Chapter from the book *Protein Structure*
Downloaded from: http://www.intechopen.com/books/protein-structure
Ligand-Binding Proteins: Structure, Stability and Practical Application

Olga Stepanenko, Alexander Fonin, Olesya Stepanenko, Irina Kuznetsova and Konstantin Turoverov

Institute of Cytology RAS,
Laboratory of Structural Dynamics,
Stability and Folding of Proteins,
Russia

1. Introduction

A tremendous diversity of ligand binding proteins exists in nature. This undoubtedly creates considerable opportunities for scientific and medicinal applications. In this chapter, we will consider a range of ligand binding proteins, with particular attention to two classes, namely the ligand-binding proteins of the bacterial periplasm and odorant-binding proteins, because these proteins are the building blocks for biosensor development.

2. Diversity of ligand-binding proteins

All functions of living organisms are related to proteins that are present in enormous numbers in the living cell. Usually, proteins are classified by their function. Often, proteins have a range of different functions. Interestingly, the functions of many proteins involved in different biological processes begin with the binding of specific molecules: carbohydrates, amino acids, anions, metal ions, ions, oligo-peptides, proteins, lipids, odorant molecules and others, collectively known as ligands. Vital cell processes, such as DNA replication, gene expression, cell signaling and so on, are initiated by the binding of specific ligands. Trafficking of molecules throughout cellular compartments is possible after the binding of such molecules to a specific carrier protein. To perform their biological functions, enzymes must bind to their cognate substrates. Though performing diverse biological functions, all of these proteins fall into the category of ligand-binding proteins. Such proteins include periplasmic binding proteins, biotin-binding proteins, lipid-binding proteins, lectins, serum albumins, immunoglobulins, and others (De Wolf & Brett, 2000). Given that ligand binding proteins have a high affinity for their ligands, ligand-binding proteins can be used in protein-based controlled delivery systems for bioactive compounds sensitive to environmental factors. Another application of ligand-binding proteins is for biosensing of different disease markers, pathogenic molecules, environmental toxins and chemically or biologically hazardous compounds. Some of these proteins do not alter their structure in response to ligand binding (lipid-binding proteins, lectins). By contrast, other proteins show significant ligand-induced conformational changes. For example, the two-lobed ligand binding proteins of bacterial periplasm switch from their open-form in the absence of
ligands to their closed-form in the presence of ligands (Stratton & Loh, 2011). In general, ligand-binding proteins, including lipid-binding proteins, some of the lectins, serum albumins and biotin-binding proteins, recognize a wide array of bound ligands. By contrast, there are proteins with narrow specificity (for example, most of the periplasmic binding proteins). Ligand-binding proteins vary in their overall structure and number of binding sites, but most such proteins have a complex multi-domain structure or exist as multi-mers. As there are excellent reviews on the aforementioned families of ligand-binding proteins, here, we will only briefly discuss ligand-binding proteins and their possible therapeutic and clinical applications.

