doi:10.1016/j.jmb.2010.03.003 J. Mol. Biol. (2010) 398, 248–263

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

NMR Solution Structure and DNA-binding Model of the DNA-binding Domain of Competence Protein A

Carey A. Hobbs[1], Benjamin G. Bobay[1,2], Richele J. Thompson[1], Marta Perego[3] and John Cavanagh[1]*

[1]*Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA*

[2]*North Carolina Research Campus, Kannapolis, NC 28081, USA*

[3]*Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA*

Received 19 December 2009; received in revised form 2 March 2010; accepted 3 March 2010
Available online 17 March 2010

Edited by M. F. Summers

Competence protein A (ComA) is a response regulator protein involved in the development of genetic competence in the Gram-positive spore-forming bacterium *Bacillus subtilis*, as well as the regulation of the production of degradative enzymes and antibiotic synthesis. ComA belongs to the NarL family of proteins, which are characterized by a C-terminal transcriptional activator domain that consists of a bundle of four helices, where the second and third helices (α8 and α9) form a helix-turn-helix DNA-binding domain. Using NMR spectroscopy, the high-resolution 3D solution structure of the C-terminal DNA-binding domain of ComA (ComAC) has been determined. In addition, surface plasmon resonance and NMR protein–DNA titration experiments allowed for the analysis of the interaction of ComAC with its target DNA sequences. Combining the solution structure and biochemical data, a model of ComAC bound to the ComA recognition sequences on the *srfA* promoter has been developed. The model shows that for DNA binding, ComA uses the conserved helix-turn-helix motif present in other NarL family members. However, the model reveals also that ComA might use a slightly different part of the helix-turn-helix motif and there appears to be some associated domain re-orientation. These observations suggest a basis for DNA binding specificity within the NarL family.

© 2010 Elsevier Ltd. All rights reserved.

Keywords: competence; ComA; NMR; two-component; response regulator

Introduction

The integration of newly acquired DNA is implicated in the continued development of the pathogenic nature of bacteria and has been shown to be responsible for development of the antibiotic resistance of bacteria.[1–5] The development of genetic competence in the Gram-positive bacterium *Bacillus subtilis* is regulated by the ComP-ComA two-component system.[6–10] Competence protein P (ComP) is a transmembrane histidine kinase that autophosphorylates in response to the extracellular signaling molecule competence protein X (ComX).[11,12] Phosphorylated ComP then activates the response regulator, competence protein A (ComA) by transferring a phosphoryl group. Upon phosphorylation, ComA functions as a transcriptional activator of a variety of genes/operons. Among these, the *srfA* operon ultimately controls the expression of the late *com* genes,[7,13,14] which express the protein machinery that binds to and takes up exogenous DNA.[15–18]

Structurally, ComA consists of two domains; an N-terminal regulatory domain and a C-terminal variable effector domain. The N-terminal domain contains a highly conserved aspartic acid residue at position 55, which is the site of phosphorylation by ComP.[19,20] Through sequence alignment and secondary structure prediction of the C-terminal effector domain, it was determined that ComA belongs to the NarL family of response regulator proteins (Fig. 1).[14,20] This family is characterized by a C-terminal transcriptional activator domain that consists of a grouping of four α-helices in which the second and third helices form a helix-turn-helix DNA-binding domain, making the C-terminal domain responsible for the interaction with and activation of the target DNA sequence. The *srfA* operon is the best characterized binding target of

*Corresponding author. E-mail address: john_cavanagh@ncsu.edu.

Abbreviations used: SPR, surface plasmon resonance; PDB, Protein Data Bank; HADDOCK, high ambiguity driven protein–protein docking; MALDI-TOF MS, matrix-assisted laser/desorption ionization time-of-flight; PCA, principal components analysis; AIR, ambiguous interaction restraint.

0022-2836/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.

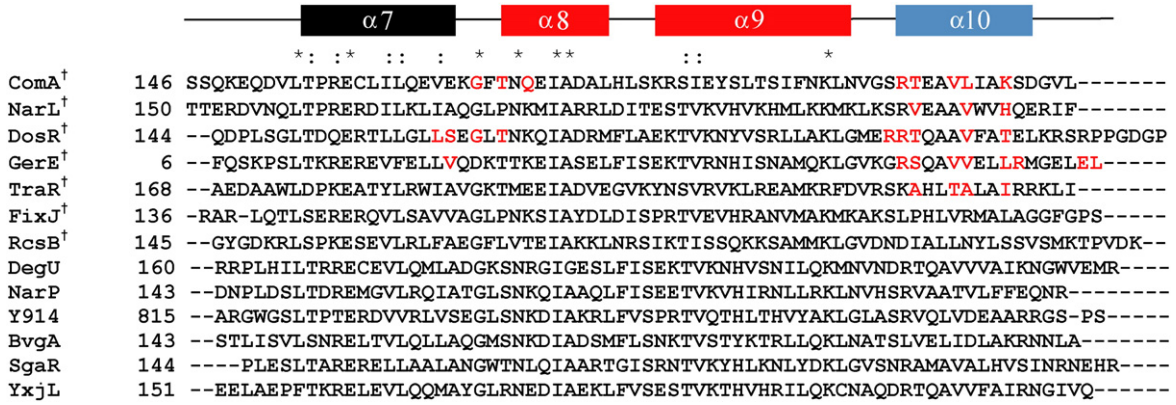


Fig. 1. Sequence alignment of the C-terminus of competence protein A (ComAC) with other members of the NarL family of proteins. The secondary structure prediction is shown on top. The red coloring of the secondary structure represents the helix-turn-helix domain and blue represents the helix involved in the dimerization interface. Residues in red highlight those residues identified as being involved specifically in the dimerization interface. An asterisk (*) denotes conserved residues, a colon (:) denotes conserved substitutions and a dagger (†) denotes proteins with known structures. The known structures include: ComAC (current work), NarL (PDB 1JE8),[35] DosR (PDB 1ZLJ),[36] GerE (PDB 1FSE),[37] RcsB (PDB 1P4W),[38] FixJC (1X3U),[39] and TraR (1H0M and 1L3L).[40]

ComA due to its crucial role in the development of competence.[7,10,13,14,16,21] However, in addition to the *srfA* operon, other operons have been identified to be regulated directly by ComA. These include *degQ* ,[22] which encodes a regulator of degradative enzyme production, and *rapA*,[23] *rapC*, [24] *rapE*,[25] and *rapF*, which encode transition phase regulatory proteins.[26–28] Whole genome DNA microarray analysis identified these genes and at least 89 genes in 35 operons to be affected by ComA independent of its role in competence.[29] Original analysis of the binding site of ComA to the *srfA* operon identified two regions of dyad symmetry crucial to the positive control of *srfA* transcription. This resulted in a proposed consensus sequence containing an inverted repeat with two 6 bp half-sites with a 4 bp spacer; 5′-TTGCGGnnnnCCGCAA-3′.[10,13,14] More recently, alignment of the ComA DNA-binding sites of the targets known to be activated directly by ComA identified a third half-site important for transcriptional activation.[30]

Despite a myriad of previous studies, a structural model of the ComA–DNA binding interaction has remained elusive. To address this, we solved the high-resolution NMR solution structure of the C-terminal DNA-binding domain of ComA (ComAC, residues 146–214).

In addition, DNA-binding studies using surface plasmon resonance (SPR) and NMR protein–DNA titration experiments were used to analyze the interaction of ComAC with its DNA targets. Combining the ComAC structure with biochemical data allowed a model to be developed of ComAC binding to the target DNA sequences. Subsequent principal component analysis (PCA) of the structural ensemble revealed an anti-correlated migration of the helix-turn-helix domains, while the protein dimerization interface possessed cooperative migration. This scissor-like motion might assist DNA binding.

Results

Characterization of ComA and its domains

Oligomerization state of ComA

The N-terminal domain of ComA (ComAN) comprises residues 1–126 of the full-length protein while the C-terminal domain of ComA (ComAC) comprises residues 146–214 of the full-length protein. The multimerization state of ComAC as well as ComAN, ComAN-D55N, and ComAFL (full-length ComA) were determined by size-exclusion chromatography (Fig. 2). At all concentrations tested (10 μM – 1 mM), ComAC eluted at a molecular mass consistent with the formation of a dimer in solution. At all concentrations tested (10 μM – 500 μM), ComAFL was found to be a dimer, corroborating earlier results.[30] ComAN, as well as ComAN-D55N, was found to exist as a dimer in solution at all concentrations tested (10 μM – 500 μM).

Thermostability of ComA and its domains

The thermal denaturation midpoint (T_m) was monitored as a change in molar ellipticity at 222 nm as the temperature was increased in 1deg. increments from 10 °C to 90 °C for ComAFL, ComAN and ComAC. The denaturation profile (Fig. 2) of both ComAFL and ComAC demonstrated a single transition, at 63 °C, whereas ComAN does not appear to have a distinct transition point. Comparison of the profiles of ComAN and ComAC indicates that the C-terminus is responsible for the stabilization of the full-length protein, while the N-terminus works to destabilize the full-length protein.

ComAC dimer subunit exchange

A combination of column chromatography and matrix-assisted laser desorption/ionization time-of-

