

A single catalytic domain of the junction-resolving enzyme T7 endonuclease I is a non-specific nicking endonuclease

Chudi Guan* and Sanjay Kumar

New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938, USA

Received July 1, 2005; Revised September 1, 2005; Accepted October 6, 2005

ABSTRACT

A stable heterodimeric protein containing a single correctly folded catalytic domain (SCD) of T7 endonuclease I was produced by means of a *trans*-splicing intein system. As predicted by a model presented earlier, purified SCD protein acts a non-specific nicking endonuclease on normal linear DNA. The SCD retains some ability to recognize and cleave a deviated DNA double-helix near a nick or a strand-crossing site. Thus, we infer that the non-specific and nicked-site cleavage activities observed for the native T7 endonuclease I (as distinct from the resolution activity) are due to uncoordinated actions of the catalytic domains. The positively charged C-terminus of T7 Endo I is essential for the enzymatic activity of SCD, as it is for the native enzyme. We propose that the preference of the native enzyme for the resolution reaction is achieved by cooperativity in the binding of its two catalytic domains when presented with two of the arms across a four-way junction or cruciform structure.

INTRODUCTION

A four-way (Holliday) junction is the essential intermediate produced during both homologous and site-specific recombination (1,2). The penultimate stage of recombination involves resolution of a four-way junction catalyzed by structure-specific nucleases termed resolvases. The resolvase from T7 phage of bacterium *Escherichia coli*, T7 endonuclease I (T7 Endo I), is one of the most studied resolvases. Different resolvases belong to several super-families. T7 Endo I belongs to the endonuclease family (3).

T7 Endo I is a stable homodimer of identical 149 amino acid subunits (4). It binds tightly (K_d 2 nM) to four-way junctions in this form (5). T7 Endo I resolves a four-way junction

by simultaneously introducing two nicks on the two non-crossing strands, at 5' sides of the junction (5). The crystal structure of T7 Endo I has been reported (6). The structural analysis shows that it forms an intimately-associated symmetrical homodimer comprising two identical catalytic domains. Each domain is composed of residues 17–44 from one subunit and residues 50–145 from the other. A bridge that forms part of the extended and tightly associated antiparallel beta-sheets (β_2) connects the two well-separated catalytic domains. The arrangement of residues at the active site of T7 Endo I is similar to that found in several well-characterized restriction enzymes (6).

Besides resolution of four-way junctions, T7 Endo I has enzymatic activity on a variety of DNA structures ranging from branched to single-base mismatched heteroduplexes (7). The broad substrate specificity of T7 Endo I makes it difficult to interpret how the enzyme selectively recognizes its substrates (8). Introduction of mutations such as amino acid deletions, insertions and substitutions to the β -bridge site does not affect the folding and activity of the catalytic domain itself, but results in a change of the reciprocal position of the two catalytic domains. This positional change leads to changes in the activity profile of the enzyme on different substrates, to changes in metal-ion dependence profile, and even to changes in the reaction mechanism and distribution of final products (9).

Based on the published 3D structure of T7 Endo I (6) and our study on the β -bridge site mutants of the enzyme (9) we proposed a mechanistic model for substrate recognition of the enzyme: the two catalytic domains of T7 Endo I, each capable of functioning as a non-specific nicking endonuclease, are juxtaposed in a way that prevents the enzyme from forming a productive complex with regular linear DNA, but enables it to simultaneously bind its two catalytic domains across a four-way junction for the resolution reaction. The enzyme can also specifically bind and cleave branched, perturbed or flexible DNA with different efficiency, depending on the distribution of conformer populations. The published experimental data so far are consistent with the model presented above. The model

*To whom correspondence should be addressed. Tel: +1 978 927 5054; Fax: +1 978 921 1350; Email: Guan@neb.com

predicts the properties of a correctly folded single catalytic domain (SCD) of T7 Endo, so characterization of such a protein would provide key supporting evidence for the model.

MATERIALS AND METHODS

Materials

Restriction enzymes, nicking enzyme N.BstNB I, DNA polymerases, T4 ligase, T4 DNA kinase, β -agarase, λ Exonuclease, DNase I, the maltose-binding protein (MBP) protein fusion expression and purification system including plasmid pMAL-c2x, the host *E.coli* strains TB1 and ER2566, Factor Xa protease, genenase, the cruciform structure-containing plasmid pUC(AT), 2 logs DNA ladder and synthetic oligonucleotides were obtained from New England Biolabs Inc. (NEB). Plasmid pNB1 was a gift from Rebecca Kucera (NEB). The *trans*-splicing intein system from *Synechocystis* sp. PCC6803 (10) including the pair of plasmids pKEB12 and pMEB2 that contain coding sequences for the C-terminal part and the N-terminal part of the intein, respectively, was provided by Dr Thomas Evans (NEB).

Recombinant DNA and mutagenesis

DNA manipulation was performed as described in *Molecular Cloning* by Sambrook *et al.* (11). Site-directed mutagenesis by two-step PCR was described previously (9). For generating En-In gene fusion by two-step PCR, two template plasmids pEndo I (9) and pMEB2 and the following four primers were used: oligo-1, GCCGAATTCATGGCAGGTTACGGCGCTAAAGGAATC; oligo-2, TTCGGTACCAAAAGACAGGGCATTGCTCGCCGGAATTAC; oligo-3, GTAATCCGGCGAGCAATGCCCTGTCTTTTGGTACCGAA; and oligo-4, CCCGGAAGCTTATTTAATTGTCCAGCGTCAAGTAATGG.

In the first step PCR plasmid pEndo I, oligo-1 and oligo-2 were used for generating the N-terminal part (En) coding sequence of T7 Endo I; pMEB2, oligo-3 and oligo-4 were used for generating the N-terminal part (In) coding sequence of the *trans*-splicing intein. The first cysteine of In was changed to alanine to prevent autoproteolysis of the desired SCD protein (10). In the second PCR step, the En-In gene fusion was assembled by using the two products (En and In) of the first PCR step as templates, oligo-1 and oligo-4 as primers. The Ic-Ec gene fusion was generated in a similar way. Plasmid pKEB12; oligo-5, ATAGACCATGGTTAAAGTTATCGGTGCTAGATCTCTGGGC; and oligo-6, AATGCTCGCCGGAATTACGTTAGCTGCGATAGCGCC, were used for generating the C-terminal portion (Ic) coding sequence of intein. Plasmid pEndo I; oligo-7, GGCGCTATCGCAGCTAACGTAATTCCGGCGAGCAAT; and oligo-8, GGCCGCTGCA-GTTATTTCTTTCCTCTTTTCTTTTAAATCT, were used for generating the C-terminal portion (Ec) coding sequence of T7 Endo I. The Ic-Ec gene fusion was assembled by PCR using both products (Ic and Ec) of the first PCR step as templates and oligo-5 and oligo-8 as primers. Expression plasmid pMEN-In which contained the coding sequence for MBP-En-In protein fusion, named MEN-In, was constructed by inserting the En-In coding sequence into the EcoRI-HindIII site of pMAL-c2x after being digested with the same restriction

enzymes. Expression plasmid pIc-Ec was constructed by replacing the DNA between the NcoI and PstI sites of pKEB12 with the Ic-Ec coding sequence after being digested with the same restriction enzymes. All cloned DNA was verified by DNA sequence analysis.

DNA sequencing

DNA sequencing was performed on Applied Biosystems automated DNA sequencers (3100) using BigDye labeled dye-terminator chemistry (Applied Biosystems).

