

# Flexibility of KorA, a plasmid-encoded, global transcription regulator, in the presence and the absence of its operator

Karthik V. Rajasekar<sup>1</sup>, Andrew L. Lovering<sup>1</sup>, Felician Dancea<sup>2</sup>, David J. Scott<sup>3</sup>, Sarah A. Harris<sup>4</sup>, Lewis E.H. Bingle<sup>1</sup>, Manfred Roessle<sup>5</sup>, Christopher M. Thomas<sup>1</sup>, Eva I. Hyde<sup>1,\*</sup> and Scott A. White<sup>1,\*</sup>

<sup>1</sup>School of Biosciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK, <sup>2</sup>School of Cancer Studies, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK, <sup>3</sup>School of Biosciences, Sutton Bonington Campus, University of Nottingham, Nottingham LE12 5RD, UK, <sup>4</sup>School of Physics and Astronomy and Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK and <sup>5</sup>EMBL, DESY, 22607 Hamburg, Germany

Received July 29, 2015; Revised March 8, 2016; Accepted March 11, 2016

## ABSTRACT

The IncP (Incompatibility group P) plasmids are important carriers in the spread of antibiotic resistance across Gram-negative bacteria. Gene expression in the IncP-1 plasmids is stringently controlled by a network of four global repressors, KorA, KorB, TrbA and KorC interacting cooperatively. Intriguingly, KorA and KorB can act as co-repressors at varying distances between their operators, even when they are moved to be on opposite sides of the DNA. KorA is a homodimer with the 101-amino acid subunits, folding into an N-terminal DNA-binding domain and a C-terminal dimerization domain. In this study, we have determined the structures of the free KorA repressor and two complexes each bound to a 20-bp palindromic DNA duplex containing its consensus operator sequence. Using a combination of X-ray crystallography, nuclear magnetic resonance spectroscopy, SAXS and molecular dynamics calculations, we show that the linker between the two domains is very flexible and the protein remains highly mobile in the presence of DNA. This flexibility allows the DNA-binding domains of the dimer to straddle the operator DNA on binding and is likely to be important in cooperative binding to KorB. Unexpectedly, the C-

terminal domain of KorA is structurally similar to the dimerization domain of the tumour suppressor p53.

## INTRODUCTION

The IncP (Incompatibility group P) plasmids are important carriers of antibiotic resistance. They are low-copy-number plasmids that can transfer to and be stably maintained in almost all Gram-negative bacteria, as well as being able to transfer to some Gram-positive bacteria and higher eukaryotic cells. Gene expression of all the plasmid backbone functions, such as replication, partitioning and plasmid transfer, is stringently controlled by a network of four plasmid-encoded repressor proteins: KorA, KorB, KorC and TrbA (1). Two repressor proteins bind to most core promoters. KorB also plays a role in plasmid partitioning as the ParB (DNA-binding protein) homologue. The KorA protein from the IncP1 plasmid RK2 binds with KorB at five of its seven operator ( $O_A$ ) sites (2) and they have been shown to act cooperatively at two of these (3). This cooperativity involves the C-terminal domain (CTD) of KorA, which has a 77.4% sequence similarity to that of TrbA (Supplementary Figure S1), that also acts cooperatively with KorB (4). Key operons controlled by KorA and KorB acting together include the operons involved in vegetative replication and conjugative gene transfer, and the central control region, encoding KorA, KorB and IncC (the ParA homologue) required for plasmid partitioning. KorA appears to be uniquely responsible for controlling gene switching be-

\*To whom correspondence should be addressed. Tel: +44 121 414 7534; Fax: +44 121 414 5925; Email: s.a.white@bham.ac.uk  
Correspondence may also be addressed to E. I. Hyde. Tel: +44 121 414 5393; Fax: +44 121 414 5925; Email: e.i.hyde@bham.ac.uk  
Present addresses:

Karthik Rajasekar, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.

Felician Dancea, First Private Investment Management, 60327 Frankfurt, Germany.

Lewis Bingle, Department of Pharmacy, Health and Wellbeing, University of Sunderland, Sunderland SR1 3SD, UK.

Manfred Roessle, Fachhochschule Lübeck, Mönkhofer Weg 239, 23562 Lübeck, Germany.

tween the two different modes of plasmid DNA replication (5,6). Intriguingly, KorB is able to repress and act cooperatively at distances more than 1000 bp upstream or downstream from the binding site of the repressor TrbA (7), similar to the effects of eukaryotic enhancers. Cooperative binding also occurs between KorB and either KorA or TrbA at short distances, and surprisingly the interaction remains cooperative even when 5 bp, introducing a twist of 180°, is inserted between the two binding sites (7). The effects at very short distances cannot be explained by DNA looping while the long-range effects cannot be explained by the protein spreading along the DNA.

In order to understand the molecular basis of the cooperativity between KorA and KorB, we are examining their structures and interactions. KorA is a 101-amino acid, homodimeric, protein with an N-terminal DNA-binding domain (DBD), containing a helix-turn-helix motif, and a C-terminal dimerization region (3), joined by a linker of four amino acids. A crystal structure of KorA bound to a 18-bp operator (O<sub>A</sub>) DNA has been determined previously (8). Here we examine the structure and mobility of KorA in the absence of DNA and structural changes upon DNA binding, using a combination of crystallography, nuclear magnetic resonance (NMR) spectroscopy, small angle X-ray scattering (SAXS) and molecular dynamics (MD). We show that the DBDs significantly change orientation upon DNA binding and that the CTD is mobile, due to the highly flexible linker, both in the free protein and DNA-bound complexes. This flexibility is likely to contribute to the cooperativity between KorA and KorB at different distances. Surprisingly, despite the lack of sequence homology, the CTD is structurally similar to the dimerization domain of the tumour suppressor p53.

## MATERIALS AND METHODS

### Protein expression and purification

The *korA* gene from RK2 was expressed and the protein purified as described previously (9). <sup>13</sup>C- and/or <sup>15</sup>N- labelled protein was expressed and purified from cells grown in M9 minimal medium with <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> as the sole nitrogen and carbon sources, respectively.

### DNA

A palindromic oligonucleotide, containing the sequence:

C<sub>1</sub>. C<sub>2</sub>. A<sub>3</sub>. A<sub>4</sub>. G<sub>5</sub>. T<sub>6</sub>. T<sub>7</sub>. T<sub>8</sub>. A<sub>9</sub>. G<sub>10</sub>. C<sub>11</sub>. T<sub>12</sub>. A<sub>13</sub>. A<sub>14</sub>. A<sub>15</sub>. C<sub>16</sub>. T<sub>17</sub>. T<sub>18</sub>. G<sub>19</sub>. G<sub>20</sub> was purchased from MWG-Biotech AG. This is the consensus 12 bp O<sub>A</sub> sequence (underlined), plus two additional flanking base pairs from the strongest O<sub>A</sub> site, plus two CG base pairs at either end. It was annealed in buffer containing 10 mM Tris-HCl pH 7.0, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl.

