2 Å Resolution Structure of DppA, a Periplasmic Dipeptide Transport/ Chemosensory Receptor[†,‡]

Alexei V. Nickitenko,[§] Sergei Trakhanov,[||,⊥] and Florante A. Quiocho*[,§,||]

Howard Hughes Medical Institute and Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

Received September 5, 1995; Revised Manuscript Received October 30, 1995[⊗]

ABSTRACT: The family of about 50 periplasmic binding proteins, which exhibit diverse specificity (e.g., carbohydrates, amino acids, dipeptides, oligopeptides, oxyanions, metals, and vitamins) and range in size from 20 to 58 kDa, is a gold mine for an atomic-level investigation of structure and molecular recognition. These proteins serve as initial receptors for active transport systems or permeases. About six of these proteins, including the dipeptide-binding protein (DppA), are also primary receptors for chemotaxis. The structure of the unbound form of DppA (M_r = 57 400) has been determined and refined to an *R*-factor of 0.169 to 2 Å resolution. DppA consists of two distinct domains (I and II) connected by two "hinge" segments which form part of the base of the wide groove between the two domains. The relative orientation of the two domains gives the protein a pearlike shape, with domain I and domain II forming the larger and smaller apical ends, respectively. From the tip to the rounded bottom measures about 85 Å, and the widest diameter is about 60 Å. Domain I, which consists of two integrated subdomains, is folded from two separate polypeptide segments from the amino- and carboxyl-terminal ends. The more compact domain II is formed from the intervening segment. Comparison of the dipeptide-binding protein structure with that of the bound form of the similar oligopeptide-binding protein [Tame, J. R. H., Murshudov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F., & Wilkinson, A. J. (1994) *Science 264,* 1578−1581] reveals the major features that differentiate the ligand specificity of the two proteins and describe the large hinge bending (about 55°) between the two domains.

The dipeptide-binding protein (DppA)[1] and oligopeptide-binding protein (OppA) encoded by *dppA* and *oppA*, respectively (Higgins & Gibson, 1986; Abouhamad et al., 1991), are two of the largest members of the family of about 50 bacterial periplasmic binding proteins [reviewed in Furlong (1987) and Tam and Saier (1993)]. Both serve as initial receptors for high-affinity active transport or *di*pep*t*ide or *o*ligo*p*eptide *p*ermease, delivering peptides to corresponding cytoplasmic membrane components for actual translocation across the membrane. Moreover, as the dipeptide-binding protein also acts as a primary chemoreceptor (Manson et al., 1986), its interaction with the membrane components for dipeptide chemotaxis initiates flagella motion. The 507-residue amino acid sequence of *Escherichia coli* DppA (Olson et al., 1991) is shorter by 10 residues with respect to the *Salmonella typhimurium* OppA sequence (Hiles et al., 1987). A sequence comparison shows only 24% identity.

Structural features of oligopeptides essential for uptake by *E. coli* include a protonated primary ammonium group, L-stereochemistry, and a modified terminal carboxylate (Payne & Gilvarg, 1978). Similar features were also observed in ligand-binding studies of purified peptide-binding protein (Guyer et al., 1986). The nature of the side chains makes little or no contribution to ligand uptake. However, in spite of the lack of side-chain specificity, the peptides are bound with high affinity (K_d values in the micromolar range). Similar ligand affinity is observed for the other binding proteins (Furlong, 1987).

Crystallization of OppA of *S. typhimurium* and DppA of *E. coli* has been previously reported (Tolley et al., 1988; Dunten et al., 1993). The structure of OppA with bound oligopeptide has recently been determined (Tame et al., 1994). Ten other binding protein structures have been previously elucidated including those with specificity for arabinose, glucose/galactose, maltodextrin (linear and cyclic), ribose, sulfate, phosphate, leucine/isoleucine/valine, leucine, lysine/ornithine/arginine, and histidine [Quiocho & Vyas, 1984; Vyas et al., 1988, 1994; Pflugrath & Quiocho, 1988; Sack et al., 1989a,b; Luecke & Quiocho, 1990; Spurlino et al., 1991; Kang et al., 1991; Mowbray & Cole, 1992; Sharff et al., 1993; Oh et al., 1993; Zou et al., 1993; Yao et al., 1994; S. Trakhanov and F. A. Quiocho, unpublished data; for reviews see Quiocho (1990, 1991)].

Here we report the crystallization of DppA of *E. coli* in a form that shows superior diffraction quality and lower symmetry than previously reported and that allows the determination of the structure at 2 Å resolution.

MATERIALS AND METHODS

Protein Purification and Crystallization. DppA, along with other binding proteins (specificity for histidine, leucine/isoleucine/valine, and leucine), was purified according to a

[†] This work was supported in part by grants from the NIH and the Welch Foundation. F.A.Q. is an Investigator of the Howard Hughes Medical Institute.

[‡] Atomic coordinates have been deposited in the Brookhaven Protein Data Bank under entry code 1DPE.

* To whom correspondence may be addressed.

[§] Howard Hughes Medical Institute.

[||] Baylor College of Medicine.

[⊥] Present address: Gladstone Institute of Cardiovascular Disease, University of California at San Francisco, San Francisco, CA.

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1995.

[1] Abbreviations: DppA, dipeptide permease or binding protein; OppA, oligopeptide-binding protein; MBP, maltodextrin-binding protein; rms, root mean square.

0006-2960/95/0434-16585$09.00/0

protocol developed by one of us (S. Trakhanov) while at the M. M. Shemyakin Institute of Bioorganic Chemistry in Moscow [unpublished data; see also Trakhanov et al. (1989)]. Further purification of the DppA to homogeneity by hydrophobic interaction chromatography in our laboratory as described as follows was important for crystallization. A protein solution (typically 20−50 mg of protein in 2.0 M ammonium sulfate) was loaded onto a Butyl-Toyo Pearl (Toyo Soda, Japan) column (1.5 × 25 cm) which had been equilibrated with 2 M ammonium sulfate and 50 mM Tris-HCl, pH 8.0. The protein was eluted with a reverse gradient of ammonium sulfate (2−1 M) in the same buffer. Fractions of the single peak containing DppA were pooled and concentrated using an Amicon concentrator fitted with a UM-30 membrane to about 15 mg/mL. Using diaflo filtration, the excess ammonium sulfate was removed and the protein was equilibrated with 10 mM sodium cacodylate buffer, pH 7.

Microdialysis and hanging drop vapor diffusion methods were used in the crystallization at 4 °C. The microdialysis cell was of similar construction as that described previously (Trakhanov, 1989), although further miniaturization allowed the use of 1 mL of outer solution and 20−25 μL of protein solution in the dialysis chamber.

Crystallization trials using the 50 solutions of Crystal Screen number 1 (Hampton Research, Inc.) failed to yield crystals, although numbers 4, 15, 18, and 32 showed signs of granular precipitate or possibly microcrystals. The precipitant in number 4 and number 32 is ammonium sulfate, number 15 is a combination of ammonium sulfate and PEG 4000, and number 18 is PEG 8000. Because excellent diffracting crystals of two other periplasmic binding proteins have been obtained in our laboratory by the addition of Cd^{2+} (Yao et al., 1994; Trakhanov & Quiocho, 1995), the four probes, along with $CdSO_4$, were used in the second round of crystal screening, varying the ammonium sulfate or PEG concentration and pH of the media in a standard matrix survey. Small crystals were obtained using ammonium sulfate as precipitant. Following refinement of the crystallization condition, suitable crystals were obtained from a drop of a mixture of 5 μL of the stock DppA solution and an equal amount of the precipitant solution of 52−55% ammonium sulfate, 0.2 M $CdSO_4$, and 50 mM Bis-Tris-propane-HCl, pH 8.5, which was suspended over a well containing 1 mL of the precipitant solution. The prismatic crystals grow to a size of 0.15 × 0.15 × 0.35 mm within 4 weeks. Approximately two times larger prismatic crystals (Figure 1) were further obtained by the microseeding technique in hanging drops. The crystals of DppA have the space group $P2_12_12_1$ with unit cell dimensions $a = 118.3$ Å, $b = 79.8$ Å, $c = 63.0$ Å, and one molecule in the asymmetric unit based on a V_m of 2.6 Å^3/Da.

