Proc. Natl. Acad. Sci. USA
Vol. 73, No. 7, pp. 2186–2190, July 1976
Biochemistry

Structure of L-arabinose-binding protein from *Escherichia coli* at 5 Å resolution and preliminary results at 3.5 Å

(sugar uptake/transport systems/x-ray crystallography/three-dimensional structure)

GEORGE N. PHILLIPS, JR., VIJAY K. MAHAJAN, ALBERT K. Q. SIU, AND FLORANTE A. QUIOCHO

Department of Biochemistry, Rice University, Houston, Texas 77001

Communicated by William N. Lipscomb, March, 29, 1976

ABSTRACT **The three-dimensional crystal structure of the L-arabinose-binding protein from *E. coli*, an essential component in the active transport of L-arabinose, has been solved at 5 Å resolution using the method of multiple isomorphous replacement. Five heavy atom derivatives were used. A preliminary 3.5 Å electron density map has also been calculated. The results indicate that the molecule is ellipsoidal with approximate dimensions 68 Å × 38 Å × 30 Å. Two similar domains within the molecule (which is a single polypeptide chain) are related by an approximate noncrystallographic rotation-translation axis. This relationship involves approximately 20% of the structure.**

L-Arabinose-binding protein is a member of a class of proteins collectively designated as binding proteins. These binding proteins are essential components of some bacterial transport systems and are engaged in the cellular accumulation of specific ligands (i.e., sugars, amino acids, and ions) recognized by the particular protein (1–3). The L-arabinose-binding protein has been shown to be involved in a high affinity uptake system in *Escherichia coli* (4). The protein has a molecular weight of about 38,000 and consists of a single polypeptide chain (4, 5). Furthermore, L-arabinose-binding protein appears to be similar to the D-galactose-binding protein (6), which besides participating in transport plays an essential role in chemotaxis (7).

The single crystal diffraction study of L-arabinose-binding protein is aimed at determining its unique three-dimensional structure and contributing to an understanding of the mechanism of its biological function.

Materials and methods

L-Arabinose-binding protein was purified from *E. coli* B/r strain UP1041 (*araA39*) according to the procedure of Parsons and Hogg (5). The strain was kindly provided by Dr. R. W. Hogg. Crystals of the binding protein suitable for high resolution studies belong to the space group $P2_12_12_1$, with unit cell dimensions: $a = 55.4 \pm 0.2$ Å, $b = 71.8 \pm 0.2$ Å and $c = 77.9 \pm 0.2$ Å and contain one molecule per asymmetric unit (8).

Derivatives were prepared by soaking native crystals in solutions made from heavy atom compounds dissolved in 60% 2-methyl-2,4-pentanediol, 10 mM potassium phosphate buffer at pH 6.5. The derivatives used and other conditions are indicated in Table 1.

Some preliminary screening of heavy atom derivatives was performed by film technique with precession cameras and an Elliot rotating anode generator. The radiation used was Ni-filtered CuKα. Three-dimensional data were collected on a Syntex $P2_1$ diffractometer with a detector to crystal distance of 64 cm. By placing an evacuated tunnel (60 torr, 8 kPa) from the detector to about 2 cm from the crystal, an improvement in peak to background ratio of counts was achieved. Diffraction intensities were measured by the procedure of Wyckoff (9) with a total omega scan of 0.2°. The intensity was then taken as the largest of three consecutive steps of 0.03°. Three reference reflections were measured every 100 reflections to monitor crystal decay. Data collection was terminated when the intensity of any of these monitors dropped below 10% of its initial value. Measured intensities were corrected for background and Lorentz and polarization effects. A semi-empirical absorption correction was also applied (10). The quality of data is indicated in Table 1, which shows results of correlations between overlaps for each derivative.

