Structure and properties of a truly apo form of AraC dimerization domain

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2006

Impact Factor: 2.63 · DOI: 10.1002/prot.21267 · Source: PubMed

4 AUTHORS, INCLUDING:

John Edward Weldon
Towson University
15 PUBLICATIONS **399** CITATIONS
SEE PROFILE

Michael Rodgers
Johns Hopkins University
12 PUBLICATIONS **269** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 66:646–654 (2007)

Structure and Properties of a Truely Apo Form of AraC Dimerization Domain

John E. Weldon, Michael E. Rodgers, Christopher Larkin, and Robert F. Schleif*
Department of Biology, Johns Hopkins University, Baltimore, Maryland

ABSTRACT **The arabinose-binding pockets of wild type AraC dimerization domains crystallized in the absence of arabinose are occupied with the side chains of Y31 from neighboring domains. This interaction leads to aggregation at high solution concentrations and prevents determination of the structure of truely apo AraC. In this work we found that the aggregation does not significantly occur at physiological concentrations of AraC. We also found that the Y31V mutation eliminates the self-association, but does not affect regulation properties of the protein. At the same time, the mutation allows crystallization of the dimerization domain of the protein with only solvent in the arabinose-binding pocket. Using a distance difference method suitable for detecting and displaying even minor structural variation among large groups of similar structures, we find that there is no significant structural change in the core of monomers of the AraC dimerization domain resulting from arabinose, fucose, or tyrosine occupancy of the ligand-binding pocket. A slight change is observed in the relative orientation of monomers in the dimeric form of the domain upon the binding of arabinose but its significance cannot yet be assessed. Proteins 2007;66:646–654.**

Key words: peptidyl arm; allosteric; arabinose; sedimentation; self-association

INTRODUCTION

AraC protein, a regulator of the L-arabinose operon in *Escherichia coli,* represses transcription of the adjacent *araBAD* genes in the absence of arabinose and activates their transcription in the presence of arabinose. The molecular mechanism by which the binding of arabinose so dramatically alters the protein's regulatory properties has been intensively studied for many years and has led to the proposal of a detailed mechanism that is supported by extensive experimental data.[1–6] In this proposal, named the light-switch mechanism, the protein's N-terminal arms play a major role in controlling the protein's response to arabinose (Fig. 1). In the absence of arabinose, genetic, physiological, and physical experiments suggest or are consistent with the arms binding to the DNA-binding domains of AraC.[1–3,7,8] These experiments indicate that the interactions between the arms and the DNA binding domains hold the DNA binding domains in a relative orientation that favors DNA looping and repression of the *araBAD genes*. In the presence of arabinose, the arms bind to the dimerization domains of AraC over the bound arabinose.[9,10] This frees the DNA binding domains and allows them to reorient so that they can easily bind to the two adjacent direct repeat DNA sites, I_1 and I_2, from which the protein activates transcription of the adjacent *ara* p_{BAD} promoter.

The genetic and physical studies referred to above leave little doubt that the N-terminal arms of AraC play a central and critical role in the protein's response to arabinose through their influence on the protein's DNA binding domains. What is not clear is whether any part of the dimerization domain of AraC other than the N-terminal arm is directly involved and significantly changes position or structure in response to the binding arabinose to the dimerization domain. Unfortunately, while the structures of AraC dimerization domain determined from crystals grown in the presence of arabinose or fucose, or in the absence of sugar,[9,11] support the light-switch mechanism, they fail to show whether parts of the dimerization domain other than the N-terminal arm are directly involved in the arabinose response and, at the same time they raise new questions. The structure of the dimerization domain determined from crystals grown in the absence of sugar shows an unexpected intermolecular interaction: the arabinose-binding pockets are occupied by Y31 residues from adjacent protein molecules. This structure, termed a β-kiss in this paper, is illustrated in Figure 2. The presence of tyrosine in the pocket could drive AraC from a true apo-structure to the one observed, one that is virtually identical to the arabinose-bound structure.

The light switch mechanism as described above does not address the possibility that arabinose-induced changes in the structure of the dimerization domain core may be involved or required for the apparent shift of the N-terminal regulatory arms from the DNA-binding domains to the dimerization domains. Understanding the physico-chemical basis of the shift in the N-terminal arms is a next step in understanding the mechanism by which AraC functions. Another fundamental question concerning regulation by

Grant sponsor: NIH; Grant number: GM18277.

*Correspondence to: Robert F. Schleif, Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218. E-mail: schleif@jhu.edu

Received 2 May 2006; Revised 14 September 2006; Accepted 19 September 2006

Published online 15 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21267

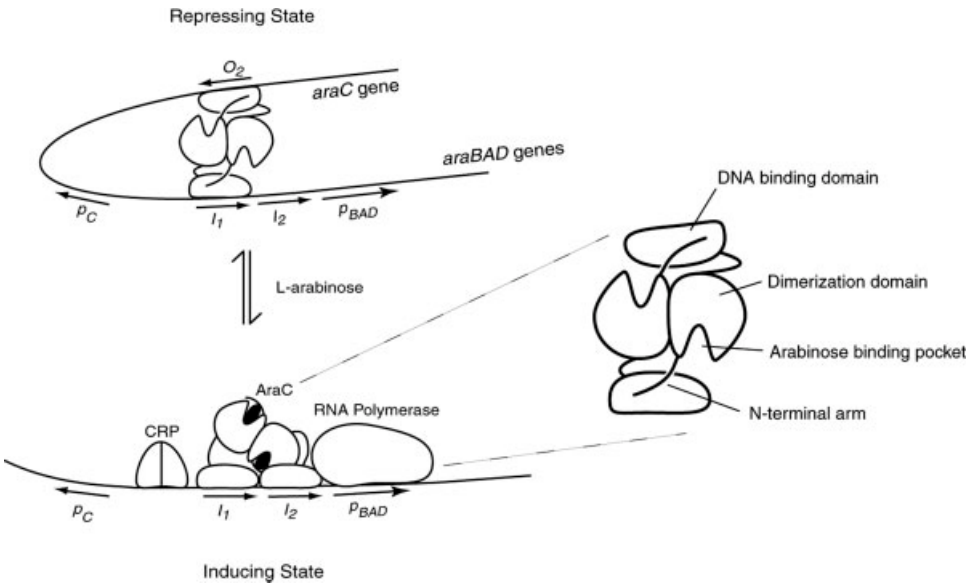

Fig. 1. Components and mechanism of the ara regulatory system. In the absence of arabinose, the N-terminal arms bind the DNA binding domains thereby stimulating DNA loop formation and repression of the p_{BAD} promoter. In the presence of arabinose, the arms bind more tightly over the sugar, freeing the DNA binding domains to bind to the $araI_1$ and $araI_2$ half-sites, from which position AraC stimulates activity of the p_{BAD} promoter. Full induction also requires binding of the catabolite repressor protein CRP.