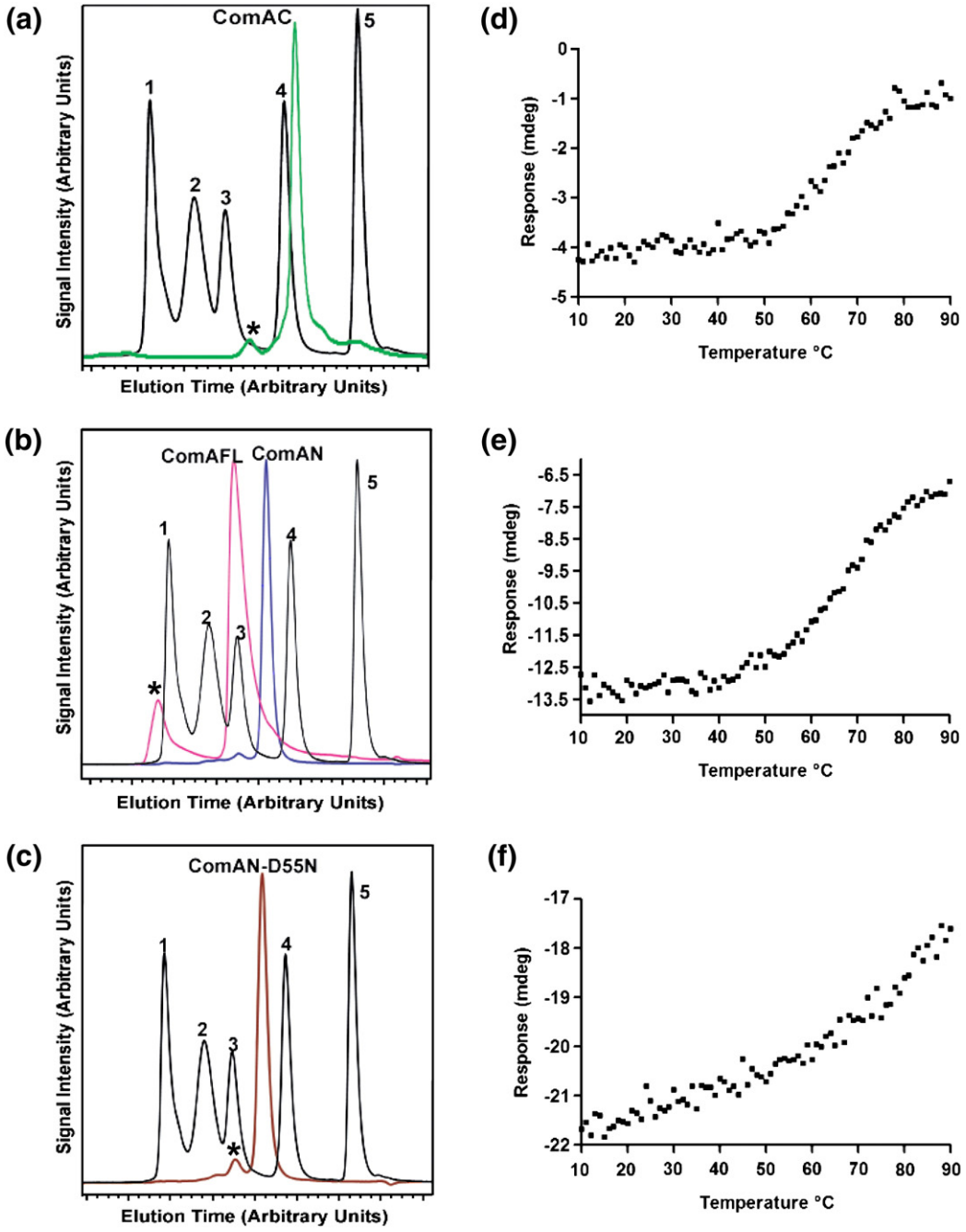


Fig. 2. Characterization of ComA. Size-exclusion chromatography of: (a) ComAC (green), (b) ComAFL (pink), ComAN (blue), and (c) ComAN-D55N (brown). Standards are in black: 1 denotes thyroglobulin (670 kDa), 2 denotes bovine γ-globulin (158 kDa), 3 denotes chicken ovalbumin (44 kDa), 4 denotes equine myoglobin (17.5 kDa) and 5 denotes vitamin B_{12} (1350 Da). An asterisk (*) identifies peaks resulting from aggregation at the highest concentration. Thermal denaturation profiles of ComA: (d) ComAC, (e) ComAFL and (f) ComAN, by monitoring the A_{222} nm over the temperature range 10 – 90 °C.

flight (MALDI-TOF) mass spectrometry was used to determine whether subunit exchange occurs between the individual subunits comprising the ComAC dimer complex, an essential aspect to NMR solution structure determination. Briefly, a sample of half His-tagged-ComAC and half untagged-ComAC were combined and applied to a Ni-NTA column and eluted as in the original protocol. Eluted fractions containing protein were combined and analyzed by MALDI-TOF mass spectrometry. Both the His-tagged-ComAC and untagged-ComAC were recovered, confirming ComAC undergoes subunit exchange (Supplementary Data Fig. S1).

Structural determination of ComAC

Heteronuclear multidimensional NMR methods were used to assign ^{1}H, ^{13}C, and ^{15}N backbone and side chain resonances, provide sequential connectiv-

ities and define distance relationships between protons, both inter- and intra-molecular, for the development of a high-resolution solution structure of ComAC.[31] The ^{1}H-^{15}N HSQC spectrum of ComAC shows good spectral dispersion; all NH resonances were assigned with the exception of Gln148 (Supplementary Data Fig. S2). Ultimately, 97% of backbone resonance assignments (HN, N, C^{α} and C′) were achieved while 85% of the total assignable side chain atoms were assigned. The dimerization interface was determined by performing $^{12}C/^{13}C$ isotope edited experiments using 3D ^{13}C nuclear Overhauser enhancement spectroscopy (NOESY)- heteronuclear single quantum coherence (HSQC) experiments with a mixing time of 120 ms.[32]

Statistics for the 10 lowest-energy, water-refined structures are given in Table 1. These structures were determined with 2316 unambiguous restraints, 1323 ambiguous restraints with 700 intraresidue, 791 sequential, 1050 medium range, and 1098 long range. The 10 structure ensemble had zero NOE and/or hydrogen bond violations and dihedral violations per structure. They showed an average C^{α}-rmsd of 0.20±0.04 Å for secondary structure (backbone), 0.56±0.06 Å secondary structure (heavy), 0.39±0.11 Å backbone and 0.90±0.06 Å heavy. The achieved energies of –1416±–9 kcal mol^{-1} for van der Waals and –5325±–98 kcal mol^{-1} for electrostatic, in addition to the Ramachandran analysis with 99.9% in the generously allowed or better conformation indicate that the reported structure for ComAC is of excellent quality (Fig. 3).

Table 1. ComAC NMR and refinement statistics

	ComAC dimer[a]
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	3,639
Intramolecular	3,150 (~87%)[b]
Intermolecular	489 (~13%)[c]
Unambiguous	2,316
Intraresidue	700
Sequential (|*i*-*j*| =1)	791
Medium range (|*i*–*j*| <4)	1050
Long range (|*i*–*j*| >5)	1098
Ambiguous	1323
Hydrogen bonds	103
Total dihedral angle restraints	
φ	172
ψ	172
Structure statistics	
Energies (kcal mol^{-1})	
van der Waals	-1416±-9
Electrostatic	-5325±-98
Average violations per structure	
NOEs and /or hydrogen bonds	0
Dihedrals	0
Violations (mean and S.D.)	
Distance constraints (Å)	0.035±0.005
Hydrogen bonds (Å)	0.054±0.002
Dihedral angle constraints (degrees)	0.672±0.024
Deviations from idealized geometry	
Bond lengths (Å)	0.00470±0.00007
Bond angles (degrees)	0.615±0.008
Impropers (degrees)	1.60±0.04
Average pairwise r.m.s.d.[d] (Å)	
Secondary structure (backbone)	0.20±0.04
Secondary structure (heavy)	0.56±0.06
Backbone	0.39±0.11
Heavy	0.90±0.06
Ramachandran analysis	
Most favored	84.2%
Additional allowed	10.3%
Generously allowed	5.4%
Disallowed	0.1%

[a] Statistics for all dimer residues.
[b] ~87% of total NOE's.
[c] ~13% of total NOE's.
[d] Pairwise r.m.s.d. was calculated among the 10 lowest energy refined structures.

Dimerization interface of ComAC

ComAC dimerizes in a parallel fashion (Fig. 3), similar to the other proteins in the NarL family. Residues involved in the dimerization interface were determined from the ^{13}C-NOESY-HSQC $^{12}C/^{13}C$ isotope edited NMR experiments. These restraints included NOEs between Val205 and Ile207′, Thr202 and Ala204′, Leu206 and Ala208′, Ile207 and Lys209′, Lys209 and Lys209′, and Ser210 and Ser210′. The amino acid residues contributing the most to the dimer interface were determined by calculating the change in surface accessibility between a monomeric subunit and the dimer complex using NACCESS†.[33] Upon dimerization, a surface area of 676.83 $Å^2$ of each monomeric subunit becomes buried, which accounts for about 15% of the solvent-accessible surface area per subunit. In α10, Thr202 and Leu206 contribute the most, 14.36% and 14.58% of the interface surface, respectively. These are followed by Arg201 (11.43%), Lys209 (10.17%) and Val205 (8.65%). Residues Gly168 (6.52%), Thr170 (8.21%) and Gln172 (6.19%), located in the coil between α7 and the initial part of α8, are also involved. ComAC is composed of four α-helices with the second and third helices (α8 and α9) assuming a conformation typical of helix-turn-helix DNA-binding domains. This domain must then be responsible for the transcriptional activation of the DNA targets of ComAC. In the unbound ComAC structure, these helices lie at an angle of 116.0° to each other.

DNA interactions

Determination of binding affinities using surface plasmon resonance

SPR analysis was used to determine the binding affinities of the wild type *srfA* consensus promoter sequence as well as two different consensus sequences (from Ref. 30). The apparent dissociation equilibrium constants (K_d) calculated using a steady-state affinity model indicated that ComAC binds to its target DNA promoter sequences with

† Hubbard, S. J. & Thornton, J. M. (1993). NACCESS, computer program. Department of Biochemistry and Molecular Biology, University College London.

