Protein Science (1995), *4*:2335-2340. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

Modeling of the structure of the *Haemophilus influenzae* heme-binding protein suggests a mode of heme interaction

PETE DUNTEN AND SHERRY L. MOWBRAY
Department of Molecular Biology, BioMedical Center, Swedish University of Agricultural Sciences, 75124 Uppsala, Sweden

(RECEIVED July 12, 1995; ACCEPTED August 30, 1995)

Abstract

The structure and function of the periplasmic heme-binding protein HbpA of *Haemophilus influenzae* were investigated. This protein is involved in the import of heme into the bacteria through the inner membrane, and thus is a key element of the organism's ability to survive in blood. A high degree of sequence similarity between HbpA and the dipeptide-binding protein of *Escherichia coli* is suggested to be the result of a functional relationship. An HbpA model built using the dipeptide-binding protein suggests a mode of heme binding that is distinct from those known in proteins of the human host. These results provide a starting point for rational drug design.

Keywords: heme; modeling; periplasmic binding protein; protein structure; transport

Haemophilus influenzae is an important cause of meningitis, as well as a major agent of lower respiratory tract and other diseases. The organism has an absolute requirement for exogenously supplied heme, the import of which utilizes a periplasmic lipoprotein, HbpA (Hanson & Hansen, 1991). This heme(haemin)-binding protein is the primary receptor for a heme transport system in the cytoplasmic membrane and, as a key element of the organism's ability to survive in blood, is a desirable target for drugs that would inhibit heme transport. As part of a transport system, it could also be used as a means of selective delivery of toxic compounds into the bacterial cell. The sequence is strongly conserved among various *Haemophilus* strains (Hanson et al., 1992). A similar protein has also been found in *Yersinia enterocolitica* (Stojiljkovic & Hantke, 1994) and proposed to exist in *Neisseria meningitidis* (Stojiljkovic et al., 1995), other heme-requiring pathogens. These facts suggest that drugs aimed at HbpA might be generally useful for a variety of disease-causing bacteria. The success of such a strategy depends largely on the ability to design drugs that do not interfere with the normal roles of heme in the host, and thus also on a knowledge of the modes of heme binding in the protein structures involved. The X-ray structures of the globins and a number of cytochromes and peroxidases are known, but no structure of a bacterial periplasmic heme-binding protein has yet been determined.

One of the more intriguing reports concerning HbpA has been one demonstrating a very high level of sequence identity (52%) between the preprotein and that of the *Escherichia coli* dipeptide-binding protein (DBP; Hanson et al., 1992). DBP is a 507-residue periplasmic protein involved in chemotaxis toward, and transport of, many dipeptides and some tripeptides (Manson et al., 1986; Abouhamad et al., 1991; Olson et al., 1991). This protein has also been shown to be related to periplasmic receptors for oligopeptides (Guyer et al., 1985; Tame et al., 1994) and nickel (Navarro et al., 1993). It was particularly striking that proteins so closely related in sequence were implicated in the transport of ligands with such different structures. With the recent solution of the structure of DBP (Dunten & Mowbray, 1995), it became possible to investigate the significance of these findings in more detail. Some of the results are presented in this report, along with a model of HbpA.

Results

Sequence comparisons

The mature sequence of HbpA contains 529 residues and that of DBP contains 507 residues. The region matching in the two proteins begins at residue 22 in HbpA and terminates 2 residues

Reprint requests to: Sherry L. Mowbray, Department of Molecular Biology, BioMedical Center, Box 590, Swedish University of Agricultural Sciences, 75124 Uppsala, Sweden; e-mail: mowbray@xray.bmc.uu.se.

Abbreviations: HbpA, the heme-binding protein of *Haemophilus influenzae*; DBP, the dipeptide-binding protein of *Escherichia coli.*

from the end of its sequence. There is a single one-residue deletion in the HbpA sequence (at residue 447 of HbpA, and 426 of DBP). Of 506 equivalent residues (covering the entire range of the sequence and X-ray structure of DBP), there are only 227 differences; that is, 55% of the residues are identical. To put this level of sequence conservation into perspective, a number of other comparisons were carried out.

