Nucleic Acids Research, 1999, Vol. 27, No. 10 **2115–2125**

Configuration of the catalytic GIY-YIG domain of intron endonuclease I-*Tev*I: coincidence of computational and molecular findings

Joseph C. Kowalski, Marlene Belfort*, Michelle A. Stapleton, Mathias Holpert, John T. Dansereau, Shmuel Pietrokovski[1,+], Susan M. Baxter and Victoria Derbyshire

Wadsworth Center, New York State Department of Health and School of Public Health, State University of New York at Albany, PO Box 22002, Albany, NY 12201-2002, USA and [1]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, WA 98109, USA

Received February 2, 1999; Revised March 17, 1999; Accepted March 26, 1999

ABSTRACT

I-*Tev*I is a member of the GIY-YIG family of homing endonucleases. It is folded into two structural and functional domains, an N-terminal catalytic domain and a C-terminal DNA-binding domain, separated by a flexible linker. In this study we have used genetic analyses, computational sequence analysis and NMR spectroscopy to define the configuration of the N-terminal domain and its relationship to the flexible linker. The catalytic domain is an αΛβ structure contained within the first 92 amino acids of the 245-amino acid protein followed by an unstructured linker. Remarkably, this structured domain corresponds precisely to the GIY-YIG module defined by sequence comparisons of 57 proteins including more than 30 newly reported members of the family. Although much of the unstructured linker is not essential for activity, residues 93–116 are required, raising the possibility that this region may adopt an alternate conformation upon DNA binding. Two invariant residues of the GIY-YIG module, Arg27 and Glu75, located in α-helices, have properties of catalytic residues. Furthermore, the GIY-YIG sequence elements for which the module is named form part of a three-stranded antiparallel β-sheet that is important for I-*Tev*I structure and function.

INTRODUCTION

Several group I self-splicing introns encode endonucleases that promote intron mobility. These endonucleases recognize and cleave intronless alleles of their host gene in the vicinity of the exon junction or homing site. The intron-containing allele is used as a template for repair synthesis resulting in acquisition of the intron. This process, which has been observed in eukaryotes, archaea, bacteria and their phages, is known as intron homing and the intron-encoded endonucleases are known as homing endonucleases.

Homing endonucleases can be classified into four families based on conserved sequence motifs (1–3). The bacteriophage T4 *td* intron-encoded endonuclease, I-*Tev*I, the subject of this study, is a member of the GIY-YIG family. The family has been defined by the presence of the sequence GIY-10/11 amino acids-YIG (4) and some other conserved residues including an Arg residue approximately 8–10 residues downstream (2).

I-*Tev*I recognizes a large target site of ~40 bp in a sequence-tolerant fashion (Fig. 1A). Its primary binding site is at the intron insertion site (IS) yet the cleavage site (CS) is 23–25 bp upstream. I-*Tev*I binds as a monomer making extensive contacts via the minor groove and phosphate backbone of its DNA substrate (5,6). The enzyme is remarkable in that it can tolerate insertions of up to 5 bp and deletions of up to 16 bp between the IS and CS (7). These studies led to a model for I-*Tev*I as a hinged monomer with a flexible linker allowing the enzyme to effect distant cleavage (Fig. 1A).

Limited proteolysis experiments showed the presence of a sensitive region in the center of the 245-amino acid protein (8). This suggested that the N- and C-terminal portions of the enzyme were independently folded structural domains separated by a protease-accessible linker region. The C-terminal domain (130C, residues 130–245) could be expressed independently and bound to the homing site with the same affinity as full-length I-*Tev*I. Attempts to overexpress the N-terminal domain (125N, residues 1–125) in *Escherichia coli* were unsuccessful as the protein was extremely toxic. However, a catalytic mutant derivative with a substitution of the conserved Arg27 to Ala facilitated expression. The same mutation in full-length I-*Tev*I rendered the enzyme binding proficient but catalytically inactive. These data also suggested that the structural domains correspond to functional domains with the N-terminal catalytic domain being separated from the C-terminal DNA-binding domain by a linker, consistent with the previous model derived from homing-site studies (Fig. 1A). The restriction enzymes *Fok*I and *Nae*I can also be divided into DNA-binding and catalytic domains separated by a protease-sensitive linker (9,10). However, I-*Tev*I's extraordinary ability to accommodate variations in spacing between the primary DNA-binding site and CS is unique.

In this paper we focus on determining the configuration of the N-terminal catalytic domain and its relationship to the flexible linker region using a combination of genetic analyses, computational

*To whom correspondence should be addressed. Tel: +1 518 473 3345; Fax: +1 518 474 3181; Email: marlene.belfort@wadsworth.org

+Present address: Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

sequence analysis and secondary structure determination by NMR spectroscopy. Together, these approaches define the catalytic domain as an α/β structure, contained within the first 92 amino acids of the protein, which corresponds precisely to the computationally derived GIY-YIG module. The remaining amino acids of the domain are unstructured, but required for activity, suggesting that they may adopt an alternate conformation upon DNA binding. Two catalytic residues are located within α-helices, but the GIY-YIG sequence elements that give the motif its name are part of a three-stranded antiparallel β-sheet which, by analogy to other enzymes, may play a role in DNA binding close to the site of cleavage.

MATERIALS AND METHODS

Construction of I-*Tev*I linker mutants

The deletion mutants were generated by PCR using two mutagenic oligonucleotide primers. The PCR template was a pSP65 vector with the I-*Tev*I sequence inserted between the *Eco*RI and *Hin*dIII sites under the control of the SP6 promoter. The primers were complementary around a single location where silent base pair changes introduced a unique restriction site and the desired base pair deletions occurred. The PCR reaction resulted in a fragment containing the entire pSP65 and I-*Tev*I sequence with the desired deletions, as well as the unique restriction site at both ends. The purified PCR fragment was digested with the restriction enzyme corresponding to the newly introduced restriction site and ligated. The resultant plasmids were transformed into JM101, colonies were screened for the presence of the mutation by restriction digest and candidate plasmids were fully sequenced.

In vitro I-*Tev*I synthesis

I-*Tev*I derivatives were synthesized *in vitro* using wheat germ extracts with mRNA synthesized from the T7 promoter of the overexpression plasmids or the SP6 promoter of plasmids constructed for the deletion analysis with 25 μCi ^{35}S-methionine (11). Aliquots were fractionated on SDS/polyacrylamide gels and the relative amounts of I-*Tev*I derivatives determined by comparison of radioactive counts in each lane as measured using a Betascope (Betagen) direct radioactivity detector.

I-*Tev*I cleavage activity assays

Cleavage assays were performed as described previously (8). The extent of cleavage was determined by visual comparison of cleavage reactions using serial dilutions of wild-type enzyme, and by Masterscan software (Scanalytics) to analyze densitometric scans of the negatives of ethidium bromide-stained gels.

Computational sequence analysis

GIY-YIG protein sequences were aligned across short conserved sequence regions using the BlockMaker (12), MEME (13) and MACAW (14) programs as previously described (15,16). The resulting block multiple alignments were used as queries by the BLIMPS multiple-alignment sequence comparison program (12) to search the NCBI databases.

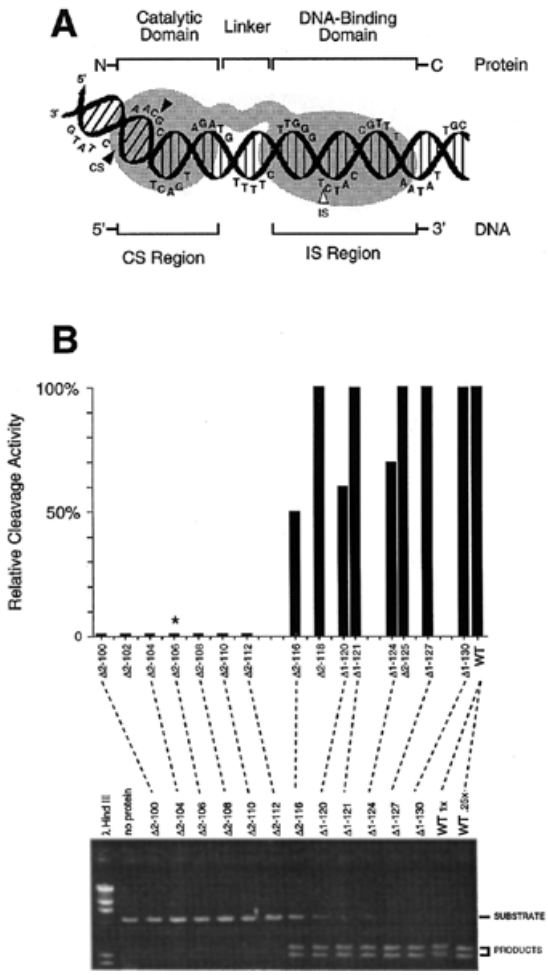


Figure 1. I-*Tev*I is a two-domain protein joined by a linker. (**A**) Structural and functional organization of I-*Tev*I and its homing site. CS, cleavage site; IS, insertion site. (**B**) Cleavage activity of the I-*Tev*I deletion derivatives. The relative activity compared to wild-type I-*Tev*I is shown in the bar chart. No activity is detectable for deletions N-terminal to residue 116, except for Δ2-106 (*) which has detectable activity, at 50–100-fold less than wild-type. The agarose gel shows standard cleavage assays with equal amounts of I-*Tev*I derivatives generated *in vitro*.

