J. Mol. Biol. (1997) **273**, 226–237

JMB

The 1.6 Å Crystal Structure of the AraC Sugar-binding and Dimerization Domain Complexed with D-Fucose

Stephen M. Soisson[1], Beth MacDougall-Shackleton[2], Robert Schleif[2] and Cynthia Wolberger[1,3]*

[1]*Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore MD 21205-2185, USA*

[2]*Biology Department, Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218, USA*

[3]*Howard Hughes Medical Institute, Johns Hopkins University School of Medicine Baltimore, MD 21205-2185 USA*

**Corresponding author*

The crystal structure of the sugar-binding and dimerization domain of the *Escherichia coli* gene regulatory protein, AraC, has been determined in complex with the competitive inhibitor D-fucose at pH 5.5 to a resolution of 1.6 Å. An in-depth analysis shows that the structural basis for AraC carbohydrate specificity arises from the precise arrangement of hydrogen bond-forming protein side-chains around the bound sugar molecule. van der Waals interactions also contribute to the epimeric and anomeric selectivity of the protein. The methyl group of D-fucose is accommodated by small side-chain movements in the sugar-binding site that result in a slight distortion in the positioning of the amino-terminal arm. A comparison of this structure with the 1.5 Å structure of AraC complexed with L-arabinose at neutral pH surprisingly revealed very small structural changes between the two complexes. Based on solution data, we suspect that the low pH used to crystallize the fucose complex affected the structure, and speculate about the nature of the changes between pH 5.5 and neutral pH and their implications for gene regulation by AraC. A comparison with the structurally unrelated *E. coli* periplasmic sugar-binding proteins reveals that conserved features of carbohydrate recognition are present, despite a complete lack of structural similarity between the two classes of proteins, suggesting convergent evolution of carbohydrate binding.

Keywords: AraC; carbohydrate recognition; fucose; arabinose; convergent evolution

Introduction

The AraC protein controls the expression of genes in *Escherichia coli* necessary for uptake and catabolism of L-arabinose (for a review, see Schleif, 1996). In the absence of arabinose, the AraC protein actively represses transcription of the ara_{BAD} operon by selectively binding to the I_1 and O_2 operator half-sites and forming a protein-mediated DNA loop of approximately 210 base-pairs (Dunn *et al.*, 1984; Hendrickson & Schleif, 1985; Lobell & Schleif, 1990; Martin *et al.*, 1986). The binding of arabinose to AraC causes the protein to cease DNA looping and favor binding to the I_1 and I_2 half sites (Hendrickson & Schleif, 1984; Lobell & Schleif, 1990), both of which are located adjacent to the binding site for RNA polymerase. Occupancy of the I_1 and I_2 half-sites results in activation of transcription from the p_{BAD} promoter (Zhang *et al.*, 1996).

The full-length AraC protein consists of 292 residues and is divided into two functional domains. The amino-terminal residues 1 to 170 constitute the sugar-binding and dimerization domain that is linked to the carboxyl-terminal DNA-binding domain (residues 178 to 292; Bustos & Schleif, 1993; Eustance *et al.*, 1994). A flexible, five residue linker that is protease-sensitive (Eustance *et al.*, 1994; Lauble *et al.*, 1989) and tolerant of amino acid substitutions (Eustance & Schleif, 1996) joins the two domains. Studies with chimeric proteins indicate that the sugar-binding and dimerization domain and the DNA-binding domain are functionally independent as they retain their respective activities when fused to other DNA-binding or dimerization domains (Bustos & Schleif, 1993). This argues against the presence of crucial interactions between the sugar-binding and dimerization domain and the DNA-binding domain. Several studies have shown that, *in vivo*, AraC

0022–2836/97/410226–12 $25.00/0/mb971314

functions as a dimer both in the presence and absence of arabinose (Hendrickson & Schleif, 1985; Lobell & Schleif, 1990; Wilcox & Meuris, 1976).

The structure of the AraC sugar-binding and dimerization domain has been determined both in the presence and absence of arabinose (Soisson *et al.*, 1997). The structure of the AraC-arabinose complex shows that AraC binds arabinose within the open end of an eight-stranded anti-parallel β-barrel. The sugar-binding site is completed by the folding of the 12 amino-terminal residues that fully enclose the arabinose within the protein. C-terminal to the β-barrel is a ninth β-strand, followed by a pair of α-helices. The arabinose-bound protein dimerizes through an anti-parallel coiled-coil interface formed by the carboxy-terminal helix of each monomer. In the absence of arabinose, the 12 amino-terminal residues of AraC are disordered, thereby uncovering the vacant sugar-binding site and allowing it to serve as a second, and much larger, dimerization interface. Arabinose binding was predicted to regulate formation of this interface by direct competition for the sugar-binding site and by inducing the amino-terminal arm to fold (Soisson *et al.*, 1997). These predictions were confirmed by velocity sedimentation experiments showing that formation of this interface is, in fact, ligand-regulated. At elevated concentrations, the AraC sugar-binding and dimerization domain in the absence of arabinose exists as a range of species from dimers to higher-order oligomers (Soisson *et al.*, 1997). Upon addition of arabinose, the protein is converted into a purely dimeric form.

Doyle and co-workers systematically tested many related monosaccharides for the ability to activate AraC and found that only L-arabinose induces the ara_{BAD} operon. In contrast to the stimulatory effect of L-arabinose, the monosaccharide D-fucose (6-deoxy-D-galactose) behaves as a competitive inhibitor, binding to AraC, but failing to induce gene expression *in vivo* or *in vitro* (Beverin *et al.*, 1971; Doyle *et al.*, 1972; Englesberg *et al.*, 1965; Greenblatt & Schleif, 1971; Wilcox *et al.*, 1974). The β-methyl-L-arabinoside was also reported to act as a competitive inhibitor of AraC *in vivo*, although to a lesser extent than D-fucose (Doyle *et al.*, 1972). L-Arabinose and D-fucose are isosteric at all positions except C5, which in fucose contains a methyl group (Figure 1). The two sugars bind competitively to the same site on the protein with similar equilibrium dissociation constants (K_d 3×10^{-3} to 6×10^{-3} M), as measured by tryptophan fluorescence quenching (Wilcox, 1974). The low affinity *in vitro* for arabinose and fucose correlates well with *in vivo* experiments showing that half-maximal operon induction is achieved at an intracellular arabinose concentration of 10^{-3} M (Doyle *et al.*, 1972; Greenblatt & Schleif, 1971; Wilcox, 1974).

We report here the crystal structure of the amino-terminal sugar-binding and dimerization domain of AraC in complex with D-fucose at pH 5.5, determined at a resolution of 1.6 Å. Comparison of this structure with the previously reported structure of AraC in complex with L-arabinose (Soisson *et al.*, 1997) shows that the carbohydrate-binding specificity of AraC arises from the precise arrangement of hydrogen bond-forming residues in the protein located in fixed orientations around the bound sugar molecule. van der Waals interactions also contribute to the epimeric and anomeric selectivity of the protein. The ability of the sugar-binding pocket to accommodate the additional methyl group of fucose is the result of small structural rearrangements of three hydrophobic side-chains and part of the amino-terminal arm. A comparison of AraC with the structurally unrelated *E. coli* periplasmic sugar-binding proteins reveals several conserved features of sugar recognition, suggesting that the two classes of proteins have convergently evolved similar mechanisms for efficiently binding monosaccharides.

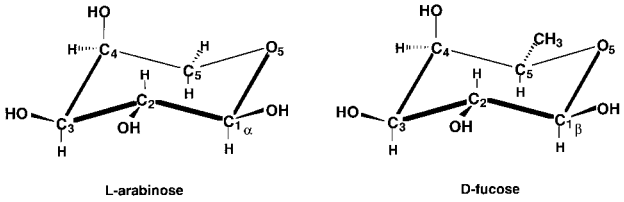

Figure 1. Comparison of L-arabinose and D-fucose. The Figure shows the structures of α-L-arabinose and β-D-fucose in the full-chair ($^4C_1$) conformation. In solution, these sugars exist as a mixture of anomers at the C1 position; however, only the equatorial configuration of the anomeric (C1) hydroxyl group is observed in the complex of these sugars with AraC.

