The phage Mu transposase (MuA) binds to the ends of the Mu genome during the assembly of higher order nucleoprotein complexes. We investigate the structure and function of the MuA end-binding domain (Iβγ). The three-dimensional solution structure of the Iβ subdomain (residues 77–174) has been determined using multidimensional NMR spectroscopy. It comprises five α-helices, including a helix–turn–helix (HTH) DNA-binding motif formed by helices 3 and 4, and can be subdivided into two interacting structural elements. The structure has an elongated disc-like appearance from which protrudes the recognition helix of the HTH motif. The topology of helices 2–4 is very similar to that of helices 1–3 of the previously determined solution structure of the MuA Iγ subdomain and to that of the homeodomain family of HTH DNA-binding proteins. We show that each of the two subdomains binds to one half of the 22 bp recognition sequence, Iβ to the more conserved Mu end distal half (β subsite) and Iγ to the Mu end proximal half (γ subsite) of the consensus Mu end-binding site. The complete Iβγ domain binds the recognition sequence with a 100- to 1000-fold higher affinity than the two subdomains independently, indicating a cooperative effect. Our results show that the Mu end DNA-binding domain of MuA has a modular organization, with each module acting on a specific part of the 22 bp binding site. Based on the present binding data and the structures of the Iβ and Iγ subdomains, a model for the interaction of the complete Iβγ domain with DNA is proposed.

Keywords: Iβ domain/MuA transposase/NMR/protein–DNA interaction/solution structure

Introduction

DNA transposition is a movement of defined segments of DNA, generally called transposons, to distant locations within a genome (Berg and Howe, 1989). It requires the formation of stable nucleoprotein complexes, known as transpososomes, which contain at a minimum an oligomeric transposase protein and the synapsed two ends of the transposon DNA (Craigie and Mizuuchi, 1987; Surette et al., 1987). Subsequent to the end synapsis, coordinated strand breakage of the two ends of the transposon and the coupled breakage and joining of the transposon ends to the host DNA follow. The ability to reconstitute an in vitro system has made transposase of the temperate phage Mu the best-studied model system for DNA transposition (reviewed by Mizuuchi, 1992; Lavoie and Chaconas, 1995).

The transposase of phage Mu, known as MuA, recognizes three distant DNA regions within the 38 kb Mu genome. It binds the two ends of the Mu genome at which DNA recombination occurs, and a transpositional enhancer (IAS, internal activation site) located ~1 kb from the left end of the Mu genome (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989; Surette et al., 1989). Three MuA-binding sites are located at each end: L1, L2 and L3 at the left end (Mu endL), and R1, R2 and R3 at the right end (Mu endR) (Craigie et al., 1984). They are related by a 22 bp consensus sequence 5’-GTTCAYNNRAARYRGCA-AAR(A/C) that shows no obvious internal symmetry (Craigie et al., 1984).

Purified MuA transposase (75 kDa) is monomeric in solution and binds reversibly to the Mu end sites in the absence of any additional cofactors (Kuo et al., 1991; Zou et al., 1991). Binding results in bending of the Mu end DNA by ~60°–90° (Adzuma and Mizuuchi, 1987; Ding et al., 1993). A schematic representation of the domains of MuA transposase, as determined by partial proteolysis, is shown in Figure 1 (Nakayama et al., 1987). The Iα domain binds to the internal activation site and is a member of the winged helix–turn–helix (HTH) family of DNA-binding proteins, albeit with a permutation of the secondary structure elements (Clubb et al., 1994, 1996). The Iβγ domain (residues 77–247) binds specifically to the sites at the ends of the Mu genome (Leung et al., 1989; Kim and Harshey, 1995). We recently have identified an additional hypersensitive protease site between residues 173 and 174 which yields two independently folded subdomains comprising residues 77–174 (Iβ) and 174–247 (Iγ) (Clubb et al., 1997). The solution structure of Iγ has been determined and shown to contain a HTH motif with a topology similar to that of homeodomains (Clubb et al., 1997). The structure of domain II (catalytic core, residues 247–574) has been solved by X-ray crystallography (Rice and Mizuuchi, 1995) and shows high structural similarity to the catalytic cores of human immunodeficiency virus (HIV) and avian sarcoma virus (ASV) integrases (Dyda et al., 1994; Bujacz et al., 1995), as well as to regions of RNase H and the Holliday junction-resolving enzyme, RuvC (see reviews by Grindley and Leschziner, 1995; Yang and Steitz, 1995). Domain II
Structure of the phage Mu transposase Iβ subdomain

