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

Robert T Clubb¹, James G Omichinski¹, Harri Savilahti², Kiyoshi Mizuuchi², Angela M Gronenborn¹* and G Marius Clore¹*

¹Laboratory of Chemical Physics and ²Laboratory 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; MuA⁷⁶) 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 MuA⁷⁶ 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 MuA⁷⁶ is similar to that of other winged HTH proteins, the connectivity of the secondary structure elements is permuted. Hence, the fold of MuA⁷⁶ represents a novel class of winged HTH DNA-binding domain.

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 (MuA⁷⁶, residues 1–76) of the Mu transposase by means of multidimensional heteronuclear NMR spectroscopy. MuA⁷⁶ 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 ¹⁵N and ¹⁵N/¹³C 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 ¹⁵N/¹³C and ¹³C/¹³C-edited nuclear Overhauser enhancement (NOE) spectra shown in Fig. 1. An iterative strategy was employed for the structure determination [11,12]. The final structure
Structure of MuA

The structure of MuA comprises a three-helical bundle buttressed by a three-stranded antiparallel β-sheet with a HTH DNA-binding motif used by a variety of proteins, although the turn is significantly longer than the usual four to five residues. 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 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 we carried out NMR studies on a complex of 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 χ₁ and 11 χ₂), 36 3J_HN coupling constant restraints, and stereo-specific 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 χ₁ torsion angle restraints were obtained by means of a conformational grid search on the basis of 3J_HN and 3J_AB coupling constants and intra-residue and sequential inter-residue NOEs [13]. Information from 3J_HNB and 3J_COHB was also used for identifying the appropriate χ₁ rotamer and for detecting rotamer averaging, and in addition, 3J_CY and 3J_CO coupling constants and the aromatic rings of Tyr25, Tyr48, Trp4 and Trp32.

The 3J coupling constant restraints, and stereospecific assignments for 17 of the 49 1H/15N coupling constant restraints were obtained by means of a conformational grid search on the basis of 3JHN and 3J_AB coupling constants and intra-residue and sequential inter-residue NOEs [13]. Information from 3J_HNB and 3J_COHB was also employed for identifying the appropriate χ₁ rotamer and for detecting rotamer averaging, and in addition, 3J_CY and 3J_CO coupling constants and the aromatic rings of Tyr25, Tyr48, Trp4 and Trp32. The χ₁ torsion angles for leucine and isoleucine residues were obtained from 3J(CC) coupling constants and the pattern of intra-residue NOEs [14]. The χ₂ angles of the two tyrosine residues were restrained to 90±30°, and the χ₂ 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 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 comprises a three-helical bundle buttressed by a three-stranded antiparallel β-sheet with an overall B1-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 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 we carried out NMR studies on a complex
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(TAGCTTTTTAGTA)–3', which contains the consensus sequence PuCTTTPyA (where Pu and Py denote purine and pyrimidine nucleotides, respectively) derived from footprinting data [3–5]. 1H–15N correlation spectra of free and bound MuA76 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 MuA76 is relatively weak (K<sub>off</sub> ~10<sup>7</sup> M<sup>–1</sup>), as judged from a qualitative assessment of footprinting data (KM, unpublished data). Although it was possible to assign the majority of backbone 15N 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 15N 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 MuA76–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 15N 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 15N resonances are broadened beyond detection (Ala21, Gly22, Val23, Ile24, Lys28, Val40, Ala45, Ile46 and Ile47). 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–15N 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

### Table 1. Structural statistics.

<table>
<thead>
<tr>
<th>Structural statistics&lt;sup&gt;a&lt;/sup&gt;</th>
<th>&lt;SA&gt;</th>
<th>(πSA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rms deviations from experimental distance restraints (Å)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>All (1210)</td>
<td>0.016 ± 0.003</td>
</tr>
<tr>
<td>Inter-residue sequential (</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Inter-residue short range (</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Inter-residue long range (</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Intra-residue (295p)</td>
<td>0.015 ± 0.006</td>
<td>0.014</td>
</tr>
<tr>
<td>H-bond (18F)</td>
<td>0.014 ± 0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>Rms deviations from (36&lt;sup&gt;b&lt;/sup&gt;) coupling constants (Hz) (1210)</td>
<td>0.36 ± 0.02</td>
<td>0.33</td>
</tr>
<tr>
<td>Rms deviations from experimental dihedral restraints (5) (174p)</td>
<td>0.381 ± 0.062</td>
<td>0.374</td>
</tr>
<tr>
<td>Deviations from idealized covalent geometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å) (1195)</td>
<td>0.004 ± 0.0002</td>
<td>0.003</td>
</tr>
<tr>
<td>Angles (°) (2175)</td>
<td>0.568 ± 0.013</td>
<td>0.555</td>
</tr>
<tr>
<td>Improper pairs (°) (606&lt;sup&gt;d&lt;/sup&gt;)</td>
<td>0.299 ± 0.051</td>
<td>0.261</td>
</tr>
<tr>
<td>δ&lt;sup&gt;e&lt;/sup&gt; ± δ&lt;sup&gt;e&lt;/sup&gt; kcal mol&lt;sup&gt;–1&lt;/sup&gt;</td>
<td>263 ± 93</td>
<td>256</td>
</tr>
</tbody>
</table>

