Structure of the *Escherichia coli* Response Regulator NarL[†,‡]

Igor Baikalov,[§] Imke Schröder,[||] Maria Kaczor-Grzeskowiak,[§] Kazimierz Grzeskowiak,[§] Robert P. Gunsalus,[§,||] and Richard E. Dickerson*[,§]

Molecular Biology Institute and Department of Microbiology and Molecular Genetics, University of California, Los Angeles, California 90095-1570

Received April 16, 1996; Revised Manuscript Received June 6, 1996[⊗]

ABSTRACT: The crystal structure analysis of the NarL protein provides a first look at interactions between receiver and effector domains of a full-length bacterial response regulator. The N-terminal receiver domain, with 131 amino acids, is folded into a 5-strand β sheet flanked by 5 α helices, as seen in CheY and in the N-terminal domain of NTRC. The C-terminal DNA-binding domain, with 62 amino acids, is a compact bundle of 4 α helices, of which the middle 2 form a helix−turn−helix motif closely related to that of *Drosophila paired* protein and other H-T-H DNA-binding proteins. The 2 domains are connected by an α helix of 10 amino acids and a 13-residue flexible tether that is not visible and presumably disordered in the X-ray structure. In this unphosphorylated form of NarL, the C-terminal domain is turned against the receiver domain in a manner that would preclude DNA binding. Activation of NarL via phosphorylation of Asp59 must involve transfer of information to the interdomain interface and either rotation or displacement of the DNA-binding C-terminal domain. Docking of a B-DNA duplex against the isolated C-terminal domain in the manner observed in *paired* protein and other H-T-H proteins suggests a stereochemical basis for DNA sequence preference: T-R-C-C-Y (high affinity) or T-R-C-T-N (low affinity), which is close to the experimentally observed consensus sequence: T-A-C-Y-N. The NarL structure is a model for other members of the FixJ or LuxR family of bacterial transcriptional activators, and possibly to the more distant OmpR and NtrC families as well.

Bacterial adaptation to a wide variety of environmental stimuli is accomplished through the coordinated regulation of many different signal processing pathways. A common mechanism of signal transduction, found in over 60 bacterial regulatory systems, has a 2-component structure: a sensor, which is an autophosphorylating histidine protein kinase, and a response regulator, which becomes activated by the phosphorylation (Bourret et al., 1991; Parkinson & Kofoid 1992; Pao & Saier 1995). A typical response regulator consists of receiver and effector domains joined by what this study indicates to be a flexible linker. While the overall organization of each of the response regulators is tailored to a specific effector function, their receiver domains are highly conserved, implying a ubiquitous mechanism of activation by Mg^{2+}-dependent protein phosphorylation.

One member of the receiver domain superfamily, bacterial chemotaxis protein CheY, has been studied extensively by X-ray crystallography (Stock et al., 1989, 1993; Volz & Matsumura, 1991; Bellsolell et al., 1994) and NMR spectroscopy (Moy et al., 1994; Lowry et al., 1994). More recently, a low-resolution NMR structure of another receiver domain, the amino-terminal domain of nitrogen regulatory protein NTRC, has been reported (Volkman et al., 1995). Although these studies provide a large amount of information on the significance of conserved residues, and conformational changes caused by Mg^{2+} binding, little is known about the active, phosphorylated state of the receiver domain, and virtually nothing is known about the propagation of activation to the effector domain of a typical response regulator. We present here the first structure analysis of a full-length response regulator possessing both receiver and effector domains: the 2.4 Å resolution crystal structure of the unphosphorylated form of NarL, a transcriptional regulator in the nitrate-signaling two-component system of *Escherichia coli*.

Among the variety of alternative respiratory substrates used by *Escherichia coli*, nitrate and formate are energetically the most favorable electron acceptor and donor under anaerobic conditions (Stewart, 1988; Gunsalus, 1992). Recent studies have identified an unusual two-component system involved in the nitrate-dependent regulation of many anaerobic electron transport and fermentative related genes (Gunsalus, 1992; Schröder et al., 1994; Stewart, 1994; Darwin & Stewart, 1995; Kaiser & Sawers, 1995). This system contains two sensor proteins, NarX and NarQ, which independently can detect both nitrate and nitrite in the environment of the cell. The response elements of this system also are 2-fold; the DNA-binding activity of the two transcriptional regulator proteins, NarL and NarP, can be modulated by either NarX or NarQ in response to nitrate and nitrite. The entire system is responsible for nitrate-dependent induction of expression of the genes for nitrate reductase (*narGHJI*), nitrite export (*narK*)(Kolisnikov & Gunsalus, 1992), and formate dehydrogenase (*fdnGHI*)-(Darwin & Stewart, 1995). In addition, it represses expression of genes for fumarate reductase (*frdABCD*)(Kalman & Gunsalus, 1989), TMAO/DMSO reductase (*dmsABC*)(Cotter

[†] Supported by National Institutes of Health Grants GM-31299 and AI-21678.

[‡] Both diffraction F_o data and final refined coordinates have been deposited with the Brookhaven Protein Data Bank, for immediate release, file name 1RNL.

* To whom correspondence should be addressed, at the Molecular Biology Institute.

[§] Molecular Biology Institute.

[||] Department of Microbiology and Molecular Genetics.

[⊗] Abstract published in *Advance ACS Abstracts,* August 1, 1996.

S0006-2960(96)00919-1 CCC: $12.00

Table 1: Diffraction Data And Phase Quality[a]

data set	native	Pt-3	Au-1	Hg-1
	Crystal Statistics			
resolution (Å)	2.6	2.4	3.0	3.0
no. of reflections				
measured	49071	60194	29652	35363
unique	8798	11188	5747	5777
$F_{obs} > 2\sigma$ (%)	96.9	97.3	95.4	96.0
R_{merge} (%)	6.7	7.1	14.4	6.4
	Phasing Statistics			
R_{iso} (%)	–	10.5	23.3	9.6
R_{Cullis} (centric)	–	0.55	0.67	0.78
phasing power	–	1.92	1.69	0.83

[a] Pt-3, platinum ethylenediamine dichloride. Au-1, potassium tetrabromoaurate. Hg-1, dimethylmercury. $R_{merge} = \sum|I - <I>|/\sum I$; I, intensity. $R_{iso} = \sum|F_{PH} - F_P|/\sum F_P$; F_{PH} and F_P, derivative and native structure factor amplitudes. $R_{Cullis} = \sum||F_{PH} - |F_P + F_H||/\sum|F_{PH} - F_P|$; F_H, calculated heavy atom structure factor. Phasing power $= <|F_H|/||F_{PH} - |F_P + F_H||>$.

