Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural and functional characterization of the C-terminal catalytic domain of SSV1 integrase

Zhengyan Zhan,[a]‡ Songying Ouyang,[b]‡ Wenguang Liang,[b] Zhenfeng Zhang,[a] Zhi-Jie Liu[b,c]* and Li Huang[a]*

[a]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, People's Republic of China, [b]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, People's Republic of China, and [c]College of Molecular and Clinical Medicine, Kunming Medical University, 1168 West Chunrong Road, Chenggong District, Kunming 650500, People's Republic of China

‡ These authors contributed equally to this work.

Correspondence e-mail: zjliu@ibp.ac.cn, huangl@sun.im.ac.cn

Received 4 January 2012
Accepted 17 February 2012

PDB References: C-terminal domain of SSV1 integrase, 3vcf; chymotrypsin-treated SSV1 integrase, 4dks.

The spindle-shaped virus SSV1 of the hyperthermophilic archaeon *Sulfolobus shibatae* encodes an integrase (SSV1 Int). Here, the crystal structure of the C-terminal catalytic domain of SSV1 Int is reported. This is the first structural study of an archaeal tyrosine recombinase. Structural comparison shows that the C-terminal domain of SSV1 Int possesses a core fold similar to those of tyrosine recombinases of both bacterial and eukaryal origin, apart from the lack of a conserved helix corresponding to αI of Cre, indicating conservation of these enzymes among all three domains of life. Five of the six catalytic residues cluster around a basic cleft on the surface of the structure and the nucleophile Tyr314 is located on a flexible loop that stretches away from the central cleft, supporting the possibility that SSV1 Int cleaves the target DNA in a *trans* mode. Biochemical analysis suggests that the N-terminal domain is responsible for the dimerization of SSV1 Int. The C-terminal domain is capable of DNA cleavage and ligation, but at efficiencies significantly lower than those of the full-length protein. In addition, neither the N-terminal domain alone nor the C-terminal domain alone shows a strong sequence preference in DNA binding. Therefore, recognition of the core-type sequence and efficient catalysis by SSV1 Int presumably requires covalent linkage and interdomain communication between the two domains.

1. Introduction

Tyrosine recombinases (*i.e.* members of the λ integrase family), named after the nucleophilic tyrosine residue that forms a transient 3′-phosphotyrosine covalent linkage to the DNA substrate in the reaction intermediate, make up one of the two large classes of site-specific recombinases. They catalyze a variety of sequence-specific DNA rearrangements through two rounds of pairwise cleavage, exchange and rejoining of single strands in the context of a tetrameric recombinase bound to two recombining DNAs, producing a Holliday-junction intermediate. These recombination reactions yield many important biological outcomes in both prokaryotes and eukaryotes, such as the integration and excision of a phage genome into and from the host genome (λ Int and HP1 Int), the maintenance of plasmid copy number (Flp) and the resolution of replicon dimers into monomers (Cre and XerC/D), as well as the mobilization of gene cassettes of integrons (IntI) (Grindley *et al.*, 2006; Esposito & Scocca, 1997; Rajeev *et al.*, 2009).

Biochemical and structural studies have provided a wealth of molecular details on tyrosine recombinases of both bacterial and eukaryal origin. Structures are now available for

λ Int, HP1 Int and XerD, as well as for protein–DNA synaptic complexes of Cre, Flp, λ Int and IntI (Kwon *et al.*, 1997; Hickman *et al.*, 1997; Subramanya *et al.*, 1997; Guo *et al.*, 1997; Biswas *et al.*, 2005; MacDonald *et al.*, 2006; Chen *et al.*, 2000). Most tyrosine recombinases adopt a two-domain structure for interaction with crossover DNA sites (also known as core-site DNA in the λ system), with a structurally conserved C-terminal catalytic domain and a variable N-terminal domain making up a C-shaped clamp around the DNA substrate. The catalytic domains lack significant sequence similarity, except for the active site, but share a similar three-dimensional fold. The active site consists of the critical nucleophile Tyr and a highly conserved pentad, *i.e.* R-K-H*xx*R-H (or W), which are directly involved in the catalysis of DNA-strand cleavage and exchange (Argos *et al.*, 1986; Abremski & Hoess, 1992; Esposito & Scocca, 1997; Cao & Hayes, 1999). The active site in tyrosine recombinases resembles those in the catalytic domains of type IB topoisomerases and protelomerases, suggesting that these enzymes are evolutionarily related and employ a similar catalytic mechanism (Cheng *et al.*, 1998; Deneke *et al.*, 2000). In tyrosine recombinases, two distinct forms of active-site organization have been found. One form permits *cis* cleavage, in which the nucleophile Tyr and the conserved catalytic pentad are provided by a single protomer, as is observed in all structurally determined protein–DNA complexes of bacterial tyrosine recombinases (Guo *et al.*, 1997; Aihara *et al.*, 2003; MacDonald *et al.*, 2006). The other supports *trans* cleavage, in which the pentad and the nucleophile Tyr are from two neighbouring protomers, as is observed in the eukaryotic Flp–DNA synaptic complex structures (Chen *et al.*, 2000). Tyrosine recombinases also exhibit half-of-the-sites activity, as alternating protomers within a synaptic tetramer are active at any given time. Structural data suggest that regulation of the half-of-the-sites activity involves swapping a portion of the C-terminus of the catalytic domain in one protomer, *i.e.* the segment immediately following the tyrosine-bearing helix in Cre, Flp, λ Int and IntI or the tyrosine-bearing helix itself in Flp, with that of a neighbouring protomer (Grindley *et al.*, 2006).

