

Secondary structure and DNA binding by the C-terminal domain of the transcriptional activator NifA from *Klebsiella pneumoniae*

Pampa Ray^{1,2}, K. John Smith¹, Rosemary A. Parslow¹, Ray Dixon² and Eva I. Hyde^{1,*}

¹School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK and ²Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, UK

Received June 13, 2002; Revised and Accepted July 25, 2002

ABSTRACT

The NifA protein of *Klebsiella pneumoniae* is required for transcriptional activation of all nitrogen fixation (*nif*) operons except the regulatory *nifLA* genes. At these operons, NifA binds to an upstream activator sequence (UAS), with the consensus TGT-N₁₀-ACA, via a C-terminal DNA-binding domain (CTD). Binding of the activator to this upstream enhancer-like sequence allows NifA to interact with RNA polymerase containing the alternative sigma factor, σ^{54} . The isolated NifA CTD is monomeric and binds specifically to DNA *in vitro* as shown by DNase I footprinting. Heteronuclear 3D NMR experiments have been used to assign the signals from the protein backbone. Three α -helices have been identified, based on secondary chemical shifts and medium range H α_i -NH $_{i+1}$, and NH $_i$ -NH $_{i+1}$ NOEs. On addition of DNA containing a half-site UAS, several changes are observed in the NMR spectra, allowing the identification of residues that are most likely to interact with DNA. These occur in the final two helices of the protein, directly confirming that DNA binding is mediated by a helix–turn–helix motif.

INTRODUCTION

The NifA protein, found in diazotrophic bacteria, is responsible for activating the transcription of genes required for nitrogen fixation. In *Klebsiella pneumoniae*, NifA activates transcription from all the operons involved in nitrogen fixation except for the *nifLA* operon, which is activated by the phosphorylated form of the transcriptional activator NtrC (1). Both NtrC and NifA work in conjunction with RNA polymerase containing the alternative σ^{54} (σ^N) factor (2). This form of RNA polymerase holoenzyme binds promoter DNA but requires an activator to initiate transcription. These σ^{54} -dependent activators use hydrolysis of nucleoside triphosphates to catalyse the isomerisation of closed complexes between RNA polymerase and DNA to form open promoter complexes (3,4).

In the γ -subdivision of proteobacteria, for example in *K.pneumoniae*, the NifL protein regulates the activity of NifA in response to molecular oxygen and fixed nitrogen (5). In nitrogen-limiting conditions and in the absence of molecular oxygen, NifA binds to one or more upstream activator sequences (UAS) in the DNA (6). The UASs are positioned typically 100 bp upstream of the binding site of σ^{54} RNA polymerase (7); however, they can function up to 2 kb upstream of the transcription start site and also downstream of the start site, thus exhibiting behaviour similar to eukaryotic enhancer elements. Promoters often contain multiple UASs, which have weak homology to the consensus sequence, namely TGT-N₁₀-ACA. Mutational studies of *nif* UASs (8) show that the optimum binding sequence of *K.pneumoniae* NifA is 5'-TGT-G/A-G/A-G-3', as expected from the consensus for the wild-type UASs. In most cases, the intervening DNA between the UAS and the promoter bends with the aid of the integration host factor, facilitating contacts between the activator and RNA polymerase (9). The UAS enhancer-like sequence thus functions to position NifA at a location appropriate for interaction with the σ^{54} RNA polymerase.

The NifA protein has a modular domain structure typical of other σ^{54} -dependent transcriptional activators (5). The N-terminal domain may regulate the inhibition of NifA by NifL and contains a GAF domain that is common to several cyclic GMP receptors (10). It is attached to the remainder of the protein by a protease sensitive, glutamine rich, linker (11). This N-terminal domain is not conserved amongst σ^{54} -dependent transcriptional activators. The central domain of NifA shows extensive homology to equivalent domains of other σ^{54} activators such as DctD, XylR and NtrC (12). These proteins are members of the AAA⁺ superfamily of ATPases, which use nucleotide hydrolysis to restructure their substrates (13). In σ^{54} -dependent activators this conserved domain is sufficient for the nucleotide-dependent interactions which drive formation of the open promoter complex (14–16). The C-terminal domain (CTD) of NifA is required for enhancer-dependent transcriptional activation. It contains a putative helix–turn–helix (HTH) motif, which is presumed to recognise the UASs (17). For the σ^{54} -dependent activators there is also a conserved Ala–Leu–X₉–Ala–Ala–X₂–Leu–Gly sequence in this domain, the terminal glycine being part of the HTH motif

*To whom correspondence should be addressed. Tel: +44 121 414 5393; Fax: +44 121 414 5982; Email: e.i.hyde@bham.ac.uk
Present address:

Pampa Ray, Department of Biological Sciences, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK

(12). The structure of a mutant CTD of another σ^{54} -dependent activator, NtrC from *Salmonella typhimurium*, has been determined using NMR spectroscopy (18). It is structurally homologous to Fis, the factor in *Escherichia coli* responsible for site-specific inversion when bound to recombinational enhancers (19). Both Fis and NtrC are dimeric and primarily α -helical in solution. Each monomer contains four helices, including a HTH motif. The dimerisation interface consists of parts of the first and second helices, forming an anti-parallel four-helix bundle. The mutant NtrC domain does not bind DNA (20), therefore it is not possible to examine the protein-DNA interactions. Although sequence homology studies have identified the DNA-binding regions of the σ^{54} -dependent activators, no direct information is currently available about the interactions of the individual amino acids with DNA. The CTD of NifA is monomeric in solution (21) and binds specifically to DNA containing UASs *in vivo* and *in vitro* (22). In a previous study, we probed the conformation of the purified C-terminal, DNA-binding domain of *K.pneumoniae* NifA by a variety of spectroscopic techniques and compared the results with the secondary structure predicted from its primary sequence (21). These studies suggest that the domain is largely α -helical in solution, with three helices, including the HTH motif. In this study, we examine the secondary structure of the protein in more detail by heteronuclear NMR spectroscopy, and examine changes in the spectrum of the protein on binding to DNA containing a half UAS. We present the first physical evidence for the interaction of protein residues within the HTH motif with its target DNA-binding site. We demonstrate that, for this particular σ^{54} -dependent activator, the majority of amino acid residues which interact with the cognate DNA sequence lie on one surface of the recognition helix.

MATERIALS AND METHODS

Protein preparation

The CTD of NifA from *K.pneumoniae* was expressed in *E.coli* BL21 (λ DE3) pLysS, from the pT7-7 vector derivative, pJES470 (22). This encodes a fusion peptide of the first nine amino acids of the pT7-7 vector (MARIRARGS) to the last 67 amino acids of NifA. The residues in this paper are labelled such that D1 corresponds to D457 of the full-length protein, with the residues from the linker as -9 to -1. ^{15}N -labelled and doubly $^{15}\text{N}/^{13}\text{C}$ -labelled protein was prepared by growing the cells in M9 medium (23) with ^{15}N -ammonium chloride and unenriched glucose or $^{13}\text{C}_6$ -glucose as the sole nitrogen and carbon sources, respectively. The protein was prepared as in our previous study, using a phosphocellulose P10 column and a Trisacryl-blue column (21). To remove traces of nucleases, the protein was further purified using a Vydac C_{18} reverse-phase column. The protein was loaded onto the column in buffer containing 25 mM ammonium acetate, 50 mM LiBr, pH 6.4, and eluted with a gradient of buffer in 0–50% acetonitrile. The protein eluted at 40% acetonitrile. Samples for NMR experiments were lyophilised, taken up in H_2O or D_2O as required, and dialysed into buffer containing 10 mM sodium phosphate, 0.5 M LiBr, 0.1 mM EDTA, 1 mM 4-(2-aminoethyl)-benzene sulphonyl fluoride (AEBSF), pH 6.4.

