Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 23 (2004) 211–221
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Hemoglobin-binding protein HgbA in the outer membrane of *Actinobacillus pleuropneumoniae*: homology modelling reveals regions of potential interactions with hemoglobin and heme

Peter D. Pawelek, James W. Coulton*

Department of Microbiology and Immunology, McGill University, 3775 University Street, Montreal, QC, Canada H3A 2B4

Received 23 February 2004; received in revised form 25 June 2004; accepted 25 June 2004
Available online 3 August 2004

Abstract

Analyses of the primary sequence of hemoglobin-binding protein HgbA from *Actinobacillus pleuropneumoniae* by comparative modelling and by a Hidden Markov Model identified its topological similarities to bacterial outer membrane receptors BtuB, FepA, FhuA, and FecA of *Escherichia coli*. The HgbA model has a globular N-terminal cork domain contained within a 22-stranded β barrel domain, its folds being similar to the structures of outer membrane receptors that have been solved by X-ray crystallography. The barrel domain of the HgbA model superimposes onto the barrel domains of the four outer membrane receptors with rmsd values less than 1.0 Å. This feature is consistent with a phylogenetic tree which indicated clustering of polypeptide sequences for three barrel domains. Furthermore, the HgbA model shares the highest structural similarity to BtuB, with the modelled HgbA barrel having approximately the same elliptical cross-section and height as that of BtuB. Extracellular loop regions of HgbA are predicted to be more extended than those of the *E. coli* outer membrane receptors, potentially facilitating a protein–protein interface with hemoglobin. Fold recognition modelling of the HgbA loop regions showed that 10 out of 11 predicted loops are highly homologous to known structures of protein loops that contribute to heme/iron or protein–protein interactions. Strikingly, HgbA loop 2 has structural homology to a loop in bovine endothelial nitric acid oxidase that is proximal to a heme-binding site; and HgbA loop 7 contains a histidine residue conserved in a motif that is involved in heme/hemoglobin interactions. These findings implicate HgbA loops 2 and 7 in recognition and binding of hemoglobin or the heme ligand.

Keywords: Homology modelling; Hemoglobin receptor; Outer membrane proteins; β Barrel; TonB.

1. Introduction

Actinobacillus pleuropneumoniae, a Gram negative bacterium belonging to the *Pasteurellaceae* family, is the etiological agent of porcine pleuropneumonia, a severe disease that causes worldwide economic loss to the swine industry [1]. Different factors contribute to virulence of *A. pleuropneumoniae* and a comprehensive review on its pathogenicity was recently published [2]. The molecular determinants whereby a micro-organism acquires iron are recognized among its virulence factors. *A. pleuropneumoniae* may use porcine hemoglobin (Hb) and heme compounds as sources of iron for growth [2–4].

We cloned and sequenced *hgbA* [5], a gene of *A. pleuropneumoniae* that encodes a hemoglobin-binding protein HgbA at the bacterial cell surface. Sequence analysis of *hgbA* identified a recognition sequence, the Ton box that is specific to TonB-dependent outer membrane (OM) receptors. Furthermore, *hgbA* is highly homologous to the gene encoding a TonB-dependent hemoglobin-binding protein in *P. multocida* [5]. To identify functional regions of HgbA that might be involved in Hb binding and heme recognition/transport, we targeted the known structures of TonB-dependent OM receptors as templates to construct a homology model of HgbA.

With the recent publication of the X-ray crystal structure for BtuB, there are now four high-resolution structures of

* Corresponding author. Tel.: +1 514 398 3929; fax: +1 514 398 7052.
E-mail address: james.coulton@mcgill.ca (J.W. Coulton).

1093-3263/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2004.06.002

TonB-dependent OM receptors which can be used as templates for homology modelling: BtuB [6], FepA [7], FhuA [8], and FecA [9]. These OM receptors share structural features. Each protein possesses an N-terminal globular cork domain inserted within a C-terminal barrel domain that is comprised of 22 antiparallel strands. Since the overall sequence identity of these TonB-dependent OM receptors is approximately 15% in the β barrel domains, the potential for homology modelling in the absence of additional information is limited. Alternative approaches including a Hidden Markov Model (HMM) that discriminates β strand sequences from genomic sequences have been utilized to predict OM barrel domains [10]. Using the HMM of Martelli et al., we assigned β strands to regions of our predicted HgbA amino acid sequence, culminating in a structure-based multiple sequence alignment to BtuB, FepA, FhuA, and FecA. The three-dimensional model generated from this alignment provides an overall topology of HgbA and identifies extracellular loop regions. Significantly, predicted extracellular loop regions exhibit high homology and fold similarity to loops in known protein structures that contribute to metal binding and to protein-protein interactions. These observations are consistent with the predicted functions of HgbA.

2. Methodology

2.1. Structure-based multiple sequence alignments

A structure-based alignment of the primary sequences of barrel domains from BtuB, FepA, FhuA, and FecA was generated by the computer program CE [11]. The PDB files 1nqe (BtuB), 1fep (FepA), 2fcp (FhuA), and 1kmo (FecA) were initially used to obtain pairwise alignments; these data were then merged to form a single structure-based multiple alignment. The multiple alignment generated by CE was visually inspected and a small number (<10 position shifts in total) of manual refinements were performed using GNU Emacs and the seq-mode module (http://www.abo.fi/ mhuhtala/seqmode.html), thereby optimizing the overall sequence identity of the alignment. Only manual refinements that resulted in improvements of positional BLOSUM62 scores of 0.50 or greater were accepted. To ensure that residues in corresponding secondary-structure elements (β strands, periplasmic turns, extracellular loops) were properly classified, the aligned structures were examined using the program SYBYL 6.9 (Tripos Corporation). Based on a candidate topology suggested by a HMM that predicts β strands in bacterial OM barrel domains [10], our predicted primary sequence of mature HgbA, 923 amino acids (GenBank entry AF468020), was merged with the structure-based alignment of BtuB, FepA, FhuA, and FecA.

A structure-based multiple sequence alignment of the cork domains of BtuB, FepA, FhuA, and FecA was generated by CE and manually refined as described above. The predicted amino acid sequence of the cork domain of HgbA was merged with this alignment based on a partial structure-based alignment of the HgbA sequence against the FhuA structure that was returned by the Jigsaw-3D Protein Homology Modelling Server (http://www.bmm.icnet.uk/servers/3djigsaw) [12]. Using CE and manual adjustments, further refinement of the predicted HgbA cork domain aligned to the cork domains of BtuB, FepA, FhuA, and FecA was performed until overall sequence identities were determined to be optimal.

The computer program Belvu was used to generate postscript outputs of the structure-based multiple alignments of the barrel and cork domains. Belvu was also used to obtain unrooted phylogenetic trees using the neighbor-joining method, and to calculate pairwise percentage sequence identity values from these alignments.