Intensity data for the native and three derivative crystals were collected at 4 °C on an SDMS dual multiwire detector system mounted on a Rigaku RU200 X-ray generator (fine focused Cu Kα) operated at 110 mA and 40 kV and equipped with a graphite monochromator. The data were processed by Chris Nielsen's program provided with the detector. Native data were also collected from a crystal frozen at −130 °C using a synchrotron source ($\lambda = 1.073$ Å) at the NSLS Howard Hughes Medical Institute station (X4A) using Fuji imaging plates. The program DENZO (Otwinowski, 1993) was used to index the images, integrate the intensities, and scale and merge the data. The statistics for all data sets are shown in Table 1.

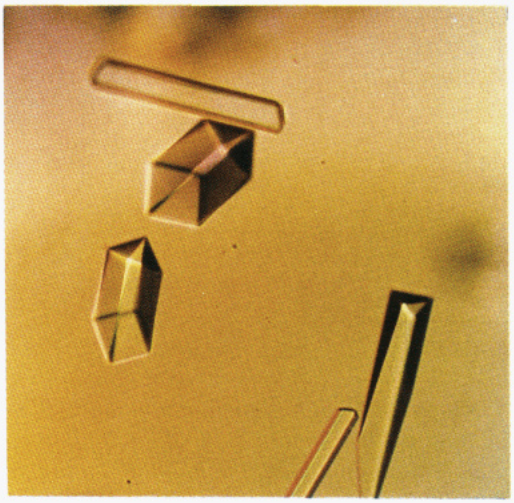

FIGURE 1: Photograph of crystals of DppA at 45× magnification.

Table 1: Summary of X-ray Structural Analysis

(a) Data Collection Statistics

crystal	source[a]	resolution (Å)	unique[b]	R_{merge}[c]	complete (%)
native	SDMS	2.3	23 722	9.8	88.2
K_2PtCl_4		3.0	10 232	8.4	84.5
PCMBS		3.0	7 739	9.5	63.9
$KAuCl_4$		3.0	11 050	9.5	91.2
native	NSLS X4A	2.0	38 709	6.7	95.3

(b) Phasing Statistics

derivative	resolution (Å)	no. of sites	R_{iso}[d]	R_{Cullis}[e]	phasing power[f]
K_2PtCl_4	10−4	7	14.6	0.84, 0.68	1.5, 1.3
PCMBS	10−4	1	10.6	0.93, 0.87	0.8, 0.6
$KAuCl_4$	10−4	1	12.5	0.95, 0.86	1.0, 0.8
⟨FOM⟩[g]	0.38, 0.59				

(c) Refinement Statistics

resolution (Å)	R-factor	R-free[h]	no. of reflections	bond (Å)[i]	angle (Å)[j]
6.0−2.0 ($F > 1\sigma$)	0.169	0.221	36 597	0.014	0.038

[a] SDMS, San Diego Multiwire System; NSLS X4A, National Synchrotron Light Source beamline X4A ($\lambda = 1.073$ Å). [b] Number of unique reflections with $I > 1\sigma$. [c] $R_{merge} = \sum(|I_i - \langle I\rangle|)/\sum|I|$. [d] $R_{iso} = \sum||F_{PH}| - |F_P||/\sum|F_P|$. [e] $R_{Cullis} = \sum||F_{PH} - F_P| - |F_H||/\sum|F_{PH} - F_P|$. [f] Phasing power = $|F_H|$/residual lack of closure error. [g] Mean figure of merit, pairs of values of R_{Cullis}, phasing power, and FOM are for the acentric and centric reflections, respectively. [h] 5% of reflections was omitted to calculate R-free. [i] rms deviation from ideal bond distance. [j] rms deviation from ideal angle related distance.

Heavy Atom Replacement. The multiple isomorphous replacement technique was used to obtain the initial electron density map. A single site for mercury (*p*-(chloromercuri)-benzoate) and for gold ($KAuCl_4$) derivatives was located from isomorphous difference Patterson maps. The relative origins were established by a cross difference Patterson map with coefficients $(F_{PH2} - F_{PH1})^2$. Seven platinum (K_2PtCl_4) sites were located from the isomorphous difference Fourier using the mercury−gold-derived MIR phases. Heavy atom parameters were refined and MIR phases at 4 Å resolution computed using the CCP4 (Collaborative Computational Project, November 4, 1994) implementation of MLPHARE (Otwinowski, 1991). As indicated in Table 1, the search for suitable derivatives resulted in only one moderately good platinum derivative. This derivative, combined with the

other two, gave acentric and centric figure of merits of 0.38, and 0.59, respectively. These initial MIR phases were improved by automated density modification and histogram matching using the DM program (CCP4, Collaborative Computational Project, November 4, 1994). The electron density map calculated with these phases was encouraging in that it showed a molecular boundary indicative of a two-domain structure expected of DppA. Moreover, we also noted that, relative to the OppA structure, the two domains are farther apart. This map showed evidence of several α-helices and β-sheet strands in both domains.

Molecular Replacement. In a parallel effort, molecular replacement was undertaken with the OppA structure (Tame et al., 1994) as the search model. As crystals of DppA were obtained in the absence of ligands, it was anticipated that the orientation of the two distinct domains would be different from that seen in the OppA structure. Indeed, a rotation search using the refined structure of OppA as a model was unsuccessful. Consequently, the model was divided into two parts (domain I, which consist of two subdomains, and domain II), and each was used independently as a model in the rotation search using the diffraction data between 8 and 4 Å with $F > 3\sigma(F)$. Using the suite of programs in X-PLOR (Brünger, 1992), the rotation search with domain I as the search model yielded, following Patterson correlation refinement, the strongest peak ($\theta_1 = 149°$, $\theta_2 = 48°$ $\theta_3 = 36.9°$). However, translation search, using the programs X-PLOR and AMORE (Navaza, 1994), gave no solution. Using domain II as the search model in the rotation search proved unsuccessful.

The 10 largest peaks in the X-PLOR rotation search with domain I as the model described above were further tried in a translation search using the 4 Å MIR phases and model phases to produce electron density maps:

$$T(\mathbf{t}) = \int_v \varrho_1(\mathbf{u},\mathbf{t})\varrho_2(\mathbf{u})\, d\mathbf{u}$$

where $\rho_1(\mathbf{u},\mathbf{t})$ and $\rho_2(\mathbf{u})$ are calculated from model and observed (MIR) electron density, respectively. The program TRPACK in the program package BLANC (Vagin, 1989) was used to calculate the $T(\mathbf{t})$ function in reciprocal space using all symmetry operators of the space group. Only the highest rotation peak indicated above yielded one very strong translation peak (9σ peak above the mean) at $x = 0.300$ Å, $y = 0.250$ Å, and $z = 0.080$ Å. The replacement solution was encouraging in that the domain I model fitted well to the solvent-flattened 4 Å MIR electron density. This was especially apparent for many of the elements of the secondary structure. Moreover, when the fitted domain I was overlapped with the identical domain of the entire model of OppA, it was clear that domain II is considerably displaced from the electron density map assigned to this domain. The placement of the domain II model to the electron density map was accomplished by rotating about a line connecting Cα 270−Cα 486 of 54° and a small twist angle in order to superimpose three β-sheet strands of OppA (residues 270−276, 411−418, and 480−485) to the density corresponding to these strands. On the basis of this fit, prominent features of the density could be easily correlated with the model, including the helices from residues 444 to 455 and 459 to 476 (OppA residue numbers). The two disulfide bonds, between Cys6 and Cys234 and between Cys422 and Cys435, in the DppA sequence clearly superimposed on two strong peaks of the density map.

