Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 1792–1796, April 1983
Biochemistry

The 3 Å resolution structure of a D-galactose-binding protein for transport and chemotaxis in *Escherichia coli*

(chemoreceptor/chain trace/structure similarity/molecular and multiple isomorphous replacement)

N. K. VYAS, M. N. VYAS, AND F. A. QUIOCHO

Department of Biochemistry, Rice University, Houston, Texas 77251

Communicated by David R. Davies, December 16, 1982

ABSTRACT **X-ray diffraction studies of a D-galactose-binding protein essential for transport and chemotaxis in *Escherichia coli* have yielded a model of the polypeptide chain backbone. An initial polyalanine backbone trace was obtained at 3.2 Å resolution by the molecular replacement technique, using a polyalanine search model derived from the refined structure of the L-arabinose-binding protein. Concurrently, a 3 Å resolution electron-density map of the D-galactose receptor was determined from multiple isomorphous replacement (MIR) phases. The properly transformed initial polyalanine model superimposed on the MIR electron-density map proved to be an excellent guide in obtaining a final trace. The few changes made in the polyalanine model to improve the fit to the density were confined primarily to the COOH-terminal peptide and some loops connecting the elements of the secondary structure. Despite the lack of significant sequence homology, the overall course of the polypeptide backbone of the D-galactose-binding protein is remarkably similar to that of the L-arabinose-binding protein, the first structure in a series to be solved from this family of binding proteins. Both structures are elongated (axial ratios of 2:1) and composed of two globular domains. For both proteins, the arrangements of the elements of the secondary structure in both domains are identical; both lobes contain a core of β-pleated sheet with a pair of helices on either side of the plane of the sheet. The four major hydrophobic clusters that stabilize the structure of the L-arabinose-binding protein are also present in the D-galactose-binding protein.**

A molecular understanding of the role of periplasmic binding proteins (or receptors) in active transport and chemotaxis in bacteria requires a detailed picture of this family of proteins. Accordingly, an x-ray structural analysis of several receptors—at least one from each of the major groups with specificity for amino acids, ions, or carbohydrates—has been initiated. This goal is imminently attainable because single crystals of five binding proteins specific for L-arabinose, D-galactose, D-maltose, sulfate, and leucine, isoleucine, or valine have been obtained (1) and recently considerable success in solving four structures has been achieved. The refinement at 1.7 Å resolution of the molecular structure of the L-arabinose-binding protein, the first of the series to be determined (2), is near completion. The protein is elongated and composed of two similar globular domains that have extensive βαβ units (2), and the binding site is located in a cleft between the two lobes (3, 4). The D-galactose chemoreceptor has been solved at 4.1 Å resolution by the multiple isomorphous replacement (MIR) technique (5), and very recently, 3 Å resolution Fourier maps calculated from the MIR phases for the sulfate-binding protein and the leucine/isoleucine/valine-binding protein were obtained (unpublished data). Moreover, amino acid sequences of several binding proteins of interest have been determined—L-arabinose-binding protein (6), D-galactose-binding protein (7), leucine/isoleucine/valine-binding protein (8), and sulfate-binding protein (9).

Phillips *et al.* (10) originally made the suggestion that the overall conformation of the L-arabinose-binding protein seen at 3.5 Å is likely to be characteristic of this family of proteins (see also ref. 11). This subsequent observation (5) that "the structure of the D-galactose binding protein at 4.1 Å resolution looks like L-arabinose-binding protein" is, at least for the class of carbohydrate-binding proteins, consistent with this proposal. Parsons and Hogg (12) have previously indicated, on the basis of antibody crossreactivity, that the two sugar receptors share some regions of similar conformation. Moreover, circular dichroic measurements of both sugar-binding proteins show virtually identical spectra, indicative of similar extensive secondary structures (refs. 3, 11; unpublished data). These results constitute the basis for the use of the molecular replacement technique to augment the structural analysis of the D-galactose-binding protein. Here we report the results of not only the application of this technique with the refined 1.9 Å structure of the L-arabinose-binding protein as the search molecule but also the extension of the MIR analysis to 3 Å. These complementary studies have enabled us to obtain a model of the polypeptide backbone of the D-galactose-binding protein. A preliminary account of this work has been presented (13).

