Primary and Secondary Modes of DNA Recognition by the NarL Two-Component Response Regulator[†,‡]

Ann E. Maris,[*,§,||] Maria Kaczor-Grzeskowiak,[⊥] Zhongcai Ma,[@] Mary L. Kopka,[⊥] Robert P. Gunsalus,[⊥,@] and Richard E. Dickerson[§,⊥]

Department of Chemistry and Biochemistry, Molecular Biology Institute, and Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, California 90095

Received April 21, 2005; Revised Manuscript Received July 31, 2005

ABSTRACT: NarL is a model response regulator for bacterial two-component signal transduction. The NarL C-terminal domain DNA binding domain alone ($NarL^C$) contains all essential DNA binding determinants of the full-length NarL transcription factor. In the full-length NarL protein, the N-terminal regulatory domain must be phosphorylated to release the DNA binding determinants; however, the first $NarL^C$−DNA cocrystal structure showed that dimerization of $NarL^C$ on DNA occurs in a manner independent of the regulatory domain [Maris, A. E., et al. (2002) *Nat. Struct. Biol. 9*, 771−778]. Dimerization via the $NarL^C$ C-terminal helix conferred high-affinity recognition of the tail-to-tail promoter site arrangement. Here, two new cocrystal structures are presented of $NarL^C$ complexed with additional 20mer oligonucleotides representative of other high-affinity tail-to-tail NarL binding sites found in upstream promoter regions. DNA structural recognition properties are described, such as backbone flexibility and groove width, that facilitate $NarL^C$ dimerization and high-affinity recognition. Lys 188 on the recognition helix accommodates DNA sequence variation between the three different cocomplexes by providing flexible specificity, recognizing the DNA major groove floor directly and/or via bridging waters. The highly conserved Val 189, which enforced significant DNA base distortion in the first cocrystal structure, enforces similar distortions in the two new cocrystal structures. Recognition also is conserved for Lys 192, which hydrogen bonds to guanines at regions of high DNA helical writhe. DNA affinity measurements for model NarL binding sites, including those that did not cocrystallize, suggest a framework for explaining the diversity of heptamer site arrangement and orientation.

The nitrate-responsive Nar two-component regulatory system in *Escherichia coli* is representative of a broad class of bacterial signal transduction systems (*1*) that modulate cell metabolism in response to environmental and/or intracellular signals. Related two-component regulatory systems are present in many bacteria and archaea and occur in isolated examples in eukaryotes (*2*). Control may be exerted at the level of enzyme activity, gene regulation, or cell motility and behavior. In *E. coli*, the Nar two-component system regulates a large family of cell respiration and fermentation processes by controlling the expression of nearly 100 genes in a nitrate-dependent fashion. In the presence of extracellular nitrate, genes needed for nitrate respiration are switched on whereas genes needed for less energetically favorable pathways such as fumarate reduction or fermentation of simple sugars are switched off. Signal detection occurs in the periplasmic space of the cell by either of the two membrane-bound sensors, NarX or NarQ (*3*). Signaling across the cell membrane and into the cytosol occurs when external nitrate levels are sufficient to activate the autokinase activity located at the C-terminus of each sensor (*4*). Either sensor then serves as a phosphoryl donor, activating the dual response regulatory proteins NarL and NarP for subsequent DNA recognition and transcription regulation. The readout of the Nar two-component signaling system by DNA recognition results in positive or negative control of transcription at multiple respiratory and fermentation pathway operons (*5*−*7*). Whereas little is known about NarP operation, NarL has been a productive model for understanding the general principles of response regulator function at the molecular level.

X-ray structures of two crystal forms of full-length NarL have been determined previously, which provided the first structural model for understanding the FixJ/NarL class of response regulators (*8*, *9*). NarL is composed of two domains (Figure 1A), the N-terminal regulatory domain ($NarL^N$)[1] and the C-terminal output domain ($NarL^C$). They are joined by a 10-residue helix (helix 6) and a 12-residue flexible linker.

[†] Supported by National Institutes of Health Grants GM31299 and AI21678.

[‡] Both diffraction F_o data and final refined coordinates have been deposited in the Protein Data Bank as entries 1ZG5 (*narG* −89/−89) and 1ZG1 (*nirB* −74/−65).

* To whom correspondence should be addressed: Physical Biosciences Division, Lawrence Berkeley National Laboratory, Building 3-350E, One Cyclotron Road, Berkeley, CA 94720. Telephone: (510) 486-7315. Fax: (510) 486-6059. E-mail: aemaris@lbl.gov.

[§] Department of Chemistry and Biochemistry.

[||] Current address: Physical Biosciences Division, Lawrence Berkeley National Laboratory, Building 3-350E, One Cyclotron Road, Berkeley, CA 94720.

[⊥] Molecular Biology Institute.

[@] Department of Microbiology, Immunology, and Molecular Genetics.

10.1021/bi050734u CCC: $30.25
Published on Web 10/15/2005

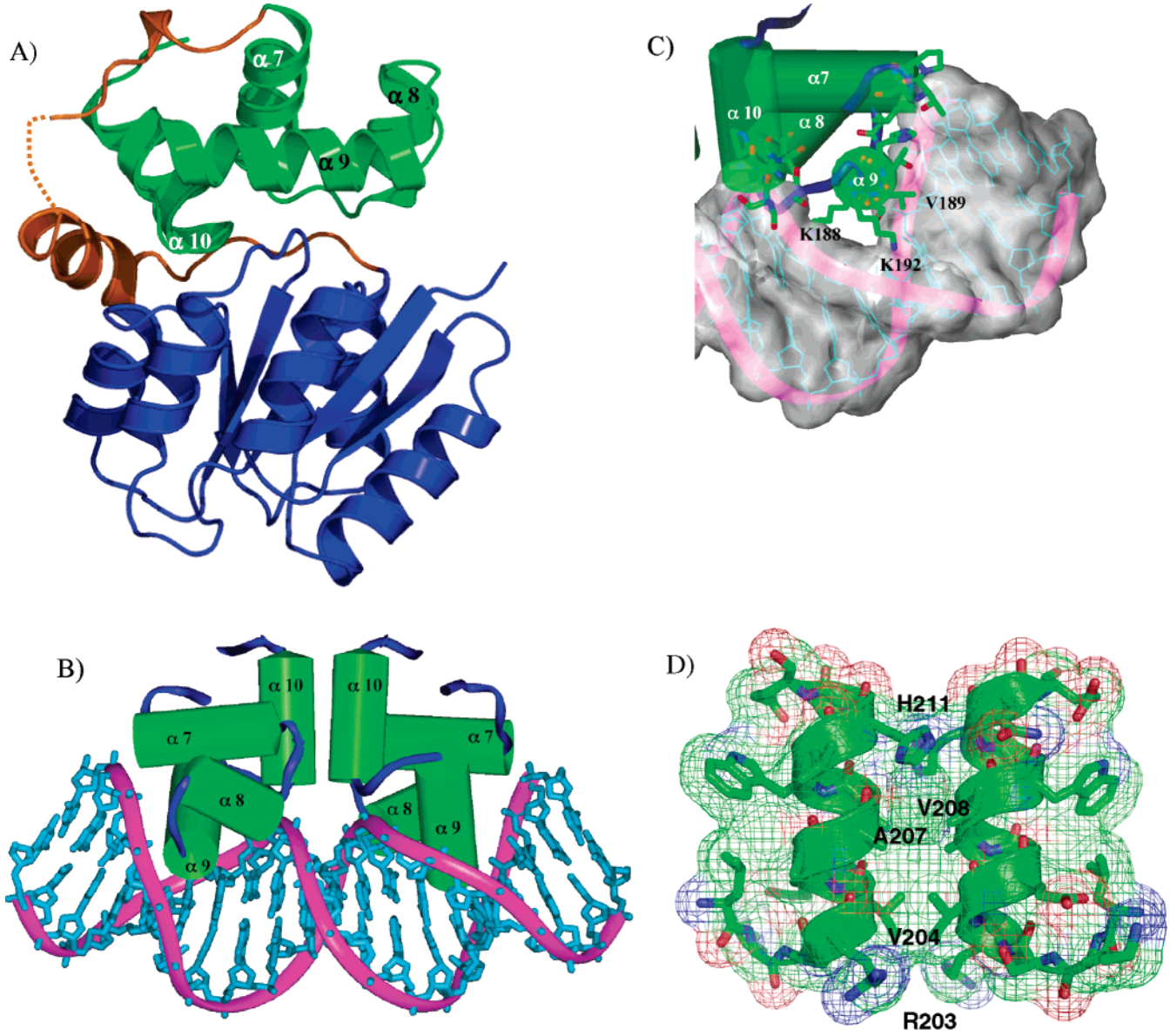

FIGURE 1: (A) Structure of the *E. coli* NarL response regulator (*9*). The NarL^N N-terminal domain is colored blue, while the NarL^C C-terminal domain is colored green. The joining region (gold) consists of 10-residue helix 6, the six preceding residues, and the 12-residue flexible linker. Residues for which no electron density was visible are indicated with a dashed line. (B) Crystal structure of the 2:1 NarL^C−DNA complex. Dimerization via helix 10 of each NarL^C domain inserts recognition helix 9 into consecutive DNA major grooves. (C) Extensive protein−DNA contacts in a view down helix 9. Residues with a main chain or side chain contacting DNA in any way are shown in stick representation. The three residues whose side chains contact the floor of the major groove, Lys 188, Val 189, and Lys 192, are labeled. (D) Close-up of the helix 10 dimer interface with the van der Waals surface representation. The loop region between helices 7 and 8 also contributes to the dimer interface. Both helix 10 and the loop region were involved in the unactivated full-length NarL interdomain interface (*9*). Panels A and D were rendered with Pymol (*36*). Panels B and C were generated using InsightII (Biosym, San Diego, CA).

NarL^N is typical of the activation domains of other response regulators, including CheY, that consist of an alternating series of five α helices and five *β* sheet strands (*10*). Phosphorylation of Asp 59, located in an acidic pocket of NarL^N (Figure 6 of ref *9*), releases NarL^C for DNA binding. The mechanism of NarL activation is poorly understood, but must involve aspects of the flexible linker region and the unmasking of the NarL^C domain that docks to the NarL^N domain in the unphosphorylated state (refs *9*, *12*, and *13* and Figure 5 of ref *11*). The NarL DNA recognition determinants reside within the four-helix bundle C-terminal domain (NarL^C) (*9*, *11*). Helices 8 and 9 form a helix−turn−helix motif common to many DNA binding proteins (Figure 1B,C). Helix 10 and the loop region between helices 7 and 8 form the dimerization interface (*11*) (Figure 1B,D). Three residues contact the floor of the DNA major groove.

The NarL consensus recognition sequence consists of the heptameric element TACYYMT (where Y is a pyrimidine and M is cytosine or adenine) (*5*, *6*, *14*−*16*). The NarL heptamer is found in different numbers and arrangements in the upstream regions of many nitrate-regulated promoters (Figure 2). Depending upon the heptamer location and arrangement, NarL can repress, activate, or derepress transcription in an operon-specific manner (*14*). The 5′-TA end

[1] Abbreviations: NarL^C, NarL C-terminal DNA binding domain; NarL^N, NarL N-terminal receiver domain; EMSA, electrophoretic mobility shift assay; M, adenine or cytosine; K, guanine or thymine; Y, pyrimidine; R, purine.