Density Fitting and Refinement. The main-chain atoms of the composite model of domains I and II of OppA described above served as a guide in electron density fitting. The locations of the two disulfide bonds provided the starting points for fitting. In the first step, the MIR (10−4 Å resolution) modified as described above was used. A partial DppA model of 365 residues was obtained after iterative cycles of refinement using PROLSQ (Hendrickson, 1985; Collaborative Computational Project, Number 4, 1994) and adjustment in electron density maps computed from combined MIR phases (10−4 Å resolution range) and partial model phases (6−2.3 Å resolution range) using the multiwire intensity data. A complete model was initially built after iterative cycles of partial model refinement and adjustment in the $(2mF_o - DF_c)\exp(\alpha_c)$ electron density map calculated by the program SIGMAA (Read, 1986) at 2.0 Å resolution using synchrotron native data. X-PLOR refinement was only used initially to refine with simulated annealing the completed model. Thereafter, the refinement to completion made use of PROLSQ. CHAIN (Sack, 1988) was used in the electron density fitting, modeling, and examining of models.

RESULTS

Structure Determination. The addition of $CdSO_4$ to the crystallization buffer was effective in promoting the formation of crystals of DppA that diffract to at least 2.3 Å using a conventional source and to 2 Å at a synchrotron facility. Structure refinement against the 36 597 reflections obtained at the synchrotron facility [95% complete data with $F > \sigma(F)$], in the range of 6−2 Å resolution, converged to an R-factor of 0.169 and R-free of 0.221 (omitting 5% of the reflections in the R-free calculation). The geometry of the final model is good with root-mean-square deviations of 0.014 and 0.038 Å from ideal bond lengths and angle distances, respectively. A representative electron density is shown in Figure 2. The model consists of 4540 atoms, including 488 water molecules and 5 cadmium ions. The average temperature factors for the entire protein structure, main-chain atoms, side-chain atoms, and water molecules are 16.2 Å^2, 15.2 Å^2, 17.3 Å^2, and 29.8 Å^2, respectively. Main-chain torsion angles for 91.4% of the residues are within the favored regions of the Ramachandran plot; no residues are in the disallowed regions. The peptide bond between residues Lys274 and Pro275 has a *cis* conformation. Two residues, Thr20 and Leu419, in the original sequence (Olson et al., 1991) were changed to Ile and Glu, respectively. These changes were dictated by unambiguous density of the Ile and Glu in all types of maps examined during structure refinement. Moreover, three water molecules make good hydrogen bonds with the Oε1 and Oε2 atoms of Glu419 (Figure 2).

DppA Structure. As shown in Figure 3, the structure consists of two distinct domains clearly separated by a cleft. (This figure, along with the information contained in Table 2, helps to follow the description below.) The relative orientation of the two domains gives rise to a pear-shaped protein, with domain I, which is composed of two integrated subdomains (a and b), forming the larger apical end and the smaller domain II assuming the other apical end. From the tip to the rounded bottom measures about 85 Å, and the

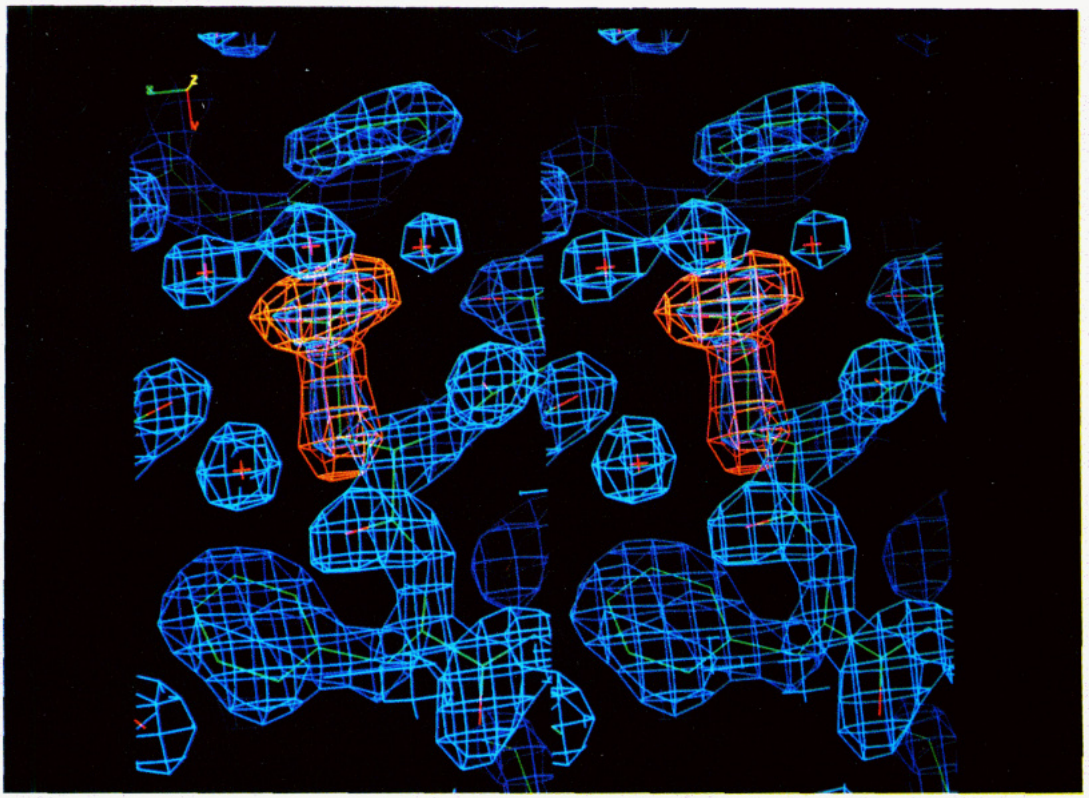

FIGURE 2: Stereoview of the electron density map and superimposed refined model of the area which includes Glu419. $2F_o - F_c$ (blue) and $F_o - F_c$ (orange) densities were calculated using structure factors obtained without the coordinates of the Glu side chain. Note the three water molecules within hydrogen-bonding distance to the carboxylate of Glu419. Glu419 in the original DppA sequence (Olson et al., 1991) is a Leu residue.

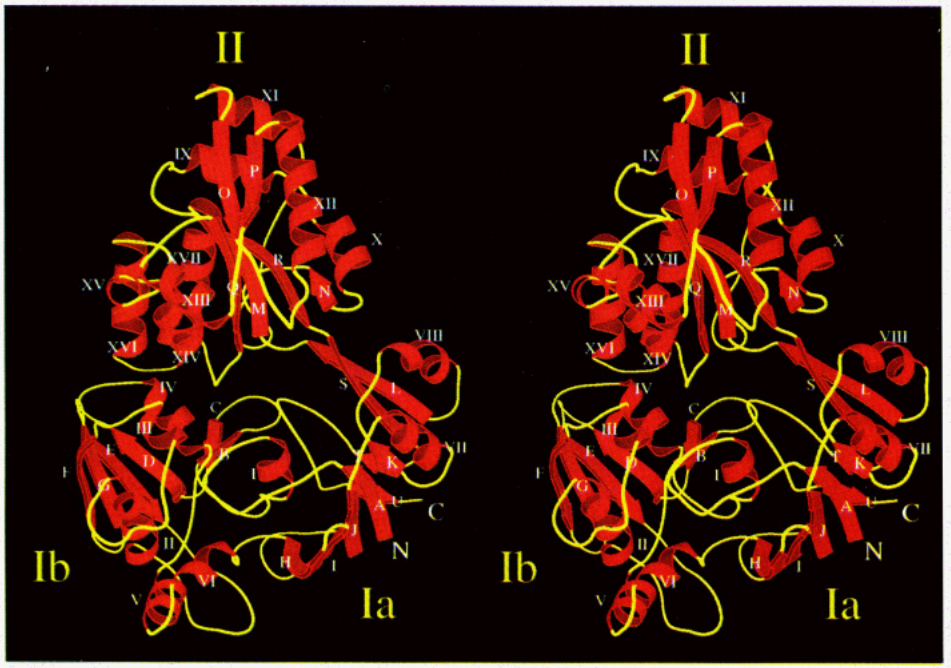

FIGURE 3: Schematic diagram [drawn with MOLSCRIPT (Kraulis, 1991)] of the DppA structure highlighting α-helices (Roman numerals) and β-strands (alphabets) and identifying subdomains Ia and Ib of domain I, domain II, the amino-terminal end (N), and the carboxyl-terminal end (C).