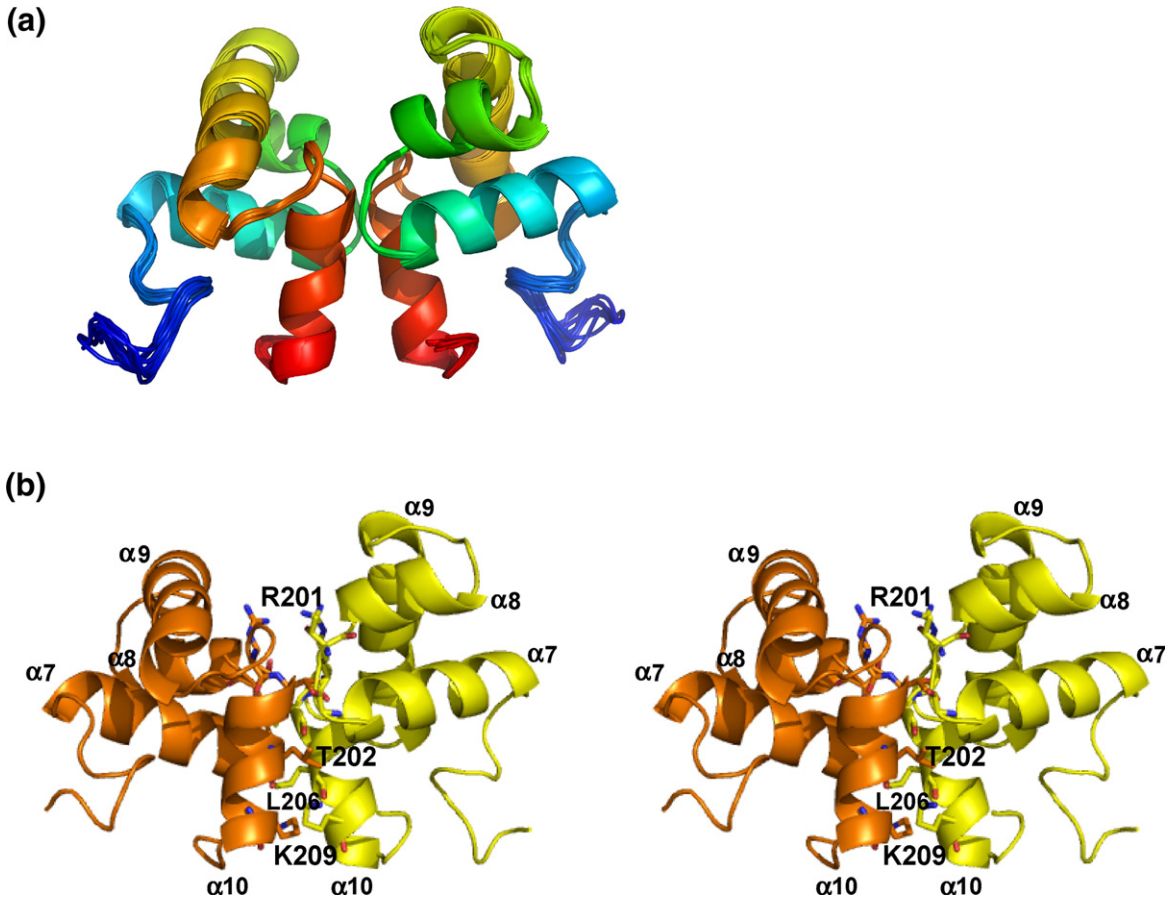


Fig. 3. Solution structure of ComAC. (a) Ensemble of the top 10 lowest-energy structures of ComAC. (b) Stereoview of ComAC dimer with subunits A and B colored orange and yellow, respectively. Residues involved in the dimer interface are represented in stick form.

extremely low nanomolar affinity (Fig. 4). ComAC bound to the wild type *srfA* DNA sequence with an average (±sd) K_d of 9.59±4.01 nM. ComAC bound to the consensus RE1–RE3 sequence with an average K_d of 11.27±2.60 nM and to the consensus RE3 sequence with an average K_d of 4.94±3.06 nM.

ComAC: DNA NMR titration experiments

A series of ComAC–DNA NMR titration experiments were done using the DNA sequences described above. Chemical shift changes, line broadening or peak disappearance were followed by $^{1}H/^{15}N$-HSQC experiments. The residues that were found to be affected by the addition of the DNA sequences include: Thr170, Asn171, Gln172, Ile174, Ala175, Ser181, Lys182, Ser184, Ile185, Thr190, Ser191, Phe193, Leu196, Val198, Gly199, Arg201 and Thr202 (Fig. 5; Supplementary Data Fig. S3). The residues most affected showed peak broadening to the point of disappearance. Arg201 disappeared after titration of a ratio of DNA:protein of 0.25:1. Thr170, Asn171, Gln172, Ser181, Lys182, Thr190, Phe193, Gly199 and Thr202 disappeared at a DNA:protein ratio of 0.5:1. Ala175, Ser184, Ile185, Ser191, Leu196 and Val198 disappeared at a DNA:protein ratio of 0.75:1. The titration was taken to a 4:1 DNA:protein ratio.

HADDOCK ComAC:DNA interaction modeling

The biomolecular modeling program HADDOCK was used to generate a structural model of the interaction of ComAC with the wild type *srfA* DNA sequence. The chemical shift perturbation experiments were used to define residues involved in the interaction. The resulting model and the corresponding refinement statistics can be seen in Fig. 6a and Table 2, respectively. An average C^{α}-rmsd of 1.43±0.55 Å was obtained for the complex interface with an average buried surface area of 1541.36±154.82 Å. PROCHECK analysis[34] of the 10 lowest-energy structures revealed that 97% were in the allowed regions of Ramachandran space. Comparison of the solved ComAC structure and the ComAC–*srfA* modeled complex indicates that the monomeric subunits maintain a similar global fold upon binding DNA. Analysis of the interaction interface from the 20 best structures, taken from the lowest-energy cluster, show the position of the ComAC dimer in consecutive major grooves with helix α9 extending into the major groove to make base-pair-specific interactions. At the dimerization interface, 61% of the intersubunit contacts were conserved upon binding DNA.

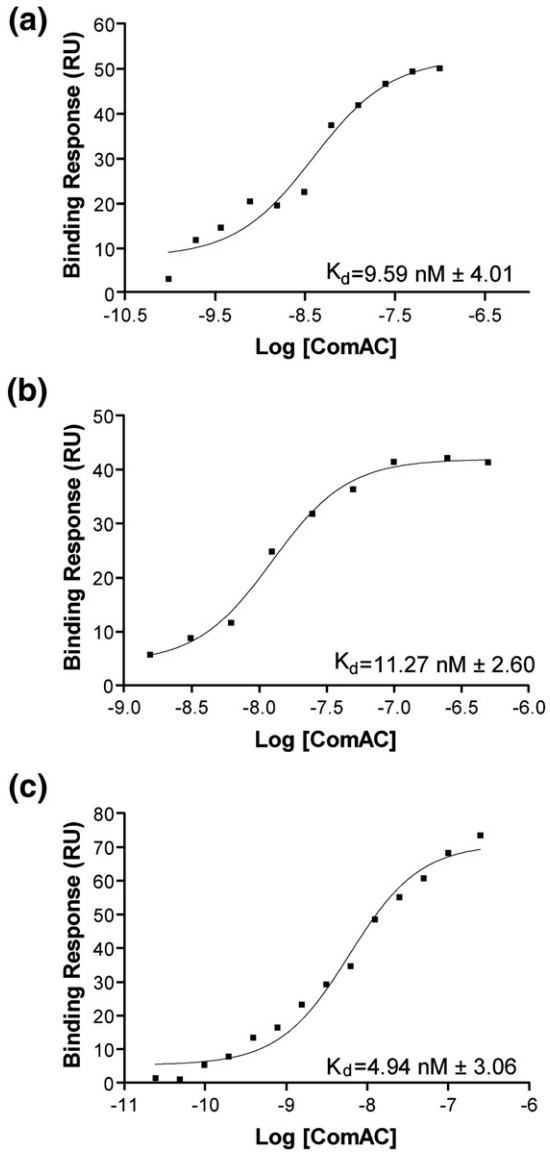


Fig. 4. Steady-state affinity model binding curves. The apparent dissociation (equilibrium) constants (K_d), calculated using a steady-state affinity model. (a) Wild type *srfA* DNA sequence with an average K_d of 9.59 nM. (b) Consensus RE1–RE3 sequence with an average K_d of 11.27 nM. (c) Consensus RE3 sequence with an average K_d of 4.94 nM.

Analysis of the most populated hydrogen bond interactions between the protein and DNA shows extensive hydrogen bonds. Phosphate backbone contacts include the HE# group of Gln172, the HA, HB# and HG groups of Ser181, the HB# and HG groups of Ser184, the HH group of Tyr187, the HB# group of Ser188 and the HG1 group of Thr190. Major groove hydrogen bond contacts primarily include the HH2# group of Arg183 and HE group of Tyr187. In addition, van der Waals interactions with the major groove include the CZ group of Arg183 and the CE group of Tyr187. Calculation of the C^{α}-rmsd of the unbound ComAC structure and the ComAC–*srfA* modeled complex indicate that the greatest deviations occur in the helix-turn-helix DNA-binding domain, α8 and α9 (Figs. 6b and 7). In α8, residues Gln172, Ile174, Asp176, Ala177, Leu178 and His179 showed the highest variation. In α9, residues Ser181, Lys182, Arg183, Tyr187, Ser188 and Leu189 showed the highest variation. This indicates that the helix-turn-helix DNA-binding domain is the most affected when ComAC interacts with its DNA binding partner.

Discussion

The competence transcription factor ComA is responsible for the transcriptional activation of several DNA targets.[30] These DNA targets include *srfA*, *degQ*, *rapA* and *rapC* promoter regions; the products of these genes control everything from competence development, production of degradative enzymes and antibiotic synthesis to sporulation. ComA is involved specifically in the transcriptional activation of the *srfA* operon, making it an essential player in the development of genetic competence in

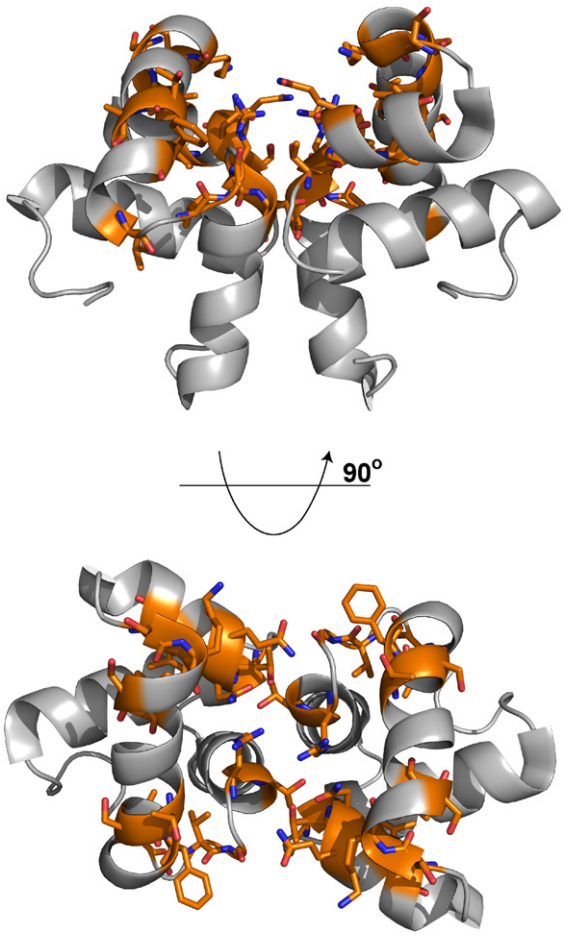


Fig. 5. ComAC NMR solution structure with residues identified by the NMR titration studies as being affected upon the addition of DNA are in stick form and highlighted orange.

