A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase

Robert T Clubb[1], James G Omichinski[1], Harri Savilahti[2], Kiyoshi Mizuuchi[2], Angela M Gronenborn[1]* and G Marius Clore[1]*

[1]Laboratory of Chemical Physics and [2]Laboratory of Molecular Biology, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA

Background: Mu transposase (MuA) is a multidomain protein encoded by the bacteriophage Mu genome. It is responsible for translocation of the Mu genome, which is the largest and most efficient transposon known. While the various domains of MuA have been delineated by means of biochemical methods, no data have been obtained to date relating to its tertiary structure.

Results: We have solved the three-dimensional solution structure of the DNA-binding domain (residues 1–76; MuA^{76}) of MuA by multidimensional heteronuclear NMR spectroscopy. The structure consists of a three-membered α-helical bundle buttressed by a three-stranded antiparallel β-sheet. Helices H1 and H2 and the seven-residue turn connecting them comprise a helix-turn-helix (HTH) motif. In addition, there is a long nine-residue flexible loop or wing connecting strands B2 and B3 of the sheet. NMR studies of MuA^{76} complexed with a consensus DNA site from the internal activation region of the Mu genome indicate that the wing and the second helix of the HTH motif are significantly perturbed upon DNA binding.

Conclusions: While the general appearance of the DNA-binding domain of MuA^{76} is similar to that of other winged HTH proteins, the connectivity of the secondary structure elements is permuted. Hence, the fold of MuA^{76} represents a novel class of winged HTH DNA-binding domain.

Structure 15 November 1994, **2**:1041–1048

Key words: DNA-binding domain, Mu transposase, NMR spectroscopy, winged helix-turn-helix motif

Introduction

Transposons are mobile genetic elements capable of translocation from one site on the DNA to another. Transposition, the recombination reaction utilized to accomplish this task, is also the mechanism employed to integrate cDNA copies of retroviral and retrotransposon RNA into the chromosomal DNA of their host. Typically, transposition involves a single protein which performs two chemical steps: cleavage of the ends of the transposon DNA, followed by strand transfer which leads to the covalent linkage of the donor DNA ends to its target host DNA site. The largest and most efficient transposon known is the bacteriophage Mu genome, which uses the phage-encoded transposase (MuA protein) to pair the ends of the phage DNA, cleave the termini, and promote strand transfer [1,2].

Mu transposase is a monomeric 75 kDa protein which functions as a tetramer during transposition. Assembly into the active tetramer requires multiple sequence elements on the donor DNA, including sites located at the ends of the genome and an enhancer-like element within the Mu genome [3–5]. Mu transposase can be divided into three structurally distinct domains, each with specific functions. The amino-terminal domain (30 kDa) is responsible for sequence-specific DNA binding and can further be subdivided into two separate subdomains, comprising residues 1–76 and 77–247 which bind an internal activation sequence and the ends of the phage genome, respectively [6,7]. In this paper we present the first structure determination of any domain related to the Mu transposase/integrase family of proteins, namely the solution structure of the amino-terminal DNA-binding domain (MuA^{76}, residues 1–76) of the Mu transposase by means of multidimensional heteronuclear NMR spectroscopy. MuA^{76} possesses a novel topology and represents a new class of winged helix-turn-helix (HTH) DNA-binding domain. In addition, we show that the flexible loop comprising the wing and the second helix of the HTH motif are significantly perturbed upon binding to a consensus DNA site derived from the internal activation sequence of the Mu genome.

Results

Structure determination

The solution structure determination was carried out using ^{15}N and $^{15}N/^{13}C$ labeled samples and involved the application of double and triple resonance three-dimensional (3D) and 4D NMR spectroscopy [8–10]. Examples of the quality of the NMR data are provided by planes of the 4D $^{15}N/^{13}C$ and $^{13}C/^{13}C$-edited nuclear Overhauser enhancement (NOE) spectra shown in Fig. 1. An iterative strategy was employed for the structure determination [11,12]. The final structure

*Corresponding authors.

calculations were based on 1192 approximate interproton distance restraints derived from the 3D and 4D heteronuclear-edited NOE spectra, supplemented by 18 distance restraints for 9 backbone hydrogen bonds (which were only introduced in the final stages of the refinement), 74 torsion angle restraints (40 ϕ, 23 χ_1 and 11 χ_2), 36 $^3J_{HN\alpha}$ coupling constant restraints, and stereospecific assignments for 17 of the 49 β-methylene groups, and for the methyl groups of four of the five valine residues and five of the nine leucine residues. Stereospecific assignments, and ϕ and χ_1 torsion angle restraints were obtained by means of a conformational grid search on the basis of $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants and intra-residue and sequential inter-residue NOEs [13]. Information from $^3J_{HN\beta}$ and $^3J_{COH\beta}$ was also employed for identifying the appropriate χ_1 rotamer and for detecting rotamer averaging, and in addition, $^3J_{C\gamma N}$ and $^3J_{C\gamma CO}$ coupling constants were used for the determination of the side chain conformation of threonine and valine residues. The 11 χ_2 torsion angle restraints comprised restraints for five leucine residues (Leu3, Leu10, Leu13, Leu16, Leu53 and Leu62), Ile24, and the aromatic rings of Tyr25, Tyr48, Trp4 and Trp32. The χ_2 torsion angles for leucine and isoleucine residues were obtained from $^3J_{CC}$ coupling constants and the pattern of intra-residue NOEs [14]. The χ_2 angles of the two tyrosine residues were restrained to 90±30°, and the χ_2 angles of Trp4 and Trp32 were restrained to 90±30° and −90±30°, respectively [15,16]. The latter two restraints were only introduced in the final stages of the refinement. The structural statistics and atomic root mean square (rms) differences for the final ensemble of 34 simulated annealing structures, based on a total of 1320 experimental NMR restraints, are summarized in Tables 1 and 2, respectively, and a stereoview of a best fit superposition of the backbone atoms and ordered side chains is shown in Fig. 2. The amino terminus (residues 1 and 2), carboxyl terminus (residues 66–76) and a protruding loop (residues 37–44) are disordered in solution. The remainder of the structure (residues 3–36 and 45–65) is well defined with a precision of 0.36±0.06 Å for the backbone atoms, 0.74±0.05 Å for all atoms and 0.36±0.05 Å for all atoms that do not exhibit conformational disorder.

