Minireview

Bacterial Periplasmic Binding Protein Tertiary Structures

Mark D. Adams and Dale L. Oxender
From the Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109-0606

The periplasmic space of Gram-negative bacteria such as Escherichia coli and Salmonella typhimurium contains a large variety of proteins with a wide range of functions. These proteins are associated with nutrient metabolism, transport, chemotaxis, antibiotic resistance, and energy utilization. There are about a dozen or so periplasmic proteins that participate in transport of small polar substrates across the bacterial inner membrane. These proteins serve to bind the solute in the periplasm and mediate the transfer of the solute across the inner membrane. They are easy to isolate and purify from the periplasmic space using osmotic shock techniques. They have been extensively characterized by the work of many investigators, and it is now becoming clear that they share a great deal of similarity in mechanism and structure.

E. coli and S. typhimurium also contain osmotic shock-insensitive permeases which are generally constitutive low affinity systems comprised of a single membrane protein that both recognizes and transports the substrate from the periplasmic space to the cytosol as well as couples the system to metabolic energy. The osmotic shock-sensitive transport systems are more complex, consisting in each case of at least one soluble periplasmic protein, which is the component for substrate recognition, and two to four inner membrane proteins that play a more direct role in the transfer of the solute across the outer membrane. Reviews of shock-sensitive transport systems describing the genetics and biochemistry of these systems have appeared recently (1-3).

Biochemistry

The osmotic shock-sensitive systems have a number of genetic and biochemical characteristics in common in addition to the structural characteristics of the periplasmic binding proteins which will be the main focus of this review. For example, all of these permease systems exhibit a $K_a$ for transport of approximately 0.1-1.0 $\mu$M. These kinetic characteristics can essentially be ascribed to the properties of the binding proteins, which exhibit $K_a$ values of approximately 0.1-0.5 $\mu$M. The periplasmic components are generally present in large excess over the membrane components, and yet they seem to constitute the rate-limiting step for transport (2, 4).

Each binding protein interacts with a unique set of membrane-bound proteins that catalyzes the actual membrane translocation of the substrate. The genes for the two- to four-membrane components are usually located immediately downstream from one of the binding protein genes on the bacterial chromosome and are often transcribed from the same promoter (2). Membrane components are usually of two types, quite hydrophobic presumably integral membrane proteins that share similar hydrophathy profiles and less hydrophobic proteins with a wide range of functions. These proteins are associated with nutrient metabolism, transport, chemotaxis, antibiotic resistance, and energy utilization.

In contrast to the shock-insensitive systems which require the proton-motive force for energization, shock-sensitive systems seem to require ATP as the energy source. Several recent publications have greatly added to our understanding of the requirements for energization of the transport process (5-8). By examining the effect of a variety of factors including ATP analogs, potential alternate energy sources, agents which dissipate the electrochemical potential, and ATPase defective (unc+) strains, it was determined that ATP is the immediate energy donor for transport and that the direct involvement of the transmembrane electrochemical potential can be excluded. In addition, in reconstituted systems consisting of right-side-out or inside-out membrane vesicles, it was possible to estimate the $K_a$ of the histidine-binding protein for the membrane complex to be $\approx 85$ $\mu$M (7) and of the complex for ATP to be $\approx 200$ $\mu$M (6). Clearly, these reconstituted vesicle systems provide great potential for the examination of the interactions of various transport components.

Structural Similarities

The periplasmic binding proteins have proved convenient for structural study because of the ease of purification from the periplasmic space and their relative stability. A great deal of structural information has been accumulated, and it is now possible to draw broad conclusions about this group of proteins that is related by overall structure and function. Florante Quiocho and co-workers have provided the detailed structural information which has made these comparisons and interpretations possible (9-13). In Table 1 a comparison is made of some of the structural and functional properties of nine periplasmic binding proteins for which three-dimensional structures are or soon will be available.