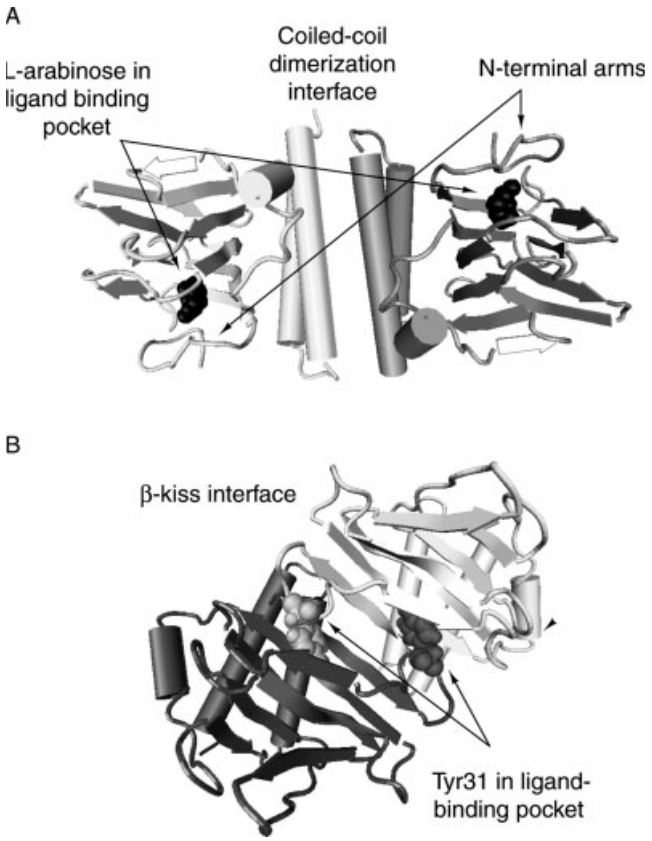

Fig. 2. Structures of AraC dimerization domain. **A**. The structure of a dimer of wild type AraC dimerization domain with arabinose in the ligand-binding pocket and the N-terminal arm over the arabinose (PDB 2ARC). **B**. Two monomers of apo-AraC of wild type protein taken from PDB 2ARA in which Y31 from one monomer is bound in the ligand-binding pocket of the other monomer. These two monomers are from different dimers.

AraC is whether indefinite polymerization mediated by the β-kiss as illustrated (Fig. 2), plays any regulatory role in normal cells.

This paper addresses both questions. We show that the behavior of the AraC dimerization domain in the absence of ligand is best described as an isodesmic, or indefinite self-association. At normal physiological concentrations of AraC, the affinity of this self-association is too weak for significant amounts of the protein to associate via the β-kiss interaction. Using structural knowledge of the β-kiss, we found that it was possible to eliminate the self-association with a targeted point mutation, Y31V. Mutant Y31V AraC shows no self-association but retains in vivo activity comparable to wild-type protein, thus showing that the β-kiss association plays no significant role in AraC regulation.

The arabinose-binding pocket of mutant Y31V dimerization domain crystallized in the absence of arabinose was found to be occupied only by solvent. This provides the desired apo-structure of AraC dimerization domain. Comparison to the other AraC dimerization domain structures shows no significant structural difference in the core of the dimerization domain in the presence or absence of arabinose.

MATERIALS AND METHODS

General Procedures, Plasmids, and Strains

Plasmid pWR03,[7] derived from pSE380, contains the full-length AraC sequence, nucleotides 1–876, between the *NcoI* and *SacI* sites. In strain SH321, $F^-\Delta araC\text{-}1022$ $\Delta lac74$ $galK^-$ str^{r12} pWR03 was used to screen AraC mutants for induction at p_{BAD} on arabinose/tetrazolium indicator plates,[7] as well as for quantitation of induction at p_{BAD} using the arabinose isomerase assay.[13] In strain

SH10, F^+ *araC:lacZ* leu^- str^r Δ*lac araD*$^-$, pWR03 was used to measure the level of repression at p_C in AraC mutants.[7] For these measurements cells were grown overnight in YT[13] medium, diluted 100-fold into minimal M9[13] medium containing vitamin B1 and ampicillin at 100 μg/mL with or without arabinose at 0.2% and grown to a density of 2×10^8/mL for enzyme measurement.[7,12,13]

Plasmid AraCTF[14] was derived from pET21b by cloning nucleotides 1–546 of AraC between the *NdeI* and *AdeI* sites. The resulting vector expresses residues 1–182 of AraC, which corresponds to the dimerization domain and linker regions of the protein followed by a C-terminal Leu-Glu linker and a His-6 tag. Wild type and mutant Y31V dimerization domain protein were expressed from the AraCTF plasmid in BL21 (DE3) and were purified essentially as described.[14] Cells were grown at 37°C in YT[13] medium containing 200 μg/mL ampicillin to an apparent $OD_{550\ nm}$ of ~0.8 to 1.0, induced with the addition of IPTG to 1 m*M* for a minimum of three hours, then harvested by centrifugation. Cells were lysed by grinding in alumina[13] or by a French press in 15 m*M* Tris-Cl, pH 8.0, 0.1 *M* NaCl, 5% glycerol, 10 m*M* $MgCl_2$, 50 m*M* L-arabinose with 10 μg/mL each DNase I and RNase A added immediately before use. Supernatant from the lysed cells was mixed with 1 mL Ni-NTA agarose beads from Qiagen (Valencia, CA) per estimated 10 mg target protein and rocked gently at 4°C for a minimum of 2 h. Typically, 5 mL of beads was incubated with extract from 1 L of cell growth medium. Unbound material was removed and the beads were washed with 15 m*M* Tris-Cl, pH 8.0, 0.1 *M* NaCl, 50 m*M* L-arabinose, and 10 m*M* imidazole until the $OD_{280\ nm}$ of the wash was <0.05. Bound protein was eluted with two column volumes of buffer containing 15 m*M* Tris-Cl, pH 8.0, 10 m*M* NaCl, 50 m*M* L-arabinose, 1 *M* imidazole. Protein was digested with 1 μg trypsin per 1 mg estimated His-6 protein overnight at 4°C to release the His-6 tag. This was followed by anion exchange chromatography on a Pharmacia Mono-Q HR 5/5, 1 mL volume in 15 m*M* Tris-Cl, pH 8.0, 50 m*M* L-arabinose using an elution gradient of 10 m*M* to 1 *M* NaCl over 40 mL.