The KorA-O<sub>A</sub> complex (~1 mM) was made by mixing KorA dimers and the operator at a ratio of 1:1.05 and dialysed into 20 mM Tris pH 7.0, 100 mM NaCl, 1 mM EDTA. For NMR studies the complex was made in 20 mM sodium phosphate buffer pH 6.2, 100 mM NaCl and 1 mM EDTA.

### X-ray crystallography

Crystals of unbound KorA were grown by vapour diffusion with 20 mg.ml<sup>-1</sup> protein mixed 1:1 with reservoir consisting of 200 mM sodium acetate, pH 4.5, 200 mM ammonium acetate and 32.5% w/v PEG-4K. KorA-O<sub>A</sub> crystals were grown by vapour-diffusion with 14 mg.ml<sup>-1</sup> KorA-O<sub>A</sub> complex mixed 1:1 with reservoir consisting of 4–8% PEG-6K, 100 mM sodium acetate, pH 4.6 and 25% ethylene glycol. Crystallographic data (Supplementary Table S1) were analysed with XDS (10), CCP4 (11) and PHENIX programs (12). The structure of free KorA was phased using MIRAS with phase improvement using SOLVE/RESOLVE (13). Two structures of DNA-bound KorA-O<sub>A</sub> complexes were determined by molecular replacement (PHASER (14)) using the DBD and CTD from the free structure as separate search models, together with a theoretical model of O<sub>A</sub>. All models were built and improved using COOT (15), and structures were refined using both REFMAC (16) and PHENIX.REFINE (12).

### NMR data acquisition and analysis

NMR relaxation experiments were recorded on a Varian Unityplus 600 MHz spectrometer. NOESY-<sup>15</sup>N-HSQC and NOESY-<sup>13</sup>C-HSQC were recorded on Varian 800 MHz NMR spectrometers. All experiments were done in 20 mM sodium phosphate buffer pH 6.2, 100 mM NaCl and 1 mM EDTA at 298K or 303K for the free protein and at 308K for the protein-DNA complex. Spectra were processed with NMRPIPE software (17) and analysed using CCPNMR (18). <sup>15</sup>N and <sup>1</sup>H chemical shift differences between free protein and DNA-bound amide resonances of KorA, combined using the formula  $\delta \text{ comb}^2 = (0.15 \delta^{15\text{N}})^2 + (\delta^{1\text{H}})^2$  (19).

Amide-<sup>15</sup>N relaxation data were collected using the method of Kay *et al.* (20). Peak heights for T<sub>1</sub> and T<sub>2</sub> data were fitted to a single exponential,  $F = a \cdot e^{-bx}$ , using CCPNMR software (18). The programme TENSOR2 was used to determine the correlation times for the two domains separately, assuming isotropic tumbling (21).

### Structure determination by NMR spectroscopy

The backbone dihedral angles were predicted from the chemical shifts using TALOS (22). Initially 1200 NOEs were assigned manually from the <sup>15</sup>N NOESY-HSQC spectrum and 460 NOEs assigned in the <sup>13</sup>C NOESY-HSQC. Automated assignment of the remaining NOEs was performed using the symmetric dimer protocols of ARIA (23), with an initial torsion angle dynamics (TAD) step followed by MD-simulated annealing (SA) procedure, starting from the protein structure in KorA-O<sub>A</sub> Complex 1 (PDB ID: 5CM3). In addition, four hydrogen bonds at the dimer interface, supported by the NOE patterns, were imposed throughout the iterative assignment stages. This led to a final list of 2588 unambiguous and 124 ambiguous NOE-derived distance restraints, including duplicates.

The final structure calculations were carried out from a random coil structure using the assigned NOE and dihedral angles, with the same TAD MD-SA protocol, along with two additional Cartesian dynamics cooling phases. The 20

energetically best structures from the calculated 200 structures were refined in water (24), during which no symmetry restraints were imposed (Supplementary Table S2). The median NMR model was calculated by a maximum likelihood superpositioning using the programme THESEUS (25).

### Small angle X-ray scattering data acquisition and analysis

SAXS data were collected at Beamline X33 at DESY (Hamburg), using three sample concentrations 2.5 mg.ml<sup>-1</sup>; 5 mg.ml<sup>-1</sup> and 10 mg.ml<sup>-1</sup>, in 20 mM Tris-HCl pH 7.0, 100 mM NaCl 1 mM EDTA buffer, at 293K. The datasets from the three samples were merged using the programme PRIMUS (26). The data were confirmed by SEC-SAXS, at Beamline BM29 at ESRF (Grenoble). About 50 µl KorA at 4 mg.ml<sup>-1</sup> was loaded onto a Superdex 200 Increase (3.2/300) column equilibrated with the above buffer, coupled to the SAXS flow cell. The Guinier plot was used to determine the radius of gyration,  $R_g$  and the intensity at zero angle,  $I_0$ , which was compared to that of bovine serum albumin to give the molecular mass of the protein. The scattering data were then examined using the dimensionless Kratky plot (27) of  $(qR_g)^2 I_q / I_0$  versus  $qR_g$  where  $q$  is the scattering vector. The distribution of conformers was determined using the EOM 2.0, Ensemble Optimization Method (28,29), to fit the scattering data, using the X-ray structures of the CTD dimer and two DBDs with the 4 aa residue linkers remaining flexible to generate the ensemble. In addition, the pair distance distribution function ( $P(r)$  versus  $r$ ) was generated using the programme GNOM (30), followed by shape reconstruction using 10 rounds of the programme GASBOR (31) to give bead models of the protein. Final reconstructed shapes were compared and averaged using the programme DAMAVER (32) and presented as an envelope using SITUS (33). In parallel, the programme BUNCH (34) was used to find the orientation of the protein domains that fitted best to the GNOM output, based on the X-ray structures of the CTD dimer and the two DBDs with 4 aa residue flexible linkers between the domains.