Comparison of the fucose-bound AraC amino-terminal domain structure with that of AraC bound to arabinose (Soisson *et al.*, 1997) reveals negligible differences in either the overall conformation of the protein or in the dimerization interface. This result is surprising in light of the very different effects that arabinose and fucose binding have on transcriptional activation. Interestingly, while the crystals of AraC bound to fucose are highly isomorphous with those of the AraC-arabinose complex, crystals of the fucose-bound molecule were obtained only at pH 5.5, while the complex with arabinose could be readily crystallized at neutral pH. Taken together with solution studies that suggest that the oligomeric state of AraC in the presence of fucose is different at pH 5.5 and 7.0 (Soisson, 1997), the possibility arises that the structure of the AraC-fucose complex at neutral pH differs in some way from the structure we have determined at pH 5.5. Considering the structural information, mutagenesis data, and solution experiments, we speculate on the nature of the pH-dependent differences in the fucose-bound structure.

Results and Discussion

Structure determination

Crystals of the sugar-binding and dimerization domain of AraC complexed with D-fucose grown at pH 5.5 diffract X-rays to greater than 1.6 Å resolution and are nearly isomorphous with crystals grown in the presence of L-arabinose at pH 7 to 8 (Soisson *et al.*, 1997; Table 1). Interestingly, crystals of AraC in the presence of D-fucose form only at around pH 5.5, while the arabinose-bound protein crystallizes in the pH range 7 to 8 (Soisson *et al.*, 1997; and unpublished results). The structure of the AraC-fucose complex was determined by a combination of rigid-body refinement and difference Fourier techniques using as a starting model the structure of AraC complexed with L-arabinose that was determined at 1.5 Å resolution (Soisson *et al.*, 1997). After one round of simulated annealing refinement at 1.8 Å resolution, clear $F_o - F_c$ density greater than 3σ was visible for the missing C6 methyl group of D-fucose (Figure 2a), thereby confirming that fucose is bound to the protein in this crystal form.

The structure was refined at a resolution of 1.6 Å with X-PLOR (Brüger, 1992b) using standard protocols to a final *R*-factor of 17.6% and a free *R*-factor (Brünger, 1992a) of 22.2%. The final model includes residues 6 to 168 of each monomer in the asymmetric unit, 405 water molecules, two acetate ions, and has excellent stereochemistry and no Ramachandran plot outliers. Figure 2b shows a portion of the final $2F_o - F_c$ electron density map in the vicinity of the sugar-binding site. The four amino-terminal residues and the ten carboxy-terminal residues are not visible in the electron density maps, and are presumed to be disordered. Similar disorder was seen in the crystals of the arabinose-bound protein. The carboxy-terminal disorder is consistent with the observation that these residues form part of a flexible linker joining the sugar-

Table 1. Data collection and refinement statistics

Wavelength (Å)	0.689
Resolution limit (Å)	1.6
Measured reflections	293,876
Unique reflections	57,378
Completeness (%)	
All data	99.6
Highest shell (1.67 to 1.60 Å)	99.9
Highest shell ($I > \sigma$)	96.9
Overall $I/\sigma(I)$	19.7
R_{merge} (%)	9.0
Refinement	
Resolution range (Å)	7.0–1.6
Number of protein atoms	2619
Number of acetate atoms	8
Number of fucose atoms	20
Number of water molecules	405
Average *B*-factor, all non-hydrogen atoms (Å^2)	12.5
R_{factor}/R_{free} ($\|F\| > 2\sigma$) (%)	17.6/22.2
Stereochemistry	
rmsd bond length (Å)	0.010
rmsd bond angles (°)	1.57
rmsd dihedral angles (°)	24.8
rmsd improper angles (°)	1.29

$R_{merge} = \Sigma_{hkl}|I - \langle I\rangle|/\Sigma_{hkl}\langle I\rangle$, *I*: observed intensity, : $\langle I\rangle$average intensity of multiple observations of symmetry-related reflections. $R_{factor} = \Sigma_{hkl}|F_{obs} - F_{calc}|/\Sigma_{hkl}F_{obs}$. R_{free} is the *R* factor for a subset of 10% of the reflection data that were not included in the crystallographic refinement.

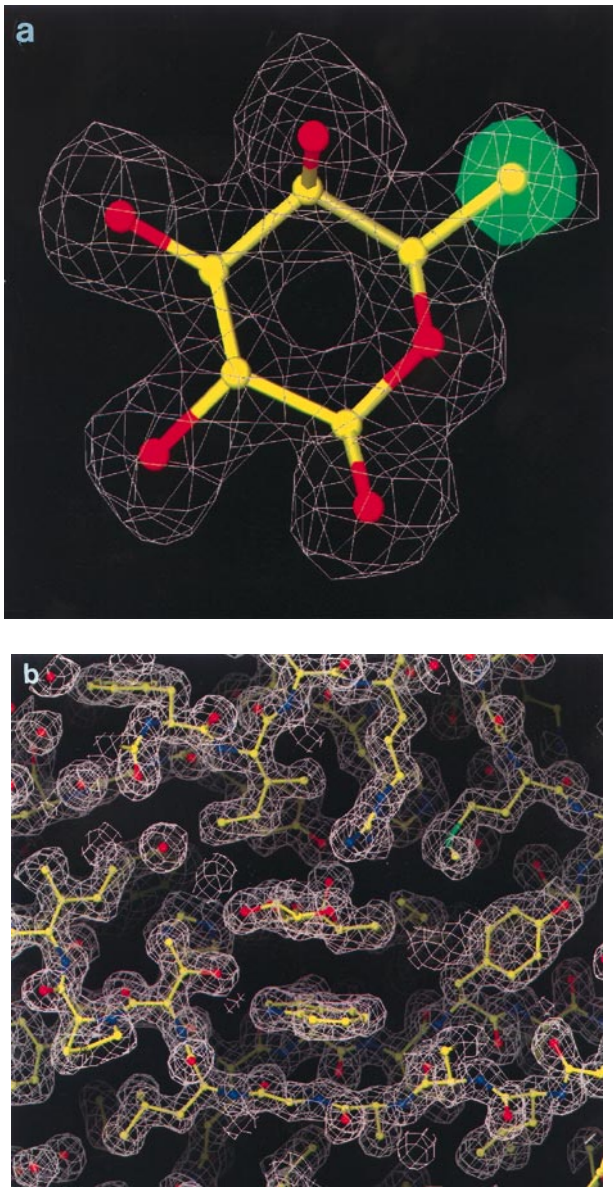

Figure 2. a, Difference electron density map showing presence of the C6 methyl group of D-fucose. After rigid-body refinement and one round of simulated annealing refinement using the structure of AraC complexed with L-arabinose, $2F_o - F_c$ electron density (8.0 to 1.8 Å, $|F| > 2\sigma$) was clear and unambiguous (gray) for the bound sugar. The missing methyl group of D-fucose appeared clearly as a peak greater than 3σ in the $F_o - F_c$ electron density (green semi-transparent ball). This Figure, as well as b, was produced using the program O (Jones *et al.*, 1991). b, Final $2F_o - F_c$ (7 to 1.6 Å, $|F| > 2\sigma$) electron density contoured at 1σ for the refined model in the vicinity of the sugar-binding site.

binding and dimerization domain to the DNA-binding domain (Bustos & Schleif, 1993; Eustance *et al*., 1994; Eustance & Schleif, 1996).

Overall structure of the sugar-binding and dimerization domain

The overall fold of the sugar-binding and dimerization domain of AraC appears to have no structural similarity to other known carbohydrate recognition domains, and has been described (Soisson *et al*., 1997). Figure 3 shows the protein sequence with secondary structure assignments derived from the crystal structures. The tertiary structure of the sugar-binding and dimerization domain (Figure 4) consists of an eight-stranded anti-parallel β-barrel with jelly-roll topology, followed by a long linker that contains two turns of 3_{10} helix and a ninth β-strand that is part of one sheet of the jelly roll. The last β-strand is followed by two α-helices, each approximately 20 amino acid residues in length, that pack against the outer surface of the barrel.