widest diameter is about 60 Å in diameter. Domain I is folded from two separate polypeptide segments from the amino- (residues 1−262) and carboxyl- (residues 479−507) terminal ends. The intervening segment (residues 263−478) comprises domain II. Whereas three segments (1−32, 183−262, and 479−507) make up subdomain Ia, only one segment (33−182) constitutes subdomain Ib.

The structure begins with strand A of subdomain Ia, and the first helix is juxtaposed between subdomains Ia and Ib. The chain continues to subdomain Ib, being composed of strands B−F and helices II−VI. Returning to domain Ia, the chain folds into four strands (H, I, J, and K), two helices (VII and VIII), and another strand (L). The chain goes on to form domain II, which is made up of a six-stranded β-sheet

Table 2: Secondary Structural Feature in the Dipeptide-Binding Protein

α-helices			β-strands		
designator	residue segment	domain location	designator	residue segment	domain location
I	22–27	Ia-b	A	2–7	Ia
II	87–98	Ib	B	37–39	Ib
III	113–117	Ib	C	46–48	Ib
IV	144–151	Ib	D	51–56	Ib
V	160–168	Ib	E	62–67	Ib
VI	172–177	Ib	F	124–129	Ib
VII	222–230	Ia	G	134–139	Ib
VIII	242–250	Ia	H	185–191	Ib
IX	279–287	II	I	195–200	Ia
X	291–299	II	J	213–218	Ia
XI	328–337	II	K	236–237	Ia
XII	361–375	II	L	254–260	Ia
XIII	386–395	II	M	263–269	II
XIV	412–420	II	N	303–305	II
XV	422–427	II	O	344–349	II
XVI	437–448	II	P	377–382	II
XVII	452–469	II	R	472–478	II
			S	479–485	Ia
			T	488–489	Ia
			U	505–506	Ia

and a total of nine helices. After strand R, the C-terminal peptide segment returns to subdomain I for strand S and the final big loop, which contains a short two-stranded (TU) sheet. Two other extraordinarily long loops connect elements of the secondary structure—strand E to helix II and strand N to helix XI (Figure 3).

DppA has essentially a β/α folding motif with β-strands and α-helices constituting about 21% and 32% of the entire structure, respectively. The arrangement of the secondary structure elements or "supersecondary structure" of domain II is somewhat different from those of the two subdomains. Domain II has a twisted central sheet sandwiched on both sides by helices. The topology of the sheet is NRMQOP with the NRMQ and OP strands in antiparallel and parallel arrangements, respectively. Of the nine helices, five (XIII, XIV, XV, XVI, XVII) are on one side of the plane of the sheet and two (X, XII) are on the opposite side. The other two helices (IX and XI) lie at the tip of the pear-shaped molecule perpendicular to the small apical end.

Although each of the two subdomains contains an extensive β-sheet, helices are confined on one side of the plane of the sheet. The two helices in subdomain Ia are on the surface, but those in Ib are found largely in the interior of domain I. Subdomain Ib is composed of two β-sheets, one with two short antiparallel strands (BC) and the other with four antiparallel strands with a DEGF topology. The larger sheet is amphiphatic, as one side is exposed to the solvent and the other is in the interior of domain I. In the sheet of subdomain Ia, with HIJAKSL topology, the middle set of three parallel strands is flanked on both sides by two antiparallel strands. The final large loop which forms a short two-stranded (TU) sheet abuts the side of the large β-sheet of subdomain Ia opposite to that occupied by helices VII and VIII.

DISCUSSION

Despite only 23% identity between *E. coli* DppA and *S. typhimurium* OppA estimated from the structure-based alignment of the amino acid sequences (Figure 4), both proteins have similar tertiary structure (Figure 5A,B). Moreover, they have several features in common with the structures of the ten smaller periplasmic receptors (M_r from 26 000 to 40 000) and with specificity that include carbohydrates [arabinose, glucose/galactose, maltodextrin (linear and cyclic), and ribose], tetrahedral oxyanions (sulfate and phosphate), and amino acids (leucine/isoleucine/valine, leucine, lysine/ornithine/arginine, and histidine). Including the DppA, nine binding protein structures have been determined in our laboratory.

Periplasmic receptor structures are basically ellipsoidal, being composed of two distinct domains (I and II) separated by a deep cleft wherein the ligand binds. In proteins with oligomeric substrates (e.g., linear and cyclic maltodextrins, dipeptides, and oligopeptides), the binding site is more like a groove. The two domains rotate with respect to each other about a hinge connecting the two domains, thus enabling the formation of an "open form" with a wide open and solvent-accessible cleft and a "closed form" with no access to the cleft. The ligand binds to the open form. The structures of the closed form with bound ligand, more of which have been determined than the open form, show that both domains interact with and engulf the ligand. We have determined the ligand-free, open form of the dipeptide-binding protein. These common structural features are consistent with the function of this family of proteins [Yao et al., 1994; for reviews see Quiocho (1990, 1991)].

The observations that the ligand-binding site of the periplasmic receptors is located between the two separate domains and that access to and from the cleft is modulated by a rigid-body domain rotation have weighed heavily in our decision to continue to describe the binding protein structures as consisting of two distinctly separate domains. This designation simplifies comparison with the other binding protein structures and the description of the rigid-body hinge-bending motion between the two domains. We had previously observed that the larger (domain II or C-domain) of the two domains of the maltodextrin-binding protein (MBP) (M_r = 40 000), the next largest of the known binding protein structures, was made up of two close-packed subdomains, of which one is significantly smaller than the other (Spurlino et al., 1991). In contrast, the two subdomains of DppA and OppA have a roughly similar large size (Figures 3 and 5A,B). [Tame et al. (1994) have considered the OppA as composed of three domains, I, II and III, which are equivalent to subdomain Ia, subdomain Ib, and domain II, respectively, of DppA (Figure 3)]. The smaller subdomain in MBP serves to extend the long groove between the two domains in order to accommodate linear oligosaccharides from maltose to maltotetraose or cyclic dextrins (Spurlino et al., 1991; Sharff et al., 1993). Subdomains do not exist in the domains of the structures of the other nine binding proteins with M_r of about 33 000 or less.