### Molecular dynamics calculations

All MD calculations used the AMBER 10 suite of programs (35). Simulations of the KorA-O<sub>A</sub> complex were performed using Complex 1 (PDB ID: 5CM3) as starting coordinates; containing residues 1–97 for one monomer and 1–93 of the second. The DNA was removed from this structure to perform simulations of the KorA dimer alone. The protein and protein/DNA complex were neutralized using Na<sup>+</sup> counterions, before sufficient TIP3P water molecules were added to ensure 10 Å of water in each orthogonal direction around the solute. The PARM99SB force field (36) was used to describe the protein, and the PARM99 force field (37), in conjunction with the PARMBSC0 forcefield modification (38), was used to describe the DNA. The system was energy minimized and then equilibrated using a standard multi-stage protocol. The temperature was maintained at 300 K using the Berendsen weak coupling scheme (with a coupling constant of 1 ps) and constant pressure (1 atm) using volume scaling (with a coupling constant of 0.2 ps). Long-range electrostatics interactions were treated using the Particle

Mesh Ewald technique. The SHAKE algorithm was used to constrain bonds to hydrogen, allowing a 2 ps time step to be used to integrate the equations of motion. MD simulations were performed for 100 ns. Conformations were recorded every 1 ps for further analysis. The trajectories were visualized using VMD and CHIMERA (39,40).

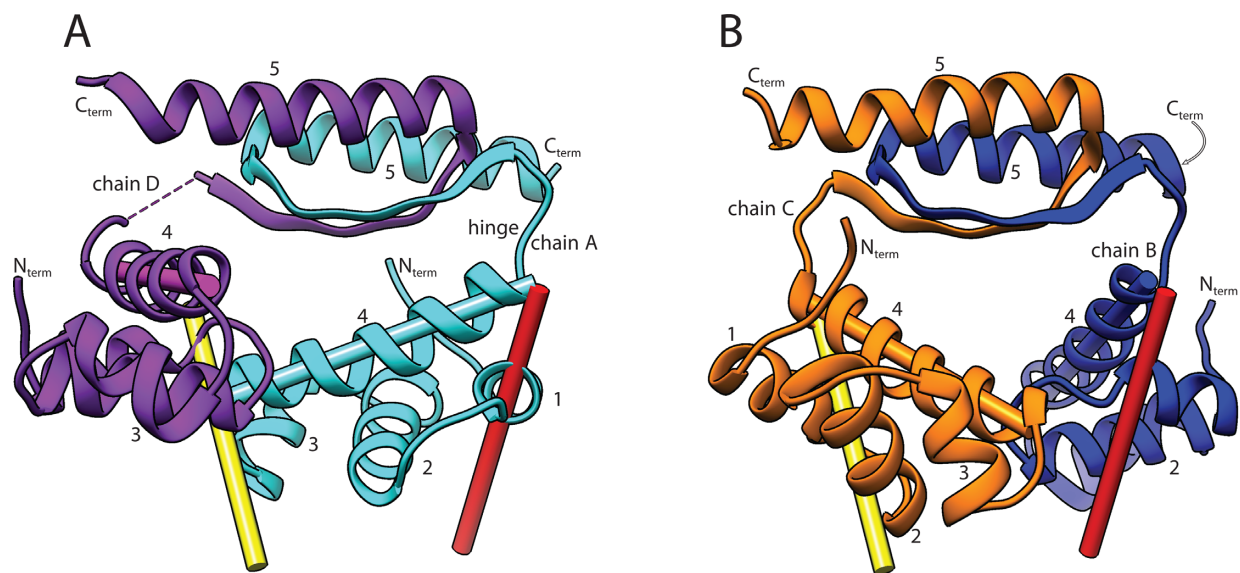
## RESULTS

### Structure of free KorA

The crystals of free KorA contained two dimers in the asymmetric unit (ASU), each with two N-terminal DNA Binding Domains with four α-helices and a C-Terminal Dimerization domain containing a single β-strand and the final helix from each monomer (Figure 1). The secondary structure of each dimer is similar to the DNA-bound forms (PDB ID: 2W7N (8) and 5CM3, 5CLV, this work), but the orientation of all four copies of the DBD relative to the CTD differs greatly, both from one another and from their positions in the DNA-bound complexes. This is shown by the solid cylinders in Figure 1 which represent the positions of helix 4, the DNA-recognition helix, in the free proteins and in the protein–DNA complexes shown in Figure 3. This variability in orientation is due to differences in conformation of the four amino acid linker between the domains (residues 66–69). This linker is highly flexible, so much so that the electron density of the linker in subunit D is not seen in the crystal structure.

Because of the variable orientation of the domains, NMR spectroscopy was used to examine the structure and dynamics of the protein in solution. The NMR spectrum of KorA was assigned previously (41), only one signal is seen for each NH group in the <sup>15</sup>N-<sup>1</sup>H HSQC spectrum confirming that the protein is symmetrical in solution on the NMR time scale (Supplementary Figure S2). <sup>15</sup>N T<sub>1</sub>, and T<sub>2</sub> relaxation times, and heteronuclear <sup>15</sup>N-<sup>1</sup>H NOEs of the main chain amide groups were measured (Figure 2A). The T<sub>1</sub> relaxation times for residues 69–93 are longer than for the N-terminal residues (apart from residue Q37 which is key for DNA binding) while the T<sub>2</sub> values show the major variation at the ends of the molecule. The T<sub>1</sub>/T<sub>2</sub> ratios suggest that the DBD of KorA has a shorter rotational correlation time ( $t_c$  11.9 ± 0.2 ns) than the CTD ( $t_c$  13.5 ± 0.31 ns), indicating that the DBD has some additional independent motion. The <sup>1</sup>H-<sup>15</sup>N heteronuclear NOEs are similar throughout the molecule except at the termini and internally around the linker at residues 66–71. The lower NOEs show that these N-H groups undergo faster internal motion than the others. The NOEs correlate with the X-ray crystallographic B factors for the protein and confirm that the linker is flexible.

The structure of each domain is well-determined from the NMR data (Figure 2B and C), and is consistent with the crystal structures (Figures 1 and 3). The structures of the DBD (residue 6–66) from both techniques fit with an RMSD of 0.92 Å, while the CTD structures (residues 71–98, dimer) fit with an RMSD of 2.9 Å. However, the relative orientation of the domains could not be determined by NMR as no NOEs were observed between the two domains, nor between the resonances of residues in the loop and the domains, indicating a high degree of disorder. Moreover, the chemical shifts of the residues 65–70 did not give strong



**Figure 1.** Ribbon diagrams of the X-ray structure of the free KorA dimers (PDB ID: 5CKT). Each subunit is labelled chain A–D at the corresponding flexible hinge region and at the N- and C-termini, while the helices are numbered 1–5 in the N to C direction. The recognition helices of the DNA-binding helix–turn–helix motifs are highlighted with a cylinder of the appropriate colour. The red and yellow cylinders indicate the approximate position of the corresponding helices in the KorA–DNA complex (PDB ID: 5CM3). (A) KorA ribbon diagram for Subunits A (cyan) and D (purple). The hinge of chain A is labelled. There is no electron density for the hinge region (residues 65–69 inclusive) in Subunit D, indicating disorder. (B) KorA ribbon diagram for Subunits B (blue) and C (orange).