Despite a striking similarity in function, size, and shape (described below), the periplasmic binding proteins generally have only very small detectable primary sequence homology to one another. The different carbohydrate-binding proteins, however, show some primary sequence relatedness to one another and to other proteins that bind carbohydrates (particularly the lac repressor) (14-16), even though specific amino acid residues that participate in sugar binding are not identical (17). Homology between LIV-BP and Mal-BP, on the other hand, is about 10%, well below statistical significance.

The structural similarity among the periplasmic binding proteins is striking. The following comparisons are generalizations that have been drawn from an examination of the structural properties of several binding proteins. Fig. 1 is a stereo diagram of the $\alpha$-carbon backbone of a "typical" periplasmic protein, the LIV-BP, in its native conformation. It provides a visual representation of some of the information presented here. The overall dimensions of the protein are about 35 x 40 x 70 Å. They are elongated structures with an axial ratio of 2:1. There are two globular domains connected by three short peptide segments. Between the domains is a large cleft approximately 18 Å deep in which the substrate binds. The two domains are separated by as much as 18 Å when no ligand is present (18). Ligand binding induces a conformational change that closes the cleft so that the ligand is completely sequestered from the solvent (9). For instance, analysis of arabinose bound to Ara-BP indicates an exposed surface area only 2.3% of that observed for the free sugar in the conformation observed in the crystals (19).

The two domains, designated N and C for the location of the amino and carboxyl termini of the polypeptide chain, are composed of residues from both the amino and carboxyl portions of the chain. The first third of the coding sequence forms the bulk of the N domain; the second third provides the majority of the C domain. The remaining residues of the polypeptide chain cross back and forth between the two domains. The N domain is generally larger than the C domain (18S versus 140 residues for the LIV-BP).

Each domain is constructed of a central parallel $\beta$-sheet surrounded on each side by two or three $\alpha$-helices. Thus in addition to structural similarities between proteins, the supersecondary structure of each domain is very similar, resulting in a pseudo 2-fold axis of symmetry within the protein. The $\alpha/\beta$ topology observed for the binding proteins is somewhat reminiscent of the nucleotide fold motif that has been described for many dehydrogenases and kinases (20). The $\beta$-sheet is...
arranged B A C D E with all "right-handed" connections (21). The α-helices run antiparallel to the sheets so that the carboxyl termini of the β-strands and the amino termini of the α-helices converge at the cleft.

A typical packing diagram depicting one domain of a binding protein is shown in Fig. 2. Squares represent β-strands and circles represent α-helices. The strand is viewed perpendicular to the plane of the paper from the amino termini of the strands toward the cleft. Shaded symbols represent secondary structure elements that are present in both domains of all binding proteins so far examined. The strand occupying position D in both domains of the sulfate-binding protein is strand J, which is antiparallel to the rest of the sheet. Also, helix IV (the structural element linking the two domains) is a β-strand, not a helix, in both domains of SO₄-BP. In half of the domains examined, the loop connecting strands D and E formed a stable α-helix. Two domains from each of four binding proteins (Ara-BP, SO₄-BP, Gal-BP, and LIV-BP) were used as models to construct the diagram. In addition to this core structure, there are one or two additional strands in the β-sheet of all the domains except the N and C domains of SO₄-BP. In all cases except the N domain of Ara-BP and Gal-BP, one strand is antiparallel to the sheet; it is always the last or next to last strand in the sheet.

The implications of this structural similarity are difficult to gauge. The similarity of structure, genetic organization, and metabolic function suggests that the binding protein-dependent transport systems may have all derived from a single evolutionary progenitor. Subsequent divergence allowed increased specificity for particular metabolites and individualized regulatory controls. During this divergence the binding proteins would be constrained to maintain their specificity for substrate, their interaction with a complex of membrane proteins of like specificity, and perhaps their interaction with a chemotaxis receptor. An alternative explanation for structural similarity in the absence of primary sequence homology is that this group of proteins has evolved with similar structures from different progenitors due to similar evolutionary pressures acting on transport systems in general. In either case, this group of proteins represents a remarkable example of physiological function dictating structural constraints.