Protein concentrations were determined using Bradford assays calibrated with bovine serum albumin (BSA) as standard (24). All concentrations are reported for the domain as a monomer, as shown in our previous study (21).

DNase I footprinting

The 341 bp *EcoRI*–*BamHI* fragment of pMB1 (25,26) containing three UASs from the *nifH*–*nifJ* intergenic region, was 5'-end labelled using $[\gamma\text{-}^{32}\text{P}]\text{dATP}$ and T4 polynucleotide kinase. Reactions for the binding assays contained 5 nM DNA, 8.1 ng/ μl poly(dI–dC) (Boehringer Mannheim), 25 mM Tris acetate pH 8, 50 mM potassium acetate, 13.5 mM ammonium acetate, 0.5 μM acetylated BSA (NEB) and 2.5–7.5 μM NifA C-terminal DNA-binding domain in a final volume of 15 μl . The reaction mixes were incubated at 30°C for 20 min. DNase I (1.1×10^{-4} U; Boehringer Mannheim) was added in 3 μl of buffer containing 10 mM Tris–HCl, 10 mM CaCl_2 , 40 mM MgCl_2 and 10% glycerol. Digestion was allowed to proceed for 30 s at 30°C and was terminated by adding 180 μl of 0.3 M sodium acetate, 10 mM EDTA. After ethanol precipitation and resuspension, the DNA fragments were separated on a 6% denaturing polyacrylamide gel, and analysed on a PhosphorImager (Molecular Dynamics).

Preparation of oligonucleotides for NMR experiments

5'-AATGTGGGAAAC-3' and 5'-GTTTCCCACATT-3' were purchased from MWG Biotech. The freeze-dried oligonucleotide samples were each dissolved in H_2O and passed through a Sephadex G-10 spin column to separate them from low molecular weight contaminants. The oligonucleotides were lyophilised and dissolved in buffer containing 10 mM sodium phosphate, 100 mM LiBr, 1 mM EDTA, pH 6.4, to a final concentration of ~20 mM. Equivalent amounts of each oligonucleotide (based on their A_{260}) were mixed, heated to 70°C for 15 min and allowed to anneal by cooling slowly.

NMR spectroscopy of free protein

NMR samples of free protein contained 0.6 ml of ~0.5 mM protein in 50 mM sodium phosphate buffer pH 6.4, 500 mM LiBr, 0.1 mM EDTA. Triple resonance and three-dimensional NMR experiments were performed on a four-channel Varian Unityplus 600 spectrometer, operating at 599.98 MHz at 27°C. Main-chain ^1HN , ^{15}N , $^{13}\text{C}\alpha$ and side-chain $^{13}\text{C}\beta$ resonances were assigned using HNCACB and CBCA(CO)NNH (27) and $\text{C}\beta$ -decoupled-HNCA (28) experiments. Main-chain $^1\text{H}\alpha$ and ^{13}CO assignments were made using CBCACO(CA)HA (29), HBHA(CBCA)(CO)NNH (30), HNCO (27) and HNHA (31) spectra. Selective ^{13}C decoupling was achieved using WURST-2 adiabatic decoupling schemes (28). ^{15}N dimensions were generally collected with Kay-type sensitivity enhanced gradient coherence selection for quadrature detection and water suppression (27), and other dimensions were collected by using the standard time-proportional phase incrementation method. A 1–1.5 s recycle delay after acquisition and before the first pulse of the program was used. Typical sweep widths were 9 (direct) and 8 kHz (indirect) for ^1H , 1.5 kHz for ^{15}N , 9.4 kHz for $^{13}\text{C}\alpha/^{13}\text{C}\beta$, 4.2 kHz for $^{13}\text{C}\alpha$, and 2.2 kHz for ^{13}CO . 1024 points were collected in the direct ^1H dimension, with 32* complex points in ^{15}N , between 48*–64* complex points in $^{13}\text{C}\alpha/^{13}\text{C}\beta$, 128* complex points in $^{13}\text{C}\alpha$, and 48* complex points in ^{13}CO . The minimum

number of scans to complete the phase cycle was collected in each case. A 3D NOESY-¹⁵N-sensitivity enhanced HSQC (32) was also collected in which four transients were acquired for each of 128* and 32* complex points in indirect ¹H and ¹⁵N dimensions, respectively.

¹H chemical shifts were referenced to the H₂O resonance, which was calibrated independently at 4.789 p.p.m. (at 27°C) relative to sodium 2,2-dimethylsilapentane-5-sulphonic acid. ¹⁵N and ¹³C chemical shifts were referenced indirectly (33). The NMR data were processed on a Silicon Graphics workstation using VNMR (Varian UK). Generally, sinebell squared apodisation shifted by 90° was used in each dimension. The intensities of cross-peaks in the 3D NOESY-¹⁵N filtered HSQC were measured in NMRCompass (Molecular Simulations). The chemical shifts of the Cα, Hα, CO and Cβ resonances were used to determine the secondary structure of the protein using the program CSI (34).

NMR spectroscopy of protein–DNA complexes

Titration experiments were performed by adding microlitre aliquots of DNA into a 40 μM protein solution in buffer containing 10 mM sodium phosphate, 100 mM LiBr, 5 mM EDTA, pH 6.4, with 10% D₂O. HSQC spectra (35) were collected on a Bruker AMX-500 spectrometer, operating at 500.13 MHz. The spectra were collected with 1024 data points with a sweep width of 14.7 p.p.m. in the direct dimension, and 256 experiments, with a sweep width of 32 p.p.m., in the indirect dimension and 128 scans per experiment. The water peak was reduced by using a modified WATERGATE pulse sequence. Phase cycling was performed in phase sensitive mode using the time-proportional phase incrementation method for quadrature detection in the indirect dimension. The spectra were processed using UXNMR software.

To estimate the apparent dissociation constant of the DNA–protein complex, the chemical shifts of the protein resonances were measured at different concentrations of DNA. For several peaks, these data were fitted simultaneously to the equations below (reviewed in 36), using non-linear regression in Sigmaplot (SPSS Inc):

$$v_{\text{obs}} = v_E + ([EL] / E_T)(v_{\text{EL}} - v_E)$$

$$[EL] = 1/2[E_T + L_T + K_d - \{(E_T + L_T + K_d)^2 - 4E_T L_T\}^{1/2}]$$

where v_{obs} is the observed shift of the resonance; v_E is its shift in the free protein; v_{EL} is its shift in the protein–DNA complex; $[EL]$ is the concentration of the protein–DNA complex; E_T is the total protein concentration; L_T is the total DNA concentration; K_d is the dissociation constant of the protein–DNA complex. Individual values for v_E , v_{EL} were obtained for each peak, with a global fit for K_d .

For peaks in intermediate exchange:

$$2\pi(v_E - v_{\text{EL}}) \approx (1 / \tau_E + 1 / \tau_{\text{EL}}) = k_{\text{off}} / (1 + P_{\text{EL}} / P_E)$$

where τ_E is the lifetime of the resonance in the free protein; τ_{EL} is the lifetime in the DNA complex = $1 / k_{\text{off}}$; P_{EL} and P_E are the fractions of protein in the bound and free state, respectively.