Gene expression and protein purification

Expression and purification of gene products using the MBP fusion and expression system were carried out as described (12). Modifications to the standard protocol and preparation of non-fusion enzyme were described previously (13). The purified enzyme, either the native enzyme or its MBP fusion form, was stored in 50% glycerol at -20°C . For purification of unfolded insoluble protein Ic-Ec or Ic-Ec($\nabla 9$), cells from one liter of culture were harvested and lysed by sonication in 50 ml of sonication buffer (10 mM Tris, pH 7.6, 50 mM NaCl and 1 mM EDTA). The insoluble pellets were saved, washed with 100 ml of high salt buffer (1.0 M NaCl) once, 100 ml distilled water twice, 100 ml 0.5% Triton X-100 twice, 100 ml distilled water twice, and then with 100 ml of sonication buffer. The pellets were saved and dissolved in 40 ml of 6 M guanidine hydrochloride containing 20 mM Tris, pH 7.6, 50 mM NaCl, 20 mM DTT and 1 mM EDTA. The solution, after clearing by centrifugation (14 000 r.p.m., JA-17 rotor, 20 min), was dialyzed against 500 ml of refolding buffer (20 mM Tris, pH 7.6, 50 mM NaCl, 2 mM DTT and 1 mM EDTA) with four changes at 4°C . The refolded soluble protein and unfolded precipitates were separated by centrifugation (12 000 r.p.m., JA-17 rotor, 20 min). About 10 mg of soluble Ic-Ec or Ic-Ec($\nabla 9$) were routinely obtained from one liter of induced culture. About 30 mg of MEN-In or MEN-In(9) were routinely purified using an amylose column from one liter of induced culture. For assembling dimeric SCD protein on an amylose column, the cells containing MEN-In or MEN-In(9) from 0.5 l of induced culture were harvested and lysed by sonication in 25 ml of sonication buffer on ice. The cell debris was removed by centrifugation (12 000 r.p.m., JA-17 rotor, 20 min) and the crude cellular extracts loaded onto the column. After briefly washing with two column volumes (40 ml) of refolding buffer, 10 mg of soluble Ic-Ec or Ic-Ec($\nabla 9$) was loaded onto the column. The flow-through fractions were collected and reloaded onto the column. The column was then washed with 500 ml of refolding buffer and eluted with the same buffer containing 10 mM maltose. SCD protein-containing fractions were pooled, dialyzed against refolding buffer containing 50% glycerol and kept at -20°C until use. About 20 mg of SCD proteins are obtained routinely by the procedure described above.

Protein analysis

Protein concentration was determined by the BioRad Protein Assay using BSA as a standard. Molecular weight and purity determinations were carried out by SDS-PAGE analysis and MALDI-ToF mass spectrometry (Voyager DE; Applied Biosystems Inc.). N-terminal protein sequence analysis was

performed on a Procise 494 Protein/Peptide Sequencer (Applied Biosystems Inc.).

Enzyme assay

For the structure-specific nuclease (resolution) activity assay, 1 μg of the cruciform-containing plasmid pUC(AT) in 20 μl of buffer (20 mM Tris, pH 7.6, 50 mM NaCl, 2 mM DTT and 0.15% Triton X-100 with either 2 mM MgCl_2 or 2 mM MnCl_2) was incubated with variable amounts of purified SCD protein or T7 Endo I at 37°C for 60 min. The digests were then analyzed by agarose gel-electrophoresis. One unit of the activity is defined as the amount of enzyme required to convert 1 μg of supercoiled pUC(AT) DNA to its linear or nicked form in 60 min. For non-specific nuclease activity, 1 U of activity is defined in the same way as described above, but using plasmid pUC19 as substrate. 2-log DNA ladder is a mixture of 20 DNA molecules with different sequences and sizes (from 10 to 0.1 kb) and is commonly used as a size marker for DNA agarose gel analysis. The large DNA molecules in 2-log DNA ladder provide a sensitive substrate for detection of non-specific endonucleases, especially for double-strand cleavage nucleases. The small DNA molecules in the ladder are useful for detection of exonucleases. We used 2-log DNA ladder substrates to detect the nuclease activities of SCD proteins. Quantitative analysis of DNA or protein bands on the gel was performed on a BioRad Phosphoimage autodensitometer. Since the maltose-binding protein-T7 endonuclease I fusion, named MBP-Endo I or ME, is fully active and has the same specific activity as non-fused enzyme after taking the molecular weights into account (9), we used the activity of ME as the standard for comparison with the activity of SCD protein.

Measuring the ratio of specific to non-specific nuclease activity of SCD protein

The structure-specific (resolution) activity on pUC(AT) should result in cleavage site specifically, at the extruded hairpin, as shown before (9). To estimate the ratio of the structure-specific nuclease activity to the non-specific activity of SCD protein, pUC(AT) was partially digested by SCD protein. The digests were resolved on an agarose gel, and the linear form plasmids were purified from the gel. The purified linear plasmids were then digested with *DrdI* restriction enzyme. The final digests were again resolved on an agarose gel, and the DNA bands of restriction fragments on the gel were quantitatively analyzed.

The ratio of specific activity to non-specific activity was calculated as follows. Suppose that a fraction (S) of pUC(AT) substrates was linearized by the specific activity of SCD protein at the cruciform site and the remaining fraction ($1 - S$) of substrates was cleaved at variable sites by the non-specific activity. The probability of a DNA sequence being cut by the non-specific nuclease activity is assumed to be proportional to the sequence length. The probability of the large *DrdI* fragment on pUC(AT) being cleaved by the non-specific activity of SCD protein is $\alpha(1 - S)$, where α = the length of the large fragment/the total length of pUC(AT) = 0.71. Since the cruciform structure is located in the small *DrdI* fragment, the probability of the small *DrdI* fragment of pUC(AT) being cleaved by either the specific or non-specific activity of SCD is $\beta(1 - S) + S$, where β = the length of the small

DrdI fragment/the total length of pUC(AT) = 0.29. The survival rate of the large *DrdI* fragment from SCD protein digestion, i.e. the probability that the large fragment is not cut by the non-specific activity and remains in the final digests, is $1 - \alpha(1 - S) = \beta + \alpha S$, since $\alpha + \beta = 1$. The survival rate for the small *DrdI* fragment, i.e. the probability that the small fragment is not cut by either the non-specific or the specific activity and remains in the final digests, is $\alpha - \alpha S$. Since the mass of a DNA fragment is proportional to its length, the mass of large *DrdI* fragment in the final digests M_l is $M_0(\alpha\beta + \alpha^2 S)$; the mass of small *DrdI* fragment in the final digests M_s is $M_0(\alpha\beta - \alpha\beta S)$, where M_0 is the amount of purified linear plasmid substrates used for each test. $M_l/M_s = \gamma = (\beta + \alpha S)/(\beta - \beta S)$, then S is determined by the equation:

$$S = \frac{\beta\gamma - \beta}{\alpha + \beta\gamma} = \frac{0.29\gamma - 0.29}{0.71 + 0.29\gamma}, \quad 1 \geq S \geq 0; \infty \geq \gamma \geq 1.$$

