

Prokaryotic 5'-3' exonucleases share a common core structure with gamma-delta resolvase

Peter J. Artymiuk, Thomas A. Ceska^{2,+}, Dietrich Suck² and Jon R. Sayers^{1,*}

Department of Molecular Biology and Biotechnology and ¹Department of Molecular and Genetic Medicine, Krebs Institute, The University of Sheffield, Sheffield S10 2JF, UK and ²European Molecular Biology Laboratory, Structural Biology Programme, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received August 7, 1997; Revised and Accepted September 22, 1997

ABSTRACT

The three dimensional crystal structure of T5 5'-3' exonuclease was compared with that of two other members of the 5'-3' exonuclease family: T4 ribonuclease H and the N-terminal domain of *Thermus aquaticus* DNA polymerase I. Though these structures were largely similar, some regions of these enzymes show evidence of significant molecular flexibility. Previous sequence analysis had suggested the existence of a helix-hairpin-helix motif in T5 exonuclease, but a distinct, though related structure is actually found to occur. The entire T5 exonuclease structure was then compared with all the structures in the complete Protein Data Bank and an unexpected similarity with gamma-delta ($\gamma\delta$) resolvase was observed. 5'-3' exonucleases and $\gamma\delta$ resolvase are enzymes involved in carrying out quite different manipulations on nucleic acids. They appear to be unrelated at the primary sequence level, yet the fold of the entire catalytic domain of $\gamma\delta$ resolvase is contained within that of the 5'-3' exonuclease. Different large-scale helical structures are used by both families to form DNA binding sites.

INTRODUCTION

$\gamma\delta$ resolvases and T5 5'-3' exonuclease are enzymes involved in the processing of nucleic acids. No significant amino acid similarity has been reported for these enzymes and they fulfil very different primary roles, in site-specific recombination and replication, respectively. Both types of enzyme bind to DNA and cut internucleotidic phosphodiester bonds in transesterification reactions (1,2). However, while T5 exonuclease simply hydrolyses the nucleic acid substrate, $\gamma\delta$ resolvase first forms a covalent protein-DNA adduct via a phosphoserine intermediate which is then attacked by an incoming 3'-hydroxyl group leading finally to strand exchange.

Resolvases bind a cointegrate, a compound plasmid carrying directly repeated copies of a transposon, and catalyse both cleavage

and subsequent religation events. The process liberates the original transposon-carrying plasmid and the target plasmid complete with inserted daughter transposon (3). The 5'-3' exonucleases such as those encoded by bacteriophages T5 and T7, and the small fragment of eubacterial DNA polymerase I homologues are capable of structure-specific endonucleolytic cleavage (4,5). They are also able to release nucleotides exonucleolytically from DNA substrates containing free 5'-ends. Bifurcations, or flap structures, with displaced 5'-single stranded tails are cleaved close to the flap junction. Thus, the resolvases act in a sequence-specific manner requiring correctly oriented copies of their cognate *res* sites in order to function whereas T5 exonuclease requires only a free 5'-end and a divalent cofactor for catalysis (2,3,6).

The program PROTEP (7) was used to compare the structure of T5 exonuclease [PDB code 1EXN, (4)] with the other structures in the Protein Data Bank (8,9). The search revealed the expected similarities between T5 5'-3' exonuclease and the two other related 5'-3' exonucleases whose structures are known: *Taq* polymerase 5'-3' exonuclease domain [PDB code 1TAQ, (10)] and T4 RNase H [PDB code 1TFR, (11)] (Fig. 1). This search also detected a less extensive but very significant similarity between the catalytic domains of T5 exonuclease and that of $\gamma\delta$ resolvase [2RSL and 1GDT, (12,13)].

METHODS

Primary sequence alignments were performed using the GAP program (14). Three dimensional structure comparisons were performed using the PROTEP program (7) which uses a maximal common subgraph algorithm (15) to compare protein structures represented as linear helices and strands in three dimensions. The structure of T5-exonuclease was compared with all unique protein structures deposited in the May 1997 release of the Protein Data Bank (a subset of 4300 structures) (8,9). Regions of helix and strand in the compared structures were assigned using the algorithm of Kabsch and Sander (16). The position and direction of each helix or strand was represented by a vector in three dimensions (the nodes of the graph) and the distances and torsional angles between them calculated (the edges of the graph) and stored in a database as a labelled graph.