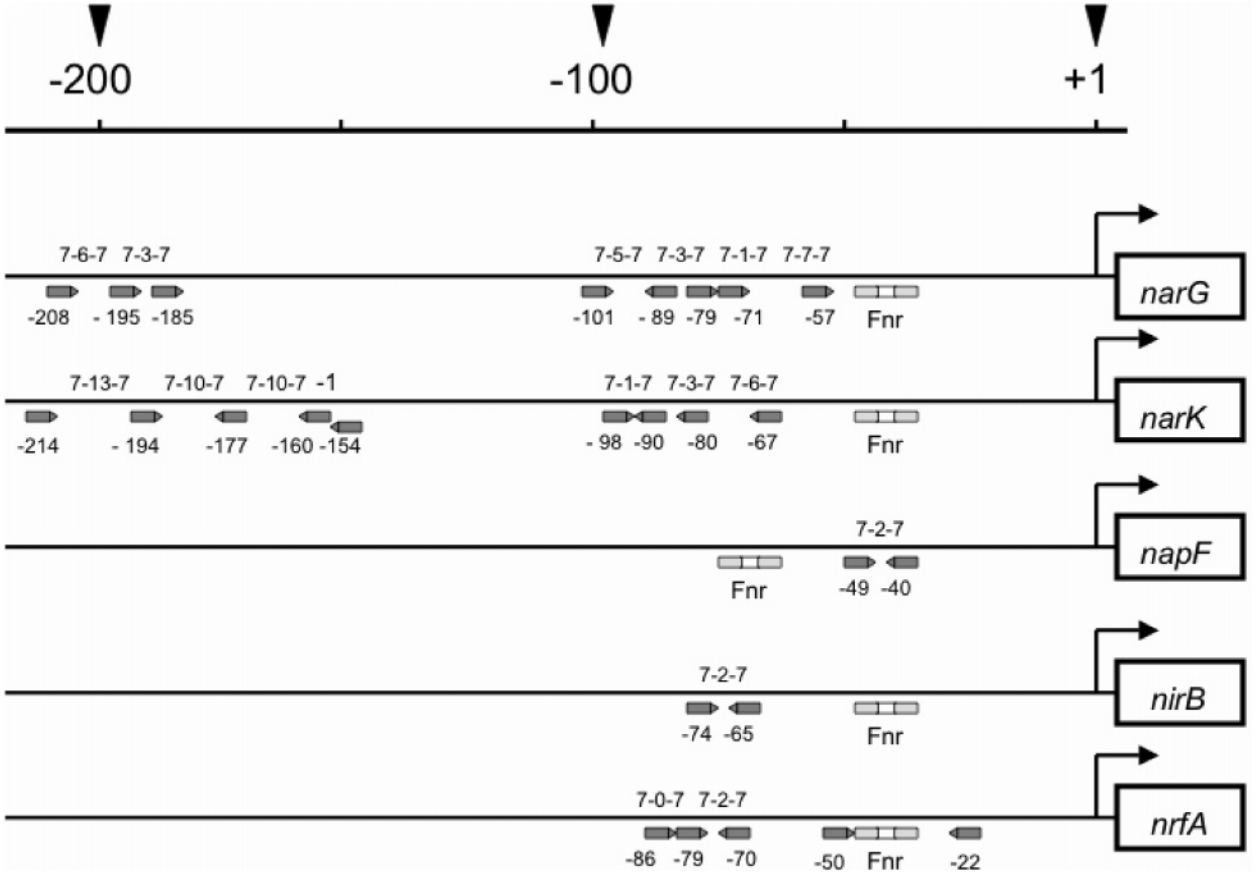

FIGURE 2: Arrangement of NarL heptamer sites at NarL-activated and NarL-repressed promoters. The number, location, and orientation of heptamer sites are indicated by the box arrows. Each box arrow below the line represents a 7 bp recognition site of the 5′TACYYMT3′ sequence. When these are grouped in pairs, then 7−n−7 above the line denotes a possible NarL oligomerization site with *n* base pairs between heptamers. The numbering is relative to the start of transcription of the indicated genes. FNR recognition sites are also indicated for anaerobic induction of the genes by the CAP homologue, FNR.

of the consensus sequence is termed the head, and the 3′-MT end is termed the tail. The NarL heptamers occur in tandem (tail-to-head), antiparallel (head-to-head or tail-to-tail), or singly (half-site). The spacing between head-to-head and tail-to-head heptamer site arrangements is often one to three bases, denoted by 7−1−7 to 7−3−7, respectively, in Figure 2, whereas tail-to-tail sites often have a spacing of two bases (the 7−2−7 arrangement). DNA sequences of this study (Table 1) represent commonly found genomic sites (Figure 2).

Unresolved is a molecular understanding of NarL recognition and binding at the diverse NarL-controlled promoters. To compare NarL recognition and binding between the different arrangements of NarL heptamer sequences, synthetic oligonucleotides representative of native NarL-regulated promoter sites were designed. Only cocrystallization trials containing the tail-to-tail arrangement resulted in cocomplex crystals. Similarly, micromolar EMSA studies used to test crystallization conditions yielded clearly shifted cocomplex bands only for the tail-to-tail arrangement (*11*). Quantitative $NarL^C$ binding affinities for the synthetic binding sites are presented here, along with two new cocrystal structures containing tail-to-tail promoter sites. The first new cocrystal structure contains a native tail-to-tail arrangement (*nirB* −74/−65), and the second contains a *narG* promoter heptamer (*narG* −89/−89) arranged in the tail-to-tail orientation (Table 1, sequences B and C, respectively). These structures, along with that reported previously (Table 1, sequence A; ref *11*), reveal how DNA sequence variation is accommodated by NarL in accomplishing specific high-affinity recognition.

MATERIALS AND METHODS

Production of NarL Protein. The N-His_6-tagged NarL (147−216) protein was prepared as described previously (*11*).

Production of Synthetic NarL Binding Sites. The DNA sequences (Table 1A−F) were synthesized at the 5 μmol scale by the solid-phase phosphoramidate method on an Eppendorf ECOSYN D300 synthesizer. After the 5′-dimethoxytrityl protecting group had been removed, the DNA solution was dried by rotating evaporation under vacuum, resuspended in 7 M urea, and purified using 32 cm × 52 cm × 0.25 cm PAGE (7 M urea, 1× TBE, and 20% acrylamide, 19:1). The desired DNA bands were visualized by brief UV shadowing and excised from the gel. Electroelution was performed using an S&S Elutrap separation system. Purified DNA was desalted by methanol elution from C8 resin. DNA was lyophilized and resuspended in water prior to annealing in 25 mM magnesium acetate. DNA annealing was performed with an oligonucleotide concentration of approximately 10 mg/mL. Palindromic sequences were not always annealed. DNA duplexes were filtered prior to complex formation with 0.45 μm spin filters.

Crystallization. Crystals were grown as described previously (*11*). Briefly, 24-well Linbro plates were used for 5−20

Table 1: Sequence and Arrangement of DNA Heptamers Used in This Study

	Sequence[b,c,d]	half-site arrangement	heptamer site	Kd[e] (uM)
A	CGTACCCATTAATGGGTACG GCATGGGTAATTACCCATGC	Tail to Tail 7-2-7	*nirB* -74/-74 = *narG* -185/-185	0.65
B	CGTACCCATTAAGGAGTACG GCATGGGTAATTCCTCATGC	Tail to Tail 7-2-7	*nirB* -74/-65 = *narG* -185/-195	0.45
C	CGTACCCCTATAGGGGTACG GCATGGGGATATCCCCATGC	Tail to Tail 7-2-7	*narG* -89/-89	0.68
D	CGTACCCCCGATCGGGGTACG GCATGGGGGCTAGCCCCATGC	Tail to Tail 7-2-7	*napF* -49/-40	0.45
E	CGAGGGGTATCTTAGGAATCG GCTCCCCATAGAATCCTTAGC	Head to Head 7-3-7	*narG* -89/-79	2.3
F	CGTACTCCTTAATACCCATCG GCATGAGGAATTATGGGTAGC	Tail to Head 7-3-7	*narG* -195/-185	1.3
G	GATACCCCTACTTAAGATCT CTATGGGGATGAATTCTAGA	Heptamer 7	*narG -89*	2.5
H	GATACCCCTAAAGGGGTATC CTATGGGGATTTCCCCATAG	Tail to Tail 7-2-7	*narG* –89/-89	0.38
I	CGTACCCATATAATGGGTACG GCATGGGTATATTACCCATGC	Tail to Tail 7-3-7	*nirB* -74/-74 = *narG* -185/-185	1.9

[a] The arrows below the boxes point in the 5′−3′ direction of each heptameric half-site. Each arrow represents a 7 bp recognition module of 5′TACYYMT3′ sequence. These are grouped in pairs where the spacer is two or three bases between heptamers. The numbering above and below the consensus 7−2−7 sequence indicates the base position of the 20mers. [b] Protein−DNA structures have been determined for sequences A−C. [c] The base pairs that differ between the determined crystal structures (A−C) are colored red. [d] Sequence G contains a single heptamer, in contrast to the heptamer pairs in all other sequences. [e] Apparent K_d values are listed for each sequence as determined by a gel shift assay with NarL^C (Figure 4).

μL hanging drops in which protein and DNA were mixed using a 2:1 protein:DNA molar ratio, with 50% reservoir solution in the initial drop. The order of addition was important. Reservoir was first mixed with DNA to maintain a minimum of 0.35 M ammonium sulfate, which prevented precipitation upon complex formation. The NarL^C concentration in the initial drop was 3.5−4.0 mg/mL to favor large crystal formation. The NarL^C concentration prior to drop addition was approximately 16 mg/mL in 9 mM Tris (pH 7.6), 70 mM KCl, 0.6 mM $MgCl_2$, and 2% glycerol. At room temperature, crystals appeared after 1 day and grew to approximately 1 mm × 1 mm × 0.07 mm plates within 1 week. The crystals were cut to fragments of approximately 0.3 mm × 0.3 mm × 0.07 mm, which reduced diffraction mosaicity arising from cryoprotection.

Data Collection and Structure Refinement. Cryoprotected crystals were prescreened in-house at the Molecular Biology Institute. The platelike crystals had few packing contacts, resulting in four different crystal packing forms showing $P2_12_12_1$ pseudosymmetry. The different packing forms were easily distinguished by comparison of their native Patterson maps (*17*). Selected crystals were shipped to Brookhaven National Laboratory beamline X8C for data collection and processing, as described previously (*11*). The two new complexes presented here were of the same packing form as the earlier structure analysis (*11*), in which the highest crystallographic symmetry was $P2_1$ with the 2_1 axis along the shortest unit cell dimension (Table 2). Each asymmetric unit (the smallest unit that can generate the entire unit cell) contained two protein−DNA complexes. Noncrystallographic symmetry between the two complexes of each asymmetric unit was not enforced. Therefore, four NarL^C protomers and two DNA 20mers were uniquely refined. Refinement was begun using CNS (*18*) rigid body refinement with the previously published NarL^C−DNA asymmetric unit (PBD entry 1JE8). DNA was initially restrained to B-form. DNA difference density was visible in sigma-a weighted ($F_o - F_c$) electron density maps, indicating that the DNA base identities were different from the initial model, as expected. The correct DNA bases were built into density. Minimization and *B*-factor refinement were performed in CNS. Density visualization and model rebuilding were performed using O (*19*). Simulated annealing was performed after each building cycle until its use did not improve refinement. Composite