RESULTS

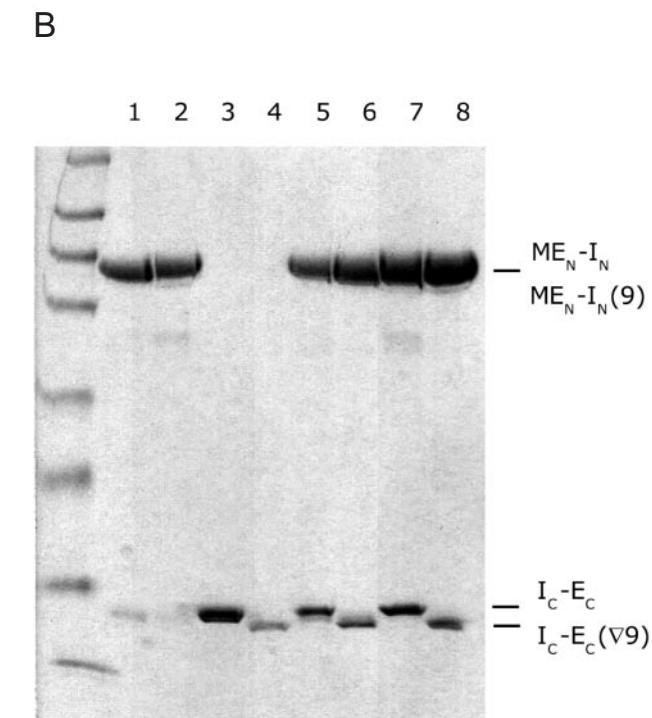
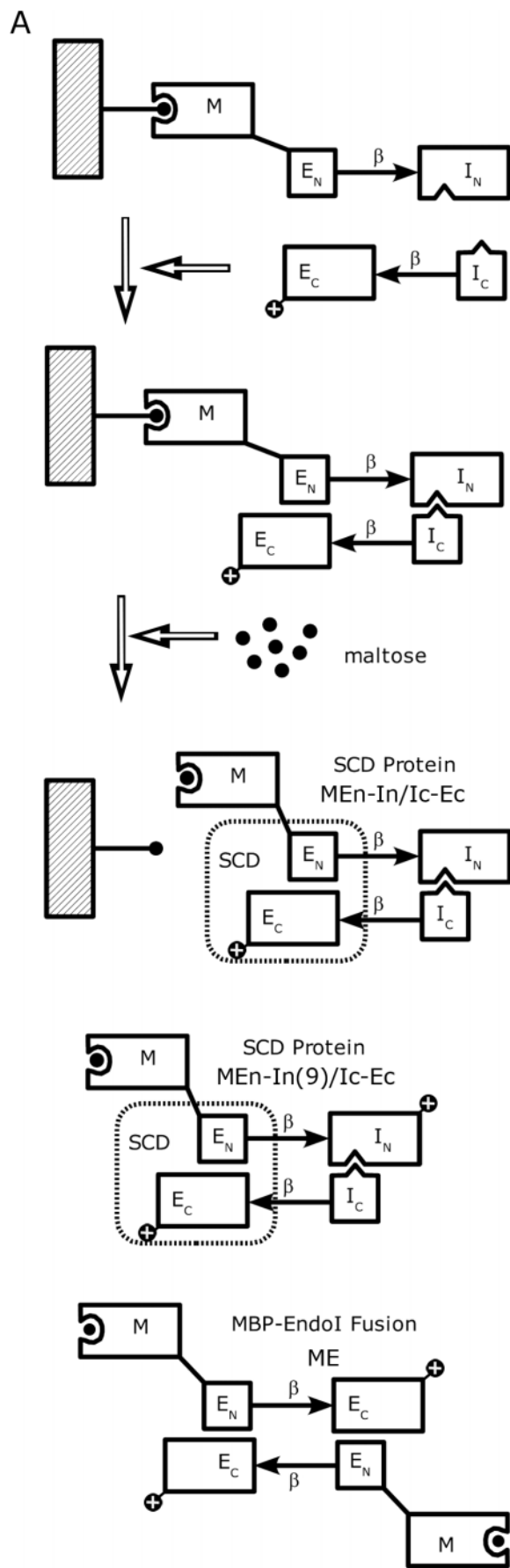
Making a stable single catalytic domain of T7 Endo I

T7 Endo I is a symmetrical homodimer. It consists of two catalytic domains connected by a β -sheet bridge. Each domain is comprised of residues 17–44 from one subunit and residues 50–145 from other subunit (6). Initial attempts were made to produce a SCD protein as a heterodimer of an N-terminal part (residues 1–46) and a C-terminal part (residues 47–149) of T7 Endo I by co-expression of separately-translated fragments. A second approach was made to fragment the full-length homodimer into two parts by introducing a protease (Genenase) cleavage site (P46H, A47Y) into the center of β -bridge of the full-length enzyme followed by cleavage of the purified mutant enzyme with Genenase. Neither approach resulted in a stable SCD protein dimer. The failure to directly produce this SCD protein might be due to instability of the heterodimeric SCD lacking the stabilizing bridge of the full-length enzyme.

In order to facilitate folding of SCD into a stable heterodimer, we added a dimerization functionality to the separately-translated catalytic-domain fragments. This approach employed a *trans*-splicing intein system from the *DnaE* gene of *Synechocystis* sp. PCC6803 (10). This system is known to mediate tight association of fusion partners, sufficient to restore catalytic activity to separated enzyme domains (14).

The experimental procedure is illustrated in Figure 1A. Briefly, two expression plasmids, pMEN-In and pIc-Ec, were constructed as described in Materials and Methods. The two plasmids were used to produce two polypeptides MEN-In and Ic-Ec. Ic-Ec was insoluble when expressed in *E.coli*. A soluble form of Ic-Ec was obtained by refolding the polypeptide *in vitro*. For assembly and purification of SCD protein, MEN-In/Ic-Ec, MEN-In was first bound to an amylose column followed by passage of excess amounts of soluble Ic-Ec through the column. After washing out the unbound Ic-Ec, the dimeric SCD protein MEN-In/Ic-Ec was eluted from the column with maltose buffer.

Derivatives of MEN-In/Ic-Ec were also made by the same method. In MEN-In/Ic-Ec(∇ 9) the last nine residues (RLKRKGGKK) of T7 Endo I (called the C-tail), were



removed from the C-terminus of Ic-Ec. In MEN-In(9)/Ic-Ec, an additional C-tail was added to the C-terminus of MEN-In. In MEN-In(9)/Ic-Ec(V9) the C-tail was transferred from Ic-Ec to the C-terminus of MEN-In. Each of the SCD proteins contains a SCD of T7 Endo I. Since MBP-Endo I is fully active (9), presumably the purified MBP-SCD fusions (Figure 1B) also would be active, provided that the SCD in these fusion proteins are folded correctly.

Non-specific nuclease activity of SCD protein

Purified SCD protein MEN-In/Ic-Ec was incubated with a DNA mixture, the 2-log DNA ladder (NEB), in either Mg^{2+} or Mn^{2+} buffer. The digests were resolved on an agarose gel (Figure 2A). The smearing of DNA bands showed that SCD protein was an active nuclease. The large DNA molecules in the substrates were cleaved sooner than the small ones, indicating that SCD protein possessed a non-specific endonuclease activity. The activity in Mn^{2+} buffer was 5–10 times higher than that in Mg^{2+} . This Mn^{2+} dependency is similar to that of the β -bridge mutants of full-length enzyme (9). Neither