The biotin-binding proteins, namely, chicken egg-white avidin, bacterial streptavidin and newly discovered tamavidins from basidiomycete fungi, have numerous medical, biological, biochemical and biotechnological applications (Lahti et al., 2006; Takakura et al., 2010; Wilchek et al., 2006). These tetrameric proteins, consisting of classical β-barrel monomers, bind biotin with exceptionally high affinity. Several peptides having the consensus HPQ tripeptide sequence are reported to be ligands of biotin-binding proteins, though with much lower affinity. The diverse family of lipid-binding proteins (LBP; Banaszak et al., 1994) is made up of extracellular LBPs (eLBPs, which are also known as lipocalins; Grzyb et al., 2006) and intracellular LBPs (iLBPs; Glatz et al., 2002; Haunerland & Spener, 2004). These low-molecular weight proteins share a remarkably similar β-barrel structure, albeit with some differences between iLBPs and lipocalins, and are found in diverse cell types. Individual LBPs can bind a wide range of small hydrophobic ligands, including fatty acids and retinol analogs. A representative member of the lipocalins is β-lactoglobulin (β-LG; Perez & Calvo, 1995), which is shown to be a promising carrier for fatty-acids, as well as a protective agent for bioactive compounds and therapeutically relevant synthetic retinoid derivatives (Liang & Subirade, 2010; Riihimaki-Lampen et al., 2010). An extensive group of lipocalins are the odorant-binding proteins (OBPs) that were successfully adapted to serve as biosensors for dangerous substances, including polyaromatic hydrocarbons (Weiz et al., 2008). Other lipocalins, such as neutrophil gelatinase-assotiated lipocalin (NGAL; Taub et al., 2010; Xu & Venge, 2000), are utilized in clinical applications. Importantly, the high structural plasticity of the lipocalin’s binding site allows, with the aid of genetic engineering, the generation of artificial lipocalins with novel ligand specificities, that is, the so-called anticalins (Skerra, 2008). Anticalins are immunologically active molecules that bind to small hapten-like compounds and to large protein antigens. Compared with antibodies, they are small (composed of just one polypeptide chain), do not require post-translational modification and exhibit robust biophysical properties. Owing to these properties, anticalins offer many potential applications, not only as reagents for biochemical research but also as a new class of drugs for medical therapy. Other ligand-binding proteins, namely serum albumins, bind to an extremely large number of diverse ligands (more than 70), including fatty acids, amino acids, therapeutic drugs and inorganic ions (Fasano et al., 2005; Varshney et al., 2010). Serum albumins are composed of three structurally homologous domains that are predominantly helical (Carter & Ho, 1994). The binding sites for a variety of ligands are distributed among distinct locations on the protein. The main advantage of serum albumins for in vivo applications is their compatibility with human blood, plasma and body components. Currently, serum albumins are playing an increasing role in the development of drug-delivery systems (Kratz, 2008) and in diverse clinical applications (Caironi & Gattinoni, 2009). A versatile family of periplasmic binding proteins originating in the periplasmic space of bacteria share a characteristic two-lobed structure and traffic different
nutrients, such as carbohydrates, amino acids, anions, metal ions, and di- and oligo-peptides (Felder et al., 1999; Fukami-Kobayashi et al., 1999; Tam & Saier, 1993). The members of this family commonly have high specificity for their cognate ligands, though there are exceptions (the case of the di/oligopeptide-binding protein). Their intrinsic ability to undergo a significant, ligand-induced conformational change has been utilized in the engineering of reagentless biosensors to monitor ligand concentration (Dwyer & Hellinga, 2004; de Lorimier et al., 2002). Members of another ubiquitous family of sugar-specific and cell-agglutinating proteins, the lectins, have been found in all kinds of organisms, from viruses to humans (Sharon & Lis, 2004). Lectins function as recognition molecules in ligand-cell and cell-cell interactions in a variety of biological systems. Mature plant lectins are divided into merolectins, hololectins, chimerolectins and superlectins, according to the number of carbohydrate-binding domains (Liu et al., 2010). There are only one or at most two carbohydrate-binding domains in merolectins and hololectins, respectively. It is important to note that different carbohydrate-binding domains of hololectins bind either to the same or structurally similar sugars. Chimerolectins are fusion proteins containing one or more carbohydrate-binding domains and an unrelated domain. Superlectins have at least two carbohydrate-binding domains with specificity for structurally unrelated sugars. Plant lectins have a similar tertiary structure, referred to as the lectin fold. This structural motif consists of a characteristically elaborate jelly roll, derived from antiparallel β-strands, and arranged as two β-sheets. This fold has also been noted in animal lectins. Animal lectins can be divided into 12 groups based on the similarity of their primary structures, including the previously discovered C-type lectins (requiring Ca$^{2+}$ for activity) and the galectins (Kilpatrick, 2002). Animal lectins are often bi-functional, with a carbohydrate-binding domain and an additional domain, which is responsible for the ability of animal lectins to bind to non-carbohydrate ligands via protein-protein, protein-lipid and protein-nucleic acid interactions. Furthermore, there are examples of animal lectins with a carbohydrate-binding domain capable of interaction with non-sugar ligands. Many studies are devoted to the anti-tumor activity of plant lectins against a variety of malignant cells. Lectins have a high potential for development of antineoplastic drugs for cancer therapy and of targeted drug delivery systems (Liu et al., 2010; Robinson et al., 2004). Immunoglobulins, with their typical tetrameric organization consisting of two light and two heavy polypeptide chains, are indispensable in basic research and diagnostics, as they can be adapted for the binding of an incredible variety of ligands (Chester & Hawkins, 1995; Sundberg, 2009). Bispecific antibodies that are capable of simultaneous binding to two different antigens are further improving the prospects for clinical applications of conventional antibodies (Fitzgerald & Lugovskoy, 2011). Some shortcomings of antibodies, such as their lengthy timeframe and high cost of production, and their large size can be overcome by development of small single-domain antibody fragments of high stability with the unique capacity to recognize molecules that are inaccessible to conventional antibodies (de Marco, 2011). Several additional classes of ligand-binding proteins, including inactivated enzymes, penicillin-binding proteins, immunophilins and others, further expand the possibilities for diverse practical applications.

In the next sections, we will focus on two classes of ligand-binding proteins, the two-domain ligand-binding proteins of the bacterial periplasm and odorant-binding proteins for which the use as the sensitive elements in socially important biosensor systems is mostly elaborated. The structure, stability and possible practical applications of these proteins will be discussed in detail.
Two-domain ligand-binding proteins of bacterial periplasm

Two-domain ligand-binding proteins of the periplasm of gram-negative bacteria (PBPs; Dwyer & Hellinga, 2004; Tam & Saier, 1993) constitute a large bacterial family of proteins serving as primary receptors for a large number of compounds. The ligands of periplasmic binding proteins are represented by carbohydrates such as glucose, maltose, ribose and arabinose; amino acids such as glutamine, leucine, valine and histidine; metal ions such as phosphate, sulfate, iron, zinc and nickel; di- and oligo-peptides and vitamins. Thus, they are involved in the active transport of the soluble molecules inside the bacterial cell. Two-domain ligand-binding proteins are constituents of the ATP-binding cassette in which ligand transport across the membrane is powered by ATP hydrolysis. In general, the ATP-binding cassette consists of two trans-membrane domains, which assist in ligand translocation across the inner membrane, and two nucleotide-binding domains, which provide the energy required for the transport process. The periplasmic binding protein is responsible for trafficking of its ligand across the periplasmic space and release of its ligand near the inner membrane. According to recent studies, the ATP-switch model for the transport function of the ATP-binding cassette has been proposed. In this model, the coupled nucleotide-binding domains switch between an ATP-dependent closed conformation and a nucleotide-free, open conformation to drive the translocation of ligand (Dawson et al., 2007; Linton, 2007). In some cases, the periplasmic binding proteins participate in chemotaxis toward different substances. For example, chemotaxis toward such attractants as some sugars (galactose, ribose and maltose) and amino acids is activated by interaction between a complex of defined binding proteins and attractant molecules recognizing specific chemoreceptors (Felder et al., 1999; Szurmant & Ordal, 2004). Periplasmic binding proteins are involved in bacterial intercellular communication processes, termed quorum sensing (Neiditch et al., 2006; Schauder & Bassler, 2001). In this case, binding of small molecules, the so-called autoinducers, leads to a series of chemical reactions that provide the bacteria with information about the cellular density of the surrounding environment. Some periplasmic binding proteins act as chaperones promoting the proper folding of denatured proteins or their fusion partners (Dalken et al., 2010; Richarme & Caldas, 1997).

