb. The extent of geminate recombination does increase to 0.40 (Fig. 6A) due to a marked 7-fold increase in the rate of covalent bond formation ($k_{\text{bond}}$) compared with WT CerHb (from ~4.0 to ~30 μs⁻¹ for the mutant, see supplemental Table S4). However, when the A55W mutation is added to this triple mutant to block exit from the apolar tunnel (Fig. 6C), there are marked ~5-fold decreases in $k_{\text{O2}}$ and $k_{\text{O2}}$, the fraction of geminate CO rebinding increases to 0.85, and the calculated value of $k_{\text{entry}}$ decreases from ~400 to ~30 μs⁻¹ s⁻¹ (Table 4 and Fig. 6A). Thus, even when the distal pocket is made apolar, ligands still appear to enter and exit the protein through the apolar tunnel, and blocking the entrance to the tunnel markedly inhibits bimolecular binding.

We determined the crystal structure of the Y11F/Q44L/T48V/A55W mutant to examine whether the apolar substitutions cause significant changes in the tertiary structure of CerHb. Electron density maps for the active site amino acids of the quadruple mutant and Trp-55 at the tunnel entrance are shown in Fig. 6B and C, respectively. A comparison of the active site of the quadruple mutant to that of the single A55W mutant is shown in Fig. 6D. The structures overlay remarkably well with a root mean square deviation of 0.17 Å for the Cα atoms, and the side chains of the B10, E7, and E11 amino acids have very similar conformations despite the changes in polarity (note that the root mean square deviation was 0.15 Å for the Cα atoms of WT CerHb and the quadruple mutant). The major difference is that there is more “free” space surrounding the bound O₂ atoms due to the loss of the O atom of Tyr-11(B10) and the N or O atoms of Gln-44(E7). This loss of steric hindrance adjacent to the bound ligand accounts for the increase in the rate and fraction of geminate recombination after laser photolysis (Fig. 6A and Table 4). Instead of being pushed away by the larger Tyr(B10) and Gln(E7) side chains, photodissociated CO can remain near the center of the heme group and more rapidly rebinds than in WT CerHb. However, the large, ≥10-
Mapping Ligand Pathways in Globins

FIGURE 7. Time courses for geminate recombination in multiple tunnel mutants of CerHb at 20 °C, pH 7.0. A, CO geminate rebinding in position Val-7(B6) mutants of CerHb. B, CO geminate rebinding in single, double, and triple Phe mutants in positions Val-7(B6), Ala-55(E18), and Leu-86(G12) mutants of CerHbCO. C, CO geminate rebinding in single, double, and triple Trp mutants in positions Val-7(B6), Ala-55(E18), and Leu-86(G12) mutants of CerHbCO.

fold decrease in the rate of entry caused by the A55W mutation at E18, even when the active site is apolar, argues strongly that most ligands (≥90%) still enter CerHb through the apolar channel ($k_{entry}$ values in Table 4).

Completely Blocking the Tunnel—As a final test of the map for CerHb, we attempted to completely block ligand movement through the apolar tunnel. To prove that entry into the distal pocket from the tunnel is at the B6 position, we mutated Val-7(B6) to Ala, Phe, and Trp. As shown in Fig. 7, introducing an indole ring at this position traps CO inside the distal pocket increasing both $F_{gem}$ and $k_{gem}$ dramatically. At the same time, the V7W replacement markedly decreases $k_{entry}$, $k_{O2}$, and $k_{NO}$ and the rate of O$_2$ dissociation (Table 4), demonstrating that this position is on the pathway for ligand escape. This conclusion is verified by the large increase in $F_{gem}$ and a decrease in $k_{entry}$ when the V7W mutation is added to the apolar pocket (Y11F/Q44L/T48V) CerHb mutant (Fig. 6A and Table 4).