Results and discussion

Structure determination

The solution structure of the Iβ subdomain was determined using double and triple resonance multidimensional NMR spectroscopy (Clore and Gronenborn, 1991, 1994; Bax and Grzesiek, 1993). Examples of the quality of the NMR data are provided by typical planes from the 4D 15N/13C- and 13C/13C-separated nuclear Overhauser enhancement (NOE) spectra shown in Figure 2. The final 30 simulated annealing structures were calculated on the basis of 1446 experimental NMR restraints. A summary of the structural statistics is provided in Table I, and a stereoview showing a best-fit superposition of the ensemble of 30 simulated annealing structures is shown in Figure 3. Residues 89–166 are well defined by the experimental data and have a backbone precision of ~0.5 Å. Although the N-terminus does fold back on the protein, as evidenced by several NOEs between the side chain of Ile77 and the aromatic rings of Trp93 and Trp96, as well as a few medium range non-sequential NOEs involving residues 81–89, the polypeptide chain from residues 76–88 is poorly defined by the NOE data. (Note: this lack of NOE data for the N-terminus is not due to overlap or other assignment problems.) At the C-terminus (residues 167–174), the polypeptide chain appears to be disordered, as evidenced by the absence of any non-sequential NOEs.

Description of the structure of Iβ

Two views of a ribbon diagram of the Iβ subdomain are shown in Figure 4A and B. The Iβ subdomain is elongated in shape, ~45 Å in length and ~18 Å in width, and is composed of five helices (residues 90–98, 101–122, 127–137, 142–151 and 160–165) connected by loops of various lengths. Iβ can be divided into two structural elements. The first element is formed by helices 1 and 5, and the N-terminal portion of helix 2, with interhelical angles of ~120°, ~75° and ~100° between helices 1 and 2, 1 and 5, and 2 and 5 respectively. The second element is formed by helices 3 and 4 and the C-terminal portion of helix 2, with interhelical angles of ~145°, ~85° and ~115° between helices 2 and 3, 2 and 4, and 3 and 4 respectively. Thus, helix 2, which spans the entire length of the Iβ subdomain, serves to bridge the two structural elements. A proline residue at position 112 introduces a small kink of ~15° in the middle of helix 2. There are also three separate
hydrophobic cores, one in each structural element, and the third at the interface of the two elements formed by helices 4 and 5 and the central portion of helix 2. The hydrophobic core within element 1 is formed by Leu92, Trp93, Trp96, Ala99, the aliphatic portion of Arg104, Ala107, Ala162 and Val165. There is a sticky patch on the surface formed by Trp93, Trp96 and Ala162 which may account for the hydrophobic residues Ile77 and Pro80 folding back on the main body of the protein. The hydrophobic core within element 2 is formed by Ala113, Ala116, Ala117, Met120, Leu121, Thr127, Ala130, Phe131, Val134, Tyr138, Leu145 and Tyr149. Finally, the hydrophobic core at the interface of the two elements is formed by Leu106, Ala107, Leu111, Val114, Val140, Thr144, Leu145, Val152, Trp160, Ala161 and Leu164.

Structural homology to Iγ and other HTH proteins

Despite the absence of any significant sequence identity, the folding topology of the second structural element (helices 2–4) of the Iβ subdomain of MuA is remarkably similar to that of helices 1–3 of the Iγ subdomain (Clubb et al., 1997), as well as to a number of other HTH proteins such as the homeodomains (Pabo and Sauer, 1992) and the DNA-binding portion (helices C–E) of the trp repressor (Schevitz et al., 1985). A best-fit superposition of Iβ and Iγ is shown in Figure 4C. Specifically, the Cα atoms of residues 111–122, 123–135 and 140–151 of Iβ can be superimposed on the Cα atoms of residues 182–193, 196–208 and 219–230 of Iγ, with an r.m.s. deviation of 1.9 Å. The HTH motif in Iβ is formed by helices 3 and 4 which are oriented at an angle of ~115° to each other. The loop
The notation of the NMR structures is as follows: <SA> are the final 30 simulated annealing structures; \( \overline{SA} \) is the mean structure obtained by averaging the coordinates of the individual SA structures best fitted to each other using residues 89–166; \( \overline{(SA)r} \) 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. The final force constants employed for the various terms in the target function used for simulated annealing are as follows: 1000 kcal/mol/Å\(^2\) for bond lengths, 500 kcal/mol/Å\(^2\) for angles and improper torsions (which serve to maintain planarity and chirality), 4 kcal/mol/Å\(^4\) for the quartic van der Waals repulsion term (with the hard sphere effective van der Waals radii set to 0.8 times their value used in the CHARMM PARAM19/20 parameters), 30 kcal/mol/Å\(^2\) for the experimental distance restraints (interproton distances and hydrogen bonds), 200 kcal/mol/Å\(^2\) for the torsion angle restraints, 1 kcal/mol/Hz\(^2\) for the coupling constant restraints, 0.5 kcal/mol/p.p.m.\(^2\) for the carbon chemical shift restraints, and 1.0 for the conformational database potential.

