Conformational changes and ligand recognition of *Escherichia coli* D-xylose binding protein revealed

Sanjeewani Sooriyaarachchi[1], Wimal Ubhayasekera[1&], Chankyu Park[2], Sherry L. Mowbray[1,3]*

[1] Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, Biomedical Center, SE-751 24, Uppsala, Sweden

[2] Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Yusong-Ku, Taejon, Korea 305-701Korea Advanced Institute of Science and Technology (KAIST), Korea.

[3] Department of Cell and Molecular Biology, Box 596, Biomedical Center, Uppsala University, SE-751 24 Uppsala, Sweden.

*Corresponding author: E-mail: mowbray@xray.bmc.uu.se; Tel: 46-18-471 4990; Fax: 46-18 53 69 71

[&]Present address: MAX-lab, Lund University, Box 118, S-221 00 Lund, Sweden, and Institute of Medicinal Chemistry, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.

Abbreviations used

ABC, ATP binding cassette; ALBP, allose binding protein; ESRF, European synchrotron radiation facility; GBP, glucose/galactose binding protein; RBP, ribose binding protein; XBP, xylose binding protein.

Abstract

ABC transport systems account for most import of necessary nutrients in bacteria. The periplasmic binding component (or an equivalent membrane-anchored protein) is critical to recognizing the cognate ligand and directing it to the appropriate membrane permease. Here we report X-ray structures of D-xylose-binding protein from *Escherichia coli* in ligand-free open, ligand-bound open and ligand-bound closed forms, at 2.15, 2.2, and 2.2-Å resolution, respectively. The ligand-bound open form is the first such structure to be reported at high resolution; the combination of the three different forms from the same protein furthermore gives unprecedented detail concerning the conformational changes involved in binding protein function. As is typical for the structural family, the protein has two similar globular domains, which are connected by a three-stranded hinge region. The open liganded structure shows that xylose binds first to the C-terminal domain, with only very small conformational changes resulting. After a 34° closing motion, additional interactions are formed with the N-terminal domain; changes in this domain are larger, and serve to make the structure more ordered near the ligand. An analysis of the interactions suggests why xylose is the preferred ligand. Further, a comparison with the most closely related proteins in the structural family shows that the conformational changes are distinct in each type of binding protein, which may have implications for how the individual proteins act in concert with their respective membrane permeases.

Keywords: protein structure, conformational changes, periplasmic binding protein, ligand, ABC transport systems

Introduction

For survival, living organisms must take up necessary substances such as sugars and amino acids through their cellular membranes. ATP binding cassette (ABC) systems[1] are actively engaged in the high-affinity transport of most of these molecules in bacterial systems. Such systems consist of three major components, i.e. the periplasmic (in Gram-negative bacteria) binding protein (the true recognition component), the membrane-bound permease, and a cytoplasmic ATP-hydrolyzing protein that provides the energy needed. In Gram-positive bacteria, the periplasmic components are replaced by membrane-anchored lipoproteins that are otherwise very similar in form and function.

Both *Salmonella enterica serovar Typhimurium (S. typhimurium)* and *Escherichia coli* can utilize D-xylose as their sole carbon source.[2 3] The sugar can enter *E. coli* cells *via* either a binding protein dependent or a low affinity transporter system.[4-6] The XylFGH operon represents the ABC transport system, with the periplasmic D-xylose-binding protein (XBP), ATP binding unit and membrane permease encoded as xylF, xylG and xylH, respectively. After entry into the cell, D-xylose is isomerized to form D-xylulose by xylulose isomerase. Xylulokinase converts this product to D-xylulose 5-phosphate, which can then enter the non-oxidative phase of the pentose phosphate pathway.[2 7]

The sugar binding protein super-family shows a common structural pattern that consists of two globular domains, in most cases connected by a two- or three-stranded hinge region, with the binding cleft located in the domain interface.[8] XBP belongs to the cluster of binding proteins that are specific for hexose and pentose sugars. Motions at the hinge allow opening/closing of these proteins, which is important in their function.[9] The periplasmic binding proteins can thus, in theory,

appear as open unliganded, open liganded, closed liganded and closed unliganded species.[9] It is the closed, ligand bound form that is linked to correct recognition of the permease; however, the binding protein must open again within the complex for transfer of the ligand to the permease to take place.

In this study, we report the X-ray structures of *E. coli* XBP in three of the four forms, i.e. the open unliganded, open liganded and closed liganded conformations. These structures thus reveal the conformational changes of XBP, as well as the basis of its ligand specificity. Comparison to other related binding proteins indicates that each has a distinct open-closed motion, which probably reflects the motions needed to act in concert with the corresponding permeases of the transport systems.

Results

Open ligand-free XBP structure (XBP-open-apo)

The structure of the ligand-free form of XBP at 2.15-Å resolution was solved by molecular replacement using individual domains of the closest available structure, *Thermoanaerobacter tengcongensis* ribose binding protein (*Tt*RBP,[10] sequence identity 29%), as search models. Statistics for data collection and refinement are summarized in Table 1. The first two amino acids are not visible in the electron density, and that for the C-terminal histidine tag is also missing. Some side chains, notably in the 13-19 loop, also have weak density.

The overall structure consists of two similar domains connected by a 3-stranded hinge (see Fig. 1a). The N-terminal domain is assembled from a long segment (residues 1-104) and a short segment (residues 242-283), accounting for 146 residues of the 307-residue mature protein. The C-terminal domain also consists of two segments (105-241 and 284-307), with a total of 161 residues. Residues 103–106,

242-244, and 283-284 serve as the hinge region; there are designated as hinge segments I, II and III, respectively. An additional feature is the hydrogen bond linking Glu52 and Glu78, indicating that at least one side chain bears a proton at the pH of crystallization (7.5).

In this open structure, a phosphate molecule interacts with Asn196 and Asp222 (Fig. 2a, b). Trp169 also lies within van der Waals contact distance of the ion. The crystallization conditions include 0.2 M ammonium dihydrogen phosphate.

Open liganded XBP structure (XBP-open-xyl)

A D-xylose-bound open structure was determined to 2.2-Å resolution, in the same space group as the XBP-open-apo structure (Table 1). Again, electron density for the first two residues is missing, and density for some side chains in the loop near residue 15 is also weak.

D-Xylose is bound as the β-pyranose, with the preferred chair conformation. Sugar-protein interactions depend solely on the C-terminal domain (Fig. 2c,d). O1 of xylose makes hydrogen bonds with Asp135 and a water molecule, but is slightly more distant from Asn137 (~3.4 Å). O2 and O3 interact with a water molecule each, as well as with Lys242 with longer distances (3.2-3.3 Å); O3 also interacts with Asp222. O4 of the xylose makes hydrogen bonds with Asn196 and Asp222. O5 makes no hydrogen bonds with protein, but is largely exposed to solvent. Xylose binding is further strengthened by hydrophobic interactions with Phe141 and, more particularly, Trp169.