& Gunsalus, 1989), alcohol dehydrogenase (*adhE*)(Kalman & Gunsalus, 1988), and pyruvate formate-lyase (*pfl*)(Kaiser & Sawers, 1995) in the presence of nitrate. The apparent redundancy of the system may be justified by the need to adjust gene expression to varying levels of environmental stimuli. This assumption is supported by differences in sensitivity of NarX and NarQ to nitrate and nitrite, as well as differences in their ability to interact with NarL, and in the specificity of the DNA target operons bound by NarL and NarP (Schröder et al., 1994; Darwin & Stewart, 1995; and unpublished work).

Response regulators that contain an N-terminal receiver domain that exhibits strong amino acid sequence homologies with CheY and related proteins are classified as the CheY superfamily (Volz, 1993). Within this group are at least four distinct families, as well as other sequences less easily classified. The CheY family, with ca. 130 amino acids, has only the N-terminal receiver domain, and possesses no output domain, DNA-binding or otherwise. The NtrC family has 2 effector domains C-terminal to the receiver domain, and ca. 460 amino acids. The OmpR has a single C-terminal domain and ca. 230 amino acids. The FixJ or LuxR family, to which NarL belongs, has a single C-terminal domain and ca. 220 amino acids in all. This C-terminal domain has sequence similarities to other transcriptional activataors, including the C-terminal region of σ factors (Kahn &Ditta, 1991).

Hence, both components of the NarL molecule have familial resemblances to other control proteins. The present three-dimensional structure of NarL extends our understanding of the structural conservatism of the receiver domain, provides the first structural template for the FixJ or LuxR family of proteins, and sheds light on the interactions between receiver and effector domains of bacterial response regulators.

MATERIALS AND METHODS

Crystallization. Recombinant protein contains residues 2−216 of the wild-type NarL sequence. Data on derivatives, phasing, and refinement are given in Table 1. A protein solution containing 20 mg/mL purified recombinant NarL, 20 mM Tris·HCl (pH 7.6), 0.5 mM $MgCl_2$, and 10% glycerol was mixed with an equal volume of the reservoir solution containing 0.1 M Tris·HCl (pH 8.5), 0.2 M sodium acetate, and 30% poly(ethylene glycol) (PEG) 4000. Sitting drops containing 20 μL of mixture were equilibrated by vapor diffusion at 4 °C against 20 mL of the reservoir solution. Crystals appeared as clusters of rectangular plates after 20 days, and single crystals isolated from the clusters were used for preparation of heavy-atom derivatives and for data collection. Crystals are orthorhombic, space group *I222*, with unit-cell dimensions $a = 61.04$ Å, $b = 78.32$ Å, and $c = 115.72$ Å. Assuming one molecule per asymmetric unit, the calculated V_M is 2.877 Å^3/Da. Three isomorphous heavy-atom derivatives were used for multiple isomorphous replacement (MIR) phasing: platinum ethylenediamine dichloride (Pt-3), potassium tetrabromoaurate (Au-1), and dimethylmercury (Hg-1), as listed in Table 1.

Data Collection. All data in this study were collected on a Rigaku RAXIS-IIC imaging plate system at 4°, with oscillation range 3°, and crystal-detector distance 120 mm. Images were processed with DENZO and scaled with SCALEPACK. Intensities were converted to structure factors using TRUNCATE from the CCP4 suite (French & Wilson, 1978; Otwinowski, 1993; Bailey, 1994). Crystallographic calculations for MIR and density modifications were also done with the CCP4 suite of programs.

Phase Analysis. The position of the major platinum-binding site was determined by inspection of a derivative difference Patterson map calculated at 4 Å resolution. A choice between space groups *I222* and $I2_12_12_1$ could not be made solely from diffraction extinctions along the crystallographic axes, because such screw axis extinctions are masked by *I*-centering. Accordingly, two sets of Harker sections were calculated, one for each possible space group. Platinum cross-peaks gave a self-consistent set of coordinates in all three Harker sections of the space group *I222*, whereas coordinates in the corresponding sections of $I2_12_12_1$ were inconsistent. A minor platinum-binding site and two gold-binding sites then were determined by the real space Patterson search routine RSPS (Knight, 1989). Positions and occupancies of heavy atoms were refined using centric reflections from 8 to 3 Å with MLPHARE (Otwinowski, 1991). Isotropic temperature factors during this and subsequent refinement were kept constant at 30 Å^2. Calculated phases with a figure of merit (FOM) 0.44 were used for difference Fourier maps to identify three more gold-binding sites and six relatively weak mercury-binding sites. Double-difference Fourier maps confirmed the mercury-binding sites, identified three more gold-binding sites (to a total of eight), but manifested triplets of density around each of the platinum-binding sites. Remodeling of platinum sites as triangles eliminated residual density from the double-difference Fourier map, reduced the Cullis *R*-factor from 0.79 to 0.55, and increased phasing power from 1.00 to 1.91 for the platinum derivative. The overall FOM for MIR phases from 65 Å to 3 Å was 0.57.

Model Building. The structure was refined using the 2.4 Å resolution platinum data set as the "native" set because of its somewhat higher resolution: 2.4 Å and 11 188 unique reflections for Pt, vs 2.6 Å and 8798 unique reflections for native (Table 1). Platinum sites are associated with methionines 28 and 175, on the back of the N-terminal and C-terminal domains, respectively (see Figure 2, below). They are far removed from any mechanistically critical region of the molecule, and the 27% increase of data in the Pt set leads to an improved image of the protein. MIR phases were

improved by a combination of solvent flattening, histogram matching, and skeletonization modes in density modification package DM (Cowtan, 1994) with an automatic mask determination and a solvent content of 56%. The free *R*-factor (R_{free}) on the last cycle of density modification was 0.319, and the overall figure of merit (FOM) increased to 0.78. An MIR map, calculated at 3 Å resolution, clearly showed molecular boundaries and secondary structure elements of the protein. Two regions of continuous electron density were interpreted as residues Pro6−Leu142 of the N-terminal domain and residues Leu156−Val210 of the C-terminal domain of NarL, and were fitted with atomic framework using FRODO (Jones, 1978). Main chain coordinates of the partial model, expanded with dummy atoms to fill a discontinuous electron density at the C-terminus, were used to calculate a protein mask. The mask was edited using MAMA and O and adjusted to cover 45% of the asymmetric unit (Jones et al., 1991). The platinum derivative was refined further to 2.6 Å resolution, and had a phasing power of 1.48 and a Cullis R-factor of 0.62. Density modification was repeated with new phases using a calculated mask and solvent content of 55%; the modified phases had FOM = 0.74 and R_{free} = 0.364. A map calculated at 2.6 Å resolution was of sufficient quality to locate most of the side chains and to extend the model to include residues Glu5−Leu142 and Gln155−Phe216, incorporating 93% of the complete NarL sequence.