First, the range of sequence conservation found for proteins known to be homologous in *H. influenzae* and *E. coli* was established. Eight proteins of clearly identical function were available for both organisms in Swiss-Prot. The average value of sequence similarity in this set was 67%, ranging from 41% for a disulfide interchange protein to 92% for the 30 s ribosomal protein, S12.

Next, the level of similarity was investigated where two periplasmic receptors interact with the same membrane component as part of their function.

An interaction with a common membrane-bound chemoreceptor is rather permissive. The *E. coli* ribose- and glucose/galactose-binding proteins interact with distinct transport systems, but with the same membrane chemoreceptor. These proteins have a sequence identity of 24%, which is only slightly higher than the similarity with a related arabinose-binding protein that shares no functional partners (19 and 20%, respectively). The surfaces of the two structures known to be important in chemotaxis are similar, but not identical (Mowbray, 1992; Björkman et al., 1994). In addition, the nickel-binding protein has been reported to mediate chemotaxis through the same chemoreceptor as the maltose-binding protein (De Pina et al., 1995), although their structures are almost completely unrelated. Thus, an interaction with a common chemoreceptor imposes little, if any, sequence identity.

An interaction of two periplasmic proteins with the same inner membrane permease places more demands on the level of sequence conservation. The *Salmonella typhimurium* histidine- and lysine/arginine/ornithine-binding proteins bind the same transporter, whereas the leucine- and leucine/isoleucine/valine-binding proteins both bind another. Members of these two pairs have 67% and 76% identity to their functional partner, respectively. The pairs are, however, structurally unrelated after the first 60 residues or so, with a sequence similarity of only 14–15% in the equivalent portions of members of different pairs. These results suggest that a rather high level of conservation is a general requirement where a common permease is involved.

The closest relatives of DBP in *E. coli* are the oligopeptide- (Guyer et al., 1985; Tame et al., 1994) and nickel-binding (Navarro et al., 1993) proteins, each of which interacts with a different membrane permease. Only the nickel-binding protein has been implicated in chemotaxis, sending signals through a different membrane chemoreceptor than DBP (De Pina et al., 1995). The sequence similarities of these related proteins are summarized in Table 1. The similarity of the DBP-HbpA pair is clearly much greater than for any other. The available sequences of the relevant membrane components are found to be related at a similar level (those for heme were not available).

Table 1. *Sequence similarity (%) between various binding proteins of the HbpA family*[a]

Protein	Heme	Nickel	Oligopeptide
Dipeptide	54.4	21.4	17.8
Heme	–	20.8	18.8
Nickel	–	–	16.2

[a] Only the regions of the proteins covered by the mature DBP sequence were used. Analysis was carried out and results defined as described in the text.

Modeling the structure of HbpA

A model was built for HbpA using the structure of DBP as a starting point (available from the Brookhaven Protein Data Bank with identity code 1dpp; Bernstein et al., 1977). The 21 amino acids at the N-terminus of mature HbpA were not included in the model because they have no counterparts in DBP; they presumably protrude from the rest of the protein and provide a point of attachment for the lipid by which it is bound to the cytoplasmic membrane. The aligned sequences of the remaining portion (506 residues) were entered into the program SOD (G. Kleywegt, unpubl.), which produced both coloring blocks for the graphics program O and a macro for generating an initial HbpA model (using the "mutate" option of O). The side-chain conformations for this model were either taken from the DBP structure, where the residues were of identical type, or assigned initially to the most common conformation in the list of Ponder and Richards (1987), where the residue types differed. The result was then viewed at the graphics terminal together with the structures of DBP and the oligopeptide-binding protein (Protein Data Bank codes 1ola and 1olb; Tame et al., 1994), and checked on a residue-by-residue basis. Side-chain conformations were changed in some cases to obtain correct packing in the protein cores, using the rotamer library and giving preference to conformations similar to, or identical with, those seen in the known binding protein structures.

