NarL Dimerization? Suggestive Evidence from a New Crystal Form[†,‡]

Igor Baikalov, Imke Schröder, Maria Kaczor-Grzeskowiak, Duilio Cascio, Robert P. Gunsalus, and Richard E. Dickerson*

Molecular Biology Institute, University of California, Los Angeles, California 90095-1570

Received September 24, 1997

ABSTRACT: The structure of the *Escherichia coli* response regulator NarL has been solved in a new, monoclinic space group, and compared with the earlier orthorhombic crystal structure. Because the monoclinic crystal has two independent NarL molecules per asymmetric unit, we now have three completely independent snapshots of the NarL molecule: two from the monoclinic form and one from the orthorhombic. Comparison of these three structures shows the following: (a) The pairing of N and C domains of the NarL molecule proposed from the earlier analysis is in fact correct, although the polypeptide chain connecting domains was, and remains, disordered and not completely visible. The new structure exhibits identical relative orientation of N and C domains, and supplies some of the missing residues, leaving a gap of only seven amino acids. (b) Examination of corresponding features in the three independent NarL molecules shows that deformations in structure produced by crystal packing are negligible. (c) The "telephone receiver" model of NarL activation is confirmed. The N domain of NarL blocks the binding of DNA to the C domain that would be expected from the helix−turn−helix structure of the C domain. Hence, binding can only occur after significant displacement of N and C domains. (d) NarL monomers have a strong tendency toward dimerization involving contacts between helixes α1 in the two monomers, and this may have mechanistic significance in DNA binding. Analogous involvement of helix α1 in intermolecular contacts is also found in UhpA and in the CheY/CheZ complex.

The X-ray crystal structure of the bacterial response regulator NarL (Figures 1 and 2), solved in orthorhombic space group *I*222 by Baikalov et al. (*1*), revealed an N-terminal signal receiver domain (or N domain) of 131 amino acids and a C-terminal DNA-binding domain (or C domain) of 62 amino acids, connected by a 10-residue linker α helix and a 12-residue flexible tether that was disordered and not visible in the X-ray analysis. The site of phosphorylation, by which NarL is activated, is Asp59 on the opposite side of the N domain from the C domain which it activates. Figures 1 and 2 show the conventional "front" side of the NarL molecule, and are repeated from reference *1* as reference or locator diagrams for the discussion that follows. Subsequent diagrams of NarL in this paper will show different views of the molecule.

Several major unsolved problems remained from that initial structure analysis:

(1) The C domain is a cluster of four α helixes resembling the familiar helix−turn−helix or HTH structure found in many DNA-binding proteins. Yet the putative DNA-binding site in that analogy is blocked by being packed against the N domain. Clearly, if the HTH analogy is correct, the C domain must move away from the N domain before it can bind DNA.

(2) It is unclear how phosphorylation of Asp59 within a cavity at one end of the N domain (bottom of Figure 1) can transmit a signal to the C domain at the other end (top of Figure 1). Some structural rearrangement presumably is necessary, but the nature of this change is unknown because the phosphorylated form of NarL to date has been too easily dephosphorylated for a crystal structure analysis to be carried out.

(3) Putative binding sites on the genes that NarL controls are confusing (Figure 3). There is reasonable agreement on a concensus binding sequence for NarL: T-R-C-Y-N (*1*, *2*). (R = purine, Y = pyrimidine, N = any base.) But sites of this class located by footprinting and mutation studies occur (a) in adjacent antiparallel pairs suggestive of NarL dimerization, (b) in closely spaced parallel sites suggestive of linear association of NarL molecules along the genome, and (c) in isolation. Does each NarL molecule act independently and alone when it binds to its control regions, or is dimerization or some other type of intermolecular association a significant part of the recognition and binding process?

(4) Since the covalent connection between N and C domains is not visible in the crystal structure, is it absolutely certain that the correct pairing of N and C domains has been selected from the electron density map? Might the blocking of the DNA-binding site on C by domain N merely reflect the circumstance that the wrong N domain had been selected from the crystal structure?

(5) As with every protein crystal structure analysis, it is possible that crystal packing forces have distorted surface features that are important to the understanding of the protein. In particular, even if the correct pairing of N and C domains has been chosen, is it possible that the packing of the two

[†] This work was supported by NSF Grant MCB-9630809.

[‡] Data and coordinates are on file with the Brookhaven Protein Data Bank and available for immediate release: atomic coordinate code 1a04, observed structure factor code r1a04sf.

* To whom correspondence should be addressed. Temporary address from 1 October 1997 through 30 September 1998: Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Rd., Oxford OX1 3QU, United Kingdom. Telephone: Oxford (01865) 275 372. Fax: Oxford (01865) 275 182. E-mail: red@iop.ox.ax.uk.

S0006-2960(97)02365-9 CCC: $15.00
Published on Web 02/25/1998

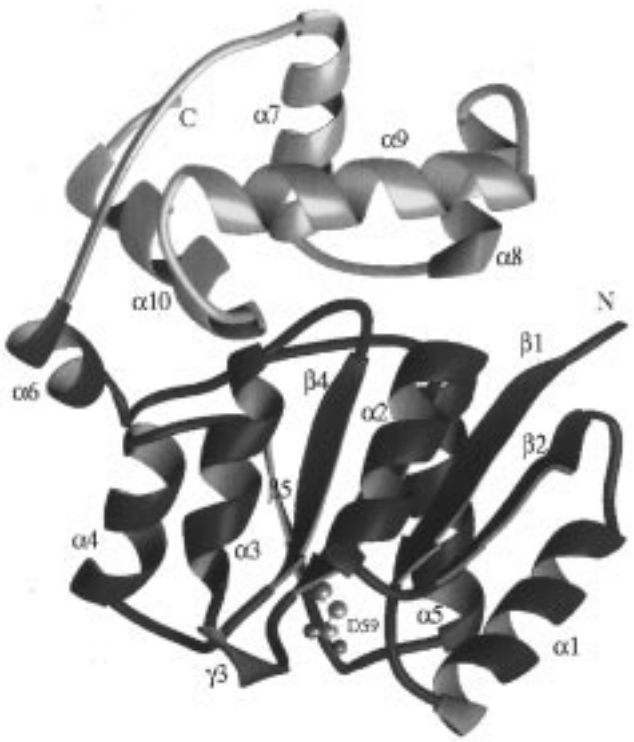

FIGURE 1: Folding of the NarL molecule in the orthorhombic crystal form. N-Terminal signal receiver domain in blue at bottom; C-terminal DNA-binding domain in yellow at top. The pink connecting chain containing residues 143−154 was disordered and not visible in the orthorhombic X-ray analysis. The phosphorylation site by which NarL is activated is at aspartic acid D59 at the very bottom. From Baikalov et al. (*1*).

domains together and the blocking of the putative DNA-binding site are a consequence of the way the domains have been pushed against one another in the crystal?

We have provisionally adopted a "telephone receiver" model for NarL activation, in which a receiver base (the N domain) and handset (the C domain) are connected by a flexible cord (the nonvisible peptide chain). In this model, activation of the N domain induces the C domain to move away, or at least to rotate, so as to free up the HTH structure for binding to DNA. The validity of this working model, of course, depends upon our having selected the proper duo of N and C domains in outlining a single NarL molecule, and on their remaining in a functionally relevant relationship when packed within the crystal. Hence, an effort has been made to obtain NarL in a different crystal form, to solve its structure, and to see whether the same close association of domains shown in Figures 1 and 2 recurs in a different crystal setting.

The original crystal form of NarL (1) was in space group *I*222 with one NarL molecule per asymmetric unit (Table 1). The new crystal form is space group *C*2 with two independent NarL molecules in the asymmetric unit. Hence, X-ray crystallography has now provided three different snapshots of the NarL molecule in different local environments. This paper describes the structure analysis and structure comparison of the new crystal form.

MATERIALS AND METHODS

Crystallization and Data Collection. Crystallization conditions are similar to those of the orthorhombic form of NarL (*1*), although a new and slightly more concentrated preparation of protein has been used. A protein solution containing 23.3 mg/mL purified NarL, 20 mM Tris·HCl (pH 7.6), 0.5 mM $MgCl_2$, and 10% glycerol was mixed with an equal volume of a 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 began to appear as clusters of plates after 11−15 days.