The functional dimer of AraC with bound arabinose at pH 7 to 8, and with bound fucose at pH 5.5, is presumed to be the dimer found in the asymmetric unit of both crystal forms. This dimer consists of two monomers interacting primarily by way of an anti-parallel coiled-coil consisting of the terminal α-helix (α2) of each monomer (Figure 4), as described (Soisson *et al*., 1997). The overall structure of AraC in the presence of both arabinose and fucose is highly conserved, with superimposed monomers having a root-mean-square difference in C^{α} positions of 0.3 Å (see Materials and Methods).

One molecule of α-L-arabinose or β-D-fucose (Figure 1) in the full-chair ($^4C_1$) conformation binds to each monomer of AraC within the open end of the β-barrel (Figure 4), occupying a small pocket lined with both polar and non-polar side-chains. The alignment of the arabinose and fucose structures (Figure 5b) shows that the positioning of D-fucose in the binding site is identical with that

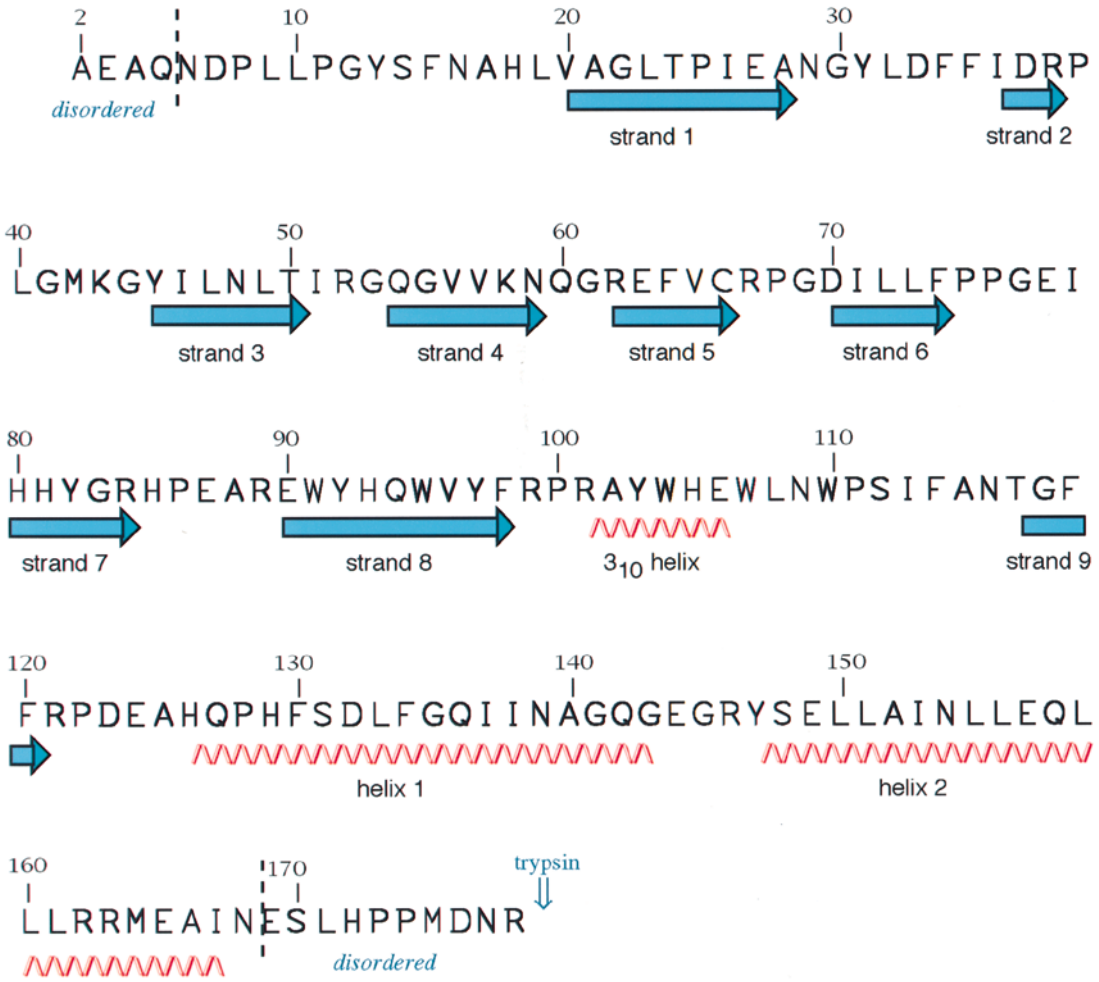

Figure 3. Amino acid sequence and corresponding secondary structure assignments derived from the crystal structures. Secondary structure assignments were made using the algorithm due to Kabsch & Sander (1983), as implemented in the program PROCHECK (Laskowski *et al*., 1993). The disordered regions at the termini of the protein are labeled and marked with a vertical broken line. The trypsin cleavage site, as determined by mass-spectrometry and amino-terminal sequencing, is marked with an arrow and labeled.

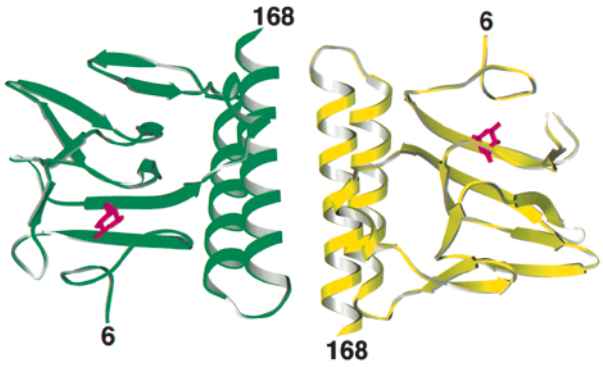

Figure 4. Ribbon diagram showing the anti-parallel coiled-coil dimer present in the asymmetric unit of the AraC plus D-fucose crystals. Monomers are in yellow and green, and the bound sugar is in purple. This Figure, as well as Figures 5b, 6, 7a and 7b was produced using the program SETOR (Evans, 1993).

seen in the complex with L-arabinose. The binding site is completed by the amino-terminal arm of the protein (residues 6 to 18), which loops around to close off the end of the β-barrel in which the sugar is bound. This arm is disordered in crystals of unliganded AraC (Soisson *et al.*, 1997), suggesting that sugar binding stabilizes the folded conformation observed in the arabinose and fucose-bound structures. This stabilization presumably results from the sugar-dependent formation of direct and water-mediated hydrogen bonds between the amino-terminal arm and the sugar-binding site, as well as the packing of Phe15 into the sugar-binding pocket (Figure 6). The folding of the amino-terminal arm in response to sugar binding plays an important role in the ligand-regulated oligomerization observed in solution and in the crystalline state (Soisson *et al.*, 1997). Subtle alterations in the sugar-binding site and conformation of the amino-terminal arm arise as a result of the methyl group on D-fucose and will be discussed below.

Carbohydrate binding: hydrogen bonding

AraC specifically binds ligand through the formation of direct hydrogen bonds with the sugar hydroxyl groups and pyranose ring oxygen atom, either through direct protein-sugar interactions or through water-mediated hydrogen bonds (Figure 5a and b). The pattern of hydrogen bonds to and from the sugar is identical between the arabinose and fucose structures, and the hydrogen bond distances are similar in the two cases (Figure 5b and Table 2). All sugar hydroxyl groups, including the anomeric hydroxyl group, serve simultaneously as both hydrogen bond donors and acceptors. This type of cooperative hydrogen bonding generally results in stronger than normal hydrogen bonds (Jeffrey & Lewis, 1978; Quiocho, 1986). Two of the sugar hydroxyl groups (OH3 and OH4) participate in hydrogen bonding patterns of the type $(NH)_n \rightarrow$ sugar-OH $\rightarrow$ O. The OH2 position receives a hydrogen bond from Thr24 and donates a hydrogen bond to a water molecule. Arg38 forms a bidentate contact to the ring oxygen (O5) and OH4, likely serving to neutralize a lone pair of electrons on the sugar ring oxygen atom. As only one hydrogen bond is made to the ring oxygen atom, the full hydrogen bond-forming capability of the two lone pairs of electrons on this atom are not utilized. It was unexpected to find that the anomeric hydroxyl group acts as both a hydrogen bond donor to the main-chain carbonyl group of Pro8 as well as a hydrogen bond acceptor from a water molecule (Figure 5a), since the anomeric hydroxyl group is thought to be a better than average hydrogen bond donor and a weaker than average hydrogen bond acceptor (Jeffrey & Lewis, 1978). By contrast, the anomeric hydroxyl in arabinose binding protein acts only as a hydrogen bond donor to the carboxylate group of an aspartic acid side-chain.