Refinement. The partial structure was subjected to positional refinement in X-PLOR (Brünger, 1992a) using platinum derivative data from 8 Å to 2.4 Å resolution; 10% of the observed reflections were sequestered in a test set for cross-validation by the free *R*-factor method (Brünger, 1992b). Refinement was carried out by simulated annealing with a slow-cooling protocol, followed by conjugate gradient minimization and temperature factor refinement. The weight applied to the crystallographic pseudo-energy term in the refinement was optimized using a free *R*-factor, and was set at about one-quarter of the value calculated by X-PLOR, imposing rather tight restraints on the stereochemistry of the model. At this point, the normal crystallographic *R* and R_{free} were 0.256 and 0.310, respectively. Examination of the ($F_o - F_c$) difference Fourier map revealed strong triangle-shaped density on Met28 and Met175, confirming the earlier assignment of triplets of platinum-binding sites made during phasing. Positioning of sites relative to the sulfur atom corresponds to a tetrahedral coordination, although the exact stereochemistry of the platinum−methionine complex cannot be refined because of limited resolution and low site occupancies, which are approximately 0.18, 0.18, and 0.12 for Met28 and 0.10, 0.10, and 0.11 for Met175 on an absolute scale. Therefore, platinum ions were included in the model as single atoms with no restraints at the positions determined at the phasing step, and their occupancies were kept fixed during refinement. After minor revision of side chain conformations and inclusion of 46 water molecules, followed by 1 round of refinement, the *R* values decreased to 0.222/0.272 (R/R_{free}). Three fragments of density, adjacent to the exposed hydrophobic side chains of Ile26, Val34, and Val88, were interpreted as glycerol molecules.

Secondary structure features are shown in Figure 1, and the complete α-carbon skeleton is displayed in stereo in Figure 2. RMS deviations from ideality in the refined structure are 0.005 Å for bond lengths and 1.17° for bond angles. More than 94% of non-glycine and non-proline residues are in most favored regions of a standard Ramachandran plot, and the others are in additional allowed regions. Furthermore, no glycines or prolines occur in unfavorable conformations. Average temperature factors are 38.5 Å^2 for main chain atoms and 42.7 Å^2 for the whole molecule, in a good agreement with 41.4 Å^2 obtained from a Wilson plot.

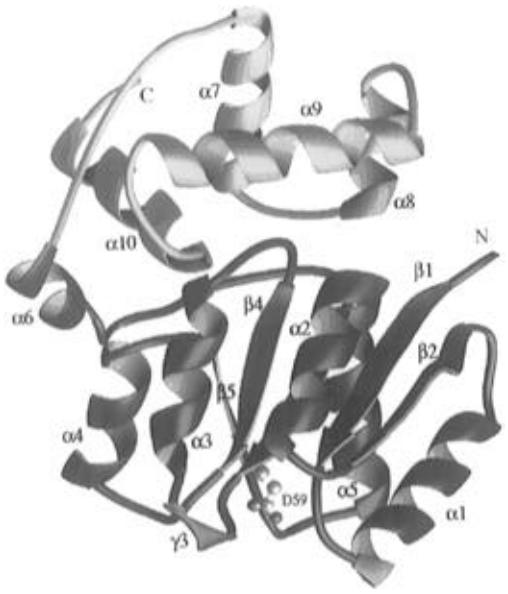

FIGURE 1: Secondary structure of NarL. The N-terminal domain is in blue and the C-terminal domain in yellow. Residues 143−154 were not visible in the electron density map and are represented schematically here by a pink loop. A ball-and-stick representation is given for the site of phosphorylation, Asp59. The figure is drawn with RIBBONS (Carson, 1987).

A representative portion of the electron density map, calculated at 2.6 Å resolution using MIR phases improved by density modification, is shown in Figure 3. The final model includes residues 5−142 and 155−216 of NarL, 6 platinum atoms in 2 clusters, 3 glycerol molecules, and 70 molecules of water. The crystallographic *R*-factor is 20.7%, and the free *R*-factor is 25.1% for 10 844 reflections from 8 to 2.4 Å. Corresponding values for 9408 strong reflections with $F_{obs} > 3\sigma$ are 19.3% and 23.6%. Both the intensity data and final coordinates have been submitted to the Brookhaven Protein Data Bank, and are available for immediate release, file name 1RNL.

RESULTS AND DISCUSSION

Overall Molecular Structure. The amino acid sequence of the NarL protein is shown in Figure 4, along with locations of α helices and strands of β sheet observed in this crystal structure analysis. The N-terminal signal receiver domain of NarL, shown in blue in Figure 1 and at the bottom in Figure 2, has the same doubly wound five-stranded α/β motif found in CheY. A parallel-chain β sheet with the topology β2−β1−β3−β4−β5 is surrounded by five α helices: α2−α3−α4 on one side of the sheet, and α1−α5 on the other. A γ turn loop is centered at position 64. The C-terminal DNA-binding domain, yellow in Figure 1 and at the top in Figure 2, has no parallel in CheY. This domain is built from four α helices: α8 and α9 form a HTH motif supported by α7, and α10 completes a hydrophobic core of the domain. The N- and C-terminal domains are connected by a linker

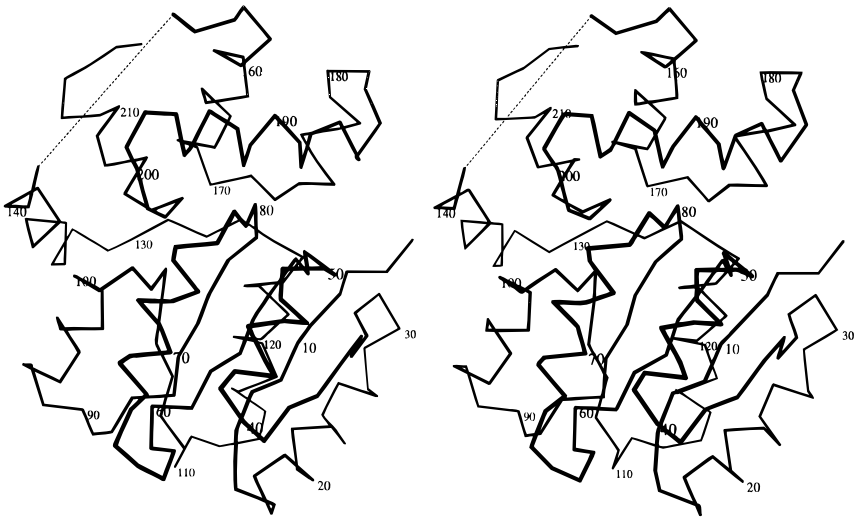

FIGURE 2: Stereoview of the polypeptide chain backbone based on coordinates of Cα atoms, labeled every 10th residue. The gap between domains is indicated by a dotted line. The figure is drawn with MOLSCRIPT (Kraulis, 1991).