Data on the old and new crystals are compared in Table 1. The new crystals are space group *C*2 with two independent molecules per asymmetric unit. Data were collected from three thin crystals on a Rigaku RAXIS-IIC imaging plate system at −180 °C, oscillation range 2, and crystal−detector distance 120 mm. Crystals were flash-frozen and mounted in different orientations. In all, 94 frames were processed with DENZO to record 93 113 observations, and merged using SCALEPACK (*3*) to yield a total of 22 871

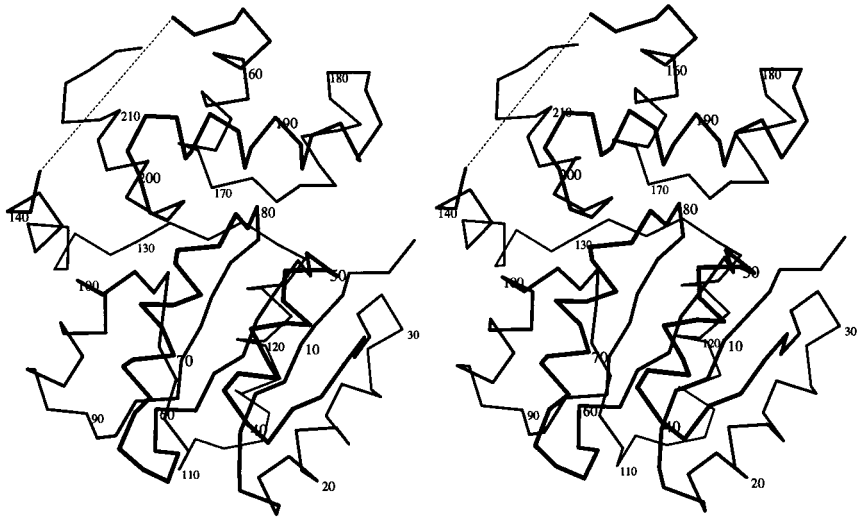

FIGURE 2: Stereoview of the polypeptide chain backbone in orthorhombic NarL, viewed from the same orientation as Figure 1. Bends mark α-carbon positions; straight lines are peptide links connecting them. Every tenth α-carbon position is numbered. The unobserved chain 143−154 connecting domains is dotted. Residues 150−154 are now visible in the current monoclinic analysis, as will be shown later in Figures 9 and 10. From reference *1*.

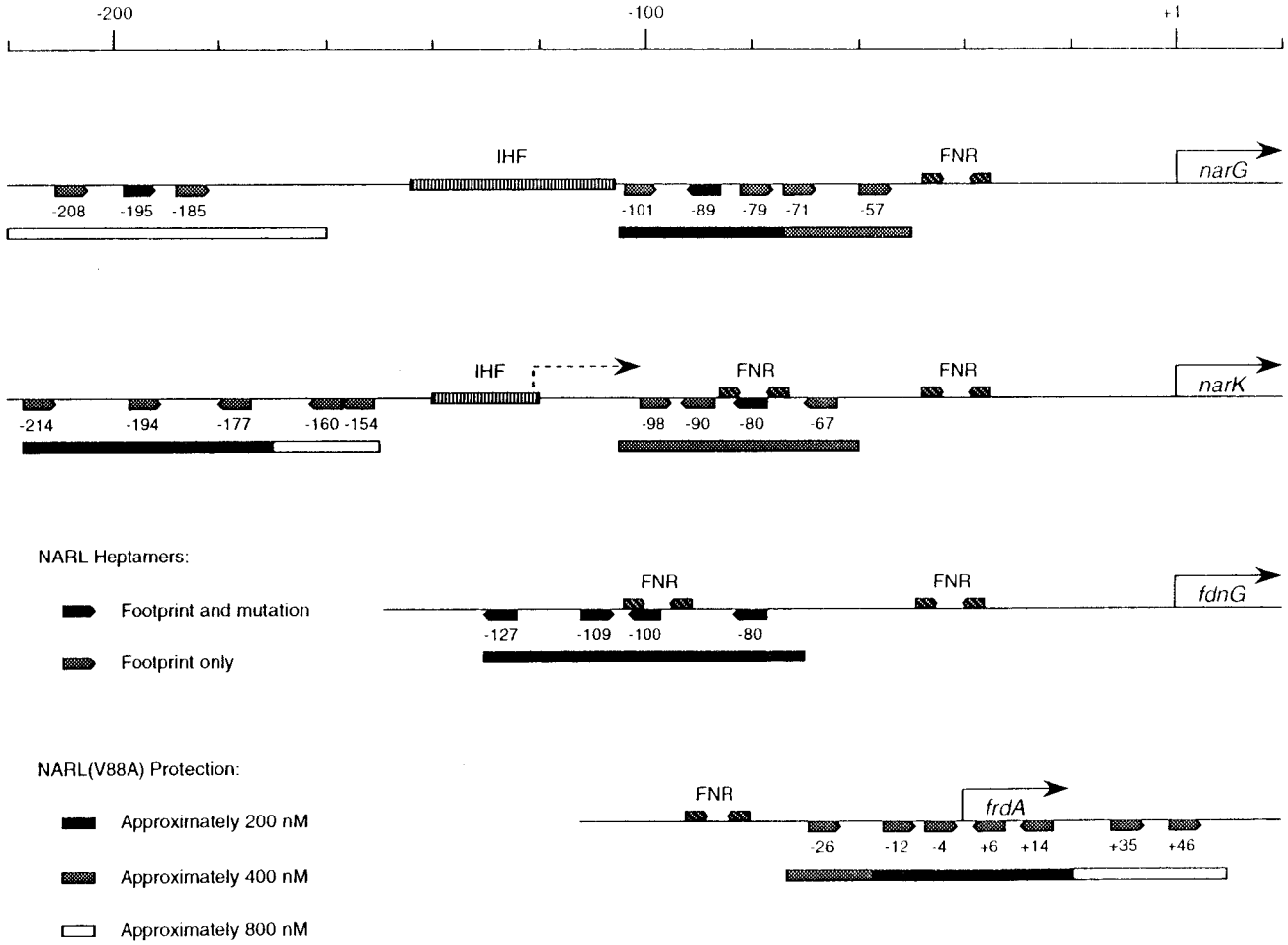

FIGURE 3: Arrangement of NarL-binding sites in four different control regions: narG, narK, fdnG, and frdA. Short arrows indicate the direction of sequences recognized by NarL, as determined by footprint and mutation studies. Presumed FNR protein and IHF protein-binding sites are also shown. Regions protected by MBP−NarL(V88A) protein from DNase I attack are indicated as rectangles below the line. From Li et al. (*23*).

Table 1: Crystal Data for the Two Forms of NarL Crystals (Molecular Mass = 24 036 Da)

crystal class	orthorhombic	monoclinic
space group	*I*222	*C*2
a (Å)	61.04	156.3
b (Å)	78.32	38.5
c (Å)	115.72	106.9
β (deg)	90.00	131.9
volume per unit cell (Å^3)	553220	478800
molecules per unit cell	8	8
molecules per asymmetric unit	1	2
V_M (Å^3/Da)	2.877	2.491
temp of data collection (°C)	4	−180
resolution of data set (Å)	2.4 (Pt deriv)	2.2
no. of unique reflections	11188	22871
R_{merge} (%)	7.1	7.8

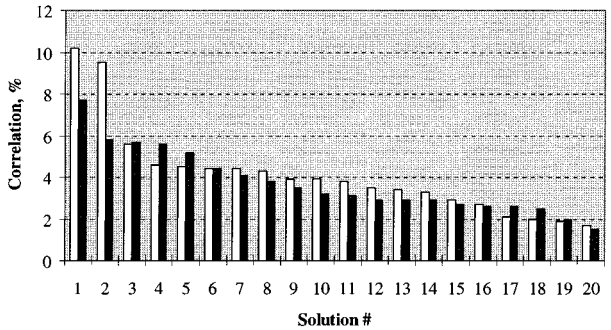

FIGURE 4: Comparison of the 20 best rotation searches using the N domain (white bars) or C domain (dark bars) as search objects. The plot shows correlation coefficients between observed and calculated intensities in space group *P*1, for data between 8.0 and 4.0 Å resolution.

unique reflections with R_{merge} = 7.8%. The data set is 93% complete to 2.2 Å resolution, and the outermost shell is still 67% complete. Intensities were converted to structure factors using TRUNCATE from the CCP4 suite (*4*, *5*).

The orthorhombic *I*222 crystals had a volume per molecule of 69 150 Å^3. The new monoclinic *C*2 crystal volume of 119 700 Å^3 per asymmetric unit immediately suggested two molecules in the asymmetric unit. This assumption gave a Matthews number (*6*), V_M, of 2.491 Å^3/Da, compared with 2.877 Å^3/Da for the orthorhombic form. Tighter crystal packing and a lower temperature of data collection together could explain the higher resolution of the new data set, despite the smaller crystal size.