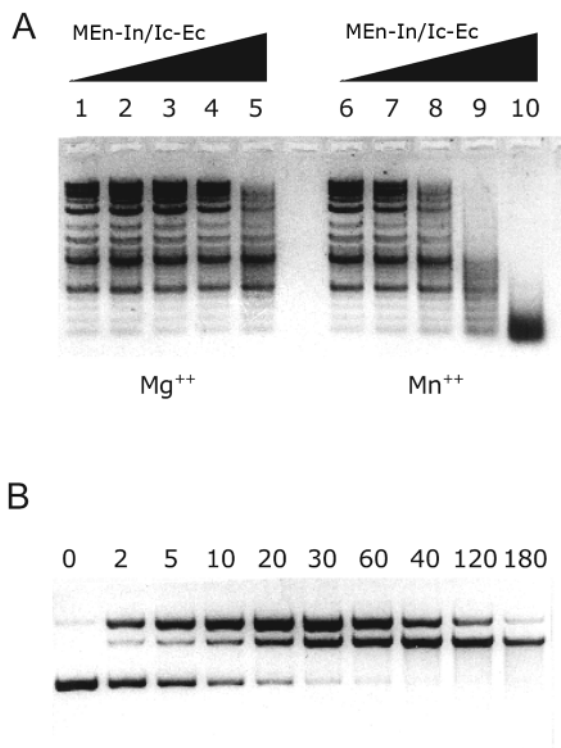


Figure 2. (A) Determination of non-specific nuclease activity of SCD protein. Variable amounts of purified MEN-In/Ic-Ec were incubated with 1 μ g of 2-log DNA ladder (NEB) in 20 μ l of either Mg²⁺ or Mn²⁺ buffer at 37°C for 1 h. The digests were resolved on a 1.2% agarose gel. From lane 1 to 5, digests with 0.00, 0.125, 0.25, 0.5 and 1.0 μ g of MEN-In/Ic-Ec in Mg²⁺ buffer, respectively. From lane 6 to 10, digests are the same as from lane 1 to 5 except for using Mn²⁺ buffer. (B) Identification of nicked strands as intermediate products of reaction by SCD protein. Plasmid pUC19 was incubated with MEN-In/Ic-Ec in Mg²⁺ buffer at 37°C. At variable time point 0, 2, 5, 10, 20, 30, 40, 60, 120 and 180 min an aliquot of sample was withdrawn from the reaction and resolved on an agarose gel. L, N and S stand for linear, nicked and supercoiled plasmids, respectively.

MEN-In nor Ic-Ec by itself had any detectable nuclease activity (data not shown).

When supercoiled pUC19 plasmids were used as substrates, the results of a time course experiment showed that the first intermediate products were nicked plasmids (Figure 2B), indicating that SCD protein was a nicking endonuclease. However, the nicked intermediates were immediately chased into linear form plasmids. This relationship suggested that SCD protein possessed not only a non-specific nicking endonuclease activity, but might also have, similar to β -bridge mutants (9), a specific cleavage activity targeted to the strand of opposite already-nicked sites.

Nicked-site cleavage activity

We have reported that T7 Endo I and its β -bridge mutants recognize a nicked-site on DNA and cleave nearby opposite the nick (9). This will be called 'nicked-site activity'. The results of the time-course experiment with SCD protein above suggested that SCD might also possess a specific nicked-site activity. To confirm the nicked-site cleavage activity for the SCD protein, we employed a site-specifically nicked DNA substrate generated previously (9). The substrate is 2.5 kb

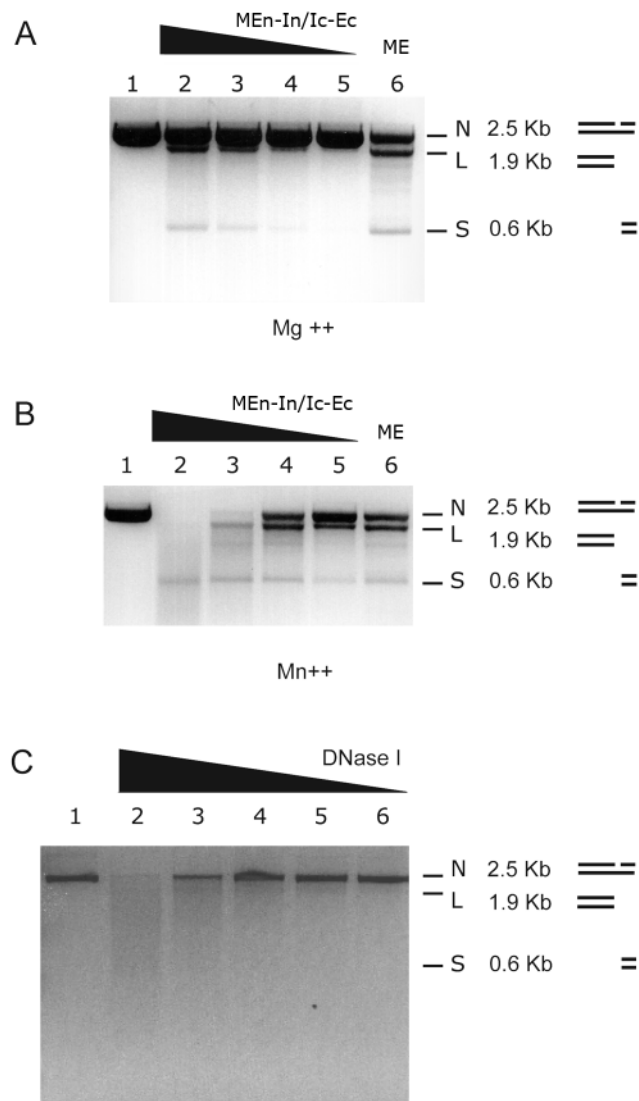


Figure 3. Determination cleavage activity opposite preexisting nicks of SCD protein. Site-specifically nicked substrate (0.5–1.0 μ g) were incubated with variable amounts of nuclease at 37°C for 60 min. The digests were resolved on an agarose gel. (A) Reaction in Mg²⁺ buffer, 0.0, 0.5, 0.25, 0.125 and 0.06 μ g of MEN-In/Ic-Ec and 0.03 μ g of ME were included in lanes 1, 2, 3, 4, 5 and 6, respectively. (B) All are the same as in (A) except for using Mn²⁺ buffer. (C) 0, 2, 1, 0.5, 0.25 and 0.125 $\times 10^{-3}$ U of bovine DNase I were included in lanes 1, 2, 3, 4, 5 and 6, respectively. Reaction took place at 37°C for 30 min in Mn²⁺ buffer.

linear DNA containing a single nick 600 bp from one end. Cleavage at the nicked-site gives rise to two distinctive 1.9 kb and 600 bp fragments. The nicked DNA substrate was incubated with MEN-In/Ic-Ec in either Mg²⁺ or Mn²⁺ buffer. The results clearly showed that SCD protein indeed possesses a specific nicked-site activity (Figure 3A). Similar to β -bridge mutants (9), the nicked-site activity in Mn²⁺ buffer was higher than that in Mg²⁺ buffer (Figure 3A and B). In a control experiment, a commonly used non-specific nicking endonuclease, bovine DNase I (15) was incubated with the site-specifically nicked substrate. The results showed that it cleaved this nicked substrate at variable positions, as no 1.9 kb and 600 bp products could be identified by agarose

Table 1. Enzyme activity (U/ μ g) of T7 Endo I and SCD proteins on different substrates

Substrate	ME	ME _N -In/ Ic-Ec	ME _N -In/ Ic-Ec(Δ 9)	ME _N -In(9)/ Ic-Ec	ME _N -In(9)/ Ic-Ec(Δ 9)
pUC19	16	1	<0.005	6	<0.02
pUC(AT)	600	1	NA	6	NA

The enzyme assays were carried out as described in Materials and Methods. Briefly, 1 μ g of plasmid substrate DNA in 20 μ l of Mg²⁺ buffer (20 mM Tris, pH 7.6, 50 mM NaCl, 2 mM DTT, 2 mM Mg²⁺ and 0.15% Triton X-100) is incubated with a variable amount of enzyme (made by 1–2 dilution) at 37°C for 60 min. The digests were resolved on an agarose gel. A unit of enzyme activity is defined as the amount of enzyme required to convert 1 μ g of supercoiled plasmid DNA to its linear or nicked form at 37°C in 60 min. Notice that pUC(AT) containing a cruciform structure can be cleaved by the resolution activity of the whole enzyme and by the structure-specific nicking endonuclease activity of SCD proteins at cruciform site or by the non-specific nicking endonuclease activity of the whole enzyme or SCD proteins at variable sites. Substrate pUC19 does not contain a cruciform structure and can only be cleaved by the non-specific nicking endonuclease activity of the whole enzyme or SCD proteins. For SCD protein, the assay is designed to measure the nicking endonuclease activity, although the nicked plasmid can be further linearized by its nicked site cleavage activity. NA, not tested.

gel-electrophoresis (Figure 3C), indicating that DNase I did not cleave the DNA specifically opposite a nicked-site, even in Mn²⁺ buffer. These results suggested that the nicked-site activity observed for the full-length enzyme and its mutants might actually be the intrinsic property of the individual catalytic domain.