Mutations were introduced into the pWR03 and AraCTF plasmids using the QuikChange Site-directed Mutagenesis kit from Stratagene (La Jolla, CA). Plasmid DNA was isolated from cells using the Wizard® Plus Miniprep DNA purification system from Promega (Madison, WI). Mutations were confirmed by DNA sequencing.

Analytical Ultracentrifugation

Purified wild-type and Y31V AraC dimerization domain were extensively dialyzed into 15 m*M* Tris-Cl, pH 8.0, 75 m*M* KCl with and without 0.2% (w/v) L-arabinose. For velocity sedimentation experiments, samples at 1.5 mg/mL and solvent were loaded into 3 mm path length cells with double-sector charcoal-filled Epon centerpieces and quartz windows. Cells were centrifuged in a Beckman XLI analytical ultracentrifuge at 50,000 rpm in an An-50Ti rotor at 20°C. Data were collected using optical adsorption detection at 280 nm with radial scans taken at approximately 2.5-min intervals. A nominal radial step size of 0.003 cm was used. Data were analyzed by the van Holde–Weischet method as implemented in the Ultrascan software.[15] The partial specific volumes were calculated from amino acid composition by the program Sednterp (v1.07).[16] The buffer density was taken as 1.00221 g/cm^3 without L-arabinose and 1.00309 g/cm^3 with 0.2% (w/v) L-arabinose present. These were estimated by Sednterp from the buffer composition using glucose in the calculation in place of arabinose because the Sednterp database lacked an entry for arabinose.

For equilibrium sedimentation experiments in the absence of L-arabinose, loading concentrations were 0.5 mg/mL in the 3 mm cell and 0.25, 0.125, and 0.0625 mg/mL in the 12 mm cells. Cells were spun at 18,000, 21,000, 24,000, and 27,000 rpm in an An-50Ti rotor at 20°C. In the presence of L-arabinose, loading concentrations were 0.5, 0.17, and 0.0625 mg/mL, and the samples were spun at 21,000, 25,000, and 28,000 rpm Equilibrium was ascertained by comparing scans taken at 3-h intervals after a 12-h initial equilibration period. Data were edited using the program WinReedit (v0.999.027; Analytical Ultracentrifugation Facility, University of Connecticut) and analyzed by nonlinear least-squares global fitting of absorbance using the NONLIN algorithm.[17] Data were fit to a variety of different models, including monomer, dimer, trimer, monomer-dimer, monomer-trimer, monomer-dimer-trimer, monomer-dimer-tetramer, and an isodesmic model. Models were rejected if the residuals of the fit were not randomly distributed or had a higher variance than another model.

Fluorescence Measurements of Arabinose Binding

Purified wild-type and Y31V AraC dimerization domain were dialyzed extensively into 15 m*M* Tris-Cl, pH 8.0, 75 m*M* KCl and then further diluted in the same buffer to a concentration of 10 μg/mL, ~475 n*M*. Fluorescence was measured in 10 mm quartz cuvettes continuously stirred with a magnetic stir bar. Measurements were made on a T-format Photon Technology International spectrofluorimeter with a 75-W xenon arc lamp. Samples were excited at 295 nm with slit widths adjusted for 1 nm spectral bandwidth and emission was observed between 320 and 370 nm with slit widths adjusted for 5 nm spectral bandwidth. L-arabinose stock solutions, dissolved in the final dialysis buffer at concentrations of 1 *M*, 0.1 *M*, and 10 m*M*, were added to the samples to achieve concentrations from 0 to 50 m*M*. Equilibrium was reached by 1 min, as interpreted by no significant change in fluorescence signal between the 1- and 2-min intervals after the addition of arabinose. Data were collected at 1 point/second at 1 nm intervals. An average emission wavelength $\langle\lambda\rangle$ at each titration point was calculated as:

$$\langle\lambda\rangle = \frac{\sum_i^n (I_i \times \lambda_i)}{\sum_i^n (I_i)}$$

where λ is the wavelength and I_i is the fluorescence intensity at the wavelength λ_i minus the background at

the same wavelength. This was plotted against arabinose concentration, and fit to a standard ligand-binding equation. Experiments were performed at least three times.

Crystallization, Data Collection, and Structure Determination of Y31V AraC Dimerization Domain

Purified Y31V AraC dimerization domain protein was dialyzed extensively against 10 m*M* Tris-Cl, pH 8.0, 75 m*M* KCl. Diffraction quality crystals were obtained with the hanging drop vapor diffusion method using a reservoir solution containing 34% PEG 3350, 0.1 *M* Hepes, pH 7.5, 0.2 *M* ammonium acetate, 2% ethylene glycol. The hanging drop was composed of 2 μL of 6 mg/mL protein mixed with 2 μL of well solution. Crystals were grown at 20°C for a minimum of 1 week before transfer to a cryoprotectant solution consisting of 38% PEG 3350, 0.1 *M* Hepes, pH 7.5, 0.2 *M* ammonium acetate, and 7% ethylene glycol. Crystals were then immediately seated in a cryo-loop and submersed in liquid nitrogen. Data were collected at 100 K in 0.5° increments for 360 frames, then indexed, processed, and scaled using the *HKL* software package.[18] All data collection was performed using a Rigaku RU-H3RHB generator equipped with an MSC Max-Flux confocal optical system, an R-AXIS IV++ imaging-plate system and an X-Stream 2000 cryogenic system.