The resulting model was entirely plausible; no major changes were needed in the main chain to accommodate the altered sequence, and the types of residues were universally consistent with their predicted locations. Hydrogen bonding partners for the new side chains were reasonable and widespread. The single one-residue deletion in the HbpA sequence was found in a surface loop (Fig. 1). This model structure was then energy minimized in X-PLOR. There were very few initial clashes and the refinement rapidly converged to an energetically favorable final structure. The resulting model showed an RMS difference (RMSD) of 0.61 Å from the α-carbons of the initial model (and therefore, from DBP), and a difference of 0.80 Å using all atoms. These values are consistent with the expected structural similarity between DBP and HbpA (Chothia & Lesk, 1986) and confirm that the model is physically reasonable. A similar energy minimization of the original DBP model without crystallographic restraints resulted in an RMSD of 0.42 Å for α-carbons only and 0.60 Å for all atoms.

The residues conserved in DBP and HbpA are located throughout the structures, at positions buried, exposed, and in the binding site with approximately equal frequency (Fig. 1). By analogy to other binding proteins, the region where the end strands of domains I and III approach seems most likely to be involved in interactions with the membrane permeases. This portion of the surface is very well conserved.

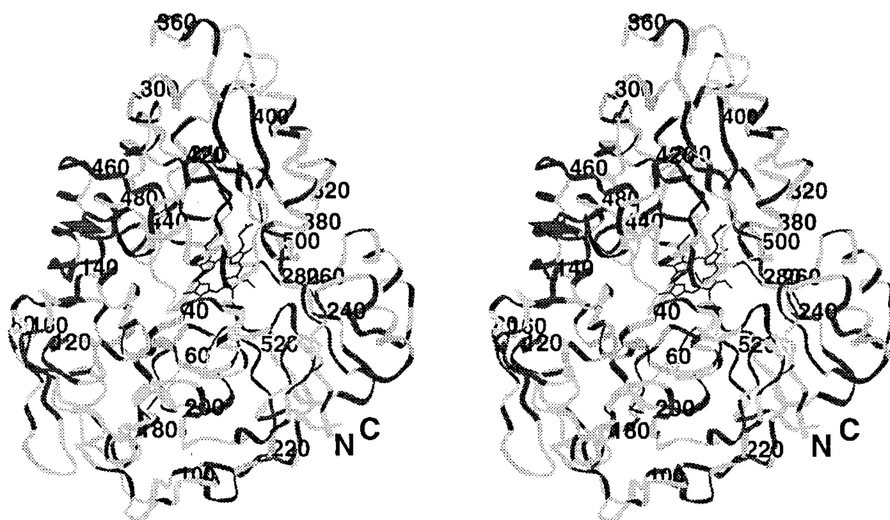

Fig. 1. The C_α-backbone of the HbpA model, with every 20th residue numbered; only the residues covered by the model are shown. The approximate dimensions of the molecule are 75 × 55 × 30 Å. Residues that are the same in HbpA and DBP are shaded lighter than those that are different; the location of the deletion in HbpA is indicated in black. Residues of domain I are 22–54, 204–281, and 499–529; those of domain II are 55–202; and those of domain III are 282–498, as defined in Dunten and Mowbray (1995).

Modeling the interaction of HbpA with heme

The shape and volume of the predicted binding site in the initial model of HbpA is shown in Figure 2A. The site is quite flat and rectangular in shape, with an almost perfect match to the size and shape of a heme. There are two, possibly three, small channels at the corners of the site by which water could approach it, but no path large enough to allow entry or exit of the ligand itself. Thus, HbpA, like DBP, must undergo a conformational change at the hinge connecting the two halves of the protein in association with ligand binding. It is also clear that HbpA should indeed bind heme or protoporphyrin, rather than some breakdown product of either, a fact that was not unambiguously established in earlier studies of HbpA using cell lysates (Hanson & Hansen, 1991).

The DBP binding site, although located in the same place, is narrower, with an elongated shape suitable to binding a di- or tripeptide, and pockets of ill-defined shape into which many types of amino acid side chains can comfortably fit (Dunten & Mowbray, 1995). Two residues seem to be most important in making additional space for the heme: where DBP has Pro 356 and Tyr 357, Ala 377 and Ser 378, respectively, are found in HbpA.