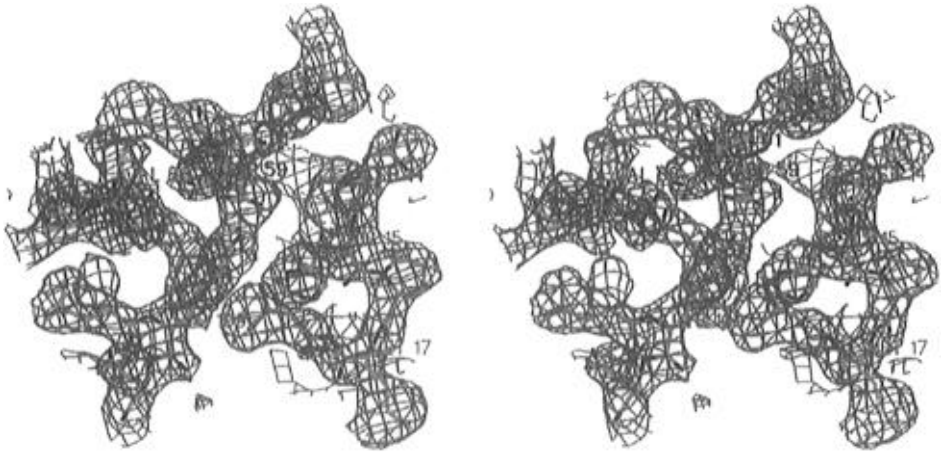

FIGURE 3: Stereoview of the NarL structure near the phosphorylation site. Representative electron density is at 2.6 Å resolution. The map, contoured at 1σ, was calculated using MIR phases after the final round of density modification and is superimposed on the refined model.

helix, α6, which has modest hydrophobic interactions with α10 of the C-terminal domain and with α4 of the N-terminal domain. Helices α6 and α7 are connected by a flexible tether that is not visible in our structure analysis, and is presumably disordered.

N-Terminal Receiver Domain. Figure 5 shows a superposition of the magnesium-free or apo-CheY molecule (Volz & Matsumura, 1991) on the N-terminal domain of NarL. Folding is nearly identical within the five-strand β sheet and helices α2−α3−α4, in the top half of Figure 5. Differences here are limited essentially to displacement of loop β3−α3 (containing the γ turn) and loop β4−α4. NMR studies of NTRC (Volkman et al., 1995) and the magnesium-bound form of CheY (Bellsolell et al., 1994) suggest that helix α4 is intrinsically unstable, but this helix is unchanged between apo-CheY and NarL. In contrast, CheY and NarL differ below the β sheet: the axis of helix α5 is rotated by ca. 21°, swinging the far end of the helix nearly 6 Å to the right in NarL, toward helix α1.

Figure 5 is a view into the phosphorylation site of NarL and CheY. A closeup view of the phosphorylatable Asp59 and its neighbors in NarL is shown in stereo in Figure 6. The Asp59 side chain is involved in a network of hydrogen bonds with the main chain NH and the side chain of Asn61, the extended side chain of Lys109, and at least one solvent molecule. Although the overall polypeptide chain fold in the phosphorylation region is conserved between NarL and CheY, several details of interactions in the active site differ, the most striking difference being the orientation of NarL side chains Asp13 and Asp14 relative to their equivalents in CheY. (Only NarL numbering is used in the text. CheY equivalents of numbered groups are given in the caption to Figure 6.) Whereas in CheY, Asp13 and Asp14 are turned toward the Asp59 phosphorylation site, in NarL they turn away from it. Voltz (1993) suggests that Arg19 may play a role as a hydrophobic cover of the α1−β1 interface. Asp13 and Asp14 in NarL are both exposed to the solvent, and could complement Arg19 in this role. However, the strong evolutionary conservatism of Arg19 (56% in the CheY superfamily and nearly 85% in the NarL family, Figure 4) might also indicate that it has a more substantial role. It may help stabilize the side chain of Asp13 in the inactive form of the enzyme. If Asp13 of CheY (NarL numbering) is replaced by the longer Glu side chain, then it, too, turns away from the phosphorylation site to hydrogen-bond with Arg19 as seen in Figure 6 (Stock et al., 1993).

The N-terminal domain of NarL belongs to the FixJ subfamily of receiver domains, which seems not to require a *cis* peptide bond between Lys109 and Pro110 like that seen in most CheY-like proteins (Volz, 1993). Even so, the