Characterization of the ends produced by cleaving the nicked DNA substrate with SCD protein was carried out by the method as described previously (9) and the results showed that SCD protein cut the DNA strand nearby and opposite the nick, producing 1–2 bp 5' overhangs (mostly 2 bp). These extensions were slightly shorter than that the 3–4 bp 5' overhangs produced by the full-length enzyme (9).

Function of C-terminal peptide

T7 Endo I is a very basic protein (pI = 9.5). The last nine amino acid residues of T7 Endo I, i.e. the C-tail, of which six are either arginine or lysine, play an essential role in the enzyme activity. Removal of the C-tails from the whole enzyme resulted in an 100-fold or more reduction in enzyme activity (16). Replacement of C-tails by other positively charged peptides or even by synthesized positively-charged polymers gave rise to an active enzyme (Dr Lixin Chen and Thomas Evans, personal communication). It was believed that the function of the C-tail was to increase the affinity of the enzyme for DNA in general. Using pUC19 as substrate in Mg²⁺ buffer, the non-specific activity of ME was estimated as ~16 U/ μ g, or 1.0 U/ pmol of domain. For ME_N-In/Ic-Ec, it was only ~1 U/ μ g, or 0.08 U/pmol of domain (Table 1). Therefore, the activity of the SCD protein is ~10-fold lower than that of the whole enzyme after normalization. To investigate if increasing the total positive charge of SCD protein could increase the activity, we mimicked the structure of the whole enzyme (6) by adding a C-tail to the C-terminal of ME_N-In (Figure 1). Interestingly, the resulting SCD protein ME_N-In(9)/Ic-Ec became about four times more active than ME_N-In/Ic-Ec in either Mg²⁺ or Mn²⁺ buffer (Table 1 and Figure 4). These results showed that non-specific binding of

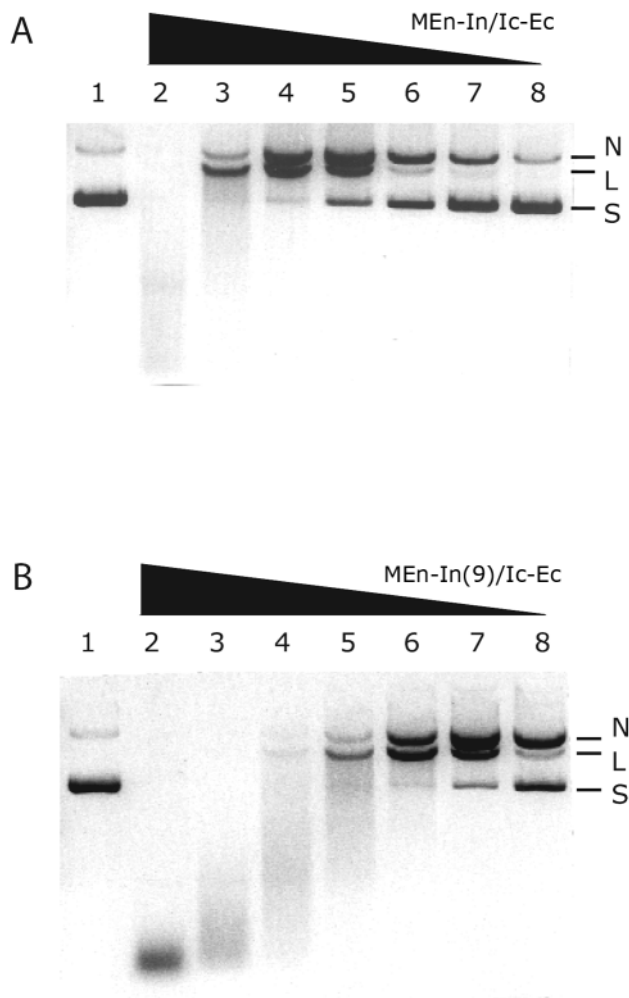


Figure 4. Determination of activity enhancement by distantly located C-tail. pUC19 (1 μ g) was incubated with variable amounts of SCD protein at 37°C for 1 h in Mn²⁺ buffer. The digests were resolved on an agarose gel. (A) 0.0, 1.0, 0.5, 0.25, 0.125, 0.063, 0.032 and 0.016 μ g of ME_N-In/Ic-Ec were included in lanes 1, 2, 3, 4, 5, 6, 7 and 8, respectively. L, N and S stand for linear, nicked and supercoiled form plasmids, respectively. (B) All are the same as in (A) except for ME_N-In(9)/Ic-Ec being used. The enzyme activities of ME_N-In/Ic-Ec and ME_N-In(9)/Ic-Ec were estimated as ~5 and 20 U/ μ g, respectively, in these tests. The figure presents the activity in presence of Mn²⁺. Enzyme activities assayed in Mg²⁺ buffer for the same proteins were qualitatively similar, but ~5-fold lower as listed in Table 1.

SCD protein to DNA by a C-tail at a site distant from the catalytic center could considerably increase the rate of formation of productive substrate–enzyme complex, suggesting that non-specific binding of T7 Endo I to substrate through the C-tail of one catalytic domain may also increase the enzymatic activity of the other catalytic domain. The relatively low affinity of SCD protein for duplex DNA suggested by this study is consistent with the previous report that binding T7 Endo I to duplex DNA is a thousand times weaker than to four-way junction DNA (4).

Removal of the C-tail from ME_N-In/Ic-Ec, however, resulted in loss of the activity of SCD protein. The activity of ME_N-In/Ic-Ec(Δ 9) was estimated to be at least a 200-fold lower than that of ME_N-In/Ic-Ec (Figure 5A and B). Adding a C-tail back to ME_N-In/Ic-Ec(Δ 9) at a different location

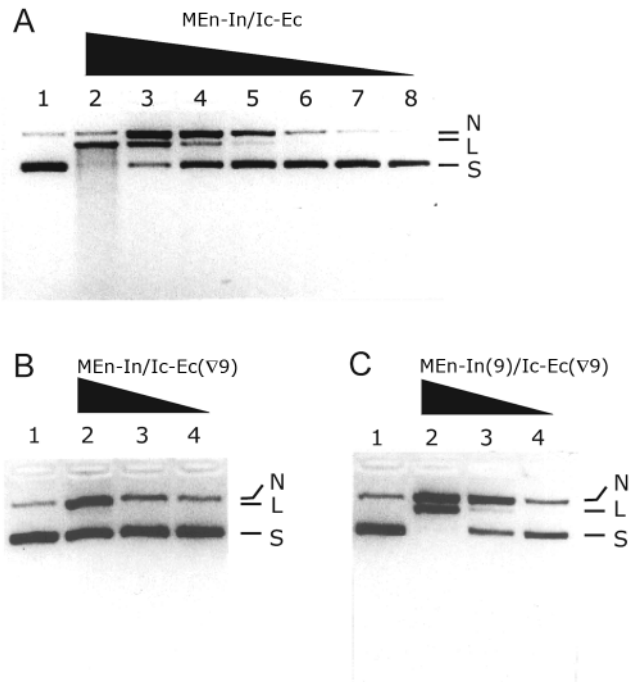


Figure 5. Comparison of activities among SCD proteins with C-tail at different locations. pUC19 (1 μg) was incubated with variable amounts of SCD protein at 37°C overnight in Mg^{2+} buffer. The digests were resolved on agarose gel. (A) 0.0, 100, 50, 25, 12.5, 6.3, 3.2 and 1.6 ng of MEN-In/Ic-Ec were included in lanes 1, 2, 3, 4, 5, 6, 7 and 8, respectively. (B) 0.0, 4.0, 2.0 and 1.0 μg of MEN-In/Ic-Ec(v9) were included in lanes 1, 2, 3 and 4, respectively. (C) 0.0, 4.0, 2.0 and 1.0 μg of MEN-In(9)/Ic-Ec(v9) were included in lanes 1, 2, 3 and 4, respectively. L, N and S stand for linear, nicked and supercoiled form plasmids, respectively.