MATERIALS AND METHODS

D-Galactose-binding protein crystals belong to the space group $P2_1$ with unit cell dimensions a = 66.00 Å, b = 37.05 Å, c = 61.57 Å, and β = 106.8° (5) whereas L-arabinose-binding protein crystals exhibit the symmetry of space group $P2_12_12_1$ with cell dimensions a = 55.46 Å, b = 71.82 Å, and c = 77.84 Å (2). The asymmetric units of both crystals contain one protein molecule. Three-dimensional diffraction intensities from native D-galactose-binding protein crystals to 2.6 Å resolution and crystals soaked in heavy atom derivatives to 3 Å resolution were measured using a Syntex $P2_1$ diffractometer and corrected for background, absorption, Lorentz polarization, and radiation decay as described for the L-arabinose-binding protein (2).

For the molecular replacement study, the known or search molecule consisted of only the main-chain and β-carbon atoms of the L-arabinose-binding protein. The coordinates of the polyalanine L-arabinose-binding protein model were obtained after refinement of the complete structure at 1.9 Å resolution by the Hendrickson–Konnert method (14) to an R factor of 24% (unpublished data). The model structure factors were calculated with the model placed in a triclinic (orthogonal) cell

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: MIR, multiple isomorphous replacement.

of dimensions a = 75 Å, b = 95 Å, c = 75 Å. The axial lengths were chosen to be large to avoid interference due to intermolecular vectors within the sphere of comparison in the rotational function. The center of mass of the model was placed at the origin of the cell.

The fast rotation function of Crowther (15) was used in this study. The rotation functions were calculated using reflections with large values of structure factor terms (1,755 for D-galactose-binding protein and 1,353 for L-arabinose-binding protein) in the resolution range 16–3.2 Å. Patterson search functions were calculated by varying Crowther's eulerian angles, α, β, and γ in 5° steps. A search radius around the Patterson function origin was limited to 16 Å, and only the asymmetric unit of the rotation space was investigated.

The solution of the galactose receptor structure at 3 Å resolution by the MIR technique was achieved by using the heavy atom derivatives indicated in Table 1. MIR refinement was carried out in the usual procedure of Blow and Crick using the PHASEREF phasing program (17). The mean figure of merit at 3 Å resolution for 5,211 reflections is 0.54. Other refinement statistics are summarized in Table 1.

RESULTS AND DISCUSSION

Two factors that could have a bearing on the molecular replacement analysis had to be taken into consideration at the outset. First, the decision to use the polyalanine model rather than the complete refined structure of the L-arabinose-binding protein as the search molecule was initially based on the finding that the amino acid compositions of the L-arabinose- and D-galactose-binding proteins, both from *E. coli* B/r, are dissimilar (12); this is in agreement with the later result indicating only 18% sequence homology between the L-arabinose-binding protein (6) and the D-galactose chemoreceptor (7). Despite these findings, both sequences show that the polypeptide chain lengths for both sugar-binding proteins are nearly the same; D-galactose-binding protein has 309 residues, only 3 more than L-arabinose-binding protein. Moreover, it has been predicted recently that both proteins have similar secondary structures (18). Second, although there is evidence for a sugar-induced conformational change in the two proteins, crystals of both were obtained from purified proteins with endogenous bound sugar substrate. L-Arabinose is bound to the L-arabinose-binding protein (4), whereas D-glucose is bound to the D-galactose-binding protein (19).

Calculation of a rotation function at 3.2 Å resolution of the polyalanine model Patterson density against the D-galactose-binding protein Patterson density showed a single large peak, which is 1.36 times the next largest peak, at eulerian angles α = 110°, β = 35°, and γ = 125°. These orientational parameters were applied to the original polyalanine search model coordinates and the oriented model was subsequently placed in the D-galactose-binding protein cell.

Table 1. Summary of 3.0 Å MIR phasing statistics for the D-galactose-binding protein

Derivative	Sites, no.	rms residual	rms F_H	R-Cullis	Reflections phased, no.	Resolution, Å
K_2PtCl_4	3	41.62	72.59	0.66	4,269	3.0
Uranyl acetate	4	34.77	59.43	0.58	1,927	4.1
Baker's mercurial	2	20.82	31.51	0.56	1,274	4.1
Cadmium acetate	1	16.86	25.89	0.68	4,394	3.0

For definitions of terms, see ref. 16.