Molecular Replacement. To avoid possible bias in tracing the missing interdomain chain, the monoclinic structure was solved by molecular replacement using only those portions of the orthorhombic NarL molecule that were clear and unambiguous: residues 6−126 of the N domain and residues 158−210 of the C domain. Moreover, to avoid biasing the monoclinic analysis with assumptions about relative domain orientation from the orthorhombic structure, separate mo-

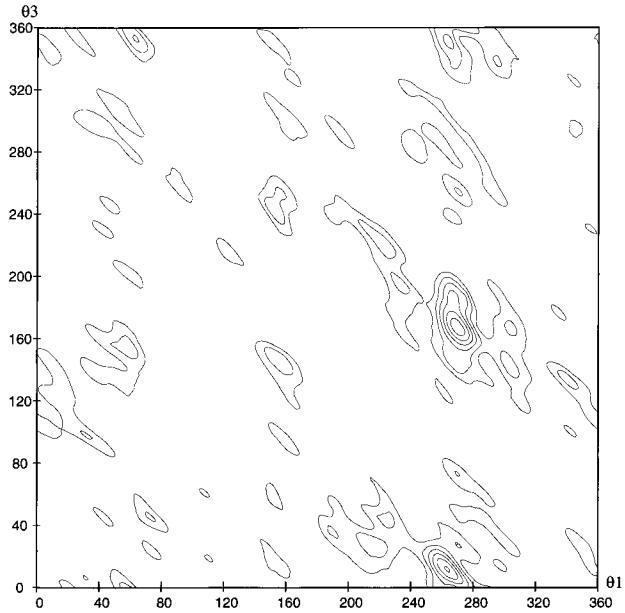

FIGURE 5: Section $\theta_2 = 36°$ of the cross-rotation map for the N-terminal domain of NarL. Contour intervals of 1 σ. Resolution of data between 8.0 and 4.0 Å. Note the two six-contour peaks at lower right. These eventually proved to be correct orientations of the two independent N domains in the asymmetric unit.

lecular replacement searches were carried out independently with N and C domains. Temperature factors were set at 20 Å^2 for all atoms of the search model, corresponding to the mean value obtained from a Wilson plot for the monoclinic data set. All molecular replacement calculations were performed with AMORE in the CCP4 package (*7*), using data between 8.0 and 4.0 Å resolution.

The rotation search unambiguously identified the orientations of the N domain for both molecules in the asymmetric unit, as shown by solutions 1 and 2 in Figure 4. By contrast, the orientation of the C domain was not clear from a rotation search, which either could arise only from the smaller size of the C domain or could indicate substantial differences in C domain orientation between the two crystal forms. Hence, the top 20 rotation solutions for the C domain were all considered in subsequent searches, anticipating that correct positioning of the N domain could ultimately define the correct orientation of the smaller fragment.

Figure 5 shows the two best rotation solutions for the N domain on section $\theta_2 = 36°$ of the cross-rotation map. Two six-contour peaks, at [$\theta_1 = 270°$, $\theta_3 = 166°$] and [$\theta_1 = 264°$, $\theta_3 = 12°$], are separated by rotation of 162° around the *z* axis. A translation search in the *xz* plane yielded significant five-contour peaks for both of these rotation solutions (Figure 6), corresponding to correlation coefficients between observed and calculated intensities of 0.209 and 0.219, respectively. A cross-translation comparison of these two solutions, in which the position of one molecule was fixed at $y = 0$ and the other molecule was moved through the unit cell to establish relative positioning, yielded a single peak with a correlation coefficient of 0.266.

After positioning of the two N domains, a translation search was carried out for the top 20 rotation solutions of the C domain. As shown in Figure 7, translation of the C domain with fixed positions of the two N domains yielded two strong solutions with correlation coefficients around 0.44.

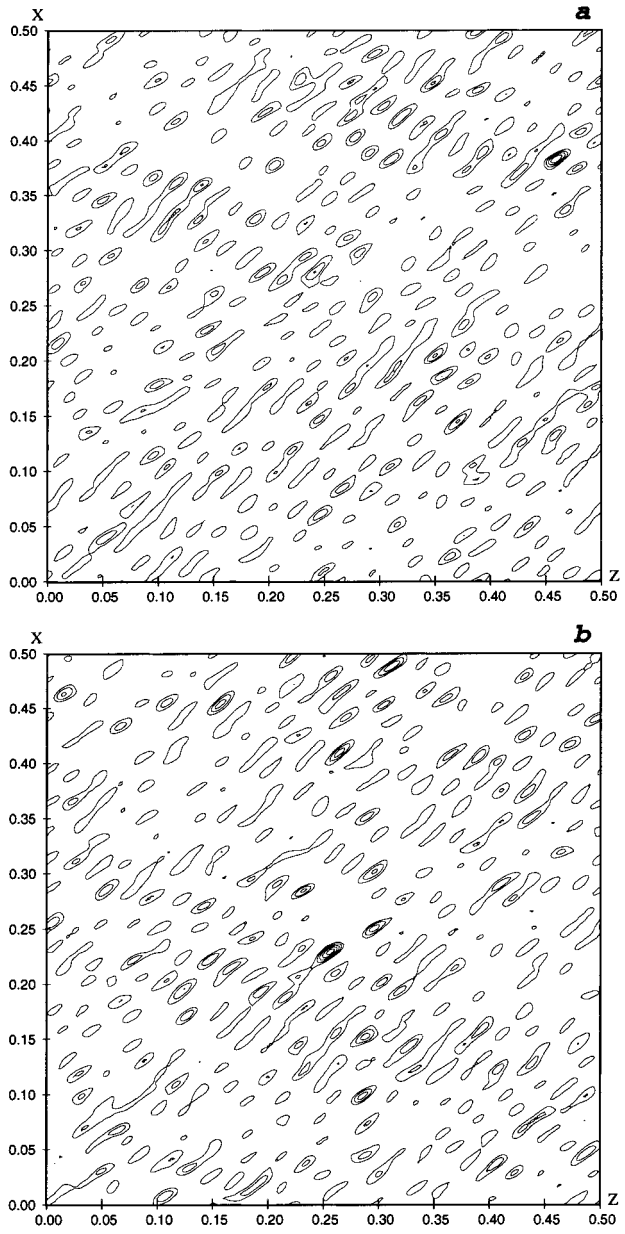

FIGURE 6: Translation search in the *xz* plane for the N domain. (a) $\theta_1 = 270°$, $\theta_3 = 166°$. (b) $\theta_1 = 264°$, $\theta_3 = 11°$. For both solutions, $\theta_2 = 36°$.

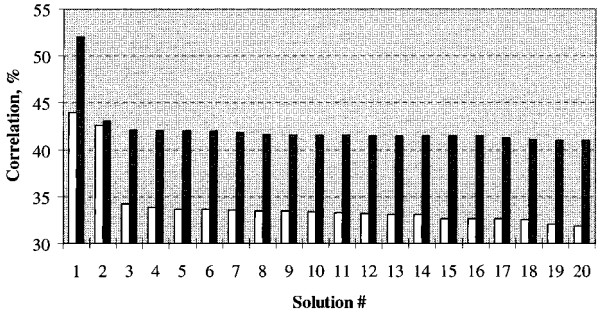

FIGURE 7: Translation search for the C domain after fixing positions of the two N domains. White bars: C domain search with only the two N domains fixed. Note two solutions, one associated with each N domain. Dark bars: Search for the second C domain, with one C and two N domains fixed. Note the unique solution.

Adding the better of these to the two N domains and repeating the translation search yielded only a single peak of height 0.52, which in fact corresponded to the second

Table 2: Torsion Angle Dynamics at 2.2 Å Resolution[a]

	R/R_{free}					
NCS restraints	1	2	3	4	5	$\langle R_{free} \rangle \pm \sigma$
none	0.329/0.432	0.326/0.426	0.327/0.419	0.324/0.423	0.327/0.424	0.425 ± 0.004
domains separately	0.344/0.403	0.343/0.395	0.342/0.399	*0.335/0.389*	0.343/0.393	0.396 ± 0.005
domains together	0.342/0.390	0.348/0.403	0.337/0.386	0.339/0.404	0.346/0.402	0.397 ± 0.008

[a] NCS restraints were imposed on all atoms of the selected residues. N domain includes residues 6−124; linker region, 128−142; C domain, 156−214. Italic type marks the solution used for further refinement.