None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5° or \( ^{1}J_{HN} \) coupling constant violations greater than 2 Hz. The torsion angles restraints comprise 94 φ, 8 ψ, 36 \( \chi_1 \) and 15 \( \chi_2 \) angles. The hydrogen bonding restraints which consist of two distances per backbone–backbone hydrogen bond were only included in the final stages of refinement using standard criteria based on amide exchange, \( ^{2}J_{HN} \) couplings and secondary \( ^{13}C \) shifts.

Only structurally useful intraresidue NOEs are included in the intraresidue interproton distance restraints. Thus, intraresidue NOEs between protons separated by two bonds or between non-stereospecifically assigned protons separated by three bonds are not incorporated in the restraints.

\( E_{L-J} \) is the Lennard–Jones van der Waals energy calculated with the CHARMM PARAM19/20 protein parameters (Brooks et al., 1983) and is not included in the target function for simulated annealing or restrained minimization.

The program PROCHECK (Laskowski et al., 1993) was used to assess the overall quality of the structures. More than 85% of residues in the most favorable region of the Ramachandran plot, <10 bad contacts per 100 residues, and a hydrogen bond energy of 0.6–1.0 are expected for a good quality structure. The dihedral angle G-factors (which should be greater than −0.5 for a good quality structure) for the φ,ψ, \( \chi_1/\chi_2 \), \( \chi_1/\psi \), and \( \psi/\chi_2 \) distributions are 0.26 ± 0.05, 0.57 ± 0.05, 0.26 ± 0.12 and 0.17 ± 0.13 respectively. The PROCHECK statistics apply to the residues 77–166.

The precision of the atomic coordinates is defined as the average r.m.s. difference between the 30 final simulated annealing structures and the mean coordinates, \( \overline{SA} \). The values given relate to residues 89–166, since the N- (residues 76–88) and C- (residues 167–174) termini are poorly defined by the experimental data. The values given for the backbone atoms relate to the N, Cα, C and O atoms; those given for all atoms refer only to non-hydrogen atoms.

Connecting the two helices of the HTH motif is four residues in length, typical of that of other prokaryotic HTH motifs, in contrast to that of \( \alpha \gamma \) which is longer (six residues in length).

However, there are a number of interesting differences between the HTH of \( \beta \) and that of other prokaryotic HTH proteins. The HTH motif of prokaryotes is generally characterized by the consensus sequence [\( _{-\text{Z-Gly}}/\text{Ala-}(\text{X})_2/\text{helix-[Z-Gly-Z-X]}_{\text{turn}}-\text{Ile/Leu/Val-_-helix, where Z is a hydrophobic aliphatic residue, X is any residue and the invariant glycine is located at position 2 of the turn} \) (Pabo and Sauer, 1992)]. \( \beta \) does have a similar sequence motif involving helices 3 and 4: specifically Phe\(_{131}\)-Ala-(X)\(_{2}\)-Ala-Gly-His-(X)\(_{3}\)-Ala\(_{142}\) which was previously used to predict the presence of an HTH motif in \( \beta \) with the turn located between Ala\(_{135}\) and Tyr\(_{138}\), and Gly\(_{136}\) at position 2 of the turn (Harshay et al., 1985; Kim and Harshay, 1995). While helices 3 (residues 127–137) and 4 (residues 142–151) comprise the HTH motif in \( \beta \), the location of the turn deviates from the prokaryotic consensus in so far as it actually extends from residues 138 to 141, with Gln\(_{139}\) at position 2 of the turn which is occupied by the invariant glycine in the prokaryotic consensus sequence. The delineation of the helices in \( \beta \) is unambiguous, not only from the NOE data, but also from the secondary \( ^{13}C \) shifts (Spera and Bax, 1991). On further inspection of the sequence of the HTH of \( \beta \), it can be seen that it actually follows the consensus