The structure of Y31V AraC dimerization domain was determined from the data set by molecular replacement using the Crystallography and NMR System, CNS, program v1.1.[19] Search models were generated from the previously determined structure of wild-type AraC dimerization domain in the presence of L-arabinose, PDB ID 2ARC. A single protomer was used as a search model. Once the five AraC monomers were placed in the asymmetric unit, rigid body refinement was performed using CNS, resulting in a R-factor of 0.4063. Subsequent simulated annealing refinement lowered the R-factor further to 0.3978. To minimize the model bias that is inherent in the molecular replacement procedure used in the first step of solving the overall structure, in subsequent stages of refinement, the structure of each monomer in the asymmetric unit has been fit into the electron-density simulated annealing omit maps. Model adjustment and building was performed with the TURBO-FRODO program[20] with alternating rounds of energy minimization refinement. B-factor refinement and the generation of Fo-Fc and 2Fo-Fc electron density maps performed using CNS. NCS restraints were dropped in later rounds of refinement. Where the protein chain was extended from the search model, omit maps were calculated to insure proper positioning of newly added residues. A final composite omit map was calculated and inspected before building solvent molecules. Prior to deposition in the RCSB Protein Data Bank, the model was assessed for stereochemical quality with Procheck.[21] The atomic coordinates and structure factors have been deposited in the Protein Data Bank[22] with the identification 1XJA.

Structure Comparisons and Energy Calculations

Distance difference lists were generated using CNS using C_α atoms of residues 20 to 167 from each monomer in the asymmetric unit of PDB ID 2ARA, 2ARC, 2AAC, and 1XJA. Distance difference matrices were plotted from the distance difference lists with the program MatrixPlot[23] where the square in row *i*, column *j* was colored according to the value of the distance difference between atoms *i* and *j* in the two structures being compared. Initial matrices showed all parts of the fucose-bound structures to differ from all the others. This was found to result from the fact that bond lengths and relative atom positions in this structure appear to be 1% too large. Consequently, all structures, 2ARA, 2ARC, 2AAC, and 1XJA were corrected for minor variations in size by using the sums of the distances between each pair of C_α atoms to normalize the distance lists. Only the fucose-bound structure required significant, ~1% correction.

Calculations of the interaction energies in the β-kiss and least squares overlaying of protein structures were performed using CHARMM 27,[24] using the CHARMM 27 potentials and explicit TIP3 water.

RESULTS

Oligomerization of AraC Dimerization Domains

The arrangement and interactions between dimers of wild type AraC protein crystallized in the absence of arabinose suggest that the protein should polymerize in the absence of arabinose.[9] Although this was observed at milligram/mL concentrations,[9] the strength of the association was not obtained. To assess the biological relevance of the polymerization, we determined the strength of the self-association. Figure 3 shows data from equilibrium sedimentation of the AraC dimerization domain in the absence of arabinose and its fit to an indefinite, or isodesmic, self-association model. Other models of similar complexity fit less well to the data. The molecular mass of the smallest interacting species was determined from the fit as 40,900 g/mol. This is close to the molecular mass of the dimerization domain dimer. It indicates that the dimer is the primary building block of the observed oligomers. The best-fit value for a dissociation constant of the interaction is approximately 5.0×10^{-5} *M*, far above the intracellular concentration of AraC, which is approximately 10^{-8} *M*.[25] These values indicate that in the absence of arabinose, virtually none of the AraC protein in cells is engaged in the β-kiss association that gives rise to the polymerization.

Elimination of Y31-Mediated Oligomerization

To help understand the mechanism of the arabinose-induced response of AraC, we tried to obtain crystals of the dimerization domain with the ligand-binding pocket devoid of arabinose or a side chain mimicking arabinose. Unfortunately, the arabinose-binding pocket in the wild type protein crystallized in the absence of arabinose

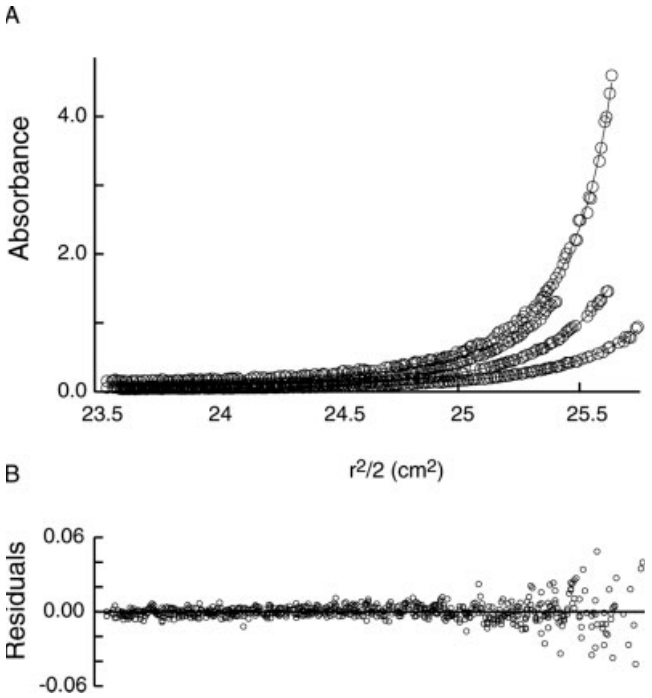

Fig. 3. Equilibrium sedimentation of AraC dimerization domain in the absence of arabinose. Data were collected using both 12 and 3 mm path length cells. All absorbance values are scaled to a 12 mm path length. **A**. Raw data from equilibrium sedimentation of the AraC dimerization domain in the absence of L-arabinose fit to an isodesmic (indefinite self-association) model. Data and curve fit from all four concentrations (0.06, 0.12, 0.25, and 0.50 mg/mL) at a single speed of 21,000 rpm are presented. **B**. Residuals for the fitted data.

TABLE I. Induction and Repression Abilities of Wild Type and Mutant AraC

	Relative level	
Protein	Repression[a]	Induction[b]
No AraC	1 ± 0.05	1 ± 0.4
Wild type	0.18 ± 0.01	17 ± 5.5
Y31Stop	1 ± 0.02	0.7 ± 0.24
Y31A	0.16 ± 0.01	13 ± 5.5
Y31V	0.18 ± 0.01	13 ± 2.5

[a]Obtained by measurement of β-galactosidase in strain SH10; F^+ p_C:*lacZ leu*$^-$ *str*^{r} Δ*lacaraD*$^-$ containing plasmid pWRO3 carrying the indicated mutations. Errors are the standard deviations of three measurements.
[b]Obtained by measurement of arabinose isomerase in strain SH321, F^-Δ*araC-1022* Δ*lac74 galK*$^-$ *str*^{r} containing plasmid pWRO3 carrying the indicated mutations.