*To whom correspondence should be addressed. Tel: +44 114 271 2327; Fax: +44 114 271 2882; Email. j.r.sayers@sheffield.ac.uk

⁺Present address: Celltech Therapeutics, 216 Bath Road, Slough, Berks SL1 4EN, UK

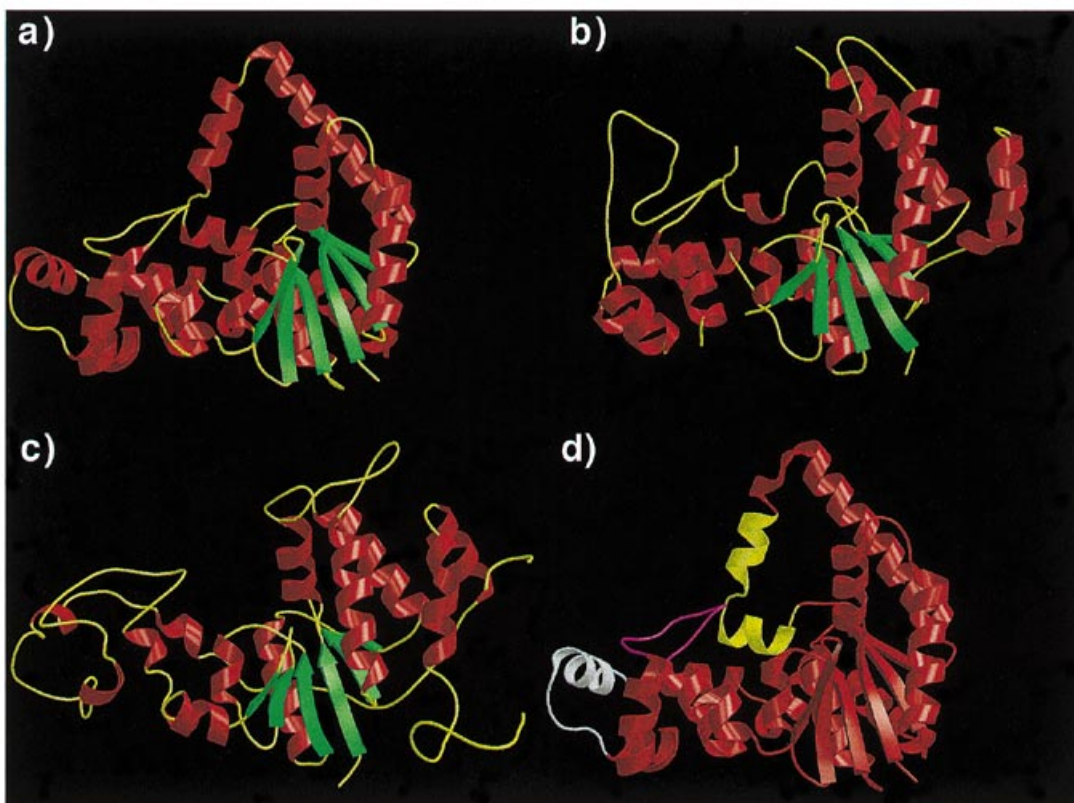


Figure 1. The family of 5'-3' exonucleases. Schematics (24) of (a) T5 exonuclease (4), (b) the ribonuclease H from bacteriophage T4 (11) and (c) the exonuclease domain of *Taq* polymerase (10). The similar catalytic domains can be seen to the right hand side around the β -sheets (green). (d) The T5 exonuclease structure colour coded according to regions which show flexibility. These regions have slightly different conformations in the asymmetric unit of the native crystal form of the nuclease. The biggest differences from the largest region of the protein are observed in the helical arch (residues 71–90, coloured yellow) with smaller differences seen in the finger contacting the arch (residues 200–212, coloured pink) and in the region under the finger (residues 233–250, coloured grey). The flexibility in the molecule observed in these regions is discussed in the text.