2.2. Homology modelling of the HgbA barrel and cork domains

The structure-based multiple sequence alignments of the barrel and cork domains of HgbA, BtuB, FepA, FhuA, and FecA were used as inputs for the computer program Modeller [13] to obtain three-dimensional homology models of the HgbA barrel and cork domains. Prior to generating the models, the alignments were checked by Modeller and revised such that no implied target Cα(i) − Cα(i + 1) distances were longer than 8.0 Å apart. Upon optimization of the sequence alignments, three-dimensional models were then generated by Modeller. During this process, the models were refined by molecular dynamics routines with incremental increases in simulation temperature from 150 K to 1000 K, followed by incremental temperature decreases from 1000 K to 300 K.

The atomic coordinates of the refined HgbA homology model were submitted to the RCSB Protein Databank (http://www.rcsb.org/pdb/) and the model was assigned the PDB code 1uxf.

2.3. Statistical analyses of the HgbA homology models

The models of the HgbA barrel and cork domains were submitted to the Biotech Validation Suite for Protein Structures (http://biotech.embl-ebi.ac.uk:8400/) for PROCHECK [14] and WHAT IF [15]. Ramachandran plots were generated by a local copy of PROCHECK on an IBM Intellistation M Pro Workstation running Redhat Linux 9.

2.4. Comparative structural alignments

Barrel and cork domains of the known structures of BtuB, FepA, FhuA, and FecA and the HgbA model structures were structurally aligned in a pairwise manner by the program LSQMAN [16] to obtain Levitt–Gerstein [17] and normalized rmsd statistics. Structures were initially aligned using coarse-fit parameter settings and then optimized with 10 cycles of Dynamic-Programming-based operator improvement.

2.5. Fold-prediction of HgbA extracellular loop regions

Amino acid sequences corresponding to predicted HgbA extracellular loops were submitted to the mGenTHREADER [18] (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html). Hits against query sequences were ranked according to sequence identity and Net Score to target protein regions returned by the server.

3. Results

3.1. Structure-based multiple sequence alignment

Initial topology of the barrel domain was predicted by an HMM [10] trained against the known structural data of OM receptors that identifies potential β strands in the barrel domain. The HMM proposed all 22 strands for the present model and the observed strand-turn-strand periodicity permitted assignment of HgbA residues to predicted extracellular loop regions. A structure-based alignment of the predicted amino acid sequence of HgbA to the sequences of TonB-dependent OM receptors was generated by the computer program CE, followed by manual refinement (Fig. 1A). Comparison of percentage sequence identities of pairwise alignments of barrel domains from HgbA to BtuB, FepA, FhuA, and FecA indicated that overall sequence identities are low, less than 20.5%, and they are generally within the same range (Table 1, rows 1–4). The HgbA barrel has slightly higher overall sequence identity to the barrel domains of BtuB and FepA.

A structure-based alignment was also constructed to align HgbA residues 7–144 against the cork domains of BtuB, FepA, FhuA, and FecA (Fig. 1B). This region of the HgbA sequence is N-terminal to the first β strand predicted by the HMM, and its overall sequence identities are comparable to those of the four TonB-dependent OM receptors (Table 1, rows 5–8). When aligned to the four TonB-dependent OM receptors, the predicted HgbA cork domain has highest sequence identity to the cork domain of FepA. In addition, HgbA residues 7–11 align to the Ton box of BtuB (residues 6–10). The Ton box of BtuB is followed by nine residues (11–19) that could not be structurally aligned with HgbA. This difference suggests that the Ton box of HgbA may be less extended into the periplasm than the Ton box of BtuB. Because there are no structural data for the Ton boxes of FhuA and FecA, they are not included in this analysis.

An unrooted phylogenetic tree based on the multiple alignments of barrel domains supports overall similarity between the barrel of HgbA and those of BtuB and FepA (Fig. 2A). The barrel domains of FecA and FhuA are more closely related to each other than to those of HgbA, BtuB, and FepA. This pattern was not observed in the unrooted phylogenetic tree based on the multiple alignments of cork domains. The HgbA cork exhibits a highest overall similarity to the FhuA cork (Fig. 2B); no phylogenetic clustering of BtuB/FepA and FhuA/FecA cork domains was observed.

Table 1
Sequence identities of structurally aligned barrel and cork domains

	HgbA	BtuB	FepA	FhuA
Barrel				
BtuB	16.9[a]	–	–	–
FepA	20.2	18.8	–	–
FhuA	14.1	18.2	13.9	–
FecA	14.2	17.0	15.9	18.2
Cork				
BtuB	23.1	–	–	–
FepA	24.1	38.8	–	–
FhuA	23.9	29.9	32.9	–
FecA	19.6	31.0	32.6	28.0

[a] Sequence identities (%) calculated from multiple alignments presented in Fig. 1A and B.

3.2. Three-dimensional homology model of HgbA

The structure-based multiple sequence alignment was submitted to the computer program Modeller and a three-dimensional model structure of HgbA was obtained. To remove bias from highly non-homologous regions corresponding to extracellular loops, the loop regions within BtuB, FepA, FhuA, and FecA were placed in non-overlapping positions within the alignment (Fig. 1A). To prevent Modeller from assigning secondary structure elements in the absence of homologous structural templates, extended sequences flanking the HgbA β-turn-β regions predicted by the HMM were replaced with pentaglycine sequences. The overall structure of the model (Fig. 3) is homologous to the known structures of TonB-dependent OM receptors: an N-terminal globular cork domain enclosed within a C-terminal barrel domain comprised of 22 antiparallel β strands. The barrel domain has an elliptical cross-section of 45 Å × 30 Å and a maximal height of 55 Å. Consistent with known structures, the model of the HgbA cork domain has a central four-stranded β sheet tilted at an angle approximately 30° relative to the cross-sectional plane of the barrel. The HgbA model has percentage β and α content similar to BtuB, FepA, FhuA, and FecA (Table 2) and the overall model conforms to acceptable stoichiometric parameters. Ramachandran analysis indicated that residue distribution is similar to those of the known structures of TonB-dependent OM receptors; no residues occur within disallowed regions. PROCHECK analysis revealed that the HgbA homology model has an overall average G factor of −0.160, significantly above the threshold value (−0.500) that reflects a protein fold with acceptable stoichiometry. Furthermore, the G factors of BtuB and FepA are negative values, similar to that of the HgbA model (Table 2).