Closed liganded XBP structure (XBP-closed-xyl)

A closed structure with bound xylose was also solved by molecular replacement, using intact *Tt*RBP as the search model to locate the three molecules of the asymmetric unit, and refined to a resolution of 2.2 Å (Fig. 1b). Only the first molecule (chain A) has good electron density for the C-terminal histidine tag, which is located completely outside the main fold of the structure. There are some differences between the molecules in the asymmetric unit. Molecules B and C are most similar, with an r.m.s. difference of 0.2 Å when all Cα atoms are compared. Comparison of these to molecule A gives r.m.s. differences of 0.4-0.5 Å. All molecules have the same conformation, but exhibit changes of close to 2 Å in some surface loops, particularly near residue 265. Although temperature factors are higher than average (~45 $Å^2$), the loops are clearly defined in the electron density.

The ligand is clearly visible in all three molecules as a chair-form β-pyranose buried in the cleft between the two domains (Fig. 1b). Residues from both domains participate in the sugar-protein interactions (Fig. 2e,f). Arg16, Asp90 and Arg91 from the N-terminal domain interact with O3, O2 and O1, respectively; Arg91 also makes a hydrogen bond with the ring oxygen. Asp135, Asn137, Asn196, Asp222 and Lys242 from the C-terminal domain make hydrogen bonds with the sugar. Altogether 12 hydrogen bonds link the protein and the sugar molecule. As for the open structure, hydrophobic interactions with Phe141 and Trp169 are observed.

Comparison of the XBP structures

The XBP-open-apo and XBP-open-xyl structures are very similar (Fig. 3), with an r.m.s distance of 0.2 Å, and 299 Cαs matching within a 0.5-Å cutoff. Only residues 15, 71, and 133-136 do not meet the cutoff criterion. However, Glu15 shows

somewhat weaker electron density in both open structures, consistent with general mobility in this loop.

Compared to either open structure, the domains in the XBP-closed-xyl structures are rotated with respect to each other by ~34° (Figs. 1 and 3). When comparing residues of the N-terminal domain alone, ~65% of the Cαs match within a 0.5 Å cutoff; those that do not lie for the most part in loops at the edge of the binding pocket (including residues 11-19 and 41-45). In the C-terminal domain, ~85% of the Cαs match using the same criteria; the exceptions are residues 132-136, 154-155, 167-170 and 208-211. The regions around residue 265 that differ most among the three XBP-closed-xyl molecules also have higher temperature factors (~55 $Å^2$) in the XBP-open structures. The differences between the open and closed forms from the standpoint of the sugar are illustrated in Fig. 4. Sugar binds first to the C-terminal domain, then closure of the protein brings residues of the N-terminal domain closer, after some reorganization, especially in the loop near residues 16 and 17.

Dihedral angles calculated using Cαs only[11] were used to identify the most relevant changes of the main chain in the hinge segments during closing. In hinge segment I, the largest change was at residues 104-105 (due to differences of ~15° at 104-φ and 105-φ). In segment II, changes in the φ angles at residues 242-244 (25-35°) are combined. The largest change in hinge segment III is a modest 11° at 284-φ.

Ramachandran outliers are observed at Asp90, Asp222 and Thr295 in all three structures. Asp90 and Asp222 are binding site residues that lie next to the first two hinge segments, and are key to their correct structure and function. Thr295 is in a distorted (π-like) helix near the C-terminus, and lacks an obvious functional role.

Comparison to other sugar-binding proteins

Alignment of related sequences from *Ochrobactrum anthropi* (58% identity), *Ralstonia solanacearum* (58%), *Anaerocellum thermophilum* (53%), *Candidatus koribacter versatilis* (53%), *Acidobacteria bacterium* (two sequences, 51%), *Bacillus licheniformis* (47%), and *T. tengcongensis* (41%; a distinct sequence from that seen in the 2IOY structure) shows conserved residues indicating that D-xylose is likely to bind to these proteins in a similar way to that observed here (Fig. 5), despite the fact that some of these (and similar proteins) are labeled as only binding proteins, with the cognate ligands not stated. Certain residues of the hinge region are also conserved (Fig. 5).

Some binding protein structures of special interest are compared to XBP in Table 2. The most similar structure available is still *Tt*RBP (PDB code 2IOY, 29% sequence identity); it should be emphasized that this is not the same as the putative D-xylose-binding protein (**NP_621981.1**) mentioned in the previous paragraph. Other sugar-binding proteins from *E. coli* are also included in Table 2; some are aligned with XBP in Fig. 5.

Ligand affinity measurements Fluorescence changes, presumably reflecting differences in the environment of Trp169, were used to follow the binding of sugars. XBP showed an excitation maximum at 283 nm and an emission maximum at 345 nm. Binding of sugars caused a fluorescence enhancement of up to 35% (xylose > ribose >> arabinose) and a slight blue shift. For D-xylose, the K_D value estimated was 0.13 μM at 25 °C. Titration with D-ribose under similar conditions resulted in a calculated K_D value of 6 μM. L-arabinose exhibited a K_D value of 11 μM in the same experimental setting.

Discussion

Binding proteins open and close continuously in solution, with the open form favored in the absence of ligand, and the closed form in its presence. This is the first time that structures of unliganded open, liganded open and liganded closed form of one individual periplasmic binding protein have been presented together, and the first time that the structure of an open, ligand-bound binding protein has been available at high resolution. (An earlier report for the leucine/isoleucine/valine binding protein at 2.8-Å resolution[12] showed that the ligand bound to the N-terminal domain of the open protein, but did not provide details of the interactions or how the protein was altered by binding.) Therefore, the three structures of XBP allow us to describe ligand mediated conformational changes with more confidence than has been possible in the past.

Our structures show that D-xylose binds first to the C-terminal domain of XBP, and that only minimal conformational changes ensue in the open form. Xylose is bound as a chair-form of the β-pyranose most common in solution.[13] The first contact between sugar and protein is thus a relatively passive event; the sugar is in a low energy form, and the protein does not need to change greatly to acquire it. In this form of xylose, all hydroxyl groups are equatorial, and so the sugar is perfectly symmetrical in shape. Two orientations were therefore considered, that shown here, and a second that is "flipped" by 180°. In both cases, 7 hydrogen bonds would exist between the ligand and open protein. The ring oxygen does not hydrogen bond to the protein in either situation, and so this could not be used to discriminate between the two possible modes of binding. The chosen docking, however, places this oxygen in a more polar environment, and C5 in a less polar setting near Trp169. Furthermore,

when the protein closes, a hydrogen bond to the ring oxygen is made (with Arg91) that is lacking in the flipped orientation. The docking shown thus has a total of 12 sugar-protein hydrogen bonds in the closed form, one more than for the alternate choice.

Closing results from rotations in main-chain dihedral angles in all three hinge strands, which are concentrated in a small number of residues. Three water molecules that interact and with O1, O2 and O3 in the open form are replaced by interactions with Arg16, Asp90, and Arg91 of the N-terminal domain. In addition, some conformational changes are observed within the N-terminal domain, particularly near the sugar-binding residue, Arg16, which make the protein more ordered; a number of the hydrogen bonds between ligand and protein become shorter and stronger, as well (Figs. 3 and 4).