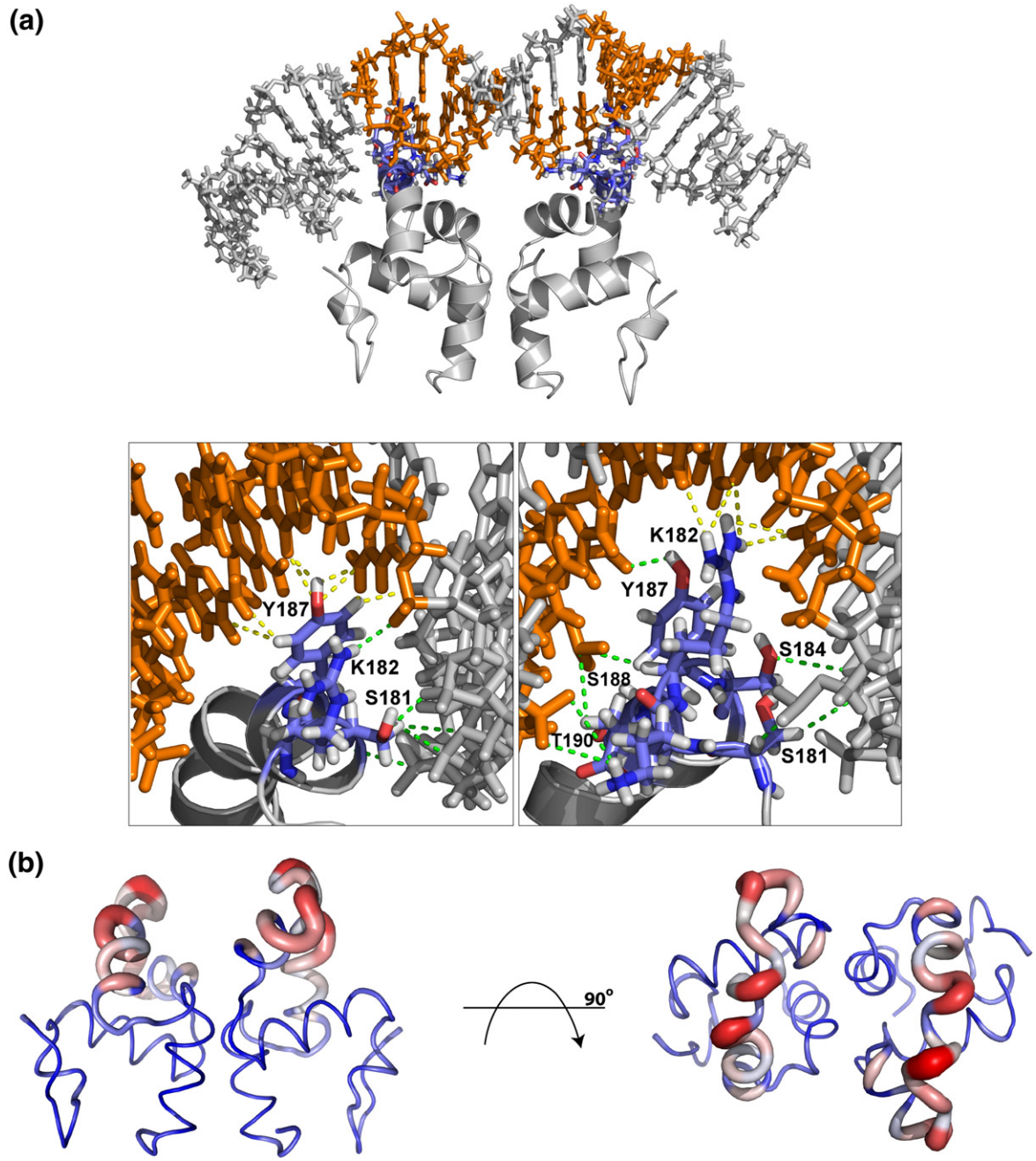


Fig. 6. Analysis of ComAC DNA interaction. (a) Structural model of the ComAC and DNA interaction. Residues and nucleotide bases involved in the interaction are highlighted in blue and orange, respectively. Insets show a detailed look at the interaction. Green dashes designate backbone interactions and yellow dashes designate major groove interactions. (b) Representation of C^{α}-rmsd values of unbound ComAC and ComAC bound to DNA. Blue represents the smallest difference, white represents the intermediate difference and red represents the greatest difference.

B. subtilis. Here, we determined the high-resolution solution structure of the C-terminal DNA-binding domain of ComA. This domain is present in solution in dimeric form, with each monomeric unit containing a helix-turn-helix DNA-binding motif (α8–α9). Each monomer consists of a grouping of four helices and the dimer is held together by interactions running along the length of α10.

ComA is a member of the NarL family of response regulators. Within this family, six structures have been solved including (Fig. 1) full-length NarL, the C-terminus of NarL (NarLC) in complex with DNA,[35] the C-terminus of DosR (DosRC),[36] the C-terminus of GerE (GerEC),[37] the C-terminus of RcsB (RcsBC),[38] the C-terminus of FixJ (FixJC)[39] and the full-length TraR.[40] ComAC has a sequence identity of 24%, 24%, 23%, 23%, 26% and 22% with each protein, respectively. As is common in many proteins with helix-turn-helix DNA-binding domains, ComAC forms a dimer in solution. This is the case also for the family members NarLC, DosRC, GerEC and TraR. Two particularly well

Table 2. HADDOCK Refinement Statistics for ComAC DNA binding model

Violations (mean and SD)	
Distance constraints (Å)	0.43±0.02
Deviations from idealized geometry	
Bond lengths (Å)	0.745±0.000008
Bond angles (°)	7.253±0.0005
Impropers (°)	0.624±0.007
Average pairwise r.m.s.d.[a] (Å)	
Interface	1.43±0.55
Buried surface area	1541.36±154.82
Ramachandran analysis	
Favored	82.1%
Allowed	14.8%
Generously allowed	0.1%
Disallowed	3.0%

[a] Pairwise r.m.s.d. calculated based on all heavy atoms at the interface.

studied structures within the NarL family are NarLC and DosRC; therefore, comparisons between ComAC and these proteins will be made where useful.

The C-terminal domains ComAC, NarLC and DosRC are each composed of four α-helices with the second and third helices comprising the helix-turn-helix DNA-binding domain. While the overall architecture is similar, slight differences in helical orientation are evident. In the unbound ComAC structure these helices lie at an angle of 116.0° with respect to each other and this angle increases to 119.6° upon binding to the *srfA* DNA sequence. In the unbound DosRC, they lie at an angle of 106.7° with respect to each other, but upon binding DNA this angle decreases to 102.0°. Two structures of NarL have been solved, unbound full-length (monomer) and the C-terminus bound to DNA (dimer). In the full-length unbound structure of NarL, these helices lie at an angle of 98.8° with respect to each other, which increases to 108.7° when bound to its DNA target. The functionality of the angle between the second and third α-helices is presumably due to the binding of inverted repeats in their DNA targets, such as 5′-TTGCGGnnnnCCGCAA-3′ in the case of ComAC.[30] The difference seen between ComAC, NarLC and DosRC is most likely due to the specific nature of these proteins binding their DNA targets in a slightly different manner.

Another shared feature of the NarL family is the organization of the dimer interface. The interface is formed mainly by α10 from each of the two

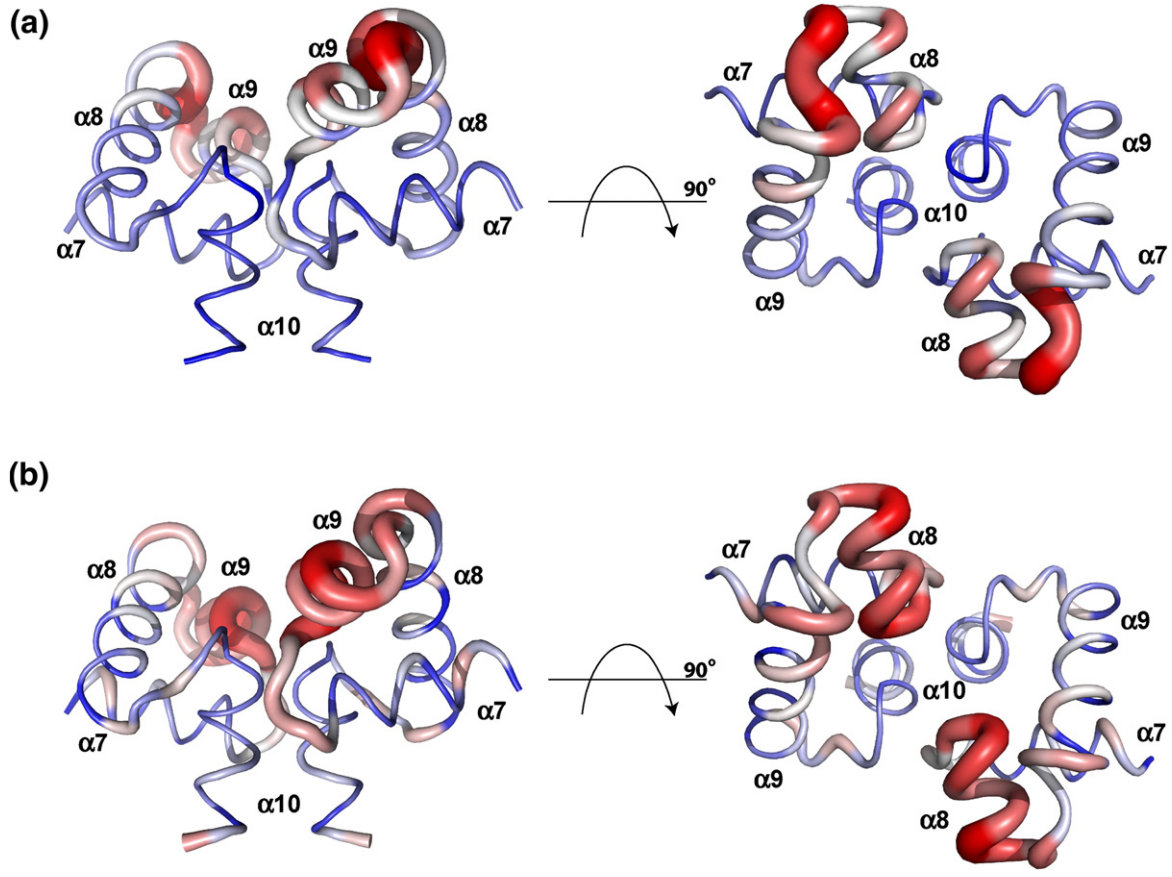


Fig. 7. Calculation of C^{α}-rmsd variation of ComAC and its homologues in their DNA-bound state. The C^{α}-rmsd differences calculated between ComAC and (a) NarLC and (b) DosRC. Blue represents the smallest difference, white represents the intermediate difference and red represents the greatest difference in addition to the increasing size of the cartoon.