Attempts to position the oriented polyalanine molecule by a Patterson space translation function suggested by Langs (20) or by a systematic search for maximum overlap of the polyalanine model coordinates into the 4.1 Å MIR electron density proved unsuccessful. (A comparison of the former method with the Q and T translation functions is found in ref. 20. The latter direct-space search method was conducted at increments of 2 Å along the three axes using a computer program that is part of the crystallographic computing package PROTEIN, kindly provided by W. Steigemann; see also ref. 17.) Reasons for this failure became evident after the translation search was successfully performed by visual examination and careful comparison of a 4.1 Å Fourier map of the oriented L-arabinose-binding protein polyalanine search model with the original 4.1 Å MIR map of the D-galactose-binding protein (5). The Fourier map of the search model in the D-galactose-binding protein cell was obtained from calculated structure factors. Both electron-density maps were contoured at the same scale, and sections down the b axis were drawn onto plastic sheets. The overall similarity of the two protein structures was again apparent by comparison of the maps. However, only when the original D-galactose-binding protein MIR map was restacked in the reverse direction (up the b axis), to display the other enantiamorph, could all eight known major helices (or four helices per domain) in the model Fourier be brought to overlap with columns of density from helices similarly located in the D-galactose-binding protein MIR map by a single translation vector joining the origins of both maps. (Axes of one map relative to the other were kept parallel during this translation.) Components of the translation vector thus obtained (x = 0.78, y = 0.80, and z = 0.85) were applied to the coordinates of the rotated polyalanine model in the D-galactose-binding protein cell. The resulting transformed model was now superimposable on the D-galactose-binding protein electron-density map recalculated with 4.1 Å MIR phases in the correct handedness. This superposition revealed that the initial rotation parameters required minor adjustment to further maximize the fit of the model to the density. From the relative orientation between the group of helices in the model density and the corresponding set of helices in the D-galactose-binding protein MIR map, we obtained a correction of the initial rotation parameter of (12, −10,0) eulerian angles and residual translational shifts of Δx = −0.005, Δy = 0.015, and Δz = 0.002.

Finally, a Patterson translation search using Langs' method (20), following modification of the initial rotational parameters, provided an independent confirmation of the correctness of the initial translation parameters. Fig. 1 shows the map of the translation function Φ ($2x$,1/2,$2z$) (defined by Langs) calculated by using 475 large terms in the 14–3.2 Å range. The single dominant peak, which is 1.24 times the second highest peak, occurs for the particular choice of the Φ function origin at $2x$ = 1.58 and $2z$ = 1.72. The required translation vector for this peak relative to the 2_1 axis (the crystal origin) is x = 0.79 and z = 0.86, in close agreement with initial values obtained by visual comparison. Though the y value can be arbitrarily chosen, it had been previously fixed at 0.0 in the MIR analysis (5). It should be noted that, for the calculation of the translation function to work, it was necessary to use the final corrected rotation parameters. On the other hand, the real-space translation search failed because of the incorrect enantiamorphy of the initial 4.1 Å MIR map compared with the known polyalanine model of the L-arabinose-binding protein.

Structure factors calculated for the final molecular replacement solution resulted in an R factor of 53% for data from 8–3

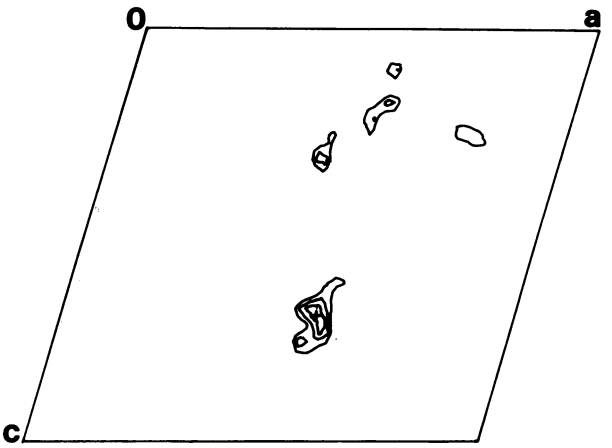

FIG. 1. Section $y = 1/2$ from the three-dimensional translation function Φ $(2x,1/2,2z)$ for D-galactose-binding protein. The contours are drawn at equal intervals of 0.5σ, starting with a value of 2σ. +, Expected peak for the particular choice of origin.