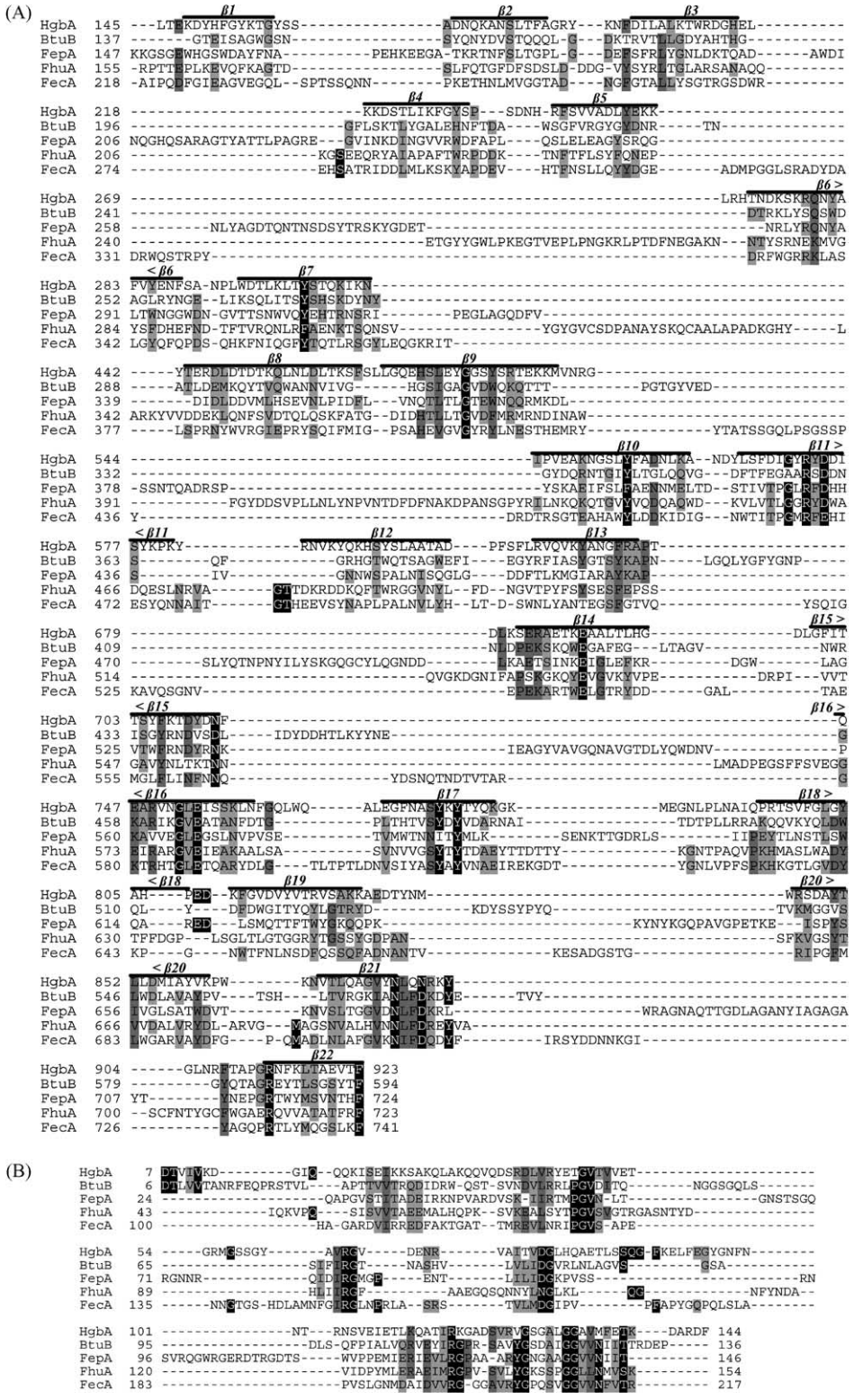

Fig. 1. Structure-based alignment of the predicted amino acid sequence of HgbA to the primary sequences of TonB-dependent OM receptors: BtuB, FepA, FhuA, and FecA. Extramembrane regions of BtuB, FepA, FhuA, and FecA that were not used to obtain the HgbA homology model are shown without alignment. Numbering of BtuB, FepA, FhuA, and FecA corresponds to residue numbers in PDB files 1nqe, 1fep, 2fcp, and 1kmo, respectively. Numbering of HgbA corresponds to the mature protein sequence. Regions corresponding to predicted extracellular loops were replaced by pentaglycine sequences in the model and are therefore not shown in this alignment. Alignment positions are highlighted according to average BLOSUM62 score: 3.00, black background and white character; 1.50, dark grey background and black character; 0.50, light grey background and black character. (A) Alignment of HgbA 145–923 to the barrel domains of BtuB, FepA, FhuA, and FecA. (B) Alignment of HgbA 7–144 to the cork domains of BtuB, FepA, FhuA, and FecA.

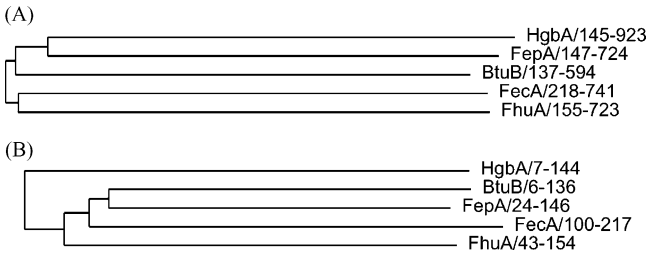

Fig. 2. Phylogenetic relationships of HgbA to the TonB-dependent OM receptors of known structure. (A) Unrooted phylogenetic tree of HgbA 145–923 and the aligned barrel domains of BtuB, FepA, FhuA, and FecA. (B) Unrooted phylogenetic tree of HgbA 7–144 and the aligned cork domains of BtuB, FepA, FhuA, and FecA. Line length is directly proportional to degree of similarity between sequences.

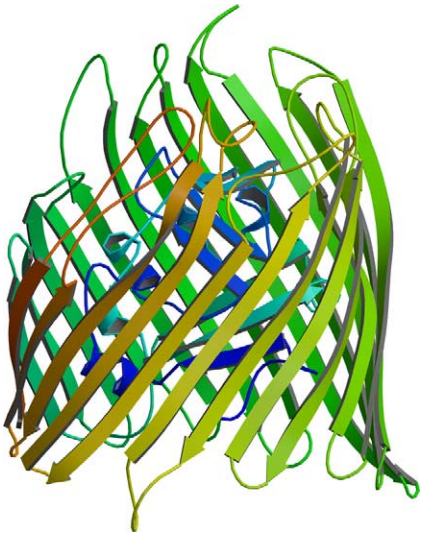

Fig. 3. Ribbon representation of the homology-based model of HgbA. Strands represented as flat arrows, loops as thin tubes, and helices as flat coils. The model color is ramped according to relative position between N-terminus (blue) and C-terminus. Figure generated by the program MOLSCRIPT [34].

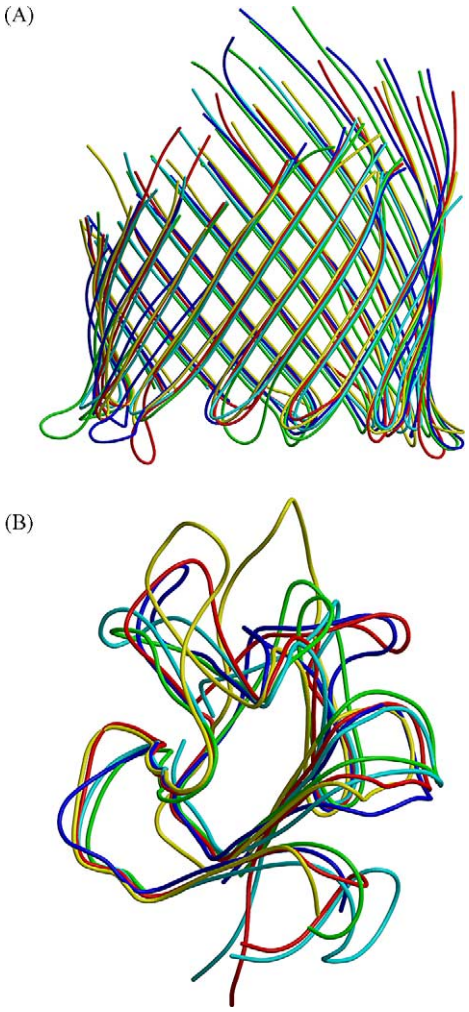

Fig. 4. Cα superpositions of the HgbA model and TonB-dependent OM receptors of known structures. Cα traces of each structure colored as follows: HgbA (red), BtuB (cyan), FepA (yellow), FhuA (green), FecA (blue). (A) Barrel domains: β strands and periplasmic turns are shown, but not extramembrane loop regions. (B) Cork domains. Figure generated by the program MOLSCRIPT [34].