(to the C-terminal of MEN-In) did not significantly restore the activity. The activity of resulting SCD protein MEN-In(9)/Ic-Ec(v9) was still ~ 50 – 100 times lower than that of MEN-In/Ic-Ec (Figure 5C). These results showed that the C-tail on a catalytic domain was essential for the activity of the domain and the distally located one could enhance the activity only when the proximal one was present, suggesting that although the distant C-tail might enhance the activity of SCD by increasing the general affinity of the protein for DNA, the proximal one is indispensable for formation of the productive substrate–enzyme complex.

Activity on cruciform structures

A cruciform structure stabilized on a negatively supercoiled plasmid is similar to a four-way junction. T7 Endo I resolves both structures with high efficiency (4). To determine whether SCD protein is able to recognize and cleave DNA at a cruciform site we performed the enzyme activity assay for the native enzyme and SCD proteins on both structure-specific substrate pUC(AT) and non-specific substrate pUC19. The results of the assays are listed in Table 1. Cruciform-containing plasmid pUC(AT) was made by replacing the multiple cloning site sequence of pUC19 with 40 A/T base pair repeats, and used to assay the junction-resolving or the structure-specific nicking endonuclease activity of enzyme; regular plasmid pUC19 was used for assaying the non-specific nicking endonuclease activity of enzyme. Using

pUC(AT) as substrate the enzyme activity (resolution activity) of ME was estimated as 600 U/ μg ; using pUC19 as substrate the non-specific nicking endonuclease activity of ME was only ~ 16 U/ μg . These results show that the native enzyme recognizes and resolves cruciform DNA efficiently. However, the measured enzyme activities of SCD protein on pUC(AT) and on pUC19 were similar. For MEN-In/Ic-Ec it was about 1 U/ μg no matter which substrate was used. For MEN-In(9)/Ic-Ec it was ~ 6 U/ μg (Table 1). These results suggest that SCD does not recognize cruciform structure efficiently. We did see that the nicking endonuclease activity of SCD protein on pUC(AT) was slightly higher than that on pUC19 after performing careful titration assays. We felt that SCD protein might have a weak specificity on cruciform structure and the enzyme activity assay (regular enzyme titration assay) used was not sensitive enough for detection of this weak cruciform-specific activity of SCD protein when masked by a major non-specific activity.

To verify the existence of a structure-specific nicking endonuclease activity for SCD protein, we carried out experiments trying to estimate the ratio of the structure-specific activity to the non-specific activity of SCD protein. First, cruciform-containing plasmid pUC(AT) was partially digested by either MEN-In/Ic-Ec or MEN-In(9)/Ic-Ec in Mg^{2+} buffer. The reactions were carried out until about equal amounts of linear and nicked plasmids were present in the digestion mixture. The digests were resolved on an agarose gel. The linear form plasmids were then purified from the gel followed by cutting with *DrdI* restriction enzyme. The final digests were resolved on an agarose gel for quantitative analysis (Figure 6A and B).

In a positive control experiment, pUC(AT) was linearized by the whole enzyme ME first. The purified linear products were then cut by *DrdI*. In a negative control experiment, pUC19 plasmids linearized by MEN-In/Ic-Ec were used as substrates for *DrdI* digestion. The results showed that the whole enzyme linearized pUC(AT) by its resolution activity almost exclusively at the cruciform site. The sequential restriction digestion by *DrdI* gave rise to three major DNA fragments sized as 2.0, 0.5 and 0.3 kb as expected (Figure 6B, lanes 3 and 4). The three fragments accounted for almost 100% of the DNA mass of the linear substrates. There was a very faint 0.8 kb band on the gel, indicating that a very small fraction of pUC(AT) was originally cleaved open by the non-specific nicking endonuclease activity plus the nicked-site cleavage activity of ME. The γ -value was estimated as ~ 50 – 100 (Materials and Methods) and S was ~ 0.95 , i.e. 95% of pUC(AT) plasmids were cleaved open at the cruciform site by ME. This number was quite close to the predicted one, 0.97 ($1 - 16\text{U}/600\text{U}$). In the negative control experiment (Figure 6B, lanes 9 and 10), only two identifiable bands were visible, sized as 2.0 and 0.8 kb, which contained about equal amounts of DNA, i.e. $\gamma = 1.0$. The calculated $S = 0$. This was expected since there was no cruciform structure on pUC19. The total DNA in the two bands was $\sim 40\%$ of the original mass of linear substrates. This was also consistent with the calculated value (42%). The results of two control experiments proved the validity of the method presented in Materials and Methods.

When the linear pUC(AT) substrates produced by MEN-In/Ic-Ec or MEN-In(9)/Ic-Ec were used for *DrdI* digestion, there were two major bands, sized as 2.0 and 0.8 kb, on the gel

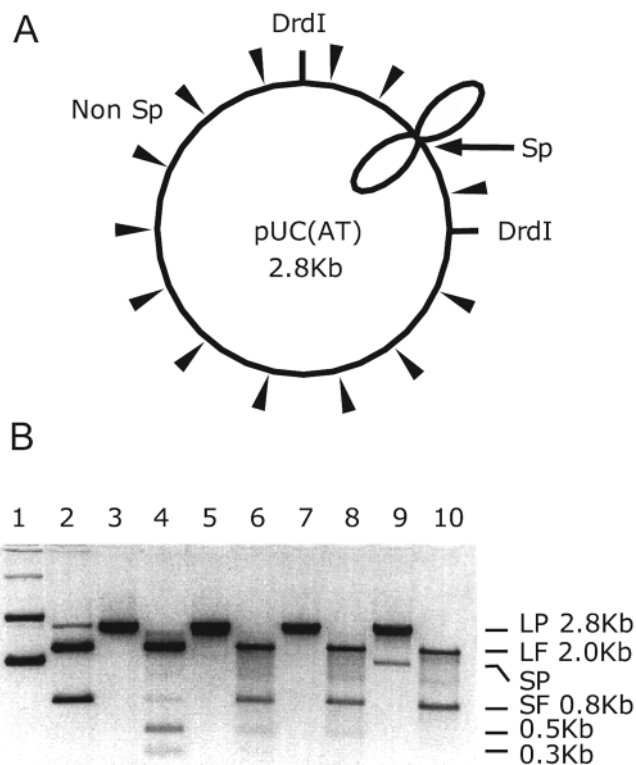


Figure 6. Determination of the ratio of structure-specific to non-specific activity of SCD protein by agarose gel-electrophoresis. (A) Schematic illustration of possible linearization sites on pUC(AT) by both specific and non-specific activities of SCD protein. The long arrow represents the specific activity (Sp) that leads the plasmid to open at the cruciform site. The short arrows represent the non-specific activity (Non sp) that leads the plasmid to open at variable sites. (B) LP and SP stand for linear and supercoiled form plasmids, respectively; LF and SF for the large and the small fragments produced by DrdI digestion respectively. Lane 1, pUC(AT); lane 2, pUC(AT) cut by DrdI (a small amount of linear plasmid was produced by incomplete digestion); lane 3, linear pUC(AT) produced by T7 Endo I; lane 4, the DNA in lane 3 cut by DrdI; lane 5, linear pUC(AT) produced by MEN-In/Ic-Ec; lane 6, the DNA in lane 5 cut by DrdI; lane 7, linear pUC(AT) produced by MEN-In(9)/Ic-Ec; lane 8, the DNA in lane 7 cut by DrdI; lane 9, linear pUC19 produced by MEN-In/Ic-Ec (a small amount of supercoiled plasmid is co-purified with the linear form); lane 10, the DNA in lane 9 cut by DrdI. Each lane contained $\sim 1 \mu\text{g}$ of DNA. All the linear plasmids used in the assay were gel-purified.