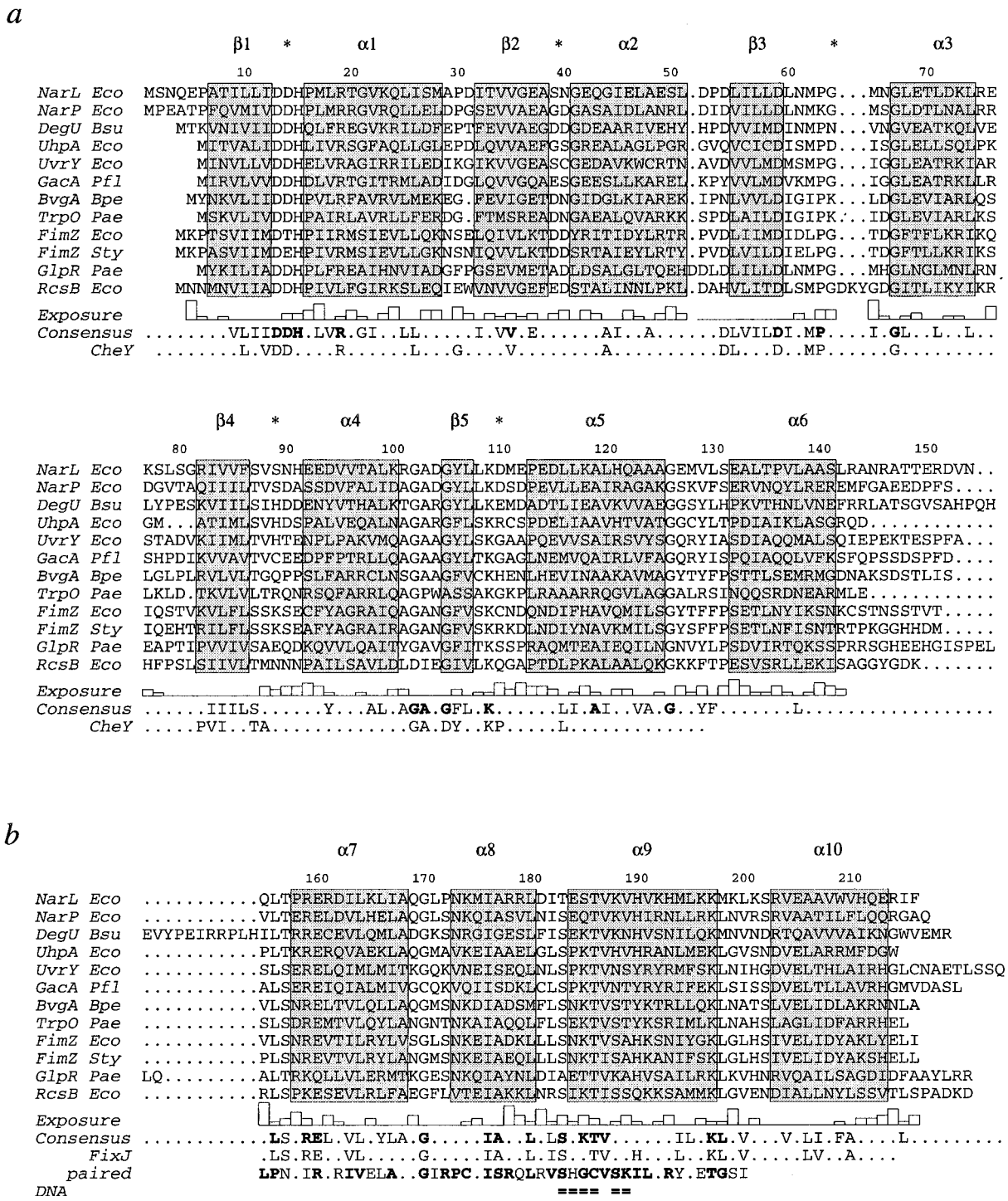

FIGURE 4: Sequence alignment of NarL with NarP and with other members of the same subfamily of transcriptional regulators. All proteins of this NarL subfamily have the same type of receiver and DNA-binding domains (Pao & Saier, 1995). Residues are numbered according to the NarL sequence. Shaded boxes represent secondary structure elements of NarL: α helices and β sheet strands. Asterisks mark loop regions adjacent to the phosphorylation site. Exposure, plotted below the aligned sequences, is calculated as the percentage of surface area of a residue accessible to the solvent. Consensus marks residues whose chemical character is conserved in at least 10 out of 12 sequences, with highly conserved residues shown in boldface type. (a) N-Terminal receiver domain, aligned with the related CheY (Volz, 1993); (b) C-terminal DNA-binding domain, aligned with the related FixJ (Kahn & Ditta, 1991). Six underlined residues in the DNA-binding region indicate a consensus sequence with *Drosophila* paired protein, interacting with DNA bases in the major groove (Xu et al., 1995).

extension of the side chain of Lys109 toward Asp59 is unchanged from CheY to NarL, preserving a salt bridge between these residues. But the conformation of the phosphorylatable Asp59 residue itself in NarL is more like that of the Mg^{2+}-bound form of CheY than of apo-CheY (Volz & Matsumura, 1991; Bellsolell et al., 1994). In NarL, the side chain of Asp59 is rotated by almost 90° compared to that of apo-CheY, and has (χ_1, χ_2) values of (−176°, 7°). This would produce steric clash with the carboxylate of Asp13 if that side chain were rotated around toward Asp59 as in apo-CheY.

The orientation of Asp59 is stabilized by hydrogen bonds to the side chain amide group and main chain amide of Asn61. That residue also interacts with the main chain of loop β4−α4 through a hydrogen bond to the carbonyl of Ser87. This provides a direct link between Asp59 and a

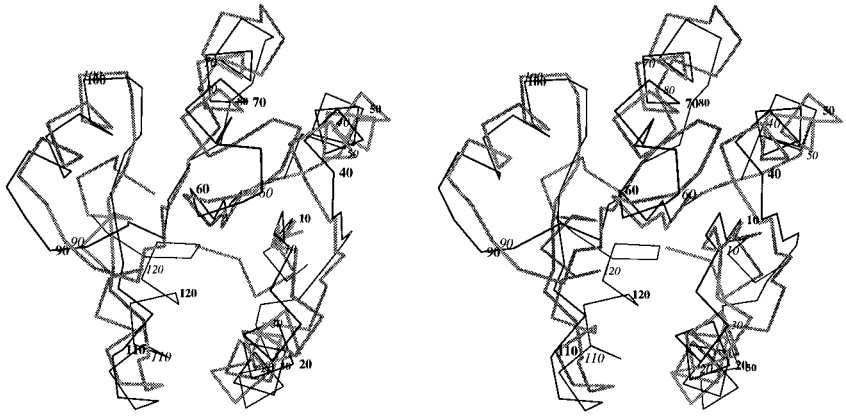

FIGURE 5: Stereoview of superimposed Cα traces of CheY (bold gray line) and the N-terminal domain of NarL (thin black line). Best fit was obtained by using residues 6−28, 37−48, 53−62, 66−77, 82−86, and 92−109 of NarL and residues 5−27, 35−46, 51−60, 64−75, 82−86, and 92−109 of CheY. RMS deviation for these 80 Cα atoms is 1.37 Å. α-helices 4−3−2 are at top from left to right, and 5−1 at bottom, also from left to right.

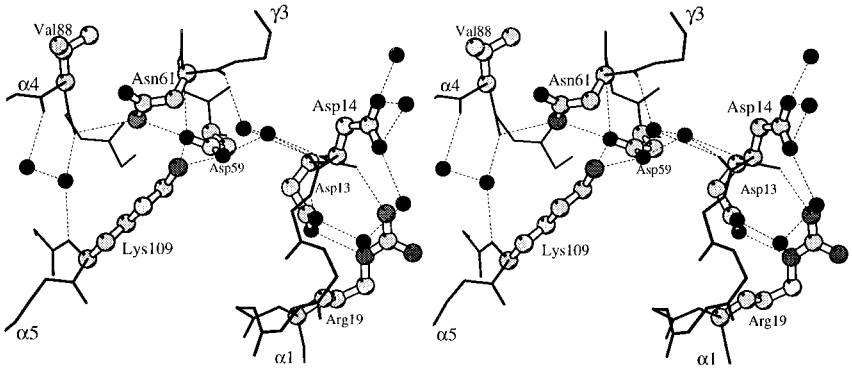

FIGURE 6: Stereoview of the NarL phosphorylation site. Residues drawn and their [CheY equivalents] are Asp13 [Asp12], Asp14 [Asp13], Arg19 [Arg18], Asp59 [Asp57], Val88 [Ala88], and Lys109 [Lys109]. Ball-and-stick representation is used for conserved residues and residues that interact with them. Isolated black spheres are oxygens of water molecules. Dotted lines indicate possible hydrogen bonds.

putative signaling surface of the receiver domain that has been inferred from analysis of spontaneous mutations (Roman et al., 1992). The link is amplified by tight packing of the carbonyl group of Asn61 against the hydrophobic side chain of Val88. This side chain has an unusually high solvent-accessible surface area for valine of ca. 70 Å^2, and in the crystal structure of NarL, it is shielded from solvent by a bound glycerol molecule. In aqueous solution, this hydrophobic residue probably would be induced to fold into the interior of the protein, and such a repositioning could affect the conformation of Asp59 through the link described above. In fact, the Val88→Ala mutant of NarL has a strong nitrate-independent constitutive phenotype, although it still needs phosphorylation for activity (Egan & Stewart, 1991; Li et al., 1994). A valine-to-alanine substitution would reduce the tendency of residue 88 to fold into the interior in aqueous solution, and could result in the same conformation as that seen in the crystal structure of NarL. Hence, a similarity in orientation of Asp59 in NarL and in the Mg^{2+}-bound form of CheY could reflect a possible role of Val88 as the "On" button of a "Bind Mg^{2+}" switch, triggered either by a Val88→Ala mutation or by docking of another protein to provide a hydrophobic pocket for Val88.