The molecular weight of periplasmic binding proteins ranges from 22 to 59 kDa. Despite significant differences in their amino acid sequences, all ligand-binding proteins share the same structural topology of the polypeptide chain. All of them display a secondary structure of the $\alpha/\beta$ type that is organized at the tertiary level into two domains linked by what is commonly referred to as a hinge region. This hinge region is formed by two or three short flexible peptide segments. The protein ligand-binding site is located in the cleft between the two domains. The periplasmic binding proteins can interconvert through a pronounced (depending on the protein) bending motion around the hinge from a ligand-free open conformation to a ligand-bound closed conformation. Both protein domains adopt a three-layered $\alpha/\beta/\alpha$ sandwich fold, and the distribution of $\beta$-strands between the two domains defines the structural subclass of the periplasmic binding protein (figure 1). One of the structural sub-classes (group I), includes *Escherichia coli* D-galactose/D-glucose-binding protein (GGBP; Borrok et al., 2007); and the other subclass (group II) includes *Escherichia coli* glutamine-binding protein (GlnBP; Hsiao et al., 1996; Sun et al., 1998; Fukami-Kobayashi et al., 1999). Both domains possess a CheY-like fold, which gave rise to the hypothesis that the ancestral protein for PBP family members is derived from a one-domain CheY-like protein
Fig. 1. Spatial pattern of PBP of group I and group II. Tertiary structures of GGBP (a) and its complex with glucose GGBP/Glc (b), representing group I, are shown. The residues of the Ca-binding site are shown in green, including the loop in the protein’s C-terminal domain (residues 134-142; 1) and Glu 205. Calcium is represented as a yellow sphere. The structures of group II GlnBP (e) and its complex with glutamine GlnBP/Gln (f) are shown. In both cases, the ligand (glucose or glutamine) is represented as a red stick union. The structures were created based on PDB data (Dutta et al., 2009); PDP codes 2GBP.ent15 (Vyas et al., 1988), 2FWO.ent16 (Borrok et al., 2007), 1GGG (Hsiao et al., 1996) and 1WDN (Sun et al., 1998) using the graphical software VMD (Hsin et al., 2008) and Raster 3D (Merritt & Bacon, 1977). The topology of group I (c) and group II (d) proteins is drawn with β-strands and α-helices indicated as arrows and cylinders, respectively. Secondary structural elements originating from a monomeric ancestral protein are represented in the same color (red or blue).

through its duplication and subsequent fusion (Lewis et al., 2000). In group I, the domain’s β-strands have a 21345 topology and have more regular organization of its secondary structure, while group II proteins, derived later on the evolutionary time scale, are characterized by a more complex topology of their β-strand distribution within each separate domain. The sheet topology of both domains of group II periplasmic binding proteins follows a 213N4 sequence, with βN becoming the first strand after crossing-over.
from the N-terminal domain to the C-terminal domain. The group II proteins are supposed to have arisen from the group I proteins through the mutual dislocation of two β-strands in each domain from their original domain to the other: one β-strand in the first domain penetrates into the parallel β-sheet of the second domain, making a new anti-parallel β-sheet, as does the β-strand of the second domain (Fukami-Kobayashi et al., 1999). It is interesting that GGBP has the most regular distribution of βαβ repeats throughout its separate domains. Additionally, there are two extra α-helices in each of the GGBP domains that are absent in all periplasmic binding proteins of group II and in some members of group I. Thus, the spatial structure of GGBP seems to be the closest to that of the ancestral periplasmic binding protein. It is worth noting that di/oligopeptide-binding proteins contain a third domain in addition to the two widely recognized domains of periplasmic binding proteins (Nickitenko et al., 1995; Sleigh et al., 1999). According to the topology of the β-strands typical of periplasmic binding proteins, two domains of the di/oligopeptide-binding proteins can be regarded as group II proteins. The third domain is organized into two hairpins that make only a few contacts with the ligand. Still, some of the periplasmic binding proteins do not fall into any of the two aforementioned structural groups. These include, for example, zinc-binding protein TroA (Lee et al., 2002) and vitamin B12-binding protein BtuF (Karpowich et al., 2003), wherein the protein’s N- and C-domains are linked by a single long helix, imposing some rigidity on the overall protein structure. These periplasmic binding proteins can be regarded as group III proteins.

Many solved structures of periplasmic binding proteins, both in the presence and in the absence of ligand, show large-scale motion of the two domains, described as bending and twisting motions around two axes. The degree of the hinge-bending motion varies from 14 to 62 ° for different periplasmic binding proteins (Shilton et al., 1996). The ligand-bound closed state of a periplasmic binding protein in complex with its ligand possesses a protein surface that is quite different from that observed in the open conformation of the protein. This difference in protein conformation is important for recognition of protein-ligand complexes by the trans-membrane proteins of the ATP-binding cassette (Hollenstein et al., 2007). For a long time, these conformational changes were assumed to be triggered by ligand binding. Recent studies revealed that ligand-free forms of periplasmic binding proteins are very flexible compared to ligand-bound forms (Bucher et al., 2011; Pang et al., 2003). This feature is common among all proteins of the PBP family, even among proteins in group III (Kandt et al., 2006; Krewulak et al., 2005). The flexible apo-form of the periplasmic binding proteins tends to oscillate along the modes that lead from the open to closed structure. Further support for this concept is provided by the existence of a dynamic equilibrium between the open and semi-closed conformations of the apo-form of the protein as revealed in the case of maltose-binding protein (MBP; Bucher et al., 2011). Additionally, the fully closed-form of the protein in the absence of ligand is not observed. To the contrary, GlnBP, in its ligand-free state, cannot achieve a partially closed conformation (Bermejo et al., 2010). Possible reasons for this are the stability of the hinge region of GlnBP – unusual in periplasmic binding proteins – and/or instability of the protein’s partially closed-unliganded conformation. Whether other members of the PBP family can adopt a semi-closed conformation in the absence of ligand is the subject of future studies.

Generally, the amino acids of the protein ligand-binding site have extensive specific interactions with their cognate ligands, resulting in a high degree of selectivity between
anomeric or epimeric carbohydrates, carbohydrates of different sizes or chemically similar anions (Bruns et al., 1997; Cuneo et al., 2009). At the same time, semi-specific ligand-binding takes place. This is the case for the di/oligopeptide-binding protein that binds to peptides ranging in size from two to nine amino acids with little discrimination between the side chains of the peptides (Sleigh et al., 1999). A network of strong hydrogen bonds is formed with the atoms of the main peptide chains, while the peptide’s side chains adapt to the binding pocket with the aid of water molecules that donate hydrogen bonds to them and shield the ligand’s charge.