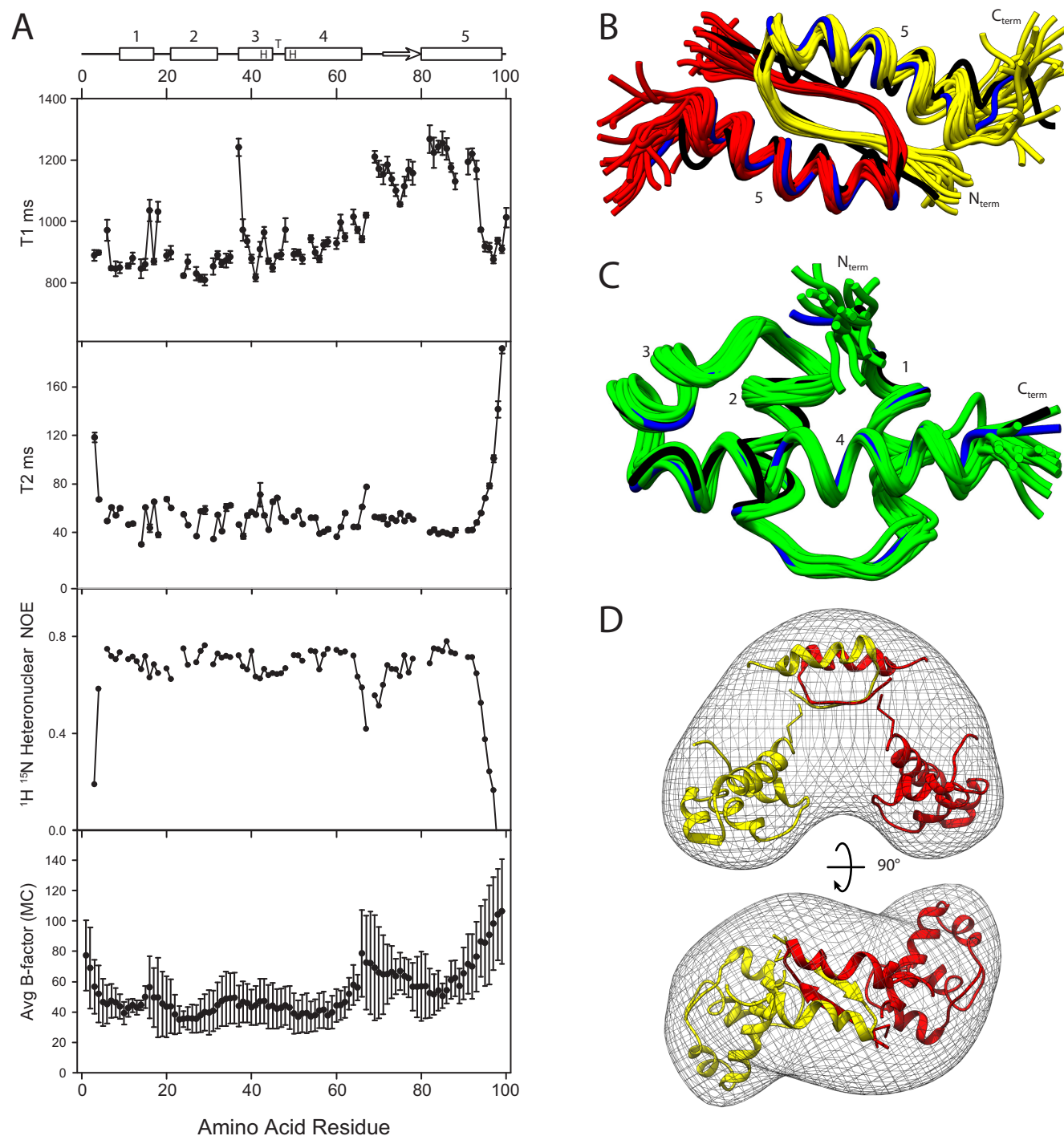
backbone angle constraints, suggesting averaging of their conformation. We therefore turned to SAXS (Supplementary Figure S3) to examine the conformational range of the protein in solution. The scattering intensity gave a linear Guinier plot, showing that the protein was monodisperse, with the scattering intensity at zero angle giving a molecular mass of 21.3 kDa, within error of the calculated value for a dimer of 22.6 kDa. The dimensionless Kratky plot (27) of  $(qR_g)^2 I_q/I_0$  versus  $qR_g$ , shows a peak with a maximum at coordinates  $(\sqrt{3}, 3/e)$ , as expected from a folded protein. However, there is also a second maximum, (approximate coordinates (4.8, 0.83)) and the plot did not go to zero, indicating that the protein has two folded domains and a flexible linker (Supplementary Figure S3D). Analysis of the scattering curve using EOM (ensemble optimization method) (28,29), shows that the protein adopts mainly compact conformations (Supplementary Figure S3E). As such, it can be useful to find a single conformation that represents an average position of the domains (42). Figure 2D shows a surface representation of a bead model of the protein based on the SAXS data, assuming P2 symmetry, using the programs GASBOR (31) and SITUS (33) while the ribbon models show the fitting of the separate domains to the SAXS data using BUNCH (34). In different BUNCH calculations, the domains rotate within the distribution but the average distance between the domains is maintained, giving a clover leaf structure.