An extensive network of bound water molecules is found between the sugar-binding pocket and the amino-terminal arm (Soisson *et al.*, 1997). Figure 5a shows waters mediating hydrogen bonds primarily with protein main-chain atoms of residues Leu9, Leu10, Gly12, and Tyr13 in the amino-terminal arm. The position of the arm is further stabilized by the main-chain carbonyl group of Pro8, which forms a direct hydrogen bond with the anomeric hydroxyl group (OH1).

Carbohydrate binding: non-polar interactions

A hallmark of protein-carbohydrate interactions is the presence of numerous van der Waals contacts, often from aromatic side-chains, that serve to position the sugar in the binding site and enhance specificity (Cygler *et al.*, 1991; Maenaka *et al.*, 1994; Mowbray & Cole, 1992; Quiocho, 1986, 1988; Vyas, 1991; Vyas *et al.*, 1988). In AraC, the bound sugar stacks directly on the indole ring of Trp95 (Figure 5b) in a manner reminiscent of that seen in the glucose/galactose-binding protein (Vyas *et al.*, 1988). The planes of the sugar ring and the tryptophan ring are nearly parallel and are separated by approximately 3.5 Å (Figure 7b), creating a large, energetically favorable van der Waals interaction surface for the bound sugar molecule. The stacking of the sugar against Trp95 could account for the decrease in tryptophan fluorescence that is observed upon sugar binding (Wilcox, 1974), analogous to the effect seen in the periplasmic sugar-binding proteins (Boos *et al.*, 1972; Miller *et al.*, 1980; Parsons & Hogg, 1974; Quiocho *et al.*, 1977). In addition to Trp95, there are 56 van der Waals interactions of distance less than 4.0 Å in the L-arabinose complex, and 67 van der Waals interactions in the D-fucose complex. The increased number of contacts seen in the fucose structure primarily results from the presence of the methyl group on this sugar.

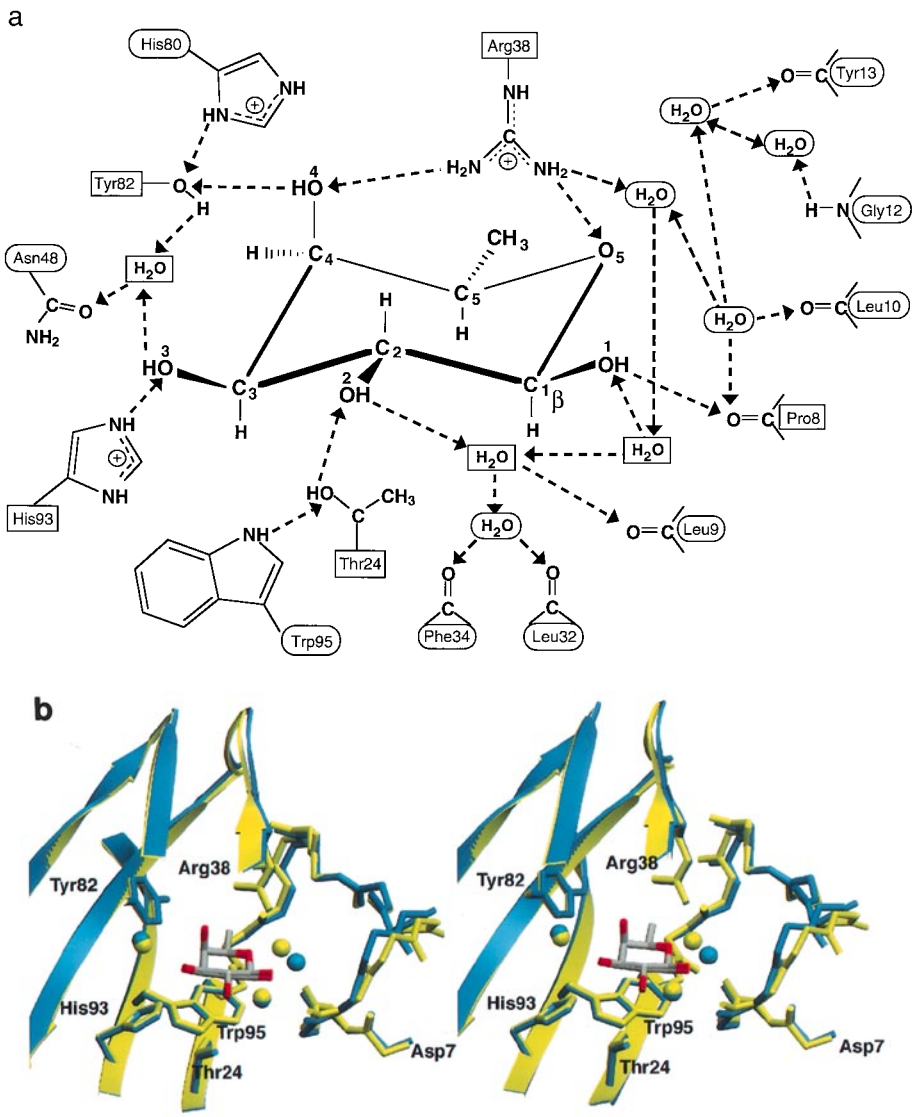

Figure 5. a, A drawing of the hydrogen bond interactions between D-fucose and AraC. Side-chains and water molecules making direct contact with the sugar molecule are boxed, while indirectly contacting groups are in ovals. Arrows point from hydrogen bond donor to acceptor. Histidine residues are assumed to be protonated, since the crystals were grown at pH 5.5. b, Stereo view of the sugar-binding site in the presence of D-fucose and L-arabinose. The D-fucose structure is in yellow, and the L-arabinose structure is in blue. Side-chains that form direct hydrogen bonds with the sugar molecule are shown in their respective colors. The side-chain of Trp95 has been included to help in orientation. The amino-terminal arm residues 7 to 18 are shown as main-chain atoms. The three bound water molecules that directly hydrogen bond with the sugar molecule are shown in their respective colors. The position of the water molecule hydrogen-bonded to the anomeric hydroxyl group shifts position by about 0.5 Å between the two structures, while the amino-terminal arm tilts away from the binding site in the D-fucose complex.

Based on the structure of the complex with L-arabinose, one would predict that the binding of D-fucose would be slightly disfavored due to steric hindrance between the C6 methyl group and the side-chain positions of Phe15, Met42 and Ile46. The structure of the complex with D-fucose shows that the methyl group of fucose is accommodated by small side-chain movements of Phe15, Met42 and Ile46 that relieve these steric constraints (Figure 6). Table 3 shows the various atoms that are in van der Waals contact with the methyl group of fucose. The interactions vary slightly between the two monomers in the asymmetric unit, most notably Met42 and Ile46. In one monomer, C^{ε} of Met42 is within van der Waals contact distance (3.8 Å) from the C6-methyl group of fucose; however, in the other monomer, this distance is increased to 4.4 Å. Likewise, the interaction distance between $C^{\delta 1}$ of Ile46 and the fucose methyl group is 3.2 Å in one monomer and 3.5 Å in the other.