Binding of the less common (~35% of the total population) α-anomer of D-xylose is not supported by the electron density; this would disturb the hydrogen bonds that O1 makes with Asp135 and Asn137. Our fluorescence measurements revealed that D-xylose binds tightly to XBP, with an estimated K_D of 0.13 μM, while D-ribose showed an ~50 times higher K_D (6 μM). Xylose and ribose are C3 epimers, and the weaker binding of the latter can be explained by the loss of three hydrogen bonds (to Arg16, Asp222 and Lys242). Further, O3 appears to clash with Gln221. Another physiologically interesting ligand, L-arabinose, was also considered, largely because of its presence in the medium during induction. This sugar is a C4 epimer of D-xylose, and binds ~100-fold less tightly (K_D 11 μM); hydrogen bonds to Asn196 and Asp222 are lost, at the same time as O4 is placed in an unfavorable environment near Leu14 and Trp17. The C2 epimer, lyxose, was not tested, because of it scarcity in nature. Binding of hexoses does not appear likely. This would give rise to clashes

near Leu14, and furthermore, the hexoses most similar to xylose are the very rare gulose and idose.

The most similar among the available structures (assessed using DALI[14]) is the *Tt*RBP used as search model in the molecular replacement (Table 2). Interestingly, XBP is more similar to this protein (29% amino acid sequence identity) than it is to *E. coli* ribose binding protein (RBP; 23% identity). *Tt*RBP and RBP themselves have 57% sequence identity. Given the large number of sequences in the databases that are annotated as unknown, or with a specificity that is probably incorrect, an analysis of the available structures and sequences was undertaken to gain a better understanding of the criteria one should use in making decisions about ligand specificity.

All of the sugars considered, whether pentose or hexose, are bound as chair forms of the pyranose, supporting the idea that binding proteins are designed to select the most common (stable) sugar form in solution. This could facilitate binding to the open forms, in which there are relatively few protein-sugar interactions. Inspection of the structures indicates that D-xylose, D-ribose and D-allose are bound in similar orientations in their respective binding proteins. The sequence identity of RBP to ALBP is 35%, substantially higher than the 23-24% of either to XBP. Some conserved residues have similar functions in binding sugar, particularly in the RBP/ALBP pair. Two highly conserved residues among the three are Asp90 and Asp222, both of which are also Ramachandran outliers. In ALBP and RBP, each residue makes interactions with two hydroxyls of the sugar, which is the case for only Asp222 in XBP. Asn196, which makes an interaction with O4 is also very well conserved (although there is sometimes a conservative replacement by aspartate). The aromatic residues are allowed some degree of variation, e.g. the binding-site

tryptophans of XBP are replaced by phenylalanines in RBP. Some hinge residues, notably Asp105 and Asn106, are also conserved.

D-glucose (or D-galactose) and L-arabinose bind to their cognate receptors in ways that differ from that seen for the other sugars, as well as from each other. The sequence identity of these proteins is not significantly lower than for the cluster discussed in the previous paragraph, i.e. ~20% as compared to ~23%. Again, the Ramachandran outliers (equivalent to Asp90 and Asp222) are highly, although not absolutely, conserved; however, these residues do not fulfill the same roles in binding particular sugar atoms. Generally, we conclude that patterns of conservation in binding site residues provide a much more reliable method of distinguishing the sugars bound than overall sequence identity.

Based on studies of RBP[15] and ALBP,[16] we expect that the open form of XBP observed here is selected by crystal-packing interactions, from a family of related conformations present in solution. There is indeed no requirement for a unique open conformation, although the open (ligand-free) and closed (ligand-bound) conformations must be clearly distinguished, if they are to function correctly within the ABC system. Coupled rotations at ϕ and φ angles in the three hinge segments of XBP will lead to rotation of the domains around a set of axes that pass near residue 104, and between residues 242 and 243. Comparison to RBP and ALBP shows that the motions in different members of this structural cluster are distinct (Fig. 6). RBP opens by rotation around axes passing closest to 104 and 241 (numbered as the equivalents in XBP), while in ALBP, the axes pass near 105, and between 240 and 241; the rotation axes for comparable pairs of open/closed structures differ by 20-30° in direction. The motion seen for a given protein is determined primarily by the sequences in and near its hinge segments. Sequence differences alter local hydrogen-

bonding patterns of both protein and water (freeing some torsion angles, and restraining others), as well as the nature of the non-polar residues that must be reorganized as motion progresses. The third hinge segment lies on the surface of the protein and is less constrained by local structure than the first two; changes here are generally spread over a larger number of main-chain dihedrals, rather than being concentrated in a few. The differences in the motions are even greater if GBP is included in the analysis (there is no open structure of the arabinose binding protein to allow comparison). Thus each of the binding proteins that has so far been investigated shows a unique type of motion that is likely to be matched to that of the opening/closing occurring when it is in complex with the membrane permease during transport. As such, the favored motions provide an additional level of specificity in the transport system.

Materials and Methods

Cloning, expression and protein purification

The gene coding for *E. coli* XBP was amplified by PCR using the primer pair 5′ATGAAAGAAGTCAAAATAGGTATGG 3′ (forward), and 5′AGCGGCCGCTTAATGATGATGATGATGATGCAGCTCGCTCTCTTTGTGG 3′ (reverse), which add codons for 6 histidine residues to the C-terminus of the protein. Taq DNA polymerase (New England Biolabs) was used for amplification. The purified DNA fragment was TOPO-cloned into pCR T7/CT-TOPO (Invitrogen). The resulting plasmid was transformed into *E. coli* Top10 cells (Invitrogen) and the correctness of the clone was verified by sequencing. *E. coli* BL21-AI was transformed with recombinant plasmids for protein production. The cells were grown

at 37 °C in 4 L of Luria-Bertani broth containing 100 µl/ml ampicillin to an OD_{600} of 0.5-1.0, and expression of the target gene was induced by adding L-arabinose to a final concentration of 0.2% (w/v). After 3 h, the cells were harvested by centrifugation, and the cell pellet was resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5% (v/v) Triton X-100, 0.01 mg/ml RNase A and 0.02 mg/ml DNase) and lysed using a cell disrupter (Constant Systems Ltd, Daventry, UK). The soluble fraction was incubated with 1.5 mL pre-equilibrated Ni-nitrilotriacetic acid agarose (Qiagen, Valencia, CA) slurry for 1 h at 4 °C. The resin was then washed with 20 column volumes of 50 mM imidazole in the same buffer, and the protein eluted with six column volumes of 250 mM imidazole in the same buffer. The eluted fractions containing the desired protein were pooled and further purified on a HiLoad 16/60 Superdex 75 preparative grade column (GE Healthcare) pre-equilibrated with 20 mM HEPES, pH 7.5, 300 mM NaCl. The eluted fractions were concentrated using a Vivaspin concentrator (10 kDa cutoff, Vivascience).

To remove any endogenously bound sugar, the purified protein sample was incubated at room temperature for 30 min with 8 M urea, then dialyzed in steps against 6, 4, 2, 1 and 0 M urea in 10 mM HEPES buffer, pH 7.5. The final, concentrated protein sample was analyzed by SDS and native PAGE to confirm the purity and homogeneity, and stored in 10 mM HEPES buffer, pH 7.5, at -20 °C.