### Structure of KorA bound to $O_A$ DNA

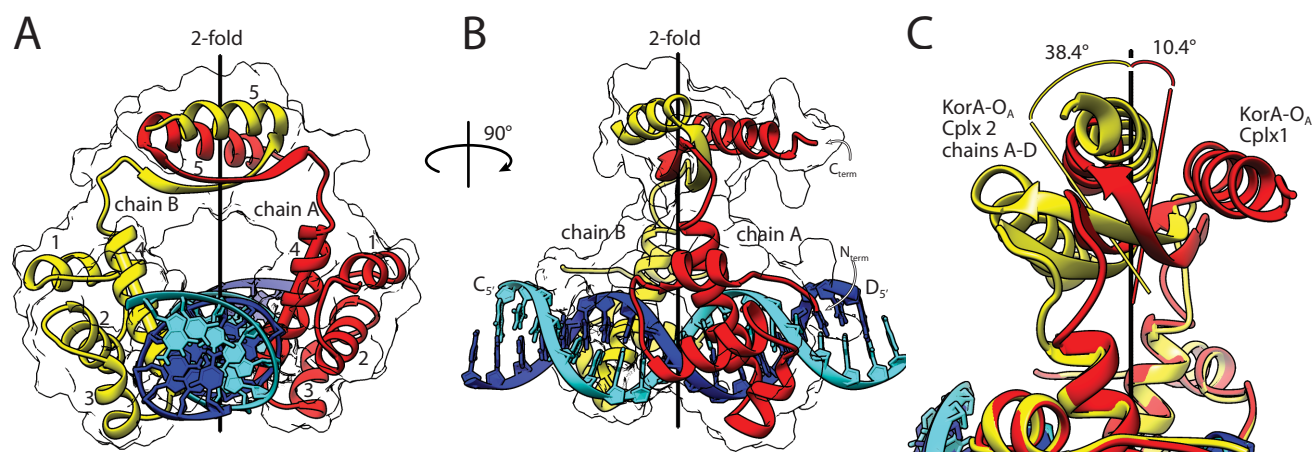
KorA was crystallized in the presence of the operator  $O_A$ . In contrast to the free protein, several crystal forms of the complex were obtained. Structures from two of these were determined. Both show the DBDs of the protein symmet-

rically placed about the DNA axis (Figure 3). In contrast, the orientation of the CTD differs in the two crystal structures. In the first complex (PDB ID: 5CM3), containing a single DNA-bound dimer in the ASU, the CTD is rotated only  $10.4^\circ$  from the local 2-fold axis. In Complex 2 (PDB ID: 5CLV), there were four copies of the dimeric protein–DNA complex in the ASU. However, only one CTD could be modelled, although all eight DBDs are clearly visible. There is only weak, partial electron density for the other three CTDs, presumably as they are highly disordered. In addition, the one CTD that is observed in 5CLV is tilted  $38.4^\circ$  away from the local 2-fold axis (Figure 3).

The protein–DNA contacts in each of these two structures are similar to each other (Supplementary Figure S4) and to that of PDB ID: 2W7N, determined previously, (8). In contrast to most proteins containing a helix–turn–helix motif, that bind to one face of the DNA; the two DBDs of KorA straddle the DNA, with helix 4 of the two subunits binding to opposite faces of the major groove (Figure 3 A and B). As in PDB ID: 2W7N (8), Q53, in the centre of helix 4, appears to be the key residue for DNA recognition as it contacts three bases, two adjacent Thymine bases on opposite strands of the DNA (T8 and T12) via its side chain amide group, and a third base (Cytosine 11) via the sidechain carbonyl group. In addition R48, at the beginning of helix 4, binds to Guanine 5 O6, but an alternative conformation of its sidechain is also seen in one half site, binding to phosphate 4. G49 is positioned so that its  $C_\alpha$  atom is in van der Waal's contact with Thymine 7-O4. This restricts both the amino acid identity and the base, as other bases, including Cytosine, would cause a steric clash. The hydroxyl group of T47 forms a water-mediated contact with Adenine 13 in both chains. The remaining protein–DNA contacts are to the phosphate groups, including many from residues 18–



**Figure 2.** NMR data for free KorA. (A) Secondary structure assignment with the 5  $\alpha$ -helices shown as rectangles, marked 1–5 and the  $\beta$ -strand shown as an arrow (HTH indicates the position of the helix-turn-helix motif);  $^{15}\text{N}$   $T_1$  and  $T_2$  relaxation times;  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear NOEs; and main chain B factors averaged over each residue of KorA in chains A–D (PDB ID: 5CKT). The error bars for the  $T_1$  and  $T_2$  measurements are the error in fitting the intensities to a single exponential, while for the B factors they show the variation in B factors for the four chains. (B) Superposition of 20 lowest energy NMR (backbone) structures for the CTD of KorA, Subunit A (red) and Subunit B (yellow). Helix 5 is numbered. The median NMR model as calculated by THESEUS is shown in blue. Also shown is the CTD of the free KorA structure (PDB ID: 5CKT) in black. (C) Superposition of 20 lowest energy NMR structures for the N-terminal DBD of KorA (green). The median NMR model is shown in blue. The DBD of 5CKT:A is shown in black. The termini and the four helices are indicated. (D) Two orientations of the GASBOR bead model of electron density that fits the SAXS distance distribution plot. For clarity the bead model is shown as an envelope with the programme SITUS (33). Superposed is the ribbon model of the optimum orientation of the X-ray structure of the domains fitted to the same distribution using the programme BUNCH (34).



**Figure 3.** X-ray crystal structures of KorA bound to a 20-bp oligonucleotide containing the  $O_A$  consensus sequence. (A) KorA- $O_A$  Complex 1 (PDB ID: 5CM3)- ribbon diagram with silhouette of protein surface and local 2-fold axis calculated using the DBDs of Subunits A and B. The protein is coloured: Subunit A in red, B in yellow. Chains C and D of the operator DNA are coloured cyan and blue, respectively. Helices are numbered 1–5 for each chain. The recognition helices of subunits A and B are highlighted using solid cylinders (see also Figure 1). (B) PDB ID: 5CM3, rotated 90° from A about the local 2-fold axis. Chains C and D are labelled at their 5' ends. (C) Part of chains A–D of KorA- $O_A$  Complex 2 (PDB ID: 5CLV) depicted in yellow (protein ribbon) and cyan (DNA) superimposed onto chains A and B of Complex 1 (PDB ID: 5CM3) shown in red (protein ribbon) and blue (DNA). Also shown are local 2-fold axes calculated from the CTDs of 5CLV (yellow) and 5CM3 (red) and the angular displacement from the DBD 2-fold axis.

23, outside the helix-turn-helix motif. There are small differences in the contacts between the subunits, eg. R57 in helix 4 contacts phosphate 9 in one chain and phosphate 10 in the other.