| Table I. Structural statistics\(^a\) |
|-----------------|-----------------|
| **Structural statistics** |
| **R.m.s. deviations from experimental distance restraints (Å)\(^b\)** |
| All (1055) | 0.019 ± 0.001 |
| interresidue sequential (|i – j| = 1) (255)\(^c\) | 0.024 ± 0.003 |
| interresidue short range (1 < |i – j| < 5) (220) | 0.023 ± 0.005 |
| interresidue long range (|i – j| > 5) (252) | 0.018 ± 0.004 |
| intraresidue (234) | 0.007 ± 0.004 |
| H-bonds (52) | 0.010 ± 0.008 |
| **R.m.s. deviations from experimental dihedral restraints (°) (153)\(^d\)** | 0.152 ± 0.075 |
| **R.m.s. deviations from \( ^{3}J_{HN} \) coupling constants (Hz) (58)\(^e\)** | 0.70 ± 0.06 |
| **R.m.s. deviations from experimental \( ^{13}C \) shifts** |
| \( ^{1}C\alpha \) (p.p.m.) (92) | 1.07 ± 0.11 |
| \( ^{13}C\beta \) (p.p.m.) (88) | 0.99 ± 0.04 |
| **Deviations from idealized covalent geometry** |
| bonds (Å) (1571) | 0.004 ± 0.0004 |
| angles (°) (2819) | 0.487 ± 0.025 |
| improppers (°) (853) | 0.510 ± 0.047 |
| **Measures of structural quality** |
| \( E_{L-J} \) (kcal/mol)\(^f\) | –420 ± 13 |
| % residues in most favorable region of Ramachandran plot | 94.0 ± 1.8 |
| No. of bad contacts/100 residues | 4.4 ± 1.6 |
| H-bond energy | 0.78 ± 0.06 |
| **Coordinate precision** |
| backbone (Å) | 0.48 ± 0.08 |
| all atoms (Å) | 1.00 ± 0.09 |

\(^a\)The values given relate to residues 89–166. The data were obtained from the PDB entry 7535.
with the exception of the replacement of the invariant Gly by Gln, i.e. [...]Val_{134}Ala-(X)_{2}helix{Tyr-Gln-Val-X}_{heun}^{-1}[(X)_{3}]Leu_{145}^{-1}...helix.

When Iβ is viewed perpendicular to the long axis of the recognition helix (helix 4) of the HTH motif (Figure 4A), it can be seen that two lobes extend from either side of it, such that the molecule has the shape of an elongated flat disc from which the DNA recognition helix protrudes. One lobe is formed by helix 1, helix 5 and the N-terminal half of helix 2, the other by the C-terminal half of helix 2 and helix 3. Thus, Iβ has the potential for extensive interactions with the DNA involving not only the recognition helix of the HTH motif but also the two lobes and the C-terminal tail of the molecule.

Evidence for the importance of the HTH motif of Iβ in DNA binding comes from mutational analysis: specifically, mutation of Phe131 to Ser or Arg146 to Asn severely reduces DNA binding affinity (Kim and Harshey, 1995).
Fig. 5. Binding of MuA Iβ (residues 77–174) and Iγ (residues 174–247) to synthetic 29 bp DNA oligonucleotides by gel affinity co-electrophoresis. The four double-stranded oligonucleotides consist of the complete 22 bp consensus sequence of the 12 end binding sites of phages Mu and D108, the consensus sequence mutated in either the γ (proximal half) or β (distal half) subsites, and a non-consensus reference oligonucleotide in which the consensus sequence has been mutated throughout with the exception of a 4 bp sequence in the center of the consensus sequence. The consensus sequence corresponds to the 24 bp sequence that is presumed to be contacted by the MuA protein as judged by footprinting (Zou et al., 1991).