<sup>a</sup> The notation of the NMR structures is as follows: SA are the final 34 simulated annealing structures; SA is the mean structure obtained by averaging the coordinates of the individual SA structures best fitted to each other (residues 3–36 and 45–65); (SA) is the restrained minimal mean structure obtained by restrained regularization of the mean structure SA. The number of terms for the various restraints is given in parentheses. bNone of the structures exhibit distance violations greater than 0.4 Å, dihedral angle violations greater than 5°, or H-bond 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. cFor each backbone hydrogen bond there are two distance restraints: r<sub>N–O</sub> ± 0.006 Å; r<sub>N–O</sub> ± 0.3 Å. These hydrogen bonding restraints were only included in the final stages of refinement. dThe improper torsion restraints serve to maintain backbone hydrogen bond there are two distance restraints: r<sub>NH–O</sub>, 1.7–2.5 Å; r<sub>NH–O</sub> ± 0.006 Å; r<sub>NH–O</sub> ± 0.3 Å. These hydrogen bonding restraints were only included in the final stages of refinement. eThe improper torsion restraints serve to maintain planarity and chirality. fThe δ<sup>e</sup> ± δ<sup>e</sup> kcal mol<sup>–1</sup> is calculated with the CHARMM159 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).<sup>a</sup>

<table>
<thead>
<tr>
<th>Backbone and ordered atoms&lt;sup&gt;a&lt;/sup&gt;</th>
<th>All atoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;SA&gt; vs SA</td>
<td>0.36 ± 0.06</td>
</tr>
<tr>
<td>&lt;SA&gt; vs (SA)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.42 ± 0.08</td>
</tr>
<tr>
<td>(SA) vs SA</td>
<td>0.23</td>
</tr>
</tbody>
</table>

<sup>a</sup> 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, C<sub>z</sub>, C, O and Cβ atoms of residues 3–36 and 45–65; the complete side chains of Leu1, Trp4, Val6, Pro7, Leu10, Leu13, Pro14, Leu16, Pro17, Thr19, Val23, Ile24, Tyr25, Val26, Thr32, Tyr46, Ser52, Leu53, Pro54, Val55, Leu62; the side chains of Glu9, Lys28, Gln30, Asn34, Glu47, Asn49, Gly56, Leu61 and Glu65 up to C7; and the side chains of Lys8 and Lys18 up to C8.
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 MuA76 (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 MuA76. 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 MuA76 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 MuA76 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, MuA76 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 MuA76 [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].
DNA-binding domain of Mu transposase Clubb et al.

(a) DNA-binding domain of Mu transposase

(b) 

Fig. 4. $^1$H-$^1$N correlation spectra of (a) free and (b) DNA-bound MuA$^{76}$. In (b) only the cross peaks whose $^1$H or $^1$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 αβ-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–$^1$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 the MuA$^{76}$–DNA complex.
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 y8 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 y8 resolvase. The Mu transposase DNA-binding domain (residues 1-76; MuA76) 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 MuA76 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 MuA76 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 MuA76 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/BamHI 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 1H-NMR spectrum of MuA76(Cys10→Leu) is virtually indistinguishable from that of wild-type MuA76 indicating that this mutation does not perturb the structure of the DNA-binding domain (JGO, GMC and AMG, unpublished data). MuA76 was uniformly labeled with either 15N or with 15N and 13C by growing the bacteria in minimal media with 15NH4Cl and/or 13C-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, 5 mM 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 MuA76 were pooled and diluted to give a final buffer concentration of 50 mM Tris, pH 7.2, 1 M urea, 5 mM EDTA and 1 mM DTT. The sample was then applied to an 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). MuA76 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. MuA76 was taken up in 0.05% aqueous trifluoroacetic acid. MuA76 was then added to a final concentration of 250 mM. Three samples were made: (i) uniformly 15N-labeled MuA76 in 90% H2O/10% D2O; (ii) uniformly 15N/13C-labeled MuA76 in 90% H2O/10% D2O; and (iii) uniformly 15N/13C-labeled
MuA\textsuperscript{76} peptide in 100% D\textsubscript{2}O. All NMR experiments were carried out at 20°C on a 2 mM sample of either \textsuperscript{15}N or \textsuperscript{13}N/\textsuperscript{13}C-labeled MuA\textsuperscript{76} (pH 5.8).

