

X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination

Norbert Sträter¹, David J. Sherratt² and Sean D. Colloms²

Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany and ²Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

¹Corresponding author
e-mail: strater@chemie.fu-berlin.de

The structure of aminopeptidase A (PepA), which functions as a DNA-binding protein in Xer site-specific recombination and in transcriptional control of the *carAB* operon in *Escherichia coli*, has been determined at 2.5 Å resolution. In Xer recombination at *cer*, PepA and the arginine repressor (ArgR) serve as accessory proteins, ensuring that recombination is exclusively intramolecular. In contrast, PepA homologues from other species have no known DNA-binding activity and are not implicated in transcriptional regulation or control of site-specific recombination. PepA comprises two domains, which have similar folds to the two domains of bovine lens leucine aminopeptidase (LAP). However, the N-terminal domain of PepA, which probably plays a significant role in DNA binding, is rotated by 19° compared with its position in LAP. PepA is a homohexamer of 32 symmetry. A groove that runs from one trimer face across the 2-fold molecular axis to the other trimer face is proposed to be the DNA-binding site. Molecular modelling supports a structure of the Xer complex in which PepA, ArgR and a second PepA molecule are sandwiched along their 3-fold molecular axes, and the accessory sequences of the two recombination sites wrap around the accessory proteins as a right-handed superhelix such that three negative supercoils are trapped.

Keywords: DNA-binding proteins/DNA looping/protein crystallography/protein–protein complexes/site-specific recombination

Introduction

Plasmid multimers formed by homologous recombination reduce the total number of plasmid molecules in the cell and increase the chance of forming plasmid-free segregants when the plasmids are distributed between daughter cells at cell division. The *Escherichia coli* Xer site-specific recombination system acts at sites found in multicopy plasmids such as ColE1 *cer* and pSC101 *psi* to monomerize plasmid multimers. This maximizes the number of independently segregating plasmid units and helps to ensure stable plasmid inheritance (Summers and Sherratt, 1984). Four host-encoded proteins (XerC, XerD, ArgR and PepA) are required for recombination at *cer* (Stirling *et al.*, 1988,

1989; Colloms *et al.*, 1990; Blakely *et al.*, 1993). The recombination reaction is catalysed by XerC and XerD, two members of the integrase class of site-specific recombinases (tyrosine recombinases). The arginine repressor (ArgR) and aminopeptidase A (PepA) are accessory proteins which are not directly involved in the strand exchange reaction but are absolutely required for recombination at *cer* *in vivo* and *in vitro* (Stirling *et al.*, 1988, 1989; Colloms *et al.*, 1996). The *cer* site consists of a 30 bp core sequence to which XerC and XerD bind, and ~180 bp of accessory sequences adjacent to the core. ArgR and PepA bind to the accessory sequences of *cer* and form a complex in which two recombination sites are inter-wrapped in a right-handed fashion (Alén *et al.*, 1997). Xer recombination at *cer* is exclusively intramolecular, resolving but not creating plasmid multimers. ArgR and PepA appear to be responsible for ensuring this resolution selectivity during recombination at *cer*.

Xer recombination also acts at *dif*, in the replication terminus region of the *E. coli* chromosome, and helps to ensure faithful segregation of newly replicated chromosomes to daughter cells at cell division (Blakely *et al.*, 1993). Xer recombination at *dif* requires only a 28 bp DNA site to which XerC and XerD bind. Accessory sequences, and the accessory proteins ArgR and PepA are not required for Xer recombination at *dif*. In contrast to recombination at *cer*, recombination at a plasmid-borne copy of *dif* is both inter- and intramolecular.

In addition to its role in Xer recombination, PepA is involved in the pyrimidine-specific transcriptional regulation of the *carAB* operon, which encodes carbamoyl-phosphate synthetase (Charlier *et al.*, 1995). PepA binds to DNA upstream of the pyrimidine regulated P1 promoter of the *E. coli carAB* operon. DNase I footprinting experiments demonstrated that PepA protects two 25–30 bp stretches, separated by 65 nucleotides in the *carAB* control region. DNA binding to these two sites is cooperative; the individual sites exhibit only very low binding affinity to PepA. Besides PepA, the integration host factor (IHF), ArgR and UMP-kinase are involved in regulation of the *E. coli carAB* operon (Charlier *et al.*, 1995; Kholti *et al.*, 1998). The pattern of DNase I protection and hypersensitivity of the *cer* sequence in the presence of PepA and ArgR shows similarities to the pattern of protection seen in the *carAB* operon. Both footprints contain an ~60 bp stretch characterized by a number of DNase I hypersensitive sites at ~10 bp intervals, suggestive of wrapping or looping of the DNA, flanked by regions of DNase I protection. However, the pattern seen at the *cer* site is somewhat more complex and extends over the entire 180 bp of accessory sequences adjacent to the *cer* recombination core site (Figure 1; Alén *et al.*, 1997). This altered pattern of DNase I cleavage was apparent in the presence of PepA alone, but was strongly enhanced when

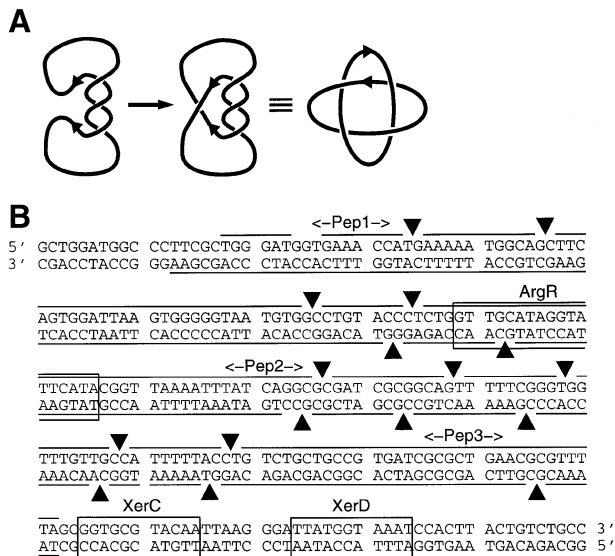


Fig. 1. (A) Proposed pathway of Xer recombination (Colloms *et al.*, 1997). In the synaptic complex the accessory sequences of the recombination sites are plectonemically interwrapped, in a right-handed sense, such that three negative interdomainal nodes are trapped. Strand exchange introduces an additional negative node. An antiparallel 4-noded catenane is formed, in agreement with experimental results. (B) DNase I protection of the *cer* site in the presence of PepA and ArgR (Alén *et al.*, 1997). Horizontal lines indicate protected regions. Filled triangles mark enhancements of DNase I cleavage. The binding sites of ArgR, XerC and XerD are boxed. The presumed sites for the interaction of PepA with the *cer* sequence in the models presented in this study are labelled Pep1, Pep2 and Pep3.

PepA and ArgR were present together. In the presence of ArgR alone only the 18 bp ARG box was protected.