Because of the differences seen in the crystal structures, we examined the mobility of the protein–DNA complex in solution, by NMR spectroscopy. In the NMR spectra the signals of bound and free protein are in slow exchange, showing tight binding. The  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of the complex (Figure 4A) was assigned using sequential NH–NH and NH- $\text{C}\alpha\text{H}$  NOEs. The differences in chemical shifts between the  $^{15}\text{N}$  and  $^1\text{H}$  amide resonances in the free and DNA-bound protein are shown in Figure 4B. The CTD and the first 17 amino acid residues show minimal changes in chemical shift on binding DNA, suggesting that they are unaffected by DNA binding. However changes in shifts are seen for the remaining residues, with the largest effects at residues 18–23, 37–38 and 47–53, in agreement with these NH groups being close to the DNA, as in the crystal structures. Heteronuclear  $^{15}\text{N}$ - $^1\text{H}$  NOE measurements of the KorA- $O_A$  complex again show a dip in the NOE intensity for residues in the linker region, showing that the linker retains some flexibility in the presence of DNA (Figure 4B).

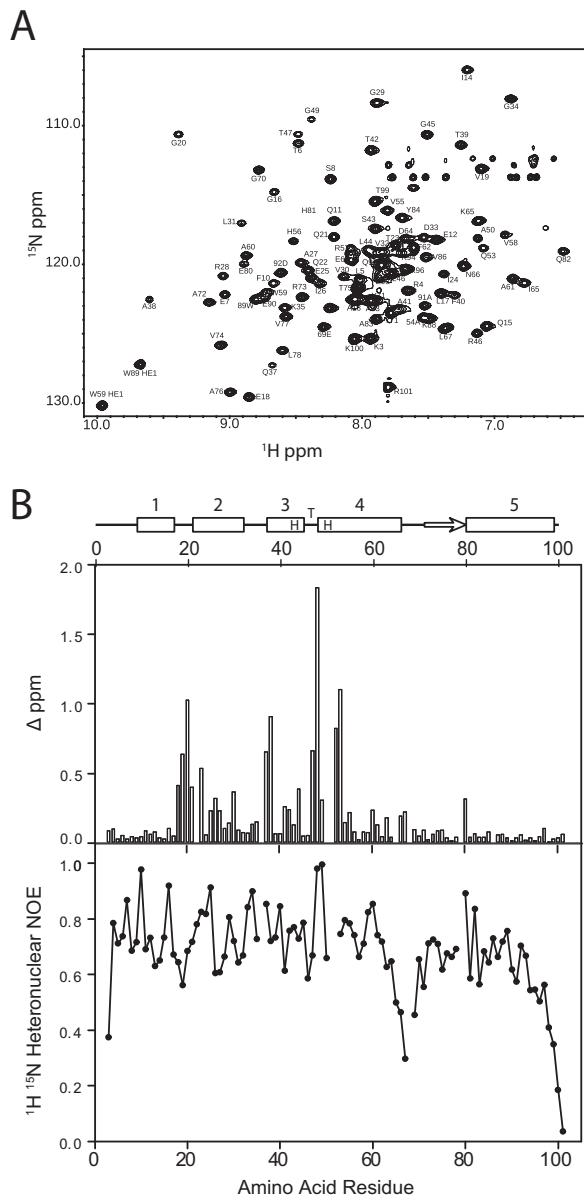
Given the highly flexible structure of the protein, MD simulations were used to explore the conformations of the free and bound protein in solution. Figure 5 shows an overlay of the conformers sampled for the KorA–DNA complex, over 100 ns of simulation; the complex remains very flexible with the CTD exploring a wider range of orientations than the two observed in the crystal structures (Supplementary Movie 1). When the protein is removed from the DNA complex, the DBDs collapse towards each other, giving additional fluctuating contacts which keep the protein compact (Supplementary Movie 2). Dynamic interactions observed between the two DBDs include salt bridges between R57 and E18 and hydrogen bonds between R48 and Q11/Q15. In the DNA complex these interactions are re-

placed: R57 and E18 both bind to the phosphate backbone, while R48 binds to Guanine 5. The relatively high T1 values for G16 and E18 in the NMR studies of the free protein may reflect their dynamics.

## DISCUSSION

The structure of KorA consists of two domains, joined by a highly flexible, four amino acid linker, which retains its mobility even when bound to DNA. Only by combining a suite of complementary biophysical techniques (X-ray crystallography, NMR spectroscopy, SAXS and MD) was it possible to study this system due to its inherently dynamic nature. The mobility of the KorA DBDs, in the absence of DNA, is likely to be due in part to the different possible electrostatic interactions between the two domains, giving rise to multiple conformers of similar energy. In addition, the recognition helix of KorA, helix 4, is longer than in classical helix-turn-helix proteins and so the CTD is held away from the operator DNA and it has no interactions with the DBDs, making it flexible even in the presence of DNA.

The consensus operator sequence of the protein covers 12 bp (G5-C16 in our sequence) and is found in the three tightest binding KorA operators (2). Three of the remaining four operators do not have Guanine at position G5 while the last has C8 instead of T8, and these four operators all bind more weakly. In our structures, as in PDB ID: 2W7N, determined previously (8), the base pairs outside the consensus sequence do not form any base contacts to the protein, despite the preference of AT base pairs here in the natural operators giving increased binding *in vitro*. In each operator half site, five of the six consensus base pairs form direct hydrogen bonds to a protein side chain. The bases of T6-A15 form no direct bonds to the protein but the phosphates 6 and 7 form a number of direct and water-mediated interactions, which likely restrict the conformation. Our structure shows essentially the same contacts as in König *et al*; how-



**Figure 4.** NMR data of KorA bound to a 20-bp oligonucleotide containing the O<sub>A</sub> consensus sequence. (A) <sup>15</sup>N-<sup>1</sup>H HSQC of KorA-O<sub>A</sub> at 308K recorded at 600 MHz. The peaks are labelled with the amino acid assignments of their NH group, the two peaks at ~10 ppm are from the indole NH groups of the two tryptophan residues. (B) Top: schematic secondary structure of KorA, the boxes show the α-helices, which are numbered and the arrow shows the β-strand. HTH indicates the position of the helix-turn-helix motif. Centre: combined <sup>15</sup>N and <sup>1</sup>H chemical shift differences between free protein and DNA-bound amide resonances of KorA. Bottom: ratio of peak intensities of <sup>15</sup>N-<sup>1</sup>H HSQC spectra of the KorA-O<sub>A</sub> complex taken without and with proton saturation of 2.9 s, giving heteronuclear <sup>15</sup>N-<sup>1</sup>H NOE.