3.3. Structural superposition of BtuB, FepA, FhuA, and FecA and the HgbA model

The three-dimensional structures of the barrel and cork domains of BtuB, FepA, FhuA, and FecA and the HgbA homology model were aligned by least-squares methods using the program LSQMAN. Visual inspection of the superimposed Cα traces of the five proteins demonstrated that the HgbA model shares a high degree of similarity to folds of the four TonB-dependent OM proteins of known crystal structures (Fig. 4A and B). However, it was clear that a higher degree of structural homology occurs in the barrel domain (Fig. 4A)

Table 2
Statistical evaluation of the HgbA homology model

Protein	Template structure[a]	% β	% α	Most favored (%)[b]	Allowed (%)[b]	Disallowed (%)[b]	Overall average G-factor[c]
HgbA	–	60.5	5.2	89.6	10.4	0	−0.160
BtuB	1nqe	64.5	5.3	90.6	9.4	0	−0.090
FepA	1fep	57.7	5.2	87.5	12.5	0	−0.030
FhuA	2fcp	59.5	4.6	84.4	16.6	0	0.190
FecA	1kmo	60.3	5.3	88.4	11.3	0.4	0.140

[a] PDB code; in each case, the A chain was used as the homology modelling template.
[b] Ramachandran analysis; percentage of total residues within designated region.
[c] PROCHECK analysis; acceptable values >−0.500.

Table 3
Structural alignments of barrel and cork domains; entries sorted by $S_{structure}$ (highest to lowest, descending)

M1	M2	Alignment length (residues)	Gaps	Normalized rmsd (Å)[a]	$S_{structure}$[b]
Barrel[c]					
HgbA	BtuB	373	18	0.714	5726
FhuA	FecA	403	31	0.949	5670
HgbA	FepA	377	26	0.731	5640
FepA	BtuB	350	19	0.914	5536
HgbA	FecA	401	29	0.854	5528
HgbA	FhuA	419	27	0.954	5409
FhuA	BtuB	401	26	0.912	5279
FepA	FecA	387	28	0.967	5261
FecA	BtuB	388	27	1.064	5101
FhuA	FepA	408	30	1.154	4968
Cork					
HgbA	BtuB	154	11	1.255	2051
HgbA	FecA	151	10	1.322	1868
HgbA	FepA	165	11	1.270	1709
FecA	BtuB	154	10	1.492	1664
FhuA	BtuB	149	12	1.435	1636
HgbA	FhuA	157	12	1.354	1633
FhuA	FecA	139	9	1.608	1587
FepA	BtuB	163	13	1.551	1555
FepA	FecA	150	16	1.626	1516
FhuA	FepA	150	17	1.482	1421

[a] rmsd Values normalized to a fixed length of 100 residues.
[b] Levitt–Gerstein structural alignment score.
[c] Aligned sequences include barrel strands and periplasmic turns, but not extracellular loops.

than in the cork domain (Fig. 4B). In addition to evaluating superposition by rmsd, we also performed Levitt–Gerstein analyses on the superposition. The Levitt–Gerstein score has been shown [17] to be a more robust expression of three-dimensional Cα superposition than the rmsd value. Analyses of pairwise alignments of the three-dimensional structures of barrel domains from BtuB, FepA, FhuA, and FecA showed that the FhuA–FecA pair has the highest Levitt–Gerstein $S_{structure}$ score; the FepA–BtuB pair has the second highest $S_{structure}$ score, suggesting that FhuA/FecA and FepA/BtuB have structurally distinct features in their barrel domains (Table 3, rows 1–10). However, the cork domains of all four proteins are not distinct from each other according to the superposition analysis (Table 3, rows 11–20). The overall rmsd values of cork domain superposition are significantly higher, reflecting structural heterogeneity in the cork domains relative to the barrel domains. Analysis of pairwise alignments of the modelled HgbA barrel and cork domains indicates that HgbA has the highest degree of structural similarity to BtuB in both domains. For all pairs examined, superposition of the modelled HgbA barrel domain with that of BtuB results in the highest $S_{structure}$ scores and lowest normalized rmsd values.

A comparison of β strand lengths within barrel domains of BtuB, FepA, FhuA, FecA along with the predicted β strands from the HgbA homology model is shown in Fig. 5. Whereas all four barrel domains have similar strand lengths (average of 12 residues) between β1–β5 and β13–β22 (Fig. 5A and B), the FhuA and FecA barrel domains have significantly longer β strands (average of 20 residues) between β6–β12 (Fig. 6A). The overall distribution of HgbA β strand lengths is more similar to BtuB/FepA than to FhuA/FecA.

3.4. Prediction of functions for extracellular loop regions in the HgbA model

Of the 11 extracellular loops predicted by our model (Table 4), 10 were found by mGenTHREADER [18] to have significant homology (generally >20% sequence identity) and a medium to high confidence level (Net Score > 0.400) to loops in target proteins that are either involved in heme/metal binding or protein–protein interactions (Tables 5 and 6). All residues in these predicted loop regions were assigned random coil identities by mGenTHREADER based on structural alignments to target proteins, consistent with their locations in extracellular loops.

Table 4
Predicted extracellular loop sequences of HgbA

HgbA loop	
L1	(158)**YSSA**(161)[a]
L2	(191)**EL**ENYDYKTADGSVQGKEREKADPYSI(217)
L3	(247)NRGHDFSYNLKPTTYINVDEYE**LRH**(271)
L4	(309)RARTDDYCDGSHCKETENLAGLQLKDGKVVDRDGNQPNLGVDELGLTTITDSKGTYTEGVNLV RAYWFDCSVFDCNKSVTAY YKDYSNNITSEEVALTKTYTDEKGRKFATLDPKSKFKSILLPGSKGYTENI**Y**(442)
L5	(485)**VNRG**GFGATSNTQWWTKRFLGMRNNFFKGTEEVITCKNATGSDQWNGLICPNEDTFSFL(543)
L6	(582)**Y**VAGQTAKIPDDMLQGLFVPLPDRPTKEQIRQNAEENIKYLS(623)
L7	(658)**P**TSDELYFTFKHPDFTVLPNVDLK(681)
L8	(713)**F**IDLKYLGPKNLMNAFGGSATARPYQIYQNVNR(745)
L9	(782)**GKMEGNLPLNAIQ**(794)
L10	(825)**AEDTYNM**YHSEEKAKDSYLK**W**(845)
L11	(874)**LQNRKY**LTWESVRSIRPFGTSNLINQATGKG**LNRFTAPG**(912)

[a] Residues contained in the three-dimensional model are indicated in bold type. Residues replaced by pentaglycine sequences are shown in plain type. Numbers in parentheses correspond to protein sequence of mature HgbA (GenBank entry AF468020).