Location and refinement of heavy atom parameters

The labeling of the single cysteinyl residue of L-arabinose-binding protein with 2-chloromercuri-4-nitrophenol, an environmentally sensitive chromophoric probe, indicated that it may be possible to obtain a single-site, high-occupancy mercury derivative (8). From Harker sections of the three-dimensional difference Patterson maps calculated from the 3.5 Å resolution data of the mercury derivative and native protein, using coefficients $||F_D| - |F_N||^2$ (11), a single mercury site was unambiguously located (Fig. 1). A 4 Å resolution "cross-sign" difference Fourier projection computed with phases from this mercury derivative and difference coefficients from the $NaAuCl_4$ derivative revealed the location of three gold sites. At this stage, three-dimensional least squares refinement at 3.5 Å resolution of the mercury and gold heavy atom parameters was undertaken. Three-dimensional 3.5 Å difference Patterson

Table 1. Heavy atom derivatives

Reagents (concentration)	Soak time	Resolution, Å	No. of crystals	No. of overlaps	R*
Native	—	3.5	4	3951	0.031
Mercurinitrophenol (2 mM)	2 weeks	3.5	4	1274	0.054
Na_2AuCl_4 (4 mM)	1 day	3.5	2	1855	0.035
$(NH_4)_2PtCl_4$ (1.2 mM)	1 week	3.5	5	2299	0.087
K_2IrCl_6 (1 mM)	11 days	5.0	1	—	—
K_2IrCl_6 (1 mM)	6 weeks	3.5	3	537	0.032

* $R = \Sigma|\bar{I}(H) - I(H)|/\Sigma|\bar{I}(H)|$, where $\bar{I}(H)$ is the average intensity of measurement H and $I(H)$ is any measurement of reflection H.

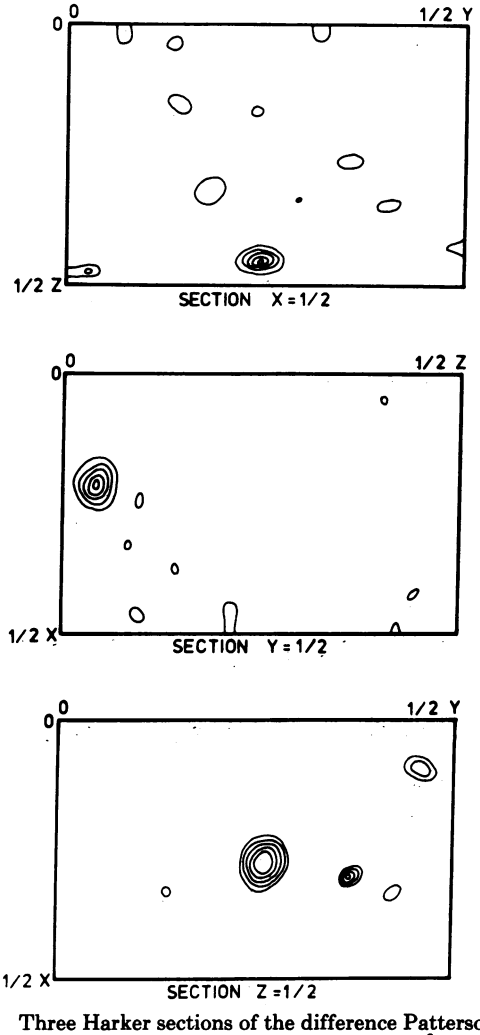

FIG. 1. Three Harker sections of the difference Patterson map of the 2-chloromercuri-4-nitrophenol derivative at 3.5 Å resolution.

maps calculated from $(NH_4)_2PtCl_4$ data proved uninterpretable. However, a difference Fourier computed with phases derived from the mercury and gold derivatives and difference coefficients from the platinum data yielded initially three dominant sites. Difference Fourier analyses using phases obtained from the mercury, gold, and platinum derivatives were principally used for solving the other heavy atoms' positions and for subsequently locating additional minor sites in all the derivatives.

Table 2. 5 Å resolution phasing statistics

Heavy atom (no. of sites)	RMS E*	RMS F_c†	R factor‡
Hg (3)	43.2	99.8	0.056
Pt (8)	79.0	135.6	0.097
Au (4)	37.6	43.9	0.049
Ir, 11 day soak (3)	33.3	53.1	0.044
Ir, 6 week soak (4)	29.8	49.6	0.039

* Root mean square lack-of-closure error $E = (\Sigma_h \epsilon_{hj}^2/n)^{1/2}$, ϵ_{hj} = lack of closure for reflection h of derivative j and n = number of reflections.

† Root mean square heavy atom contribution $F_c = (\Sigma_h f_{hj}^2/n)^{1/2}$, f_{hj} = heavy atom scattering factor.