Crystallization

XBP-open-apo was crystallized using the hanging-drop vapor diffusion method at room temperature. Drops were composed of 1 µl of protein solution (15 mg/mL, 0.4 mM) and 1 µl of mother liquor (21% w/v polyethylene glycol 3350, 0.2 M

ammonium dihydrogen phosphate, pH 7.5). Crystal formation was facilitated by streak seeding immediately after setup. Thin plate-like crystals appeared after 48 h. Prior to data collection, the crystals were stabilized by a cryoprotectant solution (35% w/v polyethylene glycol 3350 in the same buffer) and flash-cooled in liquid nitrogen. To obtain the XBP-open-xyl structure, the crystals were soaked in a 1.3 mM xylose solution (in 35% w/v polyethylene glycol 3350, 0.2 M ammonium dihydrogen phosphate, pH 7.5) for 30 seconds prior to flash cooling in liquid nitrogen.

The XBP-closed-xyl structure was obtained by co-crystallizing the protein with the relevant sugar in a 1:3 protein to ligand ratio (0.5 mM protein to 1.5 mM D-xylose). The hexagonal-shaped crystals appeared in 2.4 M sodium malonate, pH 7.0. The crystals harvested were flash-cooled directly in liquid nitrogen without introducing a cryoprotectant.

Data collection, structure solution, refinement and model building

X-ray data were collected at 100 K at the European Synchrotron Radiation Facility (ESRF), in Grenoble, France. The data were processed with MOSFLM[17] and scaled with SCALA.[18] Analysis of the unit-cell content of the XBP-open forms suggested that there would be one molecule in the asymmetric unit, consistent with a solvent content of 42% and a V_M of 2.2 $Å^3$/Da.[19] For the XBP-open-apo form, molecular replacement was performed with MOLREP,[20] as implemented in the CCP4 interface.[21] *Tt*RBP (PDB entry code 2IOY), which has 29% sequence identity, was used as the search model. The domains were used separately for the molecular replacement. The C-terminal domain of the *Tt*RBP structure, which is the largest, was used first. The combined solution with both domains was improved by rigid-body refinement in REFMAC5.[22] Model building was then carried out using ARP/wARP,[23]

and the structure rebuilt in O[24] with the improved maps, a process that was repeated several times. The protein was subjected to restrained refinement in REFMAC5 and rebuilding in a cyclical fashion. Waters were placed using the ARP/wARP-solvent command in CCP4. The XBP-open-xyl structure was solved starting with the protein only from the ligand-free model, followed by rigid-body refinement, then cyclic rebuilding and restrained refinement.

For the XBP-closed-xyl structure, unit-cell content analysis suggested that there could be 3-5 molecules in the asymmetric unit. However, molecular replacement using intact *Tt*RBP, which is also a closed form, as the search model indicated that there were three molecules with a solvent content of 66%. Rigid-body refinement was followed by cyclic rebuilding and restrained refinement, making use of ARP/wARP as described above. Statistics for the data processing and final refined models are presented in Table 1.

Binding measurements

Dissociation constant (K_D) determination was carried out using the fluorescence titration method with an SLM 4800S spectrofluorimeter (SLM instruments, Rochester, NY); excitation and emission wavelengths were set at 283 nm and 340-349 nm, and excitation and emission bandwidths set at 2 and 16 nm. The experiment was carried out at 25 °C, and the tryptophan fluorescence change accompanying ligand binding were measured. The protein and sugar solutions included 10 mM HEPES, pH 7.5. For the K_D determinations for D-xylose, a protein concentration of 2.5 μM was used and 20 readings were taken for each sugar concentration. Sugar solution was added in 1-μl increments, using a stock solution of 100 μM, until the fluorescence enhancement reach a maximum. The corrected sugar concentration at

each step was taken into account in the calculation of the K_D. The K_D determinations for D-ribose and L-arabinose were performed using a protein concentration of 5 μM, and sugar stock solutions of 500 and 665 μM, respectively. The results were verified by repeating the experiments. K_Ds were determined by nonlinear least-squares analysis,[25] fitting the data to the equilibrium binding equation $(\Delta F_{max}/(2\times n))\times((K_D/P_0+n+L_0/P_0)-\text{sqrt}((K_D/P_0+n+L_0/P_0)^2 -4\times n\times L_0/P_0)$, where ΔF_{max}= maximal fluorescence change accompanying the interaction at saturation, n = binding stoichiometry (assumed to be 1), K_D = dissociation constant for ligand binding to the protein, P_0 = total protein concentration, and L_0 = total ligand concentration.[26]

Accession numbers

Structures and reflection data have been deposited at the PDB (www.pdb.org) with entry codes 3M9W, 3M9X and 3MA0 for the XBP-open-apo, XBP-open-xyl and XBP-closed-xyl structures, respectively.

Acknowledgments

This work was supported by grants from the Swedish Research Council (VR). We thank ESRF staff members for their support during the data collection. We would especially like to thank Sophia Schedin Weiss, Institute of Medical Biochemistry and Microbiology, Uppsala University, for her very kind help in the use of the fluorescence spectrofluorimeter.

References

1. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev **72**, 317-364.
2. Lin, E. C. C. (1987). Dissimilation pathways for sugars, polyols, and carboxlates. . In: *Escherichia coli* and *Salmonella typhimurium*: Cellular and

Molecular Biology (Neidhart, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. and Umbarger, H. E. Eds.), 244-284.

3. Shamanna, D. K. & Sanderson, K. E. (1979a). Uptake and catabolism of D-xylose in *Salmonella typhimurium* LT2. Journal of Bacteriology **139**, 64-70.
4. Sumiya, M., Davis, E. O., Packman, L. C., Mcdonald, T. P. & Henderson, P. J. F. (1995). Molecular-genetics of a receptor protein for D-xylose, encoded by the gene xylF in *Escherichia coli*. Receptors & Channels **3**, 117-128.
5. Song, S. & Park, C. (1997). Organization and regulation of the D-xylose operons in *Escherichia coli* K-12: XylR acts as a transcriptional activator. J Bacteriol **179**, 7025-7032.
6. Henderson, P. J. F. (1990). Proton-linked sugar-transport systems in bacteria. Journal of Bioenergetics and Biomembranes **22**, 525-569.
7. David, J. D. & Wiesmeyer, H. (1970). Control of xylose metabolism in *Escherichia coli*. Biochim Biophys Acta **201**, 497-499.
8. Tam, R. & Saier, M. H., Jr. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol Rev **57**, 320-346.
9. Quiocho, F. A. (1990). Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. Philos Trans R Soc Lond B Biol Sci **326**, 341-351.
10. Cuneo, M. J., Tian, Y., Allert, M. & Hellinga, H. W. (2008b). The backbone structure of the thermophilic *Thermoanaerobacter tengcongensis* ribose binding protein is essentially identical to its mesophilic *E. coli* homolog. BMC Struct Biol **8**, 20.
11. Kleywegt, G. J. (1996a). Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr D Biol Crystallogr **52**, 842-857.
12. Sack, J. S., Saper, M. A. & Quiocho, F. A. (1989). Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. J Mol Biol **206**, 171-191.
13. Angyal, S. J. & Pickles, V. A. (1972). Equilibria between pyranoses and furanoses. Australian Journal of Chemistry **25**, 1695-1710.
14. Holm, L. (1998). Unification of protein families. Curr Opin Struct Biol **8**, 372-379.
15. Bjorkman, A. J. & Mowbray, S. L. (1998). Multiple open forms of ribose-binding protein trace the path of its conformational change. J Mol Biol **279**, 651-664.
16. Magnusson, U., Chaudhuri, B. N., Ko, J., Park, C., Jones, T. A. & Mowbray, S. L. (2002). Hinge-bending motion of D-allose-binding protein from *Escherichia coli*: three open conformations. J Biol Chem **277**, 14077-14084.
17. Leslie, A. G. (1999). Integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr **55**, 1696-1702.
18. Evans, P. R. (1993). Data reduction. Proceedings of CCP4 study weekend 1993 on data collection & processing, 114-122.
19. Matthews, B. W. (1968). Solvent content of protein crystals. J Mol Biol **33**, 491-497.
20. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. Journal of Applied Crystallography **30**, 1022-1025.

21. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. Acta Crystallographica Section D-Biological Crystallography **59**, 1131-1137.
22. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography **53**, 240-255.
23. Cohen, S. X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., Sixma, T. K., Lamzin, V. S., Murshudov, G. N. & Perrakis, A. (2008). ARP/wARP and molecular replacement: the next generation. Acta Crystallogr D Biol Crystallogr **64**, 49-60.
24. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A **47 (Pt 2)**, 110-119.
25. Schedin-Weiss, S., Richard, B., Hjelm, R. & Olson, S. T. (2008). Antiangiogenic forms of antithrombin specifically bind to the anticoagulant heparin sequence. Biochemistry **47**, 13610-13619.
26. Krause, J., Buhner, M. & Sund, H. (1974). Studies of glutamate-dehydrogenase - binding of NADH and NADPH to beef-liver glutamate-dehydrogenase. European Journal of Biochemistry **41**, 593-602.
27. Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallographica Section A **42**, 140-149.
28. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**, 3389-3402.
29. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology **340**, 783-795.
30. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**, 4673-4680.
31. Bjorkman, A. J., Binnie, R. A., Zhang, H., Cole, L. B., Hermodson, M. A. & Mowbray, S. L. (1994). Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. J Biol Chem **269**, 30206-30211.
32. Chaudhuri, B. N., Ko, J., Park, C., Jones, T. A. & Mowbray, S. L. (1999). Structure of D-allose binding protein from *Escherichia coli* bound to D-allose at 1.8 Å resolution. J Mol Biol **286**, 1519-1531.
33. Kleywegt, G. J. (1997). Validation of protein models from C-alpha coordinates alone. Journal of Molecular Biology **273**, 371-376.
34. Kleywegt, G. J. & Jones, T. A. (1996c). Phi/Psi-cology: Ramachandran revisited. Structure **4**, 1395-1400.
35. Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1988). Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. Science **242**, 1290-1295.

Figure Legends

Figure 1: Overall structure of XBP.
(a) In the XBP-open-apo structure, a phosphate ion bound to the C-terminal domain is shown in black. (b) In the XBP-closed-xyl structure, the bound xylose molecule is shown in black. In both cases, the protein structures are color-coded using a scheme going from blue at the N-terminus through the rainbow to red at the C-terminus; the ends are labeled.

Figure 2: The ligand-binding pocket.
Panel (a) shows the hydrogen bonding between the bound phosphate and XBP in the sugar-free structure. Panels (c) and (e) illustrate the hydrogen bonds between xylose and XBP in the open and closed structures, respectively. In the latter case, the distances given are the average in the three molecules of the asymmetric unit. Residues in blue represent the N-terminal domain, and red, the C-terminal domain. Panels (b), (d) and (f) show the electron density of the binding-site residues in the final SIGMAA-weighted $2m|F_0|-d|F_c|$ maps[27] contoured at 0.26, 0.28 and 0.29 e/Å^3, respectively. Density is colored blue for the N-terminal domain, red for the C-terminal domain, and black for the sugar and waters. Carbons are goldenrod, gold and sky blue in the open-apo, open-xyl and closed-xyl structures, respectively.

Figure 3: Conformational changes of XBP.
A stereo view shows a superposition of the Cα backbone of the N-terminal domain of the open-apo (goldenrod), open-xyl (golden) and closed-xyl (sky blue) structures of XBP, with xylose shown in similar colors.

Figure 4: Binding of xylose.
Stereo views in panels (a) and (b) represent the open-xyl (carbons and backbone in gold) and the closed-xyl (carbons and backbone in sky blue) structures; they are shown with similar views with respect to the C-terminal domain. Panel (c) shows the superimposed structures (again using the C-terminal domain), with the same coloring convention.

Figure 5: Sequence alignment.
Related sequences were identified by a BLAST search;[28] the signal sequences were identified using SIGNALP[29] and removed before alignment with ClustalW.[30] Conserved substrate-binding residues are highlighted in blue (N-terminal domain) and red (C-terminal domain), while conserved hinge residues are marked in green. The sequences were annotated as follows, with the number of residues given in parentheses: **NP_621981.1**, periplasmic sugar-binding protein from *T. tengcongensis* MB4 (328); **YP_077505.2**, putative D-xylose-binding protein *B. licheniformis* ATCC 14580 (330), **YP_002572036.1**, D-xylose ABC transporter, periplasmic ABC sugar transporter, periplasmic binding protein from *A. thermophilum DSM 6725* (327); **YP_589984.1**, ABC sugar transporter, periplasmic binding protein from *C. Koribacter versatilis Ellin345* (333); **ZP_00944131.1**, D-xylose-binding protein from *R. solanacearum* (317); **YP_001371949.1**, D-xylose ABC transporter, periplasmic substrate-binding protein from *O. anthropi* ATCC 49188 (310). The sequences from five other known structures (see Table 2) were also included: *Tt*RBP from *T. tengcongensis* (2IOY, 283), as well as the D-ribose (271), D-glucose (309), D-allose (280), and L-arabinose binding (305) proteins from *E. coli*.

Figure 6. Comparison of XBP, RBP and ALBP conformational changes.
C-terminal domains of open and closed forms of the three proteins were aligned as follows: XBP, open (gold) and closed (sky blue), opening angle 33°; RBP, open (A molecule of PDB entry 1URP[15], magenta) and closed (2DRI[31], purple), opening angle 43°; ALBP, open (B molecule of PDB entry 1GUD[16], green) and closed (1RPJ[32], dark green), opening angle 33°. Rotation axes for opening/closing, calculated using the *lsq_explicit* option in *O*, are shown in matching colors. Where multiple open forms are available, that with the most similar opening angle was chosen.

Table 1. Data collection and refinement statistics.