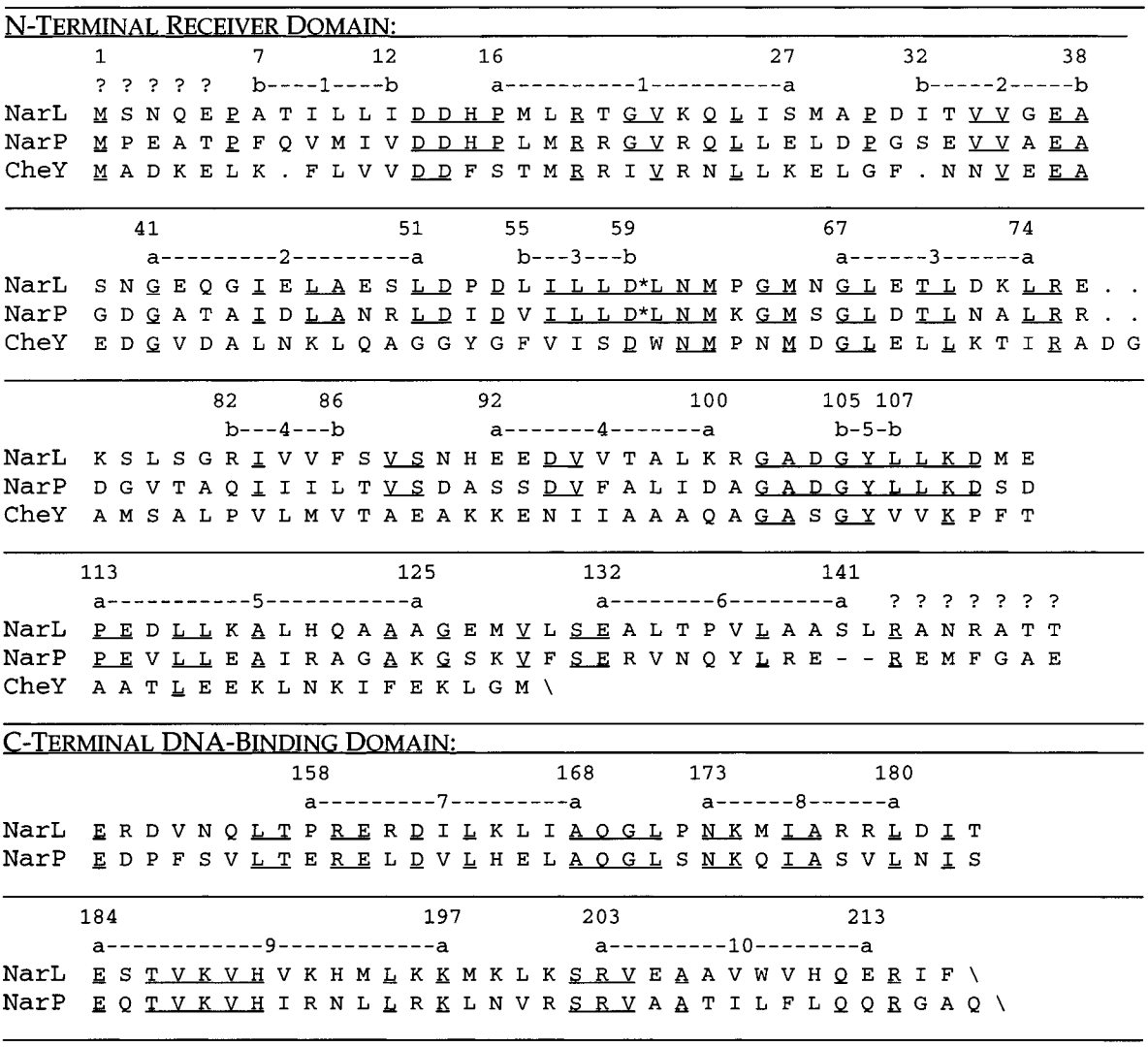

FIGURE 8: Amino acid sequence comparison of NarL, NarP, and CheY. α-Helical regions in NarL are marked and numbered by: a----3----a etc., and β sheet strands by b----3----b etc. ????? indicates the seven residues still not seen in the new monoclinic structure analysis. Residues underlined in NarL and NarP sequences are identical in both. Residues underlined in CheY are identical in all three sequences. Slants mark the carboxy terminus of each chain. The phosphorylated aspartic acid is marked D*.

solution in the prior search. The molecular replacement search was concluded with a rigid-body refinement of all four domains in X-PLOR (*8*), with an *R* factor of 0.427.

Refinement. The monoclinic structure was refined using X-PLOR with the Engh and Huber force field (*9*). Ten percent of the observed reflections were sequestered in a test set for cross-validation by the free *R* factor method (*10*). To eliminate the effect of noncrystallographic relationships on the free *R* factor, reflections for the test set were chosen from a number of thin resolution shells (*11*). Shell radius was distributed randomly in reciprocal space, and shell width was adjusted to cover 10% of the observed reflections with all reflections in each shell included.

The two independent NarL molecules in the monoclinic form will be designated as molecules A and B, and the orthorhombic molecule as O. After rigid-body refinement, orientation of the N domains of A and B differed by rotation of 160° about the *z* axis. The relative positioning of N and C domains in each molecule was similar to that observed in the orthorhombic form, confirming the correctness of the original domain pairing. But detailed positioning of N and C domains differed slightly in all three cases. Superposition of the N domain of orthorhomic molecule O onto the N domain of each of the two monoclinic molecules revealed rotation/translation displacements of [4.3°, 2.4 Å] for O on A, and [1.6°, 1.8 Å] for O on B. Superposition of A on B

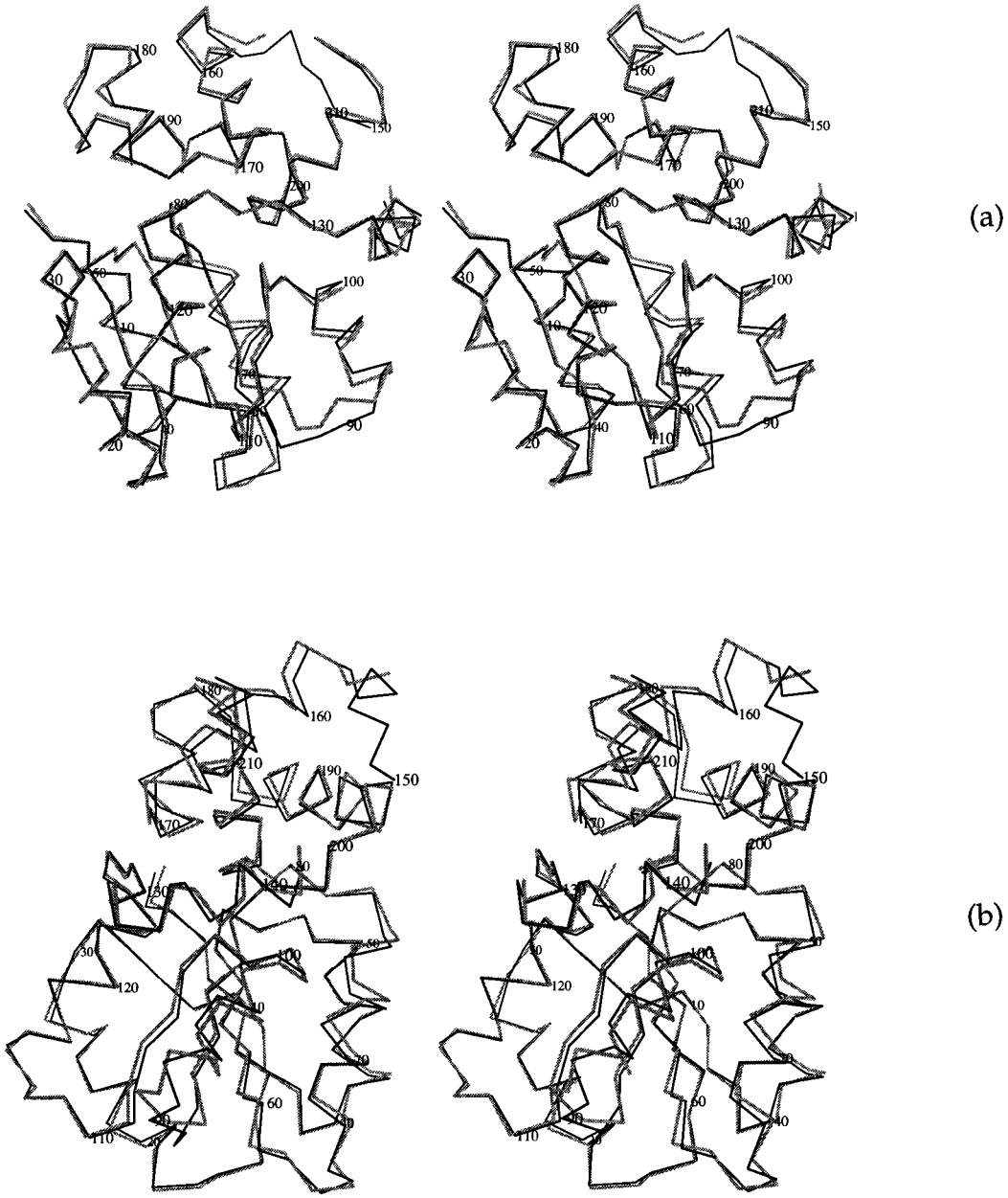

FIGURE 9: Result of superposition of the N domain of orthorhombic molecule O (grey) on monoclinic molecule A (black). Top: NarL viewed from the backside as drawn in Figures 1 and 2, with recognition helix α9 of the C domain at the back of the drawing. Bottom: NarL viewed from the left in Figures 1 and 2, with helix α9 viewed nearly on end at upper right. The monoclinic analysis revealed five more residues in the interdomain chain, 150−154. This leaves only a seven-residue gap in the tether, and confirms the association of N and C domains shown here. Small deviations between O and A are observable throughout the N domain, along with a slight motion of the C domain, which was not used for superposition. The greatest deviations between O and A occur in loop 60−66 and helix α3 which follows, and in loop 87−91 and the following helix α4.

within the monoclinic form yielded displacements of [2.8°, 3.2 Å].