The close, hydrophobic packing of the fucose-methyl group contrasts with that observed in the arabinose-binding protein-fucose complex, where several water molecules are found in close proximity to the methyl group and contribute to the weaker binding of fucose to arabinose-binding pro-

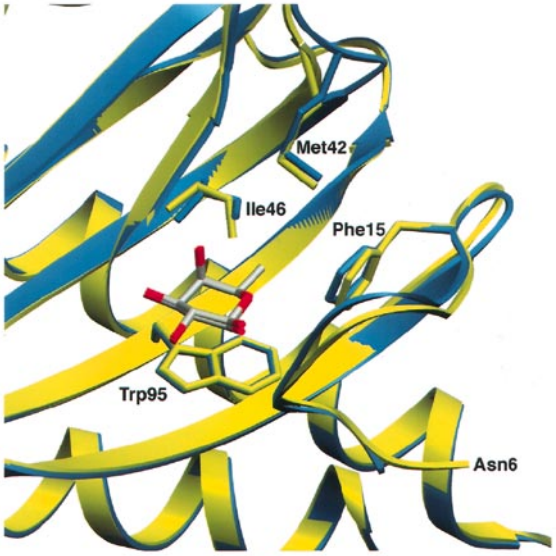

Figure 6. Alignment of the arabinose and fucose complexes showing van der Waals interactions with the C6 methyl group of D-fucose. The arabinose complex is in blue, and the fucose complex is in yellow. The side-chains of Ile46, Met42 and Phe15 all change position slightly to accommodate the methyl group of D-fucose. The movement of Phe15 (discussed in the text) causes the amino-terminal arm of the protein to slightly tilt away from the binding site.

tein as compared with the binding of arabinose (Quiocho *et al.*, 1989). The presence of these water molecules, with their inherent mobility, also accounts for the ability of the arabinose-binding protein to bind D-galactose (Quiocho *et al.*, 1989). In contrast, modeling studies (not shown) suggest that D-galactose binding to AraC would be unfavorable as a result of a steric clash between the C6-hydroxymethyl group and the non-polar side-chains of Ile46, Phe15 and Met42, a conclusion consistent with the fact that D-galactose is neither an activator nor a competitive inhibitor of AraC (Doyle *et al.*, 1972). It is likely that the tighter packing arrangement in the AraC sugar-binding pocket, as well as an unfavorable electrostatic environ-

Table 2. Comparison of hydrogen bonding in L-arabinose and D-fucose structures

Donor	Acceptor	Distance (Å)[a] L-Arabinose	D-Fucose
O1	Pro8 O	2.6/2.6	2.7/2.7
O2	Water	2.7/2.7	2.7/2.7
O3	Water	2.7/3.0	2.8/2.7
O4	Tyr82 OH	2.8/2.8	2.8/2.9
Water	O1	2.8/2.8	2.7/3.0
Thr24 $O^{\gamma 1}$	O2	2.8/2.7	2.8/2.8
His93 $N^{\varepsilon 2}$	O3	2.8/2.7	2.7/2.7
Arg38 NH1	O4	2.7/2.8	2.7/2.7
Arg38 NH2	O5	3.0/2.9	3.0/2.9

[a] Pairs of numbers represent values observed in each of the two monomers present in the crystallographic asymmetric unit. All interactions where at least one value is <3.0 Å are listed.

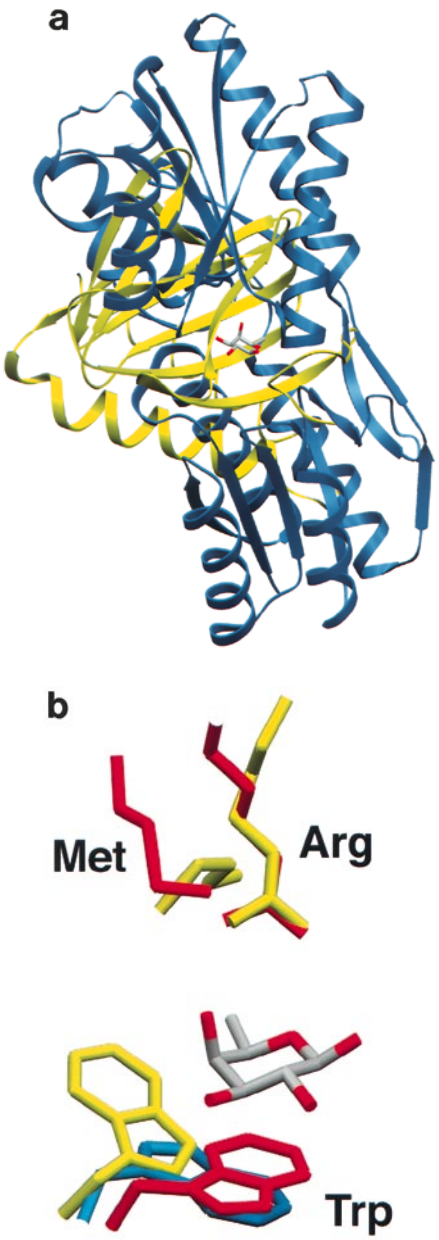

Figure 7. a, Superposition of the AraC monomer and arabinose-binding protein (Quiocho *et al.*, 1989) showing the lack of any structural similarity between the two classes of proteins. The two structures were manually aligned solely on the basis of the position of D-fucose molecules. b, Conserved interactions with D-fucose observed between arabinose-binding protein, glucose/galactose-binding protein (Vyas *et al.*, 1988), and AraC. The side-chains from AraC (Arg38, Met42, Trp95) are colored red, side-chains from arabinose binding protein (Trp16, Met108, Arg151) are colored yellow, and the side-chain from glucose/galactose-binding protein is colored blue. Although there is not a published structure of glucose/galactose-binding protein in complex with D-fucose, the structure of the complex with D-galactose (1GLG) was manually aligned to show the relative orientation of Trp16 with the pyranose ring of the structure with bound galactose.

Table 3. van der Waals interactions with the C6 methyl group of D-fucose

Protein group	Distance (Å)[a]	Mean distance (Å)
Phe15 $C^{\varepsilon 1}$	4.3/4.0	4.1
Phe15 $C^{\varepsilon 2}$	3.8/3.7	3.7
Phe15 C^{ζ}	3.5/3.6	3.6
Arg38 NH1	3.6/3.7	3.7
Met428 C^{ε}	4.4/3.8	4.1
Ile46 $C^{\delta 1}$	3.2/3.5	3.4
Trp95 $C^{\varepsilon 3}$	3.7/3.9	3.8
Trp95 $C^{\zeta 3}$	3.8/4.0	3.9

[a] Pairs of numbers represent values observed in each of the two monomers present in the crystallographic asymmetric unit. All interactions where at least one value is $\leqslant 4.0$ Å are listed.

ment, accounts for the enhanced ability of the protein to discriminate between fucose and galactose.

The local distortions induced by the binding of D-fucose to AraC have several global consequences. To make room for the methyl group of fucose, Phe15 is pushed against other side-chains, causing the amino-terminal arm to bulge away from the binding site (Figures 5b and 6). While for the purpose of discussion we have been comparing monomer B of each structure (see Materials and Methods), similar displacements are observed in other pairwise comparisons between the arabinose and fucose-bound structures. The distortion in the arm is localized between residues 12 to 16, and has a mean displacement in α-carbon positions of 0.8 Å. As a consequence of this movement, the network of water-mediated hydrogen bonds between the amino-terminal arm and the sugar-binding site is reorganized. One of these water molecules is directly hydrogen bonded to the anomeric hydroxyl group (OH1), and shifts position by about 0.5 Å in the fucose structure (Figure 5b). Although it is unclear whether this rearrangement contributes directly to the specificity of binding, it is another example of how bound water molecules can adjust their position in response to the presence of different ligands in a sugar-binding site (Quiocho *et al*., 1989).

Epimeric and anomeric specificity

Despite a relatively weak affinity for L-arabinose and D-fucose *in vitro* ($K_d \approx 10^{-3}$M, Wilcox, 1974), AraC binds these ligands with a high degree of specificity. The ability to discriminate between sugars that are of opposite chirality at a given carbon position (epimers), or between the two equilibrium positions of the anomeric OH1 hydroxyl group (anomers), arises from the disposition of hydrogen bond-forming groups in the protein, van der Waals interactions between the sugar and nonpolar side-chains in the binding site, and steric exclusion.