Å resolution. Considering that the calculation is for the polyalanine model, the R factor compares favorably with values obtained from other molecular replacement studies of proteins having identical or closely homologous sequences (17, 21–23).

Fig. 2 shows two contiguous composites of the 4.1 Å MIR electron-density map, encompassing a major portion of the D-galactose-binding protein molecule, overlaid with the polyalanine model positioned according to the final orientation and translation parameters. This figure shows the good coincidence between the model and the MIR map, especially that of the six helices found in this region, prominently.

A further useful method for assessing the correctness of the final rotation and translation parameters is reflected by the successful use of the resulting molecular replacement phases to independently locate, in 4.1 Å difference Fourier maps, major heavy atom sites from the platinum, mercury, and uranyl derivatives. (It is important to note that similar difference Fourier analyses using calculated phases from the initial uncorrected orientation and translation parameters gave negative results.) The same sites were originally located by difference Patterson or difference Fourier maps calculated with MIR phases in the 4.1 Å resolution structural analysis (5).

Although on the whole the helices and β-sheet strands from the molecular replacement model coincided with the corresponding segments of density (Fig. 2), we observed that several of the β-carbon atoms are shifted from the main density that clearly belongs to side chain residues. The fit of some of the loops connecting the elements of the secondary structure and the COOH-terminal peptide was far less satisfactory. Reciprocal space rigid-body refinement in the range 8–3.2 Å resolution of the polyalanine model using the program CORELS (24) failed to significantly improve the fit, although the R factor decreased to 35%, likely because of the absence of side chains in the model to serve as additional guides. The logical recourse at this juncture would have been to adjust the model and fit the side chain residues, but the computer graphics system (Evans and Sutherland PS300) for molecular modeling was not acquired until a year later. Moreover, we preferred to rebuild the model with respect to density maps calculated from higher resolution MIR phases or from improved phases derived from a combination of MIR and the current model (or both).