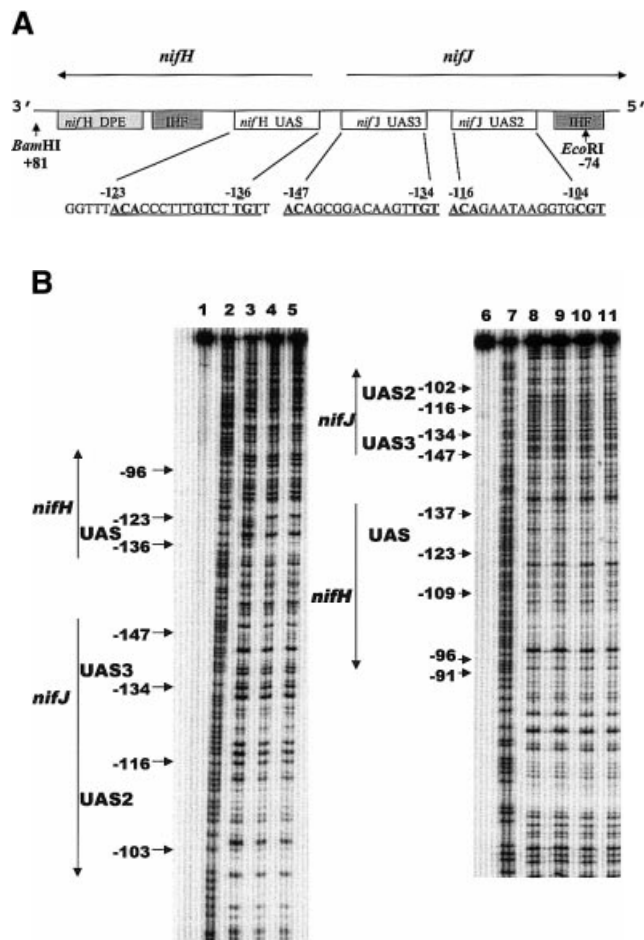


Figure 1. (A) Representation of the DNA fragment used in DNase I footprinting experiments, showing the intergenic region between the divergent *nifH* and *nifJ* promoters. The promoter elements and protein-binding sites are boxed and direction of transcription shown with arrows. DPE, downstream promoter element; IHF, binding site for integration host factor protein; UAS, upstream activator sequence. The position and sequences of the UASs are shown under the appropriate box. The numbering used is relative to the closest transcription start site. There are 30 bp between the *nifH* UAS and *nifJ* UAS3. (B) DNase I footprint of NifA CTD from *K.pneumoniae* on the *EcoRI*–*BamHI* fragment of pMB1. Lanes 1–5, fragment 5'-end labelled at the *EcoRI* end. Lane 1, DNA template (no DNase I); lane 2, G+A sequencing ladder; lanes 3–6, DNA template plus DNase I; lane 3, no protein; lane 4, 2.3 μM NifA CTD; lane 5, 4 μM NifA CTD; lanes 6–11, fragment 5'-end labelled at the *BamHI* end; lane 6, DNA template (no DNase I); lane 7, G+A sequencing ladder; lanes 8–11, DNase I plus DNA; lane 8, no protein; lane 9, 1 μM NifA CTD; lane 10, 2.3 μM NifA CTD; lane 11, 4 μM NifA CTD.

RESULTS

Assignment of the NMR spectrum of NifA CTD and secondary structure elements within the domain

The C-terminal domain of NifA was overexpressed in *E.coli* as a fusion protein from a vector encoding nine amino acids from the vector pT7-7 fused to 67 amino acids at the C-terminus of NifA (22). To test that the purification method yielded active protein, DNase I protection assays were performed using DNA containing the *K.pneumoniae nifH* promoter and the *nifH*–*nifJ* intergenic region (Fig. 1).

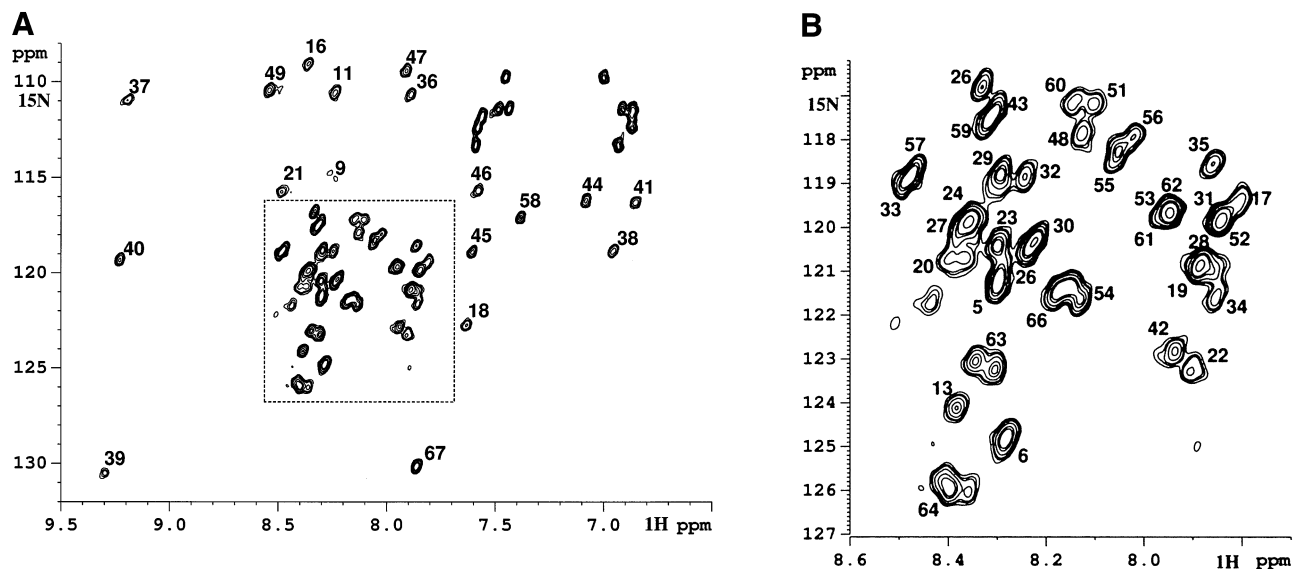


Figure 2. ^{15}N - ^1H heteronuclear single quantum coherence spectrum of NifA CTD from *K.pneumoniae* in 10 mM sodium phosphate buffer, pH 6.6, containing 500 mM LiBr, and 0.1 mM EDTA at 27°C. The peaks from the peptide amide groups are labelled with their assignments, numbered with D458 as residue 1. Unlabelled peaks arise from Asn and Gln side chains and the eight N-terminal amino acids of the protein that come from the expression vector. (A) Full amide spectrum. (B) An expansion of the dotted region of the spectrum shown in (A).

Protection from DNase I cleavage of the region -138 to -117 bp upstream of the *nifH* transcription start site is observed at 2.3 μM NifA. This shows that the protein is active and binds specifically to the *nifH* UAS, as in previous studies (23). Little protection of the *nifJ* UAS sequences is seen with the CTD of NifA at these concentrations. However, at protein concentrations >8.3 μM , some protection of bands at the *nifJ* UAS2 and UAS3 is observed (data not shown). This indicates that the CTD of NifA from *K.pneumoniae* binds more weakly to the *nifJ* UASs than to the *nifH* UAS. The same concentrations of protein were needed to obtain footprints after ion-exchange chromatography, reverse-phase HPLC and after dialysis in 0.5 M LiBr for the preparation of NMR samples, showing that the purification method did not affect the activity of the protein (data not shown).