Analysis of the product of Xer recombination at pSC101 *psi* demonstrated that the product is a right-handed, antiparallel, 4-noded catenane, and the product of recombination at ColE1 *cer* is analogous but contains a Holliday junction (Colloms *et al.*, 1997; Figure 1). These results imply that the productive synaptic complex and the strand exchange mechanism have fixed geometries. The product topology can be accounted for by a model in which the accessory sequences of two participating sites are wrapped around each other so as to trap three negative plectonemic supercoils, with an additional negative topological node being introduced by the recombination reaction. In this model, assembly of the productive synapse is required before XerC and XerD can catalyse strand exchange. The model accounts for both the product topology and the resolution selectivity of the reaction. Whereas the productive synapse can readily be formed between two sites in direct repeat on a supercoiled DNA molecule, formation of the productive synapse between sites on separate circles will not be favourable.

Based on the current knowledge of Xer recombination, Hodgman *et al.* (1998) proposed an alternative model for the synaptic nucleoprotein complex in which the DNA is wrapped around two PepA and three ArgR molecules that are sandwiched alternately along their 3-fold molecular axes.

XerC and XerD belong to the λ integrase family of site-specific recombinases, share 37% amino acid identity with each other and form a heterodimer when bound to a

recombination core site. Two such heterodimers are involved in strand exchange in the Xer synaptic complex. The X-ray structure of XerD has been determined (Subramanya *et al.*, 1997), as has the structure of the related Cre recombinase bound to a Holliday junction recombination intermediate (Gopaul *et al.*, 1998).

ArgR is an arginine-dependent DNA-binding protein that acts as a transcriptional repressor of the arginine regulon. ArgR usually binds cooperatively to two 18 bp ARG boxes which are separated by 3 bp. However, ArgR binds to a single ARG box within *cer* (Figure 1). The X-ray structure of the hexameric C-terminal oligomerization domain shows that ArgR has 32 molecular symmetry (three 2-fold axes are perpendicular to one 3-fold axis) (van Duyn *et al.*, 1996). The X-ray structure of the entire ArgR protein from *Bacillus stearothermophilus* has also recently been solved (Ni *et al.*, 1999).

The biochemical properties (Vogt, 1970) and primary sequence (Stirling *et al.*, 1989) of PepA shows that it belongs to the family of leucine aminopeptidases (LAP), which are widely distributed in mammals, plants and bacteria (Kim and Lipscomb, 1994; Sträter and Lipscomb, 1998). The hexameric aminopeptidase consists of six identical 55 kDa subunits (503 residues) and contains two metal ions in the active site, which are required for aminopeptidase activity. Leucine aminopeptidase from bovine lens, which shows 31% identity to PepA, has been well characterized by kinetic studies and X-ray crystallography (Burley *et al.*, 1990; Kim and Lipscomb, 1994; Sträter and Lipscomb, 1995a,b). However, the aminopeptidase activity of PepA is not required for its function in Xer recombination (McCulloch *et al.*, 1994) or for its role in pyrimidine-mediated repression of *carAB* transcription (Charlier *et al.*, 1995). The N-terminal domain, which is less conserved between LAP and PepA, is probably involved in the DNA-binding function of PepA. LAP has no known DNA-binding function.

Here, we report the X-ray structure of *E.coli* PepA. The structure of the hexamer shows the presence of three grooves, which are presumed to be the DNA-binding sites. Based on the present structural and biochemical data we propose a model for the *cer* synaptic complex in which the DNA is wrapped around a sandwich of ArgR and two PepA molecules. This model may also shed light on how DNA wrapping, looping and bending contribute to action at a distance in other systems such as in the control of transcription initiation in eukaryotes. Details of the aminopeptidase active site, which is perfectly conserved between LAP and PepA and superimposes closely, will be described elsewhere.

Results and discussion

Monomer structure

Crystals of PepA used in this structure analysis contain two complete hexamers in the asymmetric unit. Although minor differences are apparent in loop regions, all 12 independent subunits have a similar structure. Details of the structure determination are presented in Table I and in Materials and methods. All residues have been included in structure refinement. However, residues 146–152 were found to be disordered in all subunits. Nevertheless, since weak electron density enabled us to determine the location

Table I. Details of data collection and refinement

	Wild type	Mutant ^a
Data collection		
Space group	P3 ₂	R32
Cell parameters (Å)	$a = 178.0, c = 244.4$	$a = 197.4, c = 133.2$
Temperature (K)	100	100
Wavelength (Å)	1.23	0.98
Mosaicity (°)	0.13	0.53
Reflections collected	1 193 533	139 347
Unique reflections	298 951	76 596
Resolution (Å) ^b	30–2.5 (2.59–2.50)	30–2.8 (2.9–2.8)
R_{sym} (%) ^b	9.8 (30.7)	11.7 (35.2)
Completeness (%)	99.9	98.2
Refinement (P3 ₂)		
Reflections, cutoff	298 907, no cutoff	
Resolution (Å)	500–2.5	
R, R_{free} (%)	18.3, 21.0	
R.m.s.d. bonds (Å)	0.005	
R.m.s.d. angles (°)	1.23	
Protein atoms, average B (Å ²)	46 152, 26.8	
Zinc ions, average B (Å ²)	24, 14.6	
Carbonate ions, average B (Å ²)	48, 14.5	
Chloride ions, average B (Å ²)	4, 11.5	
Water molecules, average B (Å ²)	3733, 28.7	
Ramachandran (%)	allowed: 98.4, generous: 1.0, disallowed: 0.6	

^aSee Materials and methods.

^bValues in parentheses refer to the highest resolution shell.

of this loop, these residues were included in the refinement in order to facilitate model interpretation and computational studies.

Both domains of PepA have a mixed α/β structure. A long α -helix links the smaller N-terminal domain (residues 1–166) to the larger C-terminal domain (residues 193–503) (Figure 2). This helix has contacts with both domains. The core of the C-terminal domain has a triple-layered structure consisting of a central eight-stranded β -sheet sandwiched between five α -helices on each side. The aminopeptidase active site is located entirely within the C-terminal domain.

A six-stranded β -sheet shielded by two α -helices on one side forms the core of the N-terminal domain. On the other side of this β -sheet, strands 2, 3 and 4 are partially shielded by helix 1 and loops, whereas strands 1, 5 and 6 are in part exposed to solvent. A long loop which comprises the disordered region of residues 146–152 is present between helix 3 and strand 6.

The temperature factors (B -values) of the N-terminal domain are significantly higher than for the C-terminal domain (Figure 3). Whereas the average B -value (over all 12 monomers) of the C-terminal domain is 17.8 Å², this value is 40.0 Å² for the N-terminal domain. This indicates a higher flexibility of the N-terminal domain. Nevertheless, the electron density map indicates no differences in the relative orientation of the two domains in the 12 subunits of the asymmetric unit. The largest differences are seen in the whole region between β -strand 1 and α -helix 2 except for the residues of the β -strands, which superimpose closely. These regions also have the highest B -factors (Figure 3).

Hexamer structure

Like LAP, the PepA hexamer has 32 symmetry and may be described as a dimer of trimers (Figure 2C–E) (Burley

et al., 1990). When viewed along the 3-fold molecular axis, this hexamer has a triangular shape with a triangle edge length of ~135 Å and a thickness of ~80 Å. The catalytic domains are clustered around the 3-fold axis and are involved in interactions between the subunits of each trimer and between the two trimers. In the centre of the hexamer, a large solvent cavity of 15 Å radius and 10 Å height is present which harbours the aminopeptidase active sites. Access to this cavity is provided by three channels which are at the 2-fold molecular axis, and at the interface between two N-terminal and two C-terminal domains (Figure 2E). A gap between the N-terminal domains and the hexamer core at this interface allows for access to the cavity inside the hexamer. The N-terminal domains extend outwards to the corners of the triangle and they mediate interactions between the two trimers in the vicinity of the 2-fold axes (Figure 2E).