Site-directed mutagenesis and construction of overexpression plasmids

Amino acid substitutions were made by mutagenesis of pGEM-I-*Tev*I which contains the I-*Tev*I coding sequence cloned between the *Eco*RI and *Bam*HI sites of pGEM-3zf(–) (Promega) (8). Mutated fragments were transferred into pET-3a overexpression derivatives based on the T7 expression system as described previously (8).

Induction of expression of I-*Tev*I derivatives

To purify protein for biochemical studies, overexpression plasmids were introduced into host strain BL21(DE3)pLysS (17). Cells were grown as described previously (8). To purify uniformly labeled proteins for NMR analysis, induction was carried out in the same manner except that cells were grown in M9 minimal medium containing ^{15}N-labeled ammonium sulfate (Cambridge Isotope). For labeling of specific amino acids, overexpression

plasmids were introduced into appropriate auxotrophic host strains and were grown in defined minimal media supplemented with ^{15}N-labeled amino acid(s) (18).

Purification of I-*Tev*I derivatives

I-*Tev*I derivatives for NMR were purified essentially as described previously (8). Briefly, PEI (10%) was added to soluble cell lysates to remove nucleic acids and some contaminating proteins. In most cases a final concentration of 0.2% PEI was used to achieve optimal purification. As a final batch step in the purification, the proteins were ammonium-sulfate fractionated. These preparations were sufficiently pure for 2D HSQC analysis of specifically-labeled R27A-125N and R27A I-*Tev*I samples. Additional purification of the site-directed mutants was achieved by chromatography on a heparin–agarose column (Hi-Trap, Pharmacia Biotech) using FPLC. In addition, [*U*-^{15}N]-labeled R27A-125N samples were further purified on a Superdex 200 Hi-load 16/60 gel filtration column (Pharmacia Biotech). In all cases, samples for NMR were dialyzed against 20 mM $NaPO_4$, pH 6.0, 1 mM DTT, 0.01% NaN_3 and then concentrated using a centrifugal concentrator (Centricon).

NMR spectroscopy

NMR experiments were performed at 25°C on a Bruker Avance DRX 500 MHz spectrometer. Recycle times were usually 1.5 s. Data were processed and analyzed using FELIX 970 (MSI). 2D [^{1}H-^{15}N] HSQC (19,20) experiments resulted in final matrices of 1024 (t2) real points × 128 (t1) real points with spectral widths of 2500 Hz in the ^{1}H dimension (carrier frequency at 7.939 p.p.m.) and 1580 Hz in the ^{15}N dimension (carrier frequency at 118.0 p.p.m.). The 2D steady-state {^{1}H}^{15}N-heteronuclear NOE values (21) were determined in two separate experiments. The spectrum without NOEs was acquired with an off-resonance, presaturating ^{1}H pulse of 6 s. That pulse was on resonance in the spectrum with NOEs.

Hydrogen exchange experiments were carried out by exchanging a 1 mM 0.5 ml sample into buffer with 99.9% D_2O using a G-25 Sephadex spin column. 2D [^{1}H-^{15}N] HSQC spectra were taken every 14.65 min, starting 29.27 min after the exchange, for 2 days.

3D [^{1}H-^{15}N] NOESY-HSQC (22,23) was acquired with a mixing time of 125 ms and the final matrix was 256 × 128 × 1024 real points. The ^{1}H spectral width was 7000 Hz with a carrier frequency of 4.699 p.p.m., the ^{15}N spectral width was 1580 Hz with a carrier frequency of 118 p.p.m. and the amide ^{1}H spectral width was 2500 Hz with a carrier frequency of 7.939 p.p.m.

Two 3D [^{1}H-^{15}N] TOCSY-HSQC experiments (22,24) were taken at mixing times of 46 and 82.8 ms. The spectrum with the shorter mixing time resulted in a matrix of 128 × 128 × 1024 real points and the spectrum with the longer mixing time resulted in a matrix of 128 × 256 × 1024 real points. The spectral widths and carrier frequencies were the same as in the 3D [^{1}H-^{15}N] NOESY-HSQC above.

$^3J_{HNH\alpha}$ coupling constants were measured using a set of 10 J-modulated [^{1}H-^{15}N] HSQC experiments with delay times between 0.01 and 0.125 s (25,26). Volumes were measured and plotted as a function of delay time. Non-linear least squares fits (KaleidaGraph, Synergen) to the delay-dependent cross-peak volumes were used to obtain $^3J_{HNH\alpha}$ values (26). The final matrix was 512 × 256 points.

RESULTS

Genetic delineation of the catalytic domain/linker boundary

Limited proteolysis of I-*Tev*I had suggested the presence of an unstructured linker in the middle of the 245-amino acid protein (8). To define the extent of the linker and, more particularly, to determine the C-terminal limit of the catalytic domain, a systematic deletion analysis of the central region of I-*Tev*I was initiated. We reasoned that folded domains would be sensitive to the deletion of amino acids and would therefore yield inactive enzymes, but that the flexible linker between the two domains would likely be tolerant to deletions since there might not be defined secondary structural elements in this region.

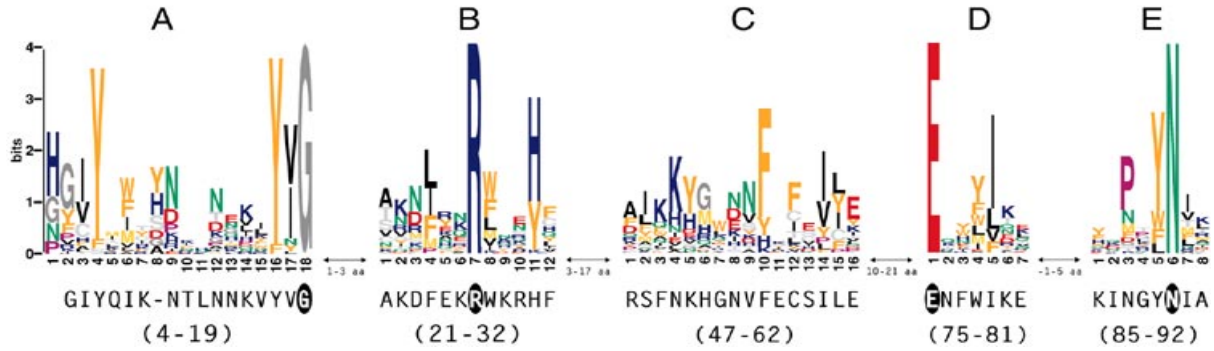
To delimit the linker/catalytic-domain boundary, a series of I-*Tev*I derivatives was constructed with one- or two-amino acid deletions from residues 100–130. The relative cleavage activity of these derivatives is shown in Figure 1B. Clearly, while retention of activity is consistent with a well-folded protein, loss of activity could be due to general structural perturbations caused by the deletions; however, all of the derivatives were similarly expressed as full-length proteins. Nomenclature of the deletion constructs is as follows: Δ1-121 is a one-amino acid deletion starting at residue 121 while Δ2-125 is a two-amino acid deletion from residue 125. Several of the derivatives retain cleavage activity indistinguishable from wild type in standard cleavage assays (i.e. Δ2-118, Δ1-121, Δ2-125, Δ1-127 and Δ1-130), while others retain significant cleavage activity (Δ2-116, Δ1-120 and Δ1-124). Notably, there is a sharp delineation in the activity of derivatives with deletions in residues N-terminal to amino acid 116, with these proteins having no appreciable cleavage activity. These data suggest that the functional catalytic domain is contained within the first 116 amino acids, while more C-terminal residues are part of the flexible linker.