In order to determine the secondary structure of the CTD by NMR spectroscopy, the protein was labelled with ^{15}N and ^{13}C . The ^1H - ^{15}N HSQC spectrum of the protein in 0.5 M LiBr is shown in Figure 2. This spectrum is very similar to that observed previously in 0.5 M NaCl (21). Normally, for a protein containing only 75 amino acid residues, the resonances can be completely assigned using ^{15}N -labelled protein alone. Due to the considerable overlap of both the ^1H and ^{15}N resonances, heteronuclear 3D triple resonance NMR experiments with $^{13}\text{C}/^{15}\text{N}$ double-labelled protein were used to determine the assignments of the backbone resonances. Main-chain ^1HN , ^{15}N , $^{13}\text{C}\alpha$ and side-chain $^{13}\text{C}\beta$ resonances were assigned sequentially using a combination of HNCACB and CBCA(CO)NNH (27) spectra. A high-resolution $\text{C}\beta$ -decoupled-HNCA (28) experiment, showing connectivities only to $^{13}\text{C}\alpha$, was used to resolve sequential $\text{C}\alpha$ resonances that were close in chemical shift. Strips from the 3D spectra illustrating the sequential assignments for residues 51-60 are shown in the Supplementary Material (Fig. S1). Main-chain

$^1\text{H}\alpha$, $^1\text{H}\beta$ and $^{13}\text{C}\text{O}$ assignments were made using HBHA(CBCA)(CO)NNH (30) experiments and HNCO (27) experiments. The assignments were confirmed using CBCACO(CA)HA (29) and HNHA (31) spectra. For Thr49 and Met64, which are adjacent to proline residues (that do not contain amides protons), the $\text{C}\beta$ resonances could not be assigned by this combination of methods. Also, no assignments were made for Pro3, which is adjacent to Pro4. (The remaining chemical shift assignments are listed in the Supplementary Material, Table S1.)

The $\text{C}\alpha$ and $\text{H}\alpha$ chemical shifts are strongly influenced by the orientation of the groups relative to the carbonyl of the amide bond. The chemical shift index uses the difference in observed shift from that expected for a random coil to predict secondary structure (34). For NifA the consensus secondary structure from the $\text{C}\alpha$, $\text{H}\alpha$, $\text{C}\beta$ and CO shifts indicates the presence of α -helices between residues 23-34, 39-45 and 51-59 (Fig. 3). The ^{15}N -HSQC edited NOESY (32), which shows which protons are close in space, was also analysed for indications of secondary structure. Cross-peaks between adjacent NH residues were observed for residues 27-33, 36-48 and 51-61, while cross-peaks between $\text{H}\alpha$ resonances and NH groups three residues apart ($\alpha_i\text{-NH}_{i+3}$) were observed for residues 27-33, 40-44 and 51-56. These NOEs also suggest the presence of α -helices within these regions of the domain. The corresponding NOEs at residues 23-26, at the beginning of the first helix, cannot be assigned unambiguously due to overlap between NH resonances and between $\text{H}\alpha$ resonances. The results from the NOEs and chemical shift indices are in agreement with predictions from the PredictProtein program, which indicated helices between residues 25-34, 39-46 and 51-59 (22). There is no evidence for any β -strands, or for any regular secondary structure in the N-terminal part of the domain before residue 23.

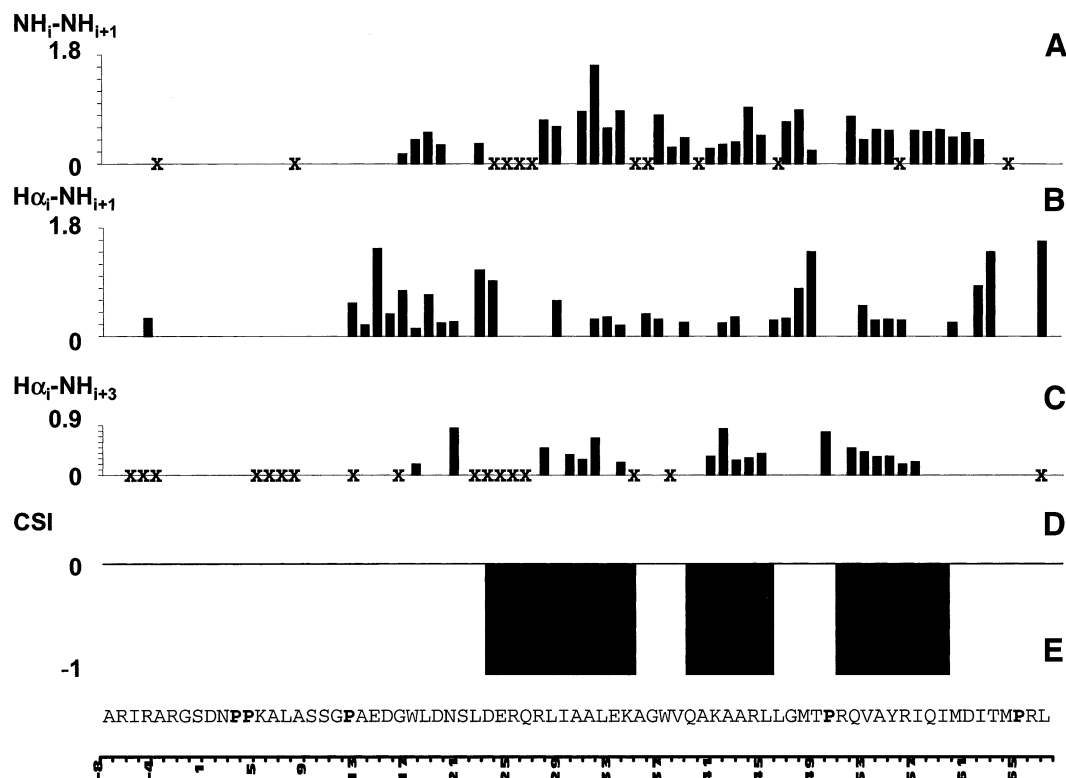


Figure 3. Histogram showing the intensities of NOEs in the ^{15}N -HSQC edited NOESY and consensus chemical shift index of the C-terminal domain of NifA from *K.pneumoniae*. (A–C) Intensities of NOEs: (A) $\text{NH}_i\text{-NH}_{i+1}$ NOEs, (B) $\text{H}\alpha_i\text{-NH}_{i+1}$ NOEs, (C) $\text{H}\alpha_i\text{-NH}_{i+3}$ NOEs. The maximum NOE intensity observed was 1.6 units. X shows positions at which the overlap of resonances prevented unambiguous assignment of NOEs. (D) The consensus chemical shift index based on the $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and $^{13}\text{C}\text{O}$ chemical shifts (34). (E) Sequence of the protein used. Residue 1 corresponds to D458 of the full-length protein, residues –8 to –1 come from the expression vector.