monomeric subunits, along with the loop between α7 and α8 and the initial part of helix α8 (Fig. 1). The angle between the two subunits in the unbound structures, measured at the α10 helices, is 38.2° in ComAC and 28.8° in DosRC. Upon DNA binding this angle decreases to 36.1° in ComAC and increases slightly to 30.5° in DosRC. A significant hydrophobic patch comprises the dimerization interface, primarily residues $_{204}$AVLIA$_{208}$ of ComAC, $_{206}$AAVWV$_{210}$ of NarLC and $_{200}$AAVFA$_{204}$ of DosRC (Fig. 1). ComAC has a buried surface area of about 677 Å^2 per monomer upon dimerization, accounting for 15% of the solvent-accessible surface per subunit, while the interface of NarLC and DosRC accounts for about 10% of the solvent-accessible surface per subunit.[35,36] The residues at the primary dimer interface of ComAC are Thr202 (Thr198 in DosRC, Val204 in NarLC), Val205 (Val202 in DosRC and Val208 in NarLC), Leu206 and Lys209. The most divergent residue in the ComAC interface compared to DosRC and NarLC is Lys209. The equivalent residue is Thr205 in DosRC and His211 in NarLC, although

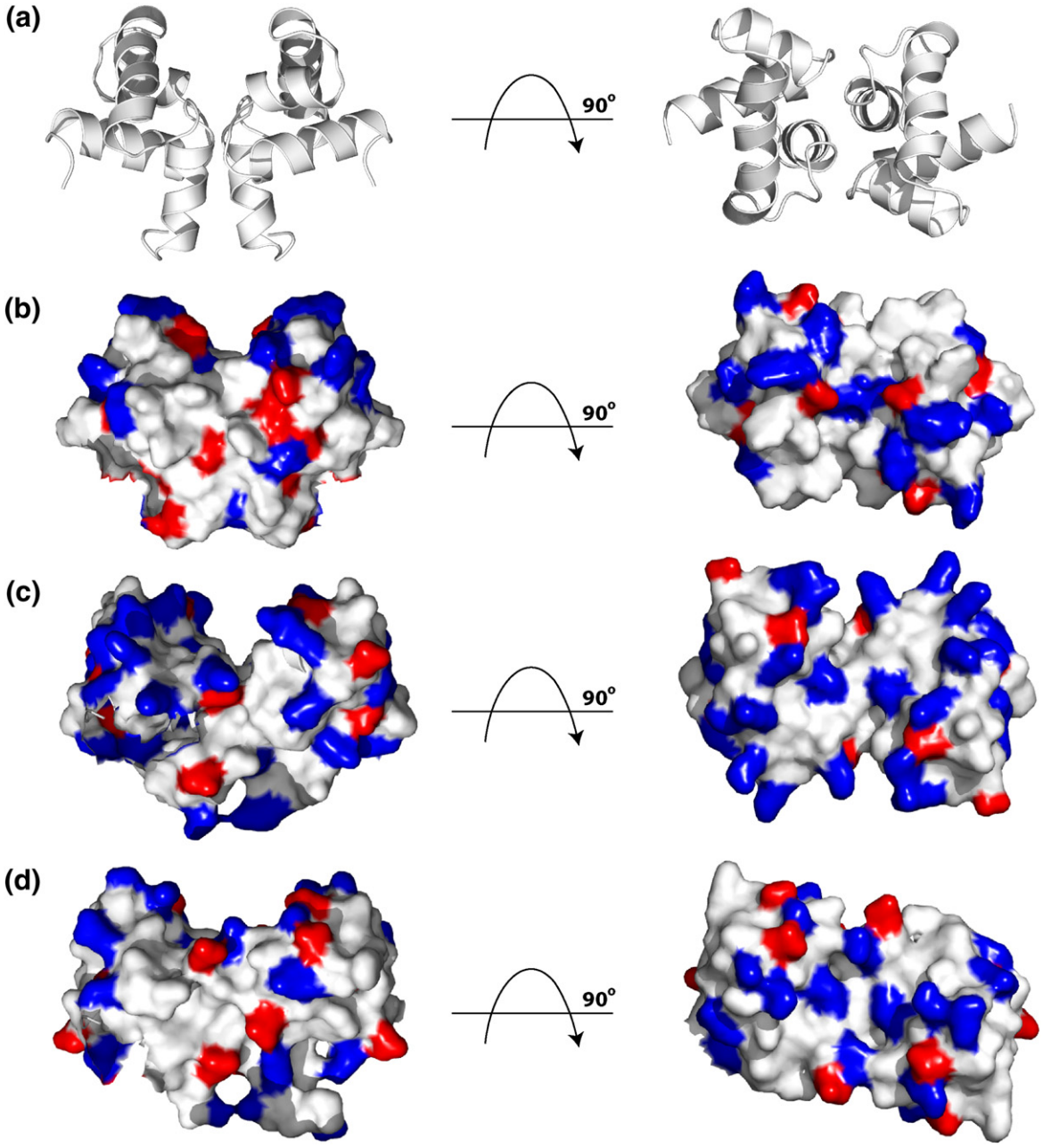


Fig. 8. Highlighting of electrostatic positive and negative amino acid residues. (a) Solution structure of ComAC showing orientation of: (b) ComAC, (c) NarLC, and (d) DosRC. Amino acid residues with positive electrostatic character (R, K and H) are highlighted in blue, those with negative electrostatic character (D and E) are highlighted in red.

these residues do share polar characteristics. Therefore, it appears that the dimerization interface is mostly conserved across these proteins.

As expected, these proteins show similar electrostatic surface characteristics surrounding their DNA-binding helix-turn-helix motifs. A positive electrostatic surface at the beginning of α10 is seen in all three proteins and can be attributed to residues Arg201, Arg203, and Arg197 for ComAC, NarLC and DosRC, respectively (Fig. 8). Indeed, a positively charged residue at this position is highly conserved (85%) across the NarL family members listed in Fig. 1. Consequently, this region is likely to contribute to the general DNA-binding mechanism for the NarL family.

Interestingly, there are slight differences in the electrostatic character of the helix-turn-helix DNA-binding motif among ComAC, NarLC and DosRC. For NarLC, the region of greatest positive electrostatic character is in the middle of α9, which accommodates the residues responsible for DNA binding ($_{188}$KVHVK$_{192}$) (Figs. 1 and 8c). For DosRC, the region of greatest positive electrostatic character is at the beginning of α9 ($_{179}$KTVKN$_{183}$), again containing the residues found to be responsible for DNA binding (Figs. 1 and 8d). While NarLC and DosRC show a shift in residues responsible for DNA binding (middle *versus* beginning of α9) as well as electrostatic characteristics, ComAC displays a further shift in electrostatic character to include the loop region between α8 and α9, as well as the beginning of α9 (Figs. 1 and 8b). This suggests that the residues in this region ($_{181}$SKRS$_{184}$) are responsible for direct contact with its DNA targets and represents yet another contribution to the recognition of different DNA targets within the NarL family.

As seen in Fig. 7, the greatest C^{α}-rmsd difference between NarLC and ComAC is present at the N-terminal end of helix α9 and slightly in helix α8, while for DosRC and ComAC the greatest C^{α}-rmsd difference occurs in helix α8. In both cases, this region comprises the DNA-binding helix-turn-helix motif. When comparing the overall dimeric structure of ComAC with those of NarLC and DosRC over all residues, C^{α}-rmsd values of 4.183 Å and 4.414 Å are observed. The C^{α}-rmsd values between NarLC and DosRC themselves are considerably lower. However, an alignment of the C^{α} atoms over only one subunit lowers the C^{α}-rmsd considerably; 1.608 Å and 1.613 Å C^{α}-rmsd observed for DosRC–ComAC and NarLC–ComAC, respectively. This suggests a slight difference in the orientation of the two ComAC monomeric subunits with respect to one another, compared to the other proteins. It is important to note that despite many structural similarities, proteins within the NarL family do not appear to recognize one another's DNA targets. It is plausible that subunit orientation within the family, along with the shifting of recognition residues within the helix-turn-helix motif discussed above, helps drive DNA specificity (Fig. 7).

The results of the study described here support this hypothesis. SPR, NMR chemical shift perturbation studies, and molecular modeling were used to examine how ComAC can interact with its target DNA sequences. SPR experiments determined the binding constants of three different DNA sequences that contained three recognition elements (RE). Consensus sequences have been determined for each of the three recognition elements. The DNA sequence designated consensus RE3 contains a consensus sequence at position 3 and the consensus RE1–RE3 sequence contains a consensus sequence in all three RE sequences. Similar binding constants, in the low nanomolar range, were obtained for both. Consensus RE3 has a K_d of 4.94±3.06 nM and consensus RE1–RE3 has a K_d of 11.26±2.60 nM. In addition, ComAC bound to the wild type *srfA* DNA sequence with an average K_d of 9.59±4.01 nM. In a DNase footprinting analysis experiment, NarLC showed a similar binding affinity for its DNA target as the full-length NarL protein, with a K_d of 0.15 nM.[35] We were unable to find a binding affinity for DosR and its target DNA sequences in the current literature.

NMR titration experiments between ComAC and the ComA consensus sequence in the *srfA* promoter: 5′-GAATC*TTTCGG*CATC*CCGCAT*GAAA-3′ identified residues essential for DNA binding. These residues include Thr170, Asn171, Gln172, Ile174 and Ala175 (in α8); Ser181 and Lys182 (in the α8–α9 loop); Ser184, Ile185, Thr190, Ser191, Phe193 and Leu196 (in α9), Val198, Gly200, Arg201 and Thr202 (in the α9–α10 loop and beginning of α10). Both NarLC and DosRC have the following residues in corresponding positions involved in their DNA binding interfaces: Thr170, Asn171, Ser181, Ser184, Thr190 and Thr202. In addition, NarLC has residues corresponding to Gln172, Val198 and Gly200, and DosRC also has a residue corresponding to Arg201. These residues all lie in the helix-turn-helix DNA-binding domain of these proteins. Just as there are similarities between the regions responsible for DNA binding between these proteins, these data seem to suggest a greater role of α8 and the α8–α9 loop in DNA recognition for ComAC when compared to DosRC and NarLC.