Mutations were introduced by choosing the least probable base among the 12 binding sites. The four double-stranded 29 bp oligonucleotides (with a single base overhang at the 5′ ends) have the following sequences (with the wild-type consensus sequence indicated by upper case bold letters and the mutated sequence in lower case letters):

- 5′-d[aggTGTTCACCTTGAAAGCCTAACATTTAAATgag] 3′
- 5′-d[aggAATTTTTCGGTTTCTCAATGAAACAACC] 3′
- 5′-d[aggTGTTCACCTTGAAtgecctggg] 3′
- 5′-d[acccaaggggcacTTTCAAGTGAAACCC] 3′
- 5′-d[acccaaggggacTTTCAAGTGAAACCC] 3′

The double-stranded oligonucleotides are shown as schematic diagrams on the left-hand side of the figure, with the boxed region corresponding to the consensus sequence. The full or half consensus oligonucleotide was loaded on the right side of every lane, the non-consensus reference oligonucleotide in which the consensus sequence (on the right-hand side of every lane) versus a control 29 bp non-consensus oligonucleotide (on the left-hand side of every lane). The small shift observed upon specific binding of MuA Iγ is due to the small size and higher negative charge of this subdomain.

The Phe131→Ser mutation would destabilize the HTH motif since Phe131 is involved in extensive hydrophobic interactions with Ala142, Arg146 and Tyr149. The Arg146→Asn mutation may not only perturb the stability of the HTH motif, since the aliphatic portion of the side chain of Arg146 is packed against Phe131, but also likely to alter the nature of the contacts made with the DNA since the guanidino group of Arg146 is solvent exposed and available for interaction with DNA. Indeed, mutation of Arg146 to Val has been shown to alter the sequence preference at the distal half of the Mu end-binding site (R.Harshey and S.-Y.Namgoong, personal communication). Only one other mutation has been shown to abolish DNA binding, namely Lys157→Gln. Lys157 is located in the loop connecting helices 4 and 5, and model building suggests that it may possibly interact with the phosphate backbone of the DNA.

**Interactions of MuA Iβγ, Iβ and Iγ with the Mu end DNA-binding site**

The Iβγ domain footprints ~24 bp of DNA (Zou et al., 1991). This Mu end DNA consensus sequence is non-palindromic (Craigie et al., 1984). The distal β subsite exhibits a higher degree of conservation among the 12 end sites of phage Mu and the closely related phage D108 than the proximal γ subsite (Craigie et al., 1984).

We initially used gel mobility shift assays and DNase I footprinting to determine whether Iβγ and the subdomains Iβ and Iγ bind to the consensus Mu end DNA-binding sequence. The intact Iβγ domain gave a distinct gel shift as well as a DNase I footprint (data not shown). We were not able, however, to detect protein–DNA interactions between the isolated subdomains and the Mu end site by these methods.

To probe protein–DNA interactions under equilibrium conditions, we performed gel affinity co-electrophoresis (Lim et al., 1991). Using this approach, we were able directly to observe complexes of both subdomains with a 29 bp DNA containing the 22 bp Mu end consensus sequence (Figure 5). Interaction between the various purified domains (Iβγ, Iβ and Iγ) and the DNA were detected by the altered mobility of synthetic oligonucleotides in agarose gels containing different concentrations of the protein. Specific binding was assessed by comparison of the gel shifts obtained with test 29 bp oligonucleotides comprising either the full consensus or half the consensus sequence (on the right-hand side of every lane) versus a control 29 bp non-consensus oligonucleotide (on the
left-hand side of every lane). In this manner, a direct comparison between the gel shifts obtained with specific and non-specific binding is obtained under absolutely identical conditions, thereby permitting small gel shift differences to be observed. A schematic diagram is shown below each gel in Figure 5 to facilitate interpretation of the results.

As anticipated from the conventional gel mobility shift assay and DNase I footprinting experiments, MuA Iβγ binds a synthetic oligonucleotide containing a 22 bp consensus Mu end-DNA-binding site with high affinity (K_D ~10^{-9} M) (data not shown). At higher protein concentrations (several hundred nM), higher order complexes were observed, presumably due to non-specific binding of additional protein monomers to the same DNA molecule. A non-consensus 29 bp oligonucleotide, on the other hand, is only bound at Iβγ concentrations >10^{-7} M, and the apparent dissociation constant for non-specific binding was estimated to be >200 nM. Alteration of the γ and β subsites of the consensus Mu end-binding site reduced the affinity of Iγ by approximately one and two orders of magnitude respectively (K_D ~10^{-6} M and 10^{-7} M respectively).