Comparison with leucine aminopeptidase

Since LAP has no known DNA-binding function, a comparison of the structures of PepA and LAP might point out regions that are different and which could be important for the DNA-binding role of PepA. As expected from the high sequence identity, the C-terminal domains of PepA and LAP superimpose well (Figure 4A). In this superposition, 260 out of 310 residues (excluding regions which have a different conformation) of the C-terminal domain of PepA have a root mean square (r.m.s.) deviation of 0.73 Å. This value increases only slightly to 0.74 and 0.78 Å when a trimer and the whole hexamer, respectively, are superimposed using these residues. Thus, the overall hexamer structure is well conserved between LAP and PepA, including the presence of the large cavity inside the hexameric molecule. No difference was observed in the orientation of the subunits within the hexamer. However, larger differences were seen in the regions 257–

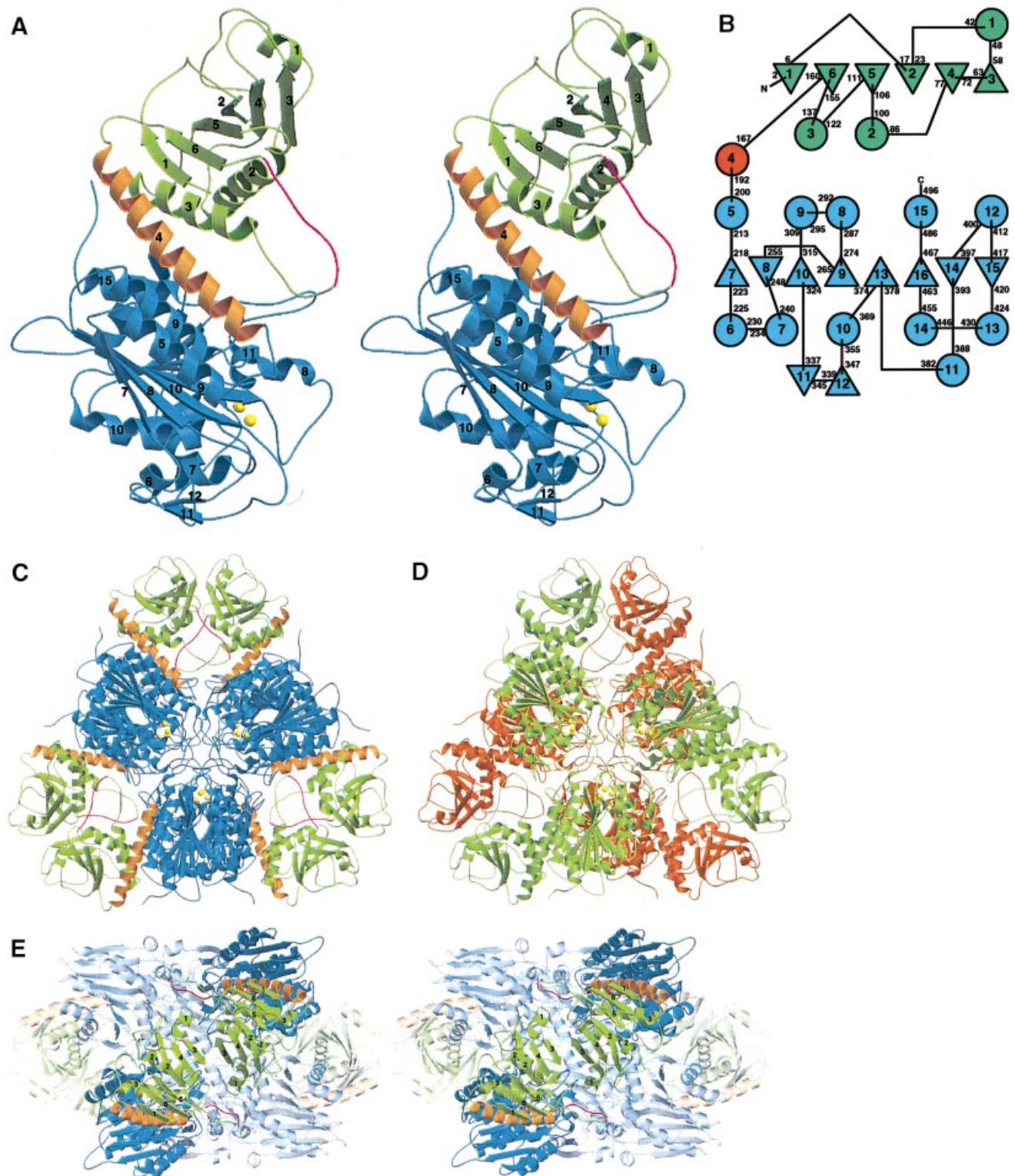


Fig. 2. Structure of PepA. **(A)** Ribbon diagram of the monomer fold (stereo figure). The N-terminal domain is coloured in green, the C-terminal domain in blue, and the domain-linking helix in orange. Also shown are the two zinc ions of the aminopeptidase active site. The loop from residue 146 to 152 (shown in purple) is disordered. **(B)** Topology diagram. The colouring scheme is the same as in (A). Helices are marked as circles and sheets as triangles (Δ or ∇ if the C-terminal end points towards or away from the viewer, respectively). The secondary structure elements are numbered and the first and last residues for each element are listed. **(C)** View of the hexamer along the 3-fold molecular axis. The colouring scheme is the same as in (A). **(D)** View of the upper (green) and lower (orange) trimers along the 3-fold axis. Two-fold axes perpendicular to the 3-fold axis relate the upper and lower trimers of the hexamer, which has 32 symmetry. **(E)** Stereo view of the hexamer along the 2-fold molecular axes. For clarity, monomers which are in the back in this view of the hexamer are shown in transparent colours. Residues at the upper and lower face of the two N-terminal domains are proposed to interact with DNA. A channel providing access to the aminopeptidase active site is between the two N-terminal domains seen in the front of the figure and the hexamer core formed by the C-terminal domains behind the two N-terminal domains.

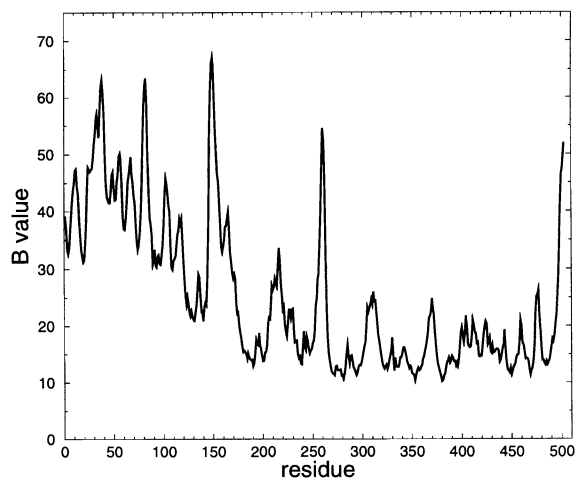


Fig. 3. Plot of the temperature factors versus residue number. Only the main chain atoms were used to calculate the average *B*-factor. The N-terminal domain has significantly higher temperature factors than the C-terminal domain.