To estimate the significance of this deviation from noncrystallographic symmetry (NCS) relating molecules A and B, a series of torsion angle dynamic simulations (*12*) was carried out using increasingly strong NCS restraints. All residues of the orthorhorhombic structure were included in the refinement, transformed according to the new domain positions, and the linker region (residues 128−142) was associated with the N domain. Temperature factors were fixed at 20 Å^2.

Table 2 lists the results of refinement. Introducing individual NCS restraints for the N domain, linker, and C domain decreased the average R_{free} from 0.425 to 0.396. Imposing a single NCS operator for all domains did not improve the model further. Hence, solution 4 from the series with separately restrained domains was chosen for subsequent refinement. Positional and *B*-factor refinement of this solution yielded R/R_{free} values of 0.302/0.362. Addition of residues 150−154 to the model after examination of the electron density maps using FRODO (*13*) and NCS restraints adjustment to account for lattice contacts reduced R/R_{free} values to 0.286/0.339.

Quality of the Model. The final model includes residues 5−142 and 150−216 in each of 2 NarL molecules, 102 pairs of NCS-related water molecules, and 157 water molecules on the intermolecular interface. Hence, the new structure analysis has not resolved the entire chain connecting the two

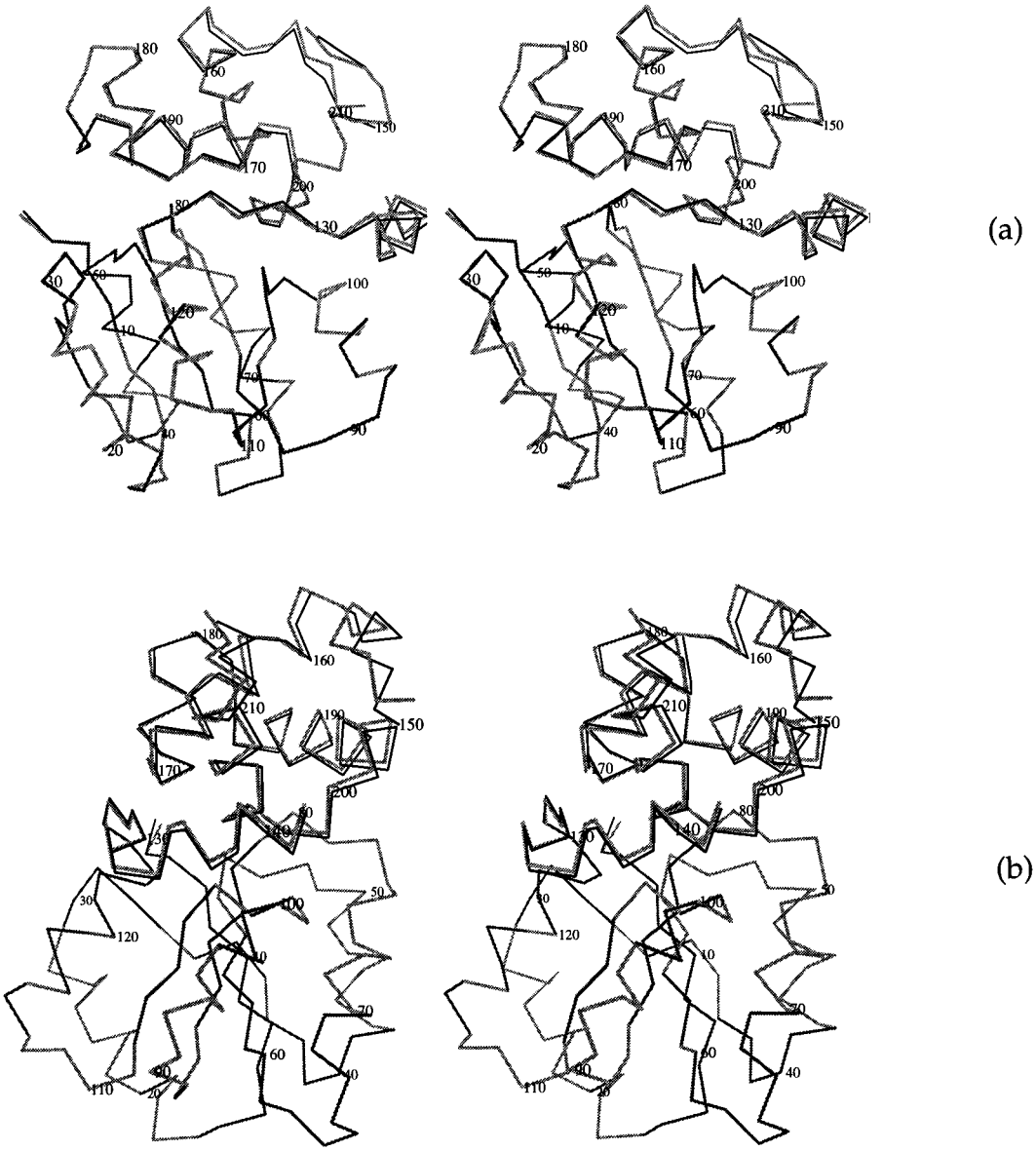

FIGURE 10: Result of comparable superposition of the N domains of monoclinic molecules A (black) on B (grey) Top and bottom views as in Figure 9. N domains are essentially identical, and C domains show only a slight rigid-body displacement from A to B, visible especially in the bottom view. Note that the greatest difference between monoclinic molecules A and B occurs at residues 150−154 of the interdomain connector chain, which were disordered and invisible in the orthorhombic analysis.

domains, but has allowed five more amino acid residues to be located on the C domain end of the connector. R/R_{free} values are 0.211/0.267 for the 22 321 reflections from 8.0 to 2.2 Å, and 0.200/0.257 for the 20 294 strong reflections with $F_o > 3\sigma$. RMS deviations from ideality in the two molecules are 0.009 Å for bond lengths and 1.30 for bond angles. Of 362 non-glycine and non-proline residues, 94.5% are in most favored regions, 1 residue is in a generously allowed region, and the others are in additional allowed regions. Moreover, no glycines or prolines sit in unfavorable conformations according to PROCHECK (14).

RESULTS AND DISCUSSION

Molecular Overlap Comparisons. Figure 8 gives the amino acid sequence of NarL, with those of the related NarP and of CheY for comparison and reference. The N domain is built from an alternation of β sheet strands and α helixes, β1-α1-β2-α2-β3-α3-β4-α4-β5-α5-α6, whereas the C domain is a bundle of four α helixes, α7-α8-α9-α10. In the N domain, the central five-strand β sheet has α helixes packed around it: α1 and α5 on one side: α2, α3, and α4 on the other. The orthorhombic molecule will be designated by O, while the two symmetrically independent NarL molecules in the monoclinic cell will be designated as A and B. Figures 9 and 10 compare stereo superpositions of molecules O upon A, and A upon B, using only the N domains for superposition. A and B are so similar that the superposition of O upon B conveys no new information and is not shown. The molecules are remarkably similar between the two crystal forms and within the monoclinic form. Slight rocking displacements are seen in the C domains, which were not used in superposition. Loops 60−66 and 75−82 at the bottom of Figure 9 are regions of intermolecular contact in both crystal forms, and are displaced slightly.

NarL molecules in their three crystal environments are far more similar to one another than is any one of them to CheY.