The results of Iβ and Iγ binding to four oligonucleotides comprising the intact 22 bp Mu end consensus sequence, the consensus sequence mutated in the β subsite, the consensus sequence mutated in the γ subsite and a control non-consensus sequence are shown in Figure 5. Iβ (residues 77–174) binds the Mu end DNA consensus sequence and the consensus sequence mutated in its γ subsite with similar affinity (K_D ~10^{-7} M). Binding of Iγ to the consensus sequence mutated in its β site was undetectable (K_D >10^{-6} M). While specific binding of Iγ to both the intact consensus sequence and the consensus sequence mutated in its β subsite is weak (K_D ~10^{-6} M) and the extent of the mobility shift upon binding is very small, it is detected reproducibly by the gel affinity co-electrophoresis method. However, no binding of Iγ to the consensus sequence mutated in the γ subsite and the non-consensus sequence could be detected.

The gel affinity co-electrophoresis data therefore indicate that Iβ binds to the β subsite of the consensus and is not influenced by the presence or absence of the γ subsite. Similarly, Iγ binds to the γ subsite and is not influenced by the presence of the β subsite. Even though the Iγ subdomain binds the Mu end DNA with relatively low affinity, it clearly contributes to the ~100-fold tighter binding or 10-fold higher sequence specificity to the full consensus site when present together with Iβ in the intact Iβγ domain. Mixing of the isolated subdomains did not show any evidence of increased DNA affinity by either of the domains. Thus, the complete Iβγ domain is required for high affinity binding of the Mu end sites.

A model for the complex of Iβγ bound to the Mu end site

In many instances, multiple DNA-binding domains are required for site-specific recognition. This may involve multimers, such as homo- or heterodimers, or tandem repeats on a single polypeptide chain. The MuA Iβγ domain clearly has a modular organization, consisting of two independently folded subdomains, Iβ and Iγ, both of which possess a HTH motif, characteristic of a large number of DNA-binding proteins. Hydroxyl radical footprinting has demonstrated that the Iβγ domain binds to two consecutive major grooves and the intervening minor groove on the same face of the DNA (Zou et al., 1991), and we have shown in this study that the Iβ subdomain interacts with the distal half (β) subsite and the Iγ subdomain with the proximal half (γ) subsite. If we assume that in both cases the recognition helix of the HTH motif interacts with the major groove, it is possible to determine the orientation of binding of Iβγ on the consensus Mu end site and to propose an approximate model for the complex. Although it is known that Iβγ binds DNA by 60–90° (Zou et al., 1991), for the sake of simplicity we have shown the model with straight DNA in Figure 6. In the model, the recognition helices of the HTH motifs of Iβ and Iγ recognize successive major grooves on the same face of the DNA in the β and γ subsites respectively. The linker that connects the well ordered regions of the two domains (residues 89–166 for Iβ and residues 180–240 for Iγ) is 13 residues in length and extends from residue 167 to residue 179. In the model, the linker can interact readily with the minor groove at the interface of the β and γ subsites. In free solution, the linker is highly flexible (Clubb et al., 1997), but is likely to become ordered upon DNA binding, in a manner analogous to that observed in the case of the Lac repressor headpiece (Sprock et al., 1996).

In the model, it can also be seen that the potential contact surface between the Iβ subdomain and the DNA is far more extensive than that available to the smaller and more compact Iγ subdomain (Figure 6), which would account for the higher DNA-binding affinity of Iβ relative to Iγ (Figure 5). Thus, while the interaction of Iγ with DNA is limited to the recognition helix of the HTH motif and its N-terminal basic tail, the interaction of Iβ with DNA includes not only the recognition helix of the HTH motif and its C-terminal tail, but also helices 1 and 2 in one lobe, and helix 3 in the other lobe. The model suggests that residues at the C-terminal end of helix 1, and the N-terminal ends of helices 2 and 3, which contain a number of Lys residues, are in position to interact with the phosphate backbone of the DNA.

At this stage, we are not able to predict the direction of the bend observed upon complexation, i.e. whether the DNA wraps around Iβγ or bends away from Iβγ. Both possibilities have been observed in complexes where the protein contacts successive major grooves on the same face of the DNA. Thus, in the complexes of the catabolite activator protein (Schultz et al., 1991) and the MATα1/ MATα2 homeodomain heterodimer (Li et al., 1995) the DNA is bent towards the protein, while in complexes of the Lac (Lewis et al., 1996) and PurR (Schumacher et al., 1994) repressors the DNA is bent away from the protein. Work currently is in progress in our laboratory to determine the solution structure of a complex of Iβγ with DNA.