262, 402–444, 458–463 and 471–481. In addition, the C-terminus in PepA is extended by five residues and interacts with the N-terminal domain. The side-chains of this C-terminal extension may also be important for DNA binding, as proposed in our model below.

The superposition of LAP and PepA based on the C-terminal domains shows that the N-terminal domains have different orientations relative to the C-terminal domains (Figure 4A). The operation relating the two domains might be described as a screw axis with a rotation angle of 19° and a translation of 2.9 Å. Differences are also seen in residues at the N-terminal side of the domain-connecting helix, which is bent in LAP whereas it is straight in PepA. Since the relative orientation of the two domains is the same in all 12 monomers of the asymmetric unit, it appears quite unlikely that the domain orientation is affected by crystal packing interactions. A comparison of the N-terminal domains alone shows that the fold of this domain is conserved between the two aminopeptidases (Figure 4B–D). An additional N-terminal β -strand is added to the edge of the central β -strand in PepA. The loop (L1 in Figure 4) between strand 2 and helix 1 of PepA is longer in LAP and contains a helix of nine residues, which is not present in PepA. Furthermore, the loop (L2) between strand 4 and helix 2 is shorter in PepA, whereas the loop (L3) between strand 5 and helix 3 is longer. These three loops may interact with DNA in the current model for DNA binding (see below). All of these differences between PepA and LAP result in a smoother and more compactly shaped N-terminal domain in PepA and give rise to a longer surface parallel to the presumed DNA-binding groove (Figure 5).

Interactions of PepA with ArgR and DNA

The molecular surface of PepA shows a large groove running from the lower trimer face across the 2-fold molecular axis to the upper trimer face (Figure 5A). This groove is large enough to accommodate a DNA duplex. Results from pentapeptide insertion mutagenesis (Hayes *et al.*, 1997; S.D.Colloms and D.J.Sherratt, unpublished observations) support the presumed location of the DNA-binding groove by showing that insertions at residues of

strand 1, of the loop between strand 2 and helix 1, of the disordered region of the loop between helix 3 and strand 6, and of strand 6 affect the DNA-binding function of PepA in Xer recombination, while these mutant variants still have peptidase activity. Since the presumed DNA-binding groove crosses a 2-fold symmetry axis of PepA, the DNA will contact equivalent N-terminal residues at the upper and lower trimer faces of PepA as well as equivalent residues of two PepA C-terminal domains, on either side of the 2-fold axis. The electrostatic potential at the protein surface of this groove shows no extended regions of a distinct positive potential throughout the groove as in some other DNA-binding proteins; however, there appear to be more regions of positive potential than negative potential, whereas nearby regions of the N-terminal domain closer to the corners of the hexamer have predominantly negative charges (Figure 5B).

PepA and ArgR both have 32 symmetry. In ArgR the DNA-binding domains are located around the core of the hexamer, which is formed by the C-terminal domains (van Duyne *et al.*, 1996). In PepA the DNA also appears to bind largely to the trimer edges and the N-terminal domains. If a direct interaction between PepA and ArgR exists, it thus seems most probable for steric reasons that the two proteins interact at the C-terminal domains of their trimer faces, as suggested previously by Hodgman *et al.* (1998). The shape and electrostatic potential of the trimer surfaces of PepA and ArgR are consistent with such an interaction. As previously noted by Hodgman *et al.* (1998), the trimer face of ArgR has a predominantly negative potential (Figure 5D). The surface potential of the trimer face formed by the C-terminal domain of PepA has a positive potential in the centre, surrounded by regions of negative potential. Both proteins have almost exclusively hydrophilic residues at their trimer faces, which are rather flat in the region around the 3-fold axis and thus have little characteristic surface complementarity. Therefore, it is difficult to predict the relative orientation of PepA and ArgR with respect to a rotation around a common 3-fold axis.

A model for the Xer complex

DNase I protection assays (Alén *et al.*, 1997) and the product topology of Xer recombination between *cer* sites on plasmid substrates (Colloms *et al.*, 1997) as well as other studies, provided important information on the structure of the Xer complex. The pattern of DNase I protection and hyperactivity of the *cer* sequence in the presence of PepA and ArgR (Figure 1) indicates that binding sites and loops of the *cer* sequence might be present in the following order: PEP1, ARG, PEP2, an ~60 bp loop, PEP3, XERC and XERD. The formation of right-handed, antiparallel, 4-noded catenanes is in agreement with a structure of a synaptic complex in which the accessory sequences are plectonemically interwrapped, in a right-handed sense around PepA and ArgR, such that three negative supercoils are trapped (Colloms *et al.*, 1997). Strand exchange introduces an additional negative node. Two alternative models have been proposed for the Xer complex, in which either one or two PepA molecules, ArgR and the recombinases interact with the two *cer* sites (Alén *et al.*, 1997). Both types of complex are proposed