The protein family of extra-cellularly bound lipoproteins homologous to the periplasmic binding proteins of gram-negative bacteria exists in gram-positive bacteria, wherein such extra-cellular proteins are covalently bound to outer cell surface (Felder et al., 1999). The spatial fold typical of the two-domain ligand-binding protein has also been found in transcriptional regulators, such as the lac-type repressors. Many eukaryotic receptors contain the PBP fold as a component of multi-domain proteins, such as glutamate/glycine-gated ion channels (for example, the ionotropic glutamate receptor GluR2), G protein-coupled receptors (for example, calcium-sensing receptors CaSRs) and atrial natriuretic peptide receptors. These receptors are regulated by conformational changes of the protein’s extracellular domain in response to ligand binding (van den Akker, 2001; Felder et al., 1999).

2.2 Odorant-binding proteins

Odorant-binding proteins in vertebrates (OBPs) belong to the family of lipocalins (Flower, 2000). Lipocalins are a functionally diverse family of small and abundant extracellular proteins that bind mainly hydrophobic molecules, including lipids, odorants, pheromones, retinoids, porphyrins, siderophores, and steroids (Akerstrom et al., 2000; Flower et al., 2000). These proteins have been primarily classified as transport proteins, but it is known currently that lipocalins are involved in many important biological processes (Grzyb et al., 2006). Indeed, many lipocalins have been implicated in the regulation of cell homeostasis: apolipoprotein D, quiescence specific protein, purpurin, alpha-1-microglobulin, and NGAL. Some lipocalins, such as alpha-1-microglobulin, glicodelin and others, participate in the regulation of the immune response (Flower, 2000). Human tear lipocalin is expressed in lacrimal glands of both sexes and released into tears. Human tear lipocalin binds to a broad array of lipophilic substances including fatty acids, fatty alcohols, phospholipids, glycolipids and cholesterol (Breustedt et al., 2005; Glasgow & Gasymov, 2011). Tear lipocalin has a high affinity for retinol, microbial and fungal siderophores and harmful lipophilic compounds. It is proposed that the physiological function of human tear lipocalin is to prevent the corneal surface from desiccating and to stabilize the tear film by binding lipids present in the outer layer of tears. Tear lipocalin is also produced in von Ebner’s lingual salivary glands, trachea, prostate, pituitary and sweat. Tear lipocalin in those tissues is supposed to protect epithelia by removing harmful hydrophobic molecules. In addition, tear lipocalin exhibits enzymatic activity such as endonuclease and cystatin-like activity (Glasgow & Gasymov, 2011; Redl, 2000; Yusifov et al., 2008). Tear lipocalin also provides anti-fungal and anti-microbial defenses by capturing siderophores (Fluckinger et al., 2004). Sex-specific pheromone-binding lipocalins and odorant-binding lipocalins (OBPs) are supposed to be associated with chemical communication and olfaction. Pheromone-binding lipocalins include major urinary proteins (MUPs) of male rat and mouse secreted predominantly in the urine and saliva (Beynon & Hurst, 2004), aphrodisin from female
hamsters isolated from the vaginal discharge (Briand et al., 2004) and lipocalins of boar saliva (Marchese et al., 1998). These proteins are believed to serve as reservoirs for delivering to the environments and sustained dissemination of pheromones. Chemical information borne by pheromones is perceived by conspecifics and invokes various behavior responses in them (Hurst & Beynon, 2004).