Multiple sequence alignment for delineation of the catalytic domain

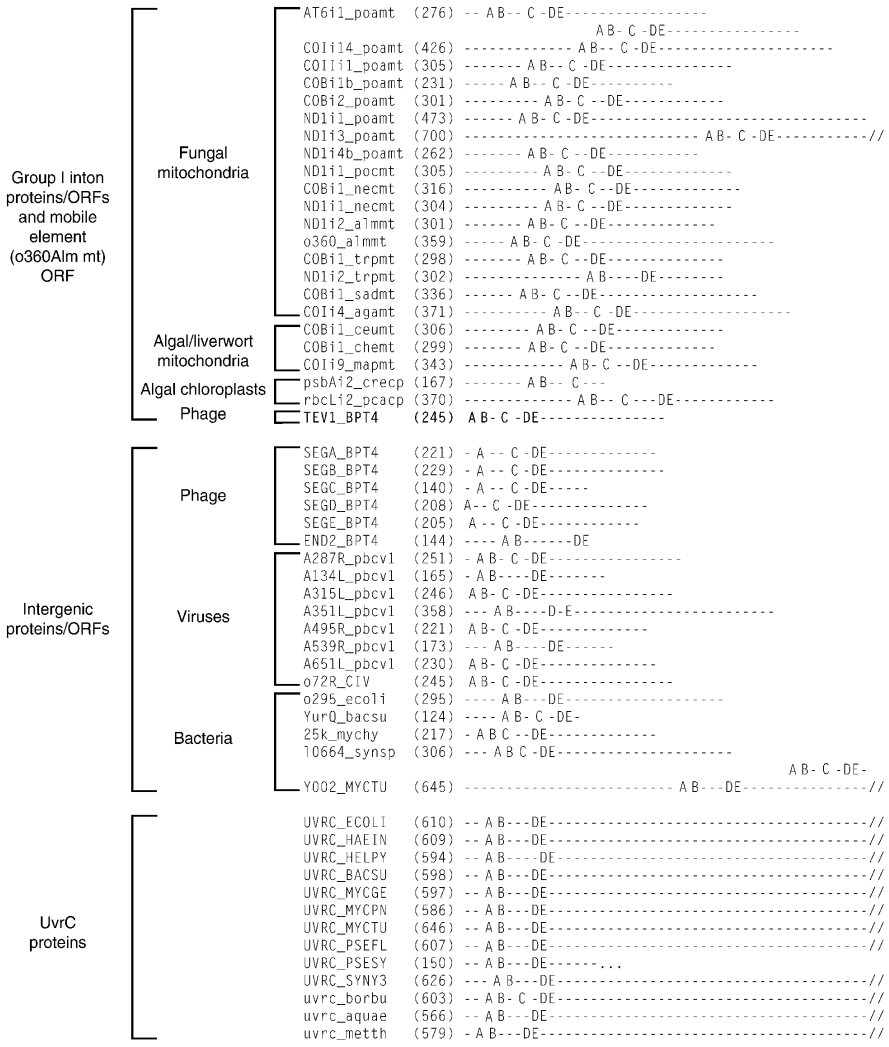
It has previously been recognized that a number of intron-encoded and intergenic endonucleases are related by conserved sequence motifs, including the characteristic GIY-YIG and some associated residues (2–4). A refined computational sequence analysis was undertaken to probe the extent of the conserved environment around these residues, to help define the nature and limits of the catalytic domain. Protein motifs were identified in GIY-YIG modules by local multiple alignments starting with known modules and using these to search iteratively for additional modules (15,27). The motifs were found to be specific for GIY-YIG modules and significantly conserved relative to what is expected by chance alignments (data not shown).

From this analysis, the GIY-YIG module is 70–100 amino acids long and contains five conserved sequence motifs (Fig. 2A). The first motif (A) is characterized by typical 'GIY' and 'YIG' patterns flanking a 10–11 amino acid segment. A second motif (B), that includes an invariant arginine, is separated by 1–3 amino acids from motif A. The third motif (C) is separated by 3–17 amino acids from the second and is somewhat less conserved than the other four. Following 10–21 non-conserved residues, the last two motifs (D and E) appear in tandem separated by up to 5 amino acids and they contain invariant glutamate and asparagine residues, respectively.

A


A
B
C
D
E
bits
GIYQIK-NTLNNKVYVG
(4-19)
AKDFEKRWKRHF
(21-32)
RSFNKHGNVFECSILE
(47-62)
ENFWIKE
(75-81)
KINGYNIA
(85-92)

B


Group I inton proteins/ORFs and mobile element (o360Alm mt) ORF
Fungal mitochondria
Algal/liverwort mitochondria
Algal chloroplasts
Phage
Intergenic proteins/ORFs
Phage
Viruses
Bacteria
UvrC proteins
AT6i1_poamt (276)
COIi14_poamt (426)
COIii1_poamt (305)
COBi1b_poamt (231)
COBi2_poamt (301)
ND1i1_poamt (473)
ND1i3_poamt (700)
ND1i4b_poamt (262)
ND1i1_pocmt (305)
COBi1_necmt (316)
ND1i1_necmt (304)
ND1i2_almmt (301)
o360_almmt (359)
COBi1_trpmt (298)
ND1i2_trpmt (302)
COBi1_sadmt (336)
COIi4_agamt (371)
COBi1_ceumt (306)
COBi1_chemt (299)
COIi9_mapmt (343)
psbAi2_crecp (167)
rbcLi2_pcacp (370)
TEV1_BPT4 (245)
SEGA_BPT4 (221)
SEGB_BPT4 (229)
SEGC_BPT4 (140)
SEGD_BPT4 (208)
SEGE_BPT4 (205)
END2_BPT4 (144)
A287R_pbcv1 (251)
A134L_pbcv1 (165)
A315L_pbcv1 (246)
A351L_pbcv1 (358)
A495R_pbcv1 (221)
A539R_pbcv1 (173)
A651L_pbcv1 (230)
o72R_CIV (245)
o295_ecoli (295)
YurQ_bacsu (124)
25k_mychy (217)
l0664_synsp (306)
YOO2_MYCTU (645)
UVRC_ECOLI (610)
UVRC_HAEIN (609)
UVRC_HELPY (594)
UVRC_BACSU (598)
UVRC_MYCGE (597)
UVRC_MYCPN (586)
UVRC_MYCTU (646)
UVRC_PSEFL (607)
UVRC_PSESY (150)
UVRC_SYNY3 (626)
uvrc_borbu (603)
uvrc_aquae (566)
uvrc_metth (579)

Based on the iterative motif search, the GIY-YIG module is found in proteins encoded by group I introns of bacteriophage T4 and in the mitochondria and chloroplast genomes of fungi, algae and liverworts; in intergenic regions of bacteria, phage and viruses; and in the UvrC subunits of bacterial and archaeal (A)BC excinucleases (Fig. 2B). While motif B is missing only from the Seg proteins, which are intergenic endonucleases of bacteriophage T4, motif C is absent from one-third of the proteins.

I-*Tev*I contains each of the five motifs that form the full module and appears to have all of the most conserved residues identified in the computational analysis (Fig. 2A). Not only does I-*Tev*I contain a typical GIY-YIG sequence (GIY-YVG in this case) characteristic of motif A, but also the absolutely conserved Arg of motif B (Arg27), highly conserved Phe (Phe56) of motif C and invariant Glu (Glu75) and Asn (Asn90) residues of motifs D and E, respectively. The five-motif module is contained within the first 92 amino acids of I-*Tev*I, well within the 116 amino acid limit placed on the catalytic domain by deletion analysis.

NMR assignment strategy for the catalytic domain of I-*Tev*I

To further define the structure of the catalytic domain, we used NMR spectroscopy to determine the topology of the first 125 amino acids of I-*Tev*I (125N) (8). Backbone resonance assignment involved the use of both uniformly and specifically ^{15}N-labeled N-terminal domain (125N) carrying the R27A mutation (^{15}N-R27A-125N). Backbone amide protons were correlated with intraresidue alpha protons using 3D clean [^{1}H-^{15}N] TOCSY-HSQC data sets (22,24). Sequential backbone amide–amide and alpha–amide connectivities were identified using a 3D [^{1}H-^{15}N] NOESY-HSQC experiment (22,23). To confirm sequential assignments, proteins selectively enriched in ^{15}N-labeled Leu, Ala/Glx, Val and Ile were produced to assign resonances, especially in unstructured regions. Using this combination of data, 99 out of 123 expected backbone amide resonances were assigned (Fig. 3).