(Figure 6B, lanes 6 and 8). Digestion of pUC(AT) with DrdI also produced the same size fragments (Figure 6B, lane 2). The differences from the direct restriction digestion were that the total mass of the two fragments produced by sequential digestion of SCD protein and DrdI accounted for only $\sim 50\%$ of the linear substrates used in the test; the other half of the products appeared as a smear background on the gel because of their variable sizes, indicating that most of the pUC(AT) was cleaved open by the non-specific nicking endonuclease activity plus the nicked-site cleavage activity at variable sites. We did see two faint 0.5 and 0.3 kb bands in SCD protein plus DrdI digests (Figure 6B, lanes 6 and 8), indicating that SCD protein had a weak cruciform-specific nuclease activity. We measured the ratio of DNA mass in the 2.0 kb band/DNA mass in the 0.8 kb band, i.e. the γ -value. It was ~ 1.5 – 2.0 . The calculated S was ~ 0.15 – 0.2 (Materials and Methods).

This meant that at least 15–20% of pUC(AT) was initially nicked at the cruciform site by the structure-specific nicking endonuclease activity of SCD protein. Nicking pUC(AT) resulted in relaxation of the plasmid and loss of the cruciform structure. The nicked plasmids were then cut open by the nicked-site cleavage activity of SCD protein described above. The S -values for MEN-In/Ic-Ec and MEN-In(9)/Ic-Ec were similar, indicating that the additional C-tail on SCD protein did not increase the cruciform recognition efficiency by SCD, although the activity of MEN-In(9)/Ic-Ec, in general, was ~ 5 – 10 times higher than that of MEN-In/Ic-Ec. Nevertheless, the structure-specific nicking endonuclease activity of SCD protein on cruciform structure was no more than 1% of the resolution activity of the whole enzyme ($6 \text{ U} \times 50\%/600 \text{ U} = 0.5\%$).

DISCUSSION

The broad substrate specificities of T7 endonuclease I make it difficult to decipher the common DNA structure that is specifically recognized by the enzyme. Based on the data from study on β -bridge mutants (9) and the published crystal structure of the whole enzyme (6) we have proposed that catalytic domain of T7 Endo I is a non-specific nicking endonuclease. SCD is unable to specifically recognize four-way junction or cruciform DNA. Two catalytic domains of T7 Endo I are juxtaposed in space by a double β -sheet linkage, or the β -bridge. The enzyme retains certain flexibility around the bridge site to allow conformational changes during substrate–enzyme interaction. Efficient resolution of a four-way junction or cruciform DNA by the enzyme is achieved by simultaneously binding the two catalytic centers to two of the arms crossing the junction. That results in the formation of productive enzyme–substrate complex (9). This model simply explains the properties of the β -bridge mutants and the broad activity of T7 Endo I on a variety of substrates with variable structures, from high activity on four-way junctions to low activity on regular double-stranded DNA. To prove this model to be the substrate–enzyme recognition mechanism of T7 Endo I, in addition to structural and mutagenesis analysis of the whole enzyme, it is necessary to isolate an SCD of the enzyme and to characterize its enzymatic activity.

The desired SCD protein is a heterodimer comprising two polypeptides (Figure 1A). In order to facilitate the folding of SCD and to enhance the dimer stability we employed a *trans*-splicing intein system (10). Using the *trans*-splicing intein to produce SCD protein was based on the following considerations. First, dimerization between the N-terminal part (In) and the C-terminal part (Ic) of the intein is spontaneous, either *in vivo* or *in vitro*, and the In/Ic dimer is very stable. Second, the N-terminal of In and the C-terminal of Ic in the In/Ic dimer, namely the splicing site of the intein, are close to each other and can bring the two polypeptides comprising SCD together in the natural position through the β -bridge to facilitate SCD folding. Changing the first cysteine residue of In to alanine can prevent the splicing reaction without interfering with dimerization of In/Ic. Finally, the size of In/Ic dimer (123 + 36 amino acids) is similar to the size of the replaced part, i.e. the other catalytic domain (43 + 100 amino acids) of T7 Endo I (Figure 1A). We still used MBP as an affinity tag for purification of SCD protein since MBP-Endo I (ME)

was fully active (9). Fusion with MBP can, in general, help the fused polypeptide with folding and increasing solubility (17).

As expected from the proposed model, purified SCD protein was a non-specific nicking endonuclease. The non-specific nuclease activity of SCD protein MEn-In(9)/Ic-Ec was comparable with that of the whole enzyme or its β -bridge mutants after normalization (1.0 U/pmol of domain in ME and 0.4 U/pmol of domain in MEn-In(9)/Ic-Ec in Mg^{2+} buffer), indicating that the SCD protein folded correctly. Similar to the β -bridge mutants (9), the activity of SCD protein in Mn^{2+} buffer was ~ 5 –10 times higher than that in Mg^{2+} buffer.

It was intriguing that SCD protein intrinsically recognizes a nicked-site and cleaves the DNA strand opposite the nick. Here we have created a non-specific nicking endonuclease with the ability to cleave DNA at a nicked-site opposite the nick. The net results of digestion are similar to that by a non-specific double-cleavage endonuclease. Many properties of SCD protein are similar to that of the β -bridge mutants, especially similar to ME(∇ PA). ME(∇ PA) also cleaves DNA substrates by a nicking plus nicking against the nick mechanism except that ME(∇ PA) still recognizes cruciform structures with a high efficiency (9). It became clear that the observed activities of T7 Endo I could be divided into two categories. Some activities, such as the non-specific nicking endonuclease and nicked-site cleavage activities, may actually be the results of uncoordinated actions of catalytic domains; others, such as resolution of four-way junctions, belong to the whole enzyme. To understand the mechanism of substrate recognition by the whole enzyme it is necessary to distinguish the activities solely belonging to the whole enzyme from those possibly belonging to the individual SCDs.

The C-tail comprising the last nine amino residues (RLKRKGGKK) of T7 Endo I polypeptide is highly positively charged. Removal of these charges from the enzyme significantly reduces the enzymatic activity (16). The 3D structural analysis of T7 Endo I has shown that the two C-tails are located on the surface of the protein and the last four residues of the C-tail are highly flexible and disordered in the crystal (6). The C-tail was believed to increase affinity of the enzyme for DNA in general since it could be replaced by other positively charged peptides, even by synthesized positively-charged polymers, without significant impairment of the enzyme activity. Our data showed that the location of the positively charged C-tail relative to the catalytic center was crucial for the activity of SCD. Although an additional distantly located C-tail could increase the activity of SCD by a 5- to 10-fold, presumably by increasing general affinity of SCD protein to DNA, the normal C-tail portion of SCD was indispensable for formation of the productive enzyme-substrate complex and its removal resulted in loss of activity. It is very probable that these different roles played by differentially located C-tails in the SCD protein hold for the whole enzyme. In the whole enzyme, a C-tail should be involved in the productive complex formation of the catalytic domain in which it is located. It should also enhance the activity of other catalytic domain by non-specifically binding to DNA. Productively binding both catalytic domains of the whole enzyme to duplex DNA should be rare and mutually exclusive. Nevertheless, both C-tails on the whole enzyme would be required for obtaining a fully active enzyme and efficiently resolving a four-way junction. The functional analysis of the

two C-tails on the whole enzyme should shed light on the mechanism used by the enzyme for searching for and recognizing a four-way junction along duplex DNA.