RESULTS

Comparison of 5'-3' exonuclease structures

The three-dimensional structures of the three 5'-3' exonucleases are shown in Figure 1. They show good agreement in their N-terminal catalytic domains (143 core C α atoms from T4 RNase H superimpose on T5 exonuclease with an RMSD of 1.64 Å; 110 core C α atoms from *Taq* superimpose with an RMSD of 1.55 Å). The only significant differences between these structures are in disordered regions not clearly determined in the X-ray structures and in a small region located toward the C-terminus. This region is poorly conserved at the sequence level (17). Sequence comparisons showed that the T4 and *Taq* enzymes were 20% identical and 48% similar, the T5 and *Taq* sequences showed 26% identity and 54% similarity and comparison of the T5 and T4 enzyme sequences showed their degree of similarity to be 45% (23% identical).

Molecular flexibility

The T5 exonuclease structure (4) contained two molecules in the asymmetric unit with near crystallographic symmetry. This

deviation from perfect symmetry was first manifested by the observation that 2-fold averaged maps were substantially degraded when a single non-crystallographic symmetry operator was used in the averaging process. By allowing independent operators in the averaging process a substantial improvement in the maps was observed. The regions that appear to be somewhat independent are in four parts of the molecule, and include the following residues. In region 1: 21–69; 92–198; 214–231; 253–288; region 2: 71–90; region 3: 200–212; and region 4: 233–250. This is shown in Figure 1d as a colour coded ribbon diagram for the four regions. The main part of the molecule consisting of the β -sheet and the surrounding helices contains the region which is structurally similar to $\gamma\delta$ resolvase. The three other regions of T5 exonuclease which do show conformational variability are the helical arch, which shows the greatest deviation, the 'finger' which touches the arch (residues 200–212) and the region at the base of the finger consisting of residues 233–250. These regions seem to move in tandem, and are dependent on the position of the helical arch. The arch shows the greatest deviation, followed by the finger region, and the 233–250 region shows the smallest deviation.

Table 1. Most significant hits found in the PROTEP search for T5 exonuclease, in decreasing order of significance

Protein name	Brookhaven ID codes	Maximum clique size	RMSD (Å)	Comments
T5 5'-3' exonuclease	1EXN	25	0.0	Self
T4 RNaseH	1TFR	18	3.1	Exceedingly similar
<i>Taq</i> DNA polymerase	1TAQ,1TAU	14	4.5	Very similar
$\gamma\delta$ resolvase	2RSL, 1GDT	9	3.3	Entire fold of catalytic domain of GDR present
Biotin carboxylase	1BNC	9	5.3	Poor helix overlap
Aconitase	1ACO ^a	8	3.2	Part of N terminal domain of aconitase
<i>HhaI</i> methyltransferase	5MHT	8	3.3	Good overlap, DNA binding protein
Aldehyde dehydrogenase	1AD3	8	3.8	Increasingly poor overlaps
Glucose-fructose oxidoreductase	1OFG	8	3.8	
Glycogen phosphorylase	1ABB ^a	8	4.0	
Malate dehydrogenase	1BMD	8	4.1	
UDP-galactose 4-epimerase	1UDP	8	4.3	
Glycine <i>N</i> -methyltransferase	1XVA	8	4.5	

This order is based primarily on clique size, and secondarily on RMS superposition. The clique size represents the largest number of sequentially ordered secondary structure elements that can be superposed between the search protein and the hit protein by the algorithm after clustering of the initial hits (7). The RMSD is the root mean square difference between pseudo atoms representing the helix and strand vectors used in the search (not C α s).

^aSimilarity with a representative member of a group of essentially identical sequences.

Absence of predicted helix-hairpin-helix domain

Homologous regions of T5 exonuclease and *Taq* Pol located near the C-termini of the nucleases were predicted to adopt a helix-hairpin-helix structure by Ponting and co-workers (18). However, as can be seen in Figure 2, this does not appear to be the case in the crystal structure reported for T5 5'-3' exonuclease. This structure does show some similarity with the HhH motif except that in place of a hairpin a small loop (residues 200-211) has been incorporated. The end of the loop contains one of the conserved aspartate residues (D204) known to be involved in cofactor binding. This region overlaps completely with the finger region referred to above. No clearly analogous structure could be identified in the *Taq* Pol structure.