PCA is a standard tool in the field of multivariate analysis and is used to extract from a set of interrelated variables a much smaller set of variables that retains most of the variation contained in the full set. Using a correlation matrix consisting of elements of the average of the pairwise products of displacements from their average position of landmarks (such as the C^{α} positions in a protein), PCA can be very helpful in identifying, from an NMR structural ensemble, correlations in conformational rearrangements within a protein.[41] Here, we used PCA on the ComAC NMR structural ensemble. Structural PCA plots of the refined NMR structures of the unbound and DNA bound model of ComAC are shown in Fig. 9a and b, respectively). The relative motions of structural elements in the unbound state are very important, because this impacts the binding event. On average, within the dimeric NMR structure ensemble: (i) α8 and α9

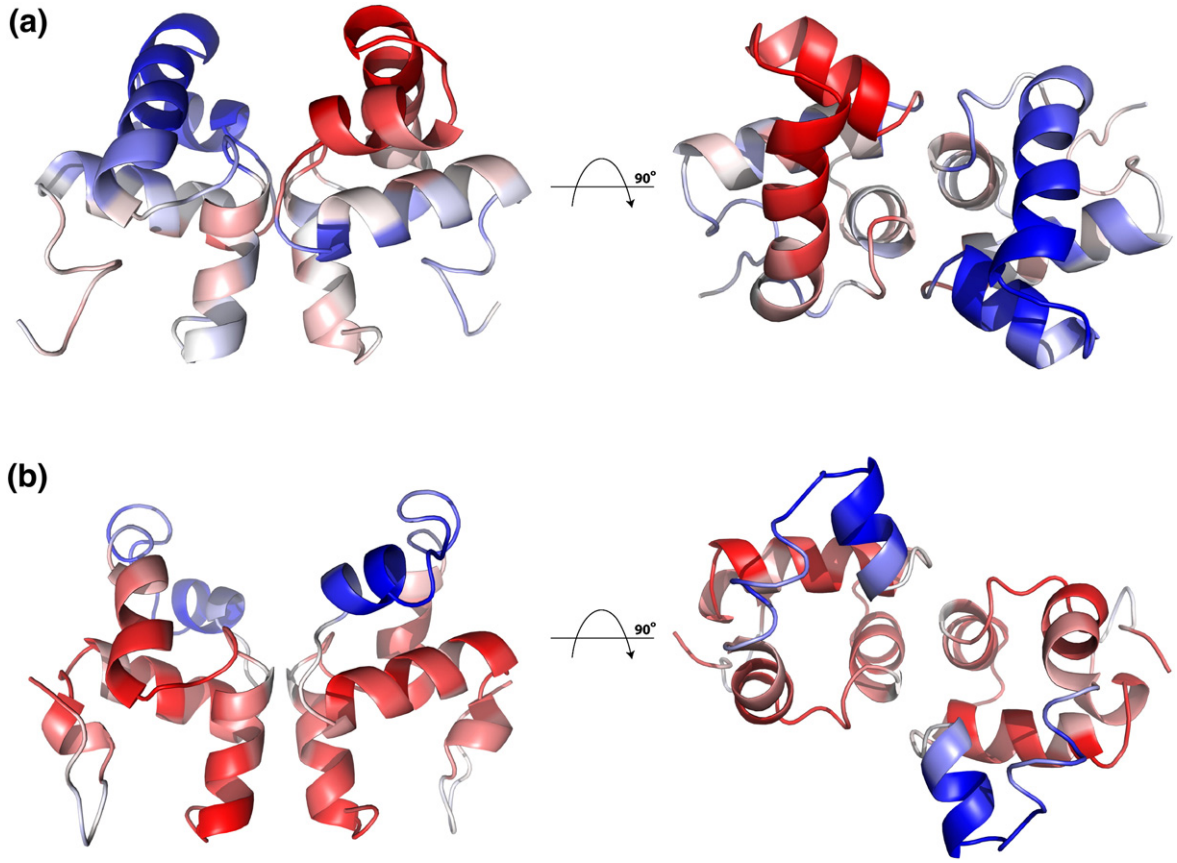


Fig. 9. THESEUS PCA analysis of the unbound and bound structure ensembles. PCA plots of the maximum likelihood correlation matrix for the structure ensembles of (a) unbound ComAC, first principal component and (b) bound ComAC, second principal component. Regions colored similarly (red or blue) are self-correlated, whereas regions colored differently (red *versus* blue) are anticorrelated.

move in the same direction within each monomer, but in an opposite sense to α8′ and α9′ from the other monomer; (ii) α10 and α10′ move in the same direction within the dimer (Fig. 9a).

The anti-correlated motion of α8/α9 and α8′/α9′ is particularly interesting as it suggests that each helix-turn-helix motif can push away from the dimeric interface (Fig. 9a) to space correctly and interact more easily with adjacent recognition elements in the DNA target sequence (Fig. 6a). This binding event might therefore resemble a scissor-like motion of the helix-turn-helix motifs. Concomitant re-orientations during complex formation with target DNA sequences might also enable residues crucial to DNA recognition and binding to make necessary contacts. Interestingly, as noted above for the ComAC–DNA model, the angle between α8 and α9 increases upon DNA binding. This type of motion is seen in the PCA analysis (Fig. 9b). Comparative PCA investigations with other members of the NarL family were not done for the following reasons. PCA studies require ensembles of structures. The structures of NarL/NarLC, DosRC, GerEC and TraR were solved by X-ray crystallography and ensembles of structures are not available from the Protein Data Bank. The NMR structure of FixJC suggests it is monomeric in solution and its mechanism of action is likely somewhat different than those of family members that exist functionally as dimers.

The ComAC:*srfA* model outlined here suggests both similarities and differences in how members of the NarL family of response regulators interact with their DNA targets. To facilitate its protein–DNA interactions, ComAC relies predominantly on the most N-terminal residues of helix α9. In comparison, NarLC mainly utilizes residues in the center of helix α9 while the binding region of DosRC falls between that of ComAC and NarLC. The experiments described here identified amino acid residues in helices α8 and α9 of ComAC as being important for binding to its target DNA sequences. PCA of the ComAC structure suggests that anti-correlated movement of each helix-turn-helix motif within each monomeric subunit of the full dimer structure may facilitate crucial protein–DNA interactions.

Materials and methods

Expression and purification of ComAC

DNA for transformation of *Escherichia coli* was isolated using the QIAprep Miniprep Spin Kit (Qiagen). The purified plasmids were transformed into competent *E.*

coli BL21(DE3)pLysS cells (Novagen). LB broth (1 L containing 100 μg mL^{-1} kanamycin) was inoculated and grown at 37 °C and agitated at 160 rpm until the cells reached an absorbance at 600 nm (A_{600nm}) of 0.8. Isopropyl-β-d-thiogalactopyranoside (IPTG) was added to a concentration of 1 mM, the temperature was reduced to 32 °C and the agitation was reduced to 120 rpm during incubation for 3 h. The cells were pelleted by centrifugation (7500 rpm for 20 min) and suspended in 25 mL of lysis buffer (25 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 5 mM imidazole). In addition to the lysis buffer, 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF; final concentration 0.25 μM) and Triton X-100 (0.01% (v/v) final concentration) were added. The cells were sonicated eight times (3 min of 8 s pulses followed by 2 s rest). The resulting suspension was centrifuged at 17,500 rpm for 15 min. The supernatant was removed and saved as the crude extract. The crude extract was applied to a Ni-NTA affinity resin (Qiagen) that was equilibrated with the lysis buffer. Once the column was loaded with the ComAC crude lysate, it was washed with 100 mL of the lysis buffer. ComAC was eluted from the column using an imidazole gradient of 5 mM to 250 mM. Fractions containing the ComAC were pooled and concentrated to 45 mL, then 100 units of thrombin were added to the concentrated sample, which was kept for 2 h at room temperature. The cleavage was stopped by the addition of 1 mL of Protease Inhibitor Cocktail (Sigma). The cleaved sample was immediately dialyzed extensively against 25 mM NaH_2PO_4, pH 6.5, 100 mM NaCl, 2 mM EDTA, 1 mM DTT. Throughout the protocol, the presence of ComAC was monitored using 12% (w/v) polyacrylamide gel electrophoresis with a Tricine buffer. Production of ^{1}H-^{15}N-^{13}C samples for NMR analysis followed the above protocol with the substitution of minimal medium (M9T) with 1 g of [^{15}N]ammonium chloride and 1 g of [*U*-$^{13}C_6$]glucose for expression.

Construction of ComAN and ComAN-D55N mutant expression vectors

The ComA vector template used for the expression of the 5′ region of the *comA* gene encoding ComAN (residues 1–126) was provided by Dr Dave Dubnau (The Public Health Research Institute, Newark, NJ). Primers were designed to incorporate NheI and SacI restriction endonuclease cleavage sites at the 5′ and 3′ ends of the coding sequence, respectively. The fragment was inserted into the pET28a vector (Novagen). This construct resulted in a thrombin-cleavable His_6 tag at the N-terminus of the recombinant ComAN protein. This results in the production of a protein with a molecular mass of approximately 17,000 Da. A gene encoding a mutant ComAN protein in which an asparagine residue replaces an aspartate residue at position 55 (ComAN-D55N) was constructed. Primer sequences were purchased from Integrated DNA Technologies:

ComAN forward: 5′- CTA GCT AGC ATG AAA AAG ATA CTA GTG-3′
ComAN backward: 5′- GCT GTT TAA AGA GCT CAA ATT ACT ATA AAA TTT CTC-3′
ComAN-D55N forward: 5′- GAT CTC ATT TTA ATG AAC CTG AAT CTA GGC-3′
ComAN-D55N backward: 5′ - GCC TAG ATT CAG GTT CAT TAA AAT GAG ATC-3′

Expression and purification of ComAN and ComAN-D55N

Expression and purification were performed as described for ComAC.

Expression and purification of ComAFL

The ComAFL vector was provided by Dr Dave Dubnau (The Public Health Research Institute, Newark, NJ). Briefly, this is a pET28a vector (Novagen) and results in a C-terminal non-thrombin cleavable His_6 tagged recombinant ComAFL protein. This produces a protein with a molecular mass of approximately 25,000 Da. Expression and purification were performed as described for ComAC.

Size-exclusion liquid chromatography

Size-exclusion liquid chromatography was used to determine the multimerization state of ComAC, ComAN, ComAN-D55N and ComAFL proteins. A protein molecular mass standard (BIO-RAD) containing bovine thyroglobulin (670,000 Da), bovine γ-globulin (158,000 Da), chicken ovalbumin (44,000 Da), equine myoglobin 17 (17,000 Da) and vitamin B12 (1350 Da) was used. Analysis was performed at room temperature using a Waters Breeze HPLC system equipped with a 4.6 mm ID × 30 cm TOSOH TSK-GEL SW2000 size-exclusion column. Sample volumes of 10 μL with concentrations ranging from 10 μM to 1 mM were injected at a flow rate of 0.3 mL min^{-1} for 20 min. A mobile phase of 25 mM NaH_2PO_4, pH 6.5, 100 mM NaCl was used and chromatographs were analyzed with Waters Breeze software version 3.30.