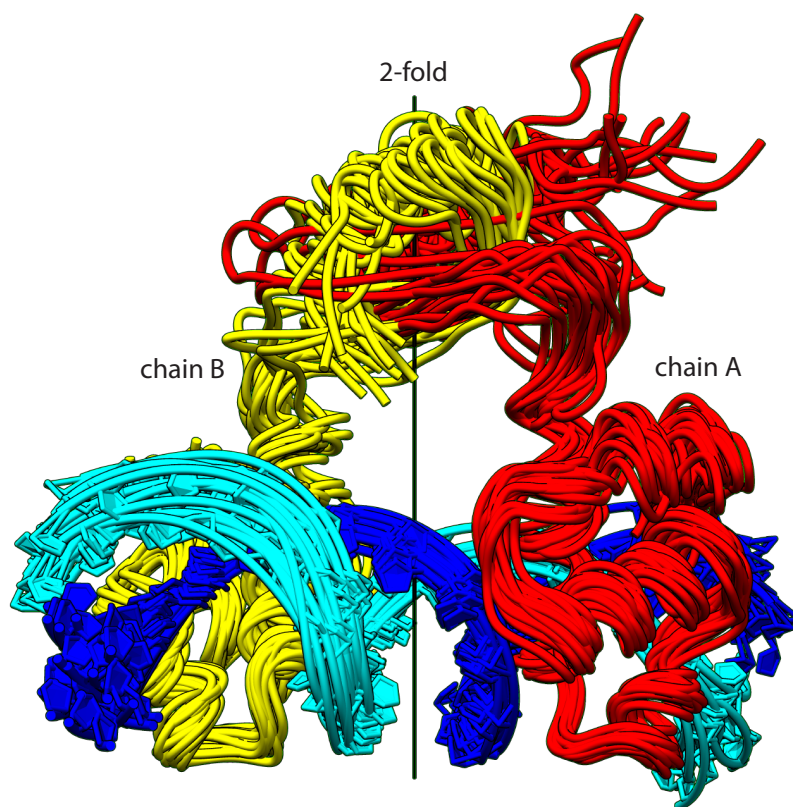
ever, in 2W7N the DNA is fitted in two orientations each with 0.5 occupancy, giving averaged protein-DNA contacts while in 5CM3 and 5CLV minor differences are observed in the DNA-protein contacts between the two half-sites within the palindromic operator, such as for the side chains of R48 and R57, again suggesting a dynamic interface.

While examining the structure of the protein, we observed a structural similarity between the CTD of KorA and that of p53 (PDB ID: 1AIE (43), Figure 6), that was not noted previously. There is no sequence similarity between the two proteins. However, in both proteins this 31-residue domain consists of a nine-residue β-strand, a three-residue turn and a 19-residue α-helix. Like p53, which has long flexible linkers, and can be considered as a dimer of dimers (44), the KorA dimer binds across the DNA duplex rather than to one face of the DNA. The mobility between DBD pairs may be necessary to allow KorA and p53 to bind in this mode. Most proteins that straddle DNA are either unstructured in the absence of DNA (such as the basic region of GCN4 (45)) or only form dimers when bound to the DNA (such as Fos and Jun (46)).

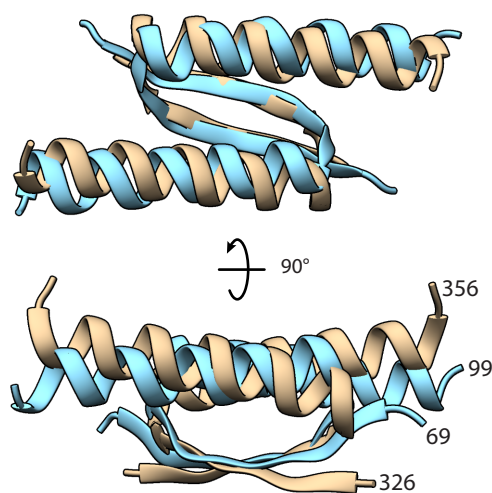
The KorA and p53 CTD surfaces differ in electrostatic properties and shape (Supplementary Figure S1). While p53 CTD forms homotetramers through a four-helix bundle, KorA CTD forms dimers. However, the CTD of KorA is involved in interactions with the co-regulator KorB, as shown by mutational analysis and NMR spectroscopy (9). The mobility of KorA CTD while bound to DNA allows it to flex considerably towards KorB. Intriguingly, KorB, is also a highly flexible protein, with two regions of intrinsic disorder, as well as two structured domains (47). The flexibility within KorA and KorB, even when bound to DNA, allows them to contact each other at different distances and relative orientations (7). Similar cooperative interactions at variable distances are a feature of many eukaryotic transcription factors which also often contain intrinsically disordered regions (reviewed in (48)), such as the transcription activator domains of yeast Gal4 and HMGA family proteins such as SOX15 that bind to the minor groove of AT-rich DNA; as well as p53, and Fos and Jun, mentioned above.

The sequence of the linker (residues 66–69) in KorA is NLPE. Sequences of 92 KorA proteins, with sequence identities >37% were aligned using BLAST (49). The alignment of residues 45–79, covering the DNA recognition helix, the linker and the beta strand shows high sequence conservation of amino acids at the beginning of the DNA recognition helix of the protein to residue 59 and a preference for A at position 61 (Supplementary Figure S5). There is also high sequence conservation from residue 70–89 in the C-terminal domain. There is much less sequence conservation of residues 62–66, that do not contact the DNA but make this helix longer than in a classical helix-turn-helix motif (50), while in the linker, P68 is conserved across more than 97% of KorA sequences, and there is a preference for V or L at position 67. This suggests that the length of the recognition helix, but not the sequence of the last turn, together with a short linker centred at P68, may be important for KorA function.

Intrinsically disordered regions of 20 or more amino acids are increasingly recognized to be important in proteins that are involved in interactions with multiple different partners, for example in cell signalling and regulation of transcription (reviewed in (51)). Several functions have been postulated for these regions, including allowing highly specific interactions with fast on- and off-kinetics (52). KorA contains only a short, four amino acid, linker that allows extensive flexibility; giving similar entropic advantages to in-



**Figure 5.** Molecular dynamics (MD) data. Overlay of 20 structures of the KorA–O<sub>A</sub> complex, sampled over 100 ns of the MD calculation. The protein backbone is coloured red and yellow for subunits A and B; with the DNA strands in blue and cyan.



**Figure 6.** Comparison of CTDs of KorA (blue) and dimerization/tetramerization domain of p53 (tan, PDB ID: 1AIE (43)).

trinsically disordered proteins. A highly flexible four amino acid linker has also been seen in the partitioning protein ParB from the plasmid P1, which allows the helix-turn helix domains to rotate essentially independently of the dimerization domains in different crystal structures (53). This allows the domains to contact several different arrangements

of DNA binding sites, required for partitioning. Since proteins with such highly flexible linkers are hard to crystallize, they are likely to be under-reported. In addition servers predicting disorder use a large window size and hence will overlook short linkers. This study shows that even a four amino acid linker can be highly dynamic and indicates that flexibility and dynamics are as crucial as 3D structure in protein function.