C-Terminal DNA-Binding Domain. When the C-terminal DNA-binding domain of NarL is compared with other protein domains containing the HTH or helix−turn−helix motif (Harrison, 1991), an especially close similarity is found with the N-terminal domain of the *Drosophila paired* protein (Xu et al., 1995). Superposition of helices α7, α8, and α9 of NarL on the corresponding helices α1, α2, and α3 of paired protein yields a RMS deviation of only 0.83 Å for the Cα atoms of 38 contiguous residues. Six residues, highly conserved within the NarL family (Figure 4), appear to be responsible for the similarity of folding. Ile163, Ile176, and Val187 build a hydrophobic cluster which fixes all three helices in their proper positions. The α8−α9 loop is anchored to the cluster by Leu180, and its position is fixed by the interaction of the main chain atoms with a small residue under the loop, Ala177. Finally, Gly170 adopts a conformation with (φ, ψ) = (85°, 15°) to assure a proper angle between helices α7 and α8. Two other highly conserved residues contribute to the stability of the fold: Glu160 at the amino end of α7 and Lys197 at the carboxyl end of α9 form a salt bridge, bringing together opposite ends of the fold. This link is stabilized further by hydrogen bonds between Lys197 and the main chain carbonyl group of residue 155, and between Glu160 and the main chain amide group of residue 193. In addition, Glu160 interacts with the hydroxyl of Thr157, providing an N-cap for helix α7.

In spite of the great folding similarity between the C-terminal domain of NarL and other HTH proteins, it cannot interact directly with DNA as they do, because the DNA-binding site is blocked by the N-terminal domain (Figure 7). However, if the C-terminal domain were to rotate away from the other domain, then a HTH complex with the DNA duplex would be quite possible. Positioning the C-terminal domain of NarL against a B-DNA duplex exactly as is observed in the DNA complex of *Drosophila paired* protein

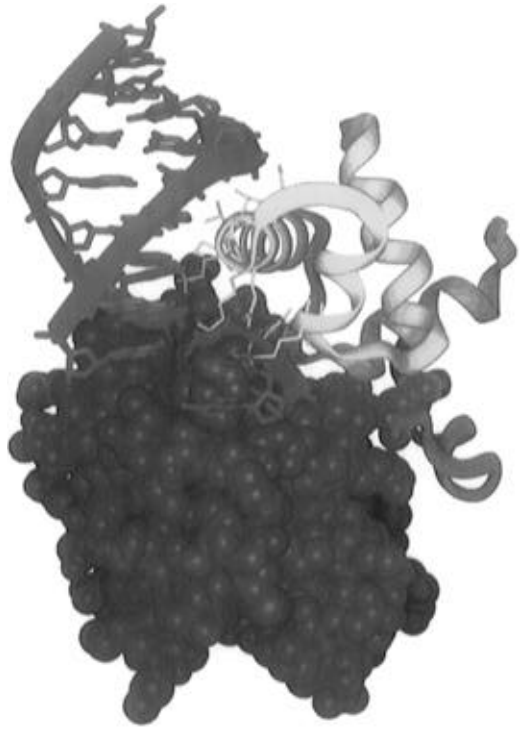

FIGURE 7: Possible mode of interaction of a DNA duplex (magenta) with the C-terminal domain of NarL (yellow), obtained by superposition of the HTH motif of the paired protein/DNA complex on the corresponding residues of NarL. This mode of binding is impossible in the present NarL molecule because the DNA helix is blocked by the N-terminal receiver domain (blue). But if the two NarL domains were to rotate or move apart, this paired protein-like HTH binding might be quite feasible. The best fit was obtained by superimposing residues 161−198 of NarL and residues 24−61 of paired protein; the RMS deviation for 38 Cα atoms of these residues is 0.83 Å. The figure is drawn with Insight II (Insight, 1995).

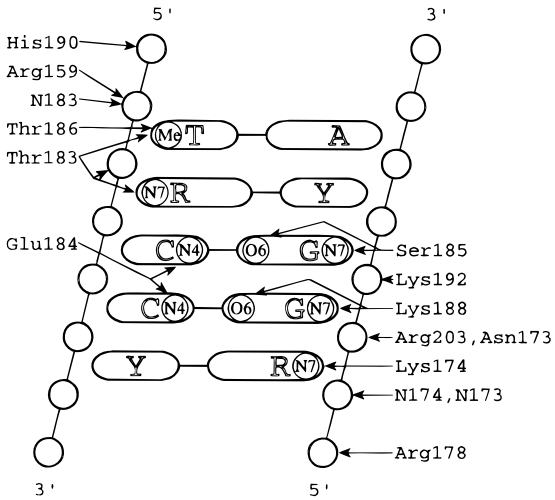

FIGURE 8: Schematic diagram of possible interactions between NarL and DNA, based on structural alignment with the paired protein/DNA complex. View into the major groove, with strand 1 running from top to bottom at left, and strand 2 from bottom to top at right. Y = pyrimidine. R = purine. N = main chain amide.

even suggests a possible stereochemical basis for NarL specificity. As schematized in Figure 8, such a *paired*-like positioning would bring the Cγ-methyl groups of Thr183 and Thr186 into perfect positions to contact and recognize the methyl group of a thymine base. Moreover, the hydroxyl group of Thr183 could donate a hydrogen bond to the N7 position of a subsequent purine base, favoring a binding sequence of $^{5'}$-T-R-$^{3'}$ along strand 1. Thr186 is invariant throughout the NarL family of sequences, but residue 183 is Ser rather than Thr in 10 of 12 sequences (Figure 4). This may mean only a somewhat weakened preference for T at the first position in these members of the family.

The hydroxyl group of Ser185 sits at the very bottom of the major groove, in a position where it could form a bifurcated hydrogen bond to the N7 and O6 atoms of a guanine on strand 2 at the third base pair along the putative recognition site. Lys188 is similarly positioned to H-bond to guanine N7 and O6 at the fourth base pair along strand 2. On the strand 1 ends of both of these base pairs, Glu184 is in position to accept hydrogen bonds from the N4 amino groups of cytosines. Hence the most favored sequence along strand 1 becomes $^{5'}$-T-R-C-C-$^{3'}$. However, a small repositioning of the long, flexible side chain of Lys188 could result in formation of a salt bridge with the carboxyl group of Glu184, which would remove both side chains from the N4-O6 pair at the fourth position, allowing a T-A base pair instead, although with lower affinity. The Lys188 interaction with N7 is preserved under this change, ultimately dictating pyrimidine at the fourth position on strand 1, with a preference for cytosine.