‡ Kraut R factor = $(\Sigma||F_D| - |F_N + f_D||)/\Sigma|F_D|$, F_D = structure factor of the derivative, F_N for the protein and f_D for the heavy atoms.

Table 3. Heavy atom parameters based on 5 Å resolution refinement

Site	x	y	z	B (Å)²*	Fractional occupancy
		Hg derivative			
1	0.1052	0.1259	0.2282	62.0	0.85
2	0.2072	0.1448	0.2543	39.4	0.22
3	0.3012	0.7417	0.2544	10.0†	0.14
		Pt derivative			
1	0.4536	0.8230	0.3092	92.5	1.05
2	0.3954	0.3103	0.3327	99.9	0.68
3	0.0350	0.1741	0.4717	72.1	0.53
4	0.0327	0.1000	0.2061	27.1	0.16
5	0.4156	0.3315	0.1739	12.2	0.15
6	0.4505	0.1388	0.3133	10.0†	0.15
7	0.4543	0.2307	0.3739	126.4	0.32
8	0.4769	0.1521	0.4512	31.7	0.21
		Au derivative			
1	0.3769	0.3135	0.3449	73.4	0.34
2	0.0048	0.0610	0.2047	68.2	0.15
3	0.2996	0.4904	0.2263	59.2	0.12
4	0.0655	0.0993	0.2353	10.0†	0.10
		Ir derivative (11 day soak)			
1	0.4704	0.2762	0.4328	103.0	0.49
2	0.4709	0.2221	0.3900	101.3	0.35
3	0.0708	0.6545	0.0047	90.7	0.18
		Ir derivative (6 week soak)			
1	0.4728	0.2780	0.4353	98.0	0.40
2	0.0090	0.1200	0.8361	62.2	0.22
3	0.4724	0.2290	0.3985	115.3	0.23
4	0.4248	0.3455	0.5109	107.7	0.23

* Isotropic temperature factor.

† Parameters that were not refined.

The final heavy atom parameters were obtained by a three-dimensional refinement procedure (12) which alternates cycles of multiple isomorphous phase angle determination (13) with least squares adjustment of the heavy atom parameters. Phase angle probabilities were calculated at 5° intervals. Each data set used in the refinement was prepared by merging overlapping data collected from several crystals (Table 1). In the 5 Å resolution refinement, three of the 22 temperature factors were fixed during refinement. During the course of the refinement, the root mean square lack of closure error E, the root mean square heavy atom contribution F_c, and the figure of merit $\langle m \rangle$ were monitored as functions of the Bragg angle

An absolute scale factor for the data was obtained initially from Wilson statistics of the 3.5 Å native protein data set but subsequently was reduced to give a 100% occupancy for the most substituted platinum site. Scale factors relating derivatives to native data were approximated initially from the ratios of the average native F_{obs} to the derivative F_{obs}, with a 5% allowance for heavy-atom scattering, and were refined during least squares refinement.

Final phasing statistics for the 5 Å resolution data are given in Table 2. The mean figure of merit at this resolution was 0.75 for 1455 reflections. Phases calculated to 3.5 Å using gold and iridium to 5 Å resolution together with mercury and platinum to 3.5 Å resulted in an $\langle m \rangle$ of 0.66 for 4072 reflections. Since the F_c contributions from gold and iridium derivatives in the range 5–3.5 Å were less than or equal to E, these derivatives were not included in the phasing to 3.5 Å resolution. The final