Description of the structure

The structure of MuA[76] comprises a three-helical bundle buttressed by a three-stranded antiparallel β-sheet with an overall B1-H1-T-H2-B2-W-B3-H3 topology (where B, H, T and W stand for β-strand, α-helix, turn and wing, respectively) (Figs 2a and 3). The overall fold of the protein is asymmetric and consists of a globular protein core and a protruding loop. Strand B1 (Trp4–Ser6) leads into the short helix H1 (Pro7–Ala11). A long seven-residue turn from Asn12 to Lys18 connects helix H1 with helix H2 (Thr19–Gln30). The arrangement of the H1-T-H2 segment is similar to the HTH DNA-binding motif used by a variety of proteins, although the turn is significantly longer than the usual four to five residues [17,18]. Strand B2 (Trp32–Arg35) follows the HTH motif and pairs with strand B3 (Ala45–Asn49) which in turn pairs with strand B1. A large partially disordered loop extending from Thr36 to Lys44, which we refer to as a wing, connects strands B2 and B3 and protrudes from the globular core. The remainder of the protein consists of strand B3 and the carboxy-terminal helix H3 (Val55–Gln65) which pack against strand B1 and helix H1 to close off the protein core. The hydrophobic core of the protein comprises Leu3 and Val5 of strand B1; Leu10 of helix H1; Leu13 and Leu16 of the turn between helices H1 and H2; Val23, Ile24, Ala27 of helix H2; Trp32 of strand B2; Tyr48 of strand B3; Leu53 of the turn between strand B3 and helix H3; and Leu61 of helix H3 (Fig. 2b).

Interaction with DNA

To identify the location of the DNA-binding site on MuA[76] we carried out NMR studies on a complex

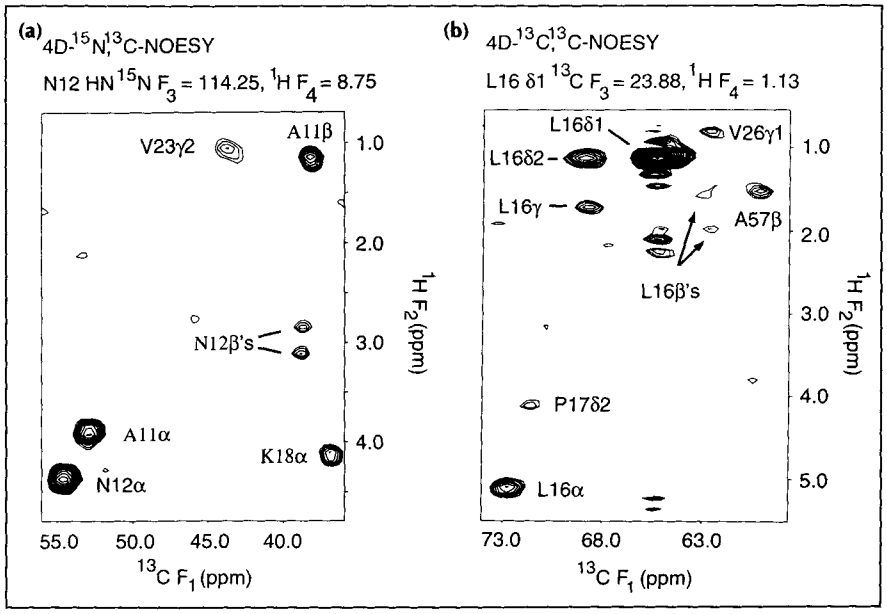

Fig. 1. Example of $^{13}C(F_1)$–$^1H(F_2)$ planes of **(a)** the 4D $^{15}N/^{13}C$-edited NOE spectrum and **(b)** the 4D $^{13}C/^{13}C$-edited NOE spectrum of MuA[76]. The destination proton in (a) is the NH group of Asn12 and in (b) is the $C^{\delta 1}H_3$ methyl group of Leu16. Unlabeled cross peaks have their maximum intensity in an adjacent plane.

Table 1. Structural statistics.

Structural statistics[a]	<SA>	$\overline{(SA)}$r
Rms deviations from experimental distance restraints (Å)[b]		
All (1210)	0.016 ± 0.003	0.012
Inter-residue sequential ($\|i-j\|=1$) (308)	0.017 ± 0.006	0.013
Inter-residue short range ($1<\|i-j\|\leq 5$) (266)	0.015 ± 0.004	0.008
Inter-residue long range ($\|i-j\|>5$) (323)	0.014 ± 0.003	0.013
Intra-residue (295)[b]	0.015 ± 0.006	0.014
H-bond (18)[c]	0.014 ± 0.008	0.001
Rms deviations from $^3J_{NH\alpha}$ coupling constants (Hz) (36)[b]	0.36 ± 0.02	0.33
Rms deviations from experimental dihedral restraints (°) (74)[b]	0.381 ± 0.062	0.374
Deviations from idealized covalent geometry		
Bonds (Å) (1195)	0.004 ± 0.0002	0.003
Angles (°) (2175)	0.568 ± 0.013	0.555
Impropers (°) (608)[d]	0.299 ± 0.051	0.261
E_{L-J}(kcal mol^{-1})[e]	−263 ± 93	−256

[a]The notation of the NMR structures is as follows: <SA> are the final 34 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual $\overline{SA}$ structures best fitted to each other (residues 3–36 and 45–65); $\overline{(SA)}$r is the restrained minimized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. The number of terms for the various restraints is given in parentheses. [b]None of the structures exhibit distance violations greater than 0.4 Å, dihedral angle violations greater than 5°, or $^3H_{NH\alpha}$ coupling constant violations greater than 2 Hz. There are no systematic interproton distance violations between 0.1 Å and 0.4 Å among the ensemble of calculated structures. All the ϕ, ψ backbone torsion angles lie within the allowed regions of the Ramachandran plot. [c]For each backbone hydrogen bond there are two distance restrains: r_{NH-O}, 1.7–2.5 Å; r_{N-O}, 2.3–3.5 Å. These hydrogen bonding restraints were only included in the final stages of refinement. [d]The improper torsion restraints serve to maintain planarity and chirality. [e]E_{L-J} is the Lennard–Jones van der Waals energy calculated with the CHARMM [59] empirical energy function and is not included in the target function for simulated annealing or restrained minimization.

