Bacterial flavohaemoglobins: a consensus sequence and identification of a discrete enterobacterial group and of further bacterial globins

Jorge Membrillo-Hernández, Robert K. Poole *

The Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, The University of Sheffield, Firr Court, Western Bank, Sheffield S10 2TN, UK

Received 2 July 1997; accepted 14 August 1997

Abstract

The amino acid sequences of haemoglobin-like proteins from the bacteria Alcaligenes eutrophus, Bacillus subtilis, Erwinia chrysanthemi, Escherichia coli, Vibrio parahemolyticus, Vitreoscilla sp. and the yeast Saccharomyces cerevisiae were studied. Phylogenetics based on distance and parsimony analysis showed that the cubacterial group can be easily distinguished from the other haemoglobin-like proteins. The construction of a consensus bacterial flavohaemoglobin based on the alignment of six bacterial and one yeast globins allowed the design of consensus primers to search for haemoglobin-like genes in other bacteria. PCR products of the expected size were found in Campylobacter jejuni, Salmonella typhimurium, Listeria monocytogenes, Rhizobium leguminosarum, Klebsiella pneumoniae, Pseudomonas aeruginosa and Staphylococcus aureus.

Keywords: Bacterial haemoglobin; hmp; Enterobacteria; Flavohaemoglobin

1. Introduction

Haemoglobin is defined as a haem protein that binds oxygen reversibly. It is widespread in many organisms including animals [1] higher plants [2], fungi [3, 4], protozoa [5, 6] and bacteria [7-11]. Wakabayashi [11] reported the first bacterial haemoglobin sequence, that of the haemoglobin from Vitreoscilla, formerly identified as a 'soluble cytochrome α'. Since then, other eight bacterial haemoglobins have been reported. Although the function of vertebrate haemoglobins and myoglobin in oxygen transport and storage is well established, the function(s) of haemoglobins in other groups of organisms is unclear. Several possible functions have been proposed for bacterial haemoglobins. For example, Vitreoscilla haemoglobin has been proposed to function in oxygen storage, diffusion or delivery in cells grown at low oxygen partial pressures [11, 12]. In addition, it can serve as a terminal oxidase under some conditions [13]. The flavohaemoglobin Fhp of Alcaligenes eutrophus has been proposed to function, either directly or indirectly, in denitrification [14] and the Escherichia coli Hmp protein has been proposed to be an oxygen sensor [15]. Recent work in our laboratory has also implicated Hmp with oxidative stress responses [16-18]. The patterns of expression of bac-
Bacterial haemoglobin types are varied [15]: for example, *Vitrioscilla* haemoglobin is maximally expressed at low oxygen tensions, whereas the regulation of expression of *E. coli* and *E. chrysanthemi* haemoglobin does not respond to oxygen availability. Bacterial haemoglobins can be classified in two groups based on their protein structure: one group consists of small haemoglobins that contain only one domain (haem); this group includes the haemoglobins of *Vitrioscilla* sp. and *Nostoc commune* [15]. The haemoglobins in *E. coli, E. chrysanthemi, A. eutrophus, V. parahaemolyticus* and *Bacillus subtilis* contain a second domain which binds flavin (FAD) and has NAD(P)H oxidase activity.

Here we report alignment of the known bacterial haemoglobins, their phylogenetic relationships and the design of a consensus sequence. We have designed two oligonucleotide 'consensus' primers allowing a search for the presence of haemoglobin-like proteins in other bacteria. Using this approach we report the presence of haemoglobin-like genes in bacteria of significance in pathogenicity, food poisoning and various natural environments, namely: *Campylobacter jejuni, Salmonella typhimurium, Listeria monocytogenes, Rhizobium leguminosarum, Klebsiella pneumoniae, Pseudomonas aeruginosa* and *Staphylococcus aureus*.

### 2. Material and methods

#### 2.1. Phylogenetic analysis

Amino acid sequences of bacterial haemoglobins used in this work are shown in Table 1. The initial alignment of amino acid sequences was made using CLUSTAL (PC-Gen; IntelliGenetics Inc., Mountain View, CA) and was further optimised with the PILEUP program (GCG, Madison, WI). To verify tree topologies, the orders of the sequence sets were jumbled. For neighbour-joining analysis, Distances (GCG) was used; an unrooted phylogenetic tree was generated with Growtree (GCG).

#### 2.2. Chromosomal DNA purification

Chromosomal DNA from *C. jejuni, S. typhimurium, L. monocytogenes, H. pylori, A. vinelandii, R. leguminosarum, K. pneumoniae, P. aeruginosa* and *S. aureus* was extracted using the Promega (Madison, WI) Genomic DNA purification kit.

### 2.3. PCR reactions

Two degenerate oligonucleotides Oligo RKPOl (2048-fold redundant) (5'-ATG(C/T)TXGCXGAX-AA(A/G)ACX(A/C)G-3') and RKPO 1 rev (512-fold redundant) (5'-AC(A/G)TXGC(G/C)AGNACNC-CA/GTA-3') were used in a PCR reaction as follows: 1 cycle of 95°C for 2 min followed by 5 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min then 30 cycles of 94°C for 1 min, 62°C for 1 min and 72°C for 1 min and finally 1 cycle of 72°C for 20 min. PCR reactions were carried out using Pfu DNA polymerase (Stratagene, La Jolla, CA) in a Hybaid Touchdown Thermocycler. Amplified DNA products were run in 0.8% agarose TBE gels.