In contrast to pheromone-binding lipocalins, OBPs are secreted in the nasal mucus of the oral sphere epithelia of both sexes. OBPs have been identified in a variety of species, including cow (Pevsner et al., 1985), pig (Dal Monte et al., 1991), rabbit (Garibotti et al., 1997), mouse (Pes et al., 1992), rat (Lobel et al., 2002), elephant (Lazar et al., 2002) and human (Briand et al., 2002). The first OBPs were discovered in the nasal mucus of the cow and several other animals according to their ability to bind pyrazine with a low detection threshold (Baldaccini et al., 1986). The studies of the binding specificity of OBPs toward several common odorants, which were performed mostly with bovine OBP (OBPb) and pig OBP-I (OBPp), have revealed that OBPs can bind a broad spectrum of hydrophobic molecules of medium size (Herent et al., 1995). Among the ligands with the highest affinity for OBPs (dissociation constants in the range of 0.1-1 μM) are heterocyclic derivatives, such as alkyl-substituted pyrazines and thiazoles, terpenoids and their derivatives, such as menthol and thymol, and medium size aliphatic alcohols and aldehydes (Herent et al., 1995). Spherically shaped terpenoids, such as camphor and its analogues, and polar compounds, such as the short chain fatty acids, exhibit poor affinity to OBPs. The low selectivity of OBPs gave rise to the hypothesis that these soluble proteins might function as chaperones for volatile hydrophobic odorants and pheromone molecules crossing the aqueous mucus layer to the olfactory receptors embedded in the membrane of olfactory neurons (Pevsner & Snyder, 1990; Steinbrecht, 1998). Later, it was proposed that OBPs probably play a more specific role in olfaction through their involvement in the first step of odorant discrimination. Indeed, the expression of several sub-types of OBP in the same animal species differing in their primary structure and having different ligand binding patterns has been observed in some animal species (Ganni et al., 1997; Garibotti et al., 1997; Utsumi et al., 1999). Three rat OBPs are specially tuned for distinct chemical classes of odorant molecules (Löbel et al., 2002). Rat OBP-1 preferentially binds heterocyclic compounds such as pyrazine derivatives. Rat OBP-2 is more specific for long-chain aliphatic aldehydes and carboxylic acids, while OBP-3 interacts strongly with odorant molecules having a saturated or unsaturated ring structure (Löbel et al., 2002). Microheterogeneity of OBPs was reported in the pig (Scaloni et al., 2001), mouse (Utsumi et al., 1999), rabbit (Garibotti et al., 1997) and porcupine (Ganni et al., 1997). Recently, it was shown that post-translational modifications can further increase the micro-diversity of OBPs (Le Danvic et al., 2009; Nagnan-Le Meillour et al., 2009). Indeed, porcine OBP can be posttranslationally modified by phosphorylation, resulting in a set of OBPP isoforms with different binding properties (Brimau et al., 2010). Moreover, the binding specificity of the VEG1 isoform, an olfactory binding protein expressed in the vomeronasal organ of the pig, to steroids is linked to the O-N-acetylglucosaminylation of the protein. The porcine VEG2 isoform, which does not undergo this modification, showed specificity for fatty acids rather than steroids (Le Danvic et al., 2009). Thus, it is supposed that phosphorylation and glycosylation could be a mechanism of regulation of OBP binding properties toward odorant and pheromone ligands, and an expanded set of OBP isoforms can serve to preliminary discriminate among ligands prior binding to the olfactory receptor (Le Danvic et al., 2009; Nagnan-Le Meillour et al., 2009).
The similarity in the amino acid sequences of different OBPs is limited with a sequence homology in the range of only 21-26%. However, few amino acid residues are absolutely conserved within all OBP classes as well as within the lipocalin family (Flower, 2000). They include a GxW motif at the N terminal element (residues 14-16; here and throughout, the section numeration is given according to OBPp), two cysteine residues located on the 4th strand β-barrel and at the C terminal element, respectively (residues 63 and 155), and a Gly (residue 119). Cysteine residues form a disulfide bridge, thus tightening the α-helix domain and the β-barrel. The motif YxxxYxG is also highly conserved (residues 78-84). Bovine OBP is the only protein in the OBP class that has only two of four conserved patches, including GxW and the YxxxYxG motifs (Bianchet et al., 1996). Members of the lipocalin family, including the OBP class, share a common β-barrel structure (figure 2). The nine-stranded β-barrel comprises residues 9-120 (strands 1-8) linked by a turn in the sequence to a short α-helical domain (residues 124-141), followed by the 9th strand of the barrel (residues 146-148) and by a C-terminal tail (residues 149-157). The β-barrel is often preceded by an N-terminal segment, containing a short 310-like helical moiety (residues 1-8). The barrel of OBPs is markedly deformed to have an elliptical shape in cross-section. The β-barrel encloses a ligand binding site composed of both an internal cavity and an external loop scaffold. Loops L2-L7 connecting strands 2-8 are all typical of short β-hairpins; the exception is loop L1 between the 1st and 2nd strands, which is a large Ω loop. Loop L1 forms a lid, partially closing the internal ligand-binding site at this end of the barrel. The internal ligand-binding cavity is formed of mainly hydrophobic and aromatic amino acids. The cavity is shielded from the solvent, and during the interaction of OBP with the ligand, an opening event takes place, allowing the ligand to enter the binding pocket (Golebiowski et al., 2007; Vincent et al., 2004). A molecular dynamics study shows that these events occur mainly at the junction between the β-strands and loops L1 and L5 with the Tyr-82 residue serving as a gate in the OBPs (Golebiowski et al., 2007). The tyrosine residue at this position is highly conserved.
within the OBP class and tolerates substitution only with a phenylalanine residue, which preserves the structure of “the door”. Investigation of the crystallographic structure of OBPb and OBPp complexed with different ligands revealed that the orientation of the ligands inside the cavity appeared to be opportunistic with no specific target patches for aromatic or charged groups and no correlation between the number of contacts and the affinity measured in solution (Vincent et al., 2004).

In contrast to classical lipocalin, dimeric bovine OBP, OBPb, is characterized by a unique folding pattern that involves the crossing of the α-helical domain of each monomer over the β-barrel of the other (Bianchet et al., 1996). In addition to the ligand binding site formed by the internal cavity of the β-barrel in each monomer, OBPb has a putative third binding site called the “central pocket”, which is located at the dimer interface in communication with the solvent (Bianchet et al., 1996; Pevsner et al., 1985). The function of the central pocket is not well understood. It has been recently proposed that the holo-form of OBPb captures the first encountered odorant molecules at the central pocket irrespective of their affinity to OBPb (Ikematsu et al., 2005). The initial ligand binding is supposed to be a prerequisite for stabilization of the OBPb structure and for the adjustment of internal binding sites for interaction with ligands. OBPb can further bind the odorant with high affinity at its internal cavity, releasing the pre-bound ligand at the central odorant-binding pocket. The internal cavity-bound odorant can be released by the binding of other odorants at the internal cavity of the second OBPb subunit or at the central pocket, depending on the nature of the odorant. It is believed that such interactions of OBPb with its ligands make it more reactive than other monomeric OBPs and enable a quick recognition of a change in the environment that is highly desirable for ruminant animals, thus permitting them to escape from danger.

Careful examination of the structure of OBPb and OBPp, which exhibits a classical lipocalin fold, revealed the presence of a single insertion of a glycine residue at the hinge between the α-helical moiety and β-barrel domain in OBPp with respect to OBPb (the stretch $L^{117}LGKG^{121}TDIED^{126}$ in OBPp; Spinelli et al., 1998). The Gly-121 residue is located exactly at the position where domain swapping occurs in OBPb, and its presence induces a shift in the alignment of the structural elements joining the swappable domains. The longer and more flexible linker segment between the α-helical and β-barrel domains in OBPp is proposed to be sufficient for preventing domain swapping in the protein (Spinelli et al., 1998), while a conserved disulfide bridge between Cys-63 and Cys-155 stabilizes this non-swapped conformation. The fact that the mutant variant of OBPb, with an insertion of a glycine residue after position 121, became monomeric has further confirmed these findings (Ramoni et al., 2002).