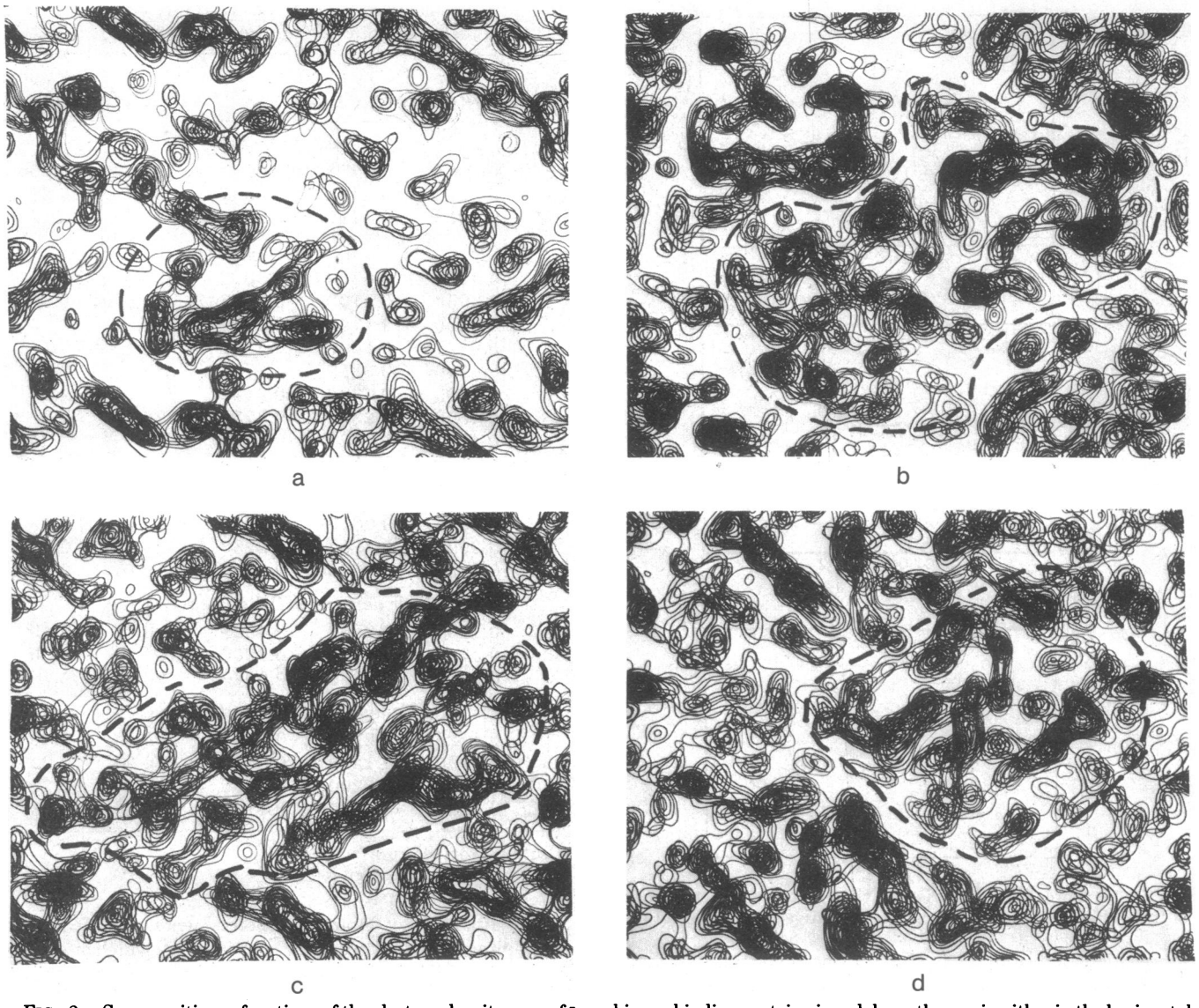

FIG. 2. Superpositions of sections of the electron density map of L-arabinose-binding protein viewed down the z axis with y in the horizontal direction. The width of an entire section corresponds to 79 Å. The photographs correspond to the following sections. (a) $z = 0.068$ to 0.136. (b) $z = 0.218$ to 0.313. This composite shows the double domain feature of the structure. (c) $z = 0.340$ to 0.462. Overlaps (polypeptide chains) between domains are shown in this superposition. (d) $z = 0.462$ to 0.600. This composite shows the top of the molecule.

heavy atom parameters from the 5 Å resolution refinement are given in Table 3.