Hydrogen bonds to the sugar originate either from side-chains that are fixed in position or from main-chain atoms with restricted mobility. This greatly limits the ability of the protein to accommodate other sugar isomers (Figure 5a and b). For instance, binding of the epimer at the C2 position (L-ribose) would be disfavored due to the inability of Thr24 to form a hydrogen bond to OH2. The position of Thr24 is restricted by the formation of a simultaneous hydrogen bond to the N^{ε} of Trp95, and so is unlikely to be able to accommodate the change in position of the hydroxyl group. Similar arguments can be made for rationalizing the inability to bind epimers at the C3 position (L-lyxose) or C4 position (D-xylose). Additional epimeric specificity arises through van der Waals interactions with Trp95. The stacking of the sugar molecule against Trp95 sterically prohibits the presence of hydroxyl groups pointing from the sugar molecule towards the indole ring. For instance, the 3-epimer of L-arabinose (L-lyxose) could not bind in the same manner as arabinose due to the steric clash between OH3 and the indole ring of Trp95.

The stacking interaction between Trp95 and the sugar ring also plays a major role in the anomeric specificity of AraC. As noted in the structure of AraC with L-arabinose (Soisson *et al*., 1997), only one sugar anomer is seen in the binding site of AraC. In solution, arabinose and fucose exist as a 40:60 mixture of anomers that differ in the position of the anomeric hydroxyl groups, OH1 (Pigman & Anet, 1972). As a result of the anomeric effect, axial anomers are slightly more stable than equatorial anomers due to interactions with the lone pairs of electrons on the sugar ring oxygen atom, and an unfavorable dipole-dipole interaction (Angyal, 1972). Despite this, simulated-annealing omit maps (Brünger, 1992b) in which the sugar was removed from the structure consistently showed the presence only of the less-favored α-L-arabinose and β-D-fucose anomers in the binding site of AraC (note that the nomenclature for D and L sugars is reversed, such that β-D-fucose and α-L-arabinose have the same absolute configuration at the anomeric (C1) position; Figure 1). The anomeric selectivity of AraC results primarily from a steric clash that would occur between Trp95 and the anomeric hydroxyl group if the axial anomers were to bind.

Another contributing factor to the anomeric specificity could be the fact that the anomeric hydroxyl group (OH1) forms a direct hydrogen bond with the main-chain carbonyl group of Pro8. In contrast, the anomeric hydroxyl group in arabinose-binding protein acts solely as a hydrogen bond donor to the carboxylate group of an aspartic acid side-chain. Interestingly, there is no carboxylate group interacting with the bound sugar in AraC, even though these are an important feature of the periplasmic sugar-binding proteins. In the latter case, carboxylate groups appear to confer the ability to bind sugar hydroxyl groups in either the epimeric or anomeric position. For instance, arabinose-binding protein can bind both the α and β anomers of L-arabinose through the proper positioning of an aspartate residue (Quiocho & Vyas, 1984) and glucose/galactose-binding protein can

recognize both epimers at the C4 position by the same mechanism (Vyas *et al.*, 1988). Glucose/galactose-binding protein also utilizes an aspartate residue to recognize the anomeric position; however, for reasons discussed below, only the equatorial anomer of galactose is bound (Vyas *et al.*, 1988).

Conserved features of sugar binding

Despite an absence of any apparent structural similarity between AraC and the *E. coli* periplasmic transport proteins (Figure 7a), several conserved features of sugar binding are observed in the structures (Figure 7b). The positioning of Arg38 in AraC is nearly identical with the arrangement seen in the structures of arabinose-binding protein complexed with either arabinose or fucose (Quiocho & Vyas, 1984; Quiocho *et al.*, 1989). Quiocho & Vyas (1984) postulated that the guanidinium group of the arginine side-chain, with its positive charge and resonance stabilization, is especially good at neutralizing the anomeric effect. Given the observation that only the less-favored sugar anomers are bound by AraC, it seems likely that Arg38 plays a similar role in AraC.

The non-polar interactions between the protein and sugar are also similar between AraC and arabinose-binding protein. Both proteins have a tryptophan side-chain that packs against the relatively non-polar face of arabinose and fucose; however, in the case of arabinose-binding protein the indole ring is not as close to parallel with the sugar ring (Figure 7b). This difference in the tryptophan orientation allows room for the axial anomers of L-arabinose and D-fucose to bind to arabinose-binding protein, while these anomers cannot bind to AraC due to steric hindrance from Trp95. In AraC, the orientation of the tryptophan relative to the sugar is more similar to that seen in the glucose/galactose-binding protein, where the sugar is stacked directly on top of the indole ring in a parallel manner (Vyas *et al.*, 1988; Figure 7b). Glucose/galactose-binding protein exhibits a preference similar to AraC for binding the equatorial anomers of its ligands (Vyas *et al.*, 1988), lending further support for the role of steric hindrance from stacking interactions in the anomeric specificity of these proteins.

Another similar interaction is seen in the van der Waals interaction between a methionine residue and the bound sugar molecule. In the arabinose-binding protein, a methionine side-chain is in van der Waals contact with the C5 position of arabinose, as well as with the C6 methyl of fucose (Quiocho *et al.*, 1989). A similar methionine residue (Met42) is present in AraC; however, in the complex with arabinose the distance between the methionine methyl group and arabinose C5 is too large for van der Waals interactions (5.6 Å). When fucose is bound, the distance between the C6 methyl group and the methionine methyl group falls more closely into the expected range for favorable van der Waals interactions (Table 3, Figures 6 and 7b).

Although AraC and arabinose-binding protein are very similar in terms of the number of hydrogen bonds and van der Waals interactions with the bound sugar molecule, AraC binds its ligands (L-arabinose and D-fucose) with an equilibrium dissociation constant ($\sim 10^{-3}$M) that is four orders of magnitude weaker than that observed for complex formation of arabinose-binding protein with the same sugars (Quiocho, 1988). Reasons for this difference in affinity are unclear at this time, but weaker binding by AraC may result, in part, from the entropic cost of ordering the amino-terminal arm and the water molecules within the sugar-binding pocket upon ligand binding.

Conclusions

The detailed analysis of carbohydrate recognition by AraC helps explain specificity of ligand binding, and reveals that the carbohydrate recognition domain of AraC utilizes many of the common themes of carbohydrate recognition seen in other systems. The structures of the arabinose-bound (at neutral pH) and fucose-bound proteins (at pH 5.5) are nearly indistinguishable (r.m.s.d = 0.28 Å) and there is no significant change in the mode of sugar binding. Comparing the complex of the AraC sugar-binding and dimerization domain with D-fucose to the complex with L-arabinose shows that the C6 methyl group of D-fucose is accommodated by small movements of several non-polar side-chains that alleviate steric clashes that would normally result as a consequence of fucose binding.

The complete lack of structural similarity between AraC and the periplasmic sugar-binding proteins is somewhat surprising, given the presence in *E. coli* of other transcriptional regulators, namely the LacI-PurR family, that bind ligand in a domain structurally similar to that of the periplasmic binding proteins (Friedman *et al.*, 1995; Lewis *et al.*, 1996; Schumacher *et al.*, 1994). A detailed comparison between the two classes of carbohydrate-binding proteins reveals several very similar features to sugar-binding, and shows how a completely different protein scaffold can present the same side-chains in approximately the same locations to a bound ligand.

The question of why D-fucose inhibits transcriptional activation by AraC still remains. We note that, although the crystals of AraC complexed with D-fucose are almost perfectly isomorphous with those of the L-arabinose complex and the two structures are highly conserved, crystals of the fucose complex grew only at about pH 5.5, while the arabinose complex crystallizes readily in the pH range 7 to 8 (Soisson *et al.*, 1997). This raises the possibility that the structure of AraC in complex with D-fucose observed at pH 5.5 differs in some way from the structure of the AraC-fucose complex in the physiological pH range. Indirectly supporting this hypothesis are velocity sedimentation experiments performed in the presence of

D-fucose at pH 5.5 and at pH 7.5. At pH 5.5, the AraC-fucose complex sediments strictly as a dimer (Soisson, 1997), much like AraC in the presence of arabinose (Soisson *et al*., 1997). At neutral pH, however, AraC in the presence of D-fucose forms dimers and higher-ordered oligomers (Soisson, 1997), analogous to those seen in the absence of arabinose (Soisson *et al*., 1997). Taken together with the fact that AraC binds fucose at neutral pH with an affinity comparable to that for arabinose (Wilcox, 1974), there must be a pH-dependent structural change that permits oligomerization of AraC at neutral pH while fucose is bound. The failure of fucose-bound AraC to activate transcription supports the hypothesis that the ability of AraC to oligomerize at high concentrations is somehow correlated with the inability of the protein to activate transcription (Soisson *et al*., 1997).