Secondary structure determination

Amide hydrogen exchange rates, $^{3}J_{HNH\alpha}$ coupling constants (25,26) and steady-state heteronuclear {^{1}H} ^{15}N NOE (21) effects were measured to further characterize the topology of R27A-125N (summarized in Fig. 4A). The catalytic domain has a ββααβα folding pattern, followed by an unstructured C-terminal tail.

Three α-helical regions (Phe24–Phe32, Ile43–Lys51 and Asp69–Ile86) were identified based on sequential amide–amide and medium-range NOEs between alpha protons and amide protons stacked along the helix, particularly $d_{\alpha N(i, i+3)}$ connectivities (28), backbone coupling constants ($^{3}J_{HNH\alpha}$) and hydrogen exchange data (Fig. 4A). Ile43 and Phe49 in the second helix are missing some medium-range NOEs, particularly $d_{\alpha N(i, i+1)}$, due to overlap, but $^{3}J_{HNH\alpha}$ coupling constants indicate that these residues have α-helical backbones. The last helical segment, residues 69–86, is largely defined by sequential NOEs and chemical shift indices, since overlap causes gaps in the medium-range NOE, coupling constant and hydrogen exchange data. Importantly, two highly conserved residues, Arg27 and Glu75, can now be located in two of the three α-helical regions of the catalytic domain.

Strong, sequential alpha to amide proton NOEs and backbone coupling constants >8 Hz, characteristic features of β-strands (28), are observed for residues Ser3–Lys9, Val16–Ala21 and Glu62–Ile64. Residues contained in the short third strand, Glu62 –Ile64, have intermediate coupling constants, but the sequence has been defined as a β-strand on the basis of alpha proton chemical shifts (29), sequential and non-sequential NOE patterns (see below).

The pattern of non-sequential amide–amide and alpha–amide NOEs defines an anti-parallel β-sheet arrangement for the three β-strands. Individual backbone NOEs and inferred hydrogen bonds that hold the β-sheet together are illustrated in Figure 4B. Hydrogen exchange rates for amide protons between strands of the β-sheet are slow and amide protons on the edge of the sheet (Ser3 and Gly4) show intermediate hydrogen exchange rates, as expected for the overall β-sheet structure described. Interestingly, the GIY-YVG residues of motif A are located in the β-sheet of the catalytic domain.

An unstructured region at the C-terminus of the catalytic domain

Random coil amide proton shifts (30), fast hydrogen exchange rates, negative or low steady-state heteronuclear NOEs and a notable lack of sequential and medium-range NOEs characterize 22 of the unassigned amide resonances. The presence of a highly mobile, unstructured region of the protein is apparent in the steady-state {^{1}H}^{15}N NOE spectrum (data not shown). About 18 resonances in the center of the proton spectrum (7.9–8.4 p.p.m.)

Figure 2. (Opposite) GIY-YIG conserved sequence motifs. (**A**) Conserved sequence motifs of the GIY-YIG module. Multiple alignment of the motifs is shown as sequence logos (12,54). The height of each amino acid is in bits of information and is proportional to its conservation at that position, after the sequences have been weighted and frequencies adjusted by the expected amino acid frequency. The minimal and maximal distances between adjacent motifs are shown between them. Beneath each logo motif is the I-*Tev*I sequence, with its position in the protein. The logos were computed from the sequences in (B). (**B**) Protein sequences containing the GIY-YIG motifs. The length of each sequence is shown in parentheses beside a dashed line showing the positions of the five conserved motifs presented in (A). Each dash represents 10 amino acids. Sequence names are according to the SwissProt convention with organism designation following the protein name. Sequence database accession numbers: AT6i1_poamt 83822, COIi6_poamt 1334538, COIi14_poamt 1334547, COIIi1_poamt 1334559, COBi1b_poamt 578862, COBi2_poamt 1334531, ND1i1_poamt 1334565, ND1i3_poamt 1334566, ND1i4b_poamt 1334568, ND1i1_pocmt 1743352, COBi1_necmt 13116, ND1i1_necmt 14129, ND1i2_almmt 2147548, o360_almmt 459018, COBi1_trpmt 732979, ND1i2_trpmt 479530, COBi1_sadmt 13617, COIi4_agamt 2738528, COBi1_ceumt 2865254, COBi1_chemt 2193888, COIi9_mapmt 786182, psbAi2_crecp 296431, rbcLi2_pcacp 3164196, TEV1_BPT4 119333, SEGA_BPT4 417766, SEGB_BPT4 464756, SEGC_BPT4 2506234, SEGD_BPT4 730735, SEGE_BPT4 140785, END2_BPT4 729416, A287R_pbcv1 1181450, A134L_pbcv1 1131478, A315L_pbcv1 1181478, A351L_pbcv1 1181514, A495R_pbcv1 1620166, A539R_pbcv1 1620210, A651L_pbcv1 2447115, o72R_CIV 2738435, o295_ecoli 1788037, YurQ_bacsu 2635759, 25k_mychy 1354226, l0664_synsp 1006601, l0441_synsp 1653896, Y002_MYCTU 1722908, UVRC_ECOLI P07028, UVRC_HAEIN P44489, UVRC_HELPY P56428, UVRC_BACSU P14951, UVRC_MYCGE P47448, UVRC_MYCPN P75350, UVRC_MYCTU P71689, UVRC_PSEFL P32966, UVRC_PSESY 3024790, UVRC_SYNY3 P73580, uvrc_borbu 2688360, uvrc_aquae 2984329, uvrc_metth 2621507.

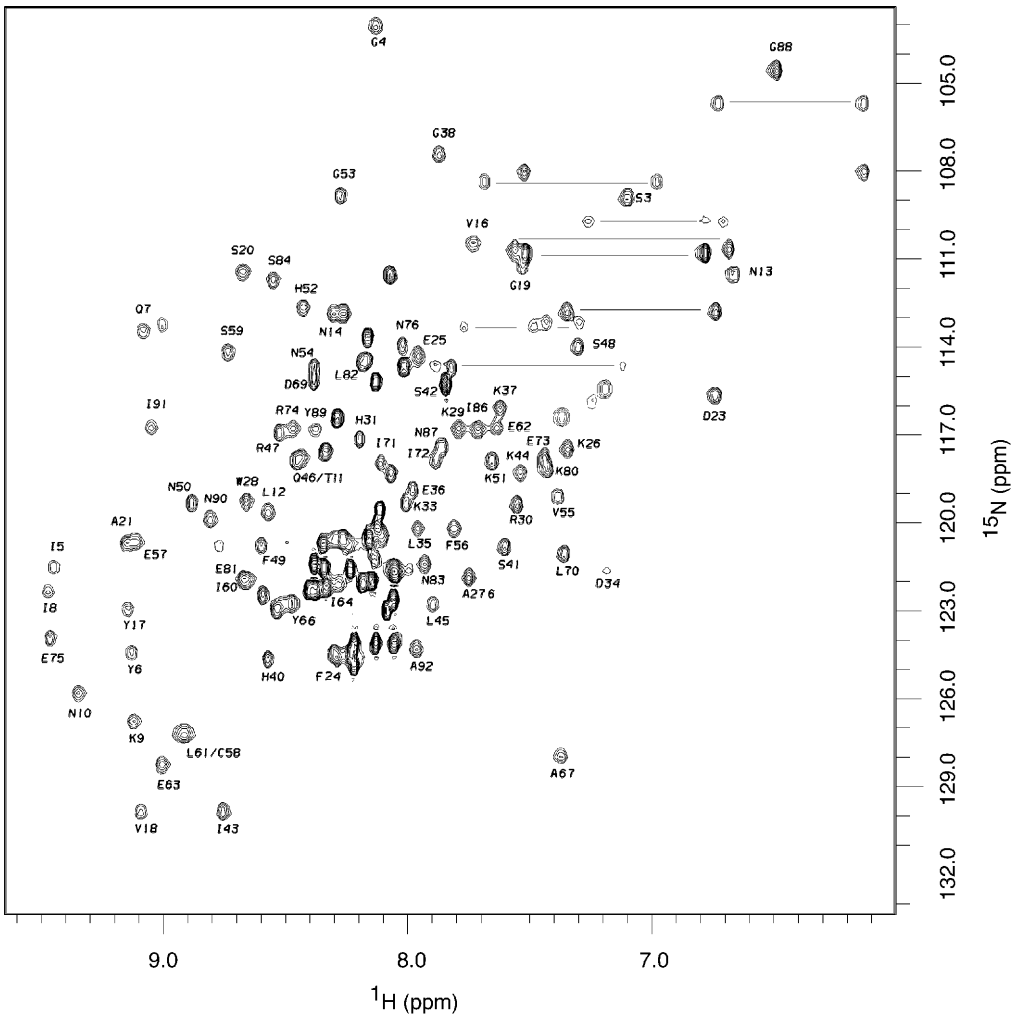


Figure 3. [^{1}H-^{15}N] HSQC spectrum of the R27A catalytic domain uniformly labeled with ^{15}N was recorded at 25°C on a Bruker Avance 500 MHz spectrometer. Each peak in the spectrum corresponds to a ^{1}H attached to an ^{15}N. The primary amide protons of the glutamine and asparagine side-chains are connected by lines (upper right of the spectrum). Assigned backbone amide protons are labeled. Aliased peaks are represented by broken contour lines.