### 3. Results and discussion

#### 3.1. Alignment of the bacterial haemoglobin protein sequences

Almost all bacterial haemoglobins belong to the two-domain flavohaemoglobin family; these haemoglobins contain a haem domain attached to a flavin domain; the only exception is the *Vitrioscilla* sp. Vhb haemoglobin that contains only one domain. In this study we considered the sequences of the five bacterial flavohaemoglobins available, *Vitrioscilla* sp. Vhb globin and the yeast flavohaemoglobin Yhb, which is a flavohaemoglobin with closer homology to bacterial haemoglobins than to animal or plant haemoglobins [4]. The amino acid sequences

<table>
<thead>
<tr>
<th>Organism</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Alcaligenes eutrophus</em></td>
<td>A53396</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>D78189</td>
</tr>
<tr>
<td><em>Erwinia chrysanthemi</em></td>
<td>X75893</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>P24232</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>L07071</td>
</tr>
<tr>
<td><em>Vibrio parahaemolyticus</em></td>
<td>U90005</td>
</tr>
<tr>
<td><em>Vitrioscilla</em></td>
<td>M27061</td>
</tr>
</tbody>
</table>
were aligned (see Section 2) and the high degree of similarity among them was evident (Fig. 1). Three-dimensional structure helices are shown based on the crystal structure of the *A. eutrophus* Fhp haemoglobin [20]. Table 2 shows the percentage of identity found among the flavohaemoglobins and also between their haem domain and the *Vitreoscilla sp.* haemoglobin. The identity varied between 35.71% and 64.81% and the homology between 55.02% and 77.21%; the two most similar flavohaemoglobins were Hmp and HmpX and the two with lowest identity were HmpX and Yhb. Yhb, the yeast haemoglobin had highest identity with the *Vitreoscilla sp.* Vhb globin. It was clear that the identities found among the enterobacteria (*E. chrysanthemi*, *E. coli* and *V. parahaemolyticus*) were the highest (Table 2), making a distinguishable group. From the alignment in Fig. 1, clusters of amino acids very homologous among the flavohaemoglobins from enterobacteria could be identified. Amino acids from positions 101–119 are highly similar among the eubacteria in comparison with the other three haemoglobins. Even though Yhb and the *Vibrio parahaemolyticus* haemoglobin showed the lowest identity, the homology was high (61.17%). Again Vhb was closest to Yhb in the homology data shown in Table 2. Inside the haem domain, helix H was the most conserved region among the bacterial haemoglobins.

### 3.2. Phylogenetic relationships among bacterial haemoglobins

Based on these alignments, the program Distances was used to create a matrix for the construction of a phylogram using the GCG program (GCG). Fig. 2 shows the results of this study. A cluster of the eubacteria globins was again clearly distinguished from the other four sequences. It has been suggested [21] that there is evidence for globin transfer within the yeast-proteobacteria cluster. A common ancestor
Table 2
Percentages of identity and homology in bacterial haemoglobins

<table>
<thead>
<tr>
<th>Percentage of identity</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vhb</td>
<td>Hmp</td>
<td>Fhp</td>
<td>HmpX</td>
</tr>
<tr>
<td>E. coli (Hmp)</td>
<td>46.52</td>
<td>45.31</td>
<td>50.68</td>
</tr>
<tr>
<td>A. eutrophus (Fhp)</td>
<td>52.74</td>
<td>37.83</td>
<td>50.61</td>
</tr>
<tr>
<td>E. chrysanthemi (HmpX)</td>
<td>43.05</td>
<td>61.92</td>
<td>57.50</td>
</tr>
<tr>
<td>S. cerevisiae (Yhb)</td>
<td>40.61</td>
<td>39.35</td>
<td>43.56</td>
</tr>
<tr>
<td>K. parahemolyticus (VbHmp)</td>
<td>43.05</td>
<td>61.92</td>
<td>57.50</td>
</tr>
<tr>
<td>B. subtilis (Hmpb)</td>
<td>50.61</td>
<td>37.78</td>
<td>40.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage of homology</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vhb</td>
<td>Hmp</td>
<td>Fhp</td>
<td>HmpX</td>
</tr>
<tr>
<td>E. coli (Hmp)</td>
<td>68.75</td>
<td>64.30</td>
<td>68.49</td>
</tr>
<tr>
<td>A. eutrophus (Fhp)</td>
<td>68.49</td>
<td>59.48</td>
<td>65.78</td>
</tr>
<tr>
<td>E. chrysanthemi (HmpX)</td>
<td>64.58</td>
<td>67.51</td>
<td>61.17</td>
</tr>
<tr>
<td>S. cerevisiae (Yhb)</td>
<td>69.78</td>
<td>63.21</td>
<td>57.3</td>
</tr>
<tr>
<td>V. parahaemolyticus (VbHmp)</td>
<td>69.78</td>
<td>63.21</td>
<td>57.3</td>
</tr>
<tr>
<td>B. subtilis (Hmpb)</td>
<td>69.78</td>
<td>63.21</td>
<td>57.3</td>
</tr>
</tbody>
</table>