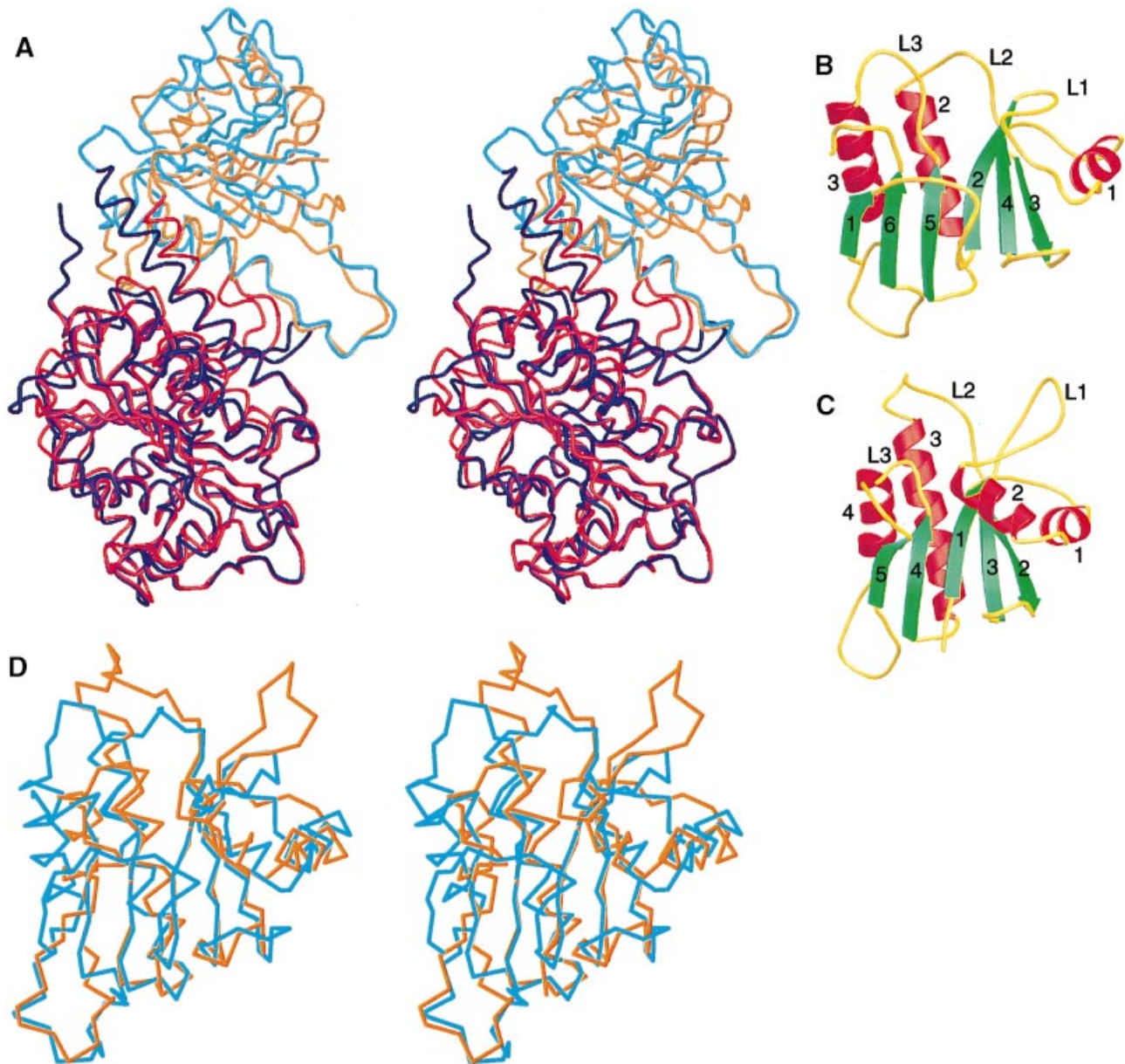


Fig. 4. Comparison of PepA and leucine aminopeptidase. **(A)** Superposition of PepA and LAP based on the C α residues of the C-terminal domains (stereo view). PepA is coloured in blue and cyan and LAP in red and orange for the C- and N-terminal domains, respectively. The N-terminal domain in PepA is rotated by 19° relative to this domain in LAP. The C-terminus is extended by five residues in PepA and interacts with the N-terminal domain. **(B)** Fold of the N-terminal domain of PepA. **(C)** Fold of the N-terminal domain of LAP. **(D)** Stereo view of a superposition of the N-terminal domains of PepA (cyan) and LAP (orange) using the DALI server (Holm and Sander, 1993).

to contain a 2-fold molecular axis, such that each *cer* site makes equivalent interactions with PepA and ArgR.

In the light of the PepA structure we can now build and analyse molecular models for these types of complexes. These models are based on the structures of the individual proteins and on biochemical data on the structure of the complex. Unfortunately, no structures with bound DNA are yet available. Our model-building approach is somewhat reminiscent of the study of Rice and Steitz (1994), who used topological constraints, structures of the proteins involved in the synaptic complex and the crystal packing arrangement to model the recombination synapse of $\gamma\delta$ resolvase.

As argued in the previous section, PepA and ArgR most likely interact with their trimer faces. We have generated a

family of models based on two PepA hexamers sandwiched around a hexamer of ArgR. The most striking feature of this type of molecular sandwich is that the presumed DNA-binding grooves of PepA form right-handed helical paths, about which two *cer* sites could be interwrapped to form a -3 synapse. In these models (Figure 6), two *cer* sites are wrapped around the common 3-fold axis of PepA and ArgR, each *cer* site interacting with PepA, ArgR and PepA again by way of the PEP1, ARG and PEP2 sequences (Figure 1). This leaves two vacant DNA-binding grooves which can bind to a third sequence (PEP3) of each *cer* site in order to juxtapose the two recombination core sites and allow Xer recombination. Each *cer* site therefore interacts with the proteins in the order PEP1-ARG-PEP2-60 bp LOOP-PEP3-XERC-XERD.

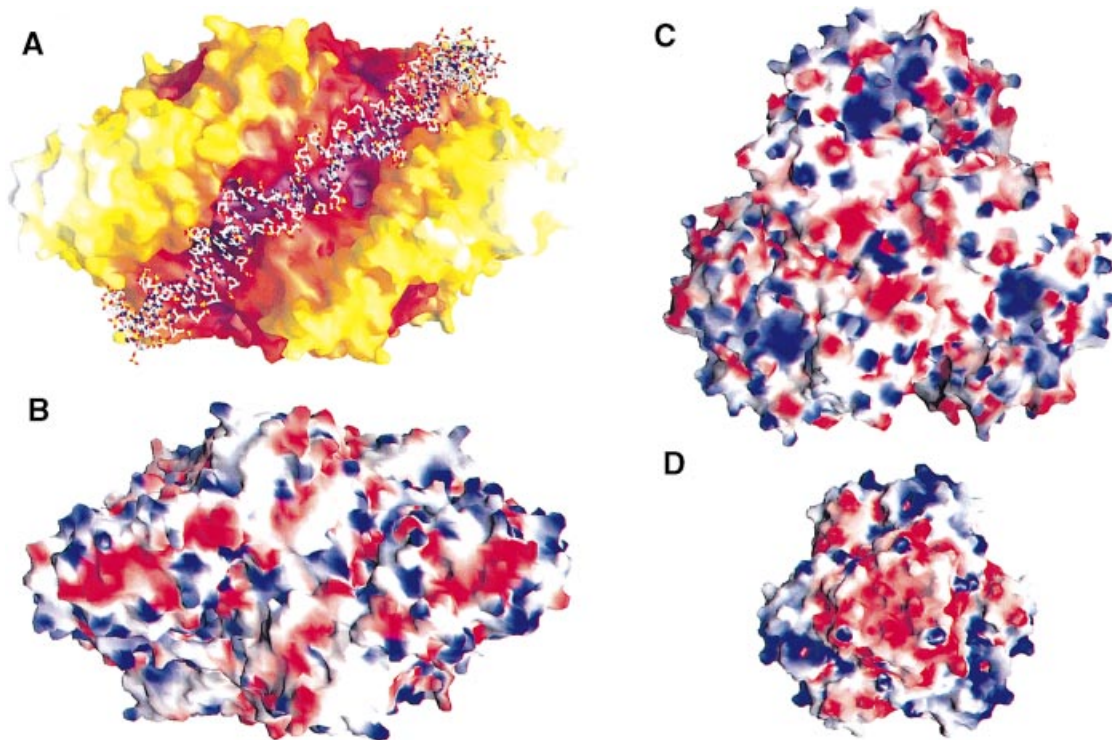


Fig. 5. Molecular surface of PepA and ArgR. (A) Surface of PepA viewed along the 2-fold molecular axis. The surface is coloured by the distance to the centre of mass. A DNA duplex has been modelled into the presumed DNA-binding groove. (B) Electrostatic potential (see Materials and methods) at the molecular surface. Potentials less than -6 kT are coloured in red and potentials >6 kT in blue. (C) and (D) Electrostatic potentials of the trimer faces of PepA (C) and ArgR (D), which are proposed to interact in the Xer nucleoprotein complex for recombination at *cer* sites. The potential for PepA is colour-coded as in (B) between -6 and $+6$ kT, whereas the potential at the surface of ArgR is coloured between -12 and $+12$ kT. The ArgR hexamer is a model which has been built as outlined in Materials and methods.