Similarities between 5'-3' exonucleases and $\gamma\delta$ resolvase

The results of the structural similarity search are shown in Table 1. Though elements of similarity between the structure of T5 exonuclease and other (non-homologous) proteins were noted, it was found to most closely resemble the catalytic domain of $\gamma\delta$ resolvase. The arrangement of β -strands in this domain is related to that of the Rossmann fold, and at a somewhat lower level of significance the search shows that this region of the structure also resembles certain dehydrogenases and the DNA-binding *HhaI* methyltransferase. It is therefore possible that this resemblance may reflect a shared convergent elaboration of a common topology and may not be indicative of a divergent evolutionary relationship. However, the $\gamma\delta$ resolvase similarity involves the entire catalytic domain of $\gamma\delta$ resolvase, whilst the other non-exonuclease hits are not only at a lower level of significance as measured by number of secondary structure elements superposed but also involve only parts of those enzymes' respective domains.

The superposition of the conserved core regions of T5 exonuclease and $\gamma\delta$ resolvase are shown (Fig. 3a) and the entire structures of both molecules are shown schematically

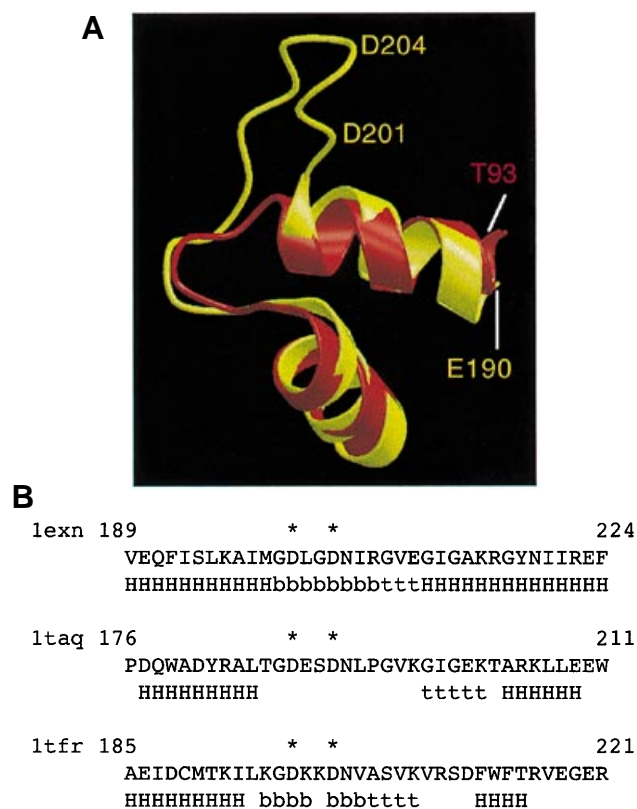


Figure 2. (A) Superposition of the helix-hairpin-helix motif from polymerase β (PDB code 1BPE, residues T93-E117 shown in red) with amino acids E190-E223 of T5 exonuclease (yellow). Positions of two conserved aspartates (D201 and D204) observed to bind metal ions in the crystal structure of T5 exonuclease (4) are shown. (B) The observed secondary structure seen in the three 5'-3' exonucleases is shown below each sequence. H, helical; b, β -strand; t, hairpin turn. The * denotes two of the conserved aspartate residues observed to bind divalent cations.