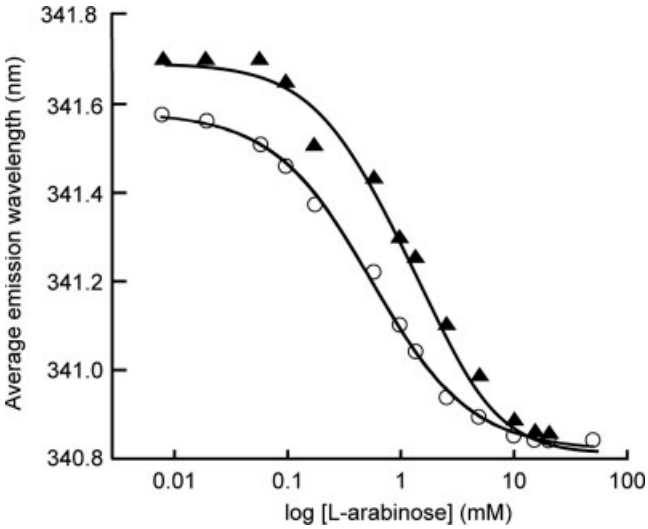

Fig. 4. Binding of L-arabinose to AraC dimerization domain. Intrinsic tryptophan fluorescence intensity was measured. The average emission wavelength at each titration point was calculated for emission spectral data spanning 320–370 nm. The excitation wavelength was 295 nm. Data was fit to a standard single-site ligand-binding model. Open circles are wild type protein, and filled triangles are Y31V protein.

turns out to be occupied by Y31 from neighboring molecules.[9] Therefore, in an attempt to prevent Y31 from occupying the pocket, we sought mutants with substitutions at this position that also retain normal repression and induction properties. Such mutants seemed likely to exist because Y31 does not appear to make important contacts in either the arabinose- or fucose-bound structures of the dimerization domain. Such mutants would also be unlikely to participate in the β-kiss polymerization because the tyrosine side chain interactions that contribute the major interaction energy in driving the β-kiss self-association of wild-type protein would no longer be possible.

Plasmid pWR03 in which the codon for residue 31 had been randomized by site-directed mutagenesis was transformed into *araC*$^-$ strain SH321. The majority of the transformants were capable of induction by arabinose, as indicated by colony color on tetrazolium indicating plates. This implies that Y31 does not play a critical role in induction by AraC. The araC gene from ten arabinose-inducible candidates was sequenced. Those with point mutations only in this tyrosine codon were found to have changed it to a codon for Ala, Gly, or Val.

The induction capability of the mutants was compared with wild type by their abilities to induce expression of arabinose isomerase from *ara* p_{BAD}. Repression capability was measured by the proteins' abilities to repress β-galactosidase synthesis in a strain of *E. coli* containing a *pC-lacZ* fusion (Table I). The Y31V mutant protein, which appears nearly normal, was chosen for further study and the Y31V mutation was introduced into plasmid AraCTF for overexpression and purification of mutant dimerization domain.

Any mutation at position 31 might appreciably alter the arabinose binding properties of the protein while still allowing the protein to induce nearly normally. Therefore, we also directly measured the arabinose affinity of wild type dimerization domain and Y31V mutant dimerization domain. This was performed using a fluorescence assay that is based on the observation that the binding of arabinose quenches the protein's intrinsic tryptophan fluorescence and also shifts the emission spectrum to a shorter wavelength.[5] Figure 4 shows arabinose binding curves for wild-type and Y31V dimerization domain based on the

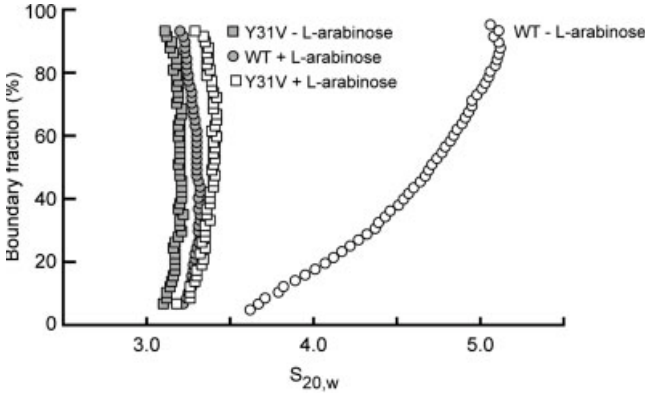

Fig. 5. The Y31V mutation eliminates self-association of the AraC dimerization domain. A Van Holde–Weischet analysis of data from velocity sedimentation of wild-type and Y31V mutant AraC dimerization domain in the presence and absence of L-arabinose.

change in the average emission wavelength. In these experiments Y31V dimerization domain was found to have a dissociation constant of 1.35 ± 0.11 m*M* for arabinose, while wild-type protein has a K_D of 0.58 ± 0.05 m*M*.

Data from velocity sedimentation of wild-type and Y31V AraC dimerization domain in the presence and absence of arabinose are presented in Figure 5. The results show that the Y31V mutation significantly reduces the propensity of the dimerization domain to form higher-order species. The mutant protein behaves like a single species in both the presence and absence of arabinose and has sedimentation coefficients similar to those of the wild-type dimerization domain in the presence of arabinose. Thus, the mutation reduces oligomerization to below a detectable level under the experimental conditions, $K_D > 1$ m*M*.

Structure Determination of the Y31V Dimerization Domain

We expressed, purified, and crystallized the Y31V mutant dimerization domain fragment of AraC. Crystals diffracted well at room temperature, but poorly when frozen. Addition of ethylene glycol improved diffraction of the frozen crystals to 2.4 Å. The structure was solved by molecular replacement using the arabinose-bound structure 2ARC followed by multiple rounds of refinement and model building, resulting in a final R-factor of 21.1 and R-free of 25.6. Table II presents the complete data collection and refinement statistics. The asymmetric unit of the dimerization domain contains five monomers. Clearly defined electron density existed to allow confident modeling of residues 8–172, 15–168, 7–175, 7–171, and 7–169, of the chains that contain residues 2–178. Additionally, 25 molecules of ethylene glycol and 259 molecules of water were positioned at the end of the building and refinement process.