Data Collection[a]

	XBP-open-apo	**XBP-open-xyl 2.2Å**	**XBP-closed-xyl**
Environment	ESRF ID14:4	ESRF ID14:1	ESRF ID29
Wavelength (Å)	0.955	0.9334	0.9762
Space group	$P2_1$	$P2_1$	$P3_221$
Unit cell constants (Å, °)	34.6 72.2 66.3 90.0 100.5 90.0	34.7 71.7 66.3 90.0 100.5 90.0	156.0 156.0 103.6 90.0 90.0 120.0
Resolution (Å)	20.00 - 2.15 (2.27 - 2.15)	30.00 - 2.20 (2.25 - 2.20)	20.00 - 2.20 (2.32 - 2.20)
Unique reflections	17,127	13,825	141,545
Average multiplicity	3.2 (3.2)	2.5 (1.9)	2.0 (2.0)
Completeness (%)	97.5 (98.8)	84.7 (58.1)	98.8 (99.0)
Wilson B value ($Å^2$)	27.4	27.9	30.3
R_{merge} (%)[b]	8.40 (37.2)	7.60 (31.6)	9.9 (40.0)
$\langle I/\sigma(I) \rangle$	10.4 (3.4)	11.8 (2.6)	6.5 (2.1)
Refinement			
Number of reflections	16,242	13,126	134,285
Completeness (%)	97.4 (98.8)	84.5 (54.6)	98.8 (98.8)
Resolution range (Å)	20.0 - 2.15 (2.2 - 2.15)	30.0 - 2.2 (2.25 - 2.2)	20.0 - 2.2 (2.25 - 2.2)
R-factor, R-free (%)	20.0, 26.0 (25.0, 30.0)	19.7, 25.5 (26.9, 33.7)	18.2, 20.2 (26.0, 28.6)
No. of protein atoms (Average B, $Å^2$)[c]			2396 (25.0)
A molecule	2318 (29.4)	2318 (26.8)	2327 (23.8)
B molecule	-		2327 (32.8)
C molecule	-		
No. of water molecules (Average B, $Å^2$)[c]	94 (28.9)	89 (25.5)	463 (30.6)
No. of ligand atoms (Average B, $Å^2$)[c]	-	10 (23.0)	30 (19.1)
No. of phosphate atoms (Average B, $Å^2$)[c]	5 (51.8)	-	-
Rms bond length (Å)	0.010	0.017	0.016
Rms bond angle (°)	1.20	1.63	1.30
Ramachandran plot outliers, number (%)[d]	6 (2.1%)	7 (2.5%)	17 (2.0%)
PDB code	3M9W	3M9X	3MA0

[a] Values in parenthesis are for the highest resolution cell.

[b] $R_{merge} = \sum_\eta \sum_i | I_{\eta i} - \langle I_\eta \rangle | / \sum_\eta \sum_i | I_{\eta i} |$.

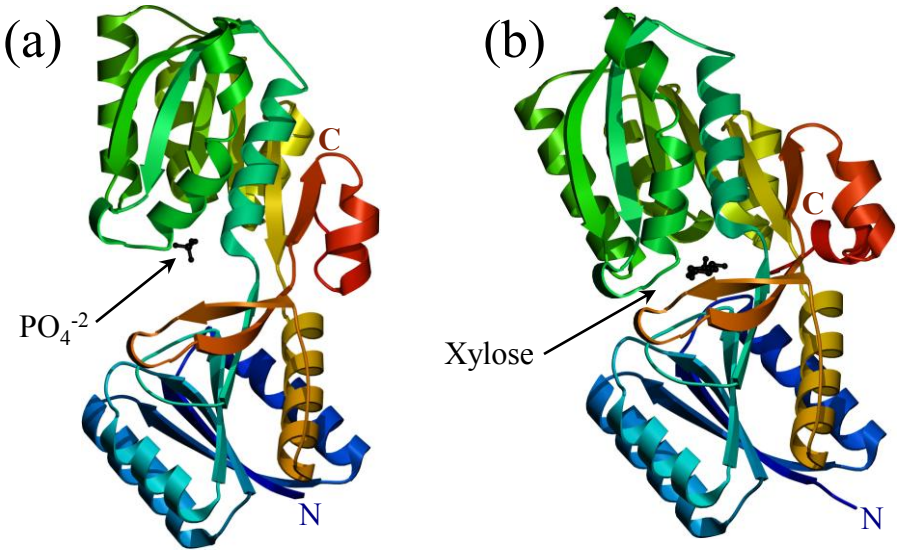
[c] Calculated using MOLEMAN.[33]

[d] A stringent-boundary Ramachandran plot was used.[34]

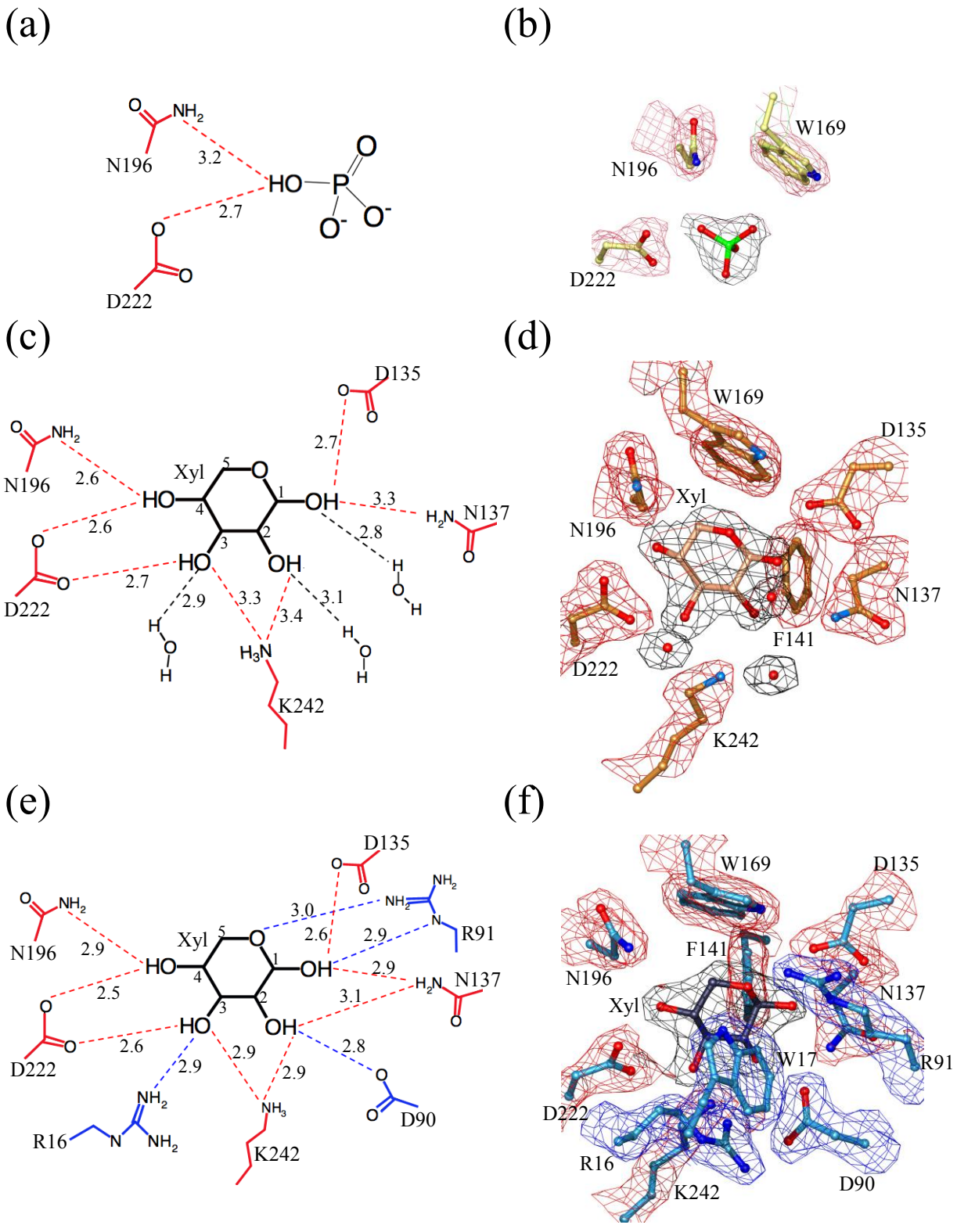
Table 2. Comparison to ligand-bound forms of other binding proteins.