Table 2. Atomic root mean square differences (Å).[a]

	Backbone atoms	Ordered atoms[a]	All atoms
<SA> *vs* $\overline{SA}$	0.36 ± 0.06	0.36 ± 0.05	0.74 ± 0.05
<SA> *vs* $\overline{(SA)}$r	0.42 ± 0.08	0.43 ± 0.07	0.90 ± 0.09
$\overline{(SA)}$r *vs* $\overline{SA}$	0.23	0.24	0.51

[a]The precision of the coordinates is defined as the average root mean square difference between the individual simulated annealing structures and the mean coordinate positions (obtained by averaging the coordinates of the individual structures best-fitted to residues 3–36 and 45–65). The atoms that do not exhibit conformational disorder comprise all N, Cα, C, O and Cβ atoms of residues 3–36 and 45–65; the complete side chains of Leu3, Trp4, Val6, Pro7, Leu10, Leu13, Pro14, Leu16, Pro17, Thr19, Val23, Ile24, Tyr25, Val26, Trp32, Tyr48, Ser52, Leu53, Pro54, Val55, Leu62; the side chains of Glu9, Lys28, Gln30, Asn34, Glu47, Asn49, Glu56, Leu61 and Gln65 up to Cγ, and the side chains of Lys8 and Lys18 up to Cδ.

with a consensus DNA sequence derived from the operator sequences within the internal activation region of the Mu genome. The DNA sequence used was 5′-d(TAGCTTTTTAGTAA)·5′-d(TTACTAAAAAGCTA) which contains the consensus sequence PuCTTTTPyA (where Pu and Py denote purine and pyrimidine nucleotides, respectively) derived from footprinting data [3–5]. ^1H–^{15}N correlation spectra of free and bound MuA^{76} are shown in Fig. 4. The exchange characteristics of the complex are intermediate on the chemical shift scale, resulting in broad lines. This is not surprising as specific binding of the intact MuA transposase to the operator DNA sequence that constitutes the binding site of MuA^{76} is relatively weak (K_{ass} ~10^{-7} M^{-1}), as judged from a qualitative assessment of footprinting data (KM, unpublished data). Although it was possible to assign the majority of backbone ^{15}N and NH resonances in the complex, the quality of the spectra was not sufficient to permit a full structure determination. As chemical shifts are extremely sensitive to any conformational or electronic influences, the analysis of the perturbation of ^{15}N and NH chemical shifts upon complex formation provides a highly sensitive tool for mapping a ligand-binding site on the surface of a protein [19–21]. The observed chemical shift differences upon formation of the MuA^{76}–DNA complex are both highly selective and localized, indicating that the overall conformation of the protein remains essentially unchanged. In the case of the NH resonances, only one residue displays a shift >0.25 ppm, namely Lys18 just before the start of helix H2 is shifted by 0.56 ppm. In the case of the ^{15}N resonances, however, there are a number of large (≥1 ppm) shifts, namely, Ser6, Thr19, Arg37, Ala38, Lys41 and Lys44. In addition, there are a number of residues whose NH and ^{15}N resonances are broadened beyond detection (Ala21, Gly22, Val23, Ile24, Lys28, Val40, Ala45, Ile46 and Glu47). Thus, the largest perturbations upon complex formation are observed at the end of strand B1, in helix H2, in the wing and at the beginning of strand B3. An estimate of the lifetime of the complex can be obtained by noting that the cross peak in the ^1H–^{15}N correlation spectrum with the largest observable frequency displacement, namely Lys18 (340 Hz in the ^1H dimension), is also severely line broadened. This yields a value in the 0.25–0.5 ms range.

Comparison with other helix-turn-helix DNA-binding proteins

The structures of a number of HTH DNA-binding proteins have now been solved [17,18]. They are all characterized by the presence of three or more helices, two of which constitute the HTH motif, and can be divided into two families: those in which the third helix directly follows or precedes the HTH motif without any intervening secondary structure elements (examples include the 434 repressor, lambda repressor, trp repressor, cro, the homeodomain, the POU domain); and those in which the third helix is separated from the HTH motif by a β-sheet. The latter family can be further subdivided into two classes depending on the presence of either a single strand or a β-hairpin between the third helix and the HTH motif (Fig. 5). The first class, which includes the catabolite gene activator protein (CAP) [22] and heat shock transcription factor [23] have a H1-B1-B2-**H2-T-H3**-B3-**W**-B4 topology, while the second class, which includes histone H5 [24], the biotin repressor [25] and the hepatocyte nuclear factor HNF-3/fork head protein [26,27] displays a

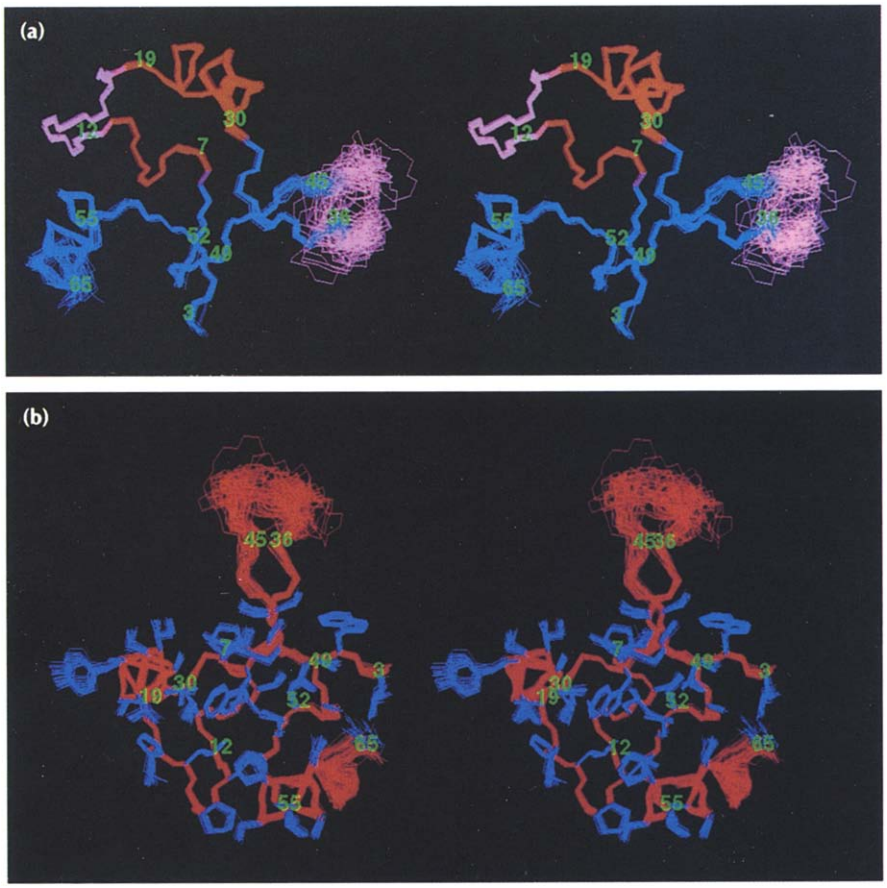

Fig. 2. Stereoviews showing the best-fit superposition of **(a)** the backbone and **(b)** the backbone and ordered side chains of the 34 simulated annealing structures of MuA^{76} (residues 3–65). The best fitting was carried out using residues 3–36 and 45–65. The color coding is as follows: in (a), helices 1 and 2, which are part of the helix-turn-helix (HTH) motif, are shown in red, the turn of the HTH motif and the flexible loop or wing (connecting strands B2 and B3) in pink, and the remainder in blue; in (b), the backbone is shown in red and the side chains in blue.