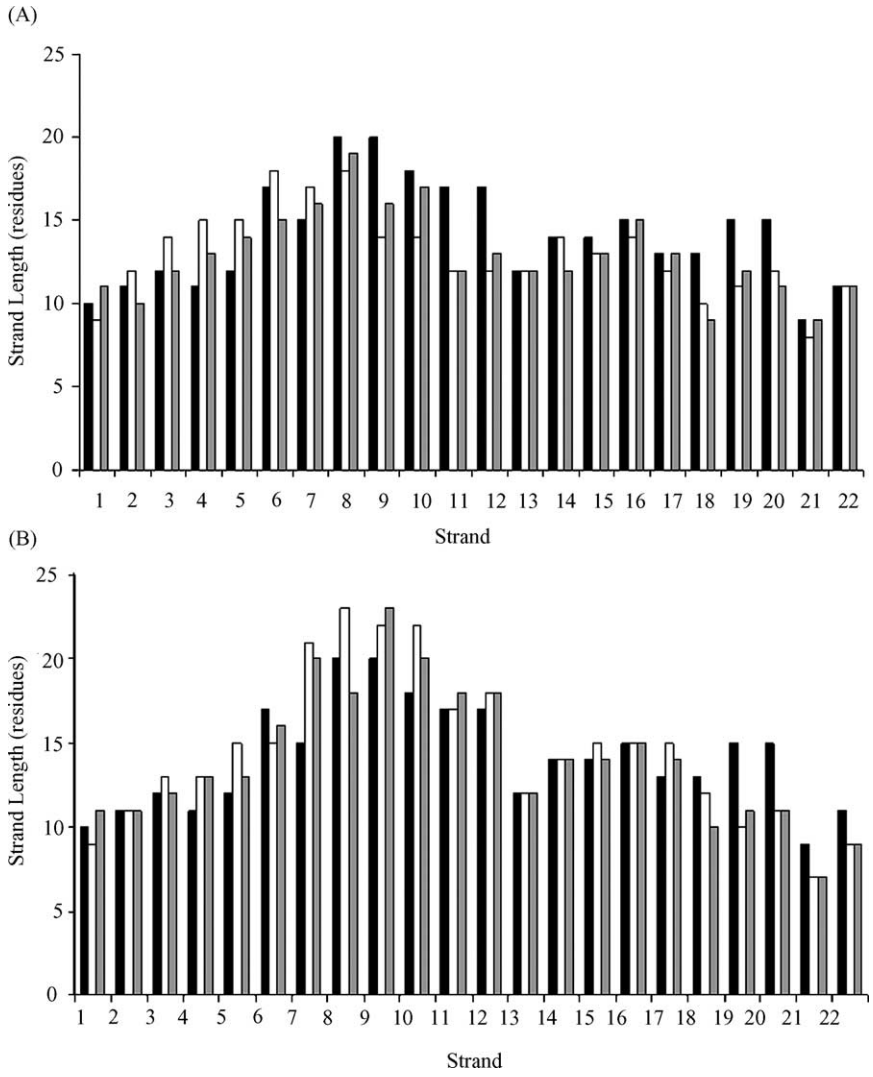

Fig. 5. Distribution of β-strand length as a function of position in the barrel. (A) Black: HgbA; white: BtuB; grey: FepA. (B) Black: HgbA; white: FhuA; grey: FecA.

4. Discussion

For the design of algorithms which scan genomic data and then predict open reading frames encoding bacterial β barrel membrane proteins, limited structural data are available. Based on known physical properties of the proteins in the database, Wimley reported [19] construction of a β barrel database to identify potential membrane-spanning regions. Martelli et al. [10] developed an HMM which uses their own database of β barrel membrane proteins to predict potential β barrel strands. The HMM relies on sequential occurrence patterns of residues within amino acid sequences that have been confirmed by high-resolution X-ray crystal structures to be β barrel strands. The authors claim [10] that their HMM is 84% accurate in predicting barrel strands that have no conspicuous sequence identity. The databases used in both methods comprise porins, other β barrel membrane proteins, and 22-stranded TonB-dependent OM receptors.

To construct a three-dimensional homology model of HgbA, we utilized the HMM described above and then comparative modelling techniques based on the data from four high-resolution structures of TonB-dependent OM receptors. Our model of HgbA includes an N-terminal cork domain and a C-terminal barrel domain comprised of 22 antiparallel β strands. The HMM correctly identified the open reading frame in the *A. pleuropneumoniae* genome which encodes HgbA. However, it predicted that the HgbA barrel domain was comprised of 30 β strands. Given that all TonB-dependent OM receptors of known structure have a maximum of 22 strands in the barrel domain, we interpreted this outcome to be unusual. Closer inspection of the sequences in the HMM-predicted strands revealed that the eight extra strands had weak sequence homology to barrel strands from porin proteins, but they had no homology to those from TonB-dependent OM receptors. Furthermore, a topology of repeating strand-loop-strand turn elements which is characteristic of the barrel domain could not be

Table 5
Comparison by mGenTHREADER of predicted HgbA extracellular loops with homologous loops from target proteins involved in metal binding

HgbA loop	Target protein	Species	Sequence identity (%)	Net Score[a]	Function of target protein
L2	Endothelial nitric oxide synthase 1nseA[b] (249)**QL**VR**Y**AG**YR**QQ**DGSVRG**------**DP**A**NV**(270) HgbA (191)**EL**EN**Y**D-**YK**TA**DGSVQG**KEREKA**DP**Y**SI**(217)	*Bos taurus*	33	0.414	Heme binding
L4	High potential iron protein 1ckuA (31)A**AR**PGLPP**E**EQ**HC**---A**N**CQF**MQ**ADAAGAT(59) HgbA (309)R**AR**TDDYC**D**GS**HC**KETE**N**LAG**LQ**LKDGKVV(338)	*Chromatium vinosum*	22	0.441	Fe·S cluster binding
L4	Cis aconitase 1amjA (381)**L**KC**KS**QF--T**I**T**PGS**(393) HgbA (420)**L**DP**KS**K**F**KSI**L**L**PGS**(434)	*Bos taurus*	28	0.419	Fe·S cluster binding
L6	Diol dehydratase 1dioG (107)R**AAE**LTA**VPDD**R**I**LE**IY**NA**LRPYRSTKEEL**LAI**ADD**LES**RY**QA(149) HgbA (583)V**AGQ**TAK**IPDD**M**L**QG**LF**VP**L-PDRPTKEQI**RQN**AEE**N-I**KY**LS(623)	*Klebsiella oxytoca*	29	0.453	Cyanocobalamin binding
L9	Fumarate reductase, γ subunit 1qlaC (195)**G**WF**DG**ET**P**DKTRA(207) HgbA (782)**G**KM**EG**N**L**P**L**NAIQ(794)	*Wolinella succinogenes*	23	0.407	Heme binding
L11	Ferridoxin oxidoreductase 1b25A (41)**I**LW**N**EA**RGVEP**LSPE**NKL**IF**AAG**P(64) HgbA (880)**LTW**ESV**RSIRP**FGTS**NL**I**NQ**ATG**K(903)	*Pyrococcus furiosis*	25	0.420	Fe·S cluster binding