Lys174 is capable of donating a hydrogen bond to a purine N7 at base 5 along strand 2, although the inherent flexibility of the lysine chain makes this interaction only a possibility. Nevertheless, the optimal sequence for NarL in a *paired*-like complex with B-DNA would be something close to $^{5'}$-T-R-C-C-Y-$^{3'}$. The remaining residues shown in Figure 8 interact with the phosphate backbone of the duplex, and most are highly variable among members of the NarL family. Precise positioning of the DNA-binding domain seems to be achieved through direct interactions of the main chain amides of residues 173 and 174 with the phosphate oxygens of one strand, and residue 183 with another strand of the DNA target. Orientation of the protein relative to DNA is fixed by the firm hydrogen bonding of conserved Arg159 at the amino end of α7 to the phosphate group at the first position on strand 1.

This simplified model of DNA recognition is based solely on direct interactions between protein and DNA bases from a *paired*-like HTH/DNA complex, and does not consider the structure and hydration of the DNA target, which certainly can affect specificity and affinity of binding (Shakked et al., 1994). Nevertheless, the deduced nucleotide binding site sequence, T-R-C-C-Y (high affinity) or T-R-C-T-N (lower affinity), is in good agreement with the first five positions of the proposed consensus sequence for the NarL-binding site, T-A-C-Y-N (Tyson et al., 1993). Hence, this model may prove useful in a mutational analysis of NarL and selection of corresponding DNA operators.

The polypeptide chain fold of NarL beyond region α7−α9 seems to exclude a possibility of additional contacts in the minor groove, of the type that were observed for the *paired* protein and Hin recombinase (Feng et al., 1994). It also is unlikely that an isolated C-terminal domain of NarL alone could dimerize to recognize pairs of either direct or inverted repeats found at different NarL-sensitive promoters, since the domain itself spans less than seven base pairs in the paired-protein modeling of NarL to DNA. This would leave a gap of at least three base pairs between domains in binding to a direct repeat, or two base pairs between domains for an inverted repeat, making protein−protein association

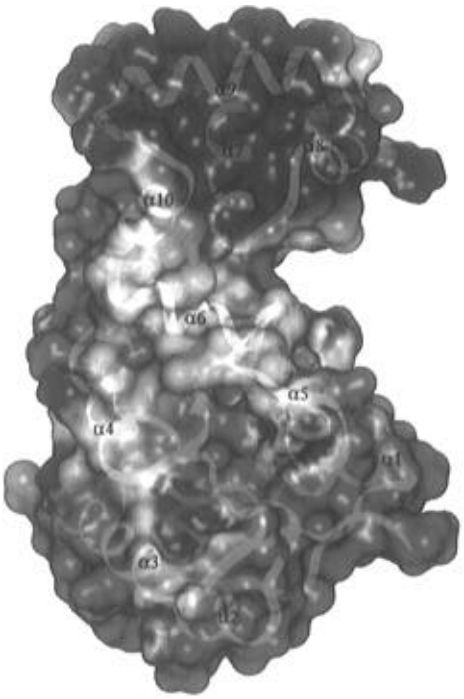

FIGURE 9: Diagram of the interdomain interface. The C-terminal domain has been rotated around helix α6 (center of drawing) to bring interface surfaces of both domains into one plane. The molecular surface, colored according to electrostatic potential, is calculated using the program DelPhi and drawn with Insight II (Insight, 1995). Red regions of the surface have a negative electrostatic potential, gray regions are uncharged, and blue regions are positive. Note a strong separation of charges between the two domains: the negatively charged N-terminal portion of the interface (bottom) is predominantly acidic, while the C-terminal interface (top) is entirely basic, as would be expected for a DNA-binding protein.

unlikely. Structural similarity with eukaryotic transcriptional activators also argues in favor of monomeric binding of NarL. Surface analysis of the C-terminal domain of NarL does identify one hydrophobic residue as a potential site for protein−protein interaction: Val189, which would be exposed in the major groove of DNA next to the first base pair of the binding site. The side chain of Val189 is too far from the bottom of the groove to make any contact with bases, but is in a favorable position to interact with another major groove-binding protein of the transcriptional complex.

Interactions between Domains. The flaw in the HTH recognition mechanism just described is that the N-terminal domain blocks DNA binding in the molecule as observed in the crystal (Figure 7). Therefore, if the observed "closed" conformation of the inactive, unphosphorylated form of NarL is not an artifact of crystal packing, then activation of the protein for DNA binding must separate N- and C-terminal domains, or at least open a cleft between them. The open conformation must present a recognition helix α9 which is not obstructed by the N-terminal receiver domain, and the entropy loss due to opening of the interdomain interface must be compensated energetically.

Surface analysis of the interface argues against complete separation of domains as a possible mechanism of activation, since about one-third of the residues constituting the interface are hydrophobic (gray side chains in Figure 9). On the other hand, the clustering of hydrophobic residues near interdomain helix α6 suggests a role for this helix as a hinge. Rotation around the axis of α6 would expose the polar regions of the interface, while the nonpolar part undergoes only minor changes. In this model, opening could be compensated partially by hydration of the polar regions of the interface. Hence, phosphorylation-induced conformational changes would need to supply only minor changes in the overall energy, in order to shift the equilibrium from a closed to an open conformation.

From this single structure, the possibility remains that the observed orientation of N- and C-terminal domains might be the result of crystal packing. Or, since the polypeptide tether connecting domains is not visible in the X-ray analysis, it could even be that the wrong N- and C-terminal domains have been chosen in our extraction of one complete NarL molecule. But both of these possibilities were ruled out recently when we obtained crystals of NarL in a different space group. Monoclinic crystals of NarL belong to space group *C*2, have two molecules per asymmetric unit, and diffract to 2.2 Å. Refinement of this structure is still in progress. The relative positions of the N-terminal and C-terminal domains in both of the two independent molecules are identical to those seen in the orthorhombic form of NarL, although local crystal environments differ. Hence, the arrangement of N- and C-terminal domains in Figure 1 is a correct picture of the unphosphorylated apo-NarL enzyme.

Conclusions. This X-ray analysis has shown us the structure of the unphosphorylated NarL, without bound Mg^{2+}, in a closed conformation that clearly cannot bind DNA. Many questions raised in part by this structure remain unanswered, and must be the subject of future investigations:

(1) How are NarL and NarP activated for binding to DNA by phosphorylation? That is, how is the phosphorylation state of Asp59 in the N-terminal domain communicated to the C-terminal DNA-binding domain? Are there other means of activating the C-terminal domain besides phosphorylation of Asp59? What structural role does Mg^{2+} play in this activation process?