Concurrent with the molecular replacement study, the MIR

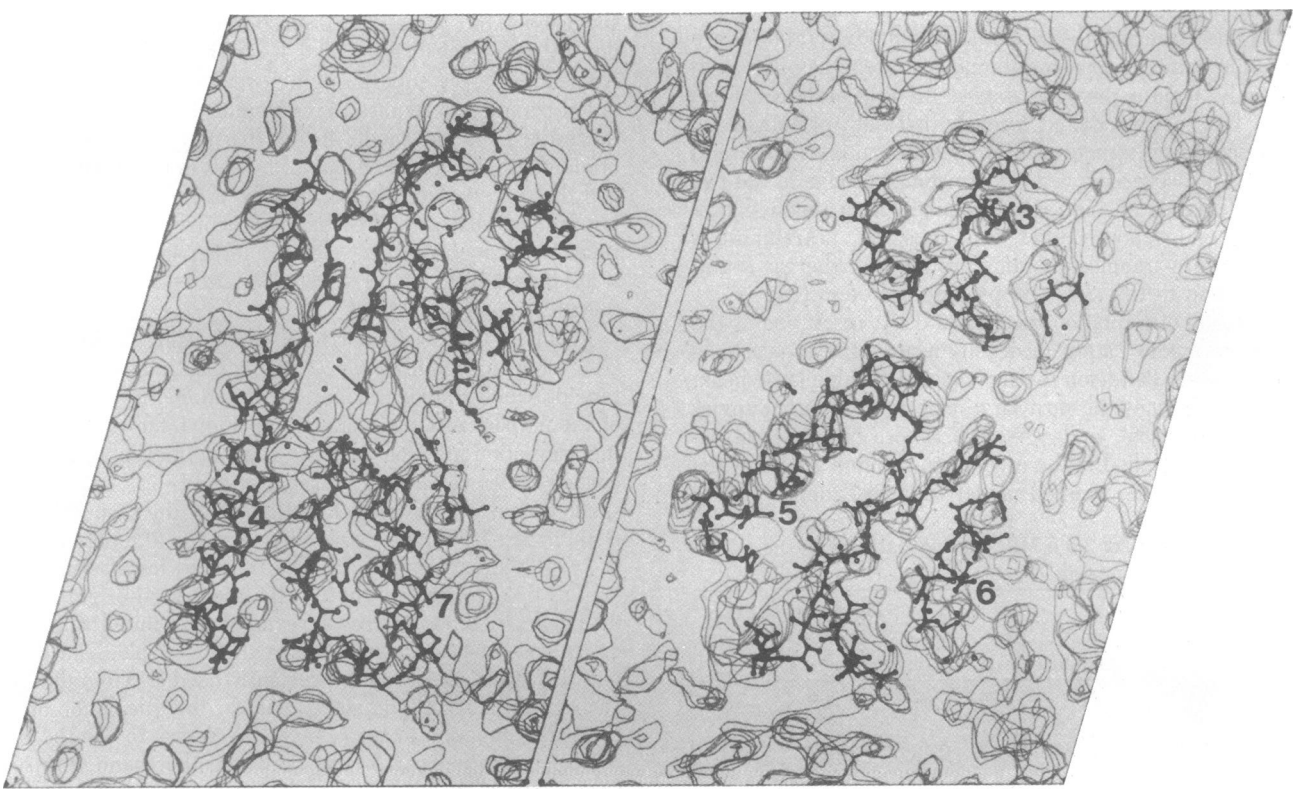

FIG. 2. Two continguous composites of the 4.1 Å MIR electron density of the D-galactose-binding protein perpendicular to the *b* axis (37.05 Å) superimposed with a properly rotated and translated polyalanine model of the L-arabinose-binding protein. Each composite is 7 Å thick. (*Left*) $y = 0.73$ to $y = 0.92$. (*Right*) $y = 0.97$ to $y = 1.16$. The numbers identifying the α-helices are identical to those originally used to designate helices in the L-arabinose-binding protein structure (2). The arrow points to an extraneous density most likely from a bound endogenous D-glucose molecule.

FIG. 3. Stereo views of the C_α backbone trace of the D-galactose-binding protein (*Upper*) and the L-arabinose-binding protein (*Lower*). The D-galactose-binding protein model was obtained from the 3 Å MIR density map using an Evans and Sutherland PS300 computer graphics system. The coordinates for the L-arabinose-binding protein model were derived from the structure refined at 1.9 Å resolution. The significance of the numbers is explained in Fig. 2. N and C, NH_2 and COOH termini.

phases of the D-galactose-binding protein were extended from 4.1 to 3 Å resolution. From the resulting electron-density map calculated with "best phases," we were able to obtain a model of the polypeptide backbone of the D-galactose-binding protein by using the computer graphics system.* The model is shown in Fig. 3. The result of the molecular replacement analysis greatly aided us in the elucidation of the model. Although the polyalanine search model properly superimposed onto the 3 Å MIR map served as the initial guide, it was, as noted above, necessary to make adjustments to better fit the density, which at this resolution showed more clarity overall. Also, as noted above, the COOH-terminal peptide, commencing at residue 286, and loops required substantial changes.

The complementary results arising from both the molecular and MIR analyses substantiate and further extend our original observation that the D-galactose chemoreceptor looks like L-arabinose-binding protein. Comparison of the backbone trace of the D-galactose-binding protein with that of the 1.9 Å refined L-arabinose-binding protein structure (Fig. 3) further underscores the remarkable structural similarity of the two sugar receptors. With the exception of the few isolated segments discussed above, the course of the polypeptide backbone of the D-galactose-binding protein generally follows that of the L-arabinose-binding protein. The relative orientations of the two similar globular domains, linked by no fewer than three peptide strands, and the topological arrangements of the helices and β-sheet strands are similar in both proteins (Fig. 3). Each domain is characterized by a core of β-sheet with a pair of helices on either side of the plane of the sheet and aligned antiparallel with the strands.