[a] mGENTHREADER Net Score indicating degree of fold similarity; scores above 0.400 are considered significant.
[b] Target protein names are given according to their PDB entry and chain identity. Numbers of target proteins in parentheses correspond to the first amino acid alignment position in the target protein sequence according to PDB numbering. Numbers of HgbA in parentheses refer to protein sequence of mature HgbA. Positions of sequence homology are shown in bold.

constructed if the eight porin-like strands were included in the model. In our model, these sequences were therefore rejected as barrel strand elements and were assigned to extracellular loop regions. We have since established that our predicted extracellular loop regions have significant sequence identity (>20%) and fold similarity to loop regions in proteins involved in heme binding, metal binding, or protein–protein interactions. These findings support our decision to assign these sequences to potential extracellular loop regions.

Table 6
Comparison by mGenTHREADER of predicted HgbA extracellular loops with homologous loops from target proteins involved in protein-protein interactions

HgbA loop	Target protein	Species	Sequence identity (%)	Net Score[a]	Function of target protein
L3	Neurotoxin Type B, catalytic domain 1f82A[b] (368)**TR**ASY**FS**D**SL**P**P**VKIK**NL**LD**NE**I(390) HgbA (247)**NR**GHD**FSYNL**K**P**TTYI**NV**--**DE**Y(267)	*Clostridium botulinum*	33	0.428	Protein substrate binding
L3	Outer surface protein A 1osp (45)E**K**NK**D**GK**YDL**IA----T**VD**KL**EL**KG(65) HgbA (247)N**R**GH**D**FS**YNL**KPTTYIN**VD**EY**EL**RH(271)	*Borrelia burgdorferi*	33	0.432	Interacts with Fab fragment
L5	Flavocetin-A 1c3aA (74)IQNKEQQC**R**SE**WS**DASSVNYE**N**LV**K**QFS**K**--K**C**YAL**K**KG**T**ELRT**W**FN**VY**C**--GTENPEV**(128) HgbA (489)GFGATSNT**Q**W-**WT**KRFLGMRN**N**FF**K**GTE**E**VIT**C**---**K**NA**T**GSDQ**W**NG**LIC**PNEDTFSF**L**(543)	*Trimeresurus flavoviridis*	15	0.487	Platelet GP Ibα-subunit binding
L7	Lipocalin (HNGAL) 1nglA (86)**P**GCQPGE**FT**LGNIKSYPGLTSYLVR**VV**(112) HgbA (658)**P**TSDELY**FT**FKHP---------DFT**VL**(675)	*Homo sapiens*	22	0.428	Metalloproteinase binding
L8	Neurotoxin Type B, catalytic domain 1f82A (402)**KDM**EKE**YRG**QNK**A**INK**QA**Y**E**E**IS**(424) HgbA (722)**KNL**MNA**FGG**SAT**A**RPY**QIYQ**N**VN**(744)	*Clostridium botulinum*	20	0.423	Protein substrate binding
L10	Superantigen Smez-2 1et6A (142)IAQHQ**LYSS**GSSY**K**SGR**L**V**F**(161) HgbA (826)EDTYN**MYHS**EEKA**K**DSY**LKW**(845)	*Streptococcus pyrogenes*	20	0.416	MHC class II binding

For a and b, see footnotes to Table 5.

Although automated identification methods such as the ones described above are effective in identifying open reading frames corresponding to TonB-dependent OM receptors, caution must be adopted in over-interpreting specific strand assignments. Here, we show that barrel domains and cork domains of TonB-dependent receptors are phylogenetically distinct. It is therefore likely that barrel strands from these proteins have evolved to accommodate or form contacts with some globular domain in the interior of the barrel. Since porins do not possess an interior cork domain, the evolutionary pressures on porin-like barrel strands are likely quite different. We therefore suggest that databases comprised of multiple classes of β barrel membrane proteins introduce bias in predicting β strand assignments for proteins of a particular class.

Since extracellular loop regions of all TonB-dependent OM receptors do not exhibit significant sequence homology, we have not included extracellular loop regions in our homology-based three-dimensional model. However, given our assignment of HgbA residues to β strands and periplasmic turns, we propose regions of the HgbA primary sequence that are likely to be contained within extracellular loop regions. According to our model, 421 out of 778 modelled residues, or 55% of the HgbA protein mass that resides C-terminal to the cork domain, are found in the extracellular space. In contrast, 170 out of 554 residues, or 31% of the corresponding protein mass of *E. coli* FhuA are found in extracellular loop regions. A higher proportion of residues in extracellular loop regions may be necessary to form a protein–protein interface with hemoglobin that is sufficient to facilitate heme recognition and transfer.

The homology model of the HgbA barrel domain generated by Modeller does not include residues predicted to belong in extracellular loop regions. However, the entry and exit points of the 22 β strands in the model delineate boundaries of 11 extracellular loop regions.

HgbA loop 2 (L2) aligns (Table 5) with 33% identity over 23 amino acids to a loop in bovine endothelial nitric oxide synthase (eNOS) that occludes the eNOS heme-binding pocket [20]. Furthermore, mGenTHREADER predicts that these loops have significant fold similarity. We therefore propose that L2 may be involved in heme recognition or transfer of heme from Hb docked to HgbA. Similarly, loop L9 aligns with 23% identity to a loop in the γ-subunit of fumarate reductase from *W. succinogenes* that is also involved in heme interactions [21], thereby implicating L9 in HgbA heme recognition. In addition to heme interactions, two proposed extracellular loop regions appear to be involved in metal binding. Loop L4, an extended region of 134 residues, was found by mGenTHREADER to have two subsites with fold similarity to target protein loops involved in Fe·S cluster binding. The second subsite, which aligns with 22% identity to a loop from the high potential iron protein from *C. vinosum*, contains a His-Cys pair which is absolutely conserved with His42 and Cys43 in the target protein, shown [22] to interact directly with bound iron–sulfur cluster. Given the length of L4, this region might fold into an additional globular domain exposed to the extracellular milieu. Indeed, six of the eight HMM-predicted porin-like strands that we assigned to the extracellular space occur within L4. It is possible that these elements may contribute to the fold of such a domain. The remaining two HMM-predicted strands occur in L5 and L6. Loop L6 aligns with 29% identity to a loop in diol dehydratase from *K. oxytoca*, which binds cyanocobalamin, the same ligand transported by BtuB [23]. We therefore propose that L6 is also involved in heme interactions.