The shape of the HbpA binding site limits the possible orientations of a heme ligand to four with a common heme plane. The nature of the amino acid side chains lining the site strongly suggests that one of these is more likely than the others (shown in Fig. 2B). One propionic acid moiety of the heme would thus form hydrogen bonds with the side-chain amide nitrogen of Gln 372 and possibly Arg 411, Tyr 410, or Ser 449, whereas the other would interact with Asn 379 and possibly Ser 378, depending on the precise structure of the actual complex. In other heme orientations, most or all of these interactions are lost, and unfavorable interactions and clashes occur. Additional nonpolar residues, including tyrosine, tryptophan, phenylalanine, isoleucine, and methionine, line the site in an appropriate fashion to provide van der Waals interactions with the rest of the heme group. The main-chain peptides linking residues 425 to 427 also appear perfectly placed to provide π-stacking interactions with two of the pyrrole rings. The equivalent residues hydrogen bond to the ligand backbone in DBP.

The proposal that one of the propionic acid moieties of the heme is located in a solvent channel close to the protein surface is consistent with the report that HbpA binds (although somewhat poorly) to hemin-agarose (Hanson & Hansen, 1991), in which heme is bound to agarose through a two-carbon spacer arm linked to the propionic acid group.

The form of heme ligand preferred by HbpA is not known. Heme free in the blood would be expected to be in the ferric state, which cannot bind oxygen, but because *Haemophilus* obtains its heme through a specific outer membrane system that literally robs the host's heme proteins (Cope et al., 1994), oxygen-bound ferrous heme seems a possibility. In theory, oxygen or water could bind on either side of the iron. In the orientation shown in Figure 2B, an oxygen or water bound to the heme could interact with Arg 376, which is in turn supported by Thr 44 and Trp 407. An additional interaction with oxygen could be provided by the main-chain amide nitrogen of residue 43, which is located at the N-terminal end of a helix. There is no apparent sixth coordination ligand for the heme iron on the opposite side of the protoporphyrin ring, nor is it obvious how the structure could be altered to place any of the usual heme-iron ligands (histidine, methionine, and asparagine) in an appropriate position. The lack of a sixth coordination ligand might help bring the K_d of HbpA for heme to the range of 10^{-6}–10^{-7} M expected for a periplasmic binding protein, rather than the tighter association found for such proteins as myoglobin and hemopexin. Deoxy heme is, of course, also a possibility, although the arginine present would be an unusual fifth coordination ligand to the iron. The future availability of purified HbpA will enable the necessary binding studies to be carried out. Energy minimization of the HbpA structure with oxygenated heme in the position shown produces a model very similar to the initial one, with excellent complementarity in the binding site, and no steric clashes.

Like DBP, HbpA seems to utilize primarily residues from domains I and III (defined as Dunten & Mowbray, 1995; see Fig. 1) in the binding of ligand. Although approximately half of the residues implicated in heme binding are, in fact, conserved in DBP, few of their roles in ligand binding appear to be maintained. The proposed orientation of the heme would place an oxygen or water bound to the heme in the same location as the