Although the non-specific nuclease activities for both SCD protein and whole enzyme are comparable, the whole enzyme recognizes and cleaves cruciform DNA at least hundreds times more efficiently than SCD protein does. An additional C-tail on SCD protein resulted in increase of the activity of the SCD in general, but it nevertheless did not change the recognition efficiency of SCD to nicked-site and cruciform structure against duplex DNA. The 3D structure of the enzyme-substrate complex of T7 endonuclease I is not available yet. Based on their experimental data and the structures of T7 endonuclease I and the BglI-DNA complex, Lilley and co-workers were able to produce an enzyme-substrate complex model by graphical modeling techniques that permits simultaneous contacts between the catalytic sites and the two continuous strands of a four-way junction. This facilitates the coordinated cleavage of the two strands, leading to a productive resolution of the junction (5). However, change of the reciprocal position of the two catalytic sites by introduction of mutations such as amino acid deletions, insertions and substitutions at the β -bridge site of the enzyme results in active mutants that resolve cruciform DNA effectively. The deletion mutant ME(∇ PA) recognizes a cruciform structure efficiently, although it can no longer coordinately cleave or resolve the structure (9). It is difficult to accommodate the experimental data of β -bridge mutants within the model produced by computer modeling. The high efficiency of recognizing a four-way junction or a cruciform structure by the whole enzyme or its active mutants is most probably realized by a cooperative effect of simultaneously and specifically binding the two catalytic domains to the two duplex DNA arms across the junction. Our results show that both the SCD protein and the whole enzyme possess similar non-specific nicking endonuclease activities and nicked-site cleavage activities on duplex DNA. Using pUC19 as a substrate, the time course experiments with both enzymes produce similar results: the first intermediate products were nicked plasmids produced by the nicking activity, the nicked plasmids were immediately chased into linear plasmids by the nicked-site cleavage activity, and finally the linear plasmids were fragmented by both the activities. The results of time-course experiments also suggest that the ratios of the nicking to the nicked-site cleavage activity for both enzymes are similar because the distribution pattern of the intermediate products in the time-course for both enzymes are similar (the data from the time course experiment with the whole enzyme are not shown in this paper). This suggests that the activity of the whole enzyme on duplex DNA actually results in the uncoordinated action of its catalytic domains. Therefore we can reasonably assume that the binding activities of the catalytic domain on duplex DNA in either SCD protein or the whole enzyme are similar since the enzyme activities of both enzymes on duplex DNA are similar. The binding activity of the whole enzyme at a four-way junction is 1000-fold higher than to duplex DNA with the same sequence (4). The binding activity of the whole enzyme at a four-way junction should also be ~ 1000 -fold higher than the binding activity of SCD protein on duplex DNA. The results of cleaving the cruciform-containing plasmid pUC(AT) with SCD protein indicate that the cruciform

structure on the plasmid competes for binding with the rest of the duplex DNA poorly, since only ~20% of the plasmids are initially cleaved by the enzyme at the cruciform site, suggesting that SCD binds to a four-way junction only slightly better than to duplex DNA. This means that the binding activity of the whole enzyme at a four-way junction should be at least a 100-fold higher than the sum of the binding activity of two SCDs on either duplex or junction DNA, suggesting the existence of cooperativity during simultaneous binding of the two catalytic domains of the whole enzyme to a four-way junction, although the nature and the energy source for this cooperativity so far are unknown. SCD protein by itself could recognize a nicked-site and, weakly, a cruciform structure, suggesting that SCD might actually recognize the structural deviation of the DNA double-helix near a nicked-site or strand-crossing site that may also play a role in the substrate–protein recognition of the whole enzyme.

T7 endonuclease I is a multifunctional enzyme. The whole enzyme can efficiently resolve four-way junctions and branched DNA produced during phage proliferation. The non-specific nuclease plus nick site cleavage activities of SCD serve well for breaking down host chromosomal DNA (18). Evolutionarily, the whole enzyme could have evolved by changing the folding pathway of an SCD enzyme that may have converted a monomeric non-specific nuclease into a dimeric structure-specific nuclease without loss of the original activity.

ACKNOWLEDGEMENTS

We thank our colleagues Lise Raleigh, Bill Jack and Paul Riggs for their helpful discussions. We also thank Don Comb for his generous support for this research. Funding to pay the Open Access publication charges for this article was provided by New England Biolabs, Inc.

Conflict of interest statement. None declared.

REFERENCES

- Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genet. Res.*, **5**, 282–304.
- Porker, C.N. and Halford, S.E. (1991) Dynamics of long-range interactions on DNA: the speed of synapsis during site-specific recombination by resolvase. *Cell*, **66**, 781–791.
- Aravind, L., Makarova, K.S. and Koonin, E.V. (2000) Holliday junction resolvase and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.
- Parkinson, M.J. and Lilley, D.M.J. (1997) The junction-resolving enzyme T7 endonuclease I: quaternary structure and interaction with DNA. *J. Mol. Biol.*, **270**, 169–178.
- Déclais, A., Fogg, J.M., Freman, A.D.J., Coste, F., Hadden, J.M., Phillips, S.E.V. and Lilley, D.M.J. (2003) The complex between a four-way junction DNA and T7 endonuclease I. *EMBO J.*, **22**, 1398–1409.
- Hadden, J.M., Convery, M.A., Déclais, A., Lilley, D.M.J. and Phillips, S.E.V. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. *Nature Struct. Biol.*, **8**, 62–67.
- Mashal, R.D., Koontz, J. and Sclar, J. (1995) Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genet.*, **9**, 177–183.
- White, M.F., Giraud-Panis, M.-J.E., Pöhler, J.R.G. and Lilley, D.M.J. (1997) Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J. Mol. Biol.*, **269**, 647–664.
- Guan, C., Kumar, S., Kucera, R. and Ewel, A. (2004) Changing the enzymatic activity of T7 endonuclease by mutations at the beta-bridge site: alteration of substrate specificity profile and metal ion requirements by mutation distant from the catalytic domain. *Biochemistry*, **43**, 4313–4322.
- Martin, D.D., Xu, M.Q. and Evan, T.C., Jr (2001) Characterization of a naturally occurring *trans*-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry*, **40**, 1393–1402.
- Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Riggs, P. (1992) Expression and purification of maltose-binding protein fusions. In Ausbel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds), *Current Protocols in Molecular Biology*. Green Publishing and Wiley-Interscience, New York, NY, pp. 16.6.1–16.6.14.
- Guan, C., Cui, T., Rao, V., Liao, W., Banner, J., Lin, C.-L. and Comb, D. (1996) Activation of glycosylasparaginase: formation of active N-terminal threonine by intramolecular autoproteolysis. *J. Biol. Chem.*, **19**, 1732–1737.
- Chen, L., Pradhan, S. and Evans, T.C., Jr (2001) Herbicide resistance from a divided EPSPS protein: the split *Synechocystis* DnaE intein as an *in vivo* affinity domain. *Gene*, **263**, 39–48.
- Weir, A.F. (1993) Deoxyribonuclease. In Burrell, M.M. (ed.), *Enzymes of Molecular Biology*. Humana Press, NJ, pp. 7–16.
- Parkinson, M.J., Pohler, J.R.G. and Lilley, D.M.J. (1999) Catalytic and binding mutants of the junction-resolving enzyme endonuclease I of bacteriophage T7: role of acidic residues. *Nucleic Acids Res.*, **27**, 682–689.
- Kupust, R.B. and Waugh, D.S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it fused. *Protein Sci.*, **8**, 1668–1674.
- Lu, M., Guo, Q., Studier, F.W. and Kallenbach, N.R. (1991) Resolution of branched DNA substrates by T7 endonuclease I and its inhibition. *J. Biol. Chem.*, **266**, 2532–2536.