Titration of NifA CTD with a DNA oligonucleotide containing a half-site UAS

Full-length NifA, like other prokaryotic HTH proteins, is oligomeric and binds to a palindromic UAS with the consensus sequence TGT-N₁₀-ACA. To study the interactions between the NifA CTD and DNA, the protein was titrated with a synthetic oligodeoxynucleotide and the effects on the HSQC spectra examined. To simplify the spectral analysis, we used an oligodeoxynucleotide containing the UAS half-site 5'-TGTGGG-3', to ensure binding of protein monomers rather than dimers. This sequence is contained in the natural *nifH* UAS that binds more tightly to the NifA CTD than the other two UASs in the *nifH-nifJ* intergenic region (Fig. 1). It also corresponds to the optimum binding sequence of the full-length NifA protein from *K.pneumoniae* (8). The oligodeoxynucleotide used in this study consisted of a dodecamer containing this sequence with additional flanking bases from the *nifH* UAS to enhance the binding of the protein and the stability of the DNA duplex.

Figure 4 shows the effect of addition of DNA on the HSQC spectrum of the protein. Most cross-peak positions remain unchanged or move very slightly on addition of DNA; for example, Val38, Gly47 and Leu46. Some peaks, such as those from Gln39, Lys41, Arg56, Tyr55 and Met60, shift considerably but remain sharp and can be followed throughout the titration. A few peaks such as those from Ala40 and Thr49 broaden out and disappear completely. Interestingly, the peak

due to residue Gln39, at 130.5 and 9.3 p.p.m., broadens on the addition of 5 equivalents of DNA but reappears at 15 equivalents of DNA, suggesting that this is now almost completely in the bound form. The chemical shift differences for the amide hydrogen and nitrogen between free protein and protein in the presence of 15 equivalents of DNA are shown in Figure 5.

The effects observed are due to different changes in chemical shifts between the resonances in the free protein and in the protein–DNA complex. The free and bound states of the protein are in dynamic equilibrium. If the chemical shift difference (in Hz) between the two states of a resonance is smaller than the rate of exchange, one sharp peak is observed. The chemical shift value of this peak is a weighted average of the chemical shifts of the resonance in the free and bound state, and so the peak moves towards the bound position as the DNA concentration increases. This is seen with most resonances. However, if the chemical shift difference between the free and bound states (in Hz) is similar in size to the exchange rate, then the resonances are in intermediate exchange. The peak broadens, and may not be visible above the noise level.

The chemical exchange behaviour observed in the NMR spectra shows that the binding is weak, consistent with the DNase I footprinting on the longer DNA fragment. Figure 6 shows the chemical shifts of several of the protein resonances plotted against the concentration of added DNA. The curves show the fit of the data to a model in which there is 1:1

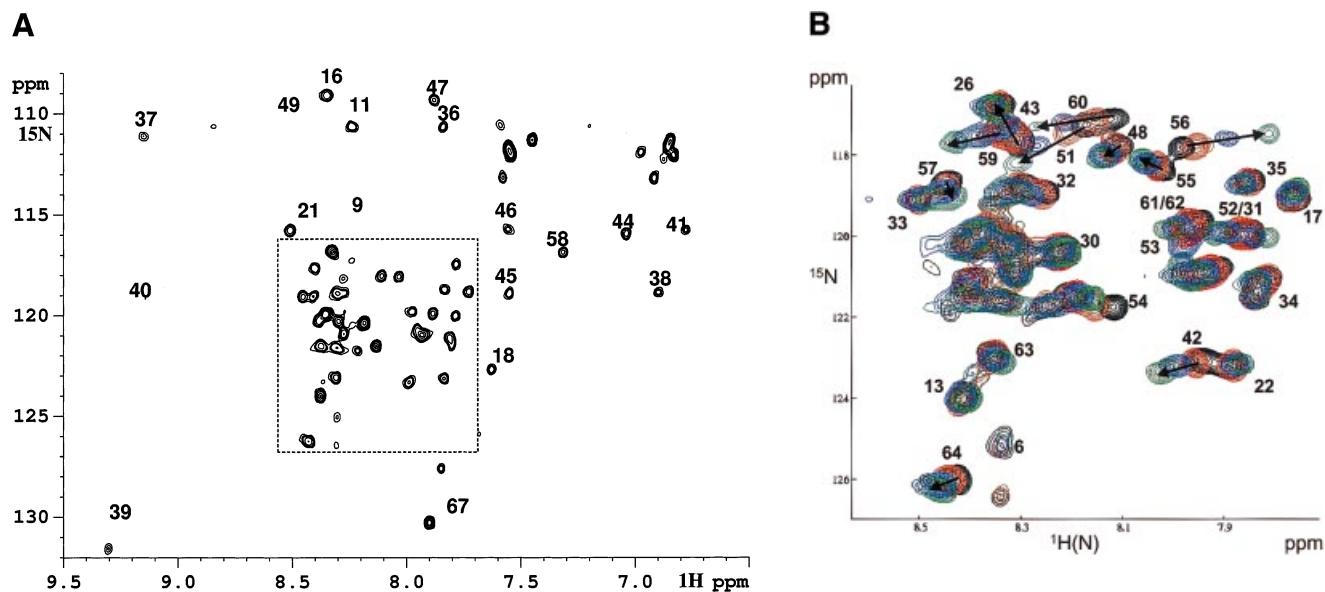


Figure 4. ^{15}N - ^1H HSQC spectra of the CTD of NifA from *K.pneumoniae* at 27°C in the presence of oligodeoxynucleotide containing a half-site UAS. The concentration of the protein was 40 μM in buffer containing 10 mM sodium phosphate, 100 mM LiBr and 5 mM EDTA pH 6.4. (A) Spectrum in the presence of 15 equivalents of oligodeoxynucleotide. Peak numbers are shown at the positions corresponding to the spectrum in the absence of DNA (as in Fig. 2A). (B) Overlay of the central part of five HSQC spectra at increasing concentrations of oligodeoxynucleotide. The expansion is of the dotted region in (A). Contour colours: black, no DNA; red, 0.75 equivalents DNA; brown, 1.5 equivalents DNA; blue, 5 equivalents DNA; and green, 15 equivalents DNA. Selected cross-peaks are labelled with their assignment and arrows indicating the changes in shift.

protein:DNA binding with the free and bound resonances in fast exchange, with a K_d of 200 μM . The data for both amide ^1H and ^{15}N peaks show good fit to this model, for various chemical shift differences. However, the K_d is poorly defined and if the peaks are not in full fast exchange, the estimate of K_d may be erroneous. Based on the largest chemical shift difference observed, the off-rate is of the order of $\sim 800\text{ s}^{-1}$ (Fig. 6).

DISCUSSION

In order to gain a better understanding of the mode of DNA binding in this family of transcription activators, we have examined the secondary structure of NifA and its interactions with DNA using ^1H NMR spectroscopy. The chemical shift indices and the patterns of NOEs show that the C-terminal domain contains three helices at residues 23–34, 39–45 and 51–59. There is no evidence for any regular secondary structure prior to residue 23.

The only other σ^{54} -dependent transcription factor for which the structure of the DNA-binding domain has been examined is that of NtrC^{3-ala} from *S.typhimurium* (18). NtrC is dimeric and contains four helices. Alignment of the secondary structural elements of the two proteins shows that the three helices in the NifA CTD align well with those in NtrC CTD (Fig. 7). In NtrC, however, there is an additional N-terminal helix between residues 402 and 414 that is absent in NifA and helix B in NtrC is longer than that in NifA. In NtrC, helices A and B form an anti-parallel four-helix bundle with the corresponding helices from a second subunit in the dimer. There are several hydrophobic contacts between residues in these helices, involving Leu421, Leu422 and Leu429 at the N-terminus of helix B. Amino acids corresponding to these

leucine residues are absent in NifA, hence it is unable to form the dimer interface.