Percentage of identity and homology among the haemoglobin amino acid sequences found in micro-organisms, using the program GAP. GAP uses the algorithm of Feng and Doolittle [23] to find the alignment of two complete sequences that maximises the number of matches and minimises the number of gaps. Percentages of identity and homology between Vhb and flavohaemoglobins are the values for the alignment of the Vhb globin with the haem domain of the flavohaemoglobins.

to E. coli, A. eutrophus, Vitreoscilla, B. subtilis and V. parahaemolyticus is thought to have acquired a copy of the chimeric flavohaemoglobin gene from an ancestor of S. cerevisiae not long after it separated from its earlier common ancestor with C. norvegensis, less than 400 Mya. The flavoprotein domain was subsequently removed in the line leading to Vitreoscilla. This scenario is compatible with the evolutionary tree based on the flavohaemoglobin haem domain (Fig. 3).

3.3. Construction of a consensus flavohaemoglobin protein sequence

Arredondo-Peter and Escamilla [19] reported a consensus sequence for plant haemoglobins based on the alignment of 16 haemoglobin amino acid sequences from different plants. Interestingly, the consensus sequence of plant haemoglobin was more homologous to Vitreoscilla Vhb haemoglobin than to animal haemoglobins [19]. Following the same protocol we determined the consensus sequence for two-domain bacterial haemoglobins (Fig. 1). The full consensus sequence corresponds to 418 amino acids. The most abundant amino acid in the consensus sequence was alanine (9.8%) and the estimated molecular mass is 46 855 Da, which is in good agreement with that of various flavohaemoglobins. The most conserved residues are at positions 6–16 (helix A), 43–56 (helices C and E), and 121–131 (helix H) in the haem domain, and at 190–194 (FP3, pyrophosphate portion of FAD), 281–295 (Nε1, ribose moiety

---

Fig. 2. Phylogenetic relationships among diverse bacterial haemoglobins and the yeast flavohaemoglobin based on the comparisons of their haem domain. The unrooted tree was generated using the PHYLIP program [25].
of NAD(P)H and 401–411 at the C-terminus in the flavin domain [4,20]. From these data we conclude that helix H is the most conserved region among the flavohaemoglobins as Arredondo-Peter and Escamilla [19] reported for plant haemoglobins. The hydrophathy profile of the consensus sequence (Fig. 3) shows that the most conserved segment of helix C is hydrophilic in contrast with those of helices A and H, the FAD pyrophosphate and NADP ribose domains and the C-terminus which are hydrophobic. A TFASTA (GCG) comparison search carried out on the consensus sequence showed that among the highest scores were plant haemoglobins rather than animal haemoglobins (data not shown).

3.4. Identification of haemoglobin-like genes in other bacteria

One of the goals of the production of a consensus flavohaemoglobin sequence is the synthesis of appropriate probes to detect haemoglobin-coding sequences in a variety of bacteria. Taking into account the alignments shown in Fig. 1, we designed two oligonucleotides that could serve as probes for searching for flavohaemoglobin genes in several bacteria. We designed the following oligonucleotides, RKP31 (5’-ATG(C/T)TXGCGAXAA(A/G)ACX(A/C)G-3’) corresponding to the N-terminal domain of the haem module and RKP32 (5’-AC(A/G)TXGCG(G/C)AGNACNCCA/GTA-3’) that corresponds to the H helix of the haem domain (Fig. 1). The expected size of the product was 400 bp. Chromosomal DNA from C. jejuni, S. typhimurium, L. monocytogenes, R. leguminosarum, K. pneumoniae, H. pylori, A. vinelandii, P. aeruginosa and S. aureus was extracted and PCR reactions were carried out as described in Section 2. As shown in Fig. 4, fragments of the expected size were found in C. jejuni, S. typhimurium, L. monocytogenes, R. leguminosarum, K. pneumoniae, P. aeruginosa and S. aureus. No products were found when H. pylori or A. vinelandii chromosomal DNA was used as a source of template even at lower annealing temperatures. Interestingly, the S. aureus PCR fragment was slightly smaller than the other PCR products found. Furthermore in R. leguminosarum, the amplified product must be different to the FixL haemoprotein previously reported [22] since the homology between the flavohaemoglobin family and the FixL protein is not high. More studies are required to identify the genes corresponding to these PCR products. However our results reveal an effective pair of probes useful for the identification of new bacterial haemoglobin genes.

Acknowledgments

We thank Alicia Gil, Susan Hill, Dave Kelly, Simon Park and Guanghui Wu for bacterial strains.
This research was supported by BBRSC grant P05184. J.M.-H. acknowledges the very useful comments of Dr. Carmen Gomez-Eichelmann.

References