have negative {^{1}H}^{15}N NOE signals, indicating that these amides undergo uncorrelated motions and are highly flexible (21). Based on sequential and medium-range interproton NOEs, the structured part of the catalytic domain ends at residue 92 (Ala92). No resonances in the C-terminal tail, residues 93–125, are sequentially assigned on the basis of interproton NOE data.

Nine amino acids in the C-terminal tail were assigned using specifically ^{15}N-labeled proteins. All lack interresidue NOEs, have fast hydrogen exchange rates and negative or low heteronuclear NOE values, providing evidence that they are largely unstructured. Two residues, Ile109 and Ile110, have non-negative, but low (0.47 and 0.04, respectively) steady-state {^{1}H} ^{15}N NOE values. These data could suggest that these isoleucines form part of a hydrophobic patch but neighboring residues, Leu105, Glu107 and Glu108, have negative {^{1}H}^{15}N NOE values, which indicate that any structure is irregular and limited. These data indicate that the C-terminal tail, residues 93–125, is largely unstructured in solution. The end of the structured segment of the catalytic domain at Ala92 is in remarkable coincidence with predictions from the computational sequence analysis of the GIY-YIG module, where sequence conservation ends at precisely this residue.

Comparison of the unstructured linker in the catalytic domain and full-length I-*Tev*I

The number of unassigned, flexible backbone amides in the R27A-125N NMR spectra raised concerns that the isolated catalytic domain might not be properly folded. Therefore, we specifically incorporated ^{15}N-valine and ^{15}N-leucine into R27A-125N and full-length R27A I-*Tev*I to characterize the domain in both contexts. R27A-125N contains 4 out of the 8 valines and 9 out of the 11 leucines of the full-length protein. Amide chemical shifts provide highly sensitive probes of tertiary protein structure. The amide chemical shift is affected by hydrogen bonding and other factors in a way that is poorly understood, but chemical shift perturbations >0.1 p.p.m. are considered to be non-random changes (31). Direct comparison of chemical shift values for the leucine and valine amide resonances of the N-terminal domain revealed only minor changes (<0.03 p.p.m.) relative to the full-length protein (Fig. 5). Sharp, intense HSQC resonance line shapes, in addition to negative steady-state heteronuclear NOE values and fast hydrogen exchange rates, indicate that the assigned amino acids (Leu105, Val117) in

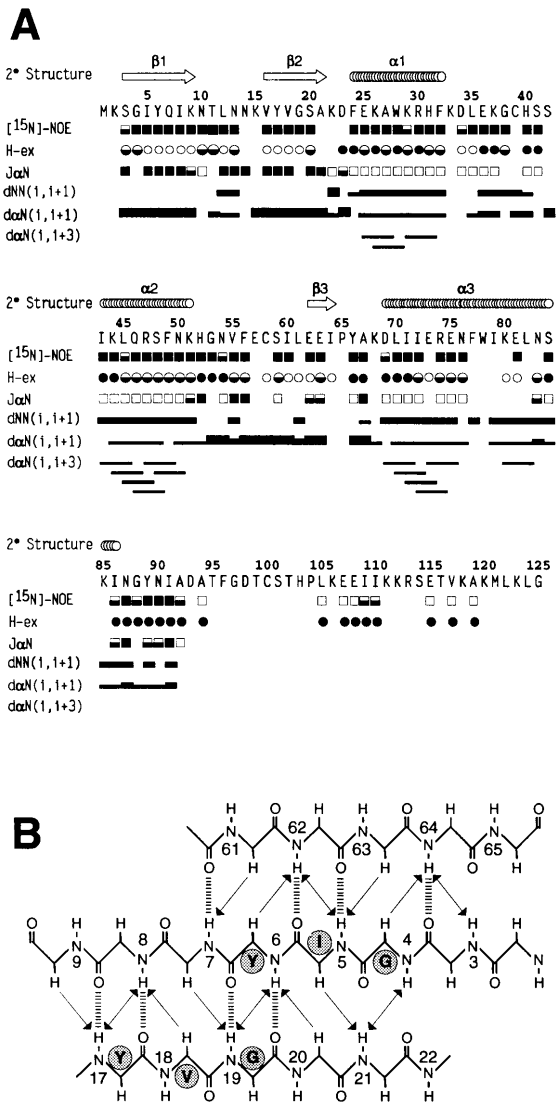


Figure 4. Topology of the catalytic domain of I-*Tev*I. (**A**) Summary of NMR data. Sequential and medium-range NOEs, $^{3}J_{HNH\alpha}$ coupling constants, hydrogen exchange rates and steady-state heteronuclear {^{1}H} ^{15}N NOE/NONOE ratios are shown. The inferred secondary structure elements are represented above the primary sequence of the catalytic domain (R27A-125N). The strength of the NOEs are represented by thick lines (strong), medium lines (intermediate) and thin lines (weak). The $^{3}J_{HNH\alpha}$ coupling constants are represented by filled squares (>8 Hz), half-filled squares (8 Hz >$^{3}J_{HNH\alpha}$ >6 Hz), or open squares (<6 Hz). The hydrogen exchange rates are represented by filled circles (fast), half-filled circles (intermediate) and open circles (slow). The {^{1}H} ^{15}N NOE/NONOE ratios are represented by filled squares (ratio >0.6), half-filled squares (ratio <0.6 and >0) and open squares (ratio <0). (**B**) Anti-parallel β-sheet structure. Interstrand NOEs are represented by arrows. Double-headed arrows indicate that NOEs from both amide protons are seen. Single-headed arrows indicate that the ^{1}H NOE is seen in the NOESY-HSQC spectrum at the ^{15}N frequency of the amide ^{1}H. Hydrogen bonds inferred from the hydrogen exchange data are represented by dashed lines. The residues of the characteristic GIY-YVG motif are labeled.

the linker of R27A-125N are quite flexible. The HSQC lineshapes and chemical shift frequencies of Leu105 and Val117 do not change when they are part of full-length I-*Tev*I (compare Fig. 5C with A and B). These data indicate that the isolated catalytic domain retains the same overall fold in full-length R27A I-*Tev*I. Importantly, the data suggest that residues in the C-terminal region of R27A-125N are also unstructured in the full-length protein.

Role of conserved residues in the GIY-YIG module

Site-directed mutagenesis of the conserved tyrosine residues of the GIY-YVG sequence (Tyr6 and Tyr17) and the four invariant residues identified in the comparative analysis of the GIY-YIG module (Gly19, Arg27, Glu75 and Asn90) was carried out to identify residues involved in cleavage activity and place them in the context of secondary structural elements. The conserved residues were converted to alanine and the mutant derivatives were expressed in an *in vitro* transcription/translation system to yield full-length proteins which were characterized for DNA-binding and cleavage activity (Fig. 6). Wild-type I-*Tev*I formed two bound complexes with a 195 bp homing-site fragment in band-shift assays (Fig. 6A and B). These complexes, U_F (upper fast) and U_S (upper slow), have been previously characterized (7). U_F is an intact complex with I-*Tev*I contacts around the IS only, whereas U_S is a nicked, bent catalytic complex, with contacts extending to the CS. Lower complex (LC) corresponds to binding of the C-terminal DNA-binding domain (compare with 130C lane) sometimes seen due to breakdown of the full-length enzyme. All six of the I-*Tev*I mutants were able to bind substrate and yield some form of U_F complex (see below) as well as LC, consistent with an intact DNA-binding domain. However, none of the six derivatives showed a band corresponding to the U_S catalytic complex. In accord with these results, cleavage assays revealed that I-*Tev*I mutants Y6A, G19A, R27A and E75A have no detectable catalytic activity (<0.1% of wild-type enzyme) while both N90A and Y17A have a greatly reduced level of cleavage compared to the wild-type enzyme (1–3%) (Fig. 6A and C).