Electron density maps and interpretation

Both the 5 Å and the preliminary 3.5 Å resolution electron density maps were calculated using centroid phases to yield the "best" Fourier map (13). Maps with boundaries $x = -0.40$ to 0.75, $y = 0.2$ to 1.3, and $z = -0.07$ to 0.60 displayed an entire molecule, intermolecular boundaries, and portions of symmetry-related molecules. Electron density maps were drawn at a scale of 0.6 cm/Å. In these maps, the molecular envelope is clear except for possibly two regions of close contact with adjacent molecules. The molecule appears to be elongated and when viewed from the $-c$ direction shows two, more or less globular, domains. Fig. 2 shows selected composites of sections of the 5 Å resolution map. The map is viewed down the z axis with fractional intervals of 0.0136. Contour levels are at equal increments of 0.06 e Å^{-3} starting at 0.12 e Å^{-3}. Fig. 2a shows the bottom of the molecule and the beginning of one domain. Fig. 2b, which comprises sections through the middle of the molecule, shows a cross section through the two domains. A stretch of what may be considered a helix parallel to the screw axis along b is also evident in this figure. Stretches of electron density that traverse the two domains are seen clearly in the next superposition, Fig. 2c. Fig. 2d shows the top of the molecule.

A styrofoam model of the 5 Å map which includes electron density down to 0.18 e Å^{-3} was constructed and is shown in Fig. 3. The approximate molecular dimensions measured from this model are 68 Å × 38 Å × 30 Å, indicating that the overall shape of the molecule is ellipsoidal. A view of the molecule along a line bisecting the a and c directions is shown in Fig. 3a. Long stretches of continuous density can be seen in various parts of the model. The double domain feature of the structure is again evident in Fig. 3b.

At least four rodlike features of electron density that most likely correspond to α-helices of at least three turns were observed in the low resolution map.

Preliminary electron density maps calculated at 3.5 Å resolution show improved detail in many regions (e.g., Fig. 4) and the interpretations are basically compatible with the low reso-

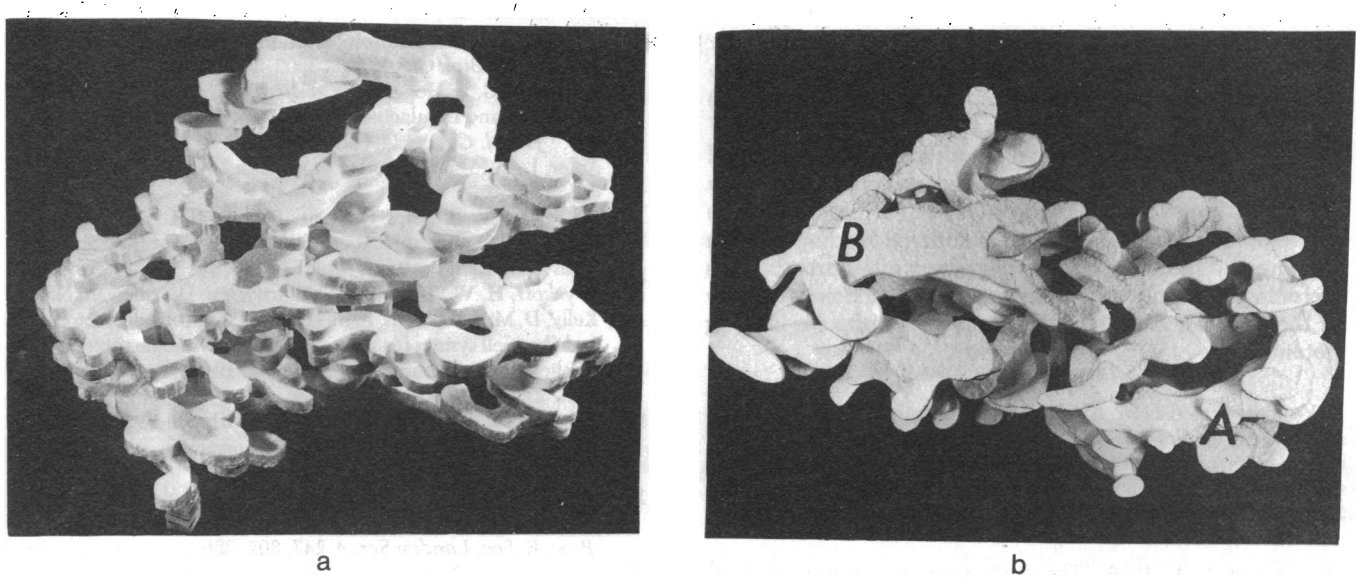

FIG. 3. Styrofoam model of L-arabinose-binding protein. (a) A view along a line bisecting the *a* and *c* axes. (b) A view up the *c* axis showing the approximate rotation-translation symmetry (see *text*). The *b* axis is approximately horizontal in both views.

lution study. Several long stretches of polypeptide chain can be traced from the 3.5 Å resolution map; however, an unambiguous trace of the entire chain is not yet possible. Other information derived from the 3.5 Å map includes the verification of the helical chains found in the low resolution map, an example of which is shown in Fig. 4. Furthermore, the correct handedness of the structure has been established on the basis of the handedness of the helices.