A further measure of the close structural similarity of the two sugar-binding proteins, though expected from molecular replacement analysis, is indicated by the unusually high percentage of equivalent pairs of α-carbon atoms. Using the procedure of Rossmann and Argos (25), we found a total of 238 (about 80%) equivalencies, with a rms deviation of 2.62 Å, between the α-carbon atoms of the preliminary trace of the D-galactose-binding protein derived from the 3 Å MIR map and the L-arabinose-binding protein obtained from its properly transformed coordinates. However, most of the nonequivalent C_α atoms deviating in the range 2.62–9.18 Å are located in loops connecting elements of the secondary structure and in the COOH-terminal peptide, which is helical in the case of the L-arabinose-binding protein.

Though a common x-ray structure of two sugar-binding proteins has been established, further detailed assessment of the entire family of binding proteins requires the inclusion of structures from the two other classes. Simultaneous with the D-galactose-binding protein structure determination, 3 Å density maps have been determined for a sulfate-binding protein from *Salmonella typhimurium* and a leucine/isoleucine/valine-binding protein from *E. coli* (unpublished data; ref. 13). These bring the number of structures solved thus far in our laboratory to a total of four, including at least one from each major class. In light of this accomplishment, a preliminary assessment of the overall structural features of the family of binding proteins can be made—all four structures are elongated (axial ratios of 2:1) and composed of two globular domains. Furthermore, the following additional results relevant to the functions of binding proteins have been obtained: (*i*) the ligand sites of the L-arabinose-binding protein (3, 4), the leucine/isoleucine/valine-binding protein (unpublished data), and the D-galactose receptor (unpublished data; Fig. 2) are located in a cleft between the two domains; (*ii*) occupation of the binding site in the L-arabinose receptor is accompanied by a conformational change (3, 4); and (*iii*) the bimolecular reactions of several binding proteins with their respective ligands are fast (association rates are of the order of 10^7 $M^{-1}\cdot s^{-1}$), and rates of ligand dissociation (off rates) are not likely to be the rate-limiting step in transport (19, 26).

It has been noted that the structural and functional integrity of the L-arabinose-binding protein is conferred by four major clusters, two in each domain, consisting almost entirely of hydrophobic residues (2). These clusters, each located at the interface between the β-sheet and a pair of helices, are also present in the D-galactose-binding protein and undoubtedly in the other structures currently being analyzed. The maintenance of these hydrophobic clusters, resulting from the aggregation of secondary structures in each domain, accounts for the structural similarity of binding proteins. In this case, a substantial sequence similarity is not necessary, and none is observed. Moreover, the domain structures commonly found in binding proteins bear strong resemblances to others seen in several unrelated proteins (27). The significance of a bilobate structure in ligand binding and concomitant conformational change has been discussed elsewhere (3, 28).

*The unique feature of the Evans and Sutherland PS300 that enables a large volume of electron density (one-third of the molecule) and complete model, or about 20,000 flicker-free vectors, to be displayed at one time facilitated the chain tracing of D-galactose-binding protein and two other proteins, all in a period of about 1 month, using a program written by J. W. Pflugrath.

In addition to binding proteins, components residing in the cytoplasmic membrane distinct for either chemotaxis or transport are required for both processes (29, 30). Furthermore, each component apparently recognizes a group of binding proteins, preferably with bound ligands. In conjunction with this, single crystals of a mutant D-galactose-binding protein purified from *E. coli* strain AW551 have recently been obtained (unpublished data). The mutant *E. coli* (provided by M. Dahl and J. Adler) is defective in chemotaxis toward D-galactose but fully competent in translocation of the sugar (31). The x-ray analysis of the mutant D-galactose-binding protein, together with the known structure of native D-galactose-binding protein, should prove helpful in assessing the structural basis and functional significance of the interaction between the binding protein and transmembrane chemotactic components.

The assistance of J. W. Pflugrath and Mark A. Saper in the use of the PS300 computer graphics system is deeply appreciated. This work was supported by Grant GM-21371 from the National Institutes of Health and by Grant C-581 from the Robert A. Welch Foundation.