Despite a much larger size of the DppA and OppA, only two segments connect the two domains (Figure 3). This feature is also shared with the structures of the phosphate-binding protein (Luecke & Quiocho, 1990), ribose-binding protein (Mowbray & Cole, 1992), lysine/ornithine/arginine-binding protein (Oh et al., 1993), and histidine-binding protein (Yao et al., 1994). The mass of the phosphate-binding protein is about 33 kDa while those of the last three proteins are the smallest (about 26 kDa). Apparently, there

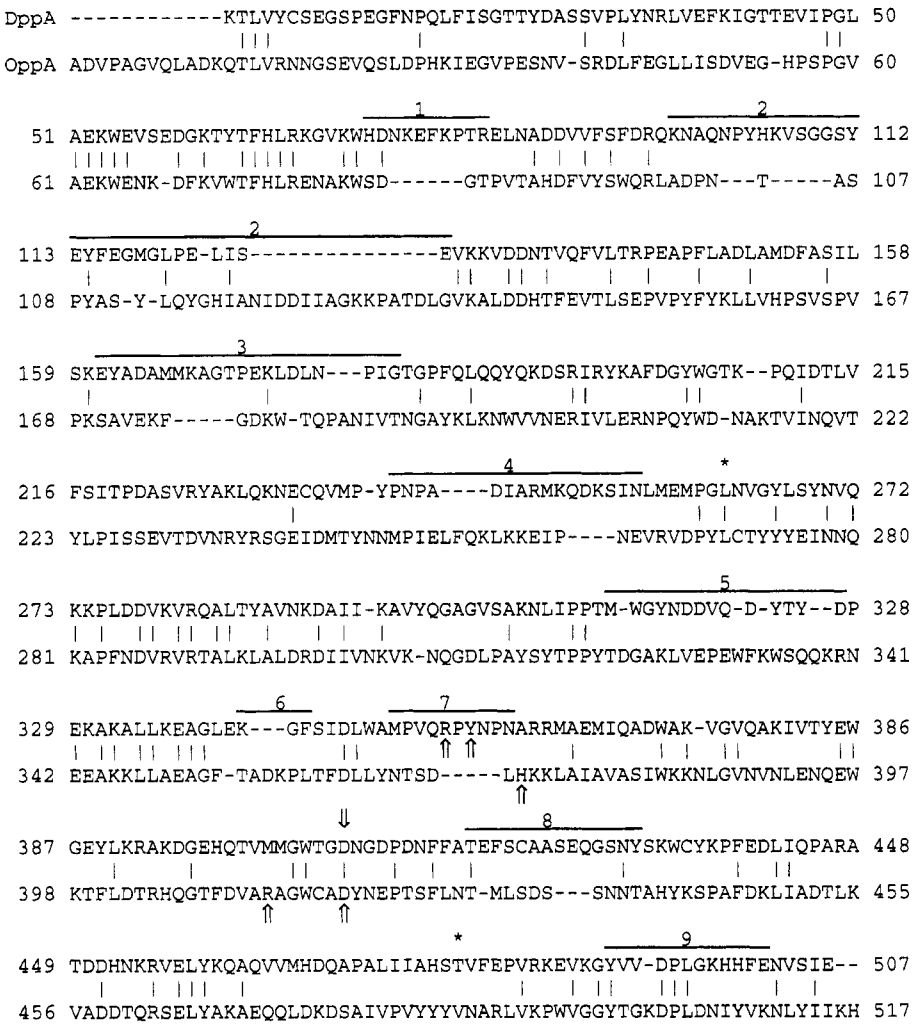

FIGURE 4: Sequence alignment of DppA and OppA based on the superimposed structures shown in Figure 5. The vertical lines indicate identical amino acids. The lines with numbers above the sequences represent segments which have considerable differences in the structures (see Table 3). Arrows point to the residues in the ligand-binding sites. The hinge residues are identified with asterisks.

is no correlation between protein size and the number of segments connecting the domains.

The manner in which the polypeptide chain is apportioned between the two domains, which also reflects the number of interdomain connecting segments, formed a new and more simple basis of classifying binding protein structures into two groups (Yao et al., 1994). Half of the known binding protein structures, those with two crossovers between the two domains (e.g., DppA and OppA; see above), belong to group II. In this group, segments from the amino- and carboxyl-terminal ends make up domain I while the segment intervening the two terminal segments constitutes domain II.

An interdomain connection of three segments is seen in the structures of the other six binding proteins (mass between about 32 and 40 kDa), belonging to group I, with specificity for arabinose, glucose/galactose, maltodextrins, sulfate, Leu/Ile/Val, and Leu. This means that each domain is folded of two separate segments from the first and the last halves of the polypeptide chain. Despite being folded of two separate segments, both domains have similar supersecondary structure which consists of a β-sheet with two or more helices on both sides of the plane of the sheet. Similar arrangement of the elements of the secondary structure exists for the two domains of binding proteins belonging to group II with the exception of DppA and OppA. Only domain II of DppA and OppA exhibits a similar supersecondary structure (Figure 3). The two subdomains of two peptide-binding proteins lack two or more helices on one face of the β-sheet to conform with the more common domain supersecondary structure in the binding proteins.

Binding protein structures has also been previously distinguished by the two most prevalent types of transitions or crossovers from one domain to the other (Spurlino et al., 1991). One prevalent type, including the pair in DppA and OppA, is one in which the crossover connects a β-strand in one domain to another strand in the other domain. Indeed, in DppA and OppA, this type of connection could lead to the formation of one long strand that penetrates both domains and creates a two-stranded sheet for the two connecting segments (e.g., see Figure 3). The other frequently observed type is one linking an α-helix of one domain to a β-sheet strand of the other domain. Very few crossovers linking helix to helix are observed. This may be related to the fact that the domains with β/α motif nearly always begin with a β-sheet strand.

Apart from the specific nature of folding associated with each periplasmic receptor, the structures of the receptors as a whole pose intriguing broad questions about folding. For instance, what determines or dictates the number of segments