Circular dichroism spectroscopy

Far UV spectra of ComAC, ComAN, ComAN-D55N and ComAFL at a concentration of 60 μM were acquired with an Applied Photophysics Pistar-180 spectrometer, using a buffer of 25 mM NaH_2PO_4, pH 6.5, 100 mM NaCl. A circular 200 μL quartz cuvette (Hellma) with a 0.02 cm path-length was used for all measurements. Spectra were recorded at a wavelength of 222 nm to monitor α-helical content. A midpoint denaturation temperature (T_m) was determined by heating samples from 10 °C to 90 °C at a rate of 1 deg. C min^{-1}. Five replicates were collected for each protein and corrected for the buffer blank.

Subunit exchange of ComAC

ComAC was expressed and purified as described above. Fractions containing ComAC were combined and excess imidazole was removed by dialysis. Half of the sample was set aside (His_6 tagged ComAC) and did not undergo thrombin cleavage. The other half underwent thrombin cleavage as described above. All uncleaved sample was removed by passing the sample through a second Ni-NTA column, resulting in a thrombin-cleaved ComAC (untagged ComAC) sample. The His_6 tagged ComAC and untagged ComAC were then combined and reapplied to the Ni-NTA column and eluted as in the original protocol. Fractions containing ComAC, evaluated by 12% polyacrylamide gel electrophoresis with Tricine

buffer, were combined. Positive ion MALDI-TOF mass spectrometry was done with an Applied Biosystems Voyager Super DE STR. This instrument is equipped with a nitrogen laser (337 nm) to desorb and ionize the samples. A close external calibration, using two points to bracket the mass range of interest, was used. A saturated solution of α-cyano-4-hydroxycinnamic acid in 45:45:10 EtOH/water/formic acid (by vol.) was used as the MALDI matrix. Spectra were obtained over a mass range of 2000 – 20,000 Da in linear mode with an average of 100 laser shots per spectrum. For the MALDI analyses, protein samples were prepared using micro-reversed phase chromatography with C4 zip tips (Millipore). The tip resin was prepared by rinsing repeatedly with 10 μl of acetonitrile containing 0.1% formic acid. This was followed by repeated washing with 0.1% formic acid. The protein was then concentrated by repeated pipetting (10 μL seven times). Salts were removed by extensive washing with 0.1% formic acid. The protein was eluted with 2.5 μl of 70% acetonitrile containing 0.1% formic acid and 0.5 μl of the protein sample was mixed with 0.5 μL of the MALDI matrix and deposited on a 100-well MALDI sample target.

NMR spectroscopy and structure calculations

All experiments were performed at 288.15 K on a Varian INOVA 600 MHz equipped with a Varian cryogenic probe. Protein samples concentrated to 1.0–2.0 mM in the following buffer were used: 90:10% or 1:99% $H_2O/^2H_2O$, 25 mM NaH_2PO_4, pH 6.5, 100 mM NaCl, 2 mM EDTA and 1 mM DTT. Sequential backbone assignments were made from HNCACB, CBCA(CO)NH, HNCA, HNCO and HN(CA)CO experiments.[42–46] Side chains were assigned from H(CCO)NH, C(CO)NH, and HCCH-TOCSY experiments.[42,43,45] Exchange protected amides were monitored by sequentially recording 100 12 min 2D ^1H-^{15}N-HSQC experiments. TALOS was used to determine coupling constants for assigning backbone ψ and φ angles.[47] NOEs were obtained from 3D ^{15}N-NOESY-HSQC experiments with 50, 100 and 120 ms mixing times and 3D ^{13}C-NOESY-HSQC experiments with a mixing time of 120 ms. $^{12}C/^{13}C$ isotope edited experiments using 3D ^{13}C-NOESY-HSQC experiments with a mixing time of 120 ms were used to determine the amino acid residues involved in the dimer interface.[32] This was done by combining ^{13}C-labeled ComAC combined with natural abundance ComAC in a 1:1 ratio. Structures were calculated with NOEs (inter- and intra-molecular), hydrogen bond restraints (inter- and intra-molecular), amide exchange experiments, and ψ and φ angles (TALOS predictions). ARIA (version 1.2) and CNS (version 1.1) programs were used to compute the solution structure starting from an extended structure with random side chain conformations.[48,49] The CNS protocols used simulated annealing with torsion angle and Cartesian space dynamics using the default parameters. Manually assigned inter- and intramolecular NOEs were input to ARIA as assigned and uncalibrated with respect to distance. The total number of ambiguous NOE restraints allowed for each peak on the NOESY spectra was set to 20. Distance restraints, derived from the manually assigned NOEs, were set to 1.8–6.0 Å. The dihedral angle restraints were taken to be ±2 sd values or at least ±20 from the average values predicted by TALOS.[47] In this study, the dihedral angles were restrained to $\varphi = -70 \pm 50°$ and $\psi = -50 \pm 50°$ for the helical regions. The noncrystallographic symmetry energy term was used to keep the C^α atoms of the monomers superimposable, and distance symmetry potential was used to ensure that the relative orientations of all of the C^α atoms of the monomers were symmetric.[49] The spectra were processed with NMRPIPE and analyzed with NMRVIEW on LINUX workstations running Fedora Core.[50,51] Molecules were visualized and aligned using PyMol‡. The 10 lowest-energy structures were further water-refined with ARIA. PCA was done on the NMR structure ensembles by using THESEUS.[41]

Surface plasmon resonance (SPR) ComAC DNA-binding experiments

An ICx Nomadics SensiQ dual channel SPR instrument was used to perform all binding experiments, in which the double-stranded DNA sequences were immobilized on the sensor surface and the protein solution was injected over the surface. To achieve this, neutravidin was immobilized onto a carboxylated surface (COOH1 sensor) over each flow cell channel through amine coupling. Briefly, amine coupling was achieved by activating the surface by injecting 100 μL of a mixture of NHS/EDC over both channels at a flow rate of 50 μL min^{-1}. This was followed by an injection of 100 μL of 50 μg mL^{-1} neutravidin in 10 mM sodium acetate, pH 4.5, over both channels at a flow rate of 5 μL min^{-1}. Then 75 μL of 1 M ethanolamine was injected at a flow rate of 25 μL min^{-1} over both channels to block remaining activated carboxyl groups. The DNA sequences used for these experiments are given below. Oligonucleotide sequences were synthesized by IDT DNA and purified by HPLC. The 5′ end of one strand was biotinylated to facilitate immobilization of the DNA to the neutravidin immobilized on the chip. Complementary strands were annealed by combining equimolar amounts and heating to 95 °C then cooling to room temperature. Sequences used in this study were determined from Ref. 30. The sequences are (italicized/bold portions correspond to a consensus recognition element):

	RE1	RE2	RE3
Wild type *srfA*:	5′-***TTTCGG***catc***CCGCAT***gaaact***TTTCAC***-3′		
Consensus RE3:	5′-***TTTCGG***catc***CCGCAT***gaaact***TTGCGG***-3′		
Consensus RE1–RE3:	5′-***TTGCGG***catc***CCGCAA***gaaact***TTGCGG***-3′		

Biotinylated double-stranded DNA sequences were injected over channel 1 only at a concentration of 1 μM until a response of 100 RUs was achieved. Steady state assay experiments were done at room temperature. A running buffer of 10 mM Hepes, pH 7.6, 2 mM $MgCl_2$, 0.1 mM EDTA, 200 mM KCl, was used throughout the experiment. The steady state assay was performed by injecting increasing concentrations of ComAC over both channels, 125 μL at a flow rate of 5 μL min^{-1}. Upon completion of the binding profiles, all responses were baseline-corrected to an injection of buffer only and subtraction of the reference flow cell (flow cell 2). This was done to correct for non-specific binding and instrument drift. The response value at equilibrium (K_d) was determined using the accompanying ICx Nomadics Q-DAT software.

‡ www.pymol.org

ComAC DNA NMR titration experiments

All experiments were done at 288.15 K on a Varian INOVA 600 MHz equipped with a Varian triple resonance cryogenic probe. The ComAC protein sample was concentrated to 250 μM in a buffer of 90:10% $H_2O/^2H_2O$, 25 mM NaH_2PO_4, pH 6.5, 100 mM NaCl, 2 mM EDTA, 1 mM DTT. DNA samples were titrated into samples that provided the following molar ratios of DNA to protein: 0:1, 0.25:1, 0.5:1, 0.75:1, 1:1, 2:1, and 4:1. A 2D ^1H-^{15}N-HSQC experiment was collected at each titration point. Data were processed using NMRPIPE and analyzed using NMRVIEW.[50,51] Oligonucleotide sequences were synthesized by IDT DNA and purified by HPLC. Complementary strands were annealed by combining equimolar amounts and heating to 95 °C then cooling to room temperature. Sequences used in this study were identical with those used in the SPR studies and determined from Ref. 30.

ComAC:DNA modeling

The DNA pdbs used in the modeling study were constructed through the X3DNA webserver§.[52] The *srfA* operon binding sequence pdb was constructed on the basis of the NarLC DNA pdb,[35] the DNA in this solved structure was a bent/curved alternating *B*- to *A*-DNA conformation. Modeling the ComAC:DNA interaction was completed using the program HADDOCK.[53] Default HADDOCK parameters were used throughout the docking procedure. Experimental restraints known as ambiguous interaction restraints (AIRs) were derived from NMR chemical shift perturbation experiments. The AIR are defined as both active and passive residues. Active residues are defined as those residues shown through the NMR protein:DNA titrations as having an effect upon binding DNA with a solvent accessibility of at least 40% as determined by NACCESS.[33] Passive residues are defined as those surrounding the active residues with at least 40% solvent accessibility.

The DNA sequence used along with its complementary strand was: 5′-GAATC*TTTCGG*CATC*CCGCAT*-GAAA-3′ which is the known DNA binding sequence of ComA for the *srfA* operon.[14] The underlined portions correspond to the response element boxes with dyad symmetry. A total of 1000 structures were generated in the static iteration, and the 200 lowest-energy structures further refined in semi-flexible iteration with the 20 lowest-energy structures undergoing water refinement. Each docking attempt was performed 10 times, and the solution with the lowest HADDOCK score was retained. The HADDOCK score is a weighted sum of intermolecular electrostatic (E_{Elec}), van der Waals (E_{vdW}), desolvation (E_{desolv}) and AIR (E_{AIR}) energies and a buried surface area (BSA) term. The C^{α}-rmsd values of the complexes were calculated by using ProFit‖.[55] Values for base pair and base pair step parameters, as well as torsion angles for the sugar-phosphate backbone and sugar pucker were obtained by using the program 3DNA. A cluster analysis was performed on the final docking solutions using a minimum cluster size of 4. The cut-off for clustering was determined manually for each docking run. The C^{α}-rmsd matrix was calculated over the backbone atoms of the interface residues of the DNA after fitting onto the interface residues of the protein.