Table 2: Data Collection and Refinement Statistics

	sequence B (*nirB* −74/−65)	sequence C (*narG* −89/−89)
space group	$P2_1$	$P2_1$
unit cell		
a (Å)	76.58	76.89
b (Å)	52.82	52.57
c (Å)	84.97	84.61
β (deg)	90.07	90.00
no. of complexes per AU	2	2
wavelength (Å)	1.1	1.1
resolution (Å)	25.0−2.25	50.0−2.20
R_{merge} (%)[a,b]	8.8 (48.1)	10.4 (68.7)
completeness (%)[a]	98.4 (98.8)	99.5 (96.4)
$I/\sigma(I)$[a]	18.3 (3.9)	16.4 (2.9)
no. of reflections[a]	32539 (1942)	34483 (2194)
refinement		
resolution (Å)	9−2.3	20−2.3
no. of protein non-H atoms	2192	2192
no. of nucleic acid non-H atoms	1628	1628
no. of water molecules	92	260
no. of sulfate ions	9	12
no. of reflections		
working set	27963	28705
test set (5%)	1490	1495
R-factor (%)[a,c]	25.6 (36.9)	21.7 (33.3)
R_{free} (%)[a,c]	28.6 (40.2)	25.2 (38.4)
rms deviation		
bonds (Å)	0.0059	0.0084
angles (deg)	1.02	1.03
dihedrals (deg)	19.3	19.0
impropers (deg)	1.01	0.99
B-factor (Å^2)		
overall	43.5	38.6
solvent	45.7	38.9
Ramachandran plot (%)[d]		
most favored	97.1	95.7
allowed	2.9	3.3
generously allowed	0.0	1.0
disallowed	0.0	0.0

[a] The numbers in parentheses are for the last shell, which was 2.25−2.20 Å for the *nirB* −75/−65 data, 2.44−2.30 Å for the *nirB* −74/−65 refinement, 2.25−2.20 Å for the *narG* −89/−89 data, and 2.44−2.30 Å for the *narG* −89/−89 refinement. [b] $R_{merge} = \sum|I - \langle I\rangle|/\sum I$, where I is the integrated intensity of a given reflection. [c] *R*-factor = $\sum(|F_o| - k|F_c|)/\sum(|F_o|)$, and R_{free} was the *R*-factor calculated using the 5% test set of randomly selected reflections. [d] Ramachandran plot parameters were calculated with PROCHECK.

omit maps (5%) were often examined to reduce any model bias in refinement. DNA backbone difference density persisted in the flexible noncontacted DNA backbone regions. Modeling alternate phosphate positions did not improve the *R*-factors or the difference density in these regions. The cocrystal structures of the nonpalindromic DNA sequence (*nirB* −74/−65) continued to display $F_o - F_c$ difference density at DNA bases that differed between the two half-sites of the sequence, because the nonpalindromic DNA was oriented in both directions. All permutations of DNA orientations in the unit cell were tested in refinement using both $P1$ and $P2_1$ data, with no markedly better fitting orientation arising. This demonstrated that the nonpalindromic DNA was packed indistinguishably in the crystal in both directions at approximately equal ratios. Using $P2_1$ data, the orientations yielding slightly lowest *R*-factors were chosen for final refinement cycles. The electron density of NarL^C protein side chains contacting the nonpalindromic DNA was clear and unambiguous, and concomitantly changed as one orientation was replaced by the other. Crystals of the nonpalindromic sequences display overall a higher level of disorder than those of the palindromic sequences (Table 2). Structure factors and final atomic coordinates for the NarL^C−*nirB* −74/−65 and NarL^C−*narG* −89/−89 cocrystal structures have been deposited with the Protein Data Bank (entries 1ZG1 and 1ZG5, respectively).

Preparation of Gel Shift Probes from Synthetic NarL Oligonucleotides. Synthetic DNA oligonucleotides containing NarL−heptamer sites in differing arrangements (Table 1) were individually cloned into the pCR4-TOPO vector using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Correct inserts were confirmed by sequencing. To generate DNA fragments for gel shift analysis, PCR amplification was performed with the recombinant pCR4-TOPO vectors as templates to give 241 bp (for the 7−2−7 sites) or 242 bp (for the 7−3−7 sites) fragments containing the NarL binding sites. The PCR products were digested with SpeI that gave a 208 or 209 bp DNA fragment. The vector was similarly amplified and digested to generate the nonspecific binding control oligonucleotides. The control oligonucleotide was the same length as test sequences because 20 additional vector bases were included instead of a NarL binding site. The *Spe*I-digested products were end labeled with [γ-^{32}P]ATP using the Klenow fragment of DNA polymerase.

NarL^C−DNA Native Gel Shift Assays. Gel shift assays were performed at room temperature by incubating 2 nM ^{32}P-labeled DNA probe with N-His$_6$-tagged NarL^C (147−216) for 10 min in the reaction buffer [10 mM Tris, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 5 mM $MgCl_2$, 15 mg/mL poly-dIdC, and 10% glycerol (v/v) (pH 7.5)]. The vector DNA was incubated with 4 μM NarL^C as the nonspecific control. The reaction mixtures were immediately subjected to nondenaturing 6% polyacrylamide DNA retardation gel electrophoresis (Invitrogen) at 100 V and 4 °C for ∼1 h. Gels were quantitatively analyzed with a PhosphorImager system (Molecular Dynamics Inc., Sunnyvale, CA).

RESULTS

General Structure Considerations

Synthetic NarL Binding Sites. Six synthetic oligonucleotides (Table 1A−F) representing functionally important heptamer sites (Figure 2) were prepared to test cocrystallization with NarL^C. Each double-stranded DNA contained two consensus-like heptamers arranged in tail-to-tail, head-to-head, or head-to-tail format (Table 1). For example, the *narG* −185 heptameric site for *narGHJI* activation (Table 1A) is also found in a tail-to-head 7−3−7 arrangement in vivo (−195/−185) (Table 1F). The *narG* −195/−185 heptamers are important in vivo, but the tail-to-head arrangement is recognized weakly in vitro by NarL^C (described below). In contrast, synthetic oligonucleotides containing the 7−2−7 tail-to-tail arrangements of *narG* −185/−195 heptamers exhibit high-affinity binding (Table 1B). The *nirB* −74 palindrome cocrystal structure (Table 1A) has been reported in a previous study (*11*). The *nirB* −74/−65 nonpalindromic sequence (Table 1B) represents the in vivo arrangement. The *narG* −89/−89 palindromic sequence (Table 1C) is a tail-to-tail arrangement of the −89 half-site, which is found in vivo in the *narG* −89/−79 head-to-head arrangement (Table 1E). The *napF* −49/−40 promoter site (Table 1D) is a naturally occurring palindrome. The *narG* −195/−185 tail-to-head arrangement is found in vivo (Table 1F).

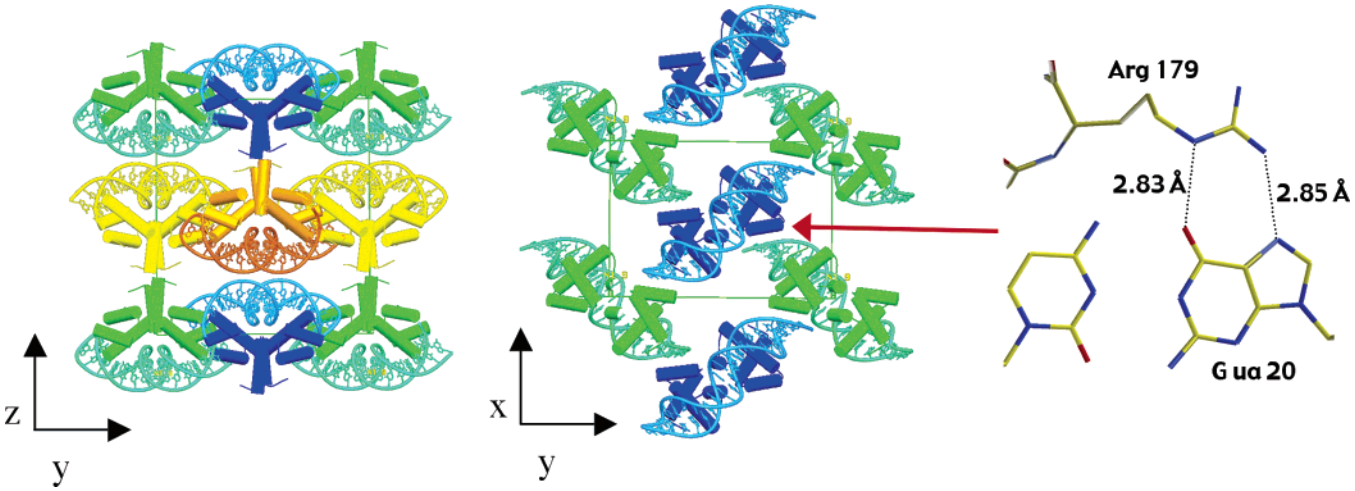

FIGURE 3: Crystal packing of NarL^C−DNA complexes. Individual complexes are isolated units in the crystal, with no direct protein−protein contacts. (A) Stacking of sheets (green and blue, and orange and yellow) along the *z* axis of the crystal. Because of pseudosymmetry, the crystal packing form was first refined in the triclinic *P*1 space group, the lowest possible symmetry (*11*, *17*). The asymmetric unit is the smallest unit that can be repeated to describe the entire crystal. The *P*1 space group had four complexes in each asymmetric unit, colored blue, green, yellow, and orange here. Repeats of the green and blue units formed one sheet in the crystal, and the yellow and orange units formed the next. No contacts are observed between sheets. (B) View of one sheet of green and blue complexes in the *x*−*y* plane. These sheets are stabilized by protein−DNA contacts. The DNA termini from each complex interact with neighboring complexes. Panels A and B were generated using SETOR (*37*). (C) In one of the crystal packing contacts, solvent-exposed R179 forms stable H-bonds with the terminal guanine of a symmetry-related complex. Other packing contacts in this region involve van der Waals interactions and DNA backbone hydrogen bonding. Panel C was rendered with Pymol (*36*).

Crystallization and Molecular Packing. Cocrystallization trials were attempted with all promoter site arrangements shown in Table 1 except G−I. Additional sequences (not shown) were also tested (*11*, *17*). Cocrystals were obtained only with sites containing the tail-to-tail 7−2−7 arrangement (Table 1A−C). Cocrystals were also obtained with sequence D, but the structure has not yet been determined because of cryoprotection issues. The CG bases on either end of the crystallization oligomers are not necessarily those found next to actual NarL sites but have been incorporated to increase termini stability, which may encourage ordered crystal packing. Similarly, the central two bases between the heptamer elements vary among NarL 7−2−7 sites.

Oligonucleotides designed for crystallization often are synthesized as 10mers or 20mers with the expectation that the helices will stack end to end, enhancing crystal growth by forming a semicontinuous helix. Such semicontinuous DNA helices have been observed previously in crystallized complexes containing both bent and straight DNA. The NarL^C−DNA 20mer complex, however, does not form a semicontinuous helix. Instead, the complexes crystallize in sheets (*17*), with no visible protein−protein contacts between sheets (Figure 3A) and few intrasheet contacts (Figure 3B,C). Because of this lack of protein−protein contacts, each NarL^C−DNA complex is almost an isolated entity. Crystal packing interactions involve mainly protein−DNA contacts (*17*). For example, each terminal guanine of the blunt-ended oligonucleotide forms an intrasheet crystal packing contact by hydrogen bonding to symmetry-related Arg 179 (Figure 3C).

NarL Recognition of the Synthetic Binding Sites. Electrophoretic mobility shift assays (EMSAs) were performed to evaluate the relative binding between the different binding site arrangements using short crystallization oligonucleotides (see Figure 1 of ref *11*), and longer radiolabeled 122 bp DNA fragments that contained the same crystallization sequences

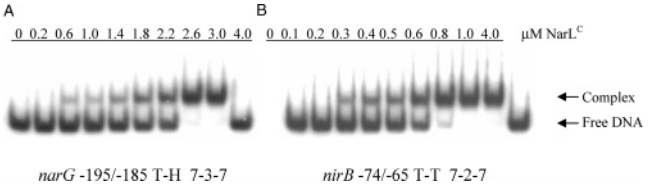

FIGURE 4: Binding site arrangement correlates to NarL^C binding affinity. Native gel shifts were used to determine relative binding affinities (Table 1) between NarL heptamer arrangements. (A) The *narG* −195/−185 tail-to-head arrangement and the *narG* −89/−79 head-to-head arrangement (not shown) are important sites in vivo, but in vitro were weakly recognized by NarL^C. (B) Recognition of the *nirB* −74/−65 tail-to-tail site is representative of the four tail-to-tail sites that were assayed, which demonstrated a 3.5−5-fold higher affinity than the *narG* −195/−185 arrangement, and a 2−3-fold higher affinity than the *narG* −89/−79 arrangement. The *narG* −89 or *narG* −185 half-site in the context of a palindromic tail-to-tail arrangement was bound in a manner similar to that of the *nirB* −74/−65 tail-to-tail site. The concentration of NarL^C is given for each lane. The DNA fragment used in lane 10 of panels A and B contains a vector plasmid that lacks a consensus heptamer element and therefore represents a nonspecific DNA control sequence.