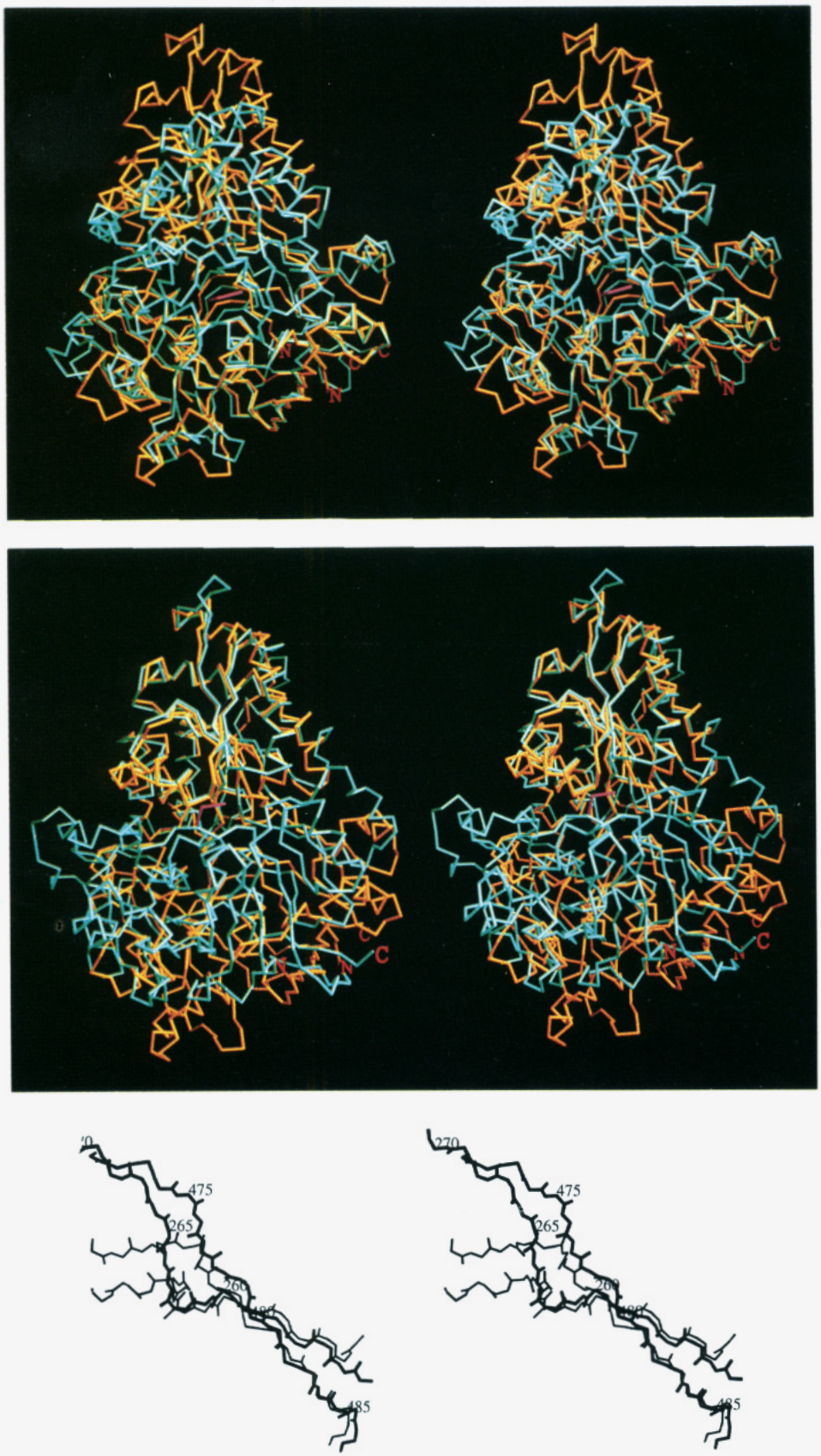

FIGURE 5: Stereoview of the superimposed Cα backbones of DppA (yellow) and OppA (cyan) with the bound tripeptide (magenta). (A) (top figure) and (B) (middle figure) are based on superpositioning domains I and II, respectively, of both structures. The orientation is identical to that shown in Figure 3. (C, bottom figure) Superimposed connecting segments or hinge between the two domains of DppA (bonds in heavy lines) and OppA (bonds in light lines).

for folding each of the two domains? Contrary to what one might suspect, the size of the protein does not appear to be a determinant. Group II structures include proteins with the largest, medium, and smallest size. The size of the ligands

Table 3: Segments of Greatest Difference between DppA and OppA

identification	segment DppA	segment OppA	comments with respect to DppA structure (refer to Figures 3−5)
1	74−83	83−86	the segment, which contains a six-residue insertion, connects strand E to helix II
2	98−126	101−135	16 residues are deleted and 10 are inserted; the segment begins at the last residue of helix II, includes helix III, and terminates at the third residue of strand F
3	161−181	170−187	the segment, which has a six- and a three-residue insertion and deletion, respectively, includes V starting from the second residue, VI, and halfway of the segment between VI and H.
4	240−255	248−263	the segment, with a four-residue insertion and deletion, encompasses half of the loop preceding VIII, VIII, and the second residue of L
5	314−327	322−340	the segment, with a five-residue deletion, connects N to XI
6	342−344	354−359	the segment has a three-residue deletion and corresponds to the loop joining XI and O
7	351−360	366−370	this segment, located in the peptide-binding site, contains a five-residue insertion which includes R355 and Y357; it connects O to XII
8	418−431	429−438	the segment, which has a four-residue insertion, starts at the third residue from the carboxy-terminal end of XIV, includes XV, and terminates halfway of the segment following XV
9	491−502	498−510	C-terminal segment precedes S

(oligomers versus monomers) does not appear also to be of any consequence. The maltodextrin-binding protein belongs to group I whereas DppA and OppA belong to group II. The ligands of the rest of binding proteins (M_r from 33 000 to 26 000) with known structures are monomers or small oxyanions. A further important question relates to the mechanism for the formation of the different types of crossovers between the two domains. Anent to this, what is the mechanism for signaling initiation of these crossovers? What dictates the topological arrangements of the secondary structure, which differ between the structures? With some more structures of the receptors in the offing, the collection of structures of the family of proteins provides an excellent resource to delve into these and other questions.

In order to compare the DppA and OppA structures, identical domains were superimposed independently by least-squares fitting of the well-matched α-carbons of 188 and 151 residues of domain I (Figure 5A) and domain II (Figure 5B), respectively. The rms deviations for the superimposed residues are 1.19 and 1.15 Å, and the maximum displacements are 3.19 and 2.49 Å for domains I and II, respectively.

From the superimposed domains, we observed at least nine peptide segments whose conformations differ the greatest between the two proteins (Table 3 and Figures 4 and 5A,B). With the exception of one segment located in the ligand-binding site (segment 7), all eight occupy the surface of the proteins (Figure 3). Insertions or deletions are mainly responsible for these differences (Table 3 and Figure 4). Superpositioning of the two peptide-binding protein structures has further uncovered features relevant to functions described as follows.

As expected, the superposed structures of the ligand-free DppA and ligand-bound OppA (Figure 5) clearly show a more open ligand-binding site cleft in DppA. Indeed, the more than 2 dozen ordered water molecules seen in the cleft are more than twice the water bound, along with the tripeptide or tetrapeptide, in the enclosed cleft of OppA (Tame et al., 1994). The presence of these water molecules provides, in part, a peptide sequence-independent mechanism for ligand binding.

The superpositioned structures also reveal the large hinge bending between the two domains. The hinge can be drawn between the α-carbons of residues 262 and 479 in the first and second crossover segments, respectively, in DppA and their counterparts (270 and 486) in OppA (Figure 5C). Using this hinge, the angle of rotation between the two domains in going from the OppA closed structure to the DppA open structure is 55.4°. This cleft opening is accompanied by a small twist of 7.7 Å.

Furthermore, the superposition of the DppA and OppA structures has shed light on likely common as well as unique features of ligand recognition. The nature of the side chains of the peptide ligands contributes little or nothing to the process of molecular recognition. Rather, it is the length of the peptide that dictates specificity. Indeed, the large size of the binding groove, combined with the abundance of water molecules in it, provides a sequence-independent mechanism for binding peptides (Tame et al., 1994). The favorable entropy contribution resulting from the displacement of these ordered water molecules upon ligand binding contributes to the formation of the complex. The structures of the complexes of OppA with tripeptide and tetrapeptide show a very similar orientation of the backbone common to both peptides (Tame et al., 1994). They also reveal three key juxtaposed charged residues which are involved in forming salt links with the primary ammonium and carboxylate termini of these peptides and, hence, are the major determinant in oligopeptide recognition. It is important to note that these residues are confined only to domain II [or domain III as identified by Tame et al. (1994)]. The α-ammonium groups of both peptides make an identical salt link with Asp419. In contrast, recognition of the terminal carboxylate group differs according to the length of the oligopeptides. The salt link with the carboxylate of the tripeptide is formed with Arg413 while that of the tetrapeptide is made with His371 which, relative to Arg413, is farther from Asp419. (The positions of Asp419, Arg413, and His371 in relation to the bound trilysine are shown in Figure 6A.) Lack of conservation of these two basic residues in DppA (Figures 3 and 6A,B) accounts for its inability to tightly bind peptides longer than dipeptides. On the other hand, Asp419 is fully conserved in DppA (Asp408) as are the backbone conformations of the segments (strand Q in DppA, Figure 3) deploying these residues as well as Arg413 of OppA and its counterpart (Met402) in DppA (Figure 6A). In addition, of particular significance is the presence of Arg355 in DppA at a distance closer to Asp408 than Arg413 is to Asp419 in OppA. Arg355 is within very close distance to the terminal carboxylate of a dipeptide, which is modeled in DppA (based on the overlap with OppA) in a conformation identical to the first two amino acids of the tripeptide bound to OppA (Figure 6B). Only a small rotation about the Cβ−Cγ bond