Studies of OBPs over a period of more than 20 years were not able to completely clarify their role. A broad range of ligands and the low number of OBPs with respect to the number of olfactory receptors in mammalian species argue in favor of OBP function as a non-specific carrier of hydrophobic molecules (Pevsner & Snyder, 1990). A role of OBPs in the termination of the olfactory signal has also been proposed, which entails the "removal" of odorants from the olfactory receptors once they have been stimulated to keep them in their active state. The low specificity of OBPs led to the hypothesis that OBPs function as a scavenger of excess odorants to prevent olfactory receptor saturation (Burchell, 1991). Recent findings imply a more active role for OBPs in olfaction, which could involve the preliminary discrimination of odorant molecules or even a direct interaction with olfactory receptors. As already mentioned, the spectrum of OBPs in animal species can be expanded.
through phosphorylation and glycosylation. Further, it has been shown with the chip based surface plasmon resonance technique that the OBP is able to modulate the activity of olfactory receptors (Vidic et al., 2008). Some evidence for subtle conformational changes in rat and porcine OBPs after ligand binding has been obtained in molecular dynamics studies wherein opening of the β-strand pair was observed (Hajjar et al., 2006). These structural dynamics of OBPs might be essential for recognition of the OBP-ligand complex by the olfactory receptor. OBPs, at least in ruminants, might fulfill a protective role. The natural ligand of the bovine OBP, 1-octen-3-ol, which is produced by endogenous ruminal microflora of ruminants, is an attractant for many insect species (Ramoni et al., 2001). Thus, the OBP can be used to capture 1-octen-3-ol and decrease the risk of infections carried by insects. OBPs have been shown to bind to high affinity aldehydes derived from lipid peroxidation. This observation gave rise to the proposal that OBPs might be used to scavenge toxic substances from nasal epithelia to protect them against oxidative stress (Grolli et al., 2006).

3. Structure and stability of ligand-binding proteins

One of the most intriguing questions in modern molecular and cell biology is how a globular protein folds into a unique, compact, highly organized and functionally active state. In the past decade, our knowledge about protein folding into the native state and even the notion of the native state itself has undergone considerable changes. At the turn of the century, publications appeared that showed that the polypeptide chains of many proteins could not, in principle, fold into a compact globular state. Although these proteins are intrinsically disordered, they are functionally active and notably are in their native state. These proteins form a compact globular state only upon interaction with their specific binding partners, such as low-molecular weight ligands, other proteins, or nucleic acids. As a consequence, understanding the effects of ligand binding on proteins is of great interest (Turoverov et al., 2010). From this point of view, ligand-binding proteins can be convenient models to investigate the role of the ligand in the structure and in the stabilization of proteins in their native state.

3.1 The role of ligands in the process of folding and stability of two-domain ligand-binding proteins

The division of periplasmic binding proteins into two different structural sub-classes according to their structural topology makes it tempting to attribute this fact to differences in folding pathways of these protein sub-classes. In fact, group I proteins in general are characterized by simple two-state folding processes (Kashiwagi et al., 2003). More complex folding, accompanied by intermediate state accumulation, is observed for periplasmic binding proteins of group II (Chun et al., 1993; Staiano et al., 2005). For example, the folding of the *Escherichia coli* glutamine-binding protein, as studied by protein intrinsic fluorescence, 1-anilinonaphthalene-8-sulfonic acid (ANS) emission fluorescence, far- and near-UV circular dichroism spectroscopy, and the parametric presentation of fluorescence data (Kuznetsova et al., 2004), proceeds through the formation of two intermediate states $I_1$ and $I_2$ (Staiano et al., 2005). Glutamine binding makes the GlnBP structure more resistant to the denaturing action of the chemical agent, guanidine hydrochloride (GdnHCl); thus, unfolding of the GlnBP complex with Gln (GlnBP/Gln) begins at higher GdnHCl concentrations, and the denaturing process becomes more cooperative. This creates the illusion of a single stage of GlnBP/Gln folding. However, GlnBP/Gln folding successively passes through the same...
intermediate states as GlnBP but in a narrower range of denaturant concentrations so that the three stages of the folding process essentially overlap. Thus, the ligand serves as an agent to protect the entire protein structure from denaturation (figure 3, a–c).

The folding mechanism for the *Escherichia coli* D-galactose/D-glucose-binding protein and its complex with glucose (GGBP/Glc), representing group I proteins, is shown to be a one-stage process, with GGBP/Glc being more stable to denaturation than GGBP alone (Stepanenko et al., 2011a). This is supported by the sigmoidal contour of all recorded characteristics, such as fluorescence and far- and near-CD. The linear shape of the parametrically represented fluorescent data and the absence of an increase in ANS fluorescence during GGBP and GGBP/Glc complex unfolding also argue for the lack of any intermediate states in protein folding (figure 3, d; Stepanenko et al., 2011a). It is noteworthy that GGBP has an extra ligand, the calcium ion, located at a distinct position from the glucose-binding site (Vyas et al., 1988). The Ca-binding site consists of the loop of the protein’s C-terminal domain (residues 134-142), and its structure resembles the “EF-hand” motif, which is typical of intracellular Ca-binding proteins (figure 1). The calcium ion forms coordination bonds with the oxygen atoms of every second residue of this loop and with the Glu 205 residue. The role of calcium in GGBP stability and folding has been evaluated. Although producing little effect on protein structure both in the absence and in the presence of glucose, calcium removal results in pronounced destabilization of GGBP even at small denaturing actions. Meanwhile, calcium depletion has practically no effect on GGBP/Glc stability (figure 3, d). Thus, the calcium ion serves as a guardian of the protein’s structure in the absence of glucose. These results indicate that the role of calcium consists of maintaining the native structure of GGBP in its open form. Such a stabilizing effect of calcium was previously observed for other calcium-containing proteins (Turoverov et al., 2010). Recently, the folding mechanism of GGBP has been shown to be more complex than expected for the folding pathways of group I proteins (Piszczek et al., 2004; Stepanenko et al., 2011b). An extensive analysis of the experimental data reveals that the two domains of GGBP have a slightly different thermal stability, which is more marked for mutant variants of GGBP carrying point amino acid substitutions in the ligand-binding site of the protein (figure 3, g–i; Stepanenko et al., 2011b). These results suggest that more careful interpretation of accumulated data are needed.