Although the heme group itself may be involved in the mechanism of recognition and binding to an OM receptor, it has recently been suggested [24] that this molecule may account for less than 20% of a total binding surface between Hb and a receptor such as HgbA; the remaining 80% of the recognition site would be formed by Hb itself. Using mGenTHREADER and comparative modelling techniques, we identify a number of HgbA extracellular loops that are likely to be involved in formation of this protein–protein interface. For example, HgbA loops L3 and L8 share significant homology to catalytic loops located in a secreted bacterial neurotoxin which bind host protein substrates. Residues 366–385 in botulinum neurotoxin B, which is homologous to HgbA L3, directly binds its protein substrate [25]. In addition to sharing homology with botulinum neurotoxin B, L3 also shares homology with an extracellular loop of the outer surface protein A from *Borrelia burgdorferi*. This protein is complexed with the Fab fragment in PDB entry 1osp, and the target homologous loop is involved in direct interaction with the Fab fragment [26]. HgbA loop L8 also aligns with 20% identity to a subsite within botulinum neurotoxin B. Where the alignment with L3 terminates at neurotoxin B residue 390, the alignment with L8 begins at neurotoxin B residue 402. Given that L3 and L8 are distant in the primary HgbA sequence, it is possible that these loops are proximal in three-dimensional space and may function in a coordinated manner. Molecular dynamics simulations recently conducted by Faraldo-Gómez et al. [27] have established that FhuA loop L8 exhibits a significant degree of structural drift over a 10 ns simulation time-frame. The authors suggested that motion of FhuA loop L8 may transmit a destabilizing effect on the switch helix via FhuA β15. If HgbA L8 (perhaps in coordination with L3) were to play a similar role, it would imply that hemoglobin binding to HgbA, not heme interaction, may provide the signal that facilitates HgbA's interaction with TonB.

Two HgbA extracellular loops share significant sequence homology and fold similarity to secreted proteins that interact with proteins in a host's blood plasma. HgbA loop L5 shares significant homology and fold similarity to a loop in the α-subunit of snake venom protein flavocetin-A. Flavocetin-A circulates in the host's blood plasma and complexes with platelet glycoprotein Ib, abrogating its ability to interact with von Willebrand factor [28]. HgbA loop L10 shares similarities to a loop in streptococcal superantigen Smez-2 which binds to the hosts MHC-II subunit, disrupting its interactions with T-cell receptors [29]. These similarities implicate HgbA

L5 and L10 in interacting with hemoglobin in the host's blood plasma.

HgbA loop L7 aligns with 22% identity and has fold similarity to a loop in the human lipocalin protein HNGAL, a human blood granulocyte component that forms a complex with matrix metalloproteinase-9 [30]. Lipocalins are proteins that transport small hydrophobic molecules, and they possess a β barrel domain comprised of eight antiparallel β strands. Interestingly, it has recently been shown [31] that the eukaryotic α-1 microglobulin, a lipocalin in blood plasma, binds hemoglobin and may be involved in its catabolism. Additionally, the terminus of HgbA β13 contains the consensus FRAP box common to heme-utilizing and hemoglobin-binding receptors [32,33]. The entry to HgbA β14 contains a sequence homologous to the NPNL box in HemR [HemR: (469)NPN**LK**A**E**KAKNW**E**(481); HgbA: (675)LP-NVD**LKSE**RAETK**E**(689)]. Positioned between the FRAP and NPNL boxes of this class of proteins is a conserved histidine that is known to be involved in hemoglobin binding/heme utilization. In HemR, this residue is His461. HgbA His669 aligns with HemR His461 and is found in the centre of predicted extracellular loop L7. Based on the observation that HemR His461Leu and His461Ala mutants were unable to utilize hemoglobin, myoglobin, or human serum albumin heme as iron sources, Bracken et al. [32] proposed that HemR His461 is contained within a surface-exposed high-affinity hemoglobin/heme-binding site. These observations strongly support our assignment of L7 as a predicted hemoglobin-binding extracellular loop.

Analyses of sequence alignments and structural similarities between BtuB, FepA, FhuA, and FecA and our three-dimensional model of HgbA reveal that the barrel domains of these proteins can be divided into two subfamilies, with each subfamily having subtle yet distinct sequence and structural homologies. The cork domains, however, cannot be clearly divided into subfamilies. Overall, they are more structurally heterogeneous while having a higher degree of sequence homology as compared to the barrel domains. These differences suggest that there was divergent evolution of gene sequences encoding cork and barrel domains, and that the genes encoding TonB-dependent OM receptors may be products of gene fusion events.

Our three-dimensional model of HgbA is most structurally similar to BtuB. The predicted HgbA barrel domain has overall dimensions of 44 Å × 28 Å × 49 Å, which compare favorably with the known dimensions [6] of the BtuB barrel (45 Å × 30 Å × 55 Å). Chimento et al. suggested that the overall architecture of BtuB may be influenced by its function to transport cyanocobalamin to the periplasm. Cyanocobalamin is the largest ligand compared to those of FepA (enterobactin), FhuA (ferrichrome), and FecA (ferric citrate). Although heme is somewhat smaller than cyanocobalamin, it is most related to this molecule in comparison to enterobactin, ferrichrome, and ferric citrate. The porphyrin ring system which binds the iron of heme is chemically similar to the corrin ring system which binds the cobalt of cyanocobalamin, the major difference being the higher degree of bond conjugation in the porphyrin moiety. It is therefore possible that there is a correlation between the overall dimensions of the HgbA and BtuB barrels and the chemical nature of the ligands transported through them. In support of this, we have shown that HgbA predicted extracellular loop L6 has a high degree of fold similarity and sequence homology to a loop in the cyanocobalamin-binding protein diol dehydratase from *K. oxytoca* (Table 5). More compellingly, Chimento et al. [6] identified five extracellular loops in BtuB that interact with cyanocobalamin. These loops map to HgbA extracellular loops L3, L4, L9, L10, and L11; we demonstrate (Tables 5 and 6) that L4, L9, and L11 have fold similarity and sequence homology to loops in target proteins involved in either heme or iron binding.

5. Conclusions

We exploited comparative modelling and Hidden Markov Modelling to assign secondary structure identities to 496 residues of 923 residues in the mature HgbA protein. The remaining 421 residues excluding HgbA residues 1–6 were assigned to extracellular loop regions and we have shown that these predicted regions likely function to bind hemoglobin or bind heme. That 10 out of 11 of these loops (L1 being too short to provide meaningful sequence information) were found by mGenTHREADER to have fold similarity to loop regions in target proteins involved in heme binding, iron binding, and protein–protein interactions, supports the topology of our model. Comparative modelling has facilitated our preliminary understanding of extracellular loop functions in HgbA. Although automated techniques aided our initial topology determination, we demonstrate that manual intervention is still required to obtain a representative model of a protein-binding TonB-dependent OM receptor. The model that we propose for the barrel domain of HgbA is of immediate use in our continuing explorations of its structural biology.

Acknowledgments

This work was initially supported by a Strategic Grant (224192) from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Recent funding was from the NSERC Canadian Research Network on Bacterial Pathogens of Swine and from its industrial partners. An award to the Montreal Integrated Genomics Group for Research on Infectious Pathogens by Canada Foundation for Innovation provided infrastructure for molecular modelling: SYBYL from Tripos and hardware from IBM. This project was launched when PDP participated in the Bologna Winter School on Biotechnology: Structural Genomics, February 2003; we acknowledge discussions and suggestions of R. Casadio and P.L. Martelli.