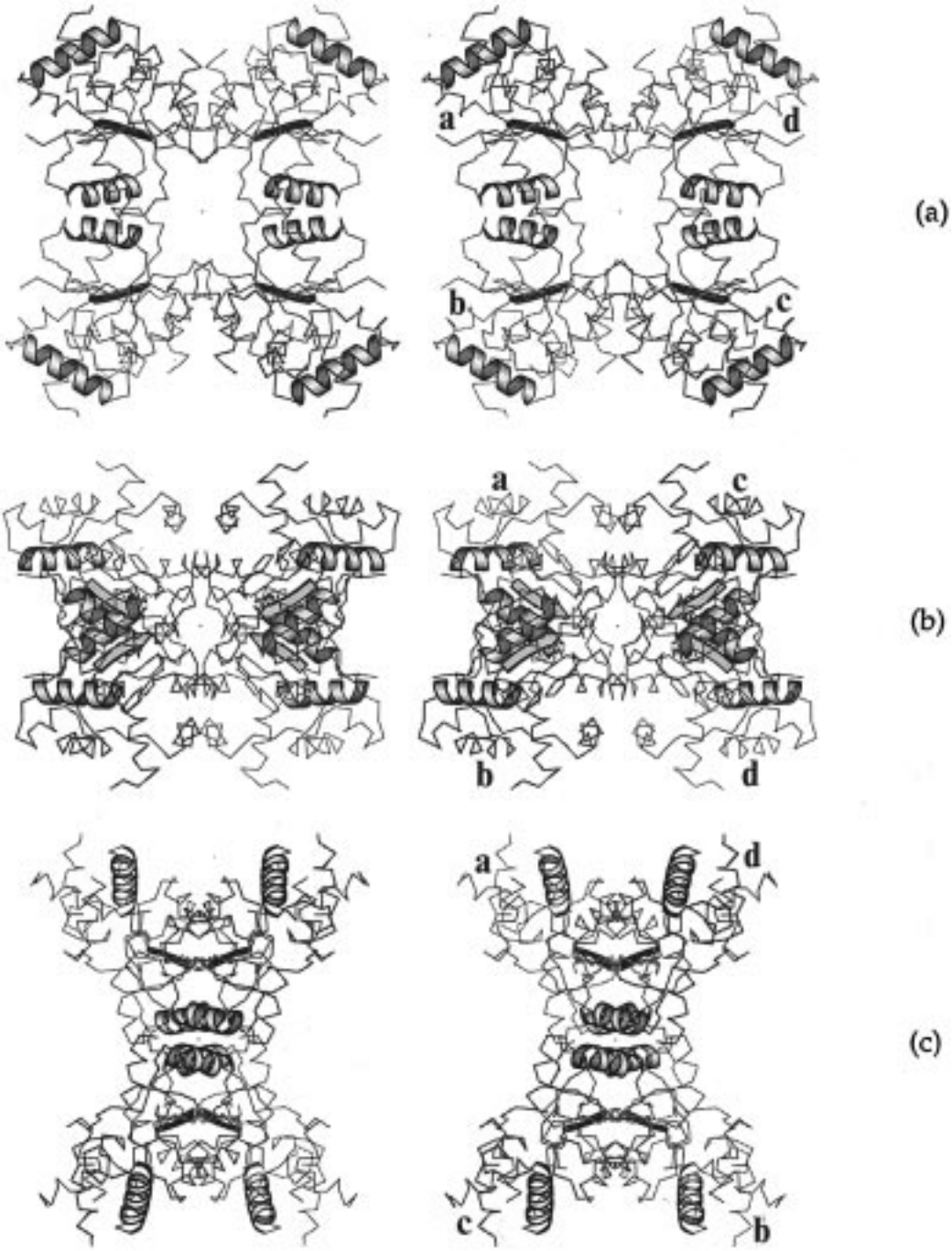

FIGURE 11: Three mutually perpendicular views of the tetrahedral cluster of NarL molecules grouped around the origin and again around the cell center in space group *I*222. The origin (0, 0, 0) or center (1/2, 1/2, 1/2) is marked by a tiny dot at the center of each stereo. Top: View down the *x* axis, with *y* vertical and *z* horizontal. Center: View down the *y* axis, with *x* vertical and *z* horizontal. Bottom: View down the *z* axis, with *x* horizontal and *y* vertical. Each NarL molecule is represented by its α-carbon backbone, with helix α9 of the C domain and helix α1 and strand *β*3 of the N domain emphasized for orientation purposes. N domains pack against one another in the tetramer; C domains extend out and away from the center. Helixes α1 in molecules *a* and *b* cross at nearly a right angle.

Figure 5 of reference *1* shows the stereo overlap of the Cα backbone of CheY with that of the N domain of orthorhombic NarL. The backbone chains, while very similar in the two molecules, show much greater deviations than are seen in Figures 9 and 10 of this paper.

The monoclinic crystal structure, although it did reveal five further ordered residues on the C-terminal end of the interdomain linker, did not localize the entire chain. But in view of the similar arrangement of subunits, discussed below, the association of N and C domains is not in doubt. The portion of the linker still unlocalized in the crystal structures extends from residue 143 to residue 149, with the sequence R-A-N-R-A-T-T. Limited proteolytic cleavage of the two domains (*15*) cuts between Asn145 and Arg146, favoring the idea that this loop of chain is exposed and labile.

In the representative comparison of close relatives of NarL shown in Figure 4 of reference *1*, DegU is seen to have an interdomain loop that is 13 residues longer than that of NarL, whereas UhpA and TrpO have loops that are 10 residues shorter. This may mean that the final turn of helix α6 or the initial turn of helix α7 are unwound in these proteins, to permit the shorter linker to stretch between domains. The

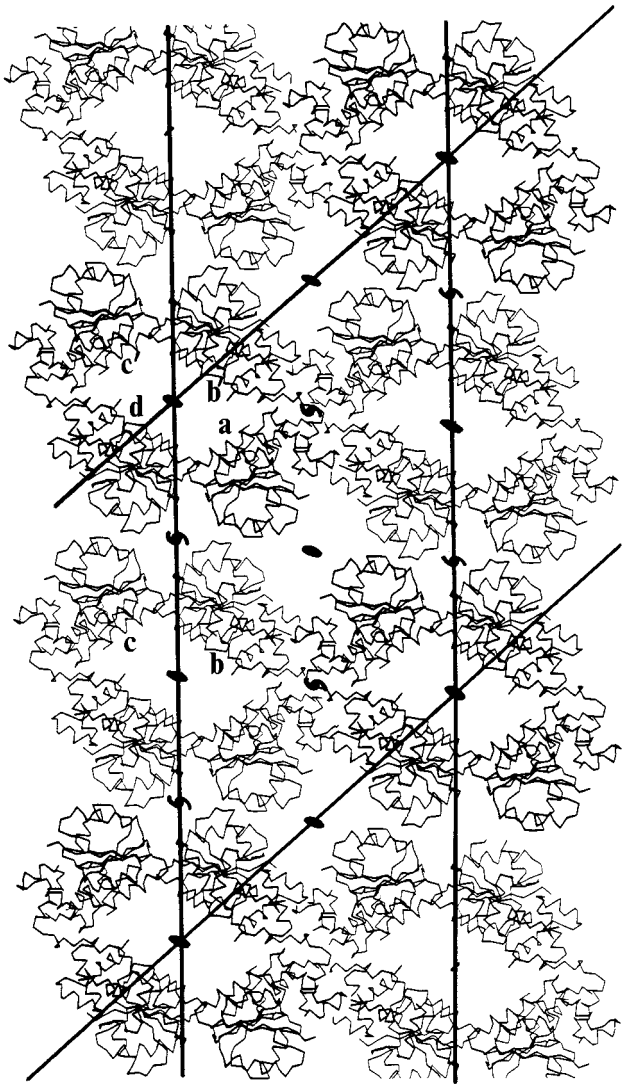

FIGURE 12: Packing diagram of the monoclinic *C*2 cell, with the *x* axis vertical, *y* axis perpendicular to the page, and *z* axis inclined from lower left to upper right. The *z** direction mentioned in the following figure caption is horizontal, perpendicular to both *x* and *y*. (*z* is not perpendicular to *x* and *y* because of the monoclinic β angle of 132°.) Twofold symmetry axes and screw axes are marked. The cell contains two oval rings of four NarL molecules, one ring centered at the origin (dark lines) and the other at (1/2, 1/2, 0) (light lines). Molecules across one ring are related by 2-fold symmetry, and are independent of their two neighbors in the ring. NarL molecules of type A (*a* and *c*′) face one another across the long axis of their oval ring, while molecules of type B (*b*′ and *d*) sit on the *z* axis. Molecules *a* and *c* are related by a 2-fold screw axis involving translation halfway up the *y* axis out of the page, as are *b* and *d*. The conventional monoclinic axes and β angle of 132° obscure the essentially rectangular packing of 4-molecule rings within the lattice.

apparently shorter linker in UhpA and TrpO does not necessarily imply any difference in N and C domain geometry from that seen in NarL.

Orthorhombic Crystal Environment. The structural similarity just evidenced occurs in the face of very different crystal packing environments. The orthorhombic form is constructed from a tetramer of NarL molecules clustered around the origin (0, 0, 0), and an identical cluster at the center of the cell (1/2, 1/2, 1/2). This tetrahedral cluster of four NarL molecules is shown in three stereoviews in Figure 11. The dimer formed between molecules **a** and **b** (or between **c** and **d**) is formed by packing helixes α1 against one another. From the top stereo in Figure 11, the loop between strand β5 and helix α5 appears to be involved as well, but the bottom stereo reveals that this loop interacts only with its own helix α5, not that of the neighbor.