Figure 6 shows the five-monomer asymmetric unit of the apo Y31V AraC dimerization domain crystal structure. Few differences aside from positioning of the N-terminal arm are apparent between molecules in this structure determination and those in the previously determined structures. Overlaying backbone representations of the

TABLE II. Crystallographic Statistics

X-Ray data	
Source	Rigaku RU-H3R Cu K
Wavelength (Å)	1.5418
Temperature (K)	100
Space group	C2
Unit cell dimensions (Å)	a = 260, b = 40.62, c = 102.59
Unit cell angles (°)	α = γ = 90, β = 105.45
Resolution range (Å)	15–2.4
R_{sym} (last shell)	10.5 (42.3)
Completeness (last shell)	94.7 (90.6)
Observed reflections	39009
Molecular replacement	
Resolution (Å)	15–4.0
Correlation coefficient	36.7
RMR(%)	49.2
Refinement	
Number of protein atoms	6615
Number of waters	259
Number of EDO molecules	25
Rwork (Rfree)	21.1 (25.6)
Mean B factor (Å^2) solvent	41.56
Mean B factor (Å^2) protein	36.72
Ramachandran statistics (%)	
Most favored regions	85.6
Additional allowed regions	13.6
Generously allowed regions	0.7
Disallowed regions	0.0

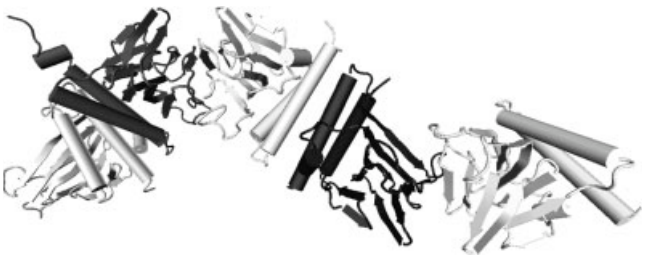

Fig. 6. Illustration of the asymmetric unit of the Y31V dimerization domain crystal structure.

structures showed them to be highly similar. Examination of the conformations of the side chains of the 15 residues possessing any atom within 4 Å of any atom of bound arabinose, that is, the residues forming the arabinose-binding pocket, showed that the χ1 dihedral angles of all resides except those of the arm were only slightly changed by the binding of arabinose. The largest change was 20° and the average of the changes was less than 10°. The β-kiss interaction seen with the wild type dimerization domain is absent from the apo Y31V crystal structure and the arabinose-binding pocket itself is empty except for solvent.

Structural Comparison Between Monomeric AraC Dimerization Domains

Neither visual comparison nor least squares superposition of structures are well suited to the detection and presentation of the subtle structural differences separating the arabinose-bound, fucose-bound, wild-type apo- (Y31-

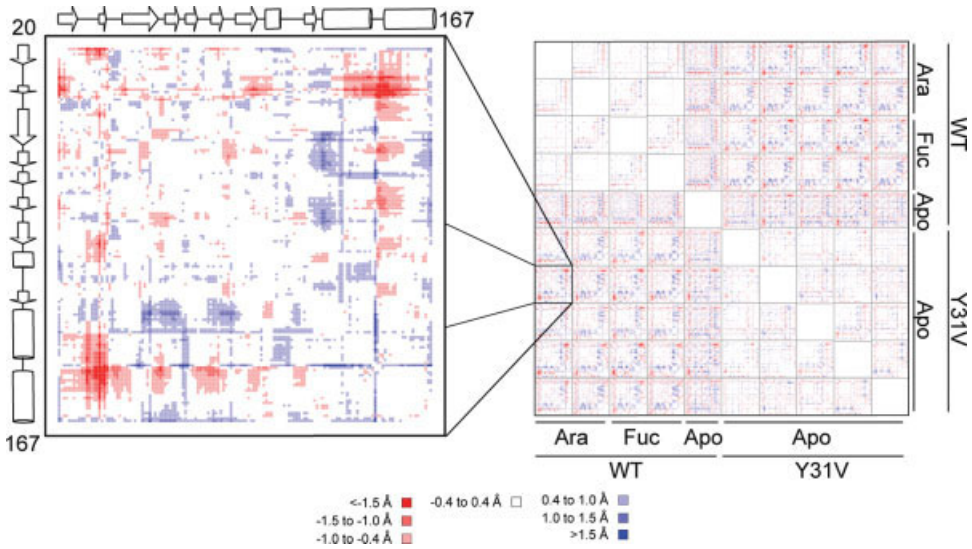

Fig. 7. Structural comparison of AraC dimerization domain monomers. A distance difference map for the C_α atoms of residues 20–167 for each pair of dimerization domain monomers is shown. Difference values were grouped and colored as indicated. An individual matrix comparing the wild-type arabinose-bound chain A monomer to the Y31V apo-AraC chain B is expanded on the left. The axes are labeled with the secondary structure elements of the AraC monomer.

bound), and Y31V apo-AraC structures. Therefore, we used distance difference matrices.[26–28] A distance difference matrix for two structures graphically displays in two dimensions the structural differences between them. The distance difference matrix element in the ith row and jth column, ΔD_{ij}^{AB}, is the difference in the distance between the Cα atoms of residues i and j in protein A, D_{ij}^{A}, and the distance between the same atoms in structure B, D_{ij}^{B}. Thus $\Delta D_{ij}^{AB} = D_{ij}^{A} - D_{ij}^{B}$. This method of comparison does not rely upon overlaying the two proteins. If the positioning of any part of one of the two proteins is shifted with respect to the rest of the protein, the distances of these atoms from the rest of the protein will be altered. Hence, the corresponding distance difference matrix elements comparing to an unshifted structure will be nonzero whereas the matrix elements representing the difference in distance between pairs of atoms in portions of the proteins with identical structure will be zero. The resulting difference matrices are then rather easily comprehended if the elements of the array are colored according to the values of the corresponding matrix elements. For example, increasing redness can indicate increasing positive values and increasing blueness can indicate decreasing negative values. An additional virtue of the distance difference representation of structural differences is that it is two dimensional, and thus all the information can be viewed simultaneously on a printed page.