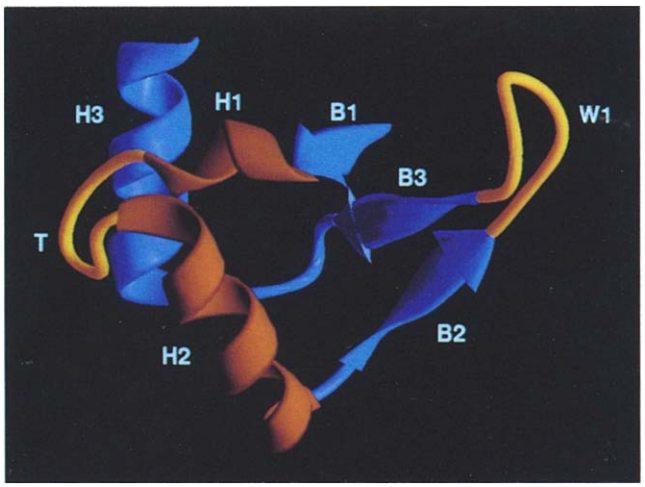

Fig. 3. Ribbon drawing of the average solution structure of MuA^{76}. The helices of the helix-turn-helix (HTH) motif are shown in red, the flexible loop (W1) and turn of the HTH motif (T) in orange, and the remainder of the protein in blue. The ribbon diagram was made with the program RIBBONS [55].

H1-B1-**H2-T-H3**-B2-**W**-B3 topology. In both cases, helices H2 and H3 constitute the two helices of the HTH motif. The angle between the helices of the HTH motif varies from 90° (as in the HNF-3/fork head motif) to 120° (as in CAP). Positioning of the helix H1 with respect to the first helix (H2) of the HTH motif, however, is conserved throughout the two classes with an angle of ~45° between H1 and H2. The topology of MuA^{76} is similar to that of the second class with the major difference being that the third helix has been transplanted from the amino terminus to the carboxyl terminus of the domain (Fig. 5). In this regard, it is worth noting that no ambiguity with regard to the connectivity of the chain can arise in a structure determined by NMR as neighboring amino acids are linked by through-bond correlations along the polypeptide backbone. This contrasts with the situation in X-ray crystallography where incorrect chain tracings in the early stages of refinement may present a major source of error [28]. Whereas the angle (110°) between the helices within the HTH motif of MuA^{76} lies within the canonical range, the orientation of the third helix (H3) is quite different from that in the other two classes. In particular, helix H3 is orthogonal to helix H1 of the HTH motif and seals the hydrophobic core by interacting with the turn of the HTH motif and strand B1. Thus, MuA^{76} constitutes the first member of a new class within this family of HTH proteins. As the DNA-binding domains of the repressor from phage Mu and the transposase and repressor from phage D108 share significant sequence identity (20–40%) with MuA^{76} [29,30], it is likely that these protein domains will also be members of this new class of HTH DNA-binding proteins. The sequence homology between the Mu transposase and repressor proteins is particularly high at positions shown to interact with DNA, as would be expected since these two proteins bind to the same DNA sites [31].