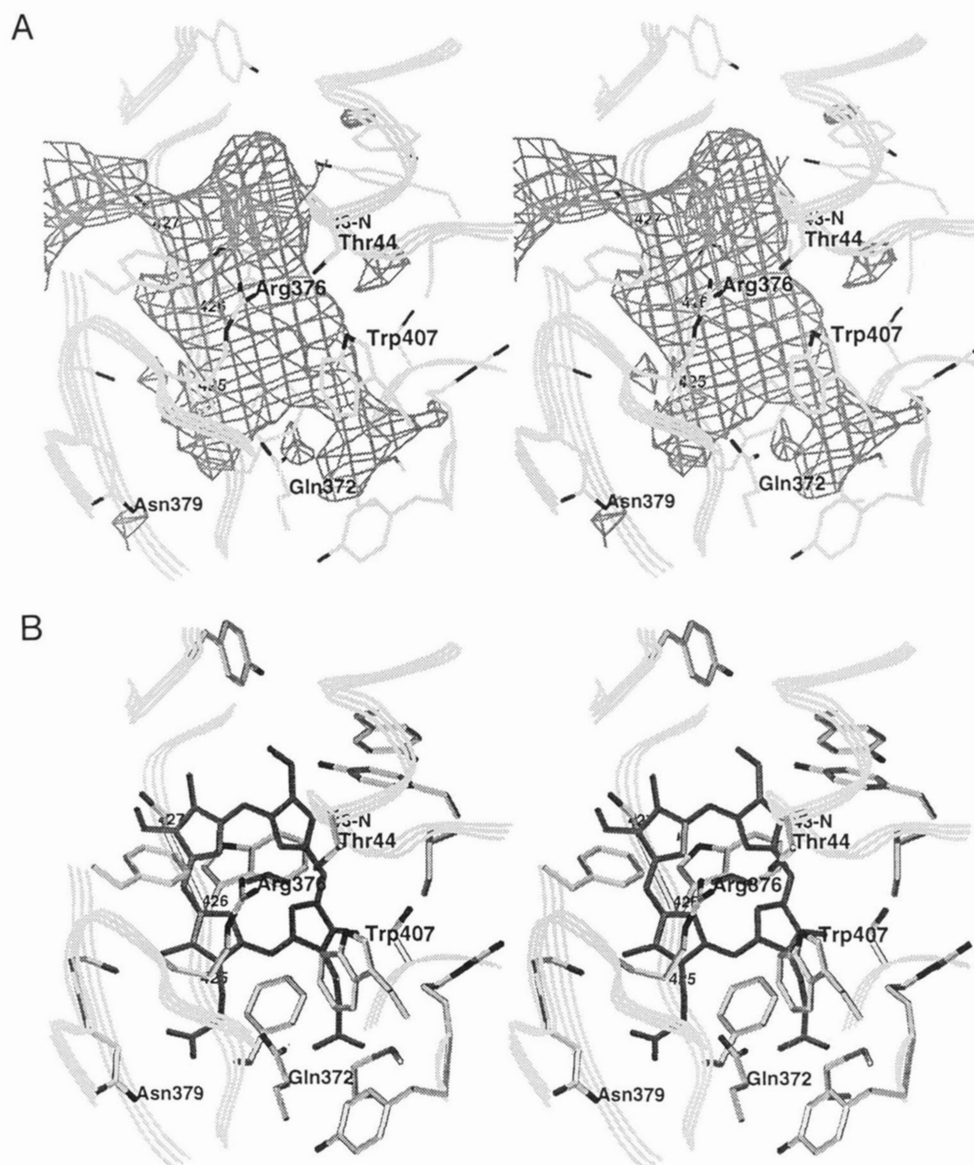

Fig. 2. The proposed binding site of the HbpA model. **A:** The molecular surface is contoured using the van der Waals radii of the protein atoms plus a distance of 1.2 Å, which has been shown to be a good representation of the size and shape of the ligand site in binding proteins (Zou et al., 1993). **B:** The proposed mode of heme binding; an oxygenated heme is shown by way of example. In both A and B, residues mentioned in the text as having critical interactions with heme are indicated, and nearby portions of the protein main chain are shown as ribbons.

C-terminus of a dipeptide bound to DBP, but no other "conserved" interactions are evident. The shape of the tripeptide bound to the oligopeptide-binding protein is, however, strongly reminiscent of heme in the proposed orientation. After structural alignment, one amino acid residue of the oligopeptide-binding protein's tripeptide ligand sits at the position of each of three pyrrole rings, with Tyr 269 closing off the fourth corner of the site.