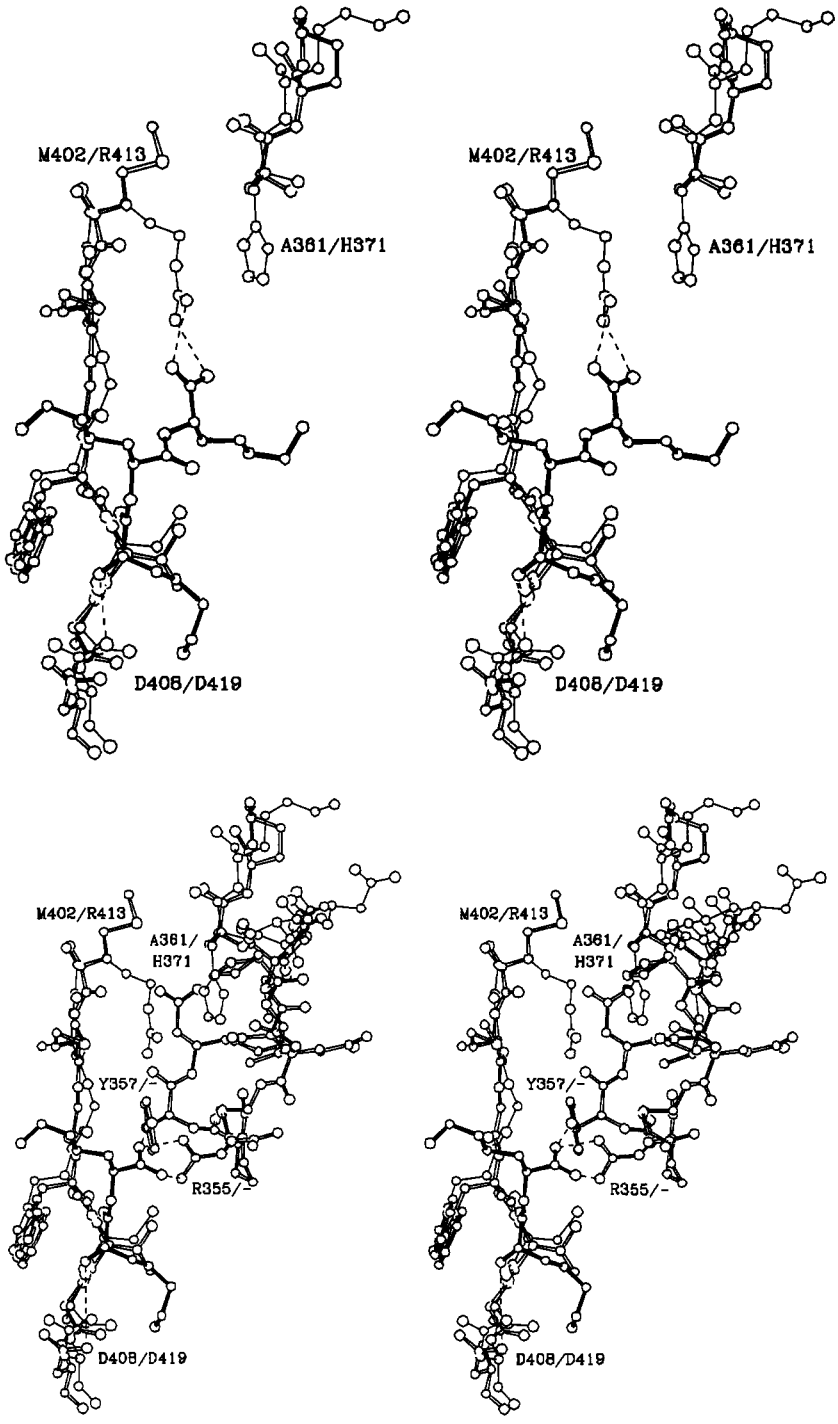

FIGURE 6: Stereo overlaps (based on Figure 5) of segments of the binding site of OppA containing key residues (bonds in solid lines) and DppA (bonds in double lines). The DppA residue is the first entry followed by that of Oppa. (A, top) Trilysine bound to OppA. Note the excellent superpositioning of the backbone of R413 to D419 of OppA with M402 to D408 of DppA and H371 of OppA with A361 of DppA. This segment in DppA corresponds almost to the entire strand Q (Figure 3). (B, bottom) Dilysine bound to DppA modeled exactly as the first two lysines of the trilysine bound to OppA shown in (A). Included in this figure is the segment of residues 352−362 of DppA, which is part of a loop between strand O and helix XII (Figure 3). This loop which contains an insertion (Figure 4) is much shorter and has a conformation significantly different from that in the OppA structure (Figure 5B). Arg355 is shown in two conformations, one as found in the refined structure and the other as obtained after rotating about the $C\beta-C\gamma$ bond in order for the guanidinium to form a salt link with the primary carboxylate of the dipeptide.

of Arg355 is required for the guanidinium group to form a good salt link with the terminal carboxylate of the modeled dipeptide (Figure 6B). Tyr357 is also within hydrogen-bonding distance to the modeled dipeptide (Figure 6B). Arg355 and Tyr357 are located in the middle of the loop (residues 351−360) linking strand O to helix XII (Figure 3). The conservation of Asp419 and its precise juxtapositioning with Arg355 provide evidence for the correctness of the modeled dipeptide bound to DppA (Figure 6B).

The fact that Arg355 and Tyr357 of DppA are not conserved in OppA (Figure 4) is compatible with the observation of extremely weak binding of dipeptide to OppA (Guyer et al., 1986). Moreover, the loop in DppA containing these two residues, which are part of a five-residue insertion

(residues 355 to 359) (Figure 4 and Table 3), is one (number 7) of the nine segments which have conformations substantially different from those in OppA (Figure 5A,B). The conformation of the loop makes the binding site in DppA more restrictive than in OppA (Figure 6B) and, hence, more exclusionary toward peptides longer than a dipeptide.

In marked contrast to DppA and OppA, the nature of the side chains of the amino acid substrates dictates the specificity of the receptors for leucine, lysine/ornithine/arginine, and histidine. Unlike several of the interactions between oligopeptide side chains and OppA which are mediated by water molecules, direct hydrogen-bonding or salt-linking interactions between protein groups and sides of the amino acid substrates characterize the complexes with the amino acid receptors (Sack et al., 1989a; Oh et al., 1993; Yao et al., 1994; S. Trakhanov and F. A. Quiocho, unpublished data).

Maltodextrin-binding protein is the other member of the periplasmic protein family which binds oligomeric substrates and whose structures, with and without bound ligands, have been previously determined (Spurlino et al., 1991; Sharff et al., 1992). Like OppA and DppA, which are unable to bind single amino acids, MBP does not bind glucose, the constituent of the oligosaccharide substrates. Further similarities have been gleamed from comparing the structures of DppA and OppA with those of MBP without sugar and with bound maltose, maltotriose, or maltotetraose (Spurlino et al., 1991; Sharff et al., 1992; J. C. Spurlino, L. Rodseth, and F. A. Quiocho, unpublished data). For instance, the binding site in MBP also consists of subsites, each consigned for binding a specific glucosyl unit of the maltotetraose or identical unit in shorter maltooligosaccharides. Much like OppA, the ability of MBP to recognize different size oligosaccharides (from maltose to maltotetraose) is achieved through specific interactions which are identical for the reducing glucosyl end but somewhat different for the nonreducing end.