Although the orientation of the three proteins with respect to a rotation around the 3-fold axis is not known, two extreme situations can be analysed, in which the two PepA molecules either have the same orientation or are rotated by 60° relative to each other (Figure 6A). In order to form a 4-noded catenane, the two *cer* sites must each wrap around the PepA–ArgR sandwich by $>360^\circ$. There are two possible locations for the PEP3 sites relative to the PEP2 sites: $+120^\circ$ (continuing the right-handed path of the DNA; Figure 6A, models A–E) or -120° (model F). In order to yield 4-noded catenanes, the PEP3 site must be in a clockwise ($+120^\circ$) direction relative to PEP2, so as to continue the right-handed interwrapping. A -120° location for PEP3 would undo some of the interwrapping of the two sites. Therefore, model F yields a 2-noded catenane, whereas the otherwise similar model A yields a 4-noded catenane.

Models A–E (Figure 6A) represent a family of models schematically, all of which yield 4-noded catenanes. The models differ in the unknown rotational orientation of the two PepA molecules and in the length of the path which each *cer* site wraps between the PEP1 and PEP2 binding sites. Whereas the DNA wraps by 480° around the 3-fold axis of the sandwich in model A, this amount is reduced in steps of 60° for each successive model to 240° in model E. Models A and B are characterized by a relatively long path of wrapping of the DNA between the PEP1 and PEP2 binding sites and by an optimal orientation of the PEP3 binding sites to position the recombination core sites for strand exchange. In models D and E, the path between PEP1 and PEP2 is significantly shorter; however,

the PEP3 binding grooves are not well positioned to align the XERC and XERD sites for recombination.

In order to estimate the length of the DNA which might be protected in these complexes and to estimate the approximate locations of the binding sites, molecular models were built for some of the complexes of Figure 6A. For model B, the length of DNA is 225 bp between first contacts with the PEP1 binding site and the midpoint of the recombination core site (between XERC and XERD). The distance between PEP1 and PEP2 is ~ 110 bp. There are 70 bp present between PEP2 and PEP3. These numbers are in rough agreement with the results from the DNase I protection assay (Figure 1). The DNA duplex in the region between PEP1 and PEP2 has a superhelix radius of ~ 47 Å and is less bent than the DNA in the co-crystal structure of the nucleosome core particle (radius 41.8 Å; Luger *et al.*, 1997). In the region of the highly bent loop between PEP2 and PEP3, the DNA curvature is also comparable to that of DNA bound to the histone octamer. In model A, ~ 130 bp are present between PEP1 and PEP2, and the total length of the DNA is ~ 250 bp. For model C these values are correspondingly smaller. Based on the regions of protection and hypersensitivity in the DNase I protection assay (Figure 1B), a complex similar to model B or C appears to fit the experimental data best.

The possibility of a Xer complex with only one PepA molecule appears less likely to us in the light of the presumed DNA-binding grooves. Since only three such grooves are present in PepA, each of which can bind ~ 40 – 50 bp of DNA, it is unclear how ~ 200 bp of two *cer* sites can be protected from DNase I by only one hexamer of

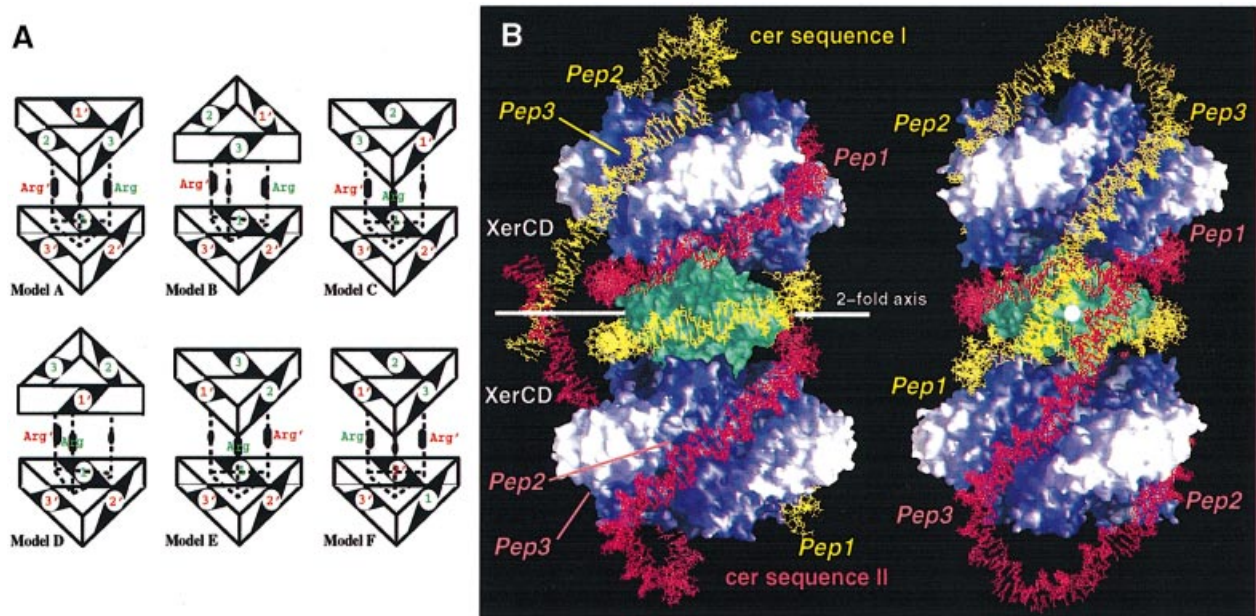


Fig. 6. Models for the Xer complex. **(A)** Scheme of six different models for the synaptic complex at Xer–*cer* recombination. As outlined in the main text, the rotational orientation of the three molecules aligned along the 3-fold axis is not known. Two extreme situations are analysed, in which the two PepA molecules either have the same orientation or differ by 60° relative to each other. PepA and ArgR are schematically shown as triangles, in which the corners mark the positions of the two N-terminal domains that interact closely at the 2-fold axis (PepA) or which bind to one ARG box (ArgR). The positions and orientations of the presumed DNA-binding grooves of PepA are marked in black. These binding sites are labelled 1, 2, 3, and 1', 2', 3' for the three interactions of the two *cer* sites with PepA. Models A–F are representative of a family of models which differ in the rotational orientation of the two PepA molecules and in the length of the path by which the DNA wraps around the 3-fold axis of the sandwich between the PEP1 and PEP2 binding sites. All of these models contain a 2-fold symmetry axis which relates the two PepA molecules and the two *cer* sites and which coincides with the 2-fold symmetry axis of ArgR. Note that the direction of this 2-fold axis of ArgR is not known, i.e. ArgR might be rotated by 180° compared with the orientation shown in these models. Model F differs in the relative location of the PEP2 and PEP3 binding sites (see text). In contrast to the other models, model F does not form 4-noded catenanes. **(B)** Molecular model of complex B viewed along and perpendicular to the 2-fold axis of the Xer complex. PepA and ArgR are represented by their molecular surfaces coloured in blue and green, respectively. The two *cer* sites are coloured in yellow and red.

PepA and one hexamer of ArgR. Furthermore, if the DNA is wrapped around PepA in the direction of the 2-fold molecular axis of PepA, the putative DNA-binding grooves direct the DNA superhelix on a left-handed helical path, which is in disagreement with the formation of a –3 synapse.