The V7A/L86A CerHb double mutant was constructed as a control for a completely open channel with Ala at the three key positions along the apolar tunnel, 7(B6), 55(E18), and 86(G12), respectively (WT CerHb has Ala at position 55(E18)). This double mutant has kinetic parameters very similar to those of WT CerHb (Table 4). In contrast, the V7F/A55F/L86F and V7W/A55W/L86W triple mutants show dramatic 10-fold decreases in $k_{O2}$, $k_{NO}$, and $k_{entry}$ and ≥10-fold increases in $F_{gem,CO}$ from ~0.05 for WT CerHb to 0.80 and 0.93 for the triple Phe and Trp mutants, respectively (Table 4 and Fig. 7). The computed $R_{enhance}$ value comparing Ala-7/Ala-55/Ala-86 versus Trp-7/Trp-55/Trp-86 is 3.1, indicating a 3-fold increase in the free energy barrier to ligand escape and entry. These data show unambiguously that the apolar tunnel in CerHb can be closed by mutagenesis, reducing the measured bimolecular rate constants for O$_2$ and NO binding and the calculated rate for ligand entry to less than 10% of the WT values.

DISCUSSION

Single versus Multiple Pathways—The experimental mutagenesis results in Figs. 2–7, Table 4, and supplemental Tables S1–S4 are very clear. Ligands enter and exit CerHb through the apolar tunnel as originally suggested by Pesce et al. (39). When three Phe or Trp replacements at positions 7(B6), 55(E18), and 86(G12) are used to fill the entire tunnel, the bimolecular rate of ligand entry decreases dramatically. An estimate of the rate of ligand entry when the apolar channel is closed was calculated from the average of the $k_{entry}$, $k_{O2}$, and $k_{NO}$ values for the V7F/A55F/L86F and V7W/A55W/L86W mutants (Table 4) and is 25 ± 12 μM$^{-1}$ s$^{-1}$. An estimate of the value for ligand entry into CerHb with an open tunnel was computed from the average of the $k_{entry}$, $k_{O2}$, and $k_{NO}$ values for WT CerHb and the V7A, L86A, and V7A/L86A mutants (Table 4) and is equal to 300 ± 130 μM$^{-1}$ s$^{-1}$.

If the triple aromatic mutations at the B6, E18, and G12 helical positions completely close the tunnel, the residual bimolecular rate of 25 μM$^{-1}$ s$^{-1}$ applies to ligand binding through alternative routes. This value suggests that in WT CerHb ≥92% of the ligand molecules enter through the apolar tunnel and only 8% by alternative routes. Interestingly, when the tunnel is blocked, the value of $k_{entry}$ is only slightly smaller than the estimated rates of ligand entry into WT (His-64(E7)) and Gln-64(E7) SwMb, which are 34 and ~60 μM$^{-1}$ s$^{-1}$, respectively (13). Thus, the limiting value of 25 μM$^{-1}$ s$^{-1}$ for the triple aromatic CerHb mutants could represent ligand movement through the Gln(E7) gate, although this conclusion is speculative.

In the case of SwMb, all the positions that regulate entry and escape are located in the distal pocket or at the heme propionate-solvent interface (Figs. 3 and 4, A–C). As pointed out by Scott et al. (13) and Olson et al. (38), when the effects of Xe binding and selected mutations are examined individually, the arguments in favor of ligand movement through the E7 gate are even stronger. There are no changes in the fraction of geminate O$_2$ recombination, the calculated bimolecular rate constants for ligand entry, or the observed association rate constants for O$_2$ and CO binding to WT SwMb when the protein solution is pressurized with 5–10 atm of Xe. Similarly, when Trp substitutions are made in the Xe1 and Xe4 sites, there are small or no changes in $k_{entry}$, $k_{escape}$, and $F_{gem}$ (13, 38).

The E7 gate pathway in SwMb is also supported strongly and independently by time-resolved x-ray crystallography mea-
Mapping Ligand Pathways in Globins

measurements with mutants at the 29(B10) position. The lifetimes of CO in the Xε1 pockets of WT, L29F, and L29W SwMb are 2–10 μs (19), 100–300 μs (14, 18), and 1–2 ms (16), respectively, and the double mutant L29Y/H64Q SwMb shows intermediate behavior (15). If ligands moved directly to solvent from the Xε1 site, the size of the amino acid at the B10 position should have had no effect on the lifetime for ligand escape from this position. Instead, ligands appear to move back to the distal pocket, pass between the B10 side chain and the plane of the heme, and then exit through the E7 channel.

When His-64(E7) is replaced with Trp in SwMb, the calculated bimolecular rate of entry decreases from 34 to 7 μM⁻¹ s⁻¹, implying that up to 25% of the ligand molecules may take alternative routes to the active site (13). However, this calculation assumes that the H64W substitution completely blocks the E7 channel. Recently, Birukou et al. (52) have shown that the indole side chain in the crystal structure of H64W deoxy-Mb does not enter the E7 channel but just partially blocks the entrance, occupying a position between the heme propionates. A similar “blocked” conformation is observed in the crystal structure of Trp-63(E7) β human HbCO subunits (52). In contrast, the E7 indole side chain in the crystal structure of Trp-58(E7) deoxy-α human Hb subunits is located in the distal pocket, completely closing the E7 channel and sterically restricting access to the heme iron (52). In this completely “closed” conformation, the Trp(E7) side chain causes 100-fold decreases in both kNO and the calculated value of kentry for isolated mutant α subunits (49, 52). Thus, it is possible that the percentage of ligands entering through the E7 channel in Mb is significantly larger than the 75% value estimated from the rate of entry for Trp-64(E7) MbO₂.

Discrepancies with MD Simulations—The mutagenesis mapping results in Figs. 3 and 4 are not in agreement with either past or recent molecular dynamics simulations, all of which suggest multiple pathways. The latest theoretical work for both CerHb and SwMb suggests 3–5 pathways with roughly equal probabilities, one of which is the E7 channel in the case of SwMb and the apolar tunnel in CerHb. Thus, the theoretical calculations predict that 20–33% of the ligands enter and exit through the paths identified experimentally by mutagenesis, implying that alternative routes show similar rates of entry. In contrast, the experimental data indicate that ligand movements through the E7 gate in SwMb and the apolar channel in CerHb are 4–10-fold faster than through the alternative routes, based on the much smaller measured bimolecular rate constants when the gate or tunnel is blocked by Trp side chains. However, these differences are relatively small on a free energy of activation scale. Thus, small inaccuracies in representing chemical and structural barriers along the individual pathways could easily affect the computed results.

One key problem is representing the barrier to apolar ligand movement through clathrate-like water clusters that cover apolar patches at the protein surface. For example, well defined crystallographic water molecules surround the highly conserved Leu-89(F4) side chain in SwMb, which is apolar and protects the iron-His-93(F8) bond from being hydrated (33). We mutated this amino acid to Gly to open a hole into the Xε1 cavity and observed that a string of 3 to 4 water molecules entered this site and were connected to well defined surface waters, forming an extended chain of “ice-like” structures (33).

The rates of bimolecular ligand binding and unimolecular release were relatively unaffected by this mutation, and there was little or no effect on the total fraction of O₂ geminate recombination, even though the Xε1 cavity was completely open to solvent. The only major change was the loss of the slow phase for geminate rebinding, which in WT SwMb reflects movement into and then back from the Xε1 cavity (13). These structural and computational results demonstrate that water enters into the exposed Xε1 cavity in the Gly-89(F4) SwMb mutant and forms a barrier that prevents diatomic apolar ligands from entering this site.

Another example of how the hydrophobic effect can influence ligand binding is the effect of the His-100(H11) to Phe and Trp mutations on ligand binding to CerHb (Fig. 5). In most of the published structures of WT and mutant CerHb, the imidazole side chain of His-100 is pointing directly out into the solvent (see Figs. 1B and 5, B and C). However, when this polar amino acid is replaced by Phe and Trp, the hydrophobic effect pushes the more apolar phenyl and indole rings back on the surface of the protein, which partially blocks the tunnel entrance, markedly increases fgem,CO by trapping ligands within the CerHb tunnel, and decreases both kentry and kescape (Fig. 5 and Table 4).

A key test of the accuracy of theoretical calculations would be to examine the effects of the L89G SwMb and H100W CerHb mutations on simulated structures, starting from the WT proteins, and on ligand trajectories, carefully looking at water structure in and around the Mb Xε1 cavity near the F4 helical position and the exit aperture of the CerHb tunnel near the H11 helical position. Other valuable tests would be simulations of the effects of the F46V SwMb mutation at the CD4 helical position on movement of the His-64(E7) side chain and of the motions of the indole side chain in Trp-64(E7) Mb. The F46V mutation accelerates ligand entry and release in SwMb and causes outward movement of the distal histidine (34). The H64W mutation slows ligand entry ~4-fold, even though only the entrance to the E7 channel is blocked (13, 52). The effects of all four of these mutations on ligand entry rates are well defined experimentally and should be replicated in any molecular dynamics simulation.

Physiological Relevance—Ironically, the fastest rates of ligand uptake and release are observed when the longer apolar tunnel in CerHb is used rather than the shorter E7 gate channel in Mb. The rates of ligand binding to CerHb are close to diffusion-controlled and very similar to model heme compounds with no distal steric hindrance or to SwMb and human Hb mutants in which the distal histidine is replaced by Gly or Ala. Thus, the Nemertean mini-Hb from C. lacteus is able to take up and release O₂ roughly 10 times more rapidly than mammalian Mbs and Hbs. CerHb surrounds the sea worm’s brain and axons and acts as an O₂ storage protein. High rates of O₂ release give the worm an advantage for high neuronal activity under hypoxic conditions in the sea floor when it hunts for prey (40).

Apolar tunnels have also evolved in the truncated HbN from Mycobacterium tuberculosis (MtTrHbN) to facilitate rapid dioxygenation of NO as part of the bacterium’s defense against...
host macrophages (58–63). In this case, there appear to be two channels that allow rapid access to the heme iron atom, a short one between the G and H helices and a longer one leading from the solvent interface at the AB and GH loops into the distal pocket (59). The rate constants for bimolecular CO and O$_2$ binding to MtTrHBhN become similar to or greater than those for CerHb when steric restrictions due to distal pocket water are removed by apolar distal pocket mutations, and the rate of NO dioxygenation of wild-type MtTrHBhN-O$_2$ is $\approx 750 \text{ M}^{-1} \text{s}^{-1}$, which is $\approx 10$-fold greater than that for either human HbO$_2$ or mammalian MBO$_2$ (61). Both results indicate that, like in CerHb, ligands can move freely into the active site of MtTrHBhN. Simulations have suggested that Phe-62(E15) closes the shorter tunnel when O$_2$ is bound, but NO can still rapidly enter the protein through the longer pathway (60). However, these ideas are controversial (62, 63) but could be verified or refuted by mutagenesis mapping approaches similar to those shown in Figs. 3 and 4.

**Conclusions**—The experimental mutagenesis mapping results shown in Figs. 3 and 4 are unambiguous. Ligands enter and exit CerHb through the interior apolar tunnel at least $\approx 90\%$ of the time, whereas ligands enter SwMb from the opposite direction $\approx 75\%$ of the time through the E7 gate. The significance of these results is 3-fold. 1) Our rational mutagenesis mapping strategy is robust and can identify interior apolar pathways. 2) Globins have evolved different routes for ligand entry, and the longer but more open apolar pathway allows more rapid rates of entry and exit. 3) The SwMb and CerHb results provide experimental benchmarks for testing the validity of molecular dynamics trajectories, one with a short pathway between the heme propionates that is gated by a polar residue and another with an internal route from the heme pocket to an opening at the opposite end of the protein.

**Acknowledgments**—We thank Loren Stagg, Nachiyl Martinez Ramón, Benjamin Shields, and Cambria Alpha from the Biochemistry and Cell Biology Department at Rice University for helping to construct several of the initial CerHb mutants as part of a summer research project, and Alyssa Kuscel from the Kenneth Olsen laboratory at Loyola University for helping to construct several of the new Mb mutants as part of a summer research project.

**REFERENCES**

Mapping Ligand Pathways in Globins