The arabinose-bound, fucose-bound, wild type apo- (tyrosine-bound) and Y31V apo- crystal structures of the AraC dimerization domain contain multiple monomers in the asymmetric unit. Since the structure of each monomer in the asymmetric unit has been constructed as independently as possible (See Methods), we used the entire complement of monomers in our structural comparison. In total, the structures of ten different monomers of AraC have been determined by x-ray diffraction: two each from the asymmetric units of wild-type protein bound to arabinose, 2ARC, and fucose, 2AAC; one from the wild-type apo- (tyrosine bound) form, 2ARA; and five monomers in the asymmetric unit of the Y31V apo-structure, 1XJA. The distance difference matrix method allows convenient comparison of the structure of each of the ten monomers with each of the nine other monomers. Thus, even if only one monomer possesses a significantly altered structure, it will be detected.

A distance difference matrix was generated for each of the forty-five unique combinations of these ten structures. They are presented in a 100-block two-dimensional matrix along with an expanded example of an individual plot, (Fig. 7). Monomers from the same asymmetric unit that have been crystallized from the same conditions are likely to possess a minimum of dissimilarity and hence a minimum of color in a distance difference matrix. Indeed, Figure 7 shows this to be so. Also as expected, regions corresponding to flexible loops in the proteins tend to be colored. The main conclusion from the matrix plot analysis is that although systematic structural changes generated by the different ligands can be detected, they are very small, on the order of several tenths of an Angstrom, and they are largely spread over the entire structure.

Comparison of Dimer Structures

The analysis presented in the previous section compared the structures of individual monomers by calculating matrix elements ΔD_{ij}^{AB} when atoms i and j are both in the same monomer. The same distance difference matrix method could also be applied to analyze the differences between dimer structures, but since the individual

TABLE III. Comparison of Monomer Orientations in Pairs of Dimers

Dimer pair compared	Degrees difference in angle between monomers of the two dimers
Arabinose-bound to Fucose-bound	2.1
Apo AB to Apo CD	2.8
Arabinose-bound to Apo AB	5.7
Arabinose-bound to Apo CD	4.2
Fucose-bound to Apo AB	7.3
Fucose-bound to Apo CD	5.4

monomers barely change in structure, it is more revealing merely to determine how the relationship between the two monomers of a dimer differs between the various dimers. We considered the dimers with bound arabinose, bound fucose, and two dimers from the asymmetric unit of Y31V with nothing bound. The comparisons were accomplished by least squares superpositioning the Cα atoms of one of the monomers from each dimer and then determining the rotation necessary to overlay the other monomers of the two dimers. The variation in the angles between the subunits is about 2° when comparing arabinose-bound or fucose-bound dimers or when comparing the two apo-dimers (Table III). The variation is about 5° when comparing sugar-bound dimers with apo-dimers. Without knowing the rigidity of the protein it is not possible to make a firm conclusion about the significance of the difference between 5 and 2°.

DISCUSSION

The N-terminal arms of the AraC protein play a key role in communicating the status of arabinose binding by the dimerization domains to the DNA-binding domains (see[29,30] for reviews). The arms interact with the DNA-binding domains in the absence of arabinose, and in the presence of arabinose they lose this interaction and bind to the dimerization domains over the bound arabinose. Are direct arm-arabinose interactions the sole effector of the arms' positioning and hence, the regulatory properties of AraC, or do changes in the dimerization/arabinose-binding domain also contribute to control of the arm or DNA binding domain positions? Structures of the dimerization domain have previously been determined from crystals grown in the presence of arabinose or fucose, and also in the absence of sugar. In all three cases, however, the arabinose-binding pocket of the protein is occupied. In the first two cases the pocket is occupied by arabinose or fucose, and in the third case it is occupied by Tyr31 from a neighboring dimer.[9,11] To determine if changes in the conformation of the dimerization domain result from occupancy of the arabinose-binding pocket, and whether the arm is the sole communicator of information to the DNA-binding domain, we sought to determine the structure of the dimerization domain when the arabinose-binding pocket is occupied only by solvent.

For several reasons, simple mutagenic replacement of the tyrosine that fortuitously binds in the arabinose-binding pocket at high protein concentrations seemed likely to yield normally functional AraC. First, the tyrosine in question does not appear to play an important role in the normal functioning of the protein. Its side chain extends into the solvent in the structure of the dimerization domain with arabinose bound. Second, this tyrosine provides a significant fraction of the interaction energy between dimerization domains in the crystals grown in the absence of arabinose or fucose and presumably in the interactions that lead to polymerization of dimers of the dimerization domains in solution in the absence of arabinose or fucose. We found this supposition to be true. Substitution of the tyrosine with valine yielded a protein that induced and repressed the *ara* p_{BAD} promoter normally. Therefore it seemed that this protein would be unlikely to polymerize via the β-kiss interaction that is seen with wild type protein. Ultracentrifuge studies demonstrated that this is the case. We found that while dimers of the wild type dimerization domain interact with a dissociation constant of 5×10^{-5} M, self-association of mutant Y31V dimerization dimers cannot be detected at all. We estimated therefore that the K_D for their self-association is greater than 10^{-3} M. At the concentrations of AraC found in cells, 10^{-8} M, no signification amount of either wild type or mutant protein can engage in the β-kiss interaction. The weakness of the β-kiss interaction in the mutant protein indicates that the interaction is unlikely to occur even at the high protein concentrations needed for crystallization. This finding provided sufficient reason to hope that if the mutant protein could be crystallized, it would do so in a crystal lattice not utilizing the β-kiss interaction and only solvent would occupy the arabinose-binding pocket.

Crystals of the mutant protein were eventually grown and the structure of the apo-dimerization domain was determined. Unlike the wild-type apo structure of the dimerization domain, residues of the arabinose-binding pocket in the Y31V protein are in contact with only solvent. Thus, it seems likely that the structure of the Y31V apo dimerization domain reflects the true apo structure. Although crystal lattice interactions with surface residues of the dimerization domain could distort this structure, generally such changes are small and local[31] and we think it unlikely that such effects have shifted a "true" apo structure to that observed here. An additional virtue of the apo-crystals was that the asymmetric unit in the crystals contained five monomers of the protein. Thus, solving the structure provided five independently determined structures of the apo-dimerization domain.