Materials and methods

Protein production

Plasmids and bacterial strains for protein expression of MuA Iβγ (residues 77–247), Iβ (residues 77–174) and Iγ (residues 174–247) were used as described in Clubb et al. (1997). Proteins were labeled uniformly (>95%) with ¹⁵N or with ¹⁵N and ¹³C by growing the bacteria in minimal
medium containing $^{15}$NH$_4$Cl and $^{13}$C$_6$-glucose as the sole nitrogen and carbon sources respectively, and were purified as follows. The soluble fraction of the *Escherichia coli* extract was applied to a DEAE-Sephrose Fast Flow (Pharmacia) column (200 ml bed volume) equilibrated with buffer A containing 100 mM Tris, pH 7.5, 5 mM EDTA and 5 mM dithiothreitol (DTT). The MuA proteins were eluted with a 0–1 M NaCl gradient in buffer A. The fractions containing MuA protein were pooled and dialyzed against buffer A. The samples were then applied to an SP Sepharose Fast Flow (Pharmacia) column (200 ml bed volume) equilibrated with buffer A. The MuA proteins were eluted with a 0–1 M NaCl gradient. MuA I$\beta$ and I$\gamma$ were applied to a G-50 Sepharose column equilibrated with 50 mM Na phosphate pH 6.8, 250 mM NaCl, 2.5 mM DTT. For the purification of MuA I$\beta\gamma$, the gel filtration column was equilibrated in 10 mM Na acetate pH 4.5, 100 mM NaCl and 2.5 mM DTT. Purified MuA I$\beta\gamma$ was dialyzed against 50 mM Na phosphate, 0.5 M NaCl, 2.5 mM DTT. The purity of the proteins was estimated to be $\geq$95% by SDS–PAGE.

Each NMR sample contained 2 mM MuA I$\beta$ in 50 mM Na phosphate, pH 6.3 and 150 mM NaCl. Five samples of MuA I$\beta$ were prepared for NMR: (i) unlabeled MuA I$\beta$ in 100% D$_2$O; (ii) uniformly $^{15}$N-labeled MuA I$\beta$ in 90% H$_2$O/10% D$_2$O; (iii) uniformly $^{15}$N/$^{13}$C-labeled MuA I$\beta$ in 90% H$_2$O/10% D$_2$O; (iv) uniformly $^{15}$N/$^{13}$C-labeled MuA I$\beta$ in 100% D$_2$O; and (v) $^{15}$C-[Tyr,Phe,Trp,His] reversed labeled $^{15}$N/$^{13}$C MuA I$\beta$ in 100% D$_2$O.

### Analysis of protein–DNA interactions by affinity co-electrophoresis

Affinity co-electrophoresis (Lim *et al.*, 1991) was used to probe interactions between MuA I$\beta\gamma$, I$\beta$ or I$\gamma$ and a Mu end consensus DNA-binding site and several variations thereof. The 1% SeaPlaque LGT gels and 3( I$\beta$) or 4.5% (I$\gamma$) NuSieve agarose gels were used essentially as described by Lim *et al.* (1991). The protein concentration in the gel was varied between 0 and 3.9 $\mu$M. The electrophoresis buffer contained 1× TAE, 50 mM NaCl, 1 mM DTT and 100 $\mu$g/ml bovine serum albumin (BSA); for I$\gamma$, the NaCl was omitted. The synthetic oligonucleotides (obtained from J.Flory, Yale University) were 5′ end-labeled using [γ-32P]ATP and T4 polynucleotide kinase. Four $\mu$l (0.4 nM) of DNA was electrophoresed through the protein-containing lanes of the gel at 60 V for 1.4 h. Two slots were formed in each lane: one slot was loaded with the control non-consensus oligonucleotide, while the other slot was loaded with a test oligonucleotide comprising either half of the consensus or the full consensus sequence (see legend to Figure 5 for sequences). The gels were dried and exposed to a Fuji imaging plate. The autoradiographs were processed with a Fuji BAS 2000 (Fuji Medical Systems). Binding constants were analyzed based on the retardation of the protein–DNA complexes and calculated using Scatchard plots.