The range of pH sensitivity suggests that a histidine residue, with a typical $pK_a \approx 6.3$, might somehow be involved in the pH effect. There are two histidine residues (His80 and His93) in the sugar-binding site (Figure 5a); however, protonation of these residues would probably not effect hydrogen bonding between the sugar and protein. This is in agreement with the competitive nature of fucose binding at neutral pH (Wilcox, 1974), and argues against histidine protonation being important for fucose binding.

The amino-terminal arm of AraC undergoes a transition from the unfolded to folded state upon arabinose binding and this transition plays a role in the ligand-regulated oligomerization observed in solution (Soisson *et al*., 1997). It is possible that the distorted conformation of the amino-terminal arm observed in the fucose-bound structure reflects an inherent instability of the arm when fucose is bound. Interestingly, there is a histidine residue in the amino-terminal arm (His18), which appears to form a favorable salt-bridge with Glu149 in the fucose structure, most likely due to protonation of the histidine residue at pH 5.5. In the arabinose-bound structure, which was determined at pH 7.5, electron density for this histidine residue is absent in one monomer and very poor in the other monomer, suggesting that protonation of His18 at pH 5.5 and salt-bridge formation with Glu149 might stabilize the conformation observed in the fucose-bound structure. This histidine residue is the last disordered residue of the amino-terminal arm in the absence of sugar and serves as the ''pivot-point'' for the turn that brings the arm over the sugar-binding site in the presence of ligand. In the presence of D-fucose at neutral pH, the His18-Glu149 salt-bridge may not form, resulting in an amino-terminal arm that might not be as wellfolded. Further evidence implicating His18 is the observation that mutation of this residue to proline results in a protein that is resistant to fucose inhibition and exhibits partial constitutive activation in the absence of arabinose (Wallace, 1982). A proline residue at position 18 may facilitate folding of the arm, thus accounting for the observed behavior of the mutant.

If, indeed, the amino-terminal arm of AraC is disordered at neutral pH in the presence of fucose, it remains to be shown how AraC can oligomerize when fucose is in the sugar-binding pocket. It is not possible to model fucose into the structure of the sugar-free AraC structure dimerized by the β-barrel interface unless the loop containing Tyr31 that inserts into the sugar-binding site undergoes a conformational change so that steric clashes do not occur. Further studies will be required to explain the source of the pH effect on fucose-bound AraC. The structure of the AraC-fucose complex provides several clues that should help guide further investigation of why D-fucose inhibits the ability of the AraC protein to activate transcription in *E. coli*.

Materials and Methods

Crystallization and data collection

The tryptic fragment of AraC containing the sugar-binding and dimerization domain was prepared and concentrated for crystallization as described (Soisson *et al*., 1997), with the following modifications. After purification of the tryptic fragment, the protein was dialyzed extensively into 15 mM Tris-HCl (pH 8.0), 75 mM KCl, 2 mM sodium azide to remove any bound arabinose. D-Fucose was then added to a final concentration of 0.2% (w/v), and the protein was concentrated in a collodion membrane (Sartorius, Goetingen, Germany) to a final concentration of 8 to 12 mg/ml. All crystallization trials were carried out by the method of hanging-drop vapor diffusion. Small crystals were initially observed in 30% (w/v) PEG 4000, 100 mM sodium citrate (pH 5.6), 0.2 % (w/v) D-fucose and 200 mM ammonium acetate. These crystals were crushed and serially diluted for use as microseeds. Large crystals suitable for diffraction analysis grew within 24 hours from a microseeded solution of 24% PEG 4000, 100 mM sodium citrate (pH 5.5), 0.2% (w/v) D-fucose and 200 mM ammonium acetate. Crystals were stabilized in the above reservoir solution to which 10% PEG 400 was added, and flash-frozen in a 100 K nitrogen stream for data collection. Interestingly, we were able to obtain crystals only in the pH range of 5 to 6, even though the arabinose-bound protein crystallizes readily at neutral pH (Soisson *et al*., 1997).

Data were collected at beamline X4A of the National Synchrotron Light Source located at Brookhaven National Labs (Upton, NY). A wavelength of 0.689 Å was chosen to minimize absorption effects from the crystal and so that higher-resolution data could be collected at the same crystal-to-film distance compared with using a longer wavelength. Data were recorded on FUJI image plates and scanned using a FUJI BAS-2000 image plate reader. Frames were then integrated, scaled, and reduced using the DENZO/SCALEPACK suite of programs (Otwinowski, 1993). Statistics of the data collection can be found in Table 1. The crystals form in the monoclinic space group $P2_1$ with two molecules in the asymmetric unit and have unit cell dimensions of $a = 39.25$ Å, $b = 93.68$ Å, $c = 49.92$ Å, $\beta = 95.57°$. These crystals are nearly isomorphous with those of the AraC sugar-binding and dimerization domain grown in the presence of L-arabinose (Soisson *et al*., 1997), which greatly simplified the structure solution process. Mass-spectrometry and amino-terminal sequencing confirm that the tryptic

fragment of AraC used to produce the crystals contains residues 2 to 178 of the protein.

Structure solution and refinement

The structure was initially determined at 1.8 Å resolution by rigid-body refinement with X-PLOR (Brünger, 1992b) using as a starting model the 1.5 Å crystal structure of the AraC sugar-binding and dimerization domain with bound arabinose (Soisson *et al*., 1997). The *R*-factor at this point was 27.2 % and the free *R*-factor (Brünger, 1992a) using 10% of the data not included in the refinement was 30.7 %. The structure was then subjected to one round of simulated-annealing refinement followed by positional refinement, and electron density maps using the coefficients $2F_o - F_c$ and $F_o - F_c$ were calculated at 1.8 Å resolution. At no point during the structure solution or refinement were non-crystallographic symmetry restraints imposed. Positive $F_o - F_c$ density >3σ was clearly visible for the missing methyl group on the bound sugar molecule, which allowed unambiguous positioning of β-D-fucose in the binding site. Simulated-annealing omit maps (Brünger, 1992b) were calculated to check the entire model by sequentially omitting 20 residue segments from both monomers in the asymmetric unit. Manual deletion and addition of side-chains were made at this time. Refinement proceeded with X-PLOR using positional refinement and individual *B*-factor refinement. The resolution was extended to 1.6 Å and water molecules were added in shells at peaks of positive $F_o - F_c$ density of greater than 4σ that were making at least one potential hydrogen bond to the protein or to an already placed water molecule. All water molecules in the final model have *B*-factors of less than 50 Å^2. In the final stages, two bound acetate groups were identified (one per monomer), and added to the model. The final *R*-factor (7.0 to 1.6 Å, $|F| > 2\sigma$) is 17.6 % and the free *R*-factor is 22.2 %. The estimated coordinate error from Luzzati plots (Luzzati, 1952) is 0.15 Å.

Structure analysis

Structural alignments were made using the program O (Jones *et al*., 1991). The root-mean-square difference in C^α positions for the two non-crystallographic symmetry-related monomers in the asymmetric unit is 0.5 Å. For alignments between arabinose-bound and fucose-bound AraC structures, ''equivalent'' monomers from the asymmetric units of both crystals were chosen for analysis. As observed in the structure of the arabinose-bound molecule, the arbitrarily designated monomer B in the fucose structure has the best electron density for the amino-terminal arm. For that reason, monomer B was used for all figures and comparisons between the arabinose and fucose-bound structures. The structure was analyzed using the program PROCHECK (Laskowski *et al*., 1993) and shows better than average scores for structures at similar resolution, with no Ramachandran plot outliers. Coordinates and structure factors for the model have been deposited at the Brookhaven Protein Data Bank, with accession code 2AAC.