E75A I-*Tev*I behaves like the previously identified catalytic mutant, R27A I-*Tev*I. It has no cleavage activity, yet it binds the homing site and is well-behaved during overexpression and purification. To address whether E75A and R27A I-*Tev*I have a common fold, we prepared E75A I-*Tev*I specifically labeled with both ^{15}N-leucine and ^{15}N-valine for NMR analysis in comparison with R27A I-*Tev*I. [^{1}H-^{15}N] HSQC spectra (Fig. 6D) show that the E75A mutation had little effect on the fold relative to R27A. Only one residue, Val16, shifts more than 0.1 p.p.m. Therefore, it would appear that Glu75 is also a catalytic residue. Interestingly, both putative catalytic residues are located in α-helices.

I-*Tev*I mutants Y6A, Y17A, G19A and N90A all appear to be structurally compromised. These derivatives are insoluble and/or unstable when overexpressed in *E.coli* and Y6A, G19A and N90A mutants could not be prepared for NMR analysis. Consistent with structural anomalies, there is variation in the U_F complex formed with the Y6A, Y17A and N90A derivatives in the band-shift analysis (Fig. 6), although the mobility of the LC complex is unchanged. This is consistent with alterations in N-terminal domain structure without perturbations in the C-terminal DNA-binding domain. In the absence of further structural information, we are limited in our ability to draw conclusions about the role of Asn90. Despite compromised preparations of Y17A I-*Tev*I, [^{1}H-^{15}N] HSQC analysis was possible for the fraction of protein that is folded. This revealed shifts in Val16 (+0.19 p.p.m.) and Leu12 (+0.13 p.p.m.) (data not shown), suggesting changes in the β-sheet. Given that the NMR data show

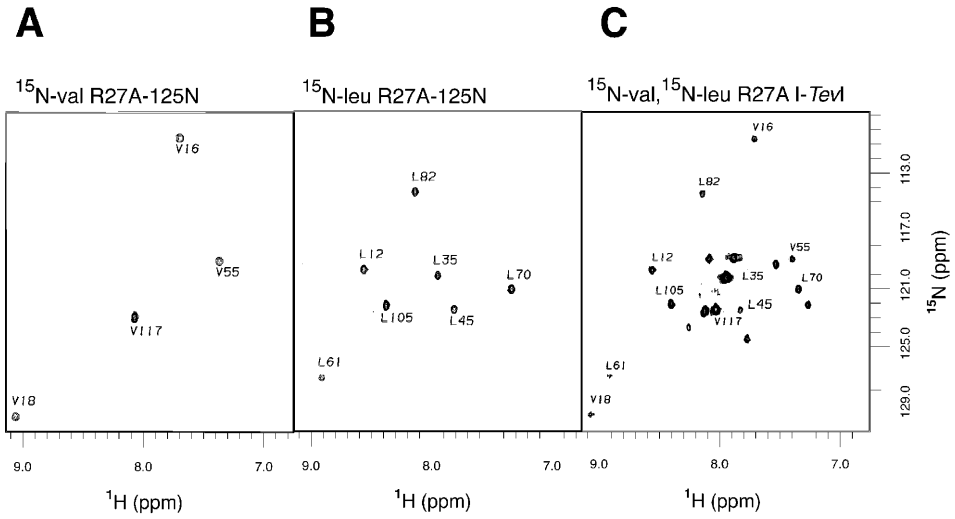


Figure 5. NMR comparison of the catalytic domain and full-length I-*Tev*I. Each resonance in the three fingerprint HSQC spectra corresponds to an amide ^{1}H-^{15}N pair from either a valine or leucine residue. (**A**) [^{1}H-^{15}N] HSQC spectrum of R27A-125N specifically labeled with ^{15}N at the valine residues. (**B**) [^{1}H-^{15}N] HSQC spectrum of R27A-125N specifically labeled with ^{15}N at the leucine residues. (**C**) [^{1}H-^{15}N] HSQC spectrum of full-length R27A I-*Tev*I specifically labeled with ^{15}N at the valine and leucine residues.

Tyr6, Tyr17 and Gly19 to be integral parts of the β-sheet and that mutation of Tyr17 perturbs this element, it is not unreasonable to propose that mutations in these residues disrupt the β-sheet, compromising both structure and activity.

DISCUSSION

Our previous work divided I-*Tev*I structurally and functionally into two domains, each comprising approximately one-half of the 245-amino acid protein. The N-terminal half contains the GIY-YIG motif characteristic of this family of endonucleases and constitutes the catalytic domain of the enzyme, while the C-terminal half is an independent recognition domain which binds the target DNA with the same affinity as full-length enzyme. We have now used a combination of approaches to delineate the catalytic domain more precisely (Fig. 7). A deletion analysis delimited the C-terminal boundary of the catalytic domain, while a rigorous sequence comparison has extended our understanding of the GIY-YIG module. Finally, a preliminary NMR structure gives us our first glimpse into the structure and function of a GIY-YIG endonuclease.

The GIY-YIG module

The computational analysis has more clearly defined the GIY-YIG module. In most cases, the module consists of five motifs with both conserved and invariant residues. We have used the module to identify additional members of the family and to locate conserved amino acids in I-*Tev*I that could be involved in cleavage activity (Fig. 2A).

The analysis identified more than 30 new members of the family. These include the UvrC subunits of bacterial and archaeal (A)BC excinucleases (Fig. 2B). (A)BC excinucleases remove damaged nucleotides by incising the damaged strand on both sides of the lesion. Curiously, although UvrC makes the 5′ incision, deletion analysis has shown that the UvrC GIY-YIG module is not essential for this activity (32). More experimental data are required to ascertain what role the module plays in UvrC function. In addition, a UvrC-like region was identified in the C-terminal half of a *Mycobacterium tuberculosis* hypothetical protein (Fig. 2), Y002-MYCTU. This region includes the GIY-YIG module and other motifs found in UvrC proteins but not the active site region (32). The N-terminal half of this protein is a close homolog of the epsilon subunit of bacterial DNA polymerase III holoenzymes, responsible for the 3′ to 5′ exonuclease proofreading activity of the polymerase (33). Again, the role of the GIY-YIG module in this ORF is unclear, but the protein it encodes would be predicted to have some nuclease activity.

Most GIY-YIG modules are diverged by sequence, having <40% identity, or they are merely similar across short sequence regions. However, sequences in some groups are more similar to each other than to other GIY-YIG modules (data not shown). These include UvrC proteins and a few groups of intron-encoded proteins, some of which could have arisen from horizontal transfer. In addition, two groups of viral proteins, the SegA to E proteins of T4 phage (34) and a group in the *Paramecium bursaria* Chlorella virus 1 (pbcv1) contain multiple proteins from a single organism which could be paralogs resulting from duplications within a genome.

Definition of the catalytic domain

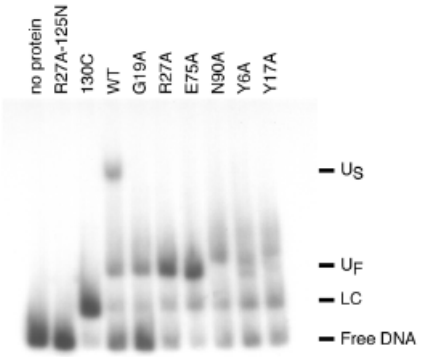
The five conserved motifs that make up the GIY-YIG module extend to residue 92 of I-*Tev*I, so that all the conserved residues are contained within the 125N derivative. The lack of sequence similarity between I-*Tev*I residues 93–245 and other members of the GIY-YIG family is consistent with the remaining portion of the enzyme being mainly involved in a non-catalytic function, such as sequence-specific DNA binding, that will vary from enzyme to enzyme.