(Table 1 and Figure 4). Affinities for the 7−2−7 arrangements were 3−5-fold higher than for the other NarL heptamer arrangements. The *narG* −89 or *narG* −185 half-site in the context of a palindromic tail-to-tail arrangement was bound similarly to the *nirB* −74/−65 tail-to-tail site. Binding of NarL^C to an isolated half-site (i.e., a single heptamer) is weak (ca. 2.5 μM, Table 1G), and is approximately 3−6-fold weaker than the corresponding 7−2−7 diheptamer arrangements (Table 1C,H). Strikingly, NarL^C serves as a platform for DNA recognition and binding, regardless of the orientation of heptamer pairs (tail-to-tail, head-to-head, or head-to-tail) or their spacing (two or three nucleotides). A nonspecific DNA fragment (Figure 4, lane 10) was not shifted by 4 μM NarL^C, thus demonstrating NarL specificity for the heptamer sequence in pairs or alone.

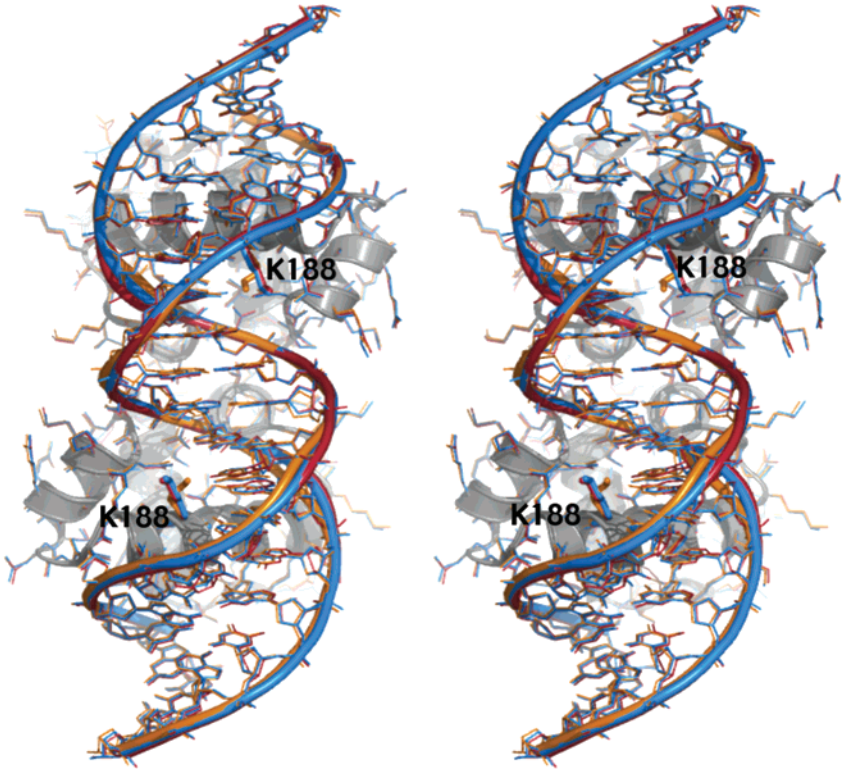

FIGURE 5: Stereo superposition of the three NarL^C−DNA complexes, viewed into the central DNA major groove. The two NarL^C domains are on the far side. The three structures are almost identical. Sequence A is colored blue, sequence B yellow, and sequence C red. The Lys 188 side chains are shown in stick representation. The bases are perpendicular to the DNA helix axis in the central region. At the NarL^C binding sites, the bases writhe around the DNA helical axis, typical of A-form DNA. Most of the DNA backbone flexibility occurs in the noncontacted regions between the NarL^C binding sites and the central dimerization region. This figure was rendered with Pymol (*36*).

Overall Molecular Structure. NarL^C is a monomer in solution (*11*; A. E. Maris, R. P. Gunsalus, et al., unpublished data) but dimerizes similarly on all three tail-to-tail binding sites (Table 1A−C). The structure of the NarL^C complex with sequence A in Table 1 has been reported previously (*11*). This paper presents the structures of complexes with sequences B and C. The three NarL^C−DNA complexes all contain a blunt-ended DNA 20mer in the 7−2−7 binding site arrangement. Their structures overlap but are not completely identical (Figure 5). The overall rmsd between identical DNA bases of the three complexes is less than 0.4 Å, and the rmsd between all equivalent atoms of the six individual NarL^C complexes of the three crystals is approximately 1.0 Å. The significant difference between protein side chains among the complexes is the ability of Lys 188 to flexibly recognize the major groove face as discussed below. Each oligonucleotide shows a gradual bend of approximately 42°, apparently induced by extensive protein contacts. Each binding site undergoes a B-form to A-form transition where the major groove floor is contacted by NarL^C (*11*).

DNA Bending in the 7−2−7-Containing 20mers. Bending in short DNA segments is conveniently examined by means of a normal vector plot (*21*), in which a unit vector is placed perpendicular to the best mean plane through each individual base pair. These vectors are brought to a common origin, and then examined in a view down the mean helix axis (Figure 6). A planar bend is indicated when a succession of points representing the tips of the unit vectors progresses linearly across the field of view. In contrast, a writhe produces a circular track of vector points within the plot. Both types of DNA bending are found in all three NarL^C−DNA 7−2−7 complexes, as shown in Figure 6. In each case, the two heptamer ends exhibit a writhe that is typical of A-DNA, while the central region shows the straight progression typical of smooth, planar bending in B-DNA. It has been known from studies of synthetic oligomers that runs of successive GC base pairs are highly compatible with the A-DNA geometry, and runs of AT base pairs with B (*20*). Hence, the bending observed in Figure 6 is most probably a consequence of the distribution of AT versus GC base pairs along the 20mer. Bends for sequences A and B are nearly identical, but that for C, which has a shorter central A-tract (Table 1), is somewhat more abrupt. Some of the sequence C central region data points are not visible because there is little bend between them, leaving them superimposed. Whether this reflects a significant difference or is to be ascribed to crystal disorder is not clear.

Local Helix Parameters and Helix Type. Figure 7 shows local helix parameters along the full 20mer helices: zp, the distance between successive phosphate atoms; MGr and mGr, the widths of major and minor grooves, respectively; P, an indicator of puckering of the sugar rings; and vroll, a measure of the relative rotation of two adjacent base pairs around their long axes (*21*−*24*). All of these parameters emphasize the A-DNA character of the GC-rich heptamer regions and the B-DNA character of the central AT-rich zone. The interphosphate distance is typically greater for A-DNA than for B-DNA. The minor groove in A-DNA is broad and shallow, whereas in B-DNA, it is deeper and narrower. Sugar puckering in A-DNA typically is C3′-endo (low P) but in B-DNA is more often C2′-endo (high P). And finally, the

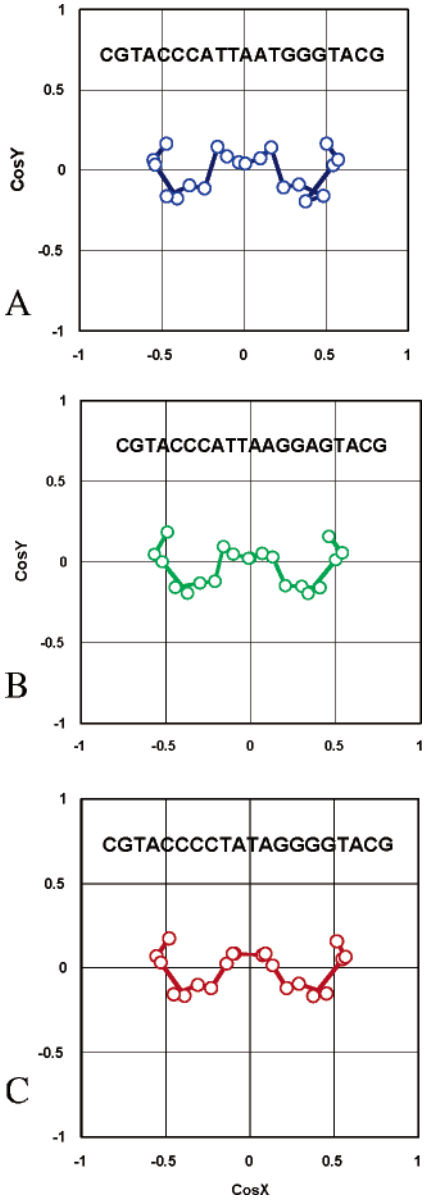

FIGURE 6: Normal vector plots (*21*) show similar overall DNA bending in the three NarL^C complexes. In each case, an overall gradual bend of approximately 42° is seen. The two ends of the DNA 20mer exhibit an A-DNA-like writhe, whereas the central region has a planar bend as often seen in B-DNA.

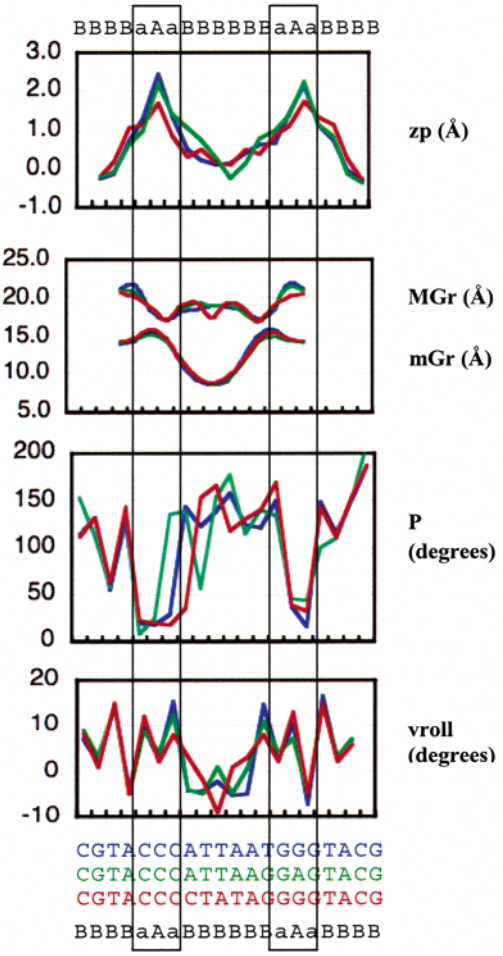

FIGURE 7: Comparison of important local DNA helix parameters for structures A (blue), B (green), and C (red). The general helical nature of the DNA is indicated by "B" for B-form, "A" for A-form, and "a" for A-like. The transition from B- to A-form is gradual, as is the overall curvature of the DNA. The zp parameter measures the separation of adjacent phosphates. Values greater than 1.5 Å are typical of A-DNA. MGr and mGr are widths of major and minor grooves, respectively. A broader minor groove width of ~15 Å is typical of A-DNA. Sugar pseudorotation angle P is a measure of sugar ring conformation, with A-form DNA having low values. The roll angle between successive base pairs is called vroll. Positive roll means the base pair is opening toward the minor groove. In general, the parameters for nonpalindromic sequence B (green) are the least well defined. Groove widths were calculated with CURVES (*23*), vroll and P with FREEHELIX (*21*, *38*), and zp values with 3DNA (*39*).

roll angle between adjacent base pairs is large in A-DNA but near zero in B-DNA. All three sequences exhibit a complete B to A transition at the center of each heptamer binding site.