The second half of helix B of NtrC and the remaining two helices align well with the helices of NifA. Glu24 of NifA CTD is equivalent to Glu430 of helix B in NtrC and this helix contains the Ala–Leu of the conserved Ala–Leu–X₉–Ala–Ala–X₂–Leu–Gly sequence (12). The four C-terminal residues of helix C of NtrC contain the Ala–Ala–X₂ part of the sequence and are identical to those in NifA. The Leu–Gly sequence is thus in the turn between helices C and D. It has not been possible to determine the secondary structure of Pro50 in NifA as it does not contain an amide proton, but residues Arg51–Asp61 in NifA CTD align with helix D.

The structure of the DNA-binding domain of TyrR from *Haemophilus influenzae* has also been determined (37). TyrR belongs to the NtrC family of proteins, as it contains an AAA⁺-like central domain, but it is a σ^{70} -dependent transcriptional activator. Like NtrC, it contains four helices but it is monomeric. The last three helices of TyrR align with those of NifA and NtrC. The first, so-called, hinge helix, is poorly formed and does not show any sequence homology to either NtrC or NifA. However, a hydrophobic patch on helix B that has been suggested to contact the central domain may be present in all three proteins. In NifA this would include residues Arg27, Ala30 and Leu33.

Figure 8 shows a schematic model of the terminal three helices of NifA, based on the orientation of the corresponding helices in NtrC (18). The side chains of the residues that give the largest change in chemical shift on addition of DNA are shown and coloured. The largest chemical shift differences on binding DNA are clustered, occurring at Gln39 and Ala40 at the N-terminus of the second helix, at Thr49 in the turn, at Arg51 at the N-terminus of the final helix and at Ala54 in the

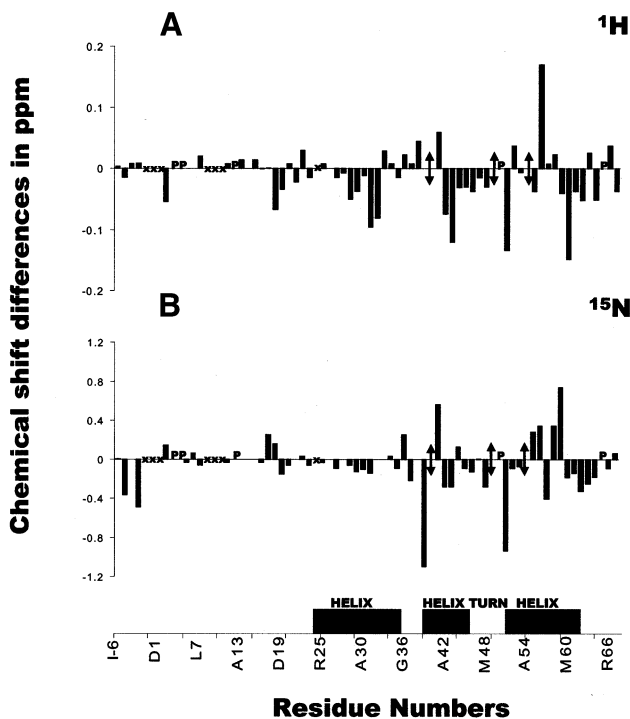


Figure 5. Histogram showing the differences in chemical shift between the NifA CTD and the domain in the presence of 15 equivalents of an oligodeoxynucleotide containing a half-site UAS. P marks the position of proline. Chemical shift difference values have not been determined for residues G-2, S-1, D1, A8, S9, S10 and D24 (marked X), due to overlap in the spectrum of the free protein. Residues A40, T49 and A54 are missing due to broadening and are marked by double-ended arrows. (A) Differences in $^1\text{H}_\text{N}$ shift. The error estimate is ± 0.03 p.p.m. (B) Differences in ^{15}N shift. The error estimate is ± 0.13 p.p.m.

centre of the final helix. We also see significant changes at Arg56 in the centre of this helix and at Ile59 and Met60 at the end of helix D. These chemical shift changes indicate changes in the chemical environment of the amide group. This can be due either to direct interactions between the residue and DNA or because the residue is undergoing a change in conformation as a result of DNA binding elsewhere. The positions in the NifA CTD affected on DNA binding are similar to those expected from direct interactions of other HTH proteins and their operator sequences. For example, interactions of bacteriophage 434 repressor with the phosphate backbone of DNA occur at the N-terminus of the first helix; base-specific DNA protein contacts occur at the N-terminus of the second helix and one turn further away; while further contacts with the DNA phosphate backbone occur at residues two turns further away (38). The binding of the protein to the DNA thus is similar to that of other HTH proteins, directly verifying the theoretical predictions of Drummond *et al.* (17). Because of the absence of the amide hydrogen at Pro50, we cannot define its secondary structure or the effects of DNA on this residue from the HSQC spectra. However, large effects of DNA binding occur at Arg51, Ala54 and Tyr55, suggesting that Pro50 (one turn from Ala54) is at the beginning of the recognition helix and so would also interact with the DNA. This would agree with sequence alignment with NtrC. The three alanine mutations in the mutant NtrC^{3-ala} CTD that

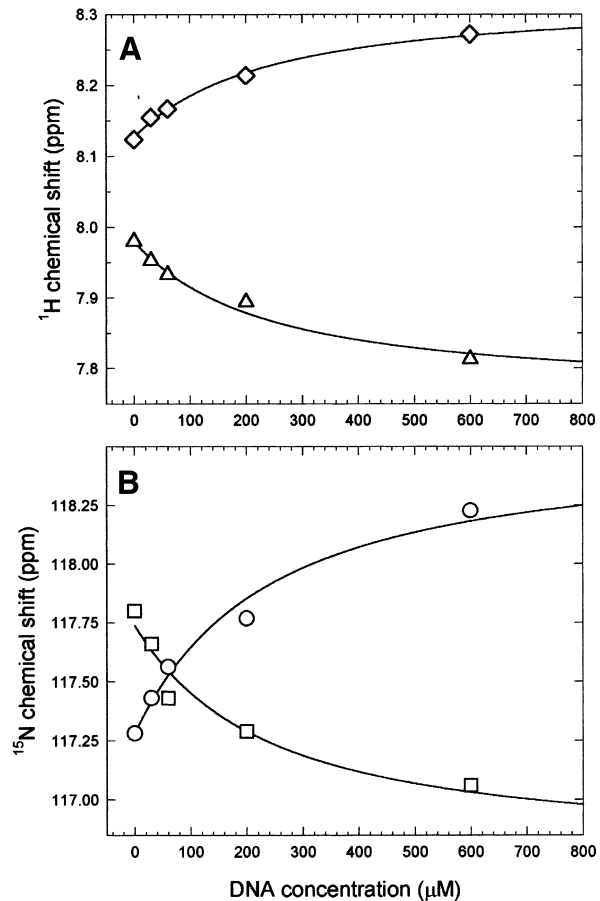


Figure 6. Chemical shifts of selected resonances at different oligodeoxynucleotide concentrations. The solid line shows the theoretical titration curves for 1:1 protein:DNA complex formation with a K_d of 200 μM in fast exchange. (A) ^1H shifts of M60 (diamonds) and R56 (triangles). (B) ^{15}N shifts of R51 (circles) and I59 (squares).