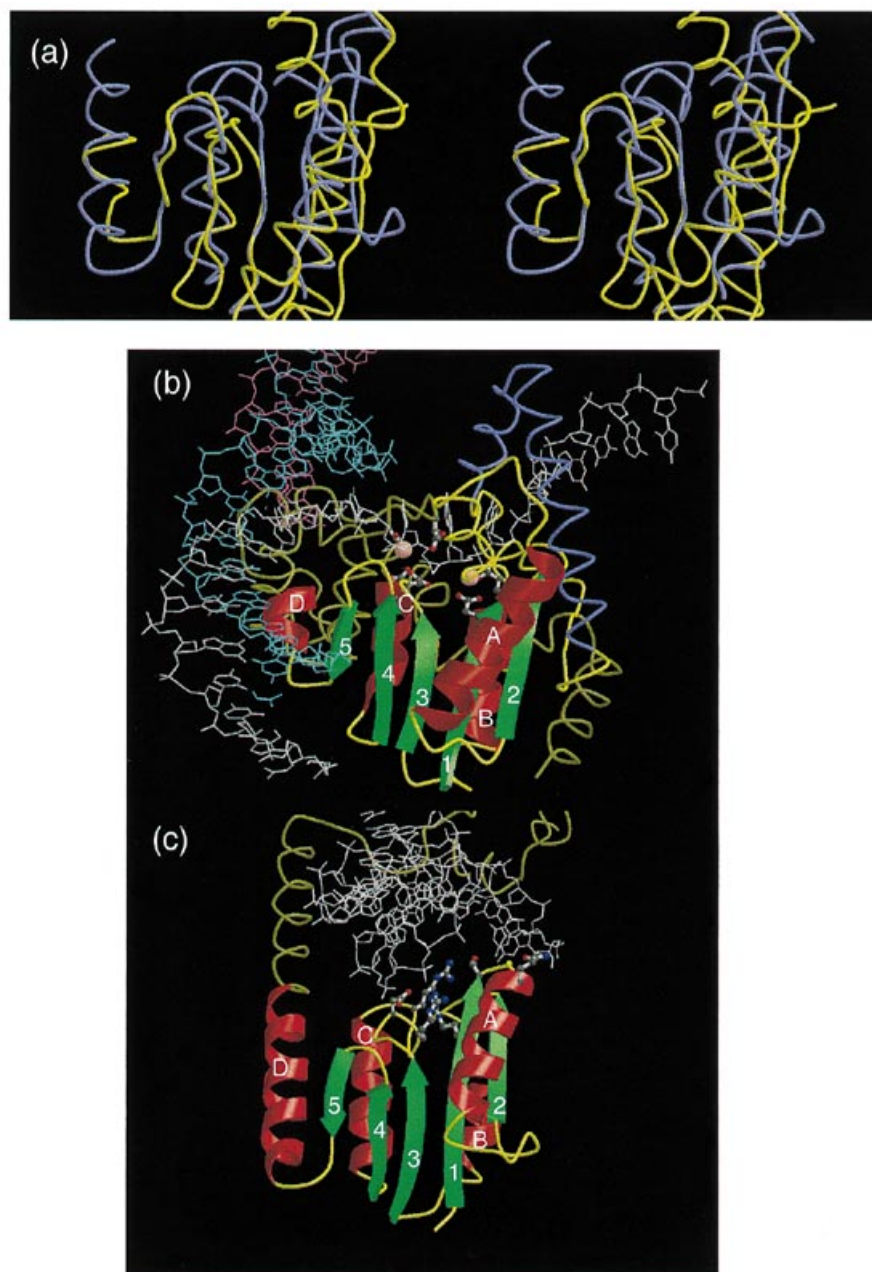


Figure 3. Comparisons of T5 exonuclease and $\gamma\delta$ resolvase. (a) Stereodigram showing the superposition of the equivalent C α atoms from the core domains of T5 exonuclease (yellow) and $\gamma\delta$ resolvase (blue). Equivalent elements (24,25,26) in (b) T5 exonuclease and (c) $\gamma\delta$ resolvase. The equivalent helices and strands in our study are represented as red coiled ribbons and green sequentially numbered arrows, respectively, the non-equivalent parts of the structure are shown as a smoothed C α trace, the arch region is indicated in blue. The active site residues, shown in (b) as ball-and-stick models, are (clockwise from top left) Asp201, Asp204, Asp131 and Asp26, Asp153 and Asp155, the silver spheres represent the two Mn²⁺ ions bound by T5 exonuclease observed in a difference electron density map. Active site residues in (c) are (clockwise from leftmost residue) Asp67, Arg71, Ser10, Gln14, Arg8 and Arg68. Bound DNA is shown in the $\gamma\delta$ resolvase structure and a flap DNA molecule is modelled into the T5 exonuclease structure (4).

(Fig. 3b and c) with the equivalent regions shown in the same orientation. The topological relationship between the two folds is clearly demonstrated in the topological diagram in Figure 4. As can be seen, the entire fold of the catalytic domain of $\gamma\delta$ resolvase, consisting of five β -strands and four α -helices is present in the T5 exonuclease catalytic domain, the latter being slightly augmented by one β -strand and an α -helix. The agreement in

3-dimensional space is good: 50 C α atoms from the core regions of $\gamma\delta$ resolvase and T5 exonuclease superimpose with an RMSD of 1.52 Å.