Discussion

HbpA is part of a family of proteins binding a very wide variety of ligands. The level of sequence conservation is not in any way correlated with the type of ligand recognized. The similarity between DBP and the oligopeptide-binding protein is only 22%, yet these two proteins use similar strategies to bind very similar ligands. The same low level of sequence conservation is found between either DBP or HbpA and the nickel-binding protein, which obviously bind distinct ligands. By contrast, the much higher identity between HbpA and DBP (55% in the mature region of DBP and the corresponding region of HbpA), through selected changes in key residues, results in the binding of very different types of ligands.

The level of sequence identity between DBP and HbpA suggests that they interact with closely related permeases or have some other functional relationship; it is unlikely to arise if the

two proteins were completely unrelated. Wild-type *E. coli* does not require or transport exogenous heme (which is not normally found in the gut), but altered strains that are able to import heme into the periplasm can then use it as a source of both porphyrin and iron (Stojiljkovic et al., 1995). Thus, an HbpA-like binding protein may allow transport of heme across *E. coli*'s inner membrane, although the usual function of such a system is not yet apparent. DBP itself does not bind haemin to any appreciable extent (P. Dunten, unpubl. data) and does not have space for any heme-like molecule in its binding site, so it cannot fulfill the heme-transport function. By analogy, a dipeptide-binding protein may exist in *Haemophilus* as well. A further clue that the heme and dipeptide transport systems are related is the fact that DBP and the associated membrane permease proteins are induced under conditions where iron and oxygen become limiting (Abouhamad & Manson, 1994). A role for any *Haemophilus* binding protein in chemotaxis is not possible, because the organism is nonmotile.

Comparison of a number of sequences and crystallographic structures (Chothia & Lesk, 1986) has suggested that a 55% sequence identity is generally associated with a structural conservation of at least 90% of the core (i.e., nonloop) residues of a given protein pair, and an RMS for backbone atoms of roughly 0.8 Å. These facts suggested that a model of the HbpA protein derived by analogy to the structure of DBP would give a fairly accurate picture of HbpA.

The model presented here gives every sign of being physically reasonable, and furthermore suggests a plausible mode of heme binding that can be used for further work. The heme is proposed to be almost completely buried in the protein, and would presumably require a conformational change associated with binding. Other heme-bearing proteins show a wide variety of binding modes, none of which is similar to that proposed for HbpA. In myoglobin and hemoglobin, a histidine is found on one side of the iron, and oxygen can bind on the other; the propionic acid groups are exposed to solvent. In cytochromes, histidine and/or methionine ligands are found on both sides of the iron, and the propionic acid moieties are hydrogen bonded to various types of protein side chains. Modification of the propionic acid groups of heme might, for example, allow it to bind to HbpA but not to the cytochromes; binding to the globins might occur, but have no physiological consequences. The possibility of designing drugs that bind to HbpA but do not interfere with the normal use of heme by the organism's host, seems, therefore, to be a real one.

Methods

Sequences were obtained from Swiss-Prot (EMBL Data Library, Version 31, May, 1995). Sequence comparisons were carried out using the Lasergene series of programs from DNASTAR, Inc., with the Clustal method of alignment and a PAM250 residue weight table; values of sequence similarities are as reported by that program.

Building of a homology-based model was aided by the use of the program SOD (Gerard Kleywegt, University of Uppsala). Visualization and adjustment of the results were carried out with the molecular graphics program, O (Jones et al., 1990), with rotamers described by Ponder and Richards (1987). Energy minimization in the program X-PLOR (Brünger, 1992) utilized the parameter set of Engh and Huber (1991) and a dielectric constant of 10.0; 500 cycles of standard Powell minimization were used. Maps for the display of molecular surfaces were calculated using the program SURFAS (Voorintholt et al., 1989).

Acknowledgments

We thank Mark Harris and Eric Hansen for helpful discussions. This work was supported by a grant from the Swedish Natural Research Council to S.M.

References

Abouhamad WN, Manson MD. 1994. The dipeptide permease of *Escherichia coli* closely resembles other bacterial transport systems and shows growth-phase-dependent expression. *Mol Microbiol 14*:1077–1092.