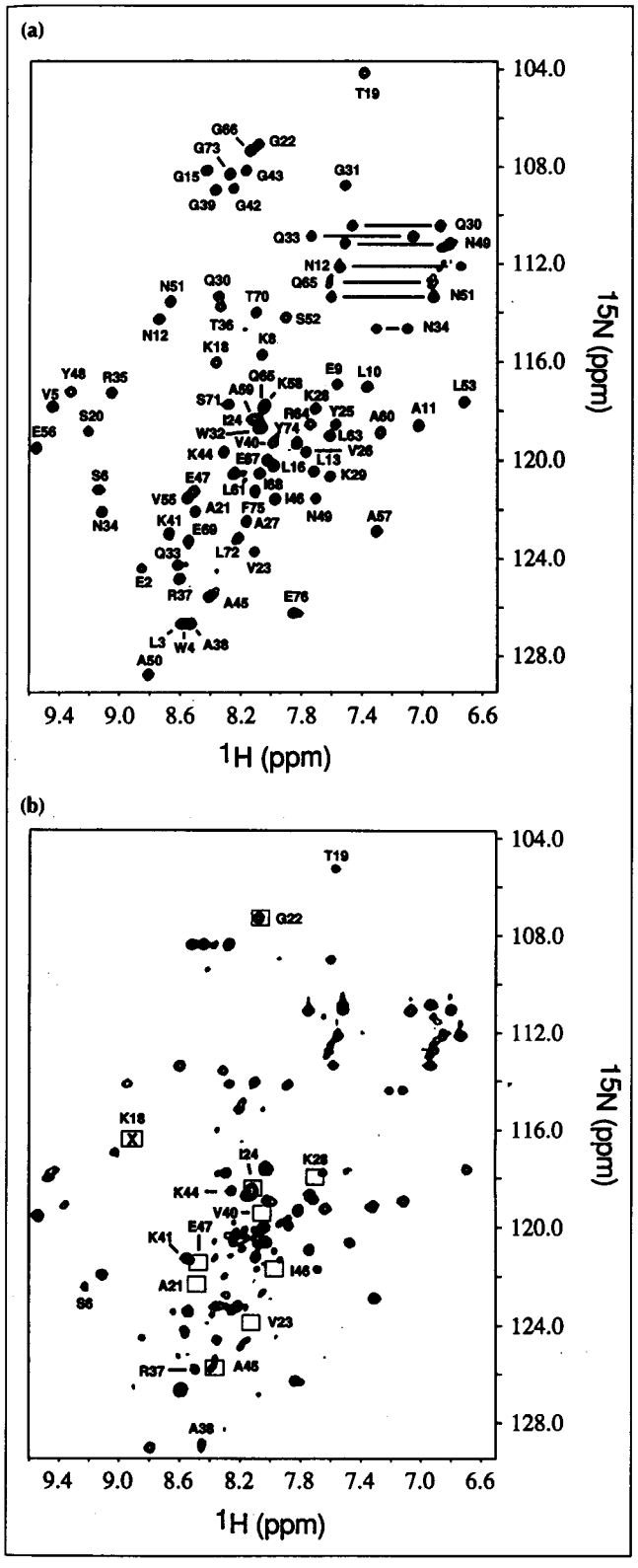

Fig. 4. ^{1}H–^{15}N correlation spectra of **(a)** free and **(b)** DNA-bound MuA^{76}. In (b) only the cross peaks whose ^{1}H or ^{15}N chemical shifts are perturbed by more than 0.25 ppm or 1.0 ppm, respectively, relative to free MuA^{76} are labeled. In addition, square boxes indicate the positions of those resonances in free MuA^{76} that are broadened beyond detectability in the complex; the square box with a cross in it indicates the position of the cross peak of Lys18 in the complex which is only visible at a lower contour level owing to severe line broadening.

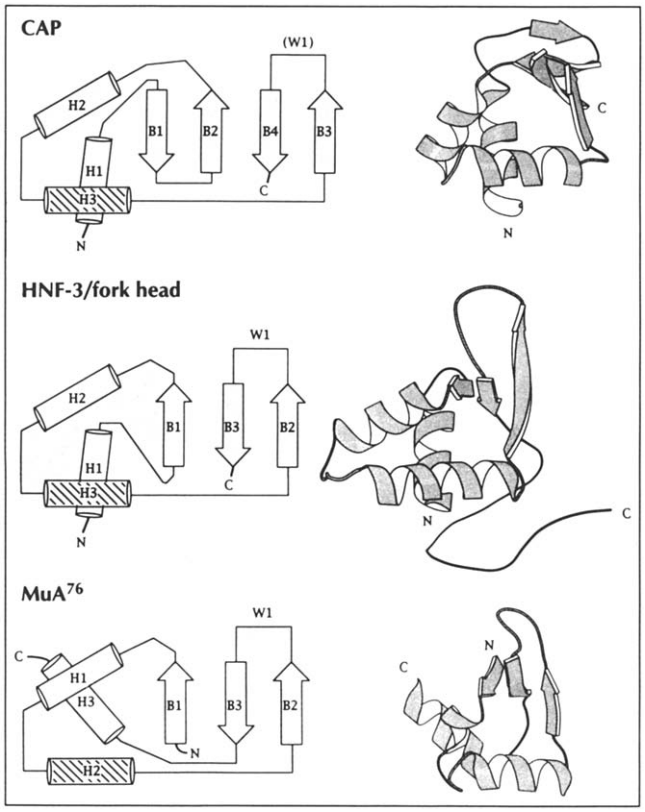

Fig. 5. Schematic (left) and ribbon (right) drawings illustrating the topological differences between the two known classes of α/β-type helix-turn-helix binding domains, typified by the catabolite gene activator protein (CAP) and the hepatocyte nuclear factor (HNF)-3/fork head protein, and MuA^{76}. In the schematic topological diagrams the recognition helix of the HTH motif is hatched. H, B and W stand for helix, strand and wing, respectively. W1, in the case of CAP, is shown in parentheses as the wing is much shorter than in the other two proteins. The ribbon drawings were made with the program MOLSCRIPT [56].

Model of the MuA^{76}–DNA complex

The relative positioning of the HTH motif and the wing between strands B2 and B3 in MuA^{76} is quite similar to the positioning of the HTH motif and analogous wing between strand B2 and B3 of the HNF-3/fork head protein (Fig. 5). In the co-crystal structure, the HNF-3/fork head domain recognizes DNA with the HTH motif and two flexible wings, located in the loop between strands B2 and B3 (W1) and at the carboxyl terminus (W2) [26]. As the NMR studies of the DNA complex with MuA^{76} indicate that the analogous regions of MuA^{76} (excluding the carboxyl terminus as no second wing is present) contact the DNA, we constructed a model of the complex based on the coordinates of the HNF-3/fork head protein–DNA complex (see Fig. 6). The regions in red indicate the sites of substantial perturbations in the ^{1}H–^{15}N correlation spectrum upon complex formation. The recognition helix H2 fits into the major groove, and the wing between strands B2 and B3 contacts the DNA. Lys8, Lys18, Thr19, Lys28, Lys29, Asn34, Thr36 and Lys44 may form hydrogen bonds or electrostatic contacts with the DNA, and other potential contacts may be made by the side chains of Ala21, Ile24, Tyr25 and Ile46.

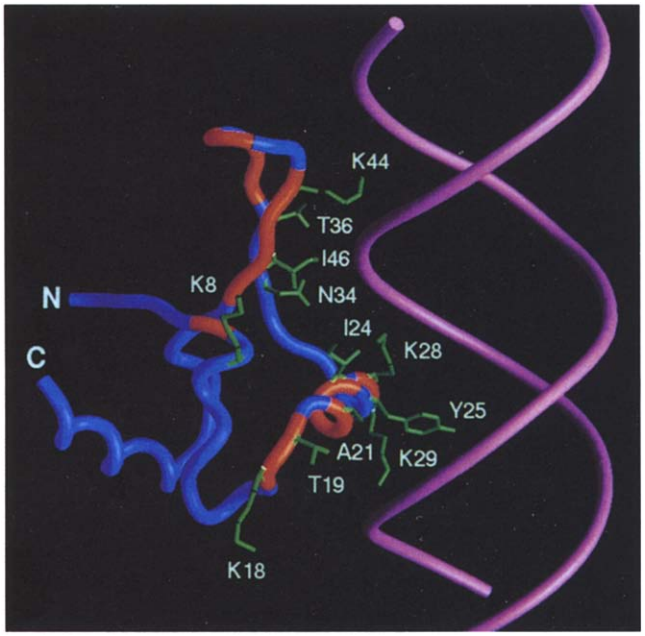

Fig. 6. Model of MuA^{76} bound to DNA that is consistent with NMR data on the complex. The model was generated by superposition of the HTH motif of MuA^{76} onto the analogous region of the HNF-3/fork head protein using the program O [57]. Since the conformation of the DNA in the MuA^{76} complex is unknown, standard B-DNA was used. Regions of the backbone of MuA^{76} that show significant perturbation upon binding DNA (as indicated by resonance differences) are colored red. Side chains close to the DNA are shown in green. The image was generated with the program GRASP [58].