The position of the active site in both enzymes is located at the C terminal ends of the first two strands of the β -sheet of the catalytic domain (Fig 3b and c). In the exonucleases, highly conserved aspartate residues from strands 1, 2 and 3 form the two

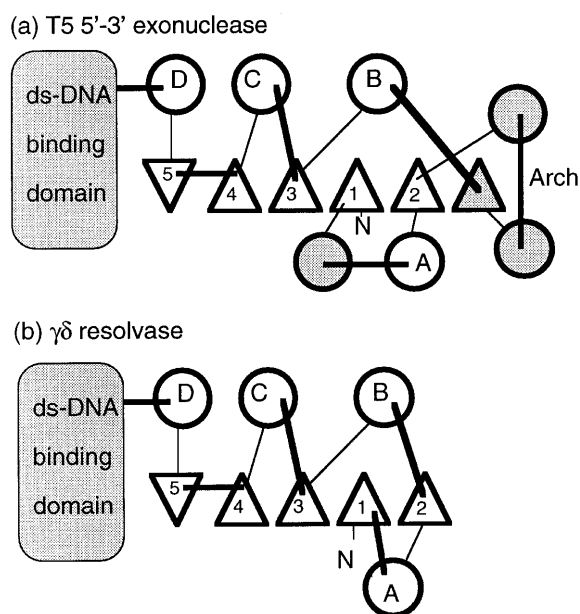


Figure 4. Topological diagrams (27) of (a) T5 exonuclease and (b) $\gamma\delta$ resolvase. Circles represent α -helices and triangles represent β -strands (apex down indicates the strand is running down into the plane of the paper, i.e., viewed from the N-terminus). Open circles and triangles indicate those secondary structure elements that were found by the PROTEP search to superimpose in 3-dimensions, and shaded ones those that do not. The equivalent strands in the β -sheets are numbered, and the equivalent helices lettered, according to their order in the sequence.

metal binding sites (4). These include Asp26, Glu128, Asp131 and Asp153 (metal site I) and Asp153, Asp155, Asp201 and Asp204 (metal site II). The active site of $\gamma\delta$ resolvase, is less well defined but is known to include residue Ser10 which makes a covalent bond with the DNA substrate (19), and there is good evidence for involvement of Arg8, Gln14, Asp67, Arg68 and Arg71 (20,21). None of these residues overlap with those conserved in the 5'-3' exonucleases, but all are concentrated in the same general region of the superimposed structures.

DISCUSSION

The three representatives of the 5'-3' exonuclease family which have been structurally characterised share a significant degree of primary sequence similarity (20–26% identity) including all 10 absolutely conserved residues postulated as defining a prokaryotic 5'-3' exonuclease motif sequence (17). There are differences between these three structures, most notably at the C-terminus, where the sequence similarity is poorest. In the region described as a helical arch in the T5 exonuclease (4) the equivalent regions in the other two nucleases are disordered. One unresolved question is the reason for this disorder in two of the three members of this family and whether it is functionally significant.

In an analysis using pre-release structural data relating to *Taq* Pol, Ponting and co-workers (18) proposed that it and T5 exonuclease should contain the HhH motif described originally by Thayer *et al.* (22). This motif is implicated in DNA binding. However, the HhH domain was not present in the three crystal structures examined. T5 exonuclease does contain a related structure, similar to the HhH domain, but with a short loop of

10 amino acids replacing the proposed hairpin. As can be seen from Figure 2 the helices of the HhH domain do superimpose very well on the helix-loop-helix observed in T5 exonuclease.

The differences between *Taq* polymerase, T4 RNase H and T5 exonuclease, and the high temperature factors observed in the C-terminal region of the 5'-3' exonuclease domain of *Taq* polymerase may be viewed as corroborating evidence for innate structural flexibility in the 5'-3' exonuclease family. Alternatively, this may reflect genuine structural differences in regions of low sequence similarity within the 5'-3' exonuclease family.