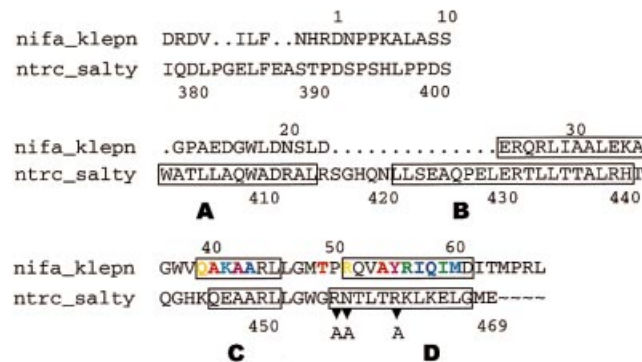


Figure 7. Sequence alignment of the C-terminal domains of NifA from *K.pneumoniae* and NtrC from *S.typhimurium*. The residues in an α -helical conformation are boxed. Helices in NtrC are labelled (A)–(D) according to Pelton *et al.* (18). The three residues mutated to Ala in the NtrC construct are indicated. Residues in NifA that show sizeable changes in chemical shift on the addition of DNA are coloured based on their weighted average change in ^1H and ^{15}N chemical shift, $\delta_{\text{av}} = [0.5\{(\delta_{\text{H}})^2 + 0.2(\delta_{\text{N}})^2\}]^{1/2}$. Red, residues in intermediate exchange; yellow, residues with $\delta_{\text{av}} = 0.3$ p.p.m.; green, residues with $\delta_{\text{av}} = 0.15$ p.p.m.; blue, residues with $\delta_{\text{av}} = 0.1$ p.p.m.; violet, residues with $\delta_{\text{av}} = 0.08$ p.p.m. The sequences were aligned using the subprogram Pile-up in GCG (Wisconsin Package version 10.2, Genetics Computer Group, Madison, WI).

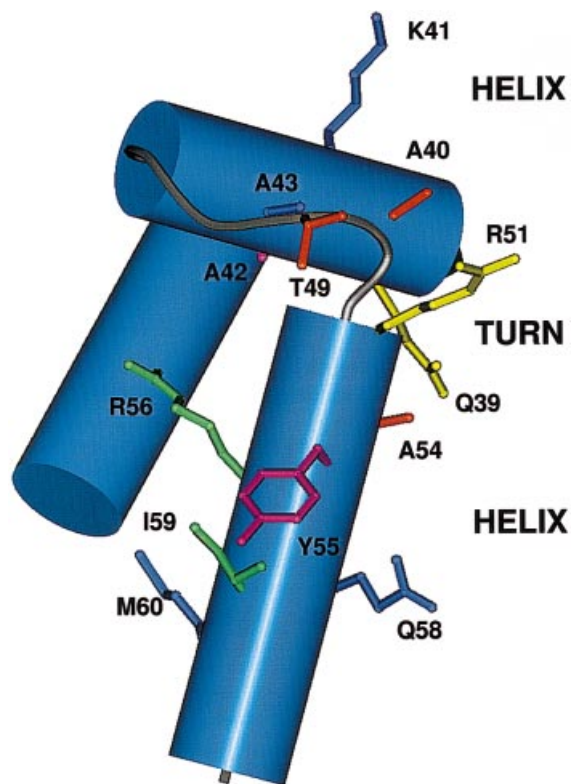


Figure 8. Schematic model of the terminal three helices of NifA, based on the orientation of the corresponding helices in NtrC (18). The helices are shown as blue cylinders with grey ribbon for the turns. The side chains of the residues that give the largest change in chemical shift on addition of DNA are shown, coloured according to their weighted average ^1H and ^{15}N chemical shift, δ_{av} as in Figure 7. The sequences of NifA and NtrC were aligned as in Figure 7 and NifA was modelled on the structure of NtrC using the program Modeller 4 (40).

replace Arg456, Asn457 and Arg461 align with Pro50, Arg51 and Tyr55, respectively, of NifA CTD, explaining why this protein does not bind DNA.

In studies by Morett *et al.* (8) on the full-length NifA protein, mutations corresponding to R51Q, R51L, Y55F and R56Q showed poor activation from the *nifH* promoter. This suggests that these residues may bind DNA. R51K and R56K showed activity equivalent to wild-type, suggesting that lysine, but not other residues, can substitute for these arginines, and that Arg51 and Arg56 may form contacts with DNA at the *nifH* UAS that are essential for transcriptional activation *in vivo*. Surprisingly, A54G also was active, although Ala54 is strongly affected by DNA binding in this study. It is possible that the backbone amide group contacts DNA and is important rather than the side chain. Q52G showed poor activity, although in our study, Gln52 is not affected by DNA binding. However, this mutant protein may be affected in folding as it did not activate the *glnAp2* promoter, which can be activated *in vivo* in the absence of the DNA-binding domain. Gly47, the conserved glycine in the turn of the motif, is unaffected by the presence of DNA, as expected, although it is structurally important for DNA binding, as shown in mutational studies of NifA from other species (39). Thus, our NMR studies support the previous biological findings.

In conclusion, our studies show that the C-terminal domain of NifA from *K.pneumoniae* contains three α -helices and provide strong evidence that the final two helices comprise a classical HTH motif required for DNA recognition. Based on the changes observed in the chemical shifts, we have demonstrated that the N-terminus of the second helix and residues on one face of the third helix interact with the cognate DNA sequence. This is the first direct evidence of the mode of DNA recognition in the σ^{54} -dependent class of enhancer-binding proteins.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank S. Kustu for the kind gift of plasmid pJES470, Y. Gao and A. J. Pemberton for maintenance of the NMR instrumentation at the University of Birmingham and A. L. Lovering for modelling NifA. This study was supported in part by a BBSRC-CASE studentship (to P.R.). The NMR instrumentation at the University of Birmingham was supported in part by grant 056432 from the Wellcome Trust (to E.I.H.), while the computing facilities were purchased by grant G460017 from the MRC. K.J.S. was supported by grant G9815098 from the MRC.

REFERENCES

- Santero,E., Hoover,T., Keener,J. and Kustu,S. (1989) *In vitro* activity of the nitrogen fixation regulatory protein NifA. *Proc. Natl Acad. Sci. USA*, **86**, 7346–7350.
- Merrick,M.J. (1993) In a class of its own—the RNA polymerase sigma factor σ^{54} (σ^N). *Mol. Microbiol.*, **10**, 903–909.
- Buck,M., Gallegos,M.T., Studholme,D.J., Guo,Y. and Gralla,J.D. (2000) The bacterial enhancer-dependent σ^{54} (σ^N) transcription factor. *J. Bacteriol.*, **182**, 4129–4136.
- Cannon,W.V., Gallegos,M.T. and Buck,M. (2000) Isomerization of a binary sigma-promoter DNA complex by transcription activators. *Nature Struct. Biol.*, **7**, 594–601.
- Dixon,R. (1998) The oxygen-responsive NifL–NifA complex: a novel two-component regulatory system controlling nitrogenase synthesis in γ -proteobacteria. *Arch. Microbiol.*, **169**, 371–380.
- Morett,E. and Buck,M. (1988) NifA-dependent *in vivo* protection demonstrates that the upstream activator sequence of *nif* promoters is a protein binding site. *Proc. Natl Acad. Sci. USA*, **85**, 9401–9405.
- Beynon,J., Cannon,M., Buchanan-Wollaston,V. and Cannon,F. (1983) The *nif* promoters of *Klebsiella* have a characteristic primary structure. *Cell*, **44**, 665–671.
- Morett,E., Cannon,W. and Buck,M. (1988) The DNA-binding domain of the transcriptional activator protein NifA resides in its carboxy terminus, recognises the upstream activator sequences of *nif* promoters and can be separated from the positive control function of NifA. *Nucleic Acids Res.*, **16**, 11469–11488.
- Hoover,T.R., Santero,E., Porter,S. and Kustu,S. (1990) The integration host factor stimulates interactions of RNA polymerase with NifA, the transcriptional activator for nitrogen fixation operons. *Cell*, **63**, 11–22.
- Ho,Y.S., Burden,L.M. and Hurley,J.H. (2000) Structure of the GAF domain, a ubiquitous signalling motif and a new class of cyclic GMP receptor. *EMBO J.*, **19**, 5288–5299.
- Wootton,J.C. and Drummond,M.H. (1989) The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.*, **2**, 535–543.
- Morett,E. and Segovia,L. (1993) The σ^{54} bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J. Bacteriol.*, **175**, 6067–6074.