While studying GdnHCl-induced unfolding – refolding of GGBP in the presence of glucose an interesting effect was observed (figure 3, e–f; Stepanenko et al., 2011a). The equilibrium curves for complex refolding-unfolding are attained only after 10 days of incubation of GGBP/Glc with GdnHCl. This effect is not revealed upon GGBP/Glc heating. Slow equilibrium acquisition between the native protein in the form of the GGBP/Glc complex and the unfolded state of the protein in the presence of GdnHCl is connected with increased viscosity of the solution at moderate and high GdnHCl concentrations, which interferes with diffusion of glucose molecules. Before equilibrium is established for an appreciable period of time, an excess concentration exists (in comparison with equilibrium) of the native complex (GGBP/Glc)N in the unfolding pathway or of the unfolded protein (GGBP)U in the refolding pathway. These imbalances are caused by the activation barrier, which must be overcome in both cases. In the unfolding pathway, the elementary process of complex dissociation does not bring about a disturbance in the configuration fit of the interacting GGBP and Glc molecules, so that the probability of the inverse reaction remains high. In the refolding pathway, (GGBP/Glc)N is formed because of the coincidence of two conditions: the formation of the native molecule (GGBP)N and the appearance of a configuration fit between (GGBP)N and Glc. Thus, the rate-limiting step in the unfolding-refolding process
for the GGBP/Glc complex is the disruption/tuning of the configuration fit between the protein in its native state and the ligand.

Fig. 3. Conformational changes of GlnBP (panels a–c) and GGBP (panels d–i). GdnHCl-induced transitions of GlnBP (blue circles) and complex with glutamine GlnBP/Gln (red triangles) were recorded by tryptophan fluorescence intensity at 365 nm (a, λ<sub>ex</sub>=297 nm), ANS fluorescence intensity (b) and parametric presentation of fluorescence intensities at 320 and 365 nm (c). GdnHCl-induced transitions of GGBP (red circles) and complex with glucose GGBP/Glc (blue circles) and their calcium-depleted forms GGBP-Ca (pink circles) and GGBP-Ca/Glc (light blue circles) were characterized by equilibrium changes in parameter A = I<sub>320</sub>/I<sub>365</sub>, λ<sub>ex</sub>=297 nm (d) and parametric presentation of fluorescence intensities at 320 and 365 nm (inset in d). In panels a–d, data for unfolding are depicted with empty symbols and data for refolding, with filled symbols. Equilibration (e) of unfolding – refolding curves as recorded by parameter A was achieved for GGBP after an incubation of less than 24 h in the presence of GdnHCl (red circles) and for GGBP/Glc, only after an incubation of 10 days (blue circles). The renaturation (closed blue squares) and denaturation (open blue squares) curves for GGBP/Glc do not coincide after 24 h incubation, but the curves tend to approach each other. The kinetics scheme characterizes the GGBP and GGBP/Glc unfolding – refolding processes (f). Heat-induced denaturation of GGBP (red) and GGBP/Glc (blue) was recorded by differential scanning calorimetry (g) and by the fluorescence intensity at 365 nm (h) and the first derivative of fluorescence intensity (i). The deconvolution of the calorimetric traces into two separate thermal transitions in panel g is shown in gray. Two sequential scans (solid and dashed lines, respectively) of the temperature dependency of the fluorescence intensity at 365 nm of the studied proteins are shown to characterize the reversibility of the thermal transitions (h).
All of the data reveal a common effect of the ligand on the folding and stability of all periplasmic binding proteins. The presence of the bound ligand does not change the folding pattern of the individual ligand-binding protein, but makes the protein more cooperative. Additionally, the structure of these proteins becomes more stable to various denaturing actions on the binding of their cognate ligands. The protein preserves its native structure up to the point of ligand dissociation induced by increased denaturing effects.

3.2 Beta-barrel scaffold of odor-binding proteins

Investigations into heat-induced and chemical denaturation of a series of OBPs revealed their high structural stability. Indeed, in the case of denaturation of porcine OBPp by the chemical denaturing agent GdnHCl, the difference in free energy between native and unfolded states of OBPp is as high as –5.95 kcal/mol (Staiano et al., 2007). The native structure of OBPp is highly resistant to heating with a transition mid-point of approximately 70 °C (Burova et al., 1999; Paolini et al., 1999; Stepanenko et al., 2008). In the absence of GdnHCl, OBPp preserves some residual structure under heating up to 80 °C as indicated by parameter A measurements (figure 4, a; Stepanenko et al., 2008). It noteworthy that parameter A being calculated as $A = \frac{I_{320}}{I_{365}}$ (where $I_{320}$ and $I_{365}$ are fluorescence intensities measured at the emission wavelengths of 320 and 365 nm, respectively), characterizes the shape and position of protein’s fluorescence spectra (Turoverov & Kuznetsova, 2003). Changes of the parameter A can provides information about even subtle perturbation of spatial structure of proteins. Complete temperature denaturation of the protein can be achieved only in the presence of GdnHCl. Aggregation of OBp that occurs at high temperatures is explained as the result of increased conformational mobility of the $\Omega$ loop with a loss of an ion pair connecting Glu 31 and Arg 152 (Stepanenko et al., 2008). As shown
by studies of the double mutant variant of OBPp with C63A and C155A substitutions, the conserved disulfide bridge between Cys-63 and Cys-155 plays an important role in the stability and function of this protein (Parisi et al., 2005).