A much looser connection exists between molecules **a** and **d** (or between **b** and **c**). This connection involves only the loop between strand β3 and helix α3 (residues 63−65) and the loop between β4 and α4 (residues 92−94). These loops, however, line the "mouth" of the cavity in which the phosphorylatable Asp59 sits (see Figures 5 and 6 of reference *1*). The four N domains sit "mouth-to-mouth" in the orthorhombic crystal, and it would be difficult to imagine phosphorylating NarL without destroying this tetrahedral arrangement. In contrast, the four C domains extend out to the four corners of the tetrahedron, and do not even make close contacts with neighboring NarL tetrahedra. Intertetrahedral contacts involve the "sides" of the tetrahedral arms, and are weak.

Monoclinic Crystal Environment. Monoclinic packing is related but different. Figure 12 shows a view down the *b* axis, which is so short (38.5 Å) that there is no molecular overlap in this direction. The *C*2 cell contains two rings of four NarL molecules each, centered about the origin (dark lines in Figure 12) and halfway along the *x* axis (light line). The dark rings in Figure 12 all lie in the same plane; neighboring light rings are half the *b* axis length or 19 Å below the plane of the page. Molecules of type A sit along the short diagonal of the cell as drawn here (two examples are labeled **a** and **c**′); symmetrically independent molecules of type B sit across the *z* axis (two examples labeled **b**′ and **d**).

The orthorhombic packing of helixes α1 against one another in adjacent molecules has a parallel in the monoclinic cell. Molecules **a** and **b** (not **b**′) in Figures 12 and 13 pack together via their helixes α1, as do **d** and **c** (not **c**′). Moreover, molecules **a** and **d** interact via the same two loops seen in the orthorhombic case: the loop between β3 and α3, and that between β4 and α4. But the parallel is not exact. Molecules **b** and **c** in Figure 13 interact, not with one another, but with the next molecule up the 2-fold screw axis. The tetrahedron of the orthorhombic cell is opened up into a spiral, although local contacts are the same. Nevertheless, the four molecules **a**, **b**, **c**, and **d** labeled in Figures 12 and 13 can be thought of as the structural "unit" of the monoclinic cell, like the tetrahedral cluster of the orthorhombic cell. Figure 13 shows this "unit" viewed from three mutually perpendicular directions.

The rings that are such striking features of Figure 12 are actually of secondary importance in the monoclinic crystal structure. Within one four-NarL ring of Figure 12, contacts between molecules **a** and **b** are as has just been described. The other contacts around the ring are more tenuous, and involve the top of the C domain. These contacts probably are responsible for the very slight variation in positions of C domains observed between the two crystal forms.

Observations between Crystal Forms. Despite their great difference in cell dimension and symmetry, the two crystal forms use exactly the same types of intermolecular contacts: strong interactions between helixes α1 in the N domains, looser contacts between β3−α3 and β4−α4 loops

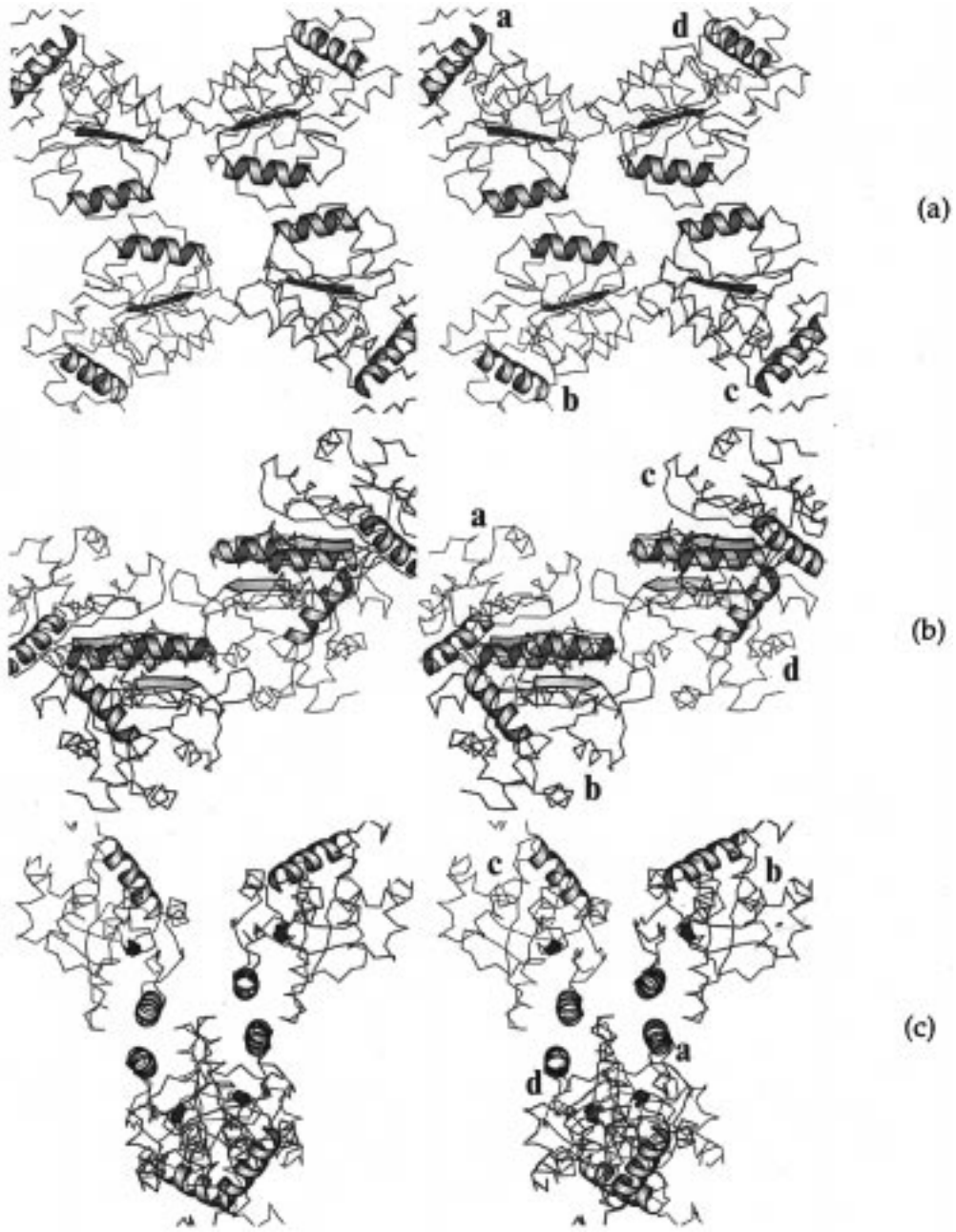

FIGURE 13: Three mutually perpendicular views of the monoclinic form, showing molecules *a*, *b*, *c*, and *d* from Figure 12. Top: View down the *y* axis, with *x* vertical and *z** horizontal. (For explanation of *z**, see previous figure caption.) Center: View down the *x* axis, with *y* vertical and *z** horizontal. Bottom: View down the *z** axis, with *x* vertical and *y* horizontal. Molecules *a* and *d* have the same loop contacts seen in the orthorhombic form. Molecules *b* and *c* make similar contacts, but with partners displaced up the 2-fold screw axis rather than with one another. Contacts between *a* and *b*, and between *c* and *d*, involve close packing of helixes α1, as in the orthorhombic case. This packing is especially visible in the lower stereo.

in the same domains, and tenuous contacts between C domains. The orthorhombic crystal form is very difficult to grow. It was obtained once at the very beginning of crystallization trials, and then never again despite repeated attempts to match the original crystallization conditions. The monoclinic cell is the preferred form, crystallizing easily under a variety of conditions.

The NarL molecule is virtually identical in its two crystal forms. Moreover, NarL seems to show a tendency toward close association, or even dimerization, involving contacts between helixes α1 (Figure 14). The dimerization is exact between two orthorhombic molecules O, whose helixes α1 cross at an angle. Dimerization is very similar although only approximate between monoclinic molecules A and B, whose helixes α1 are more nearly parallel but are displaced by two turns of helix. It is tempting to think that this tendency toward dimerization has a biological role, as discussed below. The two C domains on each dimer are widely separated, and would be good platforms for dissemination of a liberated C domain if our telephone receiver model is correct.

One can even go one step further, and ask whether the displacement along helixes α1 between the two crystal forms might bear some relationship to conformations involved in the activation of NarL by phosphorylation. This is pure speculation, of course, but speculation with a core of hard information. We have observed two stable conformations of a NarL dimer; might these also be functional conformations?

FIGURE 14: Comparison of α1-against-α1 helix packing in the orthorhombic crystal form (above, with true 2-fold axis) and the monoclinic form (below, no symmetry axis). The monoclinic molecule of type A is at the right, and the molecule of type B is at the left. C domains are at the upper left and right corners of the dimers. If these C domains are released by phosphorylation of dimerized N domains, then they could bind to recognition sites removed from one another by as much as two or three helical repeats along the DNA duplex.