One further subtlety of bending is worth mentioning. The roll between two adjacent base pairs is defined as positive if the angle between them opens toward the minor groove and negative if toward the major groove (*25*, *26*). Figures 1B and 5 show that bending of the DNA helix is achieved by inserting NarL^C domains into two adjacent major grooves, and stabilizing compression of the minor groove that lies between them. This would tend to drive the roll values in the center to even more negative values, and that is exactly what the vroll plot at the bottom of Figure 7 indicates. These local roll angles are not zero, but around −5°.

Conservation of DNA Backbone Contacts between Complexes. While only three protein side chains (Lys 188, Val 189, and Lys 192) contact the DNA major groove floor directly (Figure 1C), DNA backbone contacts are extensive. This is typical of protein−DNA complexes in which bent DNA exhibits gradual curvature. Both crystallization trials and EMSAs showed that NarL^C complexes were more stable when synthetic binding sites contained additional bases peripheral to the heptameric consensus sequences. Cyt 1 and Gua 2 were added to favor crystallization and are not necessarily found at NarL binding sites (Table 1). DNA backbone recognition occurs at two regions along the 20mer: one at the 5′-end of each heptameric consensus sequence which will be termed the heptameric site and the other at the central bend between binding sites. Contacts in the heptameric site are primarily directly between protein and DNA, whereas those in the central bend are frequently water-mediated.

Although NarL^C recognition involves extensive DNA contacts (Figures 1 and 5), certain DNA bases are not contacted. It is at these bases where large difference density is visible at the phosphate backbone, indicating flexibility of the DNA backbone at noncontacted regions (Figure 5). Conversely, in the central AT-rich sequence, the central six bases have some of the lowest *B*-factors in the entire structure, indicating a stable, well-placed structure. NarL^C dimerizes above the minor groove in this central region, compressing the groove. The nonpalindromic sequence (Table 1B) is present randomly in both orientations in the crystal, leading to increased crystal packing disorder, large backbone difference density, and *B*-factors or temperature factors in the central region which are not as low as in the palindromes.

NarL^C Dimerization at Three Different DNA Binding Sites. NarL^C dimerizes identically on each of the three synthetic tail-to-tail oligonucleotides, placing the recognition helix in consecutive major grooves on one face of the DNA helix (Figure 1B). NarL^C has extensive DNA backbone contacts (Figure 1C). The DNA backbones of the three complexes do not superimpose identically in regions where the sugar−phosphate backbone is not stabilized by protein contact (Figure 5). DNA backbone flexibility allows the three binding sites to maintain a similar 42° bend due to dimerization of NarL^C. The NarL^C dimerization interface is formed by residues Val 204, Val 208, and His 211 on helices 10 (Figure 1D) and residues Ile 167, Ala 168, and Val 170 at the loop region between helices 7 and 8 (*11*).

Secondary Recognition via DNA Backbone Interactions

Two different types of protein−DNA interaction can lead to recognition of DNA sequence. The most familiar is that in which protein side chains hydrogen bond with atoms on the major or minor groove edges of base pairs; this will be termed the "primary mode" of sequence recognition. But base sequence can also be sensed by interactions of protein with the sugar−phosphate DNA backbone, if a particular base sequence produces specific alterations in DNA helical structure. This will be termed the "secondary mode" of DNA sequence recognition. The propensity for the GC-rich centers of the heptamer sites to adopt A-like writhe and the AT-rich central region to exhibit smooth B-like planar bend are excellent examples of structure easily recognized by the secondary mode recognition. This section considers such secondary, but important, interactions between NarL^C and the sugar−phosphate DNA backbone.

Heptameric Sites. Helices 7−9 form the helix−turn−helix DNA recognition motif. Helix 7 is the support helix for helices 8 and 9 which comprise the helix−turn−helix motif. The secondary mode contacts shown in the panels of Figure 8 are practically identical among the three DNA complexes. Although all three complexes are overlaid, the high level of identity between complexes results in the apparent appearance of only one complex in panels A and B of Figure 8. In panels C and D of Figure 8, differences between the DNA backbones near the central bend of the three complexes are visible in the noncontacted regions. Figure 8A shows the interactions around DNA bases 1−3, and Figure 8B extends the view along bases 3−5 from a slightly different angle. Figure 8A involves contacts from the helix 7 N-terminus, the linker residues preceding it, helix 9, and the turn preceding helix 9: Thr 157, Pro 158, Arg 159, Ile 182, and His 190 (Figure 1C). Figure 8B involves the N-terminus of helix 9 and part of the turn preceding it: Ile 182, Thr 183, Ser 185, and Thr 186 (Figure 1C).

In Figure 8A, Thr 157 and Pro 158 occur in the linker preceding helix 7. Thr 157 is highly conserved, although Pro 158 is not always found in the NarL/FixJ subfamily members. Aliphatic side chain carbons of both residues sit in van der Waals contact with the Gua 2 phosphate, constraining its position. The Thr 157 side chain hydroxyl group is H-bonded to Gua 2 phosphate at OP1, and aliphatic side chain atoms are in van der Waals contacts with Cyt 1 sugar atoms.

Within DNA recognition helix 9, the Arg 159 side chain extends parallel to the DNA backbone and packs against sugar−phosphate atoms of both Gua 2 and Thy 3. Its main chain amide nitrogen H-bonds to OP1 of Gua 2 (3.2 Å); its ϵ-nitrogen H-bonds to OP1 of Thy 3 (2.9 Å), and its terminal amino nitrogen H-bonds to OP2 (2.9 Å).

His 190 at the center of DNA recognition helix 9 has a H-bond from its ring Nϵ2 to OP1 of Thy 3 (2.7 Å). Ring Cϵ1 is within van der Waals packing distance of both the Gua 2 sugar (4 Å) and Thy 3 exocyclic methyl C7 (3.9 Å).

Figure 8B is a continuation of secondary mode contacts, showing the DNA interactions of Ile 182, Thr 183, Ser 185, and Thr 186 of recognition helix 9 and the turn that connects it to helix 8. Ile 182, at the center of the connecting turn, is usually either Ile or Leu. It makes extensive van der Waals contacts with the DNA backbone between Thy 3 and Ade 4 phosphates, limiting their positions and facilitating H-bonding by Arg 159.

Thr 183 occurs at the C-terminus of the helix−turn−helix turn. Its amide nitrogen is H-bonded to Ade 4 OP1 (3 Å). Its hydroxyl oxygen atom is also approximately 3 Å from Ade 4 OP1, but its orientation away from the DNA backbone suggests that the threonine methyl is more important in specific DNA recognition. As with the Ile 182 side chain, the Thr 183 side chain is inserted between phosphates of two consecutive bases, Ade 4 and Cyt 5, limiting their positions.

Ser 185 and Thr 186 are found at the N-terminus of DNA recognition helix 9. Thr 186 is highly conserved. Its hydroxyl group H-bonds with Ade 4 OP1 (2.7 Å), while its methyl carbon is within van der Waals contact distance of Thy 3 backbone and base atoms (C6, 4.2 Å; C7, 4.3 Å). The Ser 185 main chain carbonyl stabilizes distorted DNA at the A-DNA transition site by a water-mediated hydrogen bond to O4 of Cyt 5 (*11*). The Ser 185 hydroxyl group hydrogen bonds directly to Cyt 5 OP1 (3 Å). The serine is not well conserved in NarL/FixJ subfamily members. Because its main chain stabilizes the DNA base distortions conferred by the highly conserved Val 189, it is likely that alternate side chain identities will play additional undescribed roles in DNA recognition.

Central AT-Rich Bend. The other important region of secondary mode interaction occurs at the AT-rich central bend (Figure 8C,D). The AT-rich nature of the central region favors a narrowed minor groove, over which NarL^C can easily dimerize. Helix 10 is the locus of dimerization. The central bend region involves base pairs 11−14, residues from

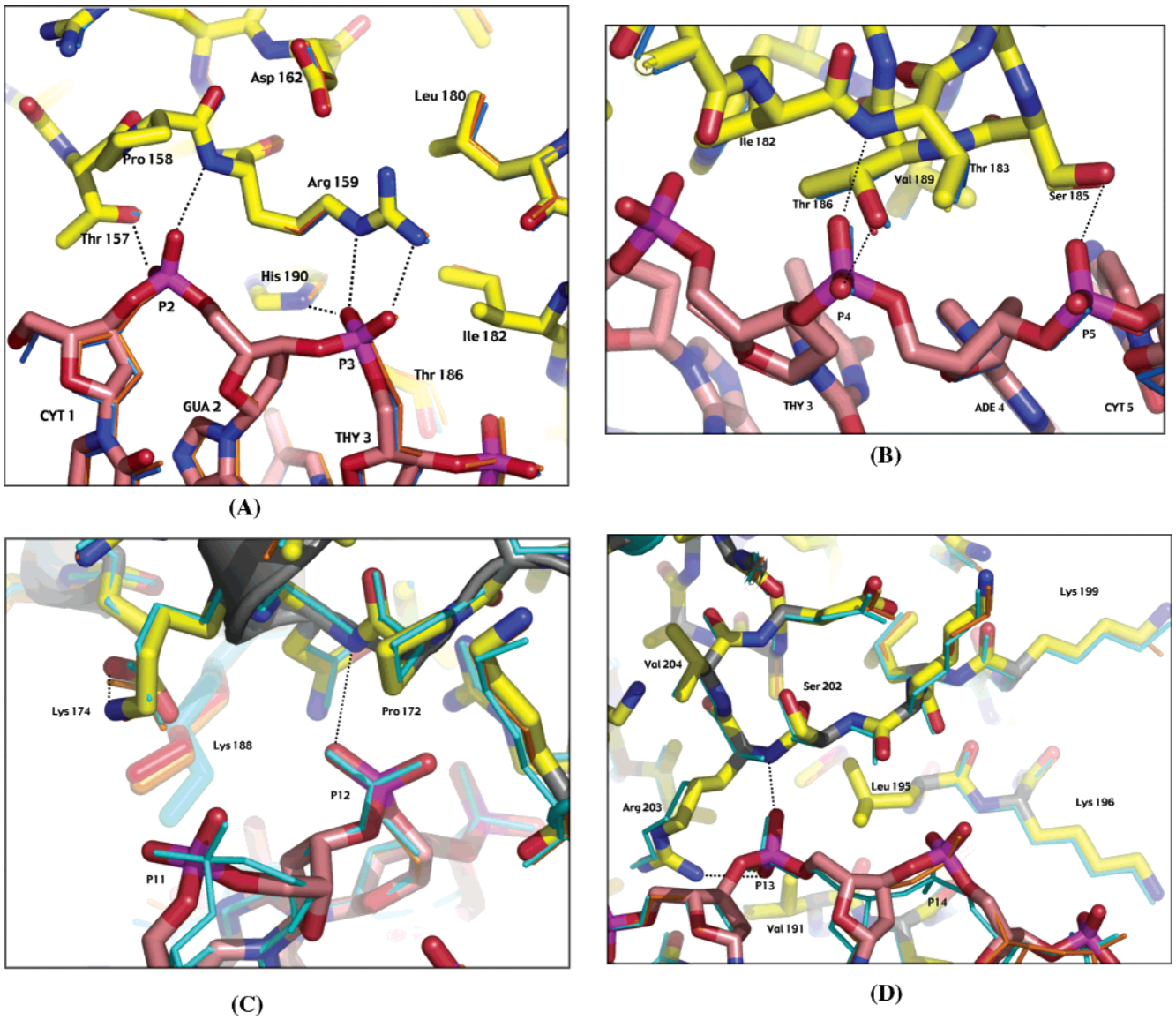

FIGURE 8: Protein−DNA interactions viewed from the minor groove side, looking into the major groove. Important hydrogen bonds are depicted as dotted lines. In these views, nonbonded OP1 recedes into the rear of the diagram and OP2 extends toward the viewer (*26*). All three complexes are superimposed, but because of their high degree of similarity, individual chains are often indistinguishable. (A) Region along bases 1−4, involving side chains Thr 157, Pro 158, Arg 159, and His 190. (B) Adjacent region along bases 3−5, with side chains Ile 182, Thr 183, Ser 185, and Thr 186. (C) Central region involving bases 11 and 12 in the AT-rich region connecting heptamer sites. The flexible Lys188 is visible in the background. (D) Adjacent central region involving bases 13 and 14. Lysines 196, 199, and 201 are visible extending away from the DNA helix into solution. This figure was rendered with Pymol (*36*).

helices 8 and 10, and the loops preceding helices 8 and 10 (Figure 1B−D).