### NMR spectroscopy

NMR experiments were performed on Bruker AMX500, AMX600, DMX600 and DMX750 spectrometers equipped with an x,y,z-shielded...
gradient triple resonance probe at 30°C. ^1H, ^13C and ^15N assignments of MuA ^β were obtained using double and triple resonance 3D NMR spectroscopy (Clore and Gronenborn, 1991; 1994; Bax and Grzesiek, 1993). These included 3D CBCANH, CBCA(CO)NH, HBHA(CO)NH, H(CC)NH, HCCH-COSY, HCCH-TOCSY, HNHA and ^13N-separated HOHAHA experiments. Approximate interproton distance restraints were derived from 3D ^15N-separated (120 ms mixing time) and ^13C-separated (110 ms mixing time) NOE spectra, 3D ^15N-separated (30 ms mixing time) and ^13C-separated (45 ms mixing time) ROE spectra, 4D ^15N/ ^13C-separated (100 ms mixing time) and ^13C-separated (100 ms mixing time) NOE spectra. In addition, approximate interproton distance restraints from ^13C-attached protons to ^12C-attached protons of aromatic residues were obtained from a 3D ^13C-separated/ ^12C-filtered NOE spectrum (100 ms mixing time) recorded on the ^12C- [Tyr,Phc,Trp,His] reversed labeled ^13C/ ^12C sample (Vuister et al., 1994). ^13C, ^13C, ^13C, ^13C and ^13C coupling constants were obtained by quantitative J correlation spectra (Bax et al., 1994; Hu and Bax, 1997a,b; Hu et al., 1997). All NMR spectra were processed with the NMR Pipe software package (Delaglio et al., 1995) and analyzed with the programs PIPP, CAPP and STAPP (Garrett et al., 1991).

**Structure calculation**

The interproton distance restraints derived from the 3D and 4D hetero-nuclear-separated NOE spectra were classified into four ranges, 1.8–2.7 A (1.8–2.9 A for NOEs involving NH protons), 1.8–3.3 A (1.8–3.5 A for NOEs involving NH protons), 1.8–5.0 A and 1.8–6.0 A, corresponding to strong, medium, weak and very weak NOEs respectively (Clore and Gronenborn, 1991). Distances involving methyl groups, aromatic ring protons and non-stereospecifically assigned methylene protons were represented as a 12C=9-1-13C=6.15 A (Nilges, 1993). Protein backbone hydrogen bonding restraints were introduced during the final stages of refinement according to standard criteria. The structures were calculated using a modified version of the hybrid distance geometry—dynamical simulated annealing protocol (Nilges et al., 1988) using the program XPLOR 3.1 (Brünger, 1993), adapted to incorporate pseudopotentials for ^13C, coupling constant (Garrett et al., 1994) and secondary ^13C and ^12C chemical shift (Kuszewski et al., 1995) restraints, and a conformational database potential derived from very high resolution (1.7 A or better) X-ray structures (Kuszewski et al., 1996, 1997). The target function that is minimized during simulated annealing comprises only quadratic harmonic potential terms for covalent geometry (i.e. bonds, angles, chirality), coupling constant and chemical shift restraints; quadratic square-well potential terms for the experimental distance and torsion angle restraints; and a quartic van der Waals repulsion term and a conformational database potential term for the non-bonded contacts. The latter biases sampling during simulated annealing refinement to conformations that are likely to be energetically possible by effectively limiting the choices of dihedral angles to those that are known to be physically realizable (Kuszewski et al., 1996). No hydrogen bonding, electrostatic or 6–12 Lennard-Jones empirical potential terms are present in the target function. The coordinates of the ensemble of 30 simulated annealing structures, the restrained regularized mean structure and the experimental restraints have been deposited in the Brookhaven Protein Data Bank (accession codes 2EZK, 2EZL and 2EZKMR).

**Acknowledgements**

We thank Drs M. Caffrey, B. Gronenborn, J. Huth, J. Omichinski, M. Starich and M. Wikström for useful discussions, H. Savilahit for help with initial DNA-binding experiments, P. Goy and R. Tschudin for technical support, and F. Delaglio and D.S. Garrett for software support. This work was supported by a predoctoral Deutscher Akademischer Austauschdienst fellowship (to S.S.), a Leukemia Society of America post-doctoral fellowship (to R.T.C.), and the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C., A.M.G. and K.M.).

**References**


Hu,J.-S. and Bax,A. (1997b) $\chi_1$ angle information from a simple two-dimensional NMR experiment which identifies trans $^{1}J_{NC}$ couplings in isotopically enriched proteins. *J. Biomol. NMR.*, **9**, 323–328.