Acknowledgements

We thank Elizabeth Reisinger for help in data collection and Craig Ogata of Beamline X4A for advice and technical support. Beamline X4A at the National Synchrotron Light Source, a DOE facility, is supported by the Howard Hughes Medical Institute. We thank Jeremy Berg for helpful comments on the manuscript. This work was supported by the Howard Hughes Medical Institute (C.W.), the David and Lucile Packard Foundation (C.W.), and grant GM18277 from the National Institutes of Health (R.S.).

References

Angyal, S. J. (1972). Conformations of sugars. In *The Carbohydrates: Chemistry and Biochemistry* (Pigman, W. & Horton, D., eds), vol. 1A, pp. 195–215, Academic Press, New York.

Beverin, S., Sheppard, D. E. & Park, S. S. (1971). D-Fucose as a gratuitous inducer of the L-arabinose operon in strains of *Escherchia coli* B/r mutant in gene *araC*. *J. Bacteriol.* **107**, 79–86.

Boos, W., Gordon, A. S., Hall, R. E. & Price, H. D. (1972). Transport properties of the galactose-binding protein of *Escherchia coli*. *J. Biol. Chem.* **247**, 917–924.

Brünger, A. T. (1992a). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472–475.

Brünger, A. T. (1992b). *X-PLOR v3.1. A System for X-ray Crystallography and NMR, Manual.* Yale University Press, New Haven.

Bustos, S. & Schleif, R. (1993). Functional domains of the AraC protein. *Proc. Natl Acad. Sci. USA,* **90**, 5638–5642.

Cygler, M., Rose, D. R. & Bundle, D. R. (1991). Recognition of a cell-surface oligosaccharide of pathogenic *Salmonella* by an antibody Fab fragment. *Science,* **253**, 442–445.

Doyle, M. E., Brown, C., Hogg, R. W. & Helling, R. B. (1972). Induction of the *ara* operon of *Escherchia coli*B/r. *J. Bacteriol.* **110**, 56–65.

Dunn, T., Hahn, S., Ogden, S. & Schleif, R. (1984). An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl Acad. Sci. USA,* **81**, 5017–5020.

Englesberg, E., Irr, J., Power, J. & Lee, N. (1965). Positive control of enzyme synthesis by gene C in the L-arabinose system. *J. Bacteriol.* **90**, 946–957.

Eustance, R. & Schleif, R. (1996). The linker region of AraC protein. *J. Bacteriol.* **178**, 7025–7030.

Eustance, R., Bustos, S. & Schleif, R. (1994). Reaching out: locating and lengthening the interdomain linker in AraC protein. *J. Mol. Biol.* **242**, 330–338.

Evans, S. V. (1993). SETOR. *J. Mol. Graph.* **11**, 134–138.

Friedman, A. M., Fischmann, T. O. & Steitz, T. A. (1995). Crystal structure of *lac* repressor core tetramer and its implications for DNA looping. *Science,* **268**, 1721–1727.

Greenblatt, J. & Schleif, R. (1971). Arabinose C protein: regulation of the arabinose operon *in vitro*. *Nature New Biol.* **233**, 166–170.

Hendrickson, W. & Schleif, R. (1984). Regulation of the *Escherchia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J. Mol. Biol.* **178**, 611–628.

Hendrickson, W. & Schleif, R. (1985). A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. *Proc. Natl Acad. Sci. USA,* **82**, 3129–3133.

Jeffrey, G. A. & Lewis, L. (1978). Cooperative aspects of hydrogen bonding in carbohydrates. *Carbohydr. Res.* **60**, 179–182.

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 Crystallog. sect A,* **A47**, 110–119.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577–2637.

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

Lauble, H., Georgalis, Y. & Heinemann, U. (1989). Studies on the domain structure of the *Salmonella typhimurium* AraC protein. *Eur. J. Biochem.* **185**, 319–325.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennen, R. G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science,* **271**, 1247–1254.

Lobell, R. & Schleif, R. (1990). DNA looping and unlooping by AraC protein. *Science,* **250**, 528–532.

Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802–810.

Maenaka, K., Kawai, G., Watanabe, K., Sunada, F. & Kumagai, I. (1994). Functional and structural role of a tryptophan generally observed in protein-carbohydrate interaction. *J. Biol. Chem.* **269**, 7070–7075.

Martin, K., Huo, L. & Schleif, R. (1986). The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression negative mutations lie in those same sites. *Proc. Natl Acad. Sci. USA,* **83**, 3654–3658.

Miller, D. M., Olson, J. S. & Quiocho, F. A. (1980). The mechanism of sugar binding to the periplasmic receptor for galactose chemotaxis and transport in *Escherchia coli*. *J. Biol. Chem.* **255**, 2465–2471.

Mowbray, S. L. & Cole, L. B. (1992). 1.7 Å X-ray structure of the periplasmic ribose receptor from *Escherichia coli*. *J. Mol. Biol.* **225**, 155–175.

Otwinowski, Z. (1993). Proceedings of the CCP4 Study Weekend: Data Collection and Processing.

Parsons, R. G. & Hogg, R. W. (1974). Crystallization and characterization of the L-arabinose binding protein of *Escherchia coli* B/r. *J. Biol. Chem.* **249**, 3602–3607.

Pigman, W. & Anet, E. F. L. J. (1972). Mutarotations and actions of acids and bases. In *The Carbohydrates: Chemistry and Biochemistry* (Pigman, W. & Horton, D., eds), vol. 1A, pp. 165–194, Academic Press, New York.

Quiocho, F. A. (1986). Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu. Rev. Biochem.* **55**, 287–315.

Quiocho, F. A. (1988). Molecular features and basic understanding of protein-carbohydrate interactions: the arabinose-binding protein-sugar complex. *Curr. Top. Microbiol. Immunol.* **139**, 135–148.

Quiocho, F. A. & Vyas, N. K. (1984). Novel stereospecificity of the L-arabinose-binding protein. *Nature,* **310**, 381–386.

Quiocho, F. A., Gilliland, G. L. & Phillips, G. N. (1977). The 2.8 Å resolution structure of the L-arabinose-binding protein from *Escherchia coli*. *J. Biol. Chem.* **252**, 5142–5149.

Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1989). Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature,* **340**, 404–407.

Schleif, R. (1996). Two positively regulated systems, *ara* and *mal*. In Escherichia coli *and* Salmonella typhimurium (Neidhardt R., F. C., III, Ingraham, J., Lin, E., Low, K., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M. & Umbarger, H., eds), pp. 1300–1309, American Society for Microbiology, Washington, DC.

Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994). Crystal structure of the LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science,* **266**, 763–770.

Soisson, S. M. (1997). Structural studies of the AraC protein from *Escherichia coli*, PhD thesis, Johns Hopkins School of Medicine, Baltimore, MD.

Soisson, S. M., MacDougall-Shackleton, B., Schleif, R. & Wolberger, C. (1997). Structural basis for ligand-regulated oligomerization of AraC. *Science,* **276**, 421–425.

Vyas, N. K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Strucut. Biol.* **1**, 732–740.

Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1988). Sugar and signal-transducer binding sites of the *Escherchia coli* galactose chemoreceptor protein. *Science,* **242**, 1290–1295.

Wallace, R. G. (1982). AraC and its constitutive mutations in *Escherichia coli* B/r: nucleotide sequences and comparison of altered activities in transcription *in vitro*, PhD thesis, University of California.

Wilcox, G. (1974). The Interaction of L-arabinose and D-fucose with AraC protein. *J. Biol. Chem.* **249**, 6892–6894.

Wilcox, G. & Meuris, P. (1976). Stabilization and size of the AraC protein. *Mol. Gen. Genet.* **145**, 97–100.

Wilcox, G., Meuris, P., Bass, R. & Englesberg, E. (1974). Regulation of the L-arabinose operon BAD *in vitro*. *J. Biol. Chem.* **249**, 2946–2952.

Zhang, X., Reeder, T. & Schleif, R. (1996). Transcription activation parameters at ara pBAD. *J. Mol. Biol.* **258**, 14–24.

Edited by D. Rees

(Received 15 April 1997; received in revised form 7 July 1997; accepted 25 July 1997)