In remarkable coincidence with the computational analysis, the NMR structural analysis of 125N shows that residues 1–92 are folded into discrete secondary structure elements forming a compact domain while residues 93–125 are unstructured. Comparative

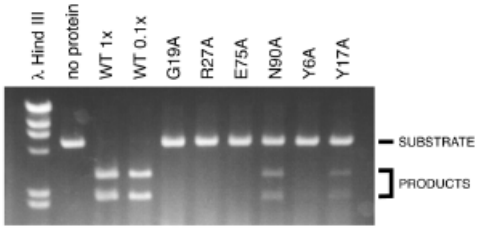
A

I-*Tev*I	Relative cleavage activity (%)	DNA-binding activity	
		U_F	U_S
WT	100	+	+
G19A	0	+	−
R27A	0	+	−
E75A	0	+	−
N90A	3	+	−
Y6A	0	+	−
Y17A	1	+	−

B



C



D

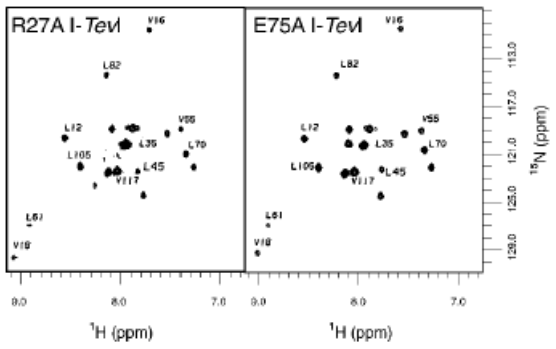


Figure 6. Mutational analysis of I-*Tev*I. (**A**) Summary of DNA-binding and cleavage data presented in (B) and (C). (**B**) Band-shift gel to show DNA-binding activity of I-*Tev*I mutant derivatives. Gel-mobility shift assays were performed as previously described using a 195 bp PCR-derived fragment (7,11). (**C**) Cleavage activity of the I-*Tev*I mutant derivatives. The agarose gel shows standard cleavage assays (37°C, 30 min) with equal amounts of I-*Tev*I derivatives generated *in vitro*. (**D**) NMR comparison of R27A I-*Tev*I and E75A I-*Tev*I. 2D [^{1}H-^{15}N] HSQC spectra of ^{15}N-Val, ^{15}N-Leu, R27A I-*Tev*I and E75A I-*Tev*I.

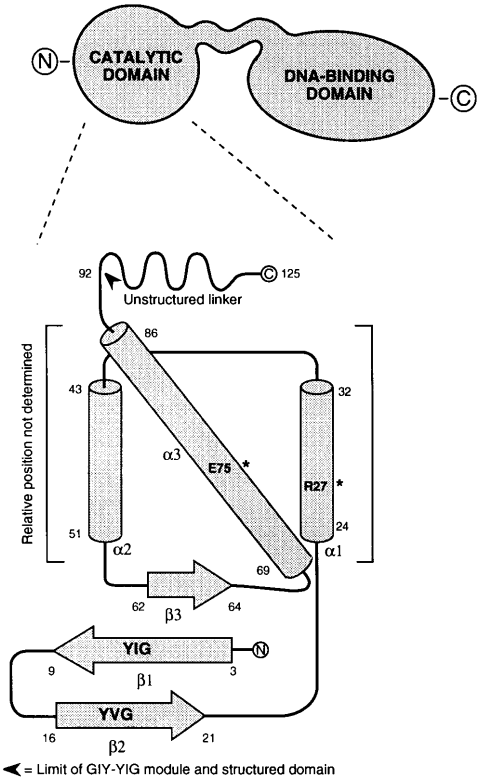


Figure 7. Overview of I-*Tev*I. Top, two-domain model of I-*Tev*I. Bottom, topology diagram of the catalytic domain of I-*Tev*I based on NMR analysis. Relative positioning of α-helices in square brackets has not been defined. Coincidence of the limit of the GIY-YIG module determined by computational analysis and the structured domain determined by NMR are indicated (arrowhead). Catalytic residues are marked by an asterisk.

HSQC spectra of specifically labeled 125N and full-length I-*Tev*I suggest that residues in the unstructured region adopt the same conformation in both contexts, and that residues 93–125 are also unstructured in the native enzyme. It is satisfying to note that all of the conserved residues in the GIY-YIG module are contained within the structured domain (residues 1–92) and that sequence conservation and structure end at the same residue. These observations are consistent with the GIY-YIG module forming a discrete, well-defined structural unit that forms a catalytic cartridge imparting endonuclease function to a variety of other protein domains.

Structure and function of residues 93–125

A systematic deletion analysis of the central portion of the molecule showed that one- and two-amino acid deletions were well tolerated C-terminal to residue 116. The apparent dispensability of many of the individual amino acids between residues 116 and 130 is consistent with their inclusion in a flexible linker that tethers the two functional domains of the protein. However, the NMR data indicated that residues beyond 92 are unstructured. The apparent discrepancy between the deletion analysis and NMR data suggests that the 'functional' catalytic domain is larger than that defined as a structural domain by NMR. The different boundaries may reflect that deletions at the linker boundary disrupt the structure of the catalytic domain, or they may reflect structural

differences between the free enzyme and the DNA-bound form. Residues 93–116 may become structured upon binding to DNA and could form essential contacts with the substrate.

Structural changes among DNA-binding proteins upon multimerization and DNA binding are not uncommon. For example, the C-terminal α-helices of the *Bam*HI dimer unravel upon DNA binding to form partially disordered arms (35) and six unstructured residues at the N-terminus of lambda repressor become ordered upon DNA binding (36). Like I-*Tev*I, γδ resolvase is divided into separate N-terminal catalytic and C-terminal DNA-binding domains, and these are separated by an arm region which becomes structured only when the resolvase dimer binds DNA (37). Additionally, it has been suggested that this arm region is the source of the enzyme's tolerance for variability in the spacing between its binding sites (38), much as the linker of I-*Tev*I imparts flexibility and the ability to cleave at different distances from its binding site.

Potential role of conserved residues in the function of the catalytic domain

Site-directed mutagenesis experiments were guided by the multiple sequence alignment (Fig. 2A). In addition to the previously identified catalytic residue Arg27 (8), Glu75 is also deemed to be a catalytic residue since E75A I-*Tev*I can bind a homing site substrate but cannot cleave, despite being well folded as judged by 2D HSQC analysis. As previously discussed, it is not unusual to find Arg residues at the active site of endo- and exonucleases where they can perform several functions including stabilization of the pentacovalent intermediate formed during the reaction (39). Similarly, it is no surprise to find a carboxylate residue at the active site of I-*Tev*I. Glu and Asp residues can act as a general base to initiate cleavage, as seen for Glu58 of ribonuclease T1 (40,41). In addition, they can bind divalent cations that can be intimately involved in phosphodiester bond cleavage. The divalent cation(s) can stabilize a transition state intermediate and/or polarize a water molecule that can attack the labile phosphate to effect cleavage as proposed for the 3′ to 5′ exonuclease of DNA polymerase I (42).

The most striking structural feature of the I-*Tev*I catalytic domain is a three-stranded anti-parallel β-sheet that contains the GIY and YVG sequence elements including Tyr6, Tyr17 and Gly19. A direct role for Gly19 in catalysis is unlikely since the side chain is unable to participate in any of the chemical steps of the cleavage reaction. In addition, the NMR data show that the Gly19 amide group is hydrogen bonded in the middle of the β-sheet and the carbonyl group is also pointed into the β-sheet array, precluding a role for these groups in the reaction mechanism (Fig. 4B). Tyrosine residues can play a role in phosphodiester bond cleavage as seen for DNA polymerase I and bovine DNase I (42–44). However, in addition to any direct catalytic role Tyr6 and Tyr17 may play, the data suggest that the GIY-YVG residues play a vital role in maintaining the integrity of the β-sheet.

What could be the role of the β-sheet? While the primary DNA-binding deteminants of I-*Tev*I are contained in the C-terminal domain, the N-terminal domain must contact the DNA to effect cleavage. This is also apparent from the ability of I-*Tev*I to identify its CS when that site is placed at a variable distance from the primary binding site (7). I-*Cre*I, PI-*Sce*I and I-*Dmo*I (members of the LAGLI-DADG family of homing endonucleases) and I-*Ppo*I (a member of the His-Cys box family) all interact with their DNA substrates via β-sheets although the families appear to use quite different catalytic mechanisms to achieve cleavage (45–49). Structural studies of the human U1a protein bound to RNA suggest that a small β-sheet serves as a platform for generic RNA binding (50,51). The DNA-binding domains of Tn*916* integrase and the GCC-box binding domain of a protein from *Arabidopsis thaliana* (AtERFI GBD) have recently been shown to bind DNA using three-stranded, anti-parallel β-sheets (52,53). The β-sheets of U1a, Tn*916* integrase and AtERFI GBD contain aromatic residues adjacent to conserved hydrophobic residues that interact with their polynucleotide substrates. The presence of highly conserved tyrosine residues and hydrophobic residues in the GIY and YVG sequences of the sheet structure suggests that the β-sheet of I-*Tev*I could play a similar role in positioning of the DNA for cleavage. Further structural analysis will provide deeper mechanistic insight into the GIY-YIG catalytic module, which represents a common functional unit in this otherwise heterogenous family of proteins.