Pro 172 in the loop preceding helix 8 (Figure 8C) serves to position the DNA backbone, but is not highly conserved in NarL/FixJ subfamily members. Pro 172 α- and β-carbons are less than 4 Å from the Ade 12 phosphate group. Because this loop between helices 7 and 8 also contributes to dimerization, substitution of one hydrophobic group for another may indicate only slight differences in dimerization geometry. For example, Ala 166 of the *E. coli* UhpA response regulator is in the position analogous to that of NarL Pro 172. Instead of proline, serine is commonly found in this position for the NarP family and for some members of the NarL family.

Asn 173 and the highly conserved Lys 174 occur at the N-terminus of helix 8. The backbone amide nitrogen of Asn 173 H-bonds with OP1 of Ade 12 (3 Å). Lys 174 forms a salt bridge with Glu 184 and has a water-mediated contact with OP1 of Thy 11.

Arg 203 is found at the N-terminus of dimerization helix 10 (Figures 1D and 8D). It is conserved in NarL and NarP, but not in other related members of the NarL/FixJ subfamily such as BvgA and UhpA. Arg 203 stabilizes the DNA backbone at the central narrowed minor groove region. Its backbone amide and NH1 H-bond with OP1 and OP2 of base 13. Aliphatic carbons of the Arg 203 side chain make van der Waals contacts with the sugar−phosphate backbone at bases 12 and 13. Val 191 is at the center of DNA recognition helix 9 and makes a 3.6 Å van der Waals contact with OP1 of Thy 13. The position is not highly conserved, but is often hydrophobic.

Leu 200, Lys 201, and Ser 202 are located at the C-terminus of the loop preceding dimerization helix 10. Leu 200 is highly conserved in the NarL family and is a

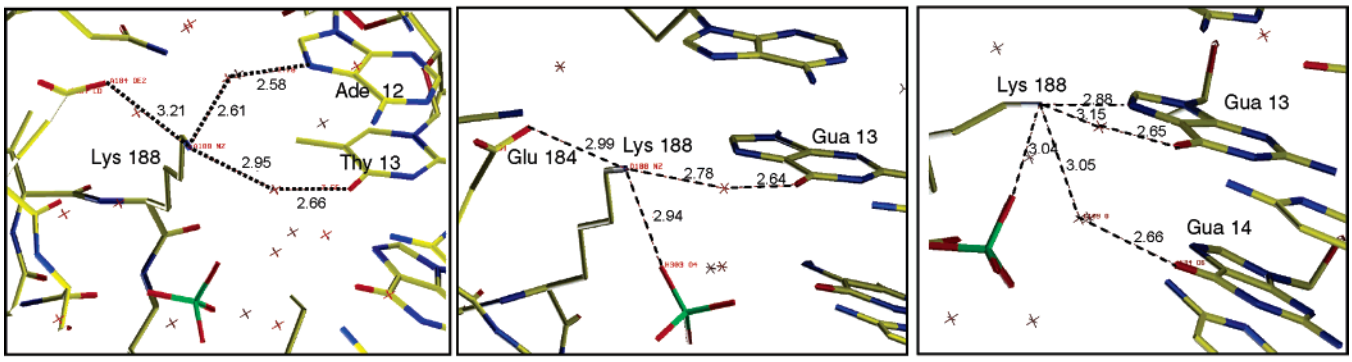

FIGURE 9: Lys 188 can flexibly recognize different heptamer sequences. (A) Lys 188 position 1, found in the previously published *nirB* −74/−74 complex (Table 1A), is shown. In this position, water-mediated recognition of Ade 12 N7 occurs. Ade 12 corresponds to the complementary strand at the seventh position of the TACCYMT consensus sequence. The structure of the *nirB* −74/−74 complex was determined at the highest resolution, resulting in more visible waters than in the other complexes. A sulfate solvent molecule (green) is seen at the bottom of panels A and B. (B) Lys 188 is again in position 1 in the nonpalindromic *nirB* −74/−65 complex (Table 1B) whether guanine or thymine is present. Base 13 corresponds to the complementary strand at the sixth position of the consensus. Few waters are visible in this complex, but a water-mediated contact with Gua 13 O6 is observed. (C) The *narG* −89/−89 complex contains a four-guanine tract in each half-site (Table 1C). Lys 188 is found in position 2, directed down into the DNA major groove and hydrogen bonded to Gua 13 N7. In this position, Lys 188 can also form a water-mediated contact with Gua 14 O6, and is slightly closer to the sulfate molecule than in position 1. This figure was generated using O (*19*).

hydrophobic residue in others. The Leu 200 backbone carbonyl has a water-mediated contact to Gua 14 OP1 (not shown). Lys 201 is not highly conserved, but is often polar or positively charged. A water-mediated contact (not shown) extends from its backbone carbonyl to the phosphate of base 14. The highly conserved Ser 202 Cα, Cβ, and backbone carbons are within van der Waals distance of the base 13 phosphate group.

Primary Recognition via Base−Edge Contacts

Sequence Recognition via Three Key Contacts within the Major Groove. Major groove recognition occurs in the GC-rich regions of the heptamers (Table 1). The largest base stacking distortions occur here along the pyrimidine strand of the GC base pairs, leaving the guanine bases stacked and poised for NarL recognition of their negatively charged major groove base edges. Three residues of helix 9, Lys 188, Val 189, and Lys 192, are involved in recognition of the consensus sequence (Figure 1C). The only significant protein difference between the three NarL^C complexes involves the adaptation of Lys 188 to the major groove face.

In all three structures, Val 189 of helix 9 perturbs Cyt 5. Cyt 5 occurs in the third position of the NarL consensus sequence TACYYMT, near its more tightly specified 5′ end. Guanine·cytosine base pairs are normally thought of as being planar because of their three Watson−Crick hydrogen bonds, and so it was surprising to observe Val 189 causing a large buckle and propeller twist at the distorted cytosine. This distorted conformation is locked in place by a water-mediated hydrogen bond from the NarL^C backbone (*11*). A thorough search of the Protein Data Bank yielded no comparable interactions (*11*).

Lys 192 recognizes the Watson−Crick complements of the third and fourth bases of the consensus sequence TACYYMT, where Y indicates a pyrimidine. With palindromic sequences A and C of Table 1, Lys 192 makes hydrogen bonds to Gua 15 (complementary to Cyt 5 of the opposing strand) and to Gua 16 (complementary to Cyt 6 of the opposing strand) (see panels A and C of Figure 3 of ref *11*). In asymmetrical nonpalindromic sequence B, the symmetrical equivalent for Gua 15 is Ade 15, and Lys 192 retains an equivalent recognition H-bond to N7 of adenine. The direct hydrogen bond from Lys 192 to O6 of Gua 16 is slightly longer in sequence C than in sequence A or B.

While H-bonding of Lys 192 to the floor of the major groove is direct, that of Lys 188 can be either direct or water-mediated (Figure 9). In panels A and B of Figure 9, the interactions are mediated by water molecules ("position 1"), while in panel C, the hydrogen bonds are both direct and water-mediated ("position 2"). In the positions observed in the three DNA complexes, Lys 188 interacts with the complementary DNA strand of the TACCYMT consensus at base positions 5−7. Lys 188 is capable of utilizing a "three-base-slide" type recognition to accommodate sequence variation.

In the *nirB* −74/−74 palindromic cocomplex, Lys 188 hydrogen bonds with O4 of Thy 13 and N7 of Ade 12 via intermediate water molecules (Figure 9A). Thy 13 and its methyl moiety prevent Lys 188 from approaching the major groove to form direct contacts, so Lys 188 in position 1 is directed toward the protein rather than the DNA. Ade 12 and Thy 13 correspond to the complementary strand at the seventh and sixth positions, respectively, of the TACCYMT consensus sequence.

O6 of Gua 13 of the nonpalindromic *nirB* −74/−65 sequence is recognized indirectly (Figure 9B). Lys 188 is again in position 1, whether guanine or thymine is present. Lys 188 is 2.99 Å from Glu 184 OE2, which is similar to the 3.2 Å distance shown in Figure 9A. Apparently, Glu 184 stabilizes the Lys 188 in position 1.

narG −89/−89 sequence C contains a four-guanine tract in each heptamer (Table 1C). Lys 188 is found in position

2, directed down into the DNA major groove and hydrogen bonded to Gua 13 N7 (Figure 9C). Thy 13 is replaced with Gua13 in sequence C, and Lys 188 is able to contact N7 of Gua 13 directly. In this position, Lys 188 also forms water-mediated contacts to the O6 atoms of Gua 13 and Gua 14, and is slightly closer to the sulfate solvent molecule than in position 1.

The nonpalindromic *nirB* −74/−65 cocrystal structure (sequence B) exemplifies the recognition flexibility of Lys 188. Because sequence B is nonpalindromic, base 13 is guanine like in sequence C, while base 13 of the opposing strand is thymine like in sequence A. Base 13 is complementary to the sixth position of the TACYYMT NarL consensus sequence, where M is A or C. Although sequence B base 13 is both guanine and thymine, in the cocrystal structure Lys 188 is always found in position 1, directed toward helix 9 rather than the DNA major groove floor (Figure 9B).

DISCUSSION

The three NarL binding sites examined here are representative of promoter arrangements found in vivo (Table 1 and ref *14*). The $NarL^C$ dimer places one recognition helix in each of two successive major DNA grooves. All three DNAs with different sequences adopt similar overall bends (Figure 6) via sequence-dependent DNA distortions (Figure 7). The minor groove of the central AT-rich region between binding sites is compressed, allowing helices 10 to approach close enough to form the dimer interface. Some of the lowest *B*-factors of the structures are found within the central bend, indicating that the minor groove compression is stable and energetically favorable. The three different sequences allow one to examine base substitution effects at five different points along the DNA helix, including a purine-to-purine substitution and a number of purine-to-pyrimidine substitutions (Table 1). The major groove hydrogen bonding pattern recognized by $NarL^C$ side chains is conserved. The overall DNA bend is unchanged, with noncontacted DNA backbone showing flexibility that could variably compensate for the energetic constraints imposed by the $NarL^C$ dimerization and binding, which stabilizes DNA conformation at regions of contact (Figures 5 and 8).

The $NarL^C$−DNA contact surface is extensive, with each base pair of the entire 20mer DNA interacting via primary or secondary mode contacts with the protein. Positively charged lysines on the DNA recognition helix form direct primary contacts to guanines in the DNA major groove binding site. The successive guanine-rich regions of the major groove binding sites are separated by adenines and thymines within the central region. The hydrogen bond donor/acceptor pattern of AT base pairs is directionally indistinguishable (see Figure 2 of ref *27*), and $NarL^C$ secondary recognition via DNA structural features in this region, such as a narrowed minor groove, is more relevant for specificity. The 7−2−7 DNA base pair spacing between heptamers allows the $NarL^C$ dimer to form above the minor groove at this adenine/thymine-rich region, directly contacting only the DNA backbone. It is possible that oligomerization occurs at tail-to-tail 7−1−7 and 7−3−7 sites, but the DNA spacing would not allow the high-affinity 7−2−7 dimer to form. Because $NarL^C$ is a monomer in solution (*11*; A. E. Maris, R. P. Gunbsalus, et al., unpublished data) and high-affinity dimerization occurs only on 7−2−7 sites, the tail-to-tail dimerization would not interfere with alternate forms of oligomerization at non-tail-to-tail sites.