The current 3.5 Å map shows some indication of phasing error in the form of high ($2\ e\ Å^{-3}$) and low ($-3\ e\ Å^{-3}$) regions of electron density concentrated at the most substituted mercury and platinum sites. The highest electron density observed elsewhere in helical regions is about $1\ e\ Å^{-3}$. Neither maximum probability phase calculation (13) nor refinement of scale and temperature factors in spherical shells helped to correct this error. This error was not observed in the 5 Å map.

The unambiguous delineation of one molecule will depend upon tracing of the polypeptide chain with the aid of the amino acid sequence and ultimately upon atomic model building. Another attempt can be made to trace the chain upon improvement of the 3.5 Å resolution map. We have solved very recently a CdI_2 derivative by difference Patterson analysis that shows the high quality required for extension to 3.5 Å resolution. Since L-arabinose-binding protein crystals diffract to about 2 Å resolution, structure determination can be extended to 2.5 Å resolution.

Similarities between two domains

A helix in one domain (labeled *A* in Fig. 3b) appears to be related to another helix in the other domain (helix *B*) by rotation of approximately 160° about *c* and translation of 11 Å along *c*. Furthermore, other regions of electron density near these two helices show some resemblance and are related by this rotation-translation axis. The structural similarity between domains encompasses approximately 20% of the structure. Whether or not similarity between two domains as observed in the structure of the L-arabinose-binding protein is a common feature of binding proteins that have been implicated in transport systems remains to be ascertained. There are some indications which suggest that such a feature may exist in other binding proteins.

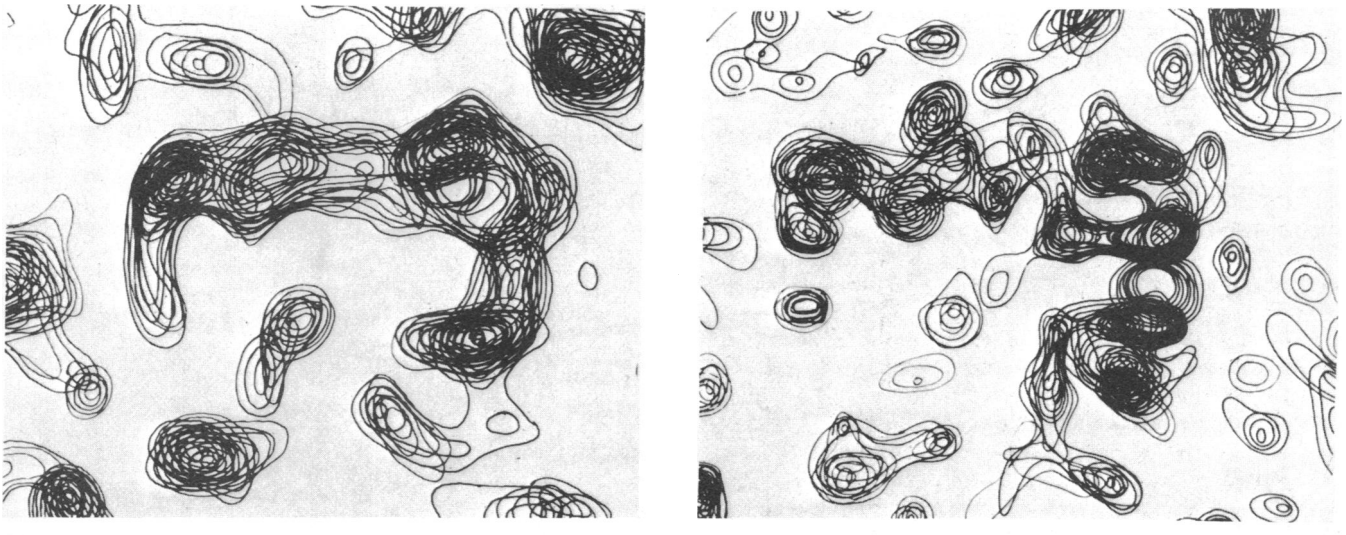

FIG. 4. Comparison of a specific region in the electron density at 5 Å and 3.5 Å resolution. The bounds are: $x = -0.15$ to 0.35, $y = 0.70$ to 1.2, and $z = 0.23$ to 0.27.