The crystals of apo-dimerization domain, like those of hemoglobin,[32] are sensitive to the presence of allosteric effector. Arabinose or fucose, but not glucose, destroys the apo-crystals. At first this seems surprising because the binding of arabinose in the dimerization domain and the binding of the N-terminal arm over the arabinose is not sterically blocked in the apo crystal. Further analysis indicates however, that changes in the crystal lattice contacts

made by the arm are likely to cause the crystals to dissolve. In the crystal, the arm from one dimerization domain interacts the core of an adjacent monomer in the next asymmetric unit and stabilizes the crystal lattice by providing a cross brace. If arabinose is added, the crystal disappears, presumably because this cross brace is eliminated when the arm shifts position to sit over arabinose in the ligand-binding pocket.

Visual inspection of the new apo-structures, and the arabinose-, fucose-, and tyrosine-bound structures or their overlays revealed no significant structural differences. We therefore sought a structure comparison method capable of impartially comparing each possible pair of structures amongst the two arabinose-bound, two fucose-bound, one tyrosine-bound, and five apo-structures, each of which is an independently (see Methods) determined structure. Distance difference matrices proved to meet the requirements.[26–28] One ten-by-ten array of distance matrices displays the entire set of conformational differences between every pair of the ten subunits. The array showed that while the apo-structures differed from those with the arabinose-binding pocket occupied, the differences were so small that they seem unlikely to be important.

Not only do the structures of the individual dimerization domains shift negligibly upon binding of arabinose or fucose, we found that the relative orientation, that is, the angle between the two monomers of a dimer, also shifts only slightly upon ligand binding (Table III). As this amount, perhaps 3°, is comparable to the 2° difference seen when comparing arabinose-bound and fucose-bound or apo-dimers, we consider it unlikely to play a significant role in the arabinose response of the protein.

In summary, the self-association of wild type AraC protein via inter-molecular interactions largely mediated by Y31 are not physiologically relevant, and upon their elimination, the dimerization domain can be crystallized with the arabinose-binding pocket containing only solvent. The structure of the empty pocket protein suggests that arabinose-arm interactions are the primary determinant of the protein's response to arabinose.

ACKNOWLEDGMENT

The authors thank Victoria Hargreaves for comments on the manuscript.

REFERENCES

1. Saviola B, Seabold R, Schleif RF. Arm-domain interactions in AraC. J Mol Biol 1998;278:539–548.
2. Seabold RR, Schleif RF. Apo-AraC actively seeks to loop. J Mol Biol 1998;278:529–538.
3. Wu M, Schleif R. Strengthened arm-dimerization domain interactions in AraC. J Biol Chem 2001;276:2562–2564.
4. Wu M, Schleif R. Mapping arm-DNA-binding domain interactions in AraC. J Mol Biol 2001;307:1001–1009.
5. Ross JJ, Gryczynski U, Schleif R. Mutational analysis of residue roles in AraC function. J Mol Biol 2003;328:85–93.
6. Gryczynski U, Schleif R. A portable allosteric mechanism. Proteins 2004;57:9–11.
7. Reed WL, Schleif RF. Hemiplegic mutations in AraC protein. J Mol Biol 1999;294:417–425.
8. Harmer T, Wu M, Schleif R. The role of rigidity in DNA looping and unlooping by AraC. Proc Natl Acad Sci USA 2001;98:427–431.
9. Soisson SM, MacDougall-Shackleton B, Schleif R, Wolberger C. Structural basis for ligand-regulated oligomerization of AraC. Science 1997;276:421–425.
10. Ghosh M, Schleif RF. Biophysical evidence of arm-domain interactions in AraC. Anal Biochem 2001;295:107–112.
11. Soisson SM, MacDougall-Shackleton B, Schleif R, Wolberger C. The 1.6 Å crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. J Mol Biol 1997;273:226–237.
12. Hahn S, Dunn T, Schleif R. Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. J Mol Biol 1984; 180:61–72.
13. Schleif RF, Wensink PC. Practical methods in molecular biology. New York: Springer-Verlag; 1981.
14. LaRonde-LeBlanc N, Wolberger C. Characterization of the oligomeric states of wild type and mutant AraC. Biochemistry 2000; 39:11593–11601.
15. Demeler B, Saber H, Hansen JC. Identification and interpretation of complexity in sedimentation velocity boundaries. Biophys J 1997;72:397–407.
16. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding EE, editor. Analytical ultracentrifugation in biochemistry and polymer science. Cambridge, UK: The Royal Society of Chemistry; 1992. pp 90–125.
17. Johnson ML, Correia JJ, Yphantis DA, Halvorson HR. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. Biophys J 1981;36:575–588.
18. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
19. Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kuntstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
20. Roussel A, Cambillau C. The TURBO-FRODO graphics package in silicon graphics geometry partners directory. Mountain View, CA: Silicon Graphics; 1991. p 81.
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
23. Gorodkin J, Staerfeldt HH, Lund O, Brunak S. MatrixPlot: visualizing sequence constraints. Bioinformatics 1999;15:769–770.
24. Brooks BR, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983; 4:178–217.
25. Kolodrubetz D, Schleif R. Identification of AraC protein and two-dimensional gels, its *in vivo* instability and normal level. J Mol Biol 1981;149:133–139.
26. Nishikawa K, Ooi T, Saito N, Isogai Y. Tertiary Structure of Proteins. I. Representation and Computation of Conformations. J Phys Soc (Japan) 1972;32:1331–1334.
27. Richards FM, Kundrot CE. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. Proteins 1998;3:71–84.
28. Srinivasan R, Rose GD. The T-to-R transformation in hemoglobin: a reevaluation. Proc Natl Acad Sci USA 1994;91:11113–11117.
29. Schleif R. Regulation of the L-arabinose operon of *Escherichia coli*. Trends Genet 2000;16:559–565.
30. Schleif R. AraC protein: a love-hate relationship. BioEssays 2003; 25:274–282.
31. Eyal E, Gerzon S, Potapov V, Edelman M, Sobolev V. The limit of accuracy of protein modeling: influence of crystal packing on protein structure. J Mol Biol 2005;351:431–442.
32. Perutz MF. Polarization dichroism, form birefringence, and molecular orientation in crystalline haemoglobins. Acta Crystallogr 1953;6:859–894.