Biological implications

Transposition is the recombination reaction whereby mobile genetic elements known as transposons are moved from one location to another [1]. The reactions involved in transposition, i.e. recognition and cleavage of transposon termini, strand transfer, and ligation of the transposon DNA at its new site, are reasonably well understood. To date, the structures of two proteins involved in site-specific recombination have been solved: namely a crystal structure of the complex of the DNA-binding domain of Hin recombinase with its target DNA site [32], and a crystal structure [33] and an NMR [34] structure of the catalytic and DNA-binding domains, respectively, of γδ resolvase. Both these proteins belong to the Tn3 family of transposons and comprise an amino-terminal catalytic domain and a carboxy-terminal DNA-binding domain.

Here we present the solution structure of the DNA-binding (amino-terminal) fragment of transposase from phage Mu, the first structure from a representative of the Mu family of transposons. Both the location of the DNA-binding domain and its overall topology are quite different from the Tn3 transposon proteins, Hin recombinase and γδ resolvase. The Mu transposase DNA-binding domain (residues 1–76; MuA^{76}) recognizes an enhancer-like sequence within the Mu genome and contains a helix-turn-helix (HTH) motif and a long flexible loop or wing, both of which are involved in DNA recognition. The HTH of MuA^{76} deviates a little from the canonical HTH since its turn comprises seven residues which is longer than the usual four or five. Moreover, it bears little sequence similarity to that of other HTH proteins explaining why the presence of an HTH motif was not correctly predicted from the primary sequence [29,30,35]. While the overall appearance of MuA^{76} is similar to that of a number of other winged HTH proteins, the connectivity of the secondary structure elements is permuted. Hence, the fold of MuA^{76} represents a new class of winged HTH protein.

Materials and methods

Protein expression and purification

The DNA coding sequence for amino acids 1–76 of Mu transposase [32] with a Cys→Leu substitution at position 10 was generated as a *NdeI*/*Bam*HI DNA fragment using the polymerase chain reaction [36], cloned into the *Escherichia coli* vector pET3C and expressed in the host strain BL21(DE3) [37]. The transposition activity of the complete Mu transposase containing the Cys10→Leu mutation is essentially identical to that of wild-type transposase but ensures that intermolecular cross-linking cannot occur (KM, unpublished data). In addition, the ^{1}H-NMR spectrum of MuA^{76}(Cys10→Leu) is virtually indistinguishable from that of wild-type MuA^{76} indicating that this mutation does not perturb the structure of the DNA-binding domain (JGO, GMC and AMG, unpublished data). MuA^{76} was uniformly labeled with either ^{15}N or with ^{15}N and ^{13}C by growing the bacteria in minimal media with $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively. Cells were grown overnight at 37°C, diluted by a quarter with salts, and then induced for 3 h at 37°C with 0.5 mM isopropyl-β-D-thiogalactoside. The cells were harvested, resuspended in 100 mM Tris buffer, pH 7.2, 5mM EDTA, 5 mM benzamidine and 1 mM dithiothreitol (DTT), lysed by passage through a French press and centrifuged at 20 000 *g* for 30 min. The supernatant was removed, centrifuged at 100 000 *g* for 1 h, and then applied to a DEAE-Sepharose Fast Flow (Pharmacia) column (200 ml bed volume) equilibrated with Buffer A (100 mM Tris, pH 7.2, 5 mM EDTA and 1 mM DTT). Fractions containing MuA^{76} were pooled and diluted to give a final buffer concentration of 50 mM Tris, pH 7.2, 1 M urea, 5mM EDTA and 1 mM DTT. The sample was then applied to a S-Sepharose Fast Flow (Pharmacia) column (200 ml bed volume) equilibrated with 50 mM Tris, pH 7.2, 5 mM EDTA and 1 mM DTT (Buffer B). MuA^{76} was eluted with a 0 to 1 M NaCl gradient in Buffer B and the fractions containing the desired product pooled. The pooled product from the S-sepharose column was further purified on a C-8 reversed-phase (Vydac) high performance liquid chromatography (HPLC) column with a 25–45% acetonitrile gradient over 20 min in 0.05% aqueous trifluoroacetic acid. MuA^{76} was taken up in water and the final pH slowly adjusted to 5.8 with NaOH. NaCl was then added to a final concentration of 250 mM. Three samples were made: (i) uniformly ^{15}N-labeled MuA^{76} in 90% H_2O/10% D_2O; (ii) uniformly ^{15}N/^{13}C-labeled MuA^{76} in 90% H_2O/10% D_2O; and (iii) uniformly ^{15}N/^{13}C-labeled

MuA^{76} peptide in 100% D_2O. All NMR experiments were carried out at 20°C on a 2 mM sample of either ^{15}N or $^{15}N/^{13}C$-labeled MuA^{76} (pH 5.8).

NMR spectroscopy

The sequential assignment of the ^1H, ^{13}C and ^{15}N chemical shifts of MuA^{76} was achieved by means of through-bond heteronuclear correlations along the backbone and side chains using the following 3D experiments: ^{15}N-separated HOHAHA, HNHA, CBCANH, CBCA(CO)NH, HBHA(CO)NH, C(CO)NH, H(CCO)NH, HCCH-COSY and HCCH-TOCSY. (For details of these experiments and original references see [8–10].) Approximate interproton distance restraints between NH protons, between NH and ^{13}C-attached protons and between ^{13}C-attached protons were obtained from 3D ^{15}N-separated (100 ms mixing time) [38,39], 4D $^{15}N/^{13}C$-separated (100 ms mixing time) [40], and 4D $^{13}C/^{13}C$-separated (100 ms mixing time) [41,42] NOE spectra, respectively. Quantitative $^3J_{HN\alpha}$, $^3J_{CC}$, $^3J_{C\gamma N}$ and $^3J_{C\gamma CO}$ couplings were measured from a 3D HNHA spectrum, a 2D long range carbon–carbon correlation spectrum, a 2D ^{13}C- $\{^{15}N\}$-spin-echo difference constant time heteronuclear single quantum coherence (HSQC) spectrum, and a 2D ^{13}C-$\{^{13}CO\}$-spin-echo difference constant time HSQC spectrum. Qualitative $^3J_{\alpha\beta}$, $^3J_{NH\beta}$ and $^3J_{COH\beta}$ couplings were obtained from 3D ^{15}N-separated HOHAHA, HNHB and HN(CO)HB experiments, respectively. (For details of these coupling constant experiments and original references see [43].) All NMR spectra were processed with the NmrPipe software [44] and analyzed with the programs PIPP, CAPP and STAPP [45].