§ w3DNA: http://w3dna.rutgers.edu/
‖ www.bioing,org.uk/software/profit

Data Bank accession codes

The atomic coordinates for ComAC have been deposited in the Protein Data bank (accession number 2krf); the chemical shift list is available at the Biological Magnetic Resonance Data Bank (accession number rcsb101480).

Acknowledgements

This research was supported by NIH grants RO1 GM65156 (to J.C.), RO1 GM55769 (to J.C.), GM55594 (to M.P.), and The V Foundation for Cancer Research (to J.C.). The authors are grateful to Dr Amanda Stewart (NCSU) for helpful comments. We thank Dr Patrick McLaughlin and Erin Regel for their help with the ComAN and ComAN-D55N constructs.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.03.003

References

1. Giammarinaro, P., Sicard, M. & Gasc, A. M. (1999). Genetic and physiological studies of the CiaH-CiaR two-component signal-transducing system involved in cefotaxime resistance and competence of Streptococcus pneumoniae. *Microbiology*, **145**, 1859–1869.
2. Gromkova, R. C., Mottalini, T. C. & Dove, M. G. (1998). Genetic transformation in Haemophilus parainfluenzae clinical isolates. *Curr. Microbiol.* **37**, 123–126.
3. Orus, P. & Vinas, M. (2000). Transfer of penicillin resistance between Neisseriae in microcosm. *Microb. Drug Resist.* **6**, 99–104.
4. Sjostrom, J. E., Lofdahl, S. & Philipson, L. (1975). Transformation reveals a chromosomal locus of the gene(s) for methicillin resistance in Staphylococcus aureus. *J. Bacteriol.* **123**, 905–915.
5. Kroll, J. S., Moxon, E. R. & Loynds, B. M. (1994). Natural genetic transfer of a putative virulence-enhancing mutation to Haemophilus influenzae type a. *J. Infect. Dis.* **169**, 676–679.
6. Dubnau, D. (1991). Genetic competence in Bacillus subtilis. *Microbiol. Rev.* **55**, 395–424.
7. Nakano, M. M. & Zuber, P. (1991). The primary role of comA in establishment of the competent state in Bacillus subtilis is to activate expression of srfA. *J. Bacteriol.* **173**, 7269–7274.
8. Hahn, J., Kong, L. & Dubnau, D. (1994). The regulation of competence transcription factor synthesis constitutes a critical control point in the regulation of competence in Bacillus subtilis. *J. Bacteriol.* **176**, 5753–5761.
9. Dubnau, D., Hahn, J., Roggiani, M., Piazza, F. & Weinrauch, Y. (1994). Two-component regulators and genetic competence in Bacillus subtilis. *Res. Microbiol.* **145**, 403–411.

10. Nakano, M. M., Xia, L. A. & Zuber, P. (1991). Transcription initiation region of the srfA operon, which is controlled by the comP-comA signal transduction system in Bacillus subtilis. *J. Bacteriol.* **173**, 5487–5493.
11. Magnuson, R., Solomon, J. & Grossman, A. D. (1994). Biochemical and genetic characterization of a competence pheromone from B. subtilis. *Cell*, **77**, 207–216.
12. Okada, M., Sato, I., Cho, S. J., Iwata, H., Nishio, T., Dubnau, D. & Sakagami, Y. (2005). Structure of the Bacillus subtilis quorum-sensing peptide pheromone ComX. *Nat. Chem. Biol.* **1**, *23–24.*
13. Nakano, M. M. & Zuber, P. (1993). Mutational analysis of the regulatory region of the srfA operon in Bacillus subtilis. *J. Bacteriol.* **175**, 3188–3191.
14. Roggiani, M. & Dubnau, D. (1993). ComA, a phosphorylated response regulator protein of Bacillus subtilis, binds to the promoter region of srfA. *J. Bacteriol.* **175**, 3182–3187.
15. Solomon, J. M., Magnuson, R., Srivastava, A. & Grossman, A. D. (1995). Convergent sensing pathways mediate response to two extracellular competence factors in Bacillus subtilis. *Genes Dev.* **9**, 547–558.
16. Nakano, M. M., Magnuson, R., Myers, A., Curry, J., Grossman, A. D. & Zuber, P. (1991). srfA is an operon required for surfactin production, competence development, and efficient sporulation in Bacillus subtilis. *J. Bacteriol.* **173**, 1770–1778.
17. Grossman, A. D. (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis. *Annu. Rev. Genet.* **29**, 477–508.
18. Turgay, K., Hahn, J., Burghoorn, J. & Dubnau, D. (1998). Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. *EMBO J.* **17**, 6730–6738.
19. Guillen, N., Weinrauch, Y. & Dubnau, D. A. (1989). Cloning and characterization of the regulatory Bacillus subtilis competence genes comA and comB. *J. Bacteriol.* **171**, 5354–5361.
20. Weinrauch, Y., Guillen, N. & Dubnau, D. A. (1989). Sequence and transcription mapping of Bacillus subtilis competence genes comB and comA, one of which is related to a family of bacterial regulatory determinants. *J. Bacteriol.* **171**, 5362–5375.
21. Hahn, J. & Dubnau, D. (1991). Growth stage signal transduction and the requirements for srfA induction in development of competence. *J. Bacteriol.* **173**, 7275–7282.
22. Msadek, T., Kunst, F., Klier, A. & Rapoport, G. (1991). DegS-DegU and ComP-ComA modulator-effector pairs control expression of the Bacillus subtilis pleiotropic regulatory gene degQ. *J. Bacteriol.* **173**, 2366–2377.
23. Mueller, J. P., Bukusoglu, G. & Sonenshein, A. L. (1992). Transcriptional regulation of Bacillus subtilis glucose starvation-inducible genes: control of gsiA by the ComP-ComA signal transduction system. *J. Bacteriol.* **174**, 4361–4373.
24. Solomon, J. M., Lazazzera, B. A. & Grossman, A. D. (1996). Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in Bacillus subtilis. *Genes Dev.* **10**, 2014–2024.
25. Jiang, M., Grau, R. & Perego, M. (2000). Differential processing of propeptide inhibitors of Rap phosphatases in Bacillus subtilis. *J. Bacteriol.* **182**, 303–310.
26. Core, L. & Perego, M. (2003). TPR-mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in Bacillus subtilis. *Mol. Microbiol.* **49**, 1509–1522.
27. Bongiorni, C., Ishikawa, S., Stephenson, S., Ogasawara, N. & Perego, M. (2005). Synergistic regulation of competence development in Bacillus subtilis by two Rap-Phr systems. *J. Bacteriol.* **187**, 4353–4361.
28. Auchtung, J. M., Lee, C. A. & Grossman, A. D. (2006). Modulation of the ComA-dependent quorum response in Bacillus subtilis by multiple Rap proteins and Phr peptides. *J. Bacteriol.* **188**, 5273–5285.
29. Comella, N. & Grossman, A. D. (2005). Conservation of genes and processes controlled by the quorum response in bacteria: characterization of genes controlled by the quorum-sensing transcription factor ComA in Bacillus subtilis. *Mol. Microbiol.* **57**, 1159–1174.
30. Griffith, K. L. & Grossman, A. D. (2008). A degenerate tripartite DNA-binding site required for activation of ComA-dependent quorum response gene expression in Bacillus subtilis. *J. Mol. Biol.* **381**, 261–275.
31. Cavanagh, J., Fairbrother, W. J., Palmer, A., Rance, M. & Skelton, N. J. (2007). *Protein NMR Spectroscopy: Principles and Practice*, 2nd edit. Elsevier Academic Press, Burlington, MA.
32. Stuart, A., Borzilleri, K. A., Withka, J. M. & Palmer, A. G., III (1999). Compensating for variations in ^{1}H-^{13}C scalar coupling constants in isotope-filtered NMR experiments. *J. Am. Chem. Soc.* **121**, 5346–5347.
33. Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS, computer program.* Department of Biochemistry and Molecular Biology, University College London.
34. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
35. Maris, A. E., Sawaya, M. R., Kaczor-Grzeskowiak, M., Jarvis, M. R., Bearson, S. M., Kopka, M. L. *et al.* (2002). Dimerization allows DNA target site recognition by the NarL response regulator. *Nat. Struct. Biol.* **9**, 771–778.
36. Wisedchaisri, G., Wu, M., Rice, A. E., Roberts, D. M., Sherman, D. R. & Hol, W. G. (2005). Structures of Mycobacterium tuberculosis DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. *J. Mol. Biol.* **354**, 630–641.
37. Ducros, V. M., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P. *et al.* (2001). Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in Bacillus subtilis. *J. Mol. Biol.* **306**, 759–771.
38. Pristovsek, P., Sengupta, K., Lohr, F., Schafer, B., von Trebra, M. W., Ruterjans, H. & Bernhard, F. (2003). Structural analysis of the DNA-binding domain of the Erwinia amylovora RcsB protein and its interaction with the RcsAB box. *J. Bacteriol.* **278**, 17752–17759.
39. Kurashima-Ito, K., Kasai, Y., Hosono, K., Tamura, K., Oue, S., Isogai, M. *et al.* (2005). Solution structure of the C-terminal transcriptional activator domain of FixJ from Sinorhizobium meliloti and its recognition of the fixK promoter. *Biochemistry*, **44**, 14835–14844.
40. Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R. *et al.* (2002). The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* **21**, 4393–4401.
41. Theobald, D. L. & Wuttke, D. S. (2008). Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput. Biol.* **4**, e43.
42. Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1992). Side chain and backbone assignments in

isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett.* **314**, 413–418.
43. Grzesiek, S. & Bax, A. (1993). Measurement of amide proton exchange rates and NOEs with water in 13C/15N-enriched calcineurin B. *J. Biomol. NMR*, **3**, 627–638.
44. Ikura, M., Kay, L. E. & Bax, A. (1990). A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, **29**, 4659–4667.
45. Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1993). A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomol. NMR*, **3**, 225–231.
46. Montelione, G. T., Emerson, S. D. & Lyons, B. A. (1992). A general approach for determining scalar coupling constants in polypeptides and proteins. *Biopolymers*, **32**, 327–334.
47. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
48. Brunger, A. T., Adams, P. D. & Rice, L. M. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr. Opin. Struct. Biol.* **8**, 606–611.
49. Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S. & King, G. F. (1996). High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. *J. Bacteriol.* **271**, 13663–13667.
50. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352.
51. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
52. Zheng, G., Lu, X. J. & Olson, W. K. (2009). Web 3DNA–a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37**, W240–W246.
53. Dominguez, C., Boelens, R. & Bonvin, A. M. (2003). HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737.