ACKNOWLEDGEMENTS

We are grateful to R. Bonocora and D. Shub for sharing their unpublished GIY-YIG sequence alignments that showed Glu75 to be highly conserved. We acknowledge the contributions of the Molecular Genetics, Biological Mass Spectroscopy and Structural Biology NMR core facilities at the Wadsworth Center. We thank Maryellen Carl for expert secretarial assistance and Lynn McNaughton for steadfast assistance during NMR data acquisition. This work was supported by NIH grants GM39422 and GM44844 to M.B. and GM56966 to Patrick Van Roey. S.P. was supported by NIH grant GM29009 to Steve Henikoff.

REFERENCES

1 Mueller,J.E., Bryk,M., Loizos,N. and Belfort,M. (1993) In Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 111–143.
2 Belfort,M. and Perlman,P.S. (1995) *J. Biol. Chem.*, **270**, 30237–30240.
3 Belfort,M. and Roberts,R.J. (1997) *Nucleic Acids Res.*, **25**, 3379–3388.
4 Michel,F. and Dujon,B. (1986) *Cell*, **46**, 323.
5 Bryk,M., Quirk,S.M., Mueller,J.E., Loizos,N., Lawrence,C. and Belfort,M. (1993) *EMBO J.*, **12**, 2141–2149.
6 Mueller,J.E., Smith,D., Bryk,M. and Belfort,M. (1995) *EMBO J.*, **14**, 5724–5735.
7 Bryk,M., Belisle,M., Mueller,J.E. and Belfort,M. (1995) *J. Mol. Biol.*, **247**, 197–210.
8 Derbyshire,V., Kowalski,J.C., Dansereau,J.T., Hauer,C.R. and Belfort,M. (1997) *J. Mol. Biol.*, **265**, 494–506.
9 Li,L., Wu,L.P. and Chandrasegaran,S. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 4275–4279.
10 Colandene,J.D. and Topal,M.D. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 3531–3536.
11 Bell-Pedersen,D., Quirk,S.M., Bryk,M. and Belfort,M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 7719–7723.
12 Henikoff,S., Henikoff,J.G., Alford,W.J. and Pietrokovski,S. (1995) *Gene*, **163**, 17–26.
13 Bailey,T.L. and Elkan,C. (1994) In *Anonymous Proceedings of Second International Conference on Intelligent Systems for Molecular Biology.* AAAI Press, Menlo Park, CA, pp. 28–36.
14 Schuler,G.D., Altschul,S.F. and Lipman,D.J. (1991) *Proteins*, **9**, 180–190.
15 Pietrokovski,S. (1994) *Protein Sci.*, **3**, 2340–2350.
16 Pietrokovski,S. (1998) *Protein Sci.*, **7**, 64–71.
17 Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
18 Muchmore,D.C., McIntosh,L.P., Russell,C.B., Anderson,D.E. and Dahlquist,F.W. (1989) *Methods Enzymol.*, **177**, 44–73.

19 Bax,A., Ikura,M., Kay,L.E., Torchia,D.A. and Tschudin,R. (1990) *J. Magn. Reson.*, **86**, 304–318.
20 Norwood,T.J., Boyd,J., Heritage,J.E., Soffe,N. and Campbell,I.D. (1990) *J. Magn. Reson.*, **87**, 488–501.
21 Kay,L.E., Torchia,D.A. and Bax,A. (1989) *Biochemistry*, **28**, 8972–8979.
22 Marion,D., Driscoll,P.C., Kay,L.E., Wingfield,P.T., Bax,A., Gronenborn,A.M. and Clore,G.M. (1989) *Biochemistry*, **28**, 6150–6156.
23 Zuiderweg,E.R.P. and Fesik,S.W. (1989) *Biochemistry*, **28**, 2387–2391.
24 Cavanagh,J. and Rance,M. (1992) *J. Magn. Reson.*, **96**, 670–678.
25 Billeter,M., Neri,D., Otting,G.D., Qian,Y.Q. and Wuthrich,K. (1992) *J. Biomol. NMR*, **2**, 257–274.
26 Neri,D., Otting,G. and Wuthrich,K. (1990) *J. Am. Chem. Soc.*, **112**, 3663–3665.
27 Pietrokovski,S., Henikoff,J.G. and Henikoff,S. (1998) *Trends Genet.*, **14**, 162–163.
28 Wuthrich,K. (1986) *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, Inc., New York, NY.
29 Wishart,D.S., Sykes,B.D. and Richards,F.M. (1992) *Biochemistry*, **31**, 1647–1651.
30 Wishart,D.S., Bigam,C.G., Holm,A., Hodges,R.S. and Sykes,B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.
31 Williamson,M.P. and Asakura,T. (1997) In Reid,D.G. (ed.), *Methods in Molecular Biology. Protein NMR Techniques.* Humana Press, Inc., Totowa, NJ.
32 Linn,J.J. and Sancar,A. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 6824–6828.
33 Kelman,Z. and O'Donnell,M. (1995) *Annu. Rev. Biochem.*, **64**, 171–200.
34 Sharma,M., Ellis,R.L. and Hinton,D.M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6658–6662.
35 Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) *Science*, **269**, 656–663.
36 Jordan,S.R. and Pabo,C.O. (1988) *Science*, **242**, 893–899.
37 Yang,W. and Steitz,T.A. (1995) *Cell*, **82**, 193–207.
38 Grindley,N.G.F. (1994) *Nucleic Acids Mol. Biol.*, **8**, 236–267.
39 Gerlt,J.A. (1993) In Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases.* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1–34.
40 Steyaert,J., Hallenga,K., Wyns,L. and Stanssens,P. (1990) *Biochemistry*, **29**, 9064–9072.
41 Heydenreich,A., Koellner,G., Choe,H., Cordes,F., Kisker,C., Schindelin,H., Adamiak,R., Hahn,U. and Saenger,W. (1993) *Eur. J. Biochem.*, **218**, 1005–1012.
42 Beese,L. and Steitz,T.A. (1991) *EMBO J.*, **10**, 25–33.
43 Lahm,A. and Suck,D. (1991) *J. Mol. Biol.*, **221**, 645–667.
44 Weston,S.A., Lahm,A. and Suck,D. (1992) *J. Mol. Biol.*, **226**, 1237–1256.
45 Duan,X., Gimble,F.S. and Quiocho,F.A. (1997) *Cell*, **89**, 555–564.
46 Heath,P.J., Stephens,K.M., Monnat,R.J.,Jr and Stoddard,B.L. (1997) *Nature Struct. Biol.*, **4**, 468–476.
47 Jurica,M.S., Monnat,R.J.,Jr and Stoddard,B.L. (1998) *Mol. Cell*, **2**, 469–476.
48 Flick,K.E., Jurica,M.S., Monnat,R.J.,Jr and Stoddard,B.L. (1998) *Nature*, **394**, 96–101.
49 Silva,G., Dalgaard,J.Z., Belfort,M. and Van Roey,P. (1999) *J. Mol. Biol.*, **286**, 1123–1136.
50 Oubridge,C., Ito,N., Evans,P.R., Teo,C.H. and Nagai,K. (1994) *Nature*, **372**, 432–438.
51 Allain,F.H., Howe,P.W., Neuhaus,D. and Varani,G. (1997) *EMBO J.*, **16**, 5764–5772.
52 Connolly,K.M., Wojiciak,J.M. and Clubb,R.T. (1998) *Nature Struct. Biol.*, **5**, 546–550.
53 Allen,M.D., Yamasaki,K., Ohme-Takagi,M., Tateno,M. and Suzuki,M. (1998) *EMBO J.*, **17**, 5484–5496.
54 Schneider,T.D. and Stephens,R.M. (1990) *Nucleic Acids Res.*, **18**, 6097–6100.