The observation that the catalytic domains of 5'-3' exonucleases possess a high degree of structural similarity to that of $\gamma\delta$ resolvase was totally unexpected. The structurally related unit involves the catalytic, rather than the DNA-binding, domain of $\gamma\delta$ resolvase. The resolvases act in a sequence specific manner while the 5'-3' exonucleases show no sequence specificity per se, thus, their major biochemical similarity lies in the ability to bind DNA and in the transesterification reaction.

The 5'-3' exonucleases recognise certain DNA structures and require only a free 5'-end. The resolvases demonstrate a high degree of sequence specificity for their DNA-cleavage and rejoining reactions. This sequence specificity resides in domain three, totally separate from the catalytic domain. The specificity domain comprises a three helix bundle arranged in a helix-turn-helix motif (13). This region of $\gamma\delta$ resolvase does not show any significant structural similarity with the 5'-3' exonucleases. The structure of the $\gamma\delta$ resolvase recognition domain in complex with a 34 bp oligonucleotide shows that DNA contacts the catalytic domain of $\gamma\delta$ resolvase (13). The structural comparisons reported here raise the intriguing possibility that the catalytic domains of 5'-3' exonucleases and $\gamma\delta$ resolvase may employ related DNA-binding mechanisms. No structural information has yet been reported for a 5'-3' exonuclease-DNA complex but a threading model was proposed by the Suck laboratory (4) (Fig. 3b).

In an analysis of the structure of the $\gamma\delta$ resolvase at 2.3 Å, Rice and Steitz observed that there were three slightly different conformations present in the crystal form they were studying (23). The major differences they observed were (i) in the twist of the β -sheet, (ii) in the conformation of the loop between β -strand 2 and α -helix B and (iii) the angle at which helix A packs against the rest of the protein. They concluded that the conformational flexibility of the molecule is likely to allow $\gamma\delta$ resolvase to form the synaptic complex (i.e., a multi-subunit complex) and to present DNA binding sites of different geometry for the catalysis of recombination.

The role of conformational flexibility in 5'-3' exonuclease function is still uncertain. There are two components to this flexibility. First the helical arch structure observed in the T5 nuclease was found to be disordered in the two other 5'-nuclease structures, and second there was an inherent molecular flexibility observed in the T5 nuclease structure (Fig. 1d). This may allow processing of diverse flap structures produced during strand displacement synthesis. Since the duplex sequence at the point of bifurcation would vary, the local DNA structure would also vary. To allow binding and cleavage irrespective of nucleotide sequence a degree of flexibility in the nuclease molecule may be required.

It is clear that in both nuclease and resolvase structures the catalytic sites are positioned in the structurally conserved core regions. The reaction mechanisms of the exonucleases and the resolvases are quite different. The resolvase mechanism requires

a first step involving cleavage of DNA strands and replacement of the phosphodiester bond by a phosphoserine bond between the 5' phosphate at the cleavage site and the hydroxyl of Ser10 in the resolvase. Subsequently strand exchange is effected by a second transesterification reaction leading to the reformation of the phosphodiester backbone. In contrast to the resolvase, the mechanism of the exonucleases requires the involvement of divalent metal ions (4,10,11) and while the active site residues are positioned in similar positions with respect to the common fold of the catalytic domain, there appears to be no immediately obvious chemical relationship between them. In particular, the initial cleavage reaction carried out by $\gamma\delta$ resolvase does not require a metal cofactor, though magnesium does stimulate the recombination step (19).

However, there are certain chemical similarities in that both enzymes cleave DNA and there is a similar transesterification step, to water or serine, involved in the mechanisms. The structural similarities may therefore indicate the existence of a remote evolutionary ancestor for the core domain of these two classes of enzymes. In this context it is most interesting to observe that the 5'-3' exonucleases exhibit ribonuclease H activity and this may constitute a direct link to their proposed common ancestral protein. One may speculate that this ancestral protein may have had a simple single-stranded nucleolytic function which has been retained, whilst the ability to bind double-stranded DNA has been added on in different ways in the two proteins: in $\gamma\delta$ resolvase by dimerisation to form a large α -helical structure and in the 5'-3' exonucleases by addition of a C-terminal region that binds the duplex part of the DNA. Thus, it seems possible that both molecules diverged from the common catalytic core domain but convergent evolution is responsible for the superficially similar flexible arch-like structures used to bind DNA.