Structural and Mechanistic Implications of the Two Crystal Forms. At the beginning of this paper, five unsolved questions were raised. Some, but not all, of these can now be answered by a comparison among the three independent pictures of the NarL molecule. The answers are listed below with the same numbering as before, although not in the same order.

(Query 4): Correctness of Pairing of N and C Domains. Comparison of the orthorhombic molecule and the two monoclinic molecules demonstrates conclusively that the correct pairing of domains is that shown in Figures 1, 2, 9, and 10. No other association of an N with a C domain will fit all three NarL images. Moreover, the visibility of five more residues of the interdomain linker makes the question of proper domain pairing moot.

(Query 5): Structural Integrity of the Two Domains. The near-identity of superposition of molecules O on A in Figure 9, and A on B in Figure 10, means that any deformation of domains resulting from crystal packing is small indeed, and of no mechanistic significance.

(Query 1): C Domain as a HTH Cluster, Blocked by Association with the N Domain. This query cannot be answered with perfect assurance until we finally have a crystal structure of a complex of the C domain with a DNA duplex. But the inference that the C domain would bind to DNA in a manner that requires it first to dissociate from the N domain is a strong one, and is consistent with the structures observed with the three intact NarL molecules. Figure 15 shows the hypothetical DNA binding that would result, if the C domain of NarL were superimposed on the DNA-binding domains of four different HTH proteins in their DNA complexes: paired protein, lambda repressor, CAP, and Trp repressor. Superposition of NarL helixes α7, α8, and α9 on helixes α1, α2, and α3 of *paired* protein leads to a rms deviation of only 0.83 Å for 38 contiguous α-carbon atoms. In each case, the N domain of NarL would have to move away from the C domain before such DNA binding became possible. But the near-identity of folding of the C domain of NarL with these other HTH proteins makes any other answer to the DNA-binding question unlikely. The "telephone receiver" model of NarL function seems inevitable.

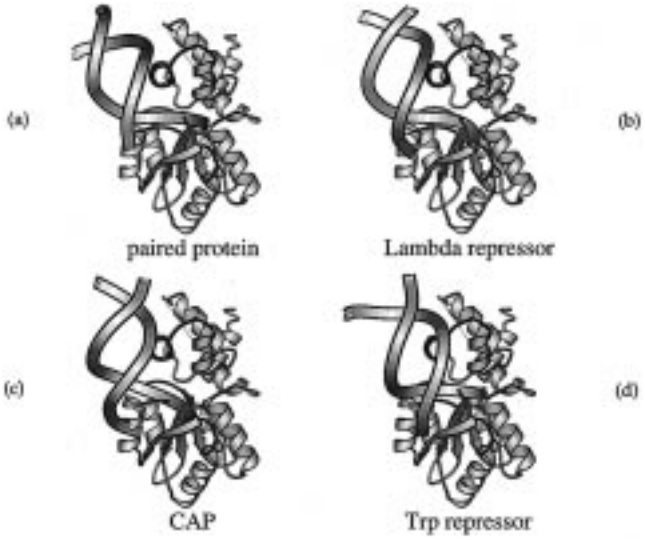

FIGURE 15: Comparison of NarL DNA interaction with that of four other helix−turn−helix or HTH proteins. In each case, the entire NarL molecule is shown, with its C domain oriented against a DNA duplex in the manner observed in the test protein in question. This has been accomplished by least-squares superposition of the HTH motif of the NarL C domain onto the DNA-bound complex of the test protein. In each case, such an association is completely impossible for an intact NarL molecule because of steric clash between DNA and the N domain, but the C domain alone could form an excellent complex. (a) Paired protein (*24*, *25*); (b) lambda repressor (*26*); (c) catabolite activator protein (*27*, *28*); (d) Trp repressor (*29*). Dotted lines in each diagram indicate the unseen tether between domains in NarL.

(Query 3): Geometry of Cooperative Binding of NarL to DNA. The query here is whether, in the light of footprinting studies such as in Figure 3, NarL could be said to bind to DNA (a) as a homodimer, leading to adjacent antiparallel or inverted DNA-binding sites; (b) as a linear array of monomers, leading to adjacent parallel or tandem binding sites; or (c) as isolated monomers, leading to scattered binding sites of no fixed relationship. This question has not been answered. But in both crystal forms, NarL has a strong tendency to form local dimers involving packing of helixes α1 against one another (Figure 14). This may be a clue that NarL binds at least in part as an N domain homodimer. If so, the flexible connection between domains means that the two C domains of the dimer could bind to antiparallel sites separated from one another by as much as two or three helical turns along the DNA duplex.

In this connection, it is worth comparing NarL with another member of the response regulator family, the *Escherichia coli* UhpA protein. UhpA has strong sequence homologies with NarL (Figure 4 of reference *1*), and presumably the same two-domain molecular structure. Like

NarL, it binds DNA following phosphorylation of the same aspartic acid: Asp54 in UhpA vs Asp59 in NarL. Webber and Kadner (*16*) show that phosphorylation of UhpA promotes oligomerization along the surface of a DNA helix, and that this oligomerization occurs via contacts between helixes α1 in the UhpA molecules.

CheY has only a single N domain, folded like that of NarL (*17*−*21*). After phosphorylation of Asp57, it dephosphorylates by two routes: a slow autodephosphorylation, and a rapid dephosphorylation involving binding to CheZ. Sanna et al. (*22*) have shown that intermolecular contacts between CheY and CheZ involve the second half of helix α1 of CheY, again implicating this particular helix in dimerization. It is highly likely that NarL dimerizes in vivo via α1 contacts, and that the intermolecular crystal contacts seen in Figure 14 reflect this tendency.

(Query 2): Mode of Signal Transmission between N and C Domains. Of all the unsolved queries, this one remains completely unresolved. It is possible to imagine that the changes in side chain positions within the activation cavity between the two crystal forms have some bearing on conformational changes induced by phosphorylation, and that such conformational changes might induce the C domain to rotate or to move away on its tether, free to bind DNA. But this is pure speculation, and must be supported (or disproved) by future research.

REFERENCES

1. Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R. P., and Dickerson, R. E. (1996) *Biochemistry 35*, 11053−11061.
2. Tyson, K. L., Bell, A. I., Cole, J. A., and Busby, S. J. W. (1993) *Mol. Microbiol. 7*, 151−157.
3. Otwinowski, Z. (1993) *CCP4 Daresbury Study Weekend*, Daresbury Laboratory, Warrington, WA4 4AD, U.K.
4. French, S., and Wilson, K. (1978) *Acta Crystallogr. A34*, 517−525.
5. Bailey, S. (1994) *Acta Crystallogr. D50*, 760−763.
6. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
7. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
8. Brünger, A. T. (1992a) *X-PLOR: A system for X-ray crystallography and NMR*, Yale University, New Haven, CT.
9. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr. A47*, 392−400.
10. Brünger, A. T. (1992b) *Nature 355*, 472−475.
11. Kleywegt, G. J., and Jones, T. A. (1995) *Structure 3*, 535−540.
12. Rice, L. M., and Brünger, A. T. (1994) *Proteins: Struct., Funct., Genet. 19*, 277−290.
13. Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268−272.
14. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
15. Rech, S., Hemschemeier, K., Jarvis, M., and Gunsalus, R. P. (1997) (in preparation).
16. Webber, C. A., and Kadner, R. J. (1997) *Mol. Microbiol. 24*, 1039−1048.
17. Stock, A. M., Mottonen, J. M., Stock, J. B., and Schutt, C. E. (1989) *Nature 337*, 745−749.
18. Volz, K., and Matsumura, P. (1991) *J. Biol. Chem. 266*, 15511−15519.
19. Bellsolell, L., Priet, J., Serrano, L., and Coll, M. (1994) *J. Mol. Biol. 238*, 489−495.
20. Ganguli, S., Wang, H., Matsumura, P., and Volz, K. (1995) *J. Biol. Chem. 270*, 17386−17393.
21. Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D., and Petsko, G. A. (1993) *Biochemistry 32*, 13375−13380.
22. Sanna, M. G., Swanson, R. V., Bourret, R. B., and Simon, M. I. (1995) *Mol. Microbiol. 15*, 1069−1079.
23. Li, J., Kustu, S., and Stewart, V. (1994) *J. Mol. Biol. 241*, 150−165.
24. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell 82*, 709−719.
25. Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell 80*, 639−650.
26. Beamer, L. J., and Pabo, C. O. (1992) *J. Mol. Biol. 227*, 177−196.
27. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1992) *Science 253*, 1001−1007.
28. Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H., and Berman, H. M. (1996) *J. Mol. Biol. 260*, 395−408.
29. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature 335*, 321−329.

BI972365A