For instance, although L-arabinose-binding protein in many aspects differs from D-galactose-binding protein, the two show similarities in structure (observed as antigenic crossreactivity) (6). Preliminary evidence suggests that the minor A form of bovine calcium-binding protein, another protein implicated in transport, also appears to possess some structural homology between different regions of the molecule (14). In view of the double-domain feature of the L-arabinose-binding protein structure and the above observations from other proteins, it is interesting to speculate that there may be a functional necessity for the existence of two similar domains in order for binding proteins to play their role in transport.

We are indebted to Dr. R. W. Hogg for the gift of the *E. coli* strain and for continued interest, and to Miss Rebecca La Brum for invaluable help in contouring electron density maps. We wish to acknowledge support of this research by National Institutes of Health Grant GM-21371 and a grant from the Robert A. Welch Foundation, C-581.

1. Heppel, L. A. (1969) "The effect of osmotic shock on release of bacterial proteins and on active transport," *J. Gen. Physiol.* **54**, 95s–109s.
2. Oxender, D. L. (1974) "Membrane transport proteins," *Biomembranes* **5**, 25–79.
3. Boos, W. (1974) "Bacterial transport," *Annu. Rev. Biochem.* **43**, 123–146.
4. Hogg, R. W. & Englesberg, E. (1969) "L-Arabinose binding protein from *Escherichia coli* B/r," *J. Bacteriol.* **100**, 423–432.
5. Parson, R. G. & Hogg, R. W. (1974) "Crystallization and characterization of the L-arabinose binding protein of *Escherichia coli* B/r," *J. Biol. Chem.* **249**, 3602–3607.
6. Parson, R. G. & Hogg, R. W. (1974) "A comparison of the L-arabinose- and D-galactose-binding proteins of *Escherichia coli* B/r," *J. Biol. Chem.* **249**, 3608–3614.
7. Adler, J. (1975) "Chemotaxis in bacteria," *Annu. Rev. Biochem.* **44**, 341–356.
8. Quiocho, F. A., Phillips, G. N., Jr., Parsons, R. G. & Hogg, R. W. (1974) "Crystallographic data of an L-arabinose-binding protein from *Escherichia coli*," *J. Mol. Biol.* **86**, 491–493.
9. Wyckoff, H. W., Doscher, M., Tsernoglou, D., Allewell, N. M., Kelly, D. M. & Richards, F. M. (1967) "Design of a diffractometer and flow cell system for x-ray analysis of crystalline proteins with applications to the crystal chemistry of ribonuclease-S," *J. Mol. Biol.* **27**, 563–578.
10. North, A. C. T., Phillips, D. C. & Matthews, F. S. (1968) "A semiempirical method of absorption correction," *Acta Crystallogr. Sect. A* **24**, 351–359.
11. Blow, D. M. (1958) "The structure of haemoglobin VII. Determination of phase angles in the non-centrosymmetric [100] zone," *Proc. R. Soc. London Ser. A* **247**, 302–336.
12. Lipscomb, W. N., Coppola, J. C., Hartsuck, J. A., Ludwig, M. L., Muirhead, H., Searl, J. & Steitz, T. A. (1966) "The structure of carboxypeptidase A. III. Molecular structure at 6 Å resolution," *J. Mol. Biol.* **19**, 423–441.
13. Blow, D. M. and Crick, F. H. C. (1959) "The treatment of errors in the isomorphous replacement method," *Acta Crystallogr.* **12**, 794–802.
14. Moffat, K., Fullman, C. S. & Wasserman, R. H., (1975) "Preliminary crystallographic data for a calcium binding protein from bovine intestine," *J. Mol. Biol.* **97**, 661–664.