Similar C-Terminal Dimerization Observed in Other Proteins. The $NarL^C$ dimer is oriented essentially identically in all three cocomplex crystal structures, with the dimerization interface comprising helix 10 and the loop region between helices 7 and 8 (Figure 1B). Until the first $NarL^C$−DNA complex structure was determined, we did not know the oligomeric state required for high-affinity DNA binding, or how dimerization might occur. Full-length unphosphorylated NarL, like $NarL^C$, is a monomer (A. E. Maris, R. P. Gunsalus, et al., unpublished data). However, the $NarL^C$ domain gave DNaseI footprinting patterns identical to those of activated full-length NarL (*11*). Specific DNA contacts were proposed on the basis of the monomeric binding mode of the paired homeodomain (*9*). However, our crystallization trials and EMSA data showed that only the 7−2−7 promoter arrangement, containing two binding sites, was recognized with high affinity. Dimerization via helix 10 was later predicted by Ducros et al. (*28*) from the crystal structure of the single-domain GerE transcription factor, with which $NarL^C$ shares 38% sequence identity. Unlike $NarL^C$, GerE is a stable dimer in solution. It is therefore not surprising that the GerE dimerization interface is more extensive than that of $NarL^C$ and involves additional C-terminal residues. Both GerE and $NarL^C$ dimerize along analogous C-terminal helices, and both interfaces involve primarily hydrophobic interactions at identical positions along the dimerization helix. The orientation of dimerization helices with respect to each other differs, with those of $NarL^C$ being almost parallel to one another (Figure 1D) and those of GerE projecting away from one another at the C-terminal ends (Figure 5 of ref *28*). Three slightly different GerE dimers were found in the asymmetric unit, demonstrating flexibility in dimerization and that, in DNA binding, a specific dimer orientation may be selected.

Dimerization via the C-terminal helix as a component of DNA recognition was also revealed in the cocrystal structures of a full-length, DNA-bound LuxR family member (*29*, *30*). In the TraR−DNA structures, the C-terminal dimerization helices once again comprise an interface similar to that of GerE and $NarL^C$. However, the dimerization helices are oriented almost parallel to each other, a case more similar to $NarL^C$ than to GerE, only rotated away from each other toward their C-termini. Dimerizations of the TraR N-terminal regulatory domain and the C-terminal DNA binding domain are essentially independent of each other, with few interdomain contacts and a flexible linker connecting the domains. Superposition of the activated TraR structure with unphosphorylated NarL suggests where the phosphorylated NarL regulatory domain could be (Figure 10). The unphosphorylated NarL structure (*9*) and $NarL^C$ dimerization on DNA demonstrate where the phosphorylated regulatory domain cannot be.

While response regulator output domains are diverse, for those containing the C-terminal dimerization helix (see Table 7 of ref *31*), similar dimerization can be expected for those that have hydrophobic residues in equivalent positions. For example, the NarL-related sister protein, NarP, has valine, isoleucine, and leucine residues at NarL positions Val 204,

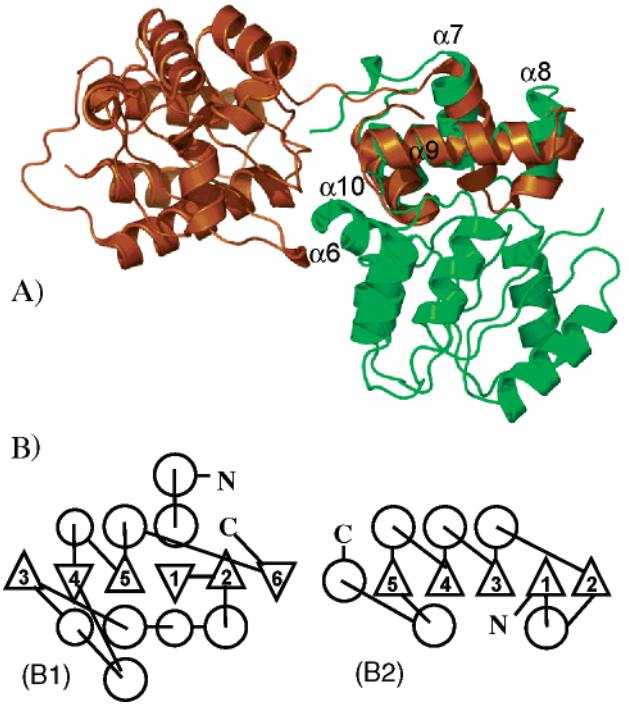

FIGURE 10: Comparison of the TraR and NarL DNA binding domains, which exhibit the same fold. (A) Superposition performed using the DNA recognition helix of each protein. The rmsd for 59 Cα atoms of the C-domain whose sequences are 20% identical is 1.8 Å (*29*). NarL (green) is in its unphosphorylated form (*8*, *9*). TraR (orange) is in its activated, DNA-bound form (*29*, *30*). No unactivated structure is yet available for TraR. In the TraR structures, both the N- and C-domains dimerize independently of one another, with a flexible linker between domains. While it is clear that activation allows the NarL N-domain to move out of the way for dimerization and DNA binding, interdomain flexibility allows many possible N-domain locations. (B) Secondary structures of the regulatory domains of TraR (B1) and NarL (B2) are shown schematically, with strands of the β sheet represented by numbered triangles and α helices by circles. Although the DNA binding domains are quite similar, the N-terminal regulatory domains differ in both structure and activation: TraR is activated by binding hormone and NarL by phosphorylation.

Val 208, and His 211, respectively (Figure 11), and recognizes a heptamer consensus similar to that of NarL (*14*). Figure 11 also shows sequence alignments for NarL and NarP orthologs found in other microorganisms also capable of harvesting energy by nitrate reduction. Hydrophobic residues found at the $NarL^C$ dimerization interface are conserved across microorganisms for both NarL and NarP. In addition, Arg 203 also is highly conserved among these orthologs and in GerE and its homologues (*28*). This arginine is important for dimerization because it forms stabilizing hydrogen bonds with the DNA backbone at the narrowed minor groove region between heptameric half-sites, over which dimerization occurs (Figures 1D and 8D). The $NarL^C$−DNA crystal structures, the synthetic DNA binding studies (Table 1 and Figure 2), and in vivo gene expression assays (*32*) demonstrate that dimerization is important. Because the $NarL^C$ helix 10 hydrophobic residues are similarly conserved in NarP, NarP could dimerize similarly on promoter DNA.

In another example of conserved secondary recognition features, a homology model based on the $NarL^C$−DNA structure was built for BvgA, a whooping cough virulence protein. Equivalent dimerization residues in BvgA are Val 197, Asp 201, and Lys 204. These residues are not exclusively nonpolar, and a dimerization interface may not have been predicted prior to determination of the $NarL^C$−DNA structure. DNA cleavage experiments based on the dimeric BvgA model successfully demonstrated the orientation and locations of BvgA promoter binding relative to the RNA polymerase α subunit C-terminal domain (*33*). The hydrophobic regions of BvgA side chains may play a role in dimerization, as His 211 of $NarL^C$ does. This may also be the case for the related response regulator UhpA (*34*), in which the predicted dimerization residues would include an arginine. Thus, members of the NarL/FixJ subfamily use varied means to achieve dimerization and DNA recognition. It is not yet known which NarL surface-exposed residues contact RNA polymerase (*35*), or how the contacts may differ at the diverse NarL-controlled operons (Figure 2).

Residues involved in the secondary DNA backbone contacts shown in Figure 5 are well conserved among NarL and NarP orthologs (Figure 11). Arg 159 hydrogen bonds to two consecutive bases via DNA backbone contacts and is highly conserved. The neighboring Thr 157 is not well conserved, but always consists of a residue capable of similarly hydrogen bonding to the DNA backbone. His 190 and Thr 186 on the DNA recognition helix hydrogen bond to the DNA backbone and are highly conserved. An exception to the many conserved residues is Ser 185, the hydroxyl group of which hydrogen bonds to the DNA backbone. The Ser 185 main chain carbonyl hydrogen bonds to a water bound to the distorted Cyt 5 at the major groove floor.

Many residues that are important for secondary DNA contacts are conserved among both NarL and NarP orthologs, but residues involved in primary DNA contacts distinguish NarL and NarP orthologs. NarL Val 189, which contacts Cyt 5 distorting the major groove floor, is highly conserved, as is Lys 192. However, the flexible Lys 188 position can be lysine, glutamine, or arginine. This primary sequence variation provides the basis of similar but slightly differing NarL and NarP consensus sequences. NarP orthologs contain conserved residues that are important for secondary DNA contacts in the $NarL^C$−DNA structures and recognize a DNA consensus sequence similar to the NarL consensus (*6*); however, no biochemical data support NarL−NarP heterodimer formation. One possible explanation arises from the position of bulky residues found on dimerization helix 10. The large NarL Trp 209, which contributes to the hydrophobic core of the helical bundle, is not conserved in the NarP family. Instead, members of the NarP family often have phenylalanine in the position adjacent, which could alter the orientation of helices with respect to each other, disfavoring heterodimerization. Amino acid sequence alignments already suggested that considerable diversity was embedded in structural features of the NarL/FixJ family (Figure 4 of ref *9*). Now, these cocomplex structures and the sequence conservations shown in Figure 11 can help define promoter interactions of nitrate-regulated two-component signaling pathways. We can begin to understand the basis of promoter recognition by diverse organisms, such as pathogenic, enteric, and soil bacteria.

Five complementary NarL recognition principles follow:

(1) Val 189 interacts with the major groove floor, distorting a cytosine base, while successive purines on the opposite

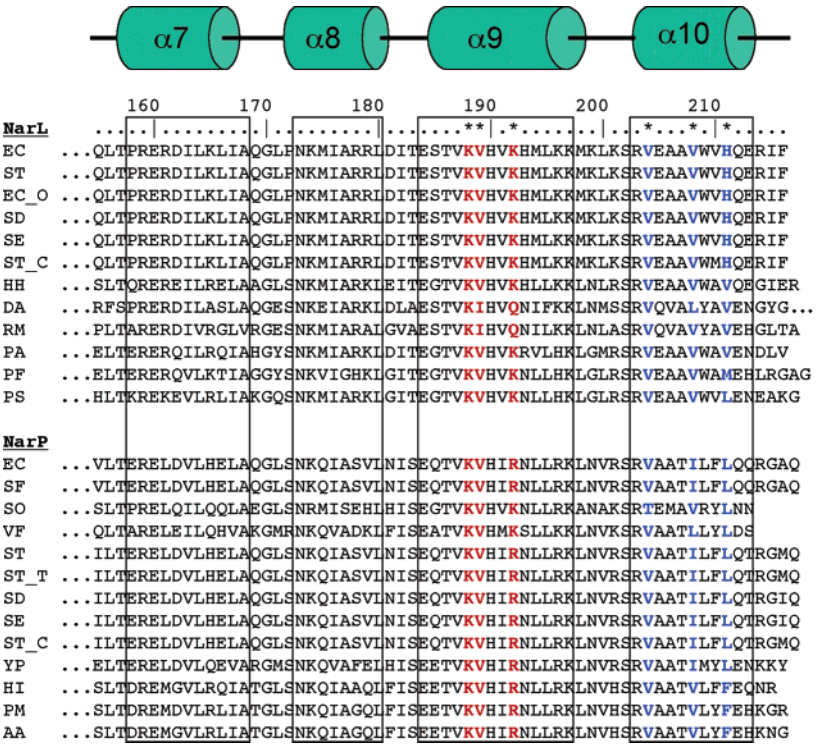

FIGURE 11: Sequence alignment of the DNA recognition domains of NarL and NarP orthologs. The secondary units are shown as cartoons and labeled. NarL numbering is shown at the top. Helices 8 and 9 form the helix−turn−helix motif. *E. coli* NarL and NarP are at the top of each list, with sequences from other nitrate-reducing microorganisms found below. The six asterisks at the top of the lists show the locations of the three residues comprising the helix 10 dimerization interface (blue) and the three residues contacting the floor of the DNA major groove (red). Ellipses indicate flanking residues that are not shown. The microorganisms are represented as follows: EC (*E. coli* K12 MG1655), ST (*Salmonella typhimurium* LT2), EC_O (*E. coli* O157), SD (*Salmonella dublin*), SE (*Salmonella enteritidis*), ST_C (*Salmonella typhi* CT18), HH (*Halomonas halodenitrificans*), DA (*Dechloromonas aromatica* RCB), RM (*Ralstonia metallidurans* CH34), PA (*Pseudomonas aeruginosa* PAO1), PF (*Pseudomonas fluorescens*), PS (*Pseudomonas stutzeri*), SF (*Shigella flexneri* 2a), SO (*Shewanella oneidensis* MR-1), VF (*Vibrio fischeri* ES114), ST_T (*Salmonella typhi* Ty2), YP (*Yersinia pseudotuberculosis*), HI (*Haemophilus influenzae* 86-028NP), PM (*Pasteurella multocida* Pm70), and AA (*Actinobacillus actinomycetemcomitans* HK1651).

strand stack stably for protein recognition (Figure 3B of ref *11*).