#### ACCESSION NUMBERS

The coordinates and diffraction data have been deposited with the Protein Data Bank: crystal structures of KorA 5CKT, KorA–O<sub>A</sub> complex 1 5CM3, KorA–O<sub>A</sub> complex 2 5CLV. The coordinates and NMR constraints for KorA have been deposited with the relevant databases with the PDB IDs: 2N5G and BMRB 6999.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We thank the ESRF for travel and access to synchrotron X-ray facilities for crystallographic data collection; NMR studies were done at the Wellcome Trust and HEFCE-funded HWB-NMR Centre at the University of Birmingham and at the MRC NIMR, London; SAXS experiments



were done at DESY, Hamburg and at ESRF, Grenoble. We thank the staff at these centres for their help and access to facilities. We thank R.A.Parslow for help with purification of KorA. S.A.H. would like to thank the UK National Grid Service and the Leeds University ARC1 facility for super-computer time.

## FUNDING

Darwin Trust PhD Studentship, University of Edinburgh to KVR. Funding for open access charge: University of Birmingham Charity Fund.

*Conflict of interest statement.* None declared.

## REFERENCES

- Bingle, L.E. and Thomas, C.M. (2001) Regulatory circuits for plasmid survival. *Curr. Opin. Microbiol.*, **4**, 194–200.
- Jadura-Burdzy, G. and Thomas, C.M. (1995) Purification of KorA protein from broad host range plasmid RK2: definition of a hierarchy of KorA operators. *J. Mol. Biol.*, **253**, 39–50.
- Kostelidou, K., Jones, A.C. and Thomas, C.M. (1999) Conserved C-terminal region of global repressor KorA of broad-host-range plasmid RK2 is required for co-operativity between KorA and a second RK2 global regulator, KorB. *J. Mol. Biol.*, **289**, 211–221.
- Zatyka, M., Bingle, L., Jones, A.C. and Thomas, C.M. (2001) Cooperativity between KorB and TrbA repressors of broad-host-range plasmid RK2. *J. Bacteriol.*, **183**, 1022–1031.
- Jadura-Burdzy, G. and Thomas, C.M. (1994) KorA protein of promiscuous plasmid RK2 controls a transcriptional switch between divergent operons for plasmid replication and conjugative transfer. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 10571–10575.
- Jadura-Burdzy, G. and Thomas, C.M. (1997) Dissection of the switch between genes for replication and transfer of promiscuous plasmid RK2: basis of the dominance of trfAp over trbAp and specificity for KorA in controlling the switch. *J. Mol. Biol.*, **265**, 507–518.
- Bingle, L.E., Macartney, D.P., Fantozzi, A., Manzoor, S.E. and Thomas, C.M. (2005) Flexibility in repression and cooperativity by KorB of broad host range IncP-I plasmid RK2. *J. Mol. Biol.*, **349**, 302–316.
- Konig, B., Muller, J.J., Lanka, E. and Heinemann, U. (2009) Crystal structure of KorA bound to operator DNA: insight into repressor cooperation in RP4 gene regulation. *Nucleic Acids Res.*, **37**, 1915–1924.
- Bingle, L.E., Rajasekar, K.V., Muntaha, S., Nadella, V., Hyde, E.I. and Thomas, C.M. (2008) A single aromatic residue in transcriptional repressor protein KorA is critical for cooperativity with its co-regulator KorB. *Mol. Microbiol.*, **70**, 1502–1514.
- Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
- Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 849–861.
- McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 32–41.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
- Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
- Mulder, F.A., Schipper, D., Bott, R. and Boelens, R. (1999) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol.*, **292**, 111–123.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
- Dosset, P., Hus, J.C., Blackledge, M. and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, **16**, 23–28.
- Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
- Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M. and Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.
- Theobald, D.L. and Wuttke, D.S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22**, 2171–2172.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.
- Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P. and Fieschi, F. (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.*, **169**, 45–53.
- Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.
- Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D., Konarev, P.V. and Svergun, D.I. (2012) New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342–350.
- Semenyuk, A.V. and Svergun, D.I. (1991) GNOM—a program package for small-angle scattering data processing. *J. Appl. Crystallogr.*, **24**, 537–540.
- Svergun, D.I., Petoukhov, M.V. and Koch, M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946–2953.
- Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **36**, 860–864.
- Wriggers, W. (2010) Using Situs for the integration of multi-resolution structures. *Biophys. Rev.*, **2**, 21–27.
- Petoukhov, M.V. and Svergun, D.I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, **89**, 1237–1250.
- Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
- Cheatham, T.E. 3rd, Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.

38. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E. 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
39. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
40. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
41. Rajasekar, K.V., Bingle, L.E., Thomas, C.M. and Hyde, E.I. (2006) <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N assignments of the KorA global transcriptional repressor protein from the low copy number IncP-1 plasmid, RK2. *J. Biomol. NMR*, **36**(Suppl. 1), 71.
42. Receveur-Brechot, V. and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Science*, **13**, 55–75.
43. Mittl, P.R., Chene, P. and Grutter, M.G. (1998) Crystallization and structure solution of p53 (residues 326–356) by molecular replacement using an NMR model as template. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 86–89.
44. Tidow, H., Melero, R., Mylonas, E., Freund, S.M., Grossmann, J.G., Carazo, J.M., Svergun, D.I., Valle, M. and Fersht, A.R. (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 12324–12329.
45. Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell*, **71**, 1223–1237.
46. Glover, J.N. and Harrison, S.C. (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature*, **373**, 257–261.
47. Rajasekar, K., Muntaha, S.T., Tame, J.R., Kommareddy, S., Morris, G., Wharton, C.W., Thomas, C.M., White, S.A., Hyde, E.I. and Scott, D.J. (2010) Order and disorder in the domain organization of the plasmid partition protein KorB. *J. Biol. Chem.*, **285**, 15440–15449.
48. Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N. and Dunker, A.K. (2006) Intrinsic disorder in transcription factors. *Biochemistry*, **45**, 6873–6888.
49. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
50. Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.
51. Tantos, A., Han, K.H. and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell. Endocrinol.*, **348**, 457–465.
52. Uversky, V.N. and Dunker, A.K. (2010) Understanding protein non-folding. *Biochim. Biophys. Acta*, **1804**, 1231–1264.
53. Schumacher, M.A. and Funnell, B.E. (2005) Structures of ParB bound to DNA reveal mechanism of partition complex formation. *Nature*, **438**, 516–519.