(2) How do NarL and NarP recognize the different base sequences of putative binding sites in their various regulons? Are all of these sites actually operative? Do NarL and NarP exhibit different affinities for one or another of these sites, and if so, what is the molecular basis for this site discrimination?

(3) How do NarL and NarP bind to DNA: as monomers, normal symmetric dimers, head-to-tail dimers, or even possibly heterodimers under some circumstances?

(4) Is the binding of NarL to DNA entirely mediated by the C-terminal output domain, with the N-terminal receiver domain acting only as a brake or inhibitor? Or does the N-terminal domain also contribute to DNA binding, perhaps via dimer formation?

(5) Is there any evidence for NarL/NarL intermolecular association, of a sort that might be expected if the DNA loops back to bring two NarL/DNA binding regions into close proximity?

This structure analysis has defined the problem inherent in question 1: Phosphorylation of Asp59 within the specificity pocket at the bottom of the N-terminal domain, as viewed in Figure 1, must somehow send a signal that affects the interdomain surface between blue and yellow subunits. The two domains must either rotate or separate, so that helix α9 is free to insert into the major groove of B-DNA. If the comparison with HTH DNA-binding proteins is valid, then the observed binding preference of T-A-C-Y-N would be

explained reasonably well by the protein/DNA interactions sketched in Figure 8. A detailed explanation of site specificity, as embodied in question 2, will have to await X-ray analyses of several DNA complexes of both NarL and NarP, and further biochemical work.

A major experimental difficulty is that, for the signal receptor to be easily reset in preparation for a new signal, NarL must be easy to dephosphorylate, and it is, with a half-life of only 30 min (Schroeder et al., 1994). We have not yet succeeded in trapping the phosphorylated, DNA-bound state of the molecule, and are searching for possible constitutive DNA-binding mutants that could be used instead. Another approach will be to attempt to cocrystallize DNA with an isolated C-terminal domain, without interference from the receptor domain. This first dual-component response regulator structure has defined the problems; many more structures under controlled conditions will be required to solve them.

ACKNOWLEDGMENT

We thank Duilio Cascio for help in data collection, and Manfred Weiss and Kevin Cowtan for helpful discussions of refinement and density modification.

REFERENCES

Bailey, S. (1994) *Acta Crystallogr. D50*, 760−763.
Bellsolell, L., Prieto, J., Serrano, L., & Coll, M. (1994) *J. Mol. Biol. 238*, 489−495.
Bourret, R. B., Borkovich, K. A., & Simon, M. I. (1991) *Annu. Rev. Biochem. 60*, 401−441.
Brünger, A. T. (1992a) *X-PLOR: A system for x-ray crystallography and NMR*, Yale University, New Haven, CT.
Brünger, A. T. (1992b) *Nature 355*, 472−475.
Carson, M. (1987) *J. Mol. Graphics 5*, 103−106.
Cotter, P. A., & Gunsalus, R. P. (1989) *J. Bacteriol. 171*, 3817−3823.
Cowtan, K. D. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31*, 34−38.
Darwin, A. J., & Stewart, V. (1995) *J. Mol. Biol. 251*, 15−29.
Egan, S. M., & Stewart, V. (1991) *J. Bacteriol. 173*, 4424−4432.
Feng, J. A., Johnson, R. C., & Dickerson, R. E. (1994) *Science 263*, 348−355.
French, S., & Wilson, K. (1978) *Acta Crystallogr. A34*, 517−525.
Gunsalus, R. P. (1992) *J. Bacteriol. 174*, 7069−7074.
Harrison, S. C. (1991) *Nature 353*, 715−719.
Insight II Users Guide (1995) Biosym/MSI, San Diego, CA.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268−272.
Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
Kahn, D., & Ditta, G. (1991) *Mol. Microbiol. 5*, 987−997.
Kaiser, M., & Sawers, G. (1995) *J. Bacteriol. 177*, 3647−3655.
Kalman, L., & Gunsalus, R. P. (1988) *J. Bacteriol. 170*, 623−629.
Kalman, L., & Gunsalus, R. P. (1989) *J. Bacteriol. 171*, 3810−3816.
Knight, S. (1989) *Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase—Structural Study*, Swedish University of Agricultural Sciences, Uppsala, Sweden.
Kolisnikov, T., & Gunsalus, R. P. (1992) *J. Bacteriol. 174*, 7104−7111.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Li, J., Kustu, S., & Stewart, V. (1994) *J. Mol. Biol. 241*, 150−165.
Lowry, D. F., Roth, A. F., Rupert, P. B., Dahlquist, F. W., Moy, F. J., Domaille, P. J., & Matsumura, P. (1994) *J. Biol. Chem. 269*, 26358−26362.
Moy, F. J., Lowry, D. F., Matsumura, P., Dahlquist, F. W., Krywko, J. E., & Domaille, P. J. (1994) *Biochemistry 33*, 10731−10742.
Otwinowski, Z. (1991) *Maximum Likelihood Refinement of Heavy Atom Parameters* 1−80−86, Daresbury Laboratory, Warrington, WA4 4AD, U.K.
Otwinowski, Z. (1993) *Oscillation Data Reduction Program* 1−56−62, Daresbury Laboratory, Warrington, WA4 4AD, U.K.
Pao, G. M., & Saier, M. H., Jr. (1995) *J. Mol. Evol. 40*, 136−154.
Parkinson, J. S., & Kofoid, E. C. (1992) *Annu. Rev. Genet. 26*, 71−112.
Roman, S. J., Meyers, M., Volz, K., & Matsumura, P. (1992) *J. Bacteriol. 174*, 6247−6255.
Schröder, I., Wolin, C. D., Cavicchioli, R., & Gunsalus, R. P. (1994) *J. Bacteriol. 176*, 4985−4992.
Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., & Sigler, P. B. (1994) *Nature 368*, 469−473.
Stewart, V. (1988) *Microbiol. Rev. 52*, 190−232.
Stewart, V. (1994) *Res. Microbiol. 145*, 450−454.
Stock, A. M., Mottonen, J. M., Stock, J. B., & Schutt, C. E. (1989) *Nature 337*, 745−749.
Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D., & Petsko, G. A. (1993) *Biochemistry 32*, 13375−13380.
Tyson, K. L., Bell, A. I., Cole, J. A., & Busby, S. J. W. (1993) *Mol. Microbiol. 7*, 151−157.
Volkman, B. F., Nohaile, M. J., Amy, N. K., Kustu, S., & Wemmer, D. E. (1995) *Biochemistry 34*, 1413−1424.
Volz, K. (1993) *Biochemistry 32*, 11741−11753.
Volz, K., & Matsumura, P. (1991) *J. Biol. Chem. 266*, 15511−15519.
Xu, W., Rould, M. A., Jun, S., Desplan, C., & Pabo, C. O. (1995) *Cell 80*, 639−650.

BI960919O