

Solution structure of the Mu end DNA-binding I β subdomain of phage Mu transposase: modular DNA recognition by two tethered domains

Silke Schumacher¹, Robert T.Clubb^{1,3},
Mengli Cai¹, Kiyoshi Mizuuchi²,
G.Marius Clore^{1,4} and
Angela M.Gronenborn^{1,4}

Laboratories of ¹Chemical Physics and ²Molecular Biology,
National Institute of Diabetes and Digestive and Kidney Diseases,
National Institutes of Health, Bethesda, MD 20892-0520, USA

³Present address: Department of Chemistry and Biochemistry,
University of California, Los Angeles at Los Angeles,
CA 90095-1570, USA

⁴Corresponding authors
e-mail: clore@vger.niddk.nih.gov; or gronenborn@vger.niddk.nih.gov

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 (I β γ). The three-dimensional solution structure of the I β 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 I γ 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, I β to the more conserved Mu end distal half (β subsite) and I γ to the Mu end proximal half (γ subsite) of the consensus Mu end-binding site. The complete I β γ 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 I β and I γ subdomains, a model for the interaction of the complete I β γ domain with DNA is proposed.

Keywords: I β 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 transposition 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'-GTTTCAYNNRAARYRCGA-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 I α 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 I β γ 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 (I β) and 174–247 (I γ) (Clubb *et al.*, 1997). The solution structure of I γ 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

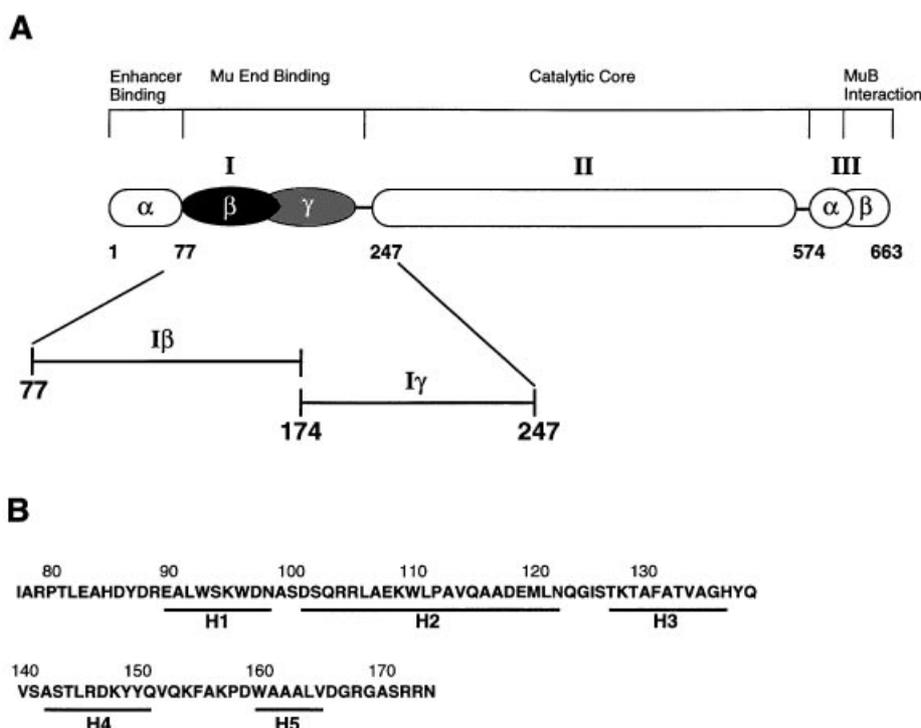


Fig. 1. (A) Schematic diagram of the domain structure of MuA transposase. Domains identified by partial proteolysis (Nagayama *et al.*, 1987) are labeled I–III and marked with residue numbers at the beginning and end of each domain. The constructs used in this study comprising subdomains I β (residues 77–174) or I γ (residues 174–247) are shown below as bars. (B) Sequence of I β (residues 77–174), with the location of the five helices determined by NMR indicated.

also displays non-specific DNA-binding activity (Nakayama *et al.*, 1987) and has a ‘DD35E’ motif which has been implicated in cation binding in the active site (Baker and Luo, 1994). Domain III α (residues 575–603) has both non-specific DNA binding and cryptic nuclease activity (Leung and Harshey, 1991; Baker *et al.*, 1993; Wu and Chaconas, 1995). Domain III β (residues 604–663) interacts with MuB protein, which promotes strand transfer to target DNA (Baker *et al.*, 1991; Leung and Harshey, 1991; Wu and Chaconas, 1994), and with the host-encoded chaperone ClpX (Levchenko *et al.*, 1995) which is involved in disassembly of the transpososome after transposition and prior to replication (Kruklytis *et al.*, 1996).

Here we present the solution structure of the I β subdomain (residues 77–174) of MuA, and characterize the role of the I β and I γ subdomains within the modular organization of the Mu end-binding domain I β γ by gel affinity co-electrophoresis.

Results and discussion

Structure determination

The solution structure of the I β subdomain was determined using double and triple resonance multidimensional NMR spectroscopy (Clare and Gronenborn, 1991, 1994; Bax and Grzesiek, 1993). Examples of the quality of the NMR data are provided by typical planes from the 4D $^{15}\text{N}/^{13}\text{C}$ - and $^{13}\text{C}/^{13}\text{C}$ -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 I β

Two views of a ribbon diagram of the I β subdomain are shown in Figure 4A and B. The I β 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. I β 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 $\sim 120^\circ$, $\sim 75^\circ$ and $\sim 100^\circ$ 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 $\sim 145^\circ$, $\sim 85^\circ$ and $\sim 115^\circ$ between helices 2 and 3, 2 and 4, and 3 and 4 respectively. Thus, helix 2, which spans the entire length of the I β subdomain, serves to bridge the two structural elements. A proline residue at position 112 introduces a small kink of $\sim 15^\circ$ in the middle of helix 2. There are also three separate

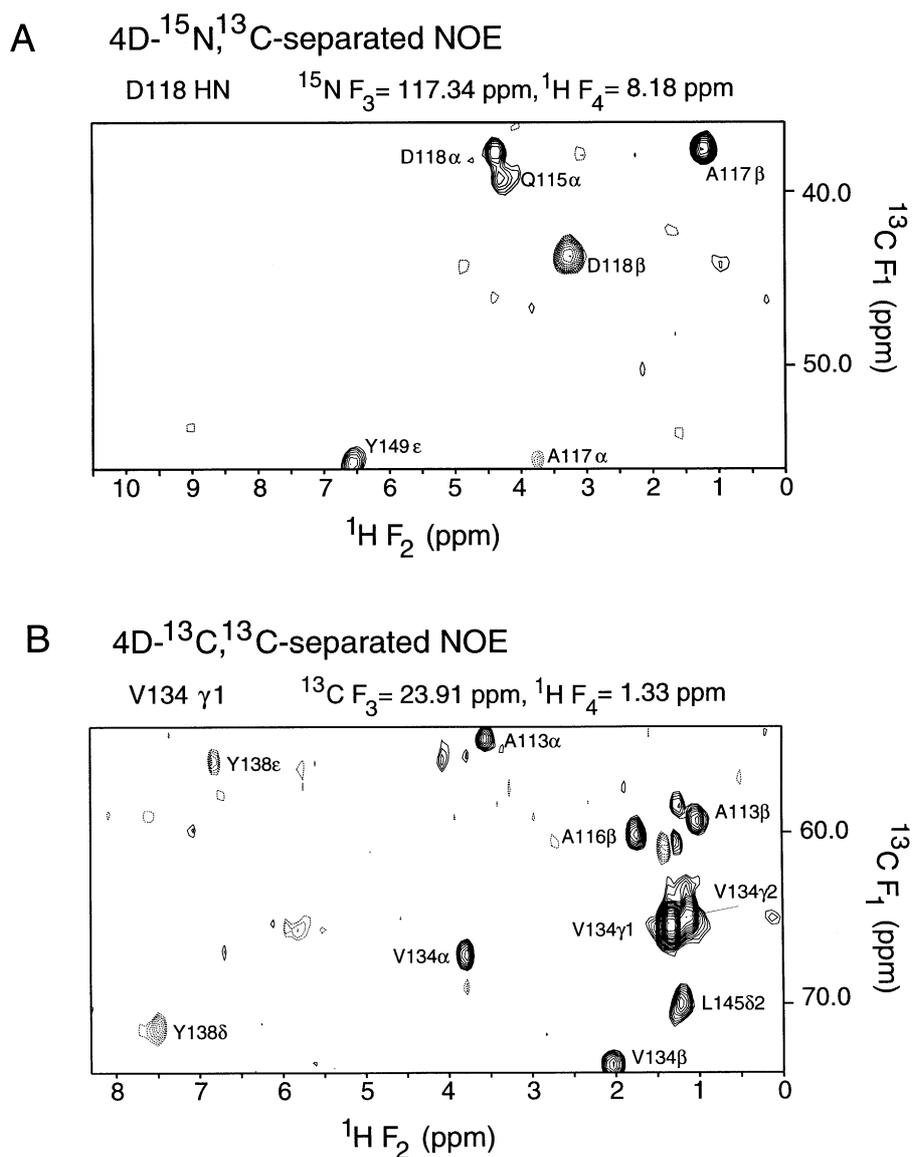


Fig. 2. Example of $^{13}\text{C}(F_1)-^1\text{H}(F_2)$ planes from the (A) 4D $^{15}\text{N}/^{13}\text{C}$ -separated and (B) $^{13}\text{C}/^{13}\text{C}$ -separated NOE spectra (100 ms mixing time) recorded on MuA I β . The destination proton in (A) is the NH group of Asp118 and in (B) the $\text{C}^\gamma\text{H}_3$ methyl group of Val134. Unlabeled cross peaks have their maximum intensity in an adjacent plane. The sweep width in the ^{13}C dimensions (F_1 and F_3) is 20.71 p.p.m. so that extensive folding is employed, with peaks folded an even and odd number of times being of opposite sign. In (A), the ^{13}C carrier was set to 46 p.p.m., and peaks folded an even and odd number of times are indicated by dashed and continuous contours respectively. In (B), the ^{13}C carrier was set to 63.71 p.p.m., and peaks folded an even and odd number of times are indicated by continuous and dashed contours respectively.

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 I γ 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 I β subdomain of MuA is remarkably similar to that of helices 1–3 of the I γ 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 I β and I γ is shown in Figure 4C. Specifically, the C α atoms of residues 111–122, 123–135 and 140–151 of I β can be superimposed on the C α atoms of residues 182–193, 196–208 and 219–230 of I γ , with an r.m.s. deviation of 1.9 Å. The HTH motif in I β is formed by helices 3 and 4 which are oriented at an angle of $\sim 115^\circ$ to each other. The loop

Table I. Structural statistics^a

	<SA>	(\overline{SA}) _r
Structural statistics		
R.m.s. deviations from experimental distance restraints (Å) ^b		
All (1055)	0.019 ± 0.001	0.023
interresidue sequential ($ i - j = 1$) (255) ^c	0.024 ± 0.003	0.025
interresidue short range ($1 < i - j \leq 5$) (220)	0.023 ± 0.005	0.016
interresidue long range ($ i - j > 5$) (252)	0.018 ± 0.004	0.019
intraresidue (234)	0.007 ± 0.004	0.025
H-bonds (52)	0.010 ± 0.008	0.000
R.m.s. deviations from experimental dihedral restraints (°) (153) ^b	0.152 ± 0.075	0.295
R.m.s. deviations from ³ J _{HNα} coupling constants (Hz) (58) ^b	0.70 ± 0.06	0.90
R.m.s. deviations from experimental ¹³ C shifts		
¹³ Cα (p.p.m.) (92)	1.07 ± 0.11	1.04
¹³ Cβ (p.p.m.) (88)	0.99 ± 0.04	1.11
Deviations from idealized covalent geometry		
bonds (Å) (1571)	0.004 ± 0.0004	0.006
angles (°) (2819)	0.487 ± 0.025	0.872
impropers (°) (853)	0.510 ± 0.047	1.029
Measures of structural quality		
E _{L-J} (kcal/mol) ^d	-420 ± 13	-370
PROCHECK ^e		
% residues in most favorable region of Ramachandran plot	94.0 ± 1.8	98.6
No. of bad contacts/100 residues	4.4 ± 1.6	5.1
H-bond energy	0.78 ± 0.06	0.80
Coordinate precision ^f		
backbone (Å)	0.48 ± 0.08	
all atoms (Å)	1.00 ± 0.09	

^aThe 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; (SA)_r is the restrained minimized mean structure obtained by restrained regularization of the mean structure 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/Å² for bond lengths, 500 kcal/mol/rad² for angles and improper torsions (which serve to maintain planarity and chirality), 4 kcal/mol/Å⁴ 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/Å² for the experimental distance restraints (interproton distances and hydrogen bonds), 200 kcal/mol/rad² for the torsion angle restraints, 1 kcal/mol/Hz² 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.

^bNone of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5° or ³J_{HNα} coupling constant violations greater than 2 Hz. The torsion angles restraints comprise 94 φ, 8 ψ, 36 χ₁ and 15 χ₂ 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, ³J_{HNα} couplings and secondary ¹³C shifts.

^cOnly 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.

^dE_{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.

^eThe 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, χ₁/χ₂, χ₁ and χ₃/χ₄ 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. Residues 167–174 appear to be completely disordered in solution.

^fThe 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, 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 Iγ which is longer (six residues in length).

However, there are a number of interesting differences between the HTH of Iβ and that of other prokaryotic HTH proteins. The HTH motif of prokaryotes is generally characterized by the consensus sequence [...Z-Gly/Ala-(X)₂]_{helix}-[Z-Gly-Z-X]_{turn}-[(X)₃-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). Iβ does have a similar sequence motif involving helices 3 and 4: specifically Phe₁₃₁-Ala-(X)₂-Ala-Gly-His-(X)₄-Ala₁₄₂ which was previously used

to predict the presence of an HTH motif in Iβ with the turn located between Ala135 and Tyr138, and Gly136 at position 2 of the turn (Harshey *et al.*, 1985; Kim and Harshey, 1995). While helices 3 (residues 127–137) and 4 (residues 142–151) comprise the HTH motif in Iβ, the location of the turn deviates from the prokaryotic consensus in so far as it actually extends from residues 138 to 141, with Gln139 at position 2 of the turn which is occupied by the invariant glycine in the prokaryotic consensus sequence. The delineation of the helices in Iβ is unambiguous, not only from the NOE data, but also from the secondary ¹³Cα and ¹³Cβ shifts (Spera and Bax, 1991). On further inspection of the sequence of the HTH of Iβ, it can be seen that it actually follows the consensus

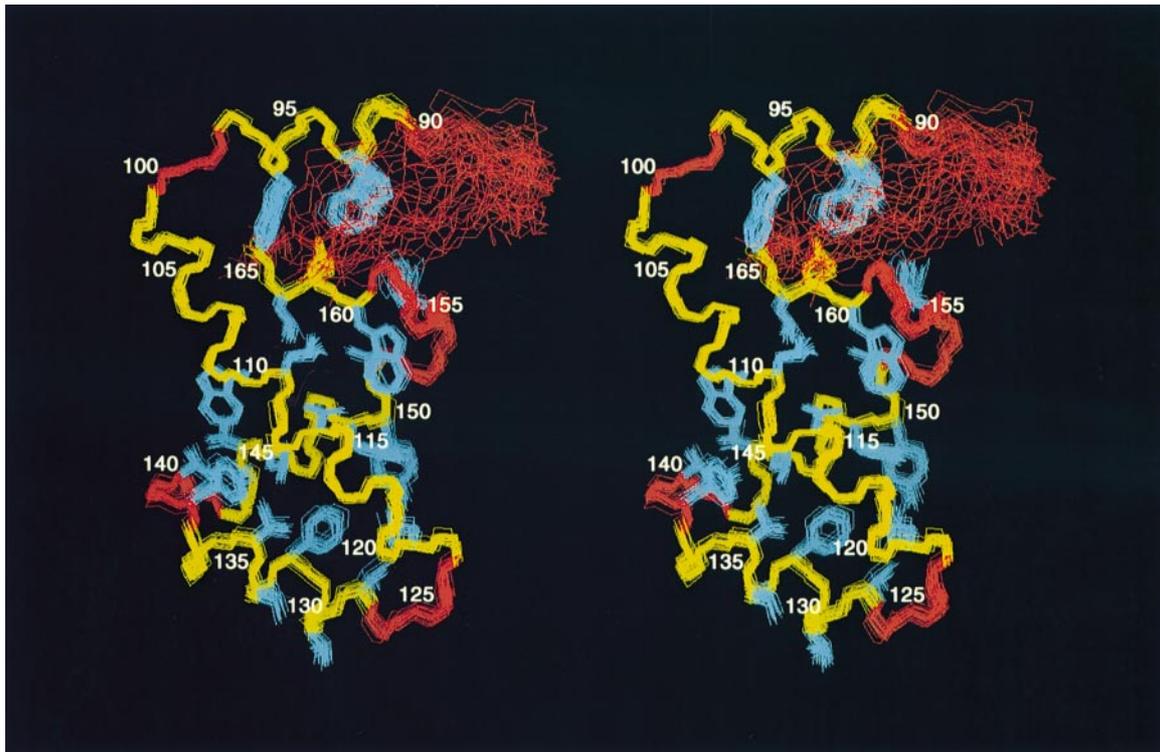


Fig. 3. Stereoview showing a best-fit superposition of the 30 final simulated annealing structures of MuA I β (residues 77–166). The backbone N, C α and C atoms are displayed in yellow for the helices and in red for the remaining residues, and selected side chains are shown in blue. The C-terminal residues (167–174) are undefined by the present data.

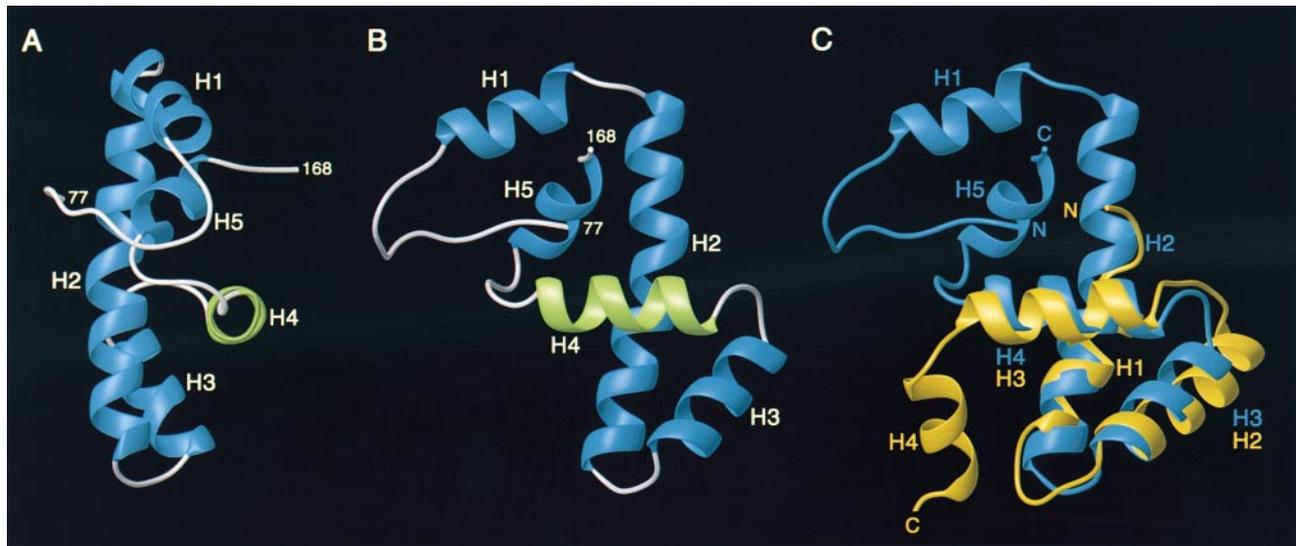


Fig. 4. (A and B) Two views of a ribbon diagram representation of the restrained regularized mean structure of MuA I β (residues 77–168). Helix 4 (marked in green) is the recognition helix of the HTH DNA-binding motif. (C) Comparison of the MuA I β (blue) and I γ (yellow) subdomains. Helices 2–4 of I β have been superimposed on helices 1–3 of I γ with a C α atomic r.m.s. difference of 1.9 Å for 37 residues. The figure was generated using the program MOLMOL (Konradi *et al.*, 1996).

with the exception of the replacement of the invariant Gly by Gln, i.e. [...-Val₁₃₄-Ala-(X)₂]_{helix}-[Tyr-Gln-Val-X]_{turn}-[(X₃)-Leu₁₄₅-...]_{helix}.

When I β 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, I β 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 I β 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).

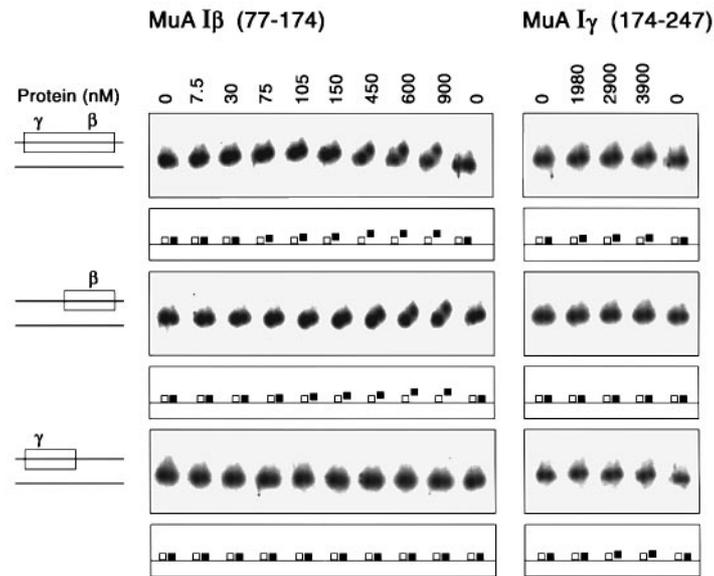


Fig. 5. Binding of MuA I β (residues 77–174) and I γ (residues 174–247) to synthetic 29 bp DNA oligonucleotides by gel affinity co-electrophoresis. The four double-stranded oligonucleotides consist of the complete 22 bp consensus sequence of the 12 end binding sites of phages Mu and D108, the consensus 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[agg**TGTTTCACTT**GAAACGCGAAAAATgg]. 5'-d[acc**ATTTTTCGCGTTTCAAGT**GAAACAcc],
 5'-d[aggaacgaggcc**GAAACGCGAAAAAT**gg]. 5'-d[acc**ATTTTTCGCGTTT**Cggccctcgtcc],
 5'-d[agg**TGTTTCACTT**GAAAGtcccccttggg]. 5'-d[acccaagggcac**TTTCAAGT**GAAACAcc],
 5'-d[aggaacgaggcc**GA- AA**gtcccccttggg]. 5'-d[acccaagggcac**TTT**Cggccctcgtcc].

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 was loaded on the left side. The upper gels show the binding of MuA I β (left-hand panels) and I γ (right-hand panels) to the full consensus sequence, the middle gels to the consensus sequence mutated in its γ (5'-half) subsite, and the bottom gels to the consensus sequence mutated in its β (3'-half) subsite. To facilitate interpretation, a schematic diagram is provided below each gel, illustrating the shift of the specific oligonucleotide on the right-hand side of each lane (■) relative to the non-consensus reference oligonucleotide on the left-hand side (□). The small shift observed upon specific binding of MuA I γ 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 is 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 I $\beta\gamma$, I β and I γ with the Mu end DNA-binding site

The I $\beta\gamma$ 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 I $\beta\gamma$ and the subdomains I β and I γ bind to the consensus Mu end DNA-binding sequence. The intact I $\beta\gamma$ 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 (I $\beta\gamma$, I β and I γ) 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 I β γ binds a synthetic oligonucleotide containing a 22 bp consensus Mu end DNA-binding site with high affinity ($K_D \sim 10^{-9}$ 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 I β γ 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 I β γ by approximately one and two orders of magnitude respectively ($K_D \sim 10^{-8}$ M and 10^{-7} M respectively).

The results of I β and I γ 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. I β (residues 77–174) binds the Mu end DNA consensus sequence and the consensus sequence mutated in its γ subsite with similar affinity ($K_D \sim 10^{-7}$ M). Binding of I β to the consensus sequence mutated in its β site was undetectable ($K_D > 10^{-6}$ M). While specific binding of I γ to both the intact consensus sequence and the consensus sequence mutated in its β subsite is weak ($K_D \sim 10^{-6}$ 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 I γ 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 I β binds to the β subsite of the consensus and is not influenced by the presence or absence of the γ subsite. Similarly, I γ binds to the γ subsite and is not influenced by the presence of the β subsite. Even though the I γ 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 I β in the intact I β γ domain. Mixing of the isolated subdomains did not show any evidence of increased DNA affinity by either of the domains. Thus, the complete I β γ domain is required for high affinity binding of the Mu end sites.

A model for the complex of I β γ 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 I β γ domain clearly has a modular organization, consisting of two independently folded subdomains, I β and I γ , both of which possess a HTH motif, characteristic of a large

number of DNA-binding proteins. Hydroxyl radical footprinting has demonstrated that the I β γ 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 I β subdomain interacts with the distal half (β) subsite and the I γ 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 I β γ on the consensus Mu end site and to propose an approximate model for the complex. Although it is known that I β γ bends 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 I β and I γ 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 I β and residues 180–240 for I γ) 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 (Spronk *et al.*, 1996).

In the model, it can also be seen that the potential contact surface between the I β subdomain and the DNA is far more extensive than that available to the smaller and more compact I γ subdomain (Figure 6), which would account for the higher DNA-binding affinity of I β relative to I γ (Figure 5). Thus, while the interaction of I γ with DNA is limited to the recognition helix of the HTH motif and its N-terminal basic tail, the interaction of I β 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 I β γ or bends away from I β γ . 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 MAT α 1/MAT α 2 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 I β γ with DNA.

Materials and methods

Protein production

Plasmids and bacterial strains for protein expression of MuA I β γ (residues 77–247), I β (residues 77–174) and I γ (residues 174–247) were used as described in Clubb *et al.* (1997). Proteins were labeled uniformly ($>95\%$) with ^{15}N or with ^{15}N and ^{13}C by growing the bacteria in minimal

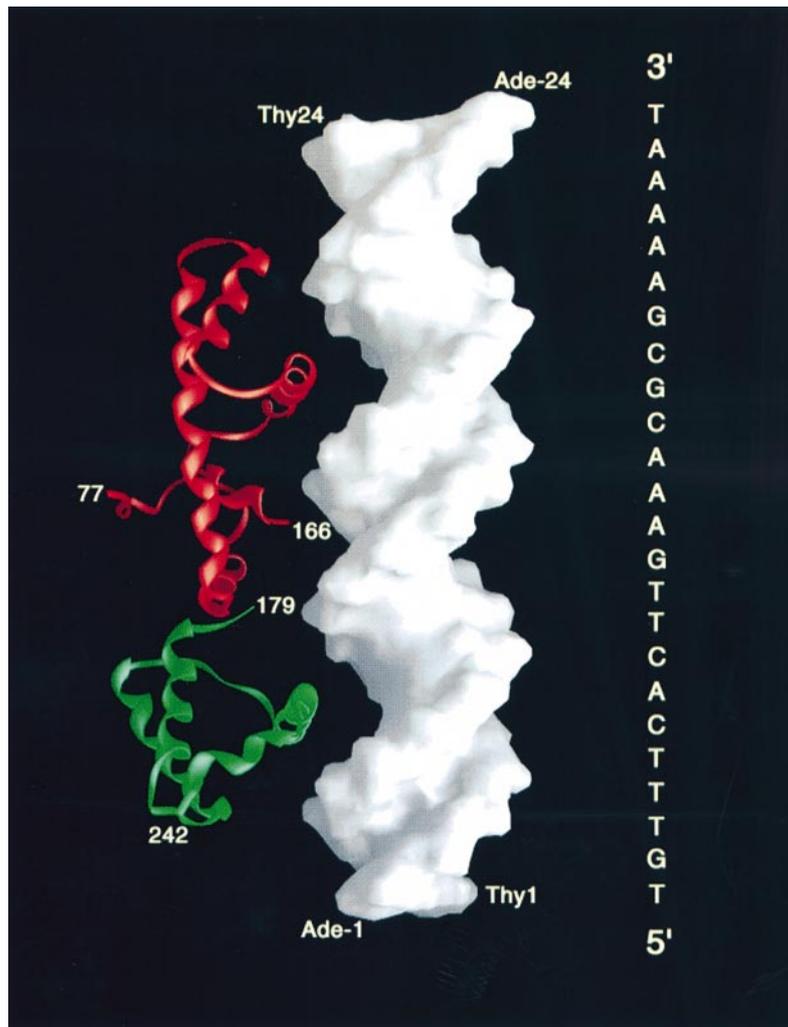


Fig. 6. Model of the interaction of the MuA I β γ with a 24 bp Mu end consensus binding site. The sequence of the DNA is shown on the right, corresponding to the consensus sequence of the Mu end sites used in the gel affinity co-electrophoresis experiments. Thy1 corresponds to Thy6 of the top strand of the Mu left end sequence. I β and I γ are displayed as red and green ribbons respectively, and the DNA is shown as a molecular surface representation in white. The linker (residues 167–178) connecting the C-terminus of I β to the N-terminus of I γ is not shown since we do not have any information as to its conformation in the complex. In addition, even though I β γ is known to bend the DNA by 60–90°, the DNA is shown for the sake of simplicity as a straight rod since we do not know whether I β γ bends the DNA towards or away from the body of the protein. The figure was generated using the program GRASP (Nicholls *et al.*, 1991).

medium containing $^{15}\text{NH}_4\text{Cl}$ and $^{13}\text{C}_6$ -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–Sepharose 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 β and I γ 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 β γ , 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 β γ was dialyzed against 50 mM Na phosphate, 0.5 M NaCl, 2.5 mM DTT. The purity of the proteins was estimated to be >95% by SDS–PAGE.

Each NMR sample contained 2 mM MuA I β in 50 mM Na phosphate, pH 6.3 and 150 mM NaCl. Five samples of MuA I β were prepared for NMR: (i) unlabeled MuA I β in 100% D_2O ; (ii) uniformly ^{15}N -labeled MuA I β in 90% $\text{H}_2\text{O}/10\%$ D_2O ; (iii) uniformly $^{15}\text{N}/^{13}\text{C}$ -labeled MuA I β in 90% $\text{H}_2\text{O}/10\%$ D_2O ; (iv) uniformly $^{15}\text{N}/^{13}\text{C}$ -labeled MuA I β in 100% D_2O ; and (v) ^{12}C -[Tyr, Phe, Trp, His] reversed labeled $^{15}\text{N}/^{13}\text{C}$ MuA I β in 100% D_2O .

Analysis of protein–DNA interactions by affinity co-electrophoresis

Affinity co-electrophoresis (Lim *et al.*, 1991) was used to probe interactions between MuA I β γ , I β or I γ and a Mu end consensus DNA-binding site and several variations thereof. The 1% SeaPlaque LGT gels (I β γ) and 3 (I β) or 4.5% (I γ) 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 μM . The electrophoresis buffer contained $1\times$ TAE, 50 mM NaCl, 1 mM DTT and 100 $\mu\text{g}/\text{ml}$ bovine serum albumin (BSA); for I γ , the NaCl was omitted. The synthetic oligonucleotides (obtained from J.Flory, Yale University) were 5' end-labeled using [γ - ^{32}P]ATP and T4 polynucleotide kinase. Four μl (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 , ^{13}C and ^{15}N assignments of MuA I β were obtained using double and triple resonance 3D NMR spectroscopy (Clare and Gronenborn, 1991, 1994; Bax and Grzesiek, 1993). These included 3D CBCANH, CBCA(CO)NH, HBHA(CO)NH, C(CO)NH, H(CCO)NH, HCCH-COSY, HCCH-TOCSY, HNHA and ^{15}N -separated HOHAHA experiments. Approximate interproton distance restraints were derived from 3D ^{15}N -separated (120 ms mixing time) and ^{13}C -separated (110 ms mixing time) NOE spectra, 3D ^{15}N -separated (30 ms mixing time) and ^{13}C -separated (45 ms mixing time) ROE spectra, 4D $^{15}\text{N}/^{13}\text{C}$ -separated (100 ms mixing time) and $^{13}\text{C}/^{13}\text{C}$ -separated (100 ms mixing time) NOE spectra. In addition, approximate interproton distance restraints from ^{13}C -attached protons to ^{12}C -attached protons of aromatic residues were obtained from a 3D ^{13}C -separated/ ^{12}C -filtered NOE spectrum (100 ms mixing time) recorded on the ^{12}C -[Tyr,Phe,Trp,His] reversed labeled $^{15}\text{N}/^{13}\text{C}$ sample (Vuister *et al.*, 1994). $^3\text{J}_{\text{HN}\alpha}$, $^3\text{J}_{\text{CC}}$, $^3\text{J}_{\text{CN}}$ and $^3\text{J}_{\text{C}\alpha\text{CO}}$ and $^3\text{J}_{\text{NH}\beta}$ 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 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 (Clare and Gronenborn, 1991). Distances involving methyl groups, aromatic ring protons and non-stereospecifically assigned methylene protons were represented as a $(\Sigma r^{-6})^{-1/6}$ sum (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 $^3\text{J}_{\text{HN}\alpha}$ coupling constant (Garrett *et al.*, 1994) and secondary $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shift (Kuszewski *et al.*, 1995) restraints, and a conformational database potential derived from very high resolution (1.7 Å 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, 2EYL and 2EZKMR).

Acknowledgements

We thank Drs M.Caffrey, B.Gronenborn, J.Huth, J.Omichinski, M.Starich and M.Wikström for useful discussions, H.Savilahti for help with initial DNA-binding experiments, G.Poy 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

Adzuma, K. and Mizuuchi, K. (1987) MuA protein-induced bending of Mu end DNA. In Sarma, R.H. (ed.), *Proceedings of the Fifth Conversation in Biomolecular Stereodynamics*, Vol. 3, pp. 97–104.

Baker, T.A. and Luo, L. (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc. Natl Acad. Sci. USA*, **91**, 6654–6658.

Baker, T.A., Mizuuchi, M. and Mizuuchi, K. (1991) MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell*, **65**, 1003–1013.

Baker, T.A., Mizuuchi, M., Savilahti, H. and Mizuuchi, K. (1993) Division of labor among monomers in the Mu transposase tetramer. *Cell*, **74**, 723–733.

Bax, A. and Grzesiek, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res.*, **26**, 131–138.

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R. and Zhu, G. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.*, **239**, 79–105.

Berg, D.E. and Howe, M.M., eds (1989) *Mobile DNA*. American Society of Microbiology, Washington, DC.

Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.*, **4**, 197–217.

Brünger, A.T. (1993) *X-PLOR Version 3.1. Manual*. Yale University Press, New Haven, CT.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1995) High resolution structure of the catalytic domain of the avian sarcoma virus integrase. *J. Mol. Biol.*, **253**, 333–346.

Clare, G.M. and Gronenborn, A.M. (1991) Structures of larger proteins in solution: three- and four dimensional heteronuclear NMR spectroscopy. *Science*, **252**, 1390–1399.

Clare, G.M. and Gronenborn, A.M. (1994) Multidimensional heteronuclear magnetic resonance of proteins. *Methods Enzymol.*, **239**, 349–363.

Clubb, R.T., Omichinski, J.G., Savilahti, H., Mizuuchi, K., Gronenborn, A.M. and Clare, G.M. (1994) A novel class of winged helix–turn–helix protein: the DNA-binding domain of Mu transposase. *Structure*, **2**, 1041–1048.

Clubb, R.T., Mizuuchi, M., Huth, J.R., Omichinski, J.G., Savilahti, H., Mizuuchi, K., Clare, G.M. and Gronenborn, A.M. (1996) The wing of the enhancer binding domain of phage Mu transposase is flexible and essential for efficient transposition. *Proc. Natl Acad. Sci. USA*, **93**, 1146–1150.

Clubb, R.T., Schumacher, S., Mizuuchi, K., Gronenborn, A.M. and Clare, G.M. (1997) Solution structure of the I γ subdomain of the Mu end DNA binding domain of phage Mu transposase. *J. Mol. Biol.*, **273**, 19–25.

Craigie, R. and Mizuuchi, K. (1985) Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell*, **41**, 867–876.

Craigie, R. and Mizuuchi, K. (1987) Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell*, **51**, 493–501.

Craigie, R., Mizuuchi, M. and Mizuuchi, K. (1984) Site-specific recognition of the bacteriophage Mu ends by the MuA protein. *Cell*, **39**, 387–394.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMR Pipe: a multidimensional spectral processing system based on UNIX PIPES. *J. Biomol. NMR*, **6**, 277–293.

Ding, A., Harshey, R.M. and Hurley, L.H. (1993) (+)-CC-1065 as a structural probe of Mu transposase-induced bending of DNA: overcoming limitations of hydroxyl-radical footprinting. *Nucleic Acids Res.*, **21**, 4281–4287.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.

Garrett, D.S., Powers, R., Gronenborn, A.M. and Clare, G.M. (1991) A common-sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Resonance*, **95**, 214–222.

Garrett, D.S., Kuszewski, J., Hancock, T.J., Lod, P.J., Vuister, G.W., Gronenborn, A.M. and Clare, G.M. (1994) The impact of direct refinement against three-bond HN-C α H coupling constants on protein structure determination by NMR. *J. Magn. Resonance*, **104**, 99–103.

Grindley, N.D.F. and Leschziner, A.E. (1995) DNA transposition: from a black box to a color monitor. *Cell*, **83**, 1063–1066.

Harshey, R.M., Getzoff, E.D., Baldwin, D.L., Miller, J.L. and Chaconas, G. (1985) Primary structure of phage Mu transposase: homology to Mu repressor. *Proc. Natl Acad. Sci. USA*, **82**, 7676–7680.

- Hu, J.-S. and Bax, A. (1997a) Determination of ϕ and χ_1 angles in proteins from ^{13}C - ^{13}C J couplings measured by three-dimensional heteronuclear NMR: how planar is the peptide bond. *J. Am. Chem. Soc.*, **27**, 6360–6368.
- Hu, J.-S. and Bax, A. (1997b) χ_1 angle information from a simple two-dimensional NMR experiment which identifies trans $^3\text{J}_{\text{NCY}}$ couplings in isotopically enriched proteins. *J. Biomol. NMR*, **9**, 323–328.
- Hu, J.-S., Grzesiek, S. and Bax, A. (1997) Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $\text{J}_{\text{C}^{\alpha}\text{C}^{\beta}}$ and J_{NCY} couplings. *J. Am. Chem. Soc.*, **119**, 1803–1804.
- Kim, K. and Harshey, R.M. (1995) Mutational analysis of the *att* DNA-binding domain of phage Mu transposase. *Nucleic Acids Res.*, **23**, 3937–3943.
- Konradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 52–55.
- Krukutlis, R., Welty, D.J. and Nakai, H. (1996) ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *EMBO J.*, **15**, 935–944.
- Kuo, C.-F., Zou, A., Jayaram, M., Getzoff, E. and Harshey, R.M. (1991) DNA–protein complexes during attachment-site synapsis in Mu DNA transposition. *EMBO J.*, **10**, 1585–1591.
- Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, G.M. (1995) The impact of direct refinement against $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts on protein structure determination by NMR. *J. Magn. Resonance*, **B106**, 92–96.
- Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.*, **5**, 1067–1080.
- Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1997) Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Resonance*, **125**, 171–177.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lavoie, B.D. and Chaconas, G. (1995) Transposition of phage Mu DNA. *Curr. Top. Microbiol. Immunol.*, **204**, 83–102.
- Leung, P.C. and Harshey, R.M. (1991) Two mutations of phage Mu transposase that affect strand transfer or interactions with B protein lie in distinct polypeptide domains. *J. Mol. Biol.*, **219**, 189–199.
- Leung, P.C., Trepow, D.B. and Harshey, R.M. (1989) Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature*, **338**, 656–658.
- Levchenko, I., Luo, L. and Baker, T.A. (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev.*, **9**, 2399–2408.
- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G. and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, **271**, 1247–1254.
- Li, T., Stark, M.R., Johnson, A.D. and Wolberger, C. (1995) Crystal structure of the MATA1/MAT α 2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.
- Lim, W.A., Sauer, R.T. and Lander, A.D. (1991) Analysis of DNA–protein interactions by affinity coelectrophoresis. *Methods Enzymol.*, **208**, 196–210.
- Mizuuchi, K. (1992) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.*, **61**, 1011–1051.
- Mizuuchi, M. and 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.
- Nakayama, C., Trepow, D.B. and Harshey, R.M. (1987) Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. *Proc. Natl Acad. Sci. USA*, **84**, 1809–1813.
- Nicholls, A., Sharp, K. and Honig, B. (1991) Protein folding and association: insights from interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.*, **11**, 281–296.
- Nilges, M. (1993) A calculational strategy for the structure determination of symmetric dimers by ^1H -NMR. *Proteins: Struct., Funct., Genet.*, **17**, 297–309.
- Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance geometry–dynamical simulated annealing calculations. *FEBS Lett.*, **229**, 317–323.
- Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.
- Rice, P.A. and Mizuuchi, K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209–220.
- Schevitz, R.W., Otwinowski, Z., Jochimiak, A., Lawson, C.L. and Sigler, P.B. (1985) The three-dimensional structure of trp repressor. *Nature*, **317**, 782–786.
- Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Crystal structure of a CAP–DNA complex; the DNA is bent by 90°. *Science*, **253**, 1001–1007.
- Schumacher, M.A., Choi, K.Y., Zalkin, H. and Brennan, R.G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science*, **266**, 763–770.
- Spera, S. and Bax, A. (1991) An empirical correlation between protein backbone conformation and C α and C β chemical shifts. *J. Am. Chem. Soc.*, **113**, 5490–5492.
- Spronk, C.A.E.M., Slijper, M., van Boom, J.H., Kaptein, R. and Boelens, R. (1996) Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. *Nature Struct. Biol.*, **3**, 916–919.
- Surette, M.G., Buch, S.J. and Chaconas, G. (1987) Transpososomes: stable protein–DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell*, **49**, 253–262.
- Surette, M.G., Lavoie, B.D. and Chaconas, G. (1989) Action at a distance in Mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). *EMBO J.*, **8**, 3483–3489.
- Vuister, G.W., Kim, S.-J., Wu, C. and Bax, A. (1994) 2D and 3D NMR study of phenylalanine residues in proteins by reverse isotope labeling. *J. Am. Chem. Soc.*, **116**, 9206–9210.
- Wu, Z. and Chaconas, G. (1994) Characterization of a region in phage Mu transposase that is involved in interaction with MuB protein. *J. Chem. Biol.*, **269**, 28829–28833.
- Wu, Z. and Chaconas, G. (1995) A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. *EMBO J.*, **14**, 3835–3843.
- Yang, W. and Steitz, T.A. (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*, **3**, 131–134.
- Zou, A., Leung, P.C. and Harshey, R.M. (1991) Transposase contacts with Mu DNA ends. *J. Biol. Chem.*, **266**, 20476–20482.

Received on August 8, 1997; revised on October 7, 1997