(2) The flexible side chain of Lys 188 allows flexible recognition of DNA bases on the major groove floor (Figure 9).

(3) Lys 192 requires purine in consensus positions 3 and 4 of the complementary strand.

(4) DNA substitutions in the minor groove below the dimerization interface (in the AT-rich region) maintain secondary backbone contacts by allowing DNA backbone flexibility in noncontacted regions.

(5) DNA base spacing other than the two bases of the high-affinity tail-to-tail 7−2−7 promoter sites (Table 1) destabilizes the high-affinity $NarL^C$ dimerization interface.

Most two-component response regulators are transcription factors, for which activation greatly enhances DNA binding. However, it is not clear exactly how oligomerization and DNA binding are related, or what roles the related aspects play in transcriptional regulation. Activation in vivo is accomplished by phosphorylation of the N-terminal receiver domain (*1*), which in several examples strengthens the oligomeric state. In the case of NarL, phosphorylation releases the DNA binding determinants and may also strengthen the oligomeric state, affecting local NarL concentration and communication with other transcription machinery. However, the DNA binding domain does not dimerize until recognition of the 7−2−7 DNA binding site. While other binding site arrangements (Figure 2) serve as recognition platforms (Table 1) for the NarL DNA binding domain, molecular details of how NarL recognizes non-7−2−7 sites are not yet known. Phosphorylation-driven oligomerization of NarL could no doubt enhance binding at non-7−2−7 sites. The DNA-dependent dimerization of $NarL^C$ allows for an additional level of transcriptional regulation beyond phosphorylation. In the TraR−DNA structures, the regulatory and DNA-binding domains oligomerize independently, supporting this mechanistic model. The TraR structures also offer a model for positioning of the activated NarL regulatory domains, although the relevant NarL structure remains to be determined.

ACKNOWLEDGMENT

A.E.M. thanks David Wemmer for his support while completing the manuscript. We thank Michael Sawaya and Duilio Cascio for assistance in data collection and processing. We also thank Michael R. Jarvis for discussions and for providing $NarL^C$.

REFERENCES

1. Parkinson, J. S., and Kofoid, E. C. (1992) Communication modules in bacterial signalling proteins, *Annu. Rev. Genet. 26*, 71−112.
2. Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction, *Annu. Rev. Biochem. 69*, 183−215.
3. Chiang, R. C., Cavicchioli, R., and Gunsalus, R. P. (1992) Identification and characterization of NarQ, a second nitrate sensor

for nitrate-dependent gene regulation in *Escherichia coli*, *Mol. Microbiol. 6*, 1913−1923.
4. Lee, A. I., Delgado, A., and Gunsalus, R. P. (1999) Signal-dependent phosphorylation of the membrane-bound NarX two-component sensor-transmitter protein of *Escherichia coli*: Nitrate elicits a superior anion ligand response compared to nitrite, *J. Bacteriol. 181*, 5309−5316.
5. Li, J., Kustu, S., and Stewart, V. (1994) *In vitro* interaction of nitrate-responsive regulatory protein NarL with DNA target sequences in the *fdnG*, *narG*, *narK* and *frdA* operon control regions of *Escherichia coli* K-12, *J. Mol. Biol. 241*, 150−165.
6. Darwin, A. J., Tyson, K. L., Busby, S. J. W., and Stewart, V. (1997) Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement, *Mol. Microbiol. 25*, 583−595.
7. Unden, G., and Bongaerts, J. (1997) Alternative respiratory pathways of *Escherichia coli*: Energetics and transcriptional regulation in response to electron acceptors, *Biochim. Biophys. Acta 1320*, 217−234.
8. Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Cascio, D., Gunsalus, R. P., and Dickerson, R. E. (1998) NarL dimerization? Suggestive evidence from a new crystal form, *Biochemistry 37*, 3665−3676.
9. Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R. P., and Dickerson, R. E. (1996) Structure of the *Escherichia coli* Response Regulator NarL, *Biochemistry 35*, 11053−11061.
10. Lee, S.-Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A., and Wemmer, D. E. (2001) Crystal structure of activated CheY.; Comparison with other activated receiver domains, *J. Biol. Chem. 276*, 16425−16431.
11. Maris, A. E., Sawaya, M. R., Kaczor-Grzeskowiak, M., Jarvis, M. R., Bearson, S. M. D., Kopka, M. L., Schroeder, I., Gunsalus, R. P., and Dickerson, R. E. (2002) Dimerization allows DNA target site recognition by the NarL response regulator, *Nat. Struct. Biol. 9*, 771−778.
12. Eldridge, A. M., Kang, H.-S., Johnson, E., Gunsalus, R. P., and Dahlquist, F. W. (2002) Effect of phosphorylation on the interdomain interaction of the response regulator NarL, *Biochemistry 41*, 15173−15180.
13. Zhang, J. H., Xiao, G., Gunsalus, R. P., and Hubbell, W. L. (2003) Phosphorylation triggers domain separation in the DNA binding response regulator NarL, *Biochemistry 42*, 2552−2559.
14. Stewart, V., and Rabin, R. S. (1995) Dual sensors and dual response regulators interact to control nitrate- and nitrite-responsive gene expression in *Escherichia coli*, in *Two-Component Signal Transduction* (Hoch, J. A., and Silhavy, T. J., Eds.) pp 233−252, American Society for Microbiology Press, Washington, DC.
15. Darwin, A. J., Li, J., and Stewart, V. (1996) Analysis of nitrate regulatory protein NarL-binding sites in the *fdnG* and *narG* operon control regions of *Escherichia coli* K-12, *Mol. Microbiol. 20*, 621−632.
16. Darwin, A. J., and Stewart, V. (1996) The Nar modulon systems: Nitrate and nitrite regulation of anaerobic gene expression, in *Regulation of gene expression in Escherichia coli* (Lin, E. C. C., and Lynch, A. S., Eds.) pp 343−359, R. G. Landes Biomedical Publishers, Austin, TX.
17. Maris, A. E. (2002) Molecular studies in protein/DNA recognition: Crystallographic trials of the retrovirus integrase protein and the crystal structure of the *Escherichia coli* response regulator $NarL^{C}$/DNA complex, Ph.D. Thesis, University of California, Los Angeles.
18. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR System (CNS): A new software system for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.
19. Jones, T. A., Zhou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these maps, *Acta Crystallogr. A47*, 110−119.
20. Dickerson, R. E., Goodsell, D. S., and Kopka, M. L. (1996) MPD and DNA Bending in Crystals and in Solution, *J. Mol. Biol. 256*, 108−125.
21. Dickerson, R. E., and Chiu, T. K. (1998) Helix bending as a factor in protein/DNA recognition, *Biopolymers 44*, 361−403.
22. El Hassan, M. A., and Calladine, C. R. (1997) Conformational characteristics of DNA: Empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps, *Philos. Trans. R. Soc. London, Ser. A 355*, 43−100.
23. Lavery, R., and Skelnar, H. (1988) The definition of generalized helicoidal parameters and the axis curvature for irregular nucleic acids, *J. Biomol. Struct. Dyn. 6*, 63−91.
24. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., and Schneider, B. (1992) The Nucleic Acids Database: A comprehensive relational database of three-dimensional structures of nucleic acids, *Biophys. J. 63*, 751−759.
25. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., and Zhurkin, V. B. (1989) Definitions and nomenclature of nucleic acid structure parameters, *EMBO J. 8*, 1−4.
26. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X.-J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.-S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry, *J. Mol. Biol. 313*, 229−237.
27. Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J., and Dickerson, R. E. (1997) Defining G−C specificity in the minor groove: Side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G, *Structure 5*, 1033−1046.
28. Ducros, V. M.-A., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P., Murshudov, G. N., Wilkinson, A. J., and Brannigan, J. A. (2001) Crystal Structure of GerE, the Ultimate Transcriptional Regulator of Spore Formation in *Bacillus subtilis*, *J. Mol. Biol. 306*, 759−771.
29. Vannini, A., Volpari, C., Gargiolo, C., Muraglia, E., Cortese, R., De Francesco, R., Neddermann, P., and Di Marco, S. (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA, *EMBO J. 21*, 4393−4401.
30. Zhang, R.-g., Pappas, T., Brace, J. L., Miller, P. C., Oulmassov, T., Molyneaux, J. M., Anderson, J. C., Bashkin, J. K., Winans, S. C., and Joachimiak, A. (2002) Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA, *Nature 417*, 971−974.
31. Grebe, T. W., and Stock, J. B. (1999) The histidine protein kinase superfamily, *Adv. Microb. Physiol. 41*, 139−227.
32. Stewart, V., and Bledsoe, P. J. (2003) Synthetic *lac* operator substitutions for studying the nitrate- and nitrite-responsive NarX-NarL and NarQ-NarP two-component regulatory systems of *Escherichia coli* K-12, *J. Bacteriol. 185*, 2104−2111.
33. Boucher, P. E., Maris, A. E., Yang, M.-S., and Stibitz, S. (2003) The response regulator BvgA and RNA polymerase α subunit C-terminal domain bind simultaneously to different faces of the same segment of promoter DNA, *Mol. Cell 11*, 163−173.
34. Olekhnovich, I. N., and Kadner, R. J. (2002) Mutational screening and affinity cleavage analysis of the UhpA binding sites in the *Escherichia coli uhpT* promoter, *J. Bacteriol. 184*, 2682−2691.
35. White, C. E., and Winans, S. C. (2005) Identification of amino acid residues of the *Agrobacterium tumefaciens* quorum-sensing regulator TraR that are critical for positive control of transcription, *Mol. Microbiol. 55*, 1473−1486.
36. DeLano, W. L. (2002) *The Pymol Molecular Graphics System*, DeLano Scientific, South San Francisco, CA.
37. Evans, S. V. (1993) SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules, *J. Mol. Graphics 11*, 134−138, 127−128.
38. Dickerson, R. E. (1998) DNA bending: The prevalence of kinkiness and the virtues of normality, *Nucleic Acids Res. 26*, 1906−1926.
39. Lu, X.-J., Shakked, Z., and Olson, W. K. (2000) A-form conformational motifs in ligand-bound DNA structures, *J. Mol. Biol. 300*, 819−840.

BI050734U