Source	Name	PDB entry code	Z-score	r.m.s.d.	# res.	Sequence identity to XBP	Ligand	Reference
T. tengcongensis	D-ribose binding protein	2IOY	36.0	1.4	274	29	β-D-ribose	10
E. coli	D-ribose binding protein	2DRI	35.2	1.5	271	23	β-D-ribose	15
E. coli	D-allose binding protein	1RPJ	33.2	1.9	288	24	β-D-allose	29
E. coli	D-glucose/galactose binding protein	2GBP	33.3	2.2	309	21	β-D-glucose	35
E. coli	L-arabinose binding protein	1ABE	30.1	2.6	305	18	α-L-arabinose β-L-arabinose	31

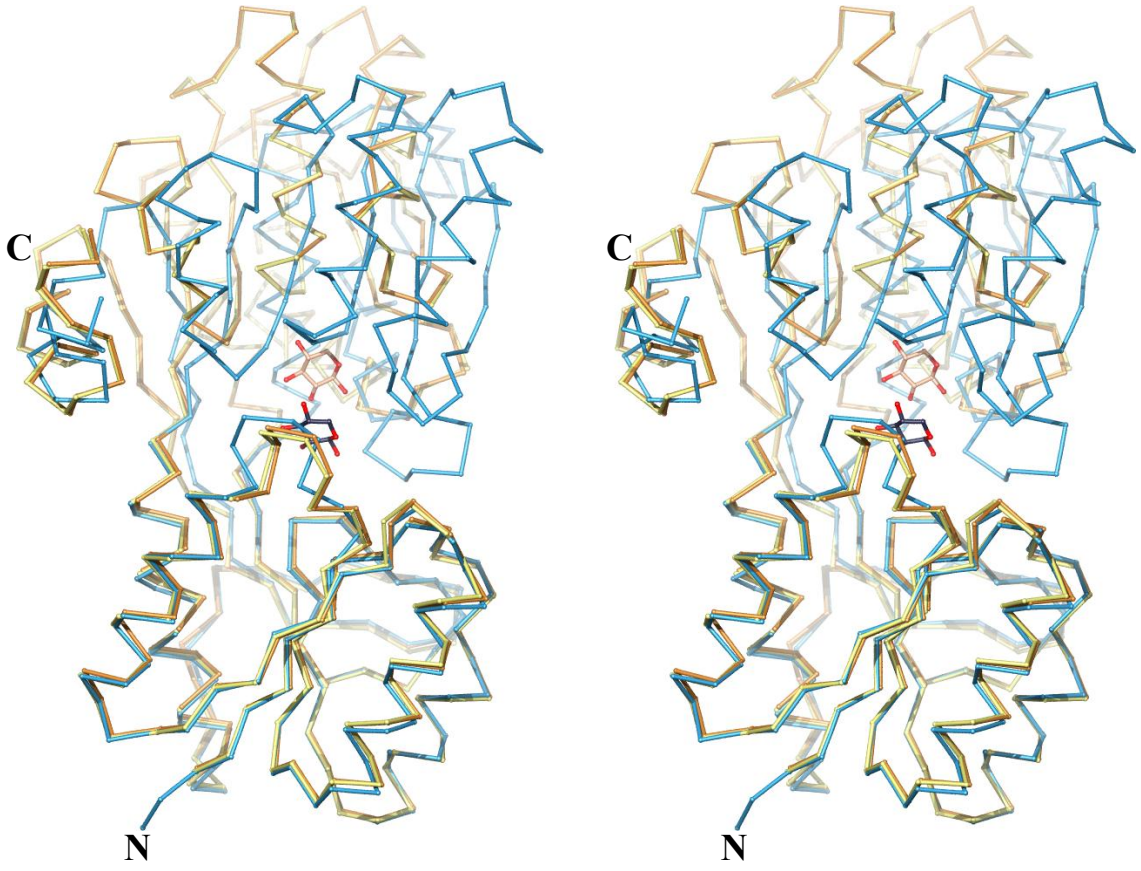
Figure



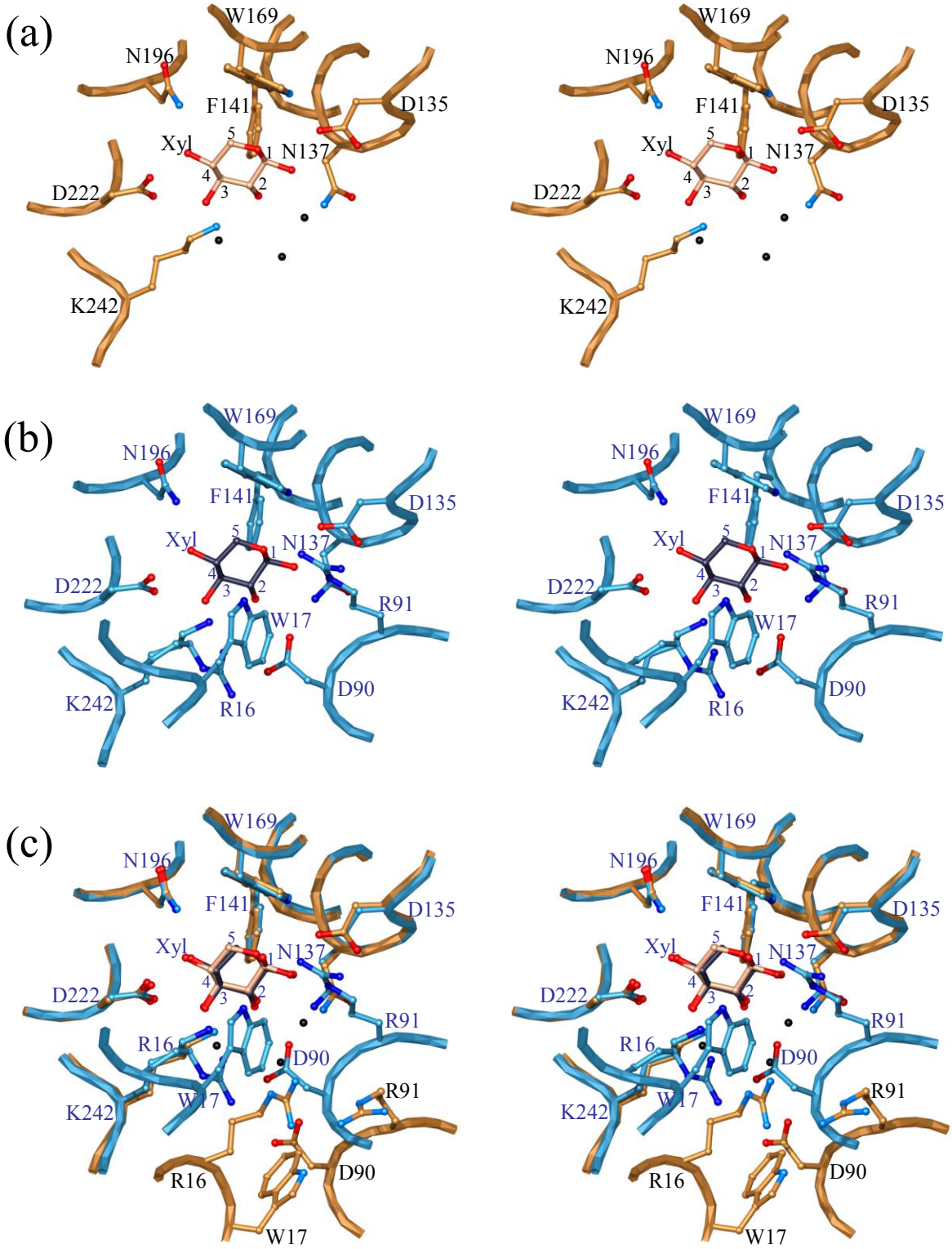
Figure


(a)
N196
3.2
HO
P
O
O-
O-
2.7
D222
(b)