Binding of the control region of the *carAB* operon to PepA is proposed to resemble the DNA binding of the region comprising PEP2, the 60 bp loop and PEP3 of our Xer model.

Our models of the Xer complex are in agreement with the topology of the plasmid products and the DNase I protection assays, as shown before. However, the current experimental data do not allow for an exact determination of the binding positions of PepA to the *cer* sequence. Chevrier *et al.* (1995) have aligned the PepA binding sites of the *carAB* operon (as determined by DNase I protection assays) to two putative PepA binding sites of the *cer* sequence. However, the positions of these binding sites are not in good agreement with the results from the DNase I footprinting experiments (Alén *et al.*, 1997). One of the proposed binding sites overlaps with the presumed ~60 bp loop which shows DNase I hyperactivity at sites separated by 10 bp, the other overlaps the XerC binding site in *cer*. PepA may be a sequence-specific binding protein, with 'direct readout' of the nucleotide sequence in the protected regions. Alternatively, PepA may recognize the sequence-dependent bendability [governed by the presence of (A + T)-rich regions at ~10 bp intervals] of the whole *cer*

sequence and the *carAB* control region, including the highly bent 60 bp loop of *cer*. The extent to which these two modes of recognition contribute to DNA binding by PepA is at present unclear, but may be revealed by a comparison of *cer* with its homologues once the exact location of the PepA binding sites on *cer* are known.

The absence of a structure or good model for the DNA-bound form of ArgR and the difficulty in docking the two molecules makes an exact prediction of the structure of the Xer complex a challenging task. Also, we cannot exclude the possibility that structural changes of PepA occur upon binding of DNA. In particular the N-terminal domains, which show the highest thermal parameters, may have a different conformation when DNA is bound. The exact positions of PepA binding on the *cer* sequence have not yet been unambiguously determined. Such results will allow us to discriminate more easily between similar models for the Xer complex. In addition, alternative models have to be considered in which the 3-fold axes of PepA and ArgR are not aligned. There may be little or no direct interactions between PepA and ArgR. One such possibility, in which only two of the three grooves of PepA are filled with *cer* DNA, has been proposed previously by Alén *et al.* (1997). Compared with this model, the Xer model presented here has higher symmetry and contains more interactions between PepA and DNA to explain the extended regions of DNase I protection (Figure 1B).

Nevertheless, the determination of the X-ray structure of PepA, a key determinant of the structure of the Xer

complex, has revealed a putative DNA-binding path and opens the possibility of building alternative models for PepA–DNA complexes and the Xer nucleoprotein complex. These alternative models can be tested by mutagenesis studies and other techniques. Here, based on the structures of PepA and ArgR and on biochemical results, we have proposed a new model for the Xer complex. This model is consistent with the experimental data to date.

Materials and methods

Purification and crystallization of PepA

PepA was purified from *E. coli* strain DS957 (DS941 *pepA::Tn5*) containing plasmid pCA9, which expresses PepA from the *trc* promoter as described previously (McCulloch *et al.*, 1994). The precipitate obtained by dialysis against 20 mM KCl, 50 mM Tris–HCl pH 8.0, 1 mM magnesium acetate was dissolved in a buffer containing 1.2 M KCl, 50 mM Tris pH 8.0 and 1 mM Mg(OAc)₂. The protein solution at a concentration of 8 mg/ml PepA was equilibrated against a buffer containing 200 mM KCl, 50 mM Tris pH 8.0 and 1 mM Mg(OAc)₂. Trigonal, needle-like or rod-shaped crystals were obtained within a few days. Crystals of a mutant protein (E502K,E503K,504G,505R,506R) were obtained by vapour diffusion (hanging drop method) against 2.0 M sodium formate and 0.1 M sodium acetate pH 4.6.

Structure determination

For data collection, crystals were transferred to a buffer containing 200 mM KCl, 50 mM Tris–HCl pH 8.0, 1 mM ZnCl₂, 20% (w/v) MPD and 10% (w/v) glycerol. Crystals of wild-type PepA belong to space group P3₂ with cell dimensions $a = 178.0$ Å and $c = 244.4$ Å for the frozen crystals and they contain two complete hexamers in the asymmetric unit. The mutant protein crystallizes in space group R32 with cell dimensions of $a = 197.4$ Å and $c = 113.3$ Å, and contains one subunit in the asymmetric unit. Data have been collected at EMBL beamline BW7A at DESY in Hamburg (Table I) on a MarResearch 345 mm imaging plate detector. For data reduction programs, DENZO and SCALEPACK were used (Otwinowski, 1993). Since no clear molecular replacement solution was obtained in space group P3₂ using LAP as a search model (Burley *et al.*, 1990; Sträter and Lipscomb, 1995a), the R32 crystal form which contained only one subunit in the asymmetric unit was used for molecular replacement with LAP using program AMORE (Navaza, 1994). A clear solution was apparent in the rotation and translation functions and the structure was partially refined in R32 without manual rebuilding. Using a hexamer of this model as the search coordinates the orientations of the two hexamers in the P3₂ crystal form were determined from the rotation and translation functions. The initial phases were refined by 12-fold electron density averaging over the subunits related by non-crystallographic symmetry (program DM; CCP4, 1994). Largely improved electron density maps were obtained, especially for the N-terminal domain. The model was rebuilt with program O (Jones *et al.*, 1991) and refined by simulated annealing and conjugate gradient minimization against standard maximum likelihood targets as implemented in program CNS (Brünger *et al.*, 1998). The models for the 12 subunits were restrained to obey the non-crystallographic symmetry, except for 73 residues for which differences are apparent between the subunits. Further details and statistics on the final model are presented in Table I and in the deposited Protein Data Bank.

Continuum electrostatic calculations

Electrostatic potential maps were calculated by solution of the Poisson–Boltzmann equation in a continuum electrostatic model as implemented in the program DelPhi (Gilson *et al.*, 1987). Focussing was used with 40% box fill for the first run and 90% box fill for the focussing run. Independence of the resulting potentials from the mapping of the molecule onto the grid (65×65×65 Å³) was assured by moving the molecule around slightly on the grid as well as by checking whether the potential maps obeyed the molecular symmetry. Single runs with a standard box fill of 60% without focussing were found not to meet these criteria for the large PepA molecule. A probe radius of 1.8 Å was used and the dielectric constant was set to 2 for the protein region and to 80 for the solvent.

Model building and generation of figures

For *E. coli* ArgR, a crystal structure of the hexameric C-terminal domain and an NMR structure for the N-terminal domain are available (van Duyn

et al., 1996; Sunnerhagen *et al.*, 1997). A model for the complete ArgR hexamer was generated by manually docking the N-terminal domains to the sides of the hexameric core, guided by the structure of the homologous *B. stearothermophilus* ArgR hexamer in the absence of DNA (Ni *et al.*, 1999). PepA and ArgR were manually docked such that their 3-fold molecular axes are aligned. Ideal B-DNA duplexes with a rise of 3.38 Å were modelled using program NAB (Macke and Case, 1998). This program inputs a set of points placed manually along the presumed DNA-binding path and interpolates the positions of the base pair origins using a cubic spline function. All manual model building was done with program O (Jones *et al.*, 1991). Figures 2 and 4 were created with MOLSCRIPT (Kraulis, 1991) and rendered with Raster3D (Merritt and Bacon, 1997), and Figures 5 and 6B were calculated using GRASP (Nichols *et al.*, 1991).