**Unique Features**

Each of the five binding protein structures that has been solved to better than 2.5-Å resolution has yielded surprising, and in some cases enigmatic, information. For example, the 1.9-Å structure of the Gal-BP shows the presence of a bound Ca²⁺ atom; both substrate anomers of the Ara-BP and Gal-BP can be accommodated by their respective binding pockets through changes in the sugar conformation, not the protein conformation; the leucine-binding proteins have virtually identical tertiary structures despite divergence in sequence and specificity.

**Leucine-binding Proteins**—Detailed structural analysis has now been performed on both LIV-BP and the leucine-specific binding protein (LS-BP) (18, 22). The genes encoding these proteins are located adjacent to one another at minute 74.5 on the E. coli chromosome and apparently arose by some sort of gene duplication event. The two proteins are approximately 80% identical at the amino acid level and have essentially identical tertiary structures (the root mean square deviation of the 317 equivalent α-carbon residues is 0.61 Å). Unlike all of the other binding proteins, the structures of the two leucine-binding proteins were solved without bound substrate. In fact, in order to grow crystals suitable for x-ray diffraction, it was necessary to remove endogenous bound ligand that co-purifies with the binding proteins. A putative leucine binding site in the N domain of LIV-BP was identified by solving the structure obtained after soaking the unliganded crystals in leucine solution. This form of the protein, termed the "open liganded form" by Sack (22) provides insights into the mechanism of leucine binding and the molecular basis for the difference in specificity between these two proteins. It also opens the way for experiments to determine the importance and mechanism of the large conformational change presumed to occur upon binding of substrate.

Binding of the leucine zwitterion to LIV-BP crystals involves a complex hydrogen bond network, utilizing both peptide and side chain hydrogen bond donors and acceptors. All of the residues that contribute hydrogen bonds to the CO or NH groups of the leucine are conserved between the two binding proteins. However, three residues...
in the two proteins in the vicinity of the leucine side chain have diverged (18, 22). These proteins present a unique opportunity to study what forces and geometries control binding of α hydrophobic side chain. All other periplasmic binding proteins so far studied bind hydrophobic, if not ionic, substrates.

The interactions that stabilize and distinguish the hydrophobic side chains of leucine, isoleucine, and valine have no precedent among the other periplasmic binding proteins. The leucine-binding site must satisfy the conflicting requirements of conserved residues that participate in recognition of leucine and diverged residues that lead to additional recognition of isoleucine and valine by the LIV-BP (or prevent recognition by the LS-BP). It would appear from transport experiments that the LIV-BP is responsible for about 10 times as much leucine transport as the LS-BP, based on leucine transport in the presence of excess valine or isoleucine (25). It will be interesting to further explore the physiological significance of this differential expression from a structural as well as a metabolic perspective (24).

Calcium Atom Bound to Galactose-binding Protein—The presence of a bound calcium atom in the galactose-binding protein was unexpected, and as yet no function has been ascribed to that ligand (25). The calcium is located about 30 Å from the sugar-binding site and about 45 Å from the putative site of interaction with the chemotaxis receptor trg. There is strong evidence that this is not an artifact of the crystallization conditions, since no Ca²⁺-containing buffers were used during growth of the crystals and because of the strong sequence and structural homology to the Ca²⁺-binding loop of known Ca²⁺-binding proteins such as calmodulin (26). The physiological role of the calcium ion in the protein is not determined, but it may be involved in the process of interaction with the ligand, leading to possible conformational changes in the protein. However, the ribose-binding protein which shares the same calcium recognition sequence, trg, does not appear to contain a similar calcium-binding loop based on the protein sequence.