Interestingly, the study of GdnHCl-induced denaturation of OBPp revealed unusual characteristics of the fluorescence of the single Trp residue (Trp 16) in this protein (Staiano et al., 2007). In OBPp, the only polar group in the vicinity of Trp 16 is Lys 120. The side chain of Lys 120 is virtually parallel to the indole ring of Trp 16 with its positive charge located just over the center of the indole ring of Trp 16. This is the most favorable position for the formation of a complex between this group and the indole ring that could be responsible for the low fluorescence intensity of the native OBPp. Pre-denaturating GdnHCl concentrations result in an increase in the fluorescence quantum yield and a decrease in the fluorescence lifetime of Trp16 in OBPp (figure 4, b). These effects are not accompanied by noticeable changes in protein secondary and tertiary structure. The accessibility of Trp16 to the solvent remains unchanged, and near UV CD spectra become more pronounced, indicating disturbances to the microenvironment of the tryptophan residue. This likely leads to the disruption of the complex between Trp 16 and the positively charged NZ atom of Lys 120 and consequently to the enhancement of fluorescence intensity. The possibility of a distance change between atom NZ of Lys 120 and the indole ring of the tryptophan residue of OBPp in water was confirmed by molecular dynamic simulations. At the same time, the increase in the asymmetry of the tryptophan residue microenvironment can also promote the formation of exciplexes of Trp 16 with bound water molecules in close vicinity to the indole ring and can explain the decrease in the fluorescence lifetime. Thus, pre-denaturating GdnHCl concentrations induce local changes in the microenvironment of Trp16 in OBPp.

Increased stability of OBPs can be attributed to their β-barrel fold. The other group of proteins with β-barrel topology is the family of fluorescent proteins (FPs) that are widely used in a variety of applications in modern biology. For FPs, an extremely high resistance to environmental conditions has also been observed (Stepanenko et al., 2010).

4. Ligand-binding proteins as the sensitive element of socially important biosensors

The selective interaction of ligand-binding proteins with their partners is the major prerequisite for their use as sensitive elements in biosensor systems for definite analytes (Deuchle et al., 2005). In some cases, binding of a ligand to a ligand-binding protein is followed by a change in total charge, refractive index or molecular mass that can be detected (Stratton and Loh, 2011). These approaches in vivo are complicated by a significant noise level, e.g., in serum blood, there are approximately 3000 proteins, and the total protein concentration is approximately 70 mg/mL (Tang et al., 2005). Two-domain ligand-binding proteins sharing the intrinsic feature of ligand-induced conformational changes and targeting a diversity of natural ligands have been exploited in protein engineering to construct different biosensor systems (Dwyer and Hellinga, 2004; Tolosa et al., 2010). The large protein domains motion about the hinge has been utilized to transduce the ligand binding event to a variety of physical signals. Though the most preferable approaches for analyte registration are fluorescent methods, intrinsic protein fluorescence is not suitable as cells and living tissues are completely impervious to UV-light. Rational introduction of FRET-compatible protein pairs and environmentally sensitive dyes into different sites on the
ligand-binding protein, as well as use of surface plasmon resonance methods, have been successful approaches to the design of biosensor systems (Ge et al., 2004; Khan et al., 2008, 2010; de Lorimier et al., 2002; Okada et al., 2009; Thomas et al., 2006). Changing the affinity and specificity of ligand-binding proteins has been achieved by the re-engineering of the protein’s ligand-binding site. Much work has been performed with PBP-based biosensors for sugar detection (e.g., glucose, ribose and other) (Amiss et al., 2007; Medintz & Deschamps, 2006; Vercillo et al., 2007). Utilization of the *Escherichia coli* D-galactose/D-glucose-binding protein as the sensitive element in the glucose biosensor is one of the most promising directions for continuous glucose monitoring (Deuchle et al., 2005; Shilton et al., 1996). In this case, the necessity to lower the affinity of GGBP to glucose should be taken into account in the development of methods to monitor the glucose level in human blood (Amiss et al., 2007). It is worth mentioning that the ligand-binding proteins represent a target for the development of antimicrobial agents. These small molecules that are antagonists to native PBP ligands act as inhibitors of the conformational changes in proteins and thus block key processes in bacteria driven by PBP (Borrok et al., 2009).

Odorant binding proteins have been proposed as promising building blocks for construction of optical biosensors for dangerous substances such as toxic, explosive molecules and so on. It has been shown that the ligand-binding site of lipocalins can be optimized to bind molecules that are structurally different from their cognate ligands. For example, mutant OBPp has been successfully utilized for the monitoring of polyaromatic hydrocarbons that are among the most dangerous pollutants in water and atmosphere (Wei et al., 2007).

Thus, the construction of biosensor systems with sensors derived from ligand-binding proteins is one of the most promising areas in modern science. The development of an increasing number of such biosensor systems is dictated by the requirements of our daily lives, as they can be applied to the control of foodstuff quality and drug purity, environmental safety, detection of explosive and dangerous substances, monitoring of clinically relevant molecules and narcotics control (Tolosa, 2010).

5. Conclusion

The ligand-binding proteins, which are diverse in their structure and function, have numerous applications. Still, there is another way to enhance their practical importance. In particular, proteins from extremophilic sources, e.g., from hyperthermophilic bacteria (Staiano et al., 2010), could broaden the exploitation of ligand-binding proteins owing to their high stability under extreme conditions. In addition, ligand-binding proteins are convenient subjects of inquiry to elucidate one of the most intriguing and perplexing questions in structural and molecular biology, namely, the problem of protein folding into a unique, compact, highly ordered and functionally active form and, especially, the role of ligands in the structure and in the stabilization of proteins in their native states.

6. Acknowledgment

Financial support from the Ministry of Education and Science (Contracts 02.740.11.5141 and 16.512.11.2114), the Program MCB RAS and from the SPb government (OVS, 2011) is gratefully acknowledged.
7. References


protein isoforms is driven by phosphorylation. *Journal of chemical ecology*, Vol.36, No.8, (June 2010), pp. 801-813, ISSN 0098-0331


Since the dawn of recorded history, and probably even before, men and women have been grasping at the mechanisms by which they themselves exist. Only relatively recently, did this grasp yield anything of substance, and only within the last several decades did the proteins play a pivotal role in this existence. In this expose on the topic of protein structure some of the current issues in this scientific field are discussed. The aim is that a non-expert can gain some appreciation for the intricacies involved, and in the current state of affairs. The expert meanwhile, we hope, can gain a deeper understanding of the topic.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: