Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate

Patrick Van Roey¹,
Christopher A. Waddling², Kristin M. Fox³,
Marlene Belfort and Victoria Derbyshire

Wadsworth Center, PO Box 509, Albany, NY 12201-0509 and

³Department of Chemistry, Union College, Schenectady,
NY 12308-3161, USA

²Present address: Howard Hughes Medical Institute, University of
California at San Francisco, 513 Parnassus Avenue, San Francisco,
CA 94143, USA

¹Corresponding author
e-mail: vanroey@wadsworth.org

I-TevI is a site-specific, sequence-tolerant intron endonuclease. The crystal structure of the DNA-binding domain of I-TevI complexed with the 20 bp primary binding region of its DNA target reveals an unusually extended structure composed of three subdomains: a Zn finger, an elongated segment containing a minor groove-binding α -helix, and a helix–turn–helix. The protein wraps around the DNA, mostly following the minor groove, contacting the phosphate backbone along the full length of the duplex. Surprisingly, while the minor groove-binding helix and the helix–turn–helix subdomain make hydrophobic contacts, the few base-specific hydrogen bonds occur in segments that lack secondary structure and flank the intron insertion site. The multiple base-specific interactions over a long segment of the substrate are consistent with the observed high site specificity in spite of sequence tolerance, while the modular composition of the domain is pertinent to the evolution of homing endonucleases.

Keywords: crystal structure/endonuclease/
helix–turn–helix/minor groove/Zn finger

Introduction

Intron-encoded endonucleases are proteins that promote the first step in the mobility of the intron at the DNA level (Belfort and Roberts, 1997). They recognize and cleave an intronless allele of their cognate gene, initiating a replicative gene conversion event that results in the recipient allele also becoming intron-plus. These enzymes are therefore termed homing endonucleases and are grouped into a number of families based on the presence of conserved sequence elements. These are the LAGLIDADG, GIY-YIG, H-N-H and His-Cys box families (Belfort *et al.*, 2001).

I-TevI, the group I intron-encoded endonuclease of the *td* gene of bacteriophage T4, is the best studied member of the GIY-YIG family (Kowalski *et al.*, 1999). The 28 kDa enzyme specifically recognizes its lengthy DNA substrate, or homing site, as a monomer (Figure 1A; Mueller *et al.*, 1995), exhibiting a high degree of sequence tolerance

(Bryk *et al.*, 1993). No single nucleotide in the 37 bp target is essential for binding and cleavage, and many multiple substitutions are well tolerated (Bryk *et al.*, 1993, 1995). Consistent with this sequence tolerance, ethylation and methylation interference studies indicated that most of the protein–DNA contacts are via the minor groove and the phosphate backbone (Figure 1B) (Bryk *et al.*, 1993). The primary binding region of the enzyme is ~20 bp in length, spanning the intron insertion site (IS), with a second region of contact close to the cleavage site (CS), 23–25 bp upstream of the IS (Figure 1A). I-TevI demonstrates remarkable flexibility, recognizing and cleaving homing site derivatives with large deletions (up to 16 bp) and insertions (up to 5 bp) between the IS and CS (Bryk *et al.*, 1995).

The two-domain nature of the homing site is mirrored by the structure of the enzyme (Figure 1A). I-TevI consists of two functionally distinct domains: an N-terminal catalytic domain and a C-terminal DNA-binding domain, separated by a long flexible linker (Derbyshire *et al.*, 1997). The catalytic domain contains the GIY-YIG sequence module and forms a discrete structural domain contained within the first 92 amino acids of the protein (Kowalski *et al.*, 1999). The DNA-binding domain, contained within residues 130–245, contacts the primary binding region of the homing site (IS region). This domain binds with the same affinity as full-length I-TevI, suggesting that it includes most, if not all, binding determinants of the enzyme (Derbyshire *et al.*, 1997). However, the basis for sequence-tolerant, site-specific binding is unresolved. The phenomenon has been hypothesized to stem from an overspecification of direct binding determinants and/or from an indirect readout of some unusual structural feature of the DNA (Bryk *et al.*, 1993).

Here we present the crystal structure of the DNA-binding domain of I-TevI in complex with a 20 bp duplex DNA that corresponds to the primary binding region of the enzyme. The structure is striking for the lengthy, intertwined nature of the protein–DNA complex, and for the modularity of the protein, consisting of three distinct subdomains, with each exhibiting unusual properties. Additionally, the structure leads to insight into the site specificity/sequence tolerance conundrum, while illuminating novel aspects of the evolution of homing endonucleases.

Results

Structure determination

The DNA-binding domain of I-TevI, in the form of a 116 amino acid polypeptide (residues 130–245) (Derbyshire *et al.*, 1997), was crystallized with a 20 bp DNA duplex extended by a one-base overhang at the 5' end

Fig. 1. I-TevI-DNA interaction. (A) Cartoon of the two-domain structure of I-TevI with its DNA homing site. (B) Protein-DNA contacts. The sequence of the DNA fragment used in the crystal structure determination with interaction sites as identified from ethylation and methylation protection assays indicated as arrows and circles (open, weak; closed, strong), respectively. Bars below indicate the DNA locations of the three types of protein-DNA contacts in the structure, showing the consistency with the biochemical data. P and H correspond to regions involved in phosphate-backbone and hydrophobic contacts, respectively, while B refers to base-specific hydrogen-bonding contacts. IS and CS indicate the insertion and cleavage sites, respectively.

of each strand. The DNA comprises the primary binding region of the homing site, as defined by footprinting experiments (Figure 1B; Bryk *et al.*, 1993), except that the overhanging base of the antisense strand is an adenine, permitting base pairing with the overhanging thymine of the sense strand of a symmetry-related molecule. The structure was determined by single isomorphous replacement/anomalous scattering methods, using a DNA molecule in which four thymines were replaced by 5-iodouridines as the heavy atom derivative, and refined at 2.2 Å resolution. Figure 2 shows the electron density map for a representative section of the structure, corresponding to the Zn finger subdomain. Table I lists the data collection and refinement statistics. The final protein model consists of residues 149–244, with no electron density observed for residues 130–148 or for residue 245. This defines His148 as the C-terminal residue of the flexible linker between the catalytic and DNA-binding domains.

A trimodular wrap-around DNA-binding domain

The protein assumes a remarkably extended structure consisting of three identifiable DNA recognition subdomains: a Zn finger (residues 149–167); an elongated subdomain (residues 168–203) that includes a minor groove-binding α -helix (residues 183–194); and a helix-turn-helix subdomain (residues 204–244) (Figure 3A and B). The protein molecule winds itself around the DNA along its full two turns (Figure 3), resulting in ~50% of the surfaces of both the protein and the DNA being rendered inaccessible to solvent molecules (Figure 3C). The DNA adopts a regular B-DNA conformation except for a widening of the minor groove at bases 7–8 and 13–14 of the sense strand, with a corresponding increase in base wobble in those regions (Figure 4A). DNA molecules translated along the (*a*,*c*)-diagonal of the unit cell interact through base pairing of the overhanging bases, resulting in a pseudo-continuous DNA molecule without significant distortion at the interface.

Fig. 2. Electron density map for the Zn finger subdomain. Stereodisplay showing the final ($2F_o - F_c$) map for residues 149–168, contoured at the 1.5σ level. The Zn ion is shown as a blue sphere in the lower center of the view and the 10 residue loop oriented towards the top of the figure. The figure was prepared using SETOR (Evans, 1993).

Table I. Crystallographic data

	Native	Iodo derivative
Data collection		
resolution (Å)	30–2.2	30–2.5
completeness (%)	99.4 (96.5)	87.9 (82.5)
redundancy	5.02	2.4
R_{sym}	0.043 (0.389)	0.070 (0.238)
$\langle I/\sigma(I) \rangle$	9.00 (3.8)	10.8 (3.4)
Refinement		
resolution (Å)	20.0–2.2	
R/R_{free}	0.215/0.248	
Deviation from ideality		
bond length (Å)	0.009	
bond angle (°)	1.46	
dihedral (°)	20.2	
Atoms/average B (Å ²)		
protein	783/33.5	
DNA	855/39.5	
Zn	1/35.7	
water	154/44.4	

Values in parentheses represent those for the highest resolution shell.

A novel Zn finger subdomain

The Zn finger of I-TevI has the unusual sequence CXCX₁₀CXXC, and its presence was therefore not predicted by computational analysis. Zn fingers that include two cysteines separated by a single amino acid are rare and do not occur in other GIY-YIG proteins (Kowalski *et al.*, 1999). Similar Zn fingers are thought to occur in the β' -subunits of bacterial RNA polymerases, which contain the sequence CXCX₁₁CXXC, but these have not been structurally characterized.

In I-TevI, the Zn finger (Figures 2, 3A and B) consists of two single-turn helices that include the cysteines with a 10 residue intervening loop. This loop lacks secondary structure other than a β -turn. The Zn finger interacts with the DNA through two hydrogen bonds, between the main chain nitrogen atoms of residues Tyr162 and Ser165 and the phosphate backbone (Table II). No base-specific contacts are observed, but O ϵ 1 of Gln158 is 3.5 Å from N1 of ADE31, the overhanging base of the translationally related duplex that forms a base pair with THY1 (Figure 1B). Although this is slightly beyond the normal

range for hydrogen bonds, it is conceivable that this would be a base-specific contact in a complex with a longer natural DNA fragment. Regardless of this, the non-specific DNA binding of the Zn finger subdomain is consistent with evidence that it can be deleted without a reduction of binding specificity or binding energy by the remainder of the DNA-binding domain (A. Dean, V. Derbyshire and M. Belfort, unpublished). Together, these results suggest that the Zn finger functions in a capacity other than interaction with the primary DNA-binding site.

The elongated domain contains all base-specific hydrogen-bonding contacts

The central, elongated section of the DNA-binding domain, residues 168–203, closely contacts the DNA in the minor groove over a 12 bp segment (Figures 3A, 4A and B). Residues 168–183 (Figure 4A) lack secondary structure, but include three β -turns: residues 171–174 (type II' β -turn), residues 172–175 (type I) and residues 177–180 (type II). Residues 184–195 (Figure 4B) form a three-turn α -helix, which is followed by another segment that lacks secondary structure (residues 196–203). All base-specific hydrogen-bonding contacts are within the elongated section and in the regions that lack secondary structure. They involve five residues preceding the α -helix and one residue following it (Figure 4A and B; Table II), contacting two regions of the DNA that flank the intron insertion site (Figure 1B). The base-specific hydrogen-bonding amino acids are separated into three distinct regions by hydrophobic residues (Figure 4A and B). Phe177 separates residues Arg168–Ser176, which interact with bp 5–7, from His182, interacting with bp 8 and 9. His182, in turn, is separated from Asn201 by the minor groove-binding helix, which has a hydrophobic surface consisting of the side chains of Thr186, Ile190 and Met194, closest to the DNA.

This elongated segment of the protein distorts the DNA by widening the minor groove, increasing the P–P distances by up to 4 Å (Figures 3A, 4A and B). The groove widening is greatest at bp 7–46 and 8–45, where Asn175, Phe178 and His182 are in hydrogen-bonding contact with the bases in the minor groove, and at bp 13–40 and 14–39 where Asn201 forms hydrogen bonds with bases from both strands of the DNA.

Fig. 3. Three-dimensional structure of the complex of the DNA-binding domain of I-TevI with its substrate. The complex is shown (A) perpendicular to the DNA axis and (B) along the DNA axis. Distortions to the DNA are limited to widening of the minor groove. (C) Space-filling model of the structure of the complex, showing the continuous tight association between the two molecules. Protein and DNA carbon atoms are colored green and gray, respectively. Figures 3 and 4 were prepared with Molscript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

The helix–turn–helix domain makes backbone and hydrophobic interactions

The 45 residue helix–turn–helix subdomain constitutes the only part of the molecule with a true globular fold, with $\beta\beta\alpha\alpha\beta$ topology (Figure 3A). It consists of a three-stranded antiparallel β -sheet, with the first strand at the center, and flanked on one side by the α -helices. Database searches with DALI (Holm and Sander, 1993) and SCOP (Murzin *et al.*, 1995) failed to identify another protein with a similar fold, although topologically the domain shows some relationship to the helix–turn–helix domains of

Table II. Hydrogen-bonding contacts		
Protein	DNA	Distance (Å)
Zn finger		
Tyr162 N	GUA50 O2P	2.8
Ser165 N	ADE51 O2P	2.9
Elongated segment		
Arg168 N η 2	ADE48 N3	3.2
Arg168 N η 1	ADE49 O4'	2.9
Arg170 N η 1	THY5 O2	2.8
Asn175 O δ 1	GUA7 N3	3.0
Ser176 O γ	CYT47 O2	2.8
Phe178 N	THY9 O2P	2.8
His182 N ϵ 2	THY9 O2	2.9
His182 N ϵ 2	CYT45 O2	3.1
Minor groove-binding helix		
Ser191 O γ	ADE12 O2P	2.6
Elongated segment		
Val195 N	CYT13 O2P	2.8
Ser200 O γ	THY41 O3'	3.2
Asn201 N δ 2	CYT14 O2	2.8
Asn201 O δ 1	GUA40 N2	3.1
Lys203 N	GUA15 O2P	2.8
Lys203 N ζ	THY16 O2P	2.7
Helix–turn–helix		
Ala216 N	GUA15 O1P	2.8
Arg220 N η 2	CYT14 O2P	2.7
Ser225 N	THY34 O1P	3.0
Ser225 O γ	THY34 O1P	2.6
Thr230 O γ 1	THY16 O1P	2.5
Tyr231 O ζ	ADE32 O1P	3.2
Arg232 N η 2	THY33 O1P	3.2
Lys237 N ζ	ADE32 O2P	3.2
Tyr242 O η	THY16 O1P	2.5

transcription factors, such as E2F-4 (Zheng *et al.*, 1999). The helices form a traditional helix–turn–helix motif with an angle of 108° between the helices, and with the second helix inserted into the major groove. Although the subdomain makes nine hydrogen-bonding contacts with the DNA, involving residues of both helices and of the third β -strand, atypically, all are to the phosphate backbone (Table II). The DNA-proximal surface of the helix inserted into the major groove is highly hydrophobic, consisting primarily of the side chains of Leu228, Thr230 and Tyr231 and the α -carbon of Gly227, and faces a mostly hydrophobic DNA surface due to the presence of the C5-methyl groups of thymine 16, 17, 18, 33 and 34 (Figure 4C). However, only the contact between the phenyl ring of Tyr231 and the methyl group of THY33 is within van der Waals distance. This constitutes the only direct contact, although there is a water-bridged hydrogen bond between the main chain nitrogen of Gly227 and N7 of ADE35. This helix–turn–helix subdomain is therefore highly unusual given the absence of base-specific hydrogen bonds, but with specificity resulting from hydrophobic surface interactions.

Discussion

A modular structure

The DNA-binding domain of I-TevI is remarkable in its extended structure that wraps around 20 bp of DNA. The domain is assembled from three distinguishable subdomains, each of which is individually related to DNA-recognition motifs found over a wide range of

DNA-binding proteins. These subdomains each have unique characteristics: the Zn finger subdomain has no apparent role in DNA binding specificity or affinity, while the elongated segment with its minor groove-binding helix is notable for its specific hydrogen-bonding interactions, and the helix–turn–helix subdomain is atypical for making

only phosphate and hydrophobic contacts. Furthermore, the three subdomains appear to represent minimal sized DNA-binding modules, with the Zn finger and helix–turn–helix subdomains representing particularly small structures of their respective types.

Correspondence between biochemical and structural results

Hydrogen-bonding contacts between the protein and the phosphate backbone are observed throughout the full length of the DNA, starting at Tyr162 in the Zn finger (Table II). However, most of these non-specific contacts are concentrated in the helix–turn–helix subdomain (Figures 3A and 4C). These data are consistent with the results of ethylation interference experiments, which indicated that I-TevI makes contacts to its DNA substrate via the phosphate backbone (Figure 1B; Bryk *et al.*, 1993). In particular, modification of the phosphates of THY16, ADE32, THY33 and THY34 was shown to have a strong effect on protein binding, these nucleotides being precisely those contacted most closely by residues in the helix–turn–helix subdomain.

As indicated previously, base-specific hydrogen-bonding interactions only occur in the extended regions of the protein (Figures 1B, 4A and B; Table II). These include six hydrogen bonds from residues in the extended segment between the Zn finger and the minor groove-binding helix with bases within a 5 bp section, THY5–THY9, and to two contacts from Asn201, which is after the helix, to the bases of GUA40 and CYT14 from two adjacent base pairs. These data are highly consistent with methylation interference experiments (Bryk *et al.*, 1993). In particular, Arg168 directly contacts the N3 of ADE48, a base shown to be sensitive to modification.

Interestingly, biochemical data for the full-length protein suggest that there are contacts to both the phosphate group and major groove of GUA50, while in the present complex only a contact between the Zn finger residue Tyr162 and the phosphate backbone is observed in this area and the major groove is exposed. However, a slight change in the position of the Zn finger domain or in the DNA conformation, which would also be required to bring Gln158 into contact with the DNA, could limit the accessibility of the major groove. However, given the dispensability of the Zn finger for DNA binding (A.Dean, V.Derbyshire and M.Belfort, unpublished), its function

Fig. 4. Individual I-TevI–DNA contact regions. (A) Elongated segment between the Zn finger and the minor groove-binding helix. The protein lacks secondary structure but residues 170–180 form a twisted structure that widens the minor groove. Base-specific hydrogen-bonding contacts (red dotted lines) throughout this segment are interrupted by the hydrophobic insertion of the phenyl ring of Phe177 (yellow). (B) The minor groove-binding α -helix. Only one hydrogen-bonding contact is seen in the helix (Ser191 to the phosphate backbone), and the surface close to the DNA consists of three hydrophobic residues. The section between the helix and the helix–turn–helix subdomain includes the remaining base-specific hydrogen-bonding contacts, Asn201 to GUA40 and CYT14. (C) The helix–turn–helix subdomain inserts its second helix into the major groove. Several of the residues of this helix make hydrogen-bonding contacts to the phosphate backbone (red dotted lines), but the surface of the helix adjacent to the DNA is mostly hydrophobic and matches the hydrophobic surface of the DNA, which presents the C5-methyl groups of thymidines 16, 17, 18, 33 and 34. The closest hydrophobic contacts are shown as blue dotted lines.

may well be related more directly to the activity of the full-length protein, in a step subsequent to that of initial DNA binding mediated by the C-terminal domain.

Site specificity in the face of sequence tolerance

It has been suggested that the sequence tolerance of I-TevI might be due to structural properties of the DNA and/or redundant DNA contacts (Bryk *et al.*, 1993). Clearly, the DNA in the co-crystal is structurally unremarkable (Figure 3). Indeed, distortions are limited to widening of the minor groove at the sites of base-specific hydrogen-bonding interaction and are therefore more likely to be a consequence of, rather than a signal for, I-TevI binding. In contrast, the multiple base-specific hydrogen bonds throughout the elongated region, as well as the hydrophobic interactions of the minor groove-binding helix and of the helix–turn–helix subdomain, provide for the high site specificity for the homing site. However, the fact that the most specific interactions occur within the elongated, and conformationally most adaptable, region of the protein is consistent with the sequence tolerance of I-TevI. This aspect of the DNA recognition by the DNA-binding domain of I-TevI appears to be analogous to the homeo-domain protein Pax6, which consists of two helix–turn–helix domains connected by a 17 residue elongated linker located in the minor groove (Xu *et al.*, 1999). As for I-TevI, several residues in the Pax6 linker make base-specific contacts with the DNA, and its C-terminal helix–turn–helix domain interacts with bases in the major groove through hydrophobic contacts and water-bridged hydrogen bonds. However, the elongated segment and the helix–turn–helix subdomain of the DNA-binding domain of I-TevI make many more contacts with the DNA and over a longer stretch of DNA than the corresponding regions of Pax6.

Functional and evolutionary insights

It is noteworthy that the central elongated region of the I-TevI DNA-binding domain is the richest in base-specific interactions and also that with the greatest impact on the DNA structure (Figure 4A). Significantly, this region spans the intron insertion site, the junction sequence between the two exons. It is specifically this junction that distinguishes the intronless target from the intron-containing donor allele and that dictates selective cleavage of the DNA of the recipient allele. It is, therefore, likely that the base-specific interactions evolved to flank this discriminatory site. Furthermore, the fact that these amino acids are in the extended regions suggests that they are afforded conformational flexibility, and/or are free to evolve rapidly, to recognize these junction sequences because they are not constrained by participation in a folded element that could give rise to conflicting interactions.

T-even phage DNA, presumably the natural substrate of I-TevI, is modified. T4 DNA contains glucosylated 5-hydroxymethylcytosine, resulting in a bulky adduct in the major groove. Because I-TevI binds DNA mostly in the minor groove, this modification should not affect most of the contacts. I-TevI would therefore appear to have evolved to avoid interactions with glucosylated hydroxymethyl groups, probably to maximize the range of natural substrates. Interestingly, the helix–turn–helix subdomain binds in the major groove, but this section of the homing

site is devoid of cytidines. While there are no base-specific hydrogen-bonding contacts in this region, the major groove binding of the subdomain confers some sequence constraints by precluding cytidines in this area and by selecting for thymidine bases through the interaction with the hydrophobic surface. Therefore, while the helix–turn–helix subdomain does not impose sequence specificity through direct hydrogen-bonding contacts, it does play a role in selecting for an AT-rich DNA substrate. Accordingly, in a randomization study of the homing site, I-TevI was less tolerant of mutations in this AT-rich region than elsewhere in its primary DNA-binding site (Bryk *et al.*, 1993).

We argued previously that the GIY-YIG domain is a catalytic cartridge that is joined to a variety of different DNA-binding domains to expand the enzyme's substrate repertoire (Derbyshire *et al.*, 1997). Consistent with this, sequence comparisons of GIY-YIG endonucleases demonstrated that similarities were limited to the catalytic domain (Cummings *et al.*, 1989; Kowalski *et al.*, 1999). However, a newly identified GIY-YIG endonuclease, I-BmoI, shares significant sequence similarity with I-TevI along the full length of the proteins (D. Edgell and D. Shub, personal communication). Comparison of the amino acid sequences of I-BmoI (DDBJ/EMBL/GenBank accession No. AF321518) and I-TevI in light of the current structure highlights the modular nature of these proteins. Both proteins have very similar C-terminal helix–turn–helix motifs, but the Zn finger subdomain is absent in I-BmoI or other GIY-YIG family members. However, I-BmoI appears to contain two copies of a module that has significant sequence similarities to the elongated segment of the I-TevI DNA-binding domain, suggesting the presence of two elongated minor groove-binding segments in I-BmoI. Thus, it would appear that this family of enzymes can rapidly evolve new specificities using two different strategies. The first is the internal flexibility offered by having the residues involved in base-specific interactions located in the extended regions. The second is the shuffling of DNA-binding modules that collectively interact with lengthy recognition sequences. In the process, multiple substrate contacts distributed over the length of the target site overcome the low information content of the minor groove and the phosphoribose backbone to promote specificity.

Materials and methods

Crystallization

The DNA-binding domain of I-TevI was expressed in *Escherichia coli* and purified as previously described (Derbyshire *et al.*, 1997). Synthetic DNA was purchased from Operon (Alameda, CA). Protein and DNA were diluted separately to low concentration, typically 0.2 µM in 20 ml, in a buffer containing 0.1 M MES pH 6.5, 0.3 M sodium formate, 0.2 M sodium chloride and 10% glycerol. After combining equimolar amounts of the two solutions, the protein–DNA complex was concentrated in Amicon stirred-cell concentrators with a 10 000 molecular weight membrane to a final protein concentration of ~4 mg/ml. Problems with aggregation required all crystallization experiments to be performed within 48 h of the completion of the protein preparation and within 6 h from when the protein–DNA complex was concentrated. The complex was crystallized using the hanging drop method at 10°C with drops containing 3 µl of protein–DNA solution and 2 µl of the well solution, which consisted of 18% PEG 3350, 15% glycerol and 0.06 M MES pH 6.5. Crystals grew within 48 h but continued to increase in size for

~1 week. The crystals are monoclinic, space group $P2_1$ with cell parameters $a = 55.14$ Å, $b = 65.21$ Å, $c = 43.67$ Å, $\beta = 93.1^\circ$.

Structure determination

The structure was determined by single isomorphous replacement/anomalous scattering phasing using a DNA in which the thymines at positions 5, 18, 33 and 34 were replaced by 5-iodouridine as the derivative. Data were measured at NSLS, beamline X12C, using a Brandeis 1K CCD detector and processed with DENZO and Scalepack (Otwinowski and Minor, 1997). The derivative data were measured at a wavelength of 1.55 Å to increase the anomalous signal, and native data were measured at 1.2 Å. Initial phases were obtained using the program SOLVE (Terwilliger and Berendzen, 1999), figure-of-merit 0.62, and improved by solvent flattening using DM (CCP4, 1994). The model was built with O (Jones *et al.*, 1991) and refined with CNS, version 0.9a (Brünger *et al.*, 1998). Ninety-two percent of the protein residues are in the most favored region of the Ramachandran plot, with the remaining 8% in the additionally allowed region (Laskowski *et al.*, 1993). The side chains of residues 152, 176, 226, 234 and 236 are disordered and have been refined with two conformations. The atomic coordinates and observed structure factor amplitudes are available from the Protein Data Bank (entry code 1I3J).

Acknowledgements

We thank John Dansereau for expert assistance with protein expression and purification, David Shub, Cheryl Eifert and David Edgell for providing the I-BmoI sequence prior to publication, and Susan Baxter, Amy Dean, David Edgell and Joe Kowalski for insightful comments on the manuscript. Research was supported by NIH grants GM56966 (P.V.R.), GM39422 and GM44844 (M.B.). The facilities of beamline X12C of NSLS are supported through grants from the DOE and the NIH.

References

- Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
- Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B. and Lambowitz, A.M. (2001) Mobile introns: pathways and proteins. In Craig, N., Craige, R., Gellert, M. and Lambowitz, A. (eds), *Mobile DNA II*. ASM Press, in press.
- Brünger, A.T. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Bryk, M., Quirk, S.M., Mueller, J.E., Loizos, N., Lawrence, C. and Belfort, M. (1993) The *td* intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J.*, **12**, 2141–2149.
- Bryk, M., Belisle, M., Mueller, J.E. and Belfort, M. (1995) Selection of a remote cleavage site by I-TevI, the *td* intron-encoded endonuclease. *J. Mol. Biol.*, **247**, 197–201.
- CCP4 (1994) The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Cummings, D.J., Michel, F. and McNally, K.L. (1989) DNA sequence analysis of the apocytochrome *b* gene of *Podospora anserina*: a new family of intronic open reading frame. *Curr. Genet.*, **16**, 407–418.
- Derbyshire, V., Kowalski, J.C., Dansereau, J.T., Hauer, C.R. and Belfort, M. (1997) Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.*, **265**, 494–506.
- Evans, S.V. (1993) SETOR: Hardware-highlighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics*, **11**, 134–138.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron-density maps and the location of errors in these maps. *Acta Crystallogr. A*, **47**, 110–119.
- Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Pietrovski, S., Baxter, S.M. and Derbyshire, V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res.*, **27**, 2115–2125.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed

and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

- Laskowski, R.A., McArthur, 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.
- Merritt, E.A. and Bacon, D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Mueller, J.E., Smith, D., Bryk, M. and Belfort, M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J.*, **14**, 5724–5735.
- Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Terwilliger, T.C. and Berendzen, J. (1999) Automated structure solution for MIR and MAD. *Acta Crystallogr. D*, **55**, 849–861.
- Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev.*, **13**, 1263–1275.
- Zheng, N., Fraenkel, E., Pabo, C.O. and Pavletich, N.P. (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.*, **13**, 666–674.

Received April 2, 2001; revised and accepted May 29, 2001