Acknowledgements

N.S. thanks W.Saenger for generous support in his laboratory. W.N. Lipscomb is acknowledged for contributions in the early phase of this project, supported by grant GM06920. We thank W.Rypniewski, I.Przylas and T.Knöfel for help with the data collection at the EMBL beamline at DESY, Hamburg. This work was supported by a grant from the Deutsche Forschungsgemeinschaft to N.S.

References

- Alén,C., Sherratt,D.J. and Colloms,S.D. (1997) Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *EMBO J.*, **16**, 5188–5197.
- Blakely,G., May,G., McCulloch,R., Arciszewska,L., Burke,M., Lovett,S. and Sherratt,D.J. (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *Escherichia coli* K12. *Cell*, **75**, 351–361.
- Brünger,A.T. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905–921.
- Burley,S.K., David,P.R., Taylor,A. and Lipscomb,W.N. (1990) Molecular structure of leucine aminopeptidase at 2.7 Å resolution. *Proc. Natl Acad. Sci. USA*, **87**, 6878–6882.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Charlier,D., Hassanzadeh,G., Kholti,A., Gigot,D., Piérard,A. and Glansdorff,N. (1995) *carP*, involved in pyrimidine regulation of the *Escherichia coli* carbamoylphosphate synthetase operon encodes a sequence-specific DNA-binding protein identical to XerB and PepA, also required for resolution of ColE1 multimers. *J. Mol. Biol.*, **250**, 392–406.
- Colloms,S.D., Sykora,P., Szatmari,G. and Sherratt,D.J. (1990) Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the λ integrase family of site-specific recombinases. *J. Bacteriol.*, **172**, 6973–6980.
- Colloms,S.D., McCulloch,R., Grant,K., Neilson,L. and Sherratt,D.J. (1996) Xer-mediated site-specific recombination *in vitro*. *EMBO J.*, **15**, 1172–1181.
- Colloms,S.D., Bath,J. and Sherratt,D.J. (1997) Topological selectivity in Xer site-specific recombination. *Cell*, **88**, 855–864.
- Gilson,M.K., Sharp,K.A. and Honig,B.H. (1987) Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.*, **9**, 327–335.
- Gopaul,D.N., Guo,F. and van Duyn,G.D. (1998) Structure of the Holliday junction intermediate in Cre-lox P site-specific recombination. *EMBO J.*, **17**, 4175–4187.
- Hayes,F., Hallett,B. and Cao,Y. (1997) Insertion mutagenesis as a tool in the modification of protein function. Extended substrate specificity conferred by pentapeptide insertions in the omega-loop of TEM-1 β -lactamase. *J. Biol. Chem.*, **272**, 28833–28836.
- Hodgman,T.C., Griffiths,H. and Summers,D.K. (1998) Nucleoprotein architecture and ColE1 dimer resolution: a hypothesis. *Mol. Microbiol.*, **29**, 545–558.
- Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
- Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Kholti,A., Charlier,D., Gigot,D., Huysveld,N., Roovers,M. and Glansdorff,N. (1998) *pyrH*-encoded UMP-kinase directly participates

- in pyrimidine-specific modulation of promoter activity in *Escherichia coli*. *J. Mol. Biol.* **280**, 571–582.
- Kim,H. and Lipscomb,W.N. (1994) Structure and mechanism of bovine lens leucine aminopeptidase. *Adv. Enzymol.*, **68**, 153–213.
- Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
- Macke,T. and Case,D.A. (1998) Modeling unusual nucleic acid structures. In Leontes,N.B. and SantaLucia,J. (eds), *Molecular Modeling of Nucleic Acids*. American Chemical Society, Washington, DC, pp.379–393.
- McCulloch,R., Burke,M.E. and Sherratt,D.J. (1994) Peptidase activity of *Escherichia coli* aminopeptidase A is not required for its role in Xer site-specific recombination. *Mol. Microbiol.*, **12**, 241–251.
- Merritt,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.
- Ni,J., Sakanyan,V., Charlier,D., Glansdorff,N. and van Duyne,G.D. (1999) Structure of the arginine repressor from *Bacillus stearothermophilus*. *Nature Struct. Biol.*, **6**, 427–432.
- Nicholls,A., Sharp,K. and Honig,B.H. (1991) Protein folding and association; insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Otwinowski,Z. (1993) *Data Collection and Processing*. SERC Daresbury DL/SCI/R34, pp. 56–62.
- Rice,P.A. and Steitz,T.A. (1994) Model for a DNA-mediated synaptic complex suggested by crystal packing of $\gamma\delta$ resolvase subunits. *EMBO J.*, **13**, 1514–1524.
- Stirling,C.J., Szatmari,G., Stewart,G., Smith,M.C.M. and Sherratt,D.J. (1988) The arginine repressor is essential for plasmid-stabilising site-specific recombination at the ColE1 *cer* locus. *EMBO J.*, **7**, 4389–4395.
- Stirling,C.J., Colloms,S.D., Collins,J.F., Szatmari,G. and Sherratt,D.J. (1989) *xerB*, an *Escherichia coli* gene required for plasmid ColE1 site-specific recombination, is identical to *pepA*, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *EMBO J.*, **8**, 1623–1627.
- Sträter,N. and Lipscomb,W.N. (1995a) Two-metal ion mechanism of bovine lens leucine aminopeptidase: active site solvent structure and binding mode of L-leucinal, a *gem*-diolate transition state analogue, by X-ray crystallography. *Biochemistry*, **34**, 14792–14800.
- Sträter,N. and Lipscomb,W.N. (1995b) Transition state analogue L-leucinephosphonic acid bound to bovine lens leucine aminopeptidase: X-ray structure at 1.65 Å resolution in a new crystal form. *Biochemistry*, **34**, 9200–9210.
- Sträter,N. and Lipscomb,W.N. (1998) Leucyl aminopeptidase (animal and plant). In Barrett,A.J., Rawlings,N.D. and Woessner,J.F. (eds), *Handbook of Proteolytic Enzymes*. Academic Press, New York, NY, pp. 1384–1389.
- Subramanya,H.S., Arciszewska,L.K., Baker,R.A., Bird,L.E., Sherratt,D.J. and Wigley,D.B. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J.*, **16**, 5178–5187.
- Summers,D.K. and Sherratt,D.J. (1984) Multimerisation of high copy number plasmids causes instability: ColE1 encodes a determinant for plasmid monomerisation and stability. *Cell*, **36**, 1097–1103.
- Sunnerhagen,M., Nilges,M., Otting,G. and Carey,J. (1997) Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. *Nature Struct. Biol.*, **4**, 819–826.
- van Duyne,G.D., Ghosh,G., Maas,W.K. and Sigler,P.B. (1996) Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. *J. Mol. Biol.*, **256**, 377–391.
- Vogt,V.M. (1970) Purification and properties of an aminopeptidase from *Escherichia coli*. *J. Biol. Chem.*, **245**, 4760–4769.

Received April 19, 1999; revised and accepted June 28, 1999