ACKNOWLEDGEMENTS

This work benefited from the use of the SEQNET facility at Daresbury Laboratory, UK. We thank the MRC for support to PJA. The Krebs Institute is a BBSRC Biomolecular Science Centre (PJA and JRS).

REFERENCES

- 1 Reed,R.R. (1981) *Cell* **25**, 713–719.
- 2 Paul,A.V. and Lehman,I.R. (1966) *J. Biol. Chem.* **241**, 3441–3451.
- 3 Sherratt,D. (1989) in Berg,D.E. and Howe,M. (Eds) *Mobile DNA*. American Society for Microbiology, Washington, DC, USA, pp. 163–184.
- 4 Ceska,T.A., Sayers,J.R., Stier,G. and Suck,D. (1996) *Nature* **382**, 90–93.
- 5 Lyamichev,V., Brow,M.A.D. and Dahlberg,J.E. (1993) *Science* **260**, 778–783.
- 6 Sayers,J.R. and Eckstein,F. (1990) *J. Biol. Chem.* **265**, 18311–18317.
- 7 Grindley,H.M., Artymiuk,P.J., Rice,D.W. and Willett,P. (1993) *J. Mol. Biol.* **229**, 707–721.
- 8 Abola,E.E., Bernstein,F.C., Bryant,S.H., Koetzle,T.F. and Weng,J. (1987) In Allen,F.H., Bergeroff,G. and Sievers,R. (Eds) *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*. Data Commission of the International Union of Crystallography, Bonn, Cambridge, Chester, pp. 107–132.
- 9 Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.* **112**, 535–542.
- 10 Kim,Y., Eom,S.H., Wang,J., Lee,D.S., Suh,S.W. and Steitz,T.A. (1995) *Nature* **376**, 612–616.
- 11 Mueser,T.C., Nossal,N.G. and Hyde,C.C. (1996) *Cell* **85**, 1101–1112.
- 12 Sanderson,M.R., Freemont,P.S., Rice,P.A., Goldman,A., Hatfull,G.F., Grindley,N.D.F. and Steitz,T.A. (1990) *Cell* **63**, 1323–1329.
- 13 Yang,W. and Steitz,T.A. (1995) *Cell* **82**, 193–207.
- 14 Devereux,J., Haeberli,P. and Smithies,O. (1985) *Nucleic Acids Res.* **12**, 387–395.
- 15 Bron,C. and Kerbosch,J. (1973) *Comm. Assoc. Comput. Machinery*, **16**, 575–577.
- 16 Kabsch,W. and Sander,C. (1983) *Biopolymers* **22**, 2577–2637.
- 17 Gutman,P.D. and Minton,K.W. (1993) *Nucleic Acids Res.* **21** 4406–4407.
- 18 Doherty,A.J., Serpell,L.C. and Ponting,C.P. (1996) *Nucleic Acids Res.* **24**, 2488–2497.
- 19 Boocock,M.R., Zhu,X. and Grindley,N.D.F. (1995) *EMBO J.* **14**, 5129–5140.
- 20 Hatfull,G.F. and Grindley,N.D.F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5429–5433.
- 21 Hughes,R.E., Hatfull,G.F., Rice,P., Steitz,T.A. and Grindley,N.D.F. (1990) *Cell* **63**, 1331–1338.
- 22 Thayer,M.M., Ahern,H., Xing,D.X., Cunningham,R.P., Tainer,J.A. (1995) *EMBO J.* **14**, 4108–4120.
- 23 Rice,P.A. and Steitz,T.A. (1994) *Structure* **2**, 371–384.
- 24 Kraulis,P.J. (1991) *J. Appl. Cryst.* **24**, 946–950.
- 25 Bacon,D. and Anderson,W.F. (1988) *J. Mol. Graphics* **6**, 219–220.
- 26 Merritt,E.A. and Murphy,M.E.P. (1994) *Acta Cryst.* **D50**, 869–873.
- 27 Flores,T.P., Moss,D.S. and Thornton,J.M. (1994) *Protein Engng* **7**, 31–37.