**NMR spectroscopy**

The sequential assignment of the \textsuperscript{1}H, \textsuperscript{13}C and \textsuperscript{15}N chemical shifts of MuA\textsuperscript{76} was achieved by means of through-bond heteronuclear correlations along the backbone and side chains using the following 3D experiments: \textsuperscript{15}N-separated HOHAHA, HNHA, CBCANH, CBCA(CO)NH, HBBHA(CO)NH, C(CO)NH, H(CCO)NH, HCC-COSY and HCH-TOCSY. (For details of these experiments and original references see [8–10].) Approximate interproton distance restraints between NH protons, between NH and 13C-attached protons and between 13C-attached protons were obtained from 3D \textsuperscript{15}N-separated (100 ms mixing time) [38,39], 4D \textsuperscript{15}N/\textsuperscript{13}C-separated (100 ms mixing time) [40], and 4D \textsuperscript{13}C/\textsuperscript{13}C-separated (100 ms mixing time) [41,42] NOE spectra, respectively. Quantitative \textsuperscript{3}J\textsubscript{HN}, \textsuperscript{3}J\textsubscript{CC} and \textsuperscript{3}J\textsubscript{CN} NOE and \textsuperscript{3}J\textsubscript{CC} NOE couplings were measured from a 3D HNHA spectrum, a 3D long range carbon–carbon correlation spectrum, a 2D \textsuperscript{13}C-\textsuperscript{15}N–(15)N–spin–echo difference constant time heteronuclear single quantum coherence (HSQC) spectrum, and a 2D \textsuperscript{13}C–\textsuperscript{13}C–spin–echo difference constant time HSQC spectrum. Qualitative \textsuperscript{3}J\textsubscript{NH} and \textsuperscript{3}J\textsubscript{COHB} couplings were obtained from 3D \textsuperscript{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 \textsuperscript{3}J\textsubscript{HN} 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 \textsuperscript{15}N and NH chemical shifts were not included. The minimum ranges employed for the $\Phi$, $\chi_1$ and $\chi_2$ torsion angle restraints were $\pm 10^\circ$, $\pm 20^\circ$ and $\pm 20^\circ$, respectively [51]. The narrow range for some of the $\Phi$ restraints was made possible by the availability of highly accurate \textsuperscript{3}J\textsubscript{HN} 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 \textsuperscript{3}J\textsubscript{HN} 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\textsuperscript{76} with DNA**

Studies were carried out on a complex of MuA\textsuperscript{76} with the 14mer oligonucleotide: 5'-d(TAGCTTTTTAGTAA)-5'-d(TTACTAAAAAGCTA) which contains the consensus sequence PuCTTTTTPyA 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 \textsuperscript{15}N or \textsuperscript{15}N/\textsuperscript{13}C labeled) at 200 mM until a 1:1 ratio of DNA to protein was reached. The sample was then concentrated with a Centricron-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\textsuperscript{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 $^2$H–\textsuperscript{15}N HSQC, 3D CBCACONH and 3D HNCO. This was sufficient to reliably obtain the majority of \textsuperscript{15}N and NH resonance assignments of complexed MuA\textsuperscript{76}, on the basis of the free MuA\textsuperscript{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; P Delaglio, DS Garrett, J Huth, P Lodi, J Qian and AC Wang for useful discussions; R Tschudin for technical support; and R Ghirlando for performing the coordinates determination. The authors acknowledge the following grants: National Institutes of Health grants GM35827 and GM34669; Public Health Service grant HD-24331; and the National Science Foundation grant PCM-9001304. The authors wish to thank the members of the Protein Crystallography Laboratory, Whitehead Institute for Biomedical Research, Cambridge, MA, for their technical and administrative support.

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.

**References**


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