Binding of Both Sugar Anomers—Ara-BP, the first periplasmic binding protein to be solved, binds both α and β anomers of arabino with equal affinity (19). An aspartate at position 90 is aligned to accept a hydrogen bond from both the α and β anomic configurations of arabino. Asp-90 is in turn held firmly in position by a salt bridge with Lys-10. An extensive hydrogen bond network surrounds the sugar molecule. It is comprised of a dozen amino acid side chains, peptide backbone carboxyl and amino groups, and several bound water molecules. That each anomer is accommodated equally well by the binding site is in agreement with other evidence that both are bound with equal affinity in solution (4).

Interestingly, the pyranose ring does not assume identical conformations when present as the α and β anomers. Rather, the C4 atoms align precisely and differences are incrementally larger toward C1 so that the anomic hydroxyls are each positioned an equal distance from Asp-90. The geometry of the ligand binding site is unchanged; only the sugar changes conformation to achieve optimal interactions with the protein.

Recently the geometry of glucose binding at the ligand binding site of the Gal-BP was reported (17). Despite strong sequence conservation, the exact hydrogen bond interactions in Gal-BP are different from those in Ara-BP. β-Glucose is the only anom is that appears in the crystals, although model-building studies indicate that both anomers could be equally well accommodated. As is the case in the Ara-BP, an aspartate residue (Asp-104) is hydrogen bonded to the anomic hydroxyl in both the α- and β-configurations. Another aspartate (Asp-14) is responsible for binding of both pyranoid epimers.

Hydrogen Bonds in Uniquely Polarized Environments—Extensive hydrogen bond networks are emerging as a principal means of ligand binding for all of the binding proteins thus far analyzed. The extreme example of this is the sulfate-binding protein, where the bound SO₄²⁻ dianion is held solely by hydrogen bonds without countercharged residues or water molecules providing primary interactions (27). Five peptide backbone NH groups plus a serine hydroxyl and an indole NH are the principal ligands for the sulfate. A complex array of hydrogen and ionic bonds acts to polarize these hydrogen bonds, providing the necessary charge neutralization for binding. In the bound form of the protein, the sulfate is completely sequestered from the solvent. This presumably facilitates the degree of polarization that occurs in the pocket and ultimately the tightness of binding of the sulfate. As is the case with the sugar-binding proteins, the shell of residues providing direct hydrogen bonds is stabilized by secondary and tertiary arrays of hydrogen bonds and salt bridges that serve to align and orient the protein side chains and bound water molecules for specific interaction with the substrate. Primary hydrogen bonding interactions with sugar substrates, however, are accomplished solely by peptide side chains rather than the backbone groups. Quiocho et al. (10) have proposed that hydrogen bonding involving highly polarized peptide units may be a general method of stabilizing isolated charges such as those of the sulfate dianion.

Prospects

With five structures solved to high resolution and more on the way, the level of understanding of periplasmic binding protein structure is well advanced. Molecular details of the mechanism of substrate binding and the conformational changes associated with binding are beginning to be elucidated. Hydrogen bonds play a critical role in substrate recognition and binding; even highly charged substrates are bound completely by hydrogen bonds. Likewise the capacity of the peptide backbone to contribute stabilizing interactions in highly polarized environments has been more fully realized. The leucine-binding proteins will likely provide interesting clues to the nature of binding of a hydrophobic substrate and discrimination among hydrophobic substrates when the liganded structures are solved. The availability of both liganded and unliganded structures of one of the binding proteins will provide further clues as to the nature of the conformational change that accompanies substrate binding. The remarkable similarity in three-dimensional structures among this group of proteins may provide insight into the relationship between primary sequence, folding pathways, and final structure. Gribkov et al. (28, 29) have shown that it is possible to make three-dimensional structural predictions when structural information is available from a family of related structures. The solution of additional structures from this group should be aided by such an approach. Likewise, this family provides an excellent test case for the refinement of algorithms for predicting structure based on sequence.

REFERENCES


26. Herzberg, O., and James, M. N. G. (1985) Common structural framework of the two Ca2+/Mg2+ binding loops of troponin C and other Ca2+ binding proteins. Biochemistry 24, 5298-5302