13. Neuwald,A.F. Aravind,L., Spouge,J.L. and Koonin,E.V. (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.*, **9**, 27–43.
14. Berger,D.K., Narberhaus,F. and Kustu,S. (1994) The isolated catalytic domain of NifA, a bacterial enhancer protein, activates transcription *in vitro*: activation is inhibited by NifL. *Proc. Natl Acad. Sci. USA*, **91**, 103–107.
15. Barrett,J., Ray,P., Sobczyk,A., Little,R. and Dixon,R. (2001) Concerted inhibition of the transcriptional activation functions of the enhancer-binding protein NIFA by the anti-activator NIFL. *Mol. Microbiol.*, **39**, 480–494.
16. Chaney,M., Grande,R., Wigneshwararaj,S.R., Cannon,W., Casaz,P., Gallegos,M.T., Schumacher,J., Jones,S., Elderkin,S., Dago,A.E., Morett,E. and Buck,M. (2001) Binding of transcriptional activators to sigma 54 in the presence of the transition state analog ADP-aluminum fluoride: insights into activator mechanochemical action. *Genes Dev.*, **15**, 2282–2294.
17. Drummond,M., Whitty,P. and Wootton,J. (1986) Sequence and domain relationships of NtrC and NifA from *Klebsiella pneumoniae*: homologies to other regulatory proteins. *EMBO J.*, **5**, 441–447.
18. Pelton,J.G., Kustu,S. and Wemmer,D.E. (1999) Solution structure of the DNA-binding domain of NtrC with three alanine substitutions. *J. Mol. Biol.*, **292**, 1095–1110.
19. Safo,M.K., Yang,W.Z., Corselli,L., Cramton,S.E., Yuan,H.S. and Johnson,R.C. (1997) The transactivation of the Fis protein that controls site-specific DNA inversion contains extended mobile β -hairpin arms. *EMBO J.*, **16**, 6860–6873.
20. Klose,K.E., North,A.K., Stedman,K.M. and Kustu,S. (1994) The major dimerisation determinants of the nitrogen regulatory protein NtrC from enteric bacteria lie in its carboxy-terminal domain. *J. Mol. Biol.*, **241**, 233–245.
21. Missailidis,S., Jaseja,M., Ray,P., Chittock,R., Wharton,C.W., Drake,A.F., Buck,M. and Hyde,E.I. (1999) Secondary structure of the C-terminal DNA-binding domain of the transcriptional activator NifA from *Klebsiella pneumoniae*: spectroscopic analyses. *Arch. Biochem. Biophys.*, **361**, 173–182.
22. Lee,H.S., Berger,D.K. and Kustu,S. (1993) Activity of purified NifA, a transcriptional activator of nitrogen fixation genes. *Proc. Natl Acad. Sci. USA*, **90**, 2266–2270.
23. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
24. Bradford,M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248–254.
25. Buck,M., Miller,S., Drummond,M. and Dixon,R. (1986) Upstream activator sequences are present in the promoter of nitrogen fixation genes. *Nature*, **320**, 374–378.
26. Buck,M., Cannon,W. and Woodcock,J. (1987) Mutational analysis of upstream sequences are required for transcriptional activation of the *Klebsiella pneumoniae nifH* promoter. *Nucleic Acids Res.*, **15**, 9945–9956.
27. Muhandiram,D.R. and Kay,L.E. (1994) Gradient-enhanced triple resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B*, **103**, 203–216.
28. Matsuo,H., Kupce,E., Li,H. and Wagner,G. (1996) Increased sensitivity in HNCA and HN(CO)CA experiments by selective C^{β} decoupling. *J. Magn. Reson. B*, **113**, 91–96.
29. Kay,L.E. (1993) Pulse-field gradient enhanced 3 dimensional NMR experiments for correlating C^{13} α - β , $C^{13\alpha}$ and $H\alpha_{i+1}$ chemical shifts in uniformly C^{13} labeled proteins dissolved in H_2O . *J. Am. Chem. Soc.*, **115**, 2055–2057.
30. Grzesiek,S. and Bax,A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly C^{13}/N^{15} enriched proteins. *J. Biomol. NMR*, **3**, 185–204.
31. Kuboniwa,H., Grzesiek,S., Delaglio,F. and Bax,A. (1994) Measurement of H_N - H_{α} J-couplings in calcium-free calmodulin using new 2D and 3D water flip-back methods. *J. Biomol. NMR*, **4**, 871–878.
32. Zhang,O.W., Kay,L.E., Olivier,J.B. and Formankay,J.D. (1994) Backbone H-1 and N-15 resonance assignments of the N-terminal SH3 domain of DRK in folded and unfolded states using enhance sensitivity pulsed-fielded gradient techniques. *J. Biomol. NMR*, **4**, 845–858.
33. Live,D.H., Davis,D.G., Agosta,W.C. and Cowburn,D. (1984) Long range hydrogen bond mediated effects in peptides: 15N NMR study of gramicidin S in water and organic solvents. *J. Am. Chem. Soc.*, **106**, 1939–1943.
34. Wishart,D.S. and Sykes,B.D. (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR*, **4**, 171–180.
35. Mori,S., Abeygunawardana,C., Johnson,M.O. and Vanzyl,P. (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B*, **108**, 94–98.
36. Lian,L.-Y. and Roberts,G.C.K. (1993) Effects of chemical exchange on NMR spectra. In Roberts,G.C.K. (ed.), *NMR of Macromolecules. A Practical Approach*. IRL Press, Oxford, pp. 153–182.
37. Wang,Y., Zhao,S., Somerville,R.L. and Jardetzky,O. (2001) Solution structure of the DNA-binding domain of the TyrR protein of *Haemophilus influenzae*. *Protein Sci.*, **10**, 592–598.
38. Anderson,W., Ptashne,M. and Harrison,S. (1987) Structure of the repressor-operator complex of bacteriophage 434. *Nature*, **355**, 87–89.
39. Juarez,K., Flores,H., Davila,S., Olvera,L., Gonzales,V. and Morett,E. (2000) Reciprocal domain evolution within a transactivator in a restricted sequence space. *Proc. Natl Acad. Sci. USA*, **97**, 3314–3318.
40. Sali,A., Yuan,F., van Vlijmen,H. and Karplus,M. (1995) Evaluation of comparative protein modelling by MODELLER. *Proteins*, **23**, 318–326.