W169
N196
D222
(c)
D135
2.7
N196
2.6
Xyl
HO
OH
3.3
H2N
N137
2.6
2.8
D222
2.7
HO
OH
2.9
3.3
3.1
3.4
H3N
K242
(d)
W169
D135
Xyl
N196
N137
F141
D222
K242
(e)
D135
3.0
2.6
2.9
R91
N196
2.9
Xyl
HO
OH
2.9
H2N
N137
2.5
3.1
D222
2.6
HO
OH
2.9
2.9
2.8
2.9
D90
R16
K242
(f)
W169
D135
F141
N196
N137
Xyl
W17
R91
D222
R16
K242
D90

Figure

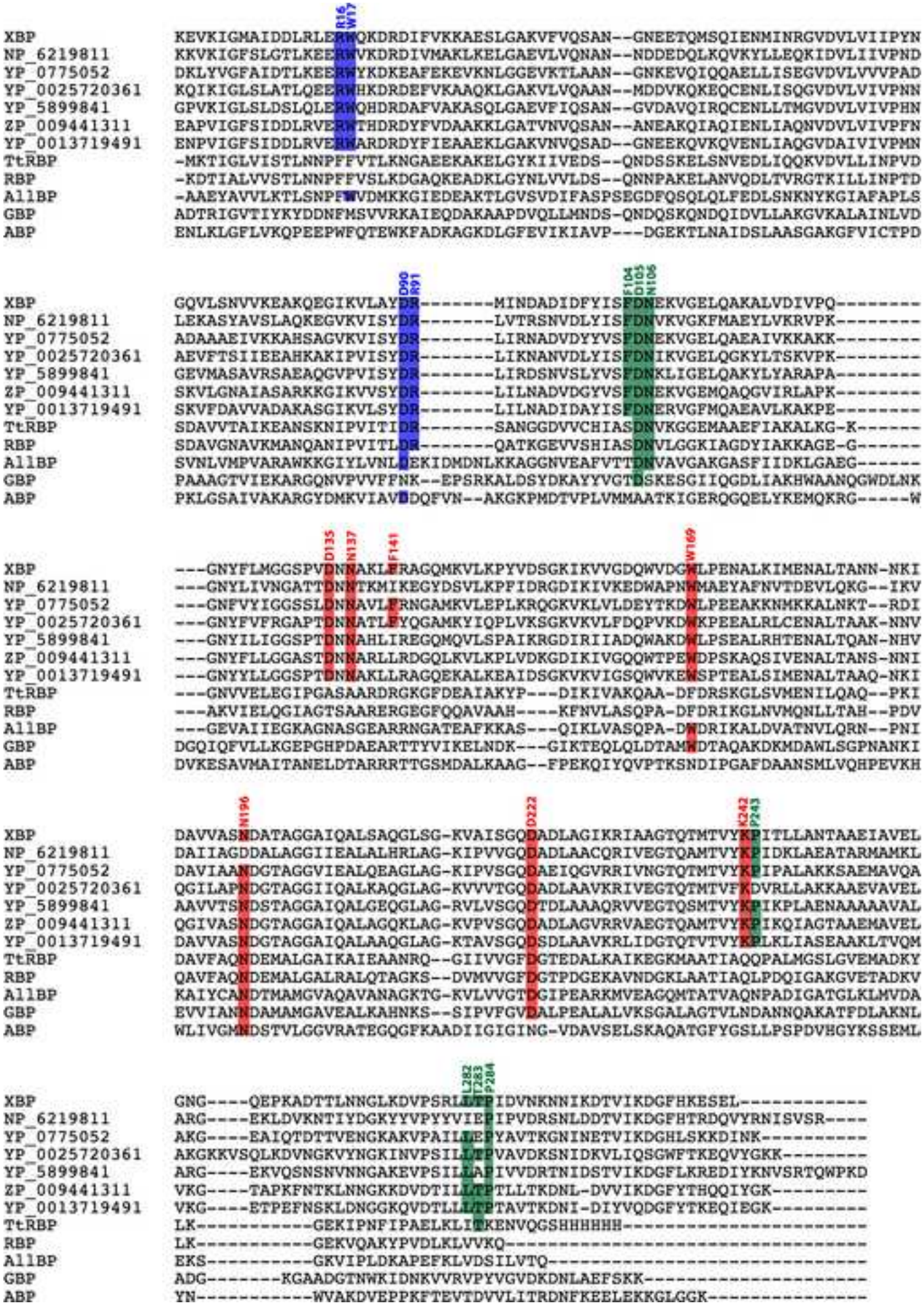


Figure


(a)
W169
N196
F141
D135
Xyl
N137
D222
K242
(b)
W169
N196
F141
D135
Xyl
N137
D222
R91
W17
K242
R16
D90
(c)
W169
N196
F141
D135
Xyl
N137
D222
R91
R16
D90
K242
W17
R16
D90
W17

Figure

Click here to download high resolution image



Figure