Conservation and diversity of ancient hemoglobins in Bacteria

JJ L. Miranda *,1
Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

Received 24 February 2006
Available online 20 March 2006

Abstract

A group of single-domain proteins in Bacteria similar to thermoglobin, an oxygen-avid hemoglobin representative of the ancestral form, reveals the primordial structure, function, and evolvability of the family. Conserved residues at specific positions function to bind ligand or participate in hydrophobic packing of the protein core during protein folding. A potential hydrogen bond network consisting of a tyrosine and glutamine residue in the distal ligand-binding site of most hemoglobins suggests that the ancestral protein bound oxygen avidly. Two divergent hemoglobins with mutations at generally conserved positions contain non-canonical ligand-binding sites, illustrating plasticity of the fold. One binds heme in a manner similar to cytochromes and may represent an evolutionary link to the precursor of the hemoglobin fold. Conservation suggests specific biochemical properties of the ancestral protein; diversity suggests an evolvability of this group of hemoglobins tolerant of mutations that perturb conserved biochemical properties for adaptation to novel functions.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Hemoglobin; Myoglobin; Globin; Evolution; Bacteria

Heme proteins of the hemoglobin (Hb), also referred to as the myoglobin or globin, family perform a myriad of functions in different forms of life, but structural conservation suggests divergence from a single ancestral protein. Although representatives of the various Hb classes are present in all three domains of life, Bacteria, Archaea, and Eucarya [1], basal placement of Bacteria in the phylogenetic tree [2,3] makes organisms of this domain the most likely to harbor proteins similar to the ancestral Hb. Many types of Hbs are found in Bacteria, including fusions to flavin-binding and sensor proteins, but all of these classes probably evolved from the same Hb ancestor [1]. Three distinct types of unfused Hbs, truncated Hbs with N-terminal and internal deletions, protoglobins with an additional N-terminal helix, and single-domain Hbs (SDHbs) with termini similar to that of eukaryotic Hbs, can be found in Bacteria. Truncated Hbs fold into a 2/2 helical arrangement that distinguishes them from the classical 3/3 helical arrangement of SDHbs and eukaryotic Hbs, suggesting the possibility of unique ancestry followed by convergent evolution [4]. Protoglobins have been suggested to represent the ancestral 3/3 Hb form based on distribution in both Bacteria and Archaea, implying presence prior to divergence of the two domains [5]. SDHbs have been suggested to represent the ancestral form of all Hbs based on the presence of one such protein, *Aquifex aeolicus* thermoglobin (AaTgb), in a hyperthermophile placed near the root of bacterial evolution by phylogenetic methods [6].

Conserved properties among proteins similar to AaTgb may reflect the primordial properties of the Hb family. A computational search identified 44 SDHbs in Bacteria (BacHbs) with sequence conservation among biochemically functional residues. BacHbs are distinct from other SDHbs found in Eucarya. Two BacHbs are divergent and contain mutations at positions otherwise conserved. Both of these BacHbs absorbed significant perturbation at biochemically functional sites that may yield different functions for the Hb fold. Structural analysis of conservation characterizes the ligand-binding and biochemical properties of the ancestral Hb before the evolution and adaptation that yielded extant Hbs.
Similar examination of diversity illustrates the capability of the ancestral Hb to perform such evolution and adaptation.

Materials and methods

Reiterative PSI-BLAST searches [7] of the nr database (http://www.ncbi.nlm.nih.gov/BLAST/) began with AaTgb [6] as the query sequence. The limits "1:200[slen] AND Bacteria[orgn]" restricted results to sequences 200 amino acids or shorter from Bacteria. In cases where two Hbs are identified in the same strain, the shorter and longer of the two sequences were denoted as x and l, respectively. Multiple sequence alignments were performed with ClustalX [8] version 1.83 and manually adjusted. Revised gene annotation for the initiation sites of seven BacHb opened reading frames significantly longer or shorter than the rest was performed by examination of the coding and upstream sequences found in the genome of each organism.

Results and discussion

PSI-BLAST searches using AaTgb as the query sequence in an attempt to find BacHbs converged after four iterations and identified a group of 44 proteins from a diverse set of 37 strains of 32 species. Neither truncated Hbs nor protoglobins emerged from the computational survey. In preliminary searches, three outlier proteins were identified and excluded from the alignments that generated a profile matrix during the final search. Synthetic constructs for recombinant bacterial expression of the Pseudomonas aeruginosa UCBPP-PA14 BacHb and Homo sapiens cytoglobin were ignored. The 80 amino acid protein from Crocosphaera watsonii WH 8501 represents only part of a functional Hb and is likely a pseudogene. All final hits (Fig. 1) retained E values of at least 1 × 10⁻¹⁹, with no other proteins meeting the 0.005 exclusion threshold. Forty-two of the 44 BacHbs are quite similar and will be discussed independently of two divergent BacHbs (DBacHbs) with mutations at positions conserved in the other proteins.

Among BacHbs, sequence conservation exists for buried residues at the F8 and CD1 positions of the heme-binding site as well as the H7 and H8 positions in a cluster formed at an early stage of protein folding. All BacHbs contain the F8 histidine that coordinates iron in the proximal heme site and a CD1 phenylalanine that participates in van der Waals interactions with heme. Both residues are conserved in all Hbs, except for substitution at the CD1 position in some truncated Hbs [4]. The H7 alanine and H8 phenylalanine, residues that play no direct role in ligand binding, are also both conserved in all BacHbs. The H8 residue is part of a conserved cluster found at the interface of the A, G, and H helices, the three of which are the first to form in the folding of Mb, and conservation at these positions suggests a role in the folding pathway [9]. The H7 alanine is not directly implicated in protein folding, but may influence the pathway due to proximity to the conserved cluster. These four positions denote basic properties of the ancestral Hb, histidine-coordinated heme binding, and evolutionarily conserved protein folding. Conservation at the B10 and E7 positions in BacHbs suggests that O₂ avidity may be an ancestral property of Hbs. The B10 tyrosine and E7 glutamine are conserved in 40 and 36 of 42 BacHbs, respectively. E7 substitutions are only found in organisms where two BacHbs are present, suggesting that if a bacterium requires a BacHb, one copy must possess both the B10 tyrosine and E7 glutamine. In the structure of oxygenated Hb from the nematode Ascaris suum, the two residues each form hydrogen bonds with bound O₂ [10], resulting in high ligand affinity. Such hydrogen bond formation has not been observed in structures of the BacHb from Vitreoscilla sp. Cl [11,12], which contains both a B10 tyrosine and E7 glutamine, but none of the structures determined contain O₂ as a ligand. Mutagenesis experiments suggest that both residues are major determinants of the high O₂ affinity observed with A. suum Hb [13,14] and that the B10 tyrosine is significantly responsible for the O₂ avidity of AaTgb [6]. The BacHbs from Clostridium perfringens NCIMB8875 and Campylobacter jejuni NCTC11168 bind O₂ with lower affinity than AaTgb even though both the B10 tyrosine and E7 glutamine are present [15]. Other residues in the distal pocket, however, can mediate the function of these two sidechains, as seen with the E11 threonine in the Hb of the nematode Cerebratulus lacteus [16]. Lack of O₂ avidity in C. perfringens NCIMB8875 and C. jejuni NCTC11168 BacHb may therefore represent loss of the ancestral property due to acquisition of other mutations. Conservation of the B10 tyrosine and E7 glutamine in BacHbs, in addition to the high O₂ affinity of AaTgb, suggests the contention that the ancestral Hb was oxygen-avid even though most extant Hbs have lost this property.

Leptospira interrogans serovar Lai str. 56601 and Jannaschia sp. CCS1 DBacHbs add diversity to the group of otherwise conserved proteins through mutations at key functional sites. The F8 histidine of the L. interrogans serovar Lai str. 56601 DBacHb is replaced with cysteine. Human and equine myoglobin are structurally tolerant to such a mutation as the introduced cysteine sidechain can ligate heme [17,18]. Perhaps cysteine ligation of iron yields a binding site similar to that of cytochromes. The conserved B10 tyrosine is replaced by phenylalanine; the CD1 phenylalanine is retained. Although the function of the ligand-binding site is not easily predicted, significant differences from that of the canonical site can be expected.

The Jannaschia sp. CCS1 DBacHb contains leucine and alanine at the F8 and CD1 positions, respectively, suggesting that heme binding does not even occur. Indeed, extant proteins that adopt the Hb fold but do not bind heme perform a variety of functions such as antibiotic binding [19] and signal transduction [20]. Substitutions are also observed at the H7 position for both DBacHbs and the H8 position for the Jannaschia sp. CCS1 DBacHb. Tolerance of all these mutations in DBacHbs significantly alters biochemical properties and represents the plasticity and evolvability necessary for adaptation of the ancestral fold to novel functions.
Sequence conservation and diversity in BacHbs and DBacHbs highlight biochemical properties and evolvability of the ancestral Hb. Characterization of AaTgb in a bacterial hyperthermophile suggested that the Hb family is of ancient origin and descended from a monomeric, stable, pentacoordinate, and oxygen-avid Hb found in the last universal common ancestor of extant life [6]. Identification of BacHbs similar to AaTgb adds further support and details.
to this hypothesis. Conservation of the B10 tyrosine and E7 glutamine in the distal ligand-binding site suggests that the ancestral Hb was indeed oxygen-avid like AaTgb; ancient Hbs also contain the H7 alanine and H8 tryptophane in the hydrophobic core. The ancestral Hb may have evolved from a heme-binding protein similar to cytochromes [21]. Whether the L. interrogans serovar Lai str. 56601 DBacHb evolved from or into the ancestral Hb is not known as sequence similarity only suggests an evolutionary link of unknown directionality. The diversity of DBacHbs represents the capability of BacHbs to tolerate changes in structure and biochemical function regardless of a conserved ancient function. The ancestral Hb possessed O2 avidity and a specific hydrophobic cluster involved in protein folding, but the fold was sufficiently evolvable to adapt to different conditions and yield the myriad of novel structures and functions of extant Hbs.

Acknowledgments

I thank Stephen C. Harrison for support throughout the course of this project. This work was supported by a National Science Foundation Graduate Research Fellowship.

References