Structure calculations

The interproton distance restraints derived from the 3D and 4D heteronuclear-separated NOE spectra were classified into four ranges, 1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), 1.8–5.0 Å and 1.8–6.0 Å, corresponding to strong, medium, weak, and very weak NOEs, respectively [46,47]. Upper distance limits for distances involving methyl protons and non-stereospecifically assigned methylene protons were corrected appropriately for center averaging [48]. In addition, 0.5 Å was added to the upper limit of distances involving methyl protons to account for the higher apparent intensity of methyl resonances [49,50]. Only structurally useful intra-residue NOEs are included in the intra-residue interproton distance restraints. Thus, NOEs between protons separated by two bonds or between non-stereospecifically assigned protons separated by three bonds are not incorporated in the restraints set.

The $^3J_{HN\alpha}$ coupling constants included directly in the refinement comprised only those that could be measured from the 3D HNHA experiment to an accuracy of 0.5 Hz or better. Thus, couplings associated with resonances that exhibit overlap of their ^{15}N and NH chemical shifts were not included. The minimum ranges employed for the ϕ, χ_1 and χ_2 torsion angle restraints were ±10°, ±20° and ±20°, respectively [51]. The narrow range for some of the ϕ restraints was made possible by the availability of highly accurate $^3J_{HN\alpha}$ coupling constant data. In all cases, the angular standard deviations of the torsion angles for the ensemble of simulated annealing structures were much smaller than the ranges employed for the corresponding torsion angle restraints.

The structures were calculated using the hybrid distance geometry-simulated annealing protocol [52] which makes use of the program X-PLOR (version 3.1) [53]. The target function that is minimized during simulated annealing comprises only quadratic harmonic potential terms for covalent geometry (that is bonds, angles and chirality), square-well quadratic potentials for the experimental distance and torsion angle restraints, a harmonic potential for the $^3J_{HN\alpha}$ coupling constant restraints [54], and a quartic van der Waals repulsion term for the non-bonded contacts. No hydrogen bonding, electrostatic or 6–12 Lennard–Jones empirical potential terms are present in the target function.

Interaction of MuA^{76} with DNA

Studies were carried out on a complex of MuA^{76} with the 14mer oligonucleotide: 5′-d(TAGCTTTTTAGTAA)·5′-d(TTACTAAAAAGCTA) which contains the consensus sequence PuCTTTTPyA where Pu and Py denote purine and pyrimidine nucleotides, respectively [3–5]. Samples of the complex were made by slowly adding the DNA to the protein (which was either ^{15}N or $^{15}N/^{13}C$ labeled) at 200 mM until a 1:1 ratio of DNA to protein was reached. The sample was then concentrated with a Centricon-3 (Amicon) to give a final concentration of 2 mM complex in 250 μM NaCl. The high salt was required to stabilize the protein fold, as at low salt MuA^{76} unfolds even in the presence of DNA. As a result, the strength of the electrostatic interactions will be reduced and this may be a significant contributory factor to the unfavorable exchange regime resulting in line broadening. The following spectra were recorded: 2D ^1H–^{15}N HSQC, 3D CBCACONH and 3D HNCO. This was sufficient to reliably obtain the majority of ^{15}N and NH resonance assignments of complexed MuA^{76}, on the basis of the free MuA^{76} assignments.

The coordinates of the 34 simulated annealing structures, as well as the restrained minimized mean structure, and the complete set of experimental NMR restraints have been deposited in the Brookhaven Protein Data Bank.

Acknowledgements: We thank E Appella for synthesizing the oligonucleotides used in the present study; F Delaglio, DS Garrett, J Huth, PJ Lodi, J Qian and AC Wang for useful discussions; R Tschudin for technical support; and R Ghirlando for performing sedimentation equilibrium centrifugation studies of MuA^{76} and S Burley for providing the coordinates of the HNF-3/fork head protein–DNA complex. This work was supported by a Leukemia Society of America post-doctoral fellowship (to RTC) and the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to GMC, AMG and KM).

References

1. Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61**, 1011–1051.
2. Symonds, N., Toussaint, A., van de Putte, P. & Howe, M.M. (1987). *Phage Mu.* (1st edn.), Cold Spring Harbor Laboratory, New York.
3. Mizuuchi, M., Baker, T.A. & Mizuuchi, K. (1992). Assembly of the active form of the transposase–Mu DNA complex: a critical control point in Mu transposition. *Cell* **70**, 303–311.
4. Baker, T.A. & Mizuuchi, K. (1992). DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev.* **6**, 2221–2132.
5. Surette, M.G. & Chaconas, G. (1991). The Mu transpositional enhancer can function in *trans*: requirement of the enhancer for synapsis but not strand cleavage. *Cell* **68**, 1101–1108.
6. Leung, P.C., Teplow, D.B. & Harshey, R.M. (1989). Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature* **338**, 656–658.
7. Mizuuchi, M. & Mizuuchi, K. (1989). Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. *Cell* **58**, 399–408.
8. Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Accounts Chem. Res.* **26**, 131–138.