1. Quiocho, F. A., Meador, W. E. & Pflugrath, J. W. (1979) *J. Mol. Biol.* **133**, 181–184.
2. Gilliland, G. L. & Quiocho, F. A. (1981) *J. Mol. Biol.* **146**, 341–362.
3. Newcomer, M. E., Miller, D. M., III & Quiocho, F. A. (1979) *J. Biol. Chem.* **254**, 7529–7533.
4. Newcomer, M. E., Gilliland, G. L. & Quiocho, F. A. (1981) *J. Biol. Chem.* **256**, 13213–13217.
5. Quiocho, F. A. & Pflugrath, J. W. (1980) *J. Biol. Chem.* **255**, 6559–6561.
6. Hogg, R. W. & Hermodson, M. A. (1977) *J. Biol. Chem.* **252**, 5135–5141.
7. Mahoney, W. C., Hogg, R. W. & Hermodson, M. A. (1981) *J. Biol. Chem.* **256**, 4350–4356.
8. Ovchinnikov, R. A., Aldanova, N. A., Grinkevich, V. A., Arzamazova, N. M. & Moroz, I. N. (1977) *FEBS Lett.* **78**, 313–316.
9. Isihara, H. & Hogg, R. W. (1980) *J. Biol. Chem.* **255**, 4614–4618.
10. Phillips, G. N., Jr., Mahajan, V. K., Siu, A. K. Q. & Quiocho, F. A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2186–2190.
11. Quiocho, F. A., Gilliland, G. L., Miller, D. M. & Newcomer, M. E. (1977) *J. Supramol. Struct.* **6**, 503–518.
12. Parsons, R. G. & Hogg, R. W. (1974) *J. Biol. Chem.* **249**, 3608–3614.
13. Quiocho, F. A., Vyas, N. K., Pflugrath, J. W., Saper, M. A. & Vyas, M. N. (1982) *American Crystallographic Association Summer Meeting Program and Abstracts* (La Jolla, CA), Vol. 10, p. 12.
14. Konnert, J. H. & Hendrickson, W. A. (1980) *Acta Crystallogr., Sect. A* **36**, 344–350.
15. Crowther, R. A. (1972) in *The Molecular Replacement Method. A Collection of Papers on the Use of Noncrystallographic Symmetry*, ed. Rossmann, M. G. (Gordon & Breach, New York), pp. 173–178.
16. Blundel, T. L. & Johnson, L. N. (1976) *Protein Crystallography* (Academic, New York).
17. Remington, S., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* **158**, 111–152.
18. Argos, P., Mahoney, W. C., Hermodson, M. A. & Hanei, M. (1981) *J. Biol. Chem.* **256**, 4357–4361.
19. Miller, D. M., III, Olson, J. S. & Quiocho, F. A. (1980) *J. Biol. Chem.* **255**, 2465–2471.
20. Langs, D. A. (1975) *Acta Crystallogr., Sect. A* **31**, 543–550.
21. Navia, M. A., Segal, D. M., Padlan, E. A., Davies, D. R., Rao, N., Rudikoff, S. & Potter, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4071–4074.
22. Buehner, M., Lifchitz, A., Bally, R. & Mornon, J. P. (1981) *J. Mol. Biol.* **159**, 353–358.
23. Murthy, M. R. N., Garavito, R. M., Johnson, J. E. & Rossmann, M. G. (1980) *J. Mol. Biol.* **138**, 859–872.
24. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. (1977) *Acta Crystallogr., Sect. A* **33**, 800–804.
25. Rossmann, M. G. & Argos, P. (1975) *J. Biol. Chem.* **250**, 7525–7532.
26. Miller, D. M., III (1981) Dissertation (William Marsh Rice Univ., Houston, TX).
27. Gilliland, G. L. (1979) Dissertation (William Marsh Rice Univ., Houston, TX).
28. Newcomer, M. E., Lewis, B. A. & Quiocho, F. A. (1981) *J. Biol. Chem.* **256**, 13218–13222.
29. Higgins, C. F., Haas, P. D., Nikaido, K., Ardeshir, F., Garcia, G. & Ames, G. F.-L. (1982) *Nature (London)* **298**, 723–727.
30. Koshland, D. E., Jr. (1981) *Annu. Rev. Biochem.* **50**, 765–782.
31. Hazelbauer, G. L. & Adler, J. (1971) *Nature (London)* **230**, 101–104.