Abouhamad WN, Manson M, Gibson MM, Higgins CF. 1991. Peptide transport and chemotaxis in *Escherichia coli* and *Salmonella typhimurium*: Characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein. *Mol Microbiol 5*:1035–1047.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Björkman AJ, Binnie RA, Cole LB, Zhang H, Hermodson MA, Mowbray SL. 1994. Identical mutations at corresponding positions in two homologous proteins with nonidentical effects. *J Biol Chem 269*:11196–11200.

Brünger AT. 1992. *X-PLOR: A system for X-ray crystallography and NMR.* New Haven, Connecticut: Yale University Press.

Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J 5*:823–826.

Cope LD, Thomas SE, Latimer JT, Slaughter CA, Mullereberhard U, Hansen EJ. 1994. The 100 kDa haem:haemopexin-binding protein of *Haemophilus influenzae*: Structure and localization. *Mol Microbiol 13*:863–873.

De Pina K, Navarro C, McWalter L, Boxer DH, Price NC, Kelly SM, Mandrand-Berthelot MA, Wu LF. 1995. Purification and characterization of the periplasmic nickel-binding protein NikA of *Escherichia coli* K12. *Eur J Biochem 227*:857–865.

Dunten P, Mowbray SL. 1995. Crystal structure of the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Protein Sci 4*:2327–2334.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A 47*:392–400.

Guyer CA, Morgan DG, Osheroff N, Staros JV. 1985. Purification and characterization of a periplasmic oligopeptide binding protein from *Escherichia coli*. *J Biol Chem 260*:10812–10818.

Hanson MS, Hansen EJ. 1991. Molecular cloning, partial purification, and characterization of a haemin-binding lipoprotein from *Haemophilus influenzae* type b. *Mol Microbiol 5*:267–278.

Hanson MS, Slaughter C, Hansen EJ. 1992. The *hbpA* gene of *Haemophilus influenzae* type b encodes a heme-binding lipoprotein conserved among heme-dependent *Haemophilus* species. *Infect Immunol 60*:2257–2266.

Jones TA, Bergdoll M, Kjeldgaard M. 1990. O: A macromolecular modeling environment. In: Bugg C, Ealick S, eds. *Crystallographic and modeling methods in molecular design*. New York: Springer-Verlag. pp 189–195.

Manson MD, Blank V, Brade G, Higgins CF. 1986. Peptide chemotaxis in *E. coli* involves the Tap signal transducer and the dipeptide permease. *Nature 321*:253–256.

Mowbray SL. 1992. The ribose and glucose/galactose receptors: Competitors in bacterial chemotaxis. *J Mol Biol 227*:418–440.

Navarro C, Wu LF, Mandrand-Berthelot MA. 1993. The *nik* operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. *Mol Microbiol 9*:1181–1191.

Olson ER, Dunyak DS, Jurss LM, Poorman RA. 1991. Identification and characterization of *dppA*, an *Escherichia coli* gene encoding a periplasmic dipeptide transport protein. *J Bacteriol 173*:234–244.

Ponder JW, Richards FM. 1987. Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol 193*:775–791.

Stojiljkovic I, Hantke K. 1994. Transport of haemin across the cytoplasmic membrane through a haemin-specific periplasmic binding protein-

dependent transport system in *Yersinia enterocolitica*. *Mol Microbiol 13*:719–732.

Stojiljkovic I, Hwa V, de Saint Martin L, O'Gaora P, Nassif X, Heffron F, So M. 1995. The *Neisseria meningitidis* haemoglobin receptor: Its role in iron utilization and virulence. *Mol Microbiol 15*:531–541.

Tame J, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF, Wilkinson AJ. 1994. The structural basis of sequence-independent peptide binding by OppA protein. *Science 264*:1578–1580.

Voorintholt R, Kosters MT, Vetger G, Vriend G, Hol WGJ. 1989. A very fast program for visualizing protein surfaces, channels and cavities. *J Mol Graph 7*:243–245.

Zou J, Flocco MM, Mowbray SL. 1993. The 1.7 Å refined X-ray structure of the periplasmic glucose/galactose receptor of *Salmonella typhimurium*. *J Mol Biol 233*:739–752.