References

[1] B. Fenwick, S. Henry, Porcine pleuropneumonia, J. Am. Vet. Med. Assoc. 204 (1994) 1334–1340.

[2] J.T. Bosse, H. Janson, B.J. Sheehan, A.J. Beddek, A.N. Rycroft, S.J. Kroll, P.R. Langford, *Actinobacillus pleuropneumoniae*: pathobiology and pathogenesis of infection, Microbes. Infect. 4 (2002) 225–235.

[3] H.G. Deneer, A.A. Potter, Effect of iron restriction on the outer membrane proteins of *Actinobacillus (Haemophilus) pleuropneumoniae*, Infect. Immun. 57 (1989) 798–804.

[4] C.G. D'Silva, F.S. Archibald, D.F. Niven, Comparative study of iron acquisition by biotype 1 and biotype 2 strains of *Actinobacillus pleuropneumoniae*, Vet. Microbiol. 44 (1995) 11–23.

[5] R. Srikumar, L.G. Mikael, P.D. Pawelek, A. Khamessan, B.F. Gibbs, M. Jacques, J.W. Coulton, Molecular cloning of hemoglobin-binding protein HgbA in the outer membrane of *Actinobacillus pleuropneumoniae*, Microbiology 150 (2004) 1723–1734.

[6] D.P. Chimento, A.K. Mohanty, R.J. Kadner, M.C. Wiener, Substrate-induced transmembrane signaling in the cobalamin transporter BtuB, Nat. Struct. Biol. 10 (2003) 394–401.

[7] S.K. Buchanan, B.S. Smith, L. Venkatramani, D. Xia, L. Esser, M. Palnitkar, R. Chakraborty, D. van der Helm, J. Deisenhofer, Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*, Nat. Struct. Biol. 6 (1999) 56–63.

[8] A.D. Ferguson, E. Hofmann, J.W. Coulton, K. Diederichs, W. Welte, Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide, Science 282 (1998) 2215–2220.

[9] A.D. Ferguson, R. Chakraborty, B.S. Smith, L. Esser, D. van der Helm, J. Deisenhofer, Structural basis of gating by the outer membrane transporter FecA, Science 295 (2002) 1715–1719.

[10] P.L. Martelli, P. Fariselli, A. Krogh, R. Casadio, A sequence-profile-based HMM for predicting and discriminating beta barrel membrane proteins, Bioinformatics 18 (Suppl. 1) (2002) S46–S53.

[11] I.N. Shindyalov, P.E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, Protein Eng. 11 (1998) 739–747.

[12] P.A. Bates, L.A. Kelley, R.M. MacCallum, M.J. Sternberg, Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM, Proteins Suppl. 5 (2001) 39–46.

[13] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.

[14] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Cryst. 26 (1993) 283–291.

[15] G. Vriend, WHAT IF: a molecular modeling and drug design program, J. Mol. Graph. 8 (1990) 52–56.

[16] G.J. Kleywegt, T.A. Jones, Detecting folding motifs and similarities in protein structures, Meth. Enzymol. 277 (1997) 208–230.

[17] M. Levitt, M. Gerstein, A unified statistical framework for sequence comparison and structure comparison, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 5913–5920.

[18] D.T. Jones, GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences, J. Mol. Biol. 287 (1999) 797–815.

[19] W.C. Wimley, Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures, Protein Sci. 11 (2002) 301–312.

[20] C.S. Raman, H. Li, P. Martasek, V. Kral, B.S. Masters, T.L. Poulos, Crystal structure of constitutive endothelial nitric oxide synthase: a paradigm for pterin function involving a novel metal center, Cell 95 (1998) 939–950.

[21] C.R. Lancaster, A. Kroger, M. Auer, H. Michel, Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution, Nature 402 (1999) 377–385.

[22] E. Parisini, F. Capozzi, P. Lubini, V. Lamzin, C. Luchinat, G.M. Sheldrick, Ab initio solution and refinement of two high-potential iron protein structures at atomic resolution, Acta Crystallogr. D. Biol. Crystallogr. 55 (1999) 1773–1784.

[23] N. Shibata, J. Masuda, T. Tobimatsu, T. Toraya, K. Suto, Y. Morimoto, N. Yasuoka, A new mode of B12 binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase, Struct. Fold. Des. 7 (1997) 997–1008.

[24] C.A. Genco, D.W. Dixon, Emerging strategies in microbial haem capture, Mol. Microbiol. 39 (2001) 1–11.

[25] M.A. Hanson, R.C. Stevens, Cocrystal structure of synaptobrevin-II bound to botulinum neurotoxin type B at 2.0 Å resolution, Nat. Struct. Biol. 7 (2000) 687–692.

[26] H. Li, J.J. Dunn, B.J. Luft, C.L. Lawson, Crystal structure of Lyme disease antigen outer surface protein A complexed with an Fab, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 3584–3589.

[27] J.D. Faraldo-Gómez, G.R. Smith, M.S. Sansom, Molecular dynamics simulations of the bacterial outer membrane protein FhuA: a comparative study of the ferrichrome-free and bound states, Biophys. J. 85 (2003) 1406–1420.

[28] K. Fukuda, H. Mizuno, H. Atoda, T. Morita, Crystal structure of flavocetin-A, a platelet glycoprotein Ib-binding protein, reveals a novel cyclic tetramer of C-type lectin-like heterodimers, Biochemistry 39 (2000) 1915–1923.

[29] V.L. Arcus, T. Proft, J.A. Sigrell, H.M. Baker, J.D. Fraser, E.N. Baker, Conservation and variation in superantigen structure and activity highlighted by the three-dimensional structures of two new superantigens from *Streptococcus pyogenes*, J. Mol. Biol. 299 (2000) 157–168.

[30] M. Coles, T. Diercks, B. Muehlenweg, S. Bartsch, V. Zolzer, H. Tschesche, H. Kessler, The solution structure and dynamics of human neutrophil gelatinase-associated lipocalin, J. Mol. Biol. 289 (1999) 139–157.

[31] M. Allhorn, T. Berggard, J. Nordberg, M.L. Olsson, B. Akerstrom, Processing of the lipocalin alpha(1)-microglobulin by hemoglobin induces heme-binding and heme-degradation properties, Blood 99 (2002) 1894–1901.

[32] C.S. Bracken, M.T. Baer, A. Abdur-Rashid, W. Helms, I. Stojiljkovic, Use of heme-protein complexes by the *Yersinia enterocolitica* HemR receptor: histidine residues are essential for receptor function, J. Bacteriol. 181 (1999) 6063–6072.

[33] E.R. Murphy, R.E. Sacco, A. Dickenson, D.J. Metzger, Y. Hu, P.E. Orndorff, T.D. Connell, BhuR, a virulence-associated outer membrane protein of *Bordetella avium*, is required for the acquisition of iron from heme and hemoproteins, Infect. Immun. 70 (2002) 5390–5403.

[34] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Cryst. 24 (1991) 946–950.