9. Clore, G.M. & Gronenborn, A.M. (1991). Application of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Progr. Nucl. Magn. Reson. Spectr.* **23**, 43–92.
10. Clore, G.M. & Gronenborn, A.M. (1994). Multidimensional heteronuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363.
11. Clore, G.M. & Gronenborn, A.M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399.
12. Kraulis, P.J., *et al.*, & Gronenborn, A.M. (1989). Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*: a study using nuclear magnetic resonance and hybrid distance geometry–dynamical simulated annealing. *Biochemistry* **28**, 7241–7257.
13. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1990). ^{1}H-NMR stereospecific assignments by conformational database searches. *Biopolymers* **29**, 813–822.
14. Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. & Clore, G.M. (1993). The high resolution three-dimensional solution structure of human interleukin-4 determined by multi-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* **32**, 6744–6762.
15. Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339.
16. Dunbrack, R.L. & Karplus, M. (1993). Backbone dependent rotamer library for proteins: application to sidechain prediction. *J. Mol. Biol.* **230**, 543–571.
17. Harrison, S.C. & Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* **59**, 933–969.
18. Brennan, R.G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* **74**, 773–776.
19. Chen, Y., Reizer, J., Saier, M.H., Fairbrother, W.J. & Wright, P.E. (1993). Mapping of the binding interfaces of proteins of the bacterial phosphotransferase system HPr and IIA^{glc}. *Biochemistry* **32**, 32–37.
20. van Nuland, N.A.J., Kroon, G.J.A., Dijkstra, K., Wolters, G.K., Scheek, R.M. & Robillard, G.T. (1993). The NMR determination of the IIa^{mtl} binding site on HPr of the *Escherichia coli* phosphoenol pyruvate-dependent phosphotransferase system. *FEBS Lett.* **315**, 11–15.
21. Gronenborn, A.M. & Clore, G.M. (1993). Identification of the contact surface of a Streptococcal G domain complexed with a human Fc fragment. *J. Mol. Biol.* **233**, 331–335.
22. Schultz, S.C., Shields, G.C. & Steitz, T.A. (1991). Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001–1007.
23. Harrison, C.J., Bohm, A.A. & Nelson, H.C.M. (1994). Crystal structure of the DNA binding domain of the heat shock transcriptional factor. *Science* **263**, 224–227.
24. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. & Sweet, R.M. (1993). Crystal structure of the globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219–223.
25. Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, A.J. & Matthews, B.W. (1992). *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. Natl. Acad. Sci. USA* **89**, 9257–9261.
26. Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420.
27. Lai, E., Clark, K.L., Burley, S.K. & Darnell, J.E.,Jr. (1993). Hepatocyte nuclear factor 3/fork head or 'winged helix' proteins: a family of transcriptional factors of diverse biologic function. *Proc. Natl. Acad. Sci. USA* **90**, 10421–10423.
28. Brändén, C.I. & Jones, T.A. (1990). Between objectivity and subjectivity. *Nature* **343**, 687–689.
29. Mizuuchi, M., Weisberg, R.A. & Mizuuchi, K. (1986). DNA sequences of the control region of phage D108: the amino-terminal amino acid sequences of repressor and transposase are similar both in phage D108 and in its relative, phage Mu. *Nucleic Acids Res.* **14**, 3813–3825.
30. Mizuuchi, M. & Mizuuchi, K. (1989). Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implication for regulation. *Cell* **58**, 399–408.
31. Craigie, R., Mizuuchi, M. & Mizuuchi, K. (1984). Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* **39**, 387–394.
32. Feng, J.-A., Johnson, R.C. & Dickerson, R.E. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science* **263**, 348–355.
33. Sanderson, M.R., *et al.*, & Steitz, T.A. (1990). The crystal structure of the catalytic domain of the site-specific recombination enzyme γδ resolvase at 2.7 Å resolution. *Cell* **63**, 1323–1329.
34. Liu, T., DeRose, E.F. & Mullen, G.P. (1994) Determination of the structure of the DNA binding domain of γδ resolvase in solution. *Protein Sci.* **3**, 1286–1295.
35. Harshey, R.M., Getzoff, E.D., Baldwin, D.L., Miller, J.L. & Chaconas, G. (1985). Primary structure of phage Mu transposase: homology to Mu repressor. *Proc. Natl. Acad. Sci. USA* **82**, 7676–7680.
36. Saiki, R.K., *et al.*, & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354.
37. Sudier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorf, J.W. (1990). Gene expression technology. *Methods Enyzmol.* **185**, 60–89.
38. Zuiderweg, E.R.P. & Fesik, S.W. (1989). Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28**, 2387–2391.
39. Marion, D., *et al.*, & Clore, G.M. (1989). Overcoming the overlap in the assignment of ^{1}H-NMR spectra of larger proteins using three-dimensional heteronuclear ^{1}H–^{15}N Hartmann–Hahn and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin-1β. *Biochemistry* **28**, 6150–6156.
40. Kay, L.E., Clore, G.M., Bax, A. & Gronenborn, A.M. (1990). Four-dimensional heteronuclear triple resonance NMR spectroscopy of interleukin-1β in solution. *Science* **249**, 411–414.
41. Clore, G.M., Kay, L.E., Bax, A. & Gronenborn, A.M. (1991). Four dimensional ^{13}C/^{13}C-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin-1β. *Biochemistry* **30**, 12–18.
42. Vuister, G.W., *et al.*, & Bax, A. (1993). Increased resolution and improved spectral quality in four-dimensional ^{13}C/^{13}C-separated HMQC-NOE-HMQC spectra using pulsed field gradients. *J. Magn. Reson. B* **101**, 210–213.
43. Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R. & Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79–105.
44. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, J., Pfeifer, J. & Bax, A. (1994). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. In *Proceedings of the 35th Experimental Nuclear Magnetic Resonance Conference*. Abstract No. WP108, p. 262.
45. Garrett, D.S., Powers, R., Gronenborn, A.M. & Clore, G.M. (1991). A common sense approach to peak picking two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–222.
46. Williamson, M.P., Havel, T.F. & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315.
47. Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1986). The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5**, 2729–2735.
48. Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.
49. Clore, G.M., Gronenborn, A.M., Nilges, M. & Ryan, C.A. (1987). The three-dimensional structure of potato carboxypeptidase inhibitor in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Biochemistry* **26**, 8012–8023.
50. Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
51. Lodi, P.J., Garrett, D.S., Kuszewski, J., Tsang, M.L.S., Weatherbee, J.A., Leonard, W.J., Gronenborn, A.M. & Clore, G.M. (1994). High resolution solution structure of the β chemokine hMIP-1β by multi-dimensional NMR. *Science* **263**, 1762–1767.
52. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry–dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324.
53. Brünger, A.T. (1993). *X-PLOR Version 3.1 Manual.* Yale University, New Haven, CT.
54. Garrett, D.S., *et al.*, & Clore, G.M. (1994). The impact of direct refinement against three-bond HN–CαH coupling constants on protein structure determination by NMR. *J. Magn. Reson. B* **104**, 99–103.
55. Carson, M. (1987). Ribbon models of macromolecules. *J. Mol. Graphics* **5**, 103–106.
56. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
57. Jones, T.A. & Kjeldgaard, M. (1992). *O — The manual. Version 5.8.1.* University of Uppsala, Sweden.
58. Nicholls, A.J. (1993). *GRASP Manual.* Columbia University, New York.
59. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Received: **15 Aug 1994**; revisions requested: **12 Sep 1994**; revisions received: **19 Sep 1994**. Accepted: **22 Sep 1994**.