A comparison of the sugar-free and sugar-bound MBP structures has also provided the first structural evidence for a hinge bending between the two domains (Spurlino et al., 1991; Sharff et al., 1992). With the addition of the DppA and OppA tertiary structures, this motion which modulates the accessibility to and from the ligand-binding cleft has now been observed in periplasmic receptors with widely differing size. Interestingly, the hinge-bending angle of about 55° observed for the DppA−OppA combination is similar to that seen in the lysine/ornithine/arginine-binding protein, the smallest of the structures (Oh et al., 1993). This domain rotation is the largest seen thus far among the binding proteins. Apparently, the extent of domain rotation, which could be influenced by crystal packing, is not dependent on protein size. The maltodextrin-binding protein, which has three interdomain connecting segments, exhibits a hinge-bending angle of about 35° (Spurlino et al., 1991; Sharff et al., 1992). Domain rotation is an important component of the function of many bilobate proteins, including enzymes [for a recent review, see Gerstain et al. (1994)]. This motion for a large number of these proteins is associated expressly with ligand or substrate binding. In addition to ligand binding, hinge bending in the periplasmic receptors fulfills a key role in signal transduction. It enables the formation of a ligand-stabilized closed structure which is presumably recognized, in preference to the open structure, by the appropriate membrane-bound components and, thus, initiates nutrient translocation or chemotaxis. Productive docking with the membrane components for transport or chemotaxis requires the participation of both domains of the protein (Spurlino et al., 1991; Sharff et al., 1992; Oh et al., 1993). These dual key roles (ligand binding and signal transduction) of domain rotation are most likely also manifested in the *pur*R and *lac*R of the *lac*I family of repressors, which coincidentally have structures and inducer-binding sites very similar to those of the periplasmic binding proteins (Schumacher et al., 1994; Friedman et al., 1995).

ACKNOWLEDGMENT

We thank William E. Meador for assistance in the intensity data collection in our laboratory. We also thank Nanhua Yao, David K. Wilson, and Dr. Kenneth A. Johnson of our laboratory and, especially, Dr. Craig Ogata of the HHMI group at Brookhaven, for assistance in data collection at the Howard Hughes Medical Institute Beam Line X4A at the Brookhaven National Synchrotron Light Source.

REFERENCES

Abouhamad, W. N., Manson, M., Gibson, M. M., & Higgins, C. F. (1991) *Microbiology 5*, 1035−1047.

Brünger, A. T. (1992) *X-PLOR Manual Version 3.1,* Yale University, New Haven, CT.

Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography, *Acta Crystallogr. D50*, 760−763.

Dunten, P. W., Harris, J. H., Feiz, V., & Mowbray, S. L. (1993) *J. Mol. Biol. 231*, 145−147.

Friedman, A. M., Fischmann, T. O., & Steitz, T. A. (1995) *Science 268*, 1721−1727.

Furlong, C. E. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaecher, M., & Umbarger, H. E., Eds.) pp 786−796, American Society for Microbiology, Washington, DC.

Gerstain, M., Lesk, A. M., & Chothia, C. (1994) *Biochemistry 33*, 6739−6749.

Guyer, C. A., Morgan, D. G., & Staros, J. V. (1986) *J. Bacteriol. 168*, 775−779.

Hendrickson, W. (1985) *Methods Enzymol. 115*, 252−270.

Higgins, C. F., & Gibson, M. M. (1986) *Methods Enzymol. 125*, 365−377.

Hiles, I. D., Gallagher, M. P., Jamieson, D. J., & Higgins, C. F. (1987) *J. Mol. Biol. 195*, 125−142.

Kang, C.-H., Shin, W.-C., Yamagata, Y., Gokcen, S., Ames, G. F.-L., & Kim, S.-H. (1991) *J. Biol. Chem. 266,* 23893−23899.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Luecke, H., & Quiocho, F. A. (1990) *Nature 347*, 402−406.

Manson, M. D., Blank, V., Brade, G., & Higgins, C. F. (1986) *Nature 321*, 253−256.

Mowbray, S. L., & Cole, L. B. (1992) *J. Mol. Biol. 225*, 155−175.

Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.

Oh, B., Pandit, J., Kang, C.-H., Nikaido, K., Gokcen, S., Ames, G. F.-L., & Kim, S.-H. (1993) *J. Biol. Chem. 268*, 11348−11355.

Olson, E. R., Dunyak, D. S., Jurss, L. M., & Poorman, R. A. (1991) *J. Bacteriol. 173*, 234−244.

Otwinowski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend* (Wolf, W., Evans, P. R., & Leslie, A. G. W., Eds.) pp 80−86, SERC, Daresbury Laboratory, Warrington, U.K.

Otwinowski, Z. (1993) in *Data Collection and Processing: Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaacs, N., & Bailey, S., Eds.) pp 56−62, SERC, Daresbury Laboratory, Warrington, U.K.

Payne, J. W., & Gilvarg, C. (1978) in *Bacterial Transport* (Rosen, B. P., Ed.) pp 325−383, Marcel Dekker, New York.

Pflugrath, J. W., & Quiocho, F. A. (1988) *J. Mol. Biol. 200*, 163−180.
Quiocho, F. A. (1990) *Philos. Trans. R. Soc. London 326*, 341−351.
Quiocho, F. A. (1991) *Curr. Opin. Struct. Biol. 1*, 922−933.
Quiocho, F. A., & Vyas, N. K. (1984) *Nature 310*, 381−386.
Read, R. J. (1986) *Acta Crystallogr. A42*, 140−149.
Sack, J. S. (1988) *J. Mol. Graphics 6*, 224−225.
Sack, J. S., Saper, M. A., & Quiocho, F. A. (1989a) *J. Mol. Biol. 206*, 171−191.
Sack, J. S., Trakhanov, S., Tsigannik, I. H., & Quiocho, F. A. (1989b) *J. Mol. Biol. 206*, 193−207.
Schumacher, M. A., Choi, K. Y., Zalkin, H., & Brennan, R. G. (1994) *Science 266*, 763−770.
Sharff, A. J., Rodseth, L. E., Spurlino, J. C., & Quiocho, F. A. (1992) *Biochemistry 31*, 10657−10663.
Sharff, A. J., Rodseth, L. E., Spurlino, J. C., & Quiocho, F. A. (1993) *Biochemistry 32*, 10553−10559.
Spurlino, J. C., Lu, G.-Y., & Quiocho, F. A. (1991) *J. Biol. Chem. 266*, 5202−5219.
Tam, R., & Saier, M. H., Jr. (1993) *Microbiol. Rev. 57*, 320−346.
Tame, J. R. H., Murshudov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F., & Wilkinson, A. J. (1994) *Science 264*, 1578−1581.
Tolley, S. P., Derewenda, Z., Hyde, S. C., Higgins, C. F., & Wilkinson, A. J. (1988) *J. Mol. Biol. 204*, 493−494.
Trakhanov, S. D. (1989) *Bioorg. Khim. 15*, 725−730.
Trakhanov, S., & Quiocho, F. A. (1995) *Protein Sci. 4*, 1914−1919.
Trakhanov, S. D., Chirgadze, N.-Y., & Yusifov, E. F. (1989) *J. Mol. Biol. 204*, 847−849.
Vagin, A. A. (1989) *News letter on protein crystallography*, Vol. 24, pp 24−26, Daresbury Laboratory, Warrington, U.K.
Vyas, N. K., Vyas, M. N., & Quiocho, F. A. (1988) *Science 242*, 1290−1295.
Vyas, N. K., Vyas, M. N., & Quiocho, F. A. (1991) *J. Biol. Chem. 266*, 5226−5237.
Vyas, M. N., Vyas, N. K., & Quiocho, F. A. (1994) *Biochemistry 33*, 4762−4768.
Wang, Z., Choudhary, A., Ledvina, P. S., & Quiocho, F. A. (1994) *J. Biol. Chem. 269*, 25091- 25094.
Yao, N., Trakhanov, S., & Quiocho, F. A. (1994) *Biochemistry 33*, 4769−4779.
Zou, J., Flocco, M. M., & Mowbray, S. L. (1993) *J. Mol. Biol. 233*, 739−752.

BI952098D