Much less is known about tyrosine recombinases in Archaea. No structures are currently available for the archaeal enzymes. An integrase (SSV1 Int) encoded by the spindle-shaped virus SSV1 isolated from the hyperthermophilic crenarchaeon *Sulfolobus shibatae* was the first archaeal member of the tyrosine recombinase family to be biochemically characterized (Muskhelishvili *et al.*, 1993). SSV1 Int catalyzes site-specific integration of the viral DNA into the host chromosome using viral and chromosomal attachment sites, *i.e.* *att*P and *att*A (also denoted *att*B), respectively, and the excision of the integrated provirus using the prophage attachment sites *att*L and *att*R. The integration and excision reactions catalyzed by SSV1 Int do not appear to involve any other accessory proteins (Muskhelishvili *et al.*, 1993; Muskhelishvili, 1994). Sequence analysis suggests that SSV1 Int employs an active site similar to those in other tyrosine recombinases (Argos *et al.*, 1986; Esposito & Scocca, 1997). Previous studies have demonstrated that SSV1 Int utilizes the invariant nucleophile Tyr314 to form a covalent 3′-phosphoprotein intermediate (Serre *et al.*, 2002). However, SSV1 Int contains substitutions at several positions in the conserved catalytic pentad and therefore cannot be phylogenetically classified into any known subgroup of the tyrosine recombinase family. These changes include the substitution of Arg for the first conserved Lys and that of Lys for the first canonical His, as well as that of Arg for the second moderately conserved His (or Trp in Cre and Flp). Site-directed mutagenesis has shown that these substitutions are probably conservative changes (Letzelter *et al.*, 2004). In addition, complementation experiments *in vitro* have raised the possibility that SSV1 Int adopts a *trans* cleavage mechanism (Letzelter *et al.*, 2004). Here, we report the crystal structure and biochemical characterization of the C-terminal catalytic domain of SSV1 Int. We show that the catalytic domain of SSV1 Int shares a conserved core fold with those of tyrosine recombinases from Bacteria and Eukarya, suggesting a common evolutionary origin of these enzymes in the three domains of life. The C-terminal domain of the enzyme is capable of DNA cleavage and ligation, whereas the N-terminal domain is responsible for dimerization of SSV1 Int.

2. Materials and methods

2.1. Construction of expression vectors

The wild-type SSV1 integrase gene (ORF D335) and its truncated mutants N173 (residues 1–173) and C174 (residues 174–335) were amplified by high-fidelity PCR using SSV1 genomic DNA as the template (for primer sequences, see Supplementary Table S1[1]) and cloned into expression plasmid pET30a(+) (Novagen) between the *Nde*I and *Bam*HI sites. The sequences of the inserts in all constructs were verified by DNA sequencing.

2.2. Protein overproduction and purification

Recombinant proteins were overproduced in *Escherichia coli* strain BL21 (DE3) after induction with 1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) when the culture had grown to an optical density of 1.0 at 600 nm in LB supplemented with kanamycin (50 μg ml^{-1}) at 310 K. Following 3 h induction, the cells were harvested by centrifugation, resuspended in buffer *A* [20 m*M* Tris–HCl pH 6.8, 1 m*M* DTT, 0.1 m*M* EDTA, 0.3 *M* NaCl, 10%(*w*/*v*) glycerol] supplemented with protease inhibitors (50 μg ml^{-1} phenylmethylsulfonyl fluoride, 1 μg ml^{-1} aprotinin, 0.2 μg ml^{-1} benzamidine, 0.1 μg ml^{-1} leupeptin, 0.5 μg ml^{-1} pepstatin A) and sonicated on ice. The lysate was centrifuged at 20 000*g* for 30 min at 277 K and the supernatant was heat-treated at 343 K for 30 min. The sample was centrifuged at 20 000*g* for 30 min at 277 K. The supernatant was adjusted to an NaCl concentration of 1 *M*. Polyethylenimine (PEI; Sigma) was added to a final concentration of 0.3%. After stirring for 1 h, the mixture was

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: RR5015). Services for accessing this material are described at the back of the journal.

Table 1
Data-collection and refinement statistics for SSV1 Int.

Values in parentheses are for the highest resolution shell.

		Chymotrypsin-treated SSV1 Int	
Derivatization	C174	Native	SeMet
PDB code	3vcf	4dks	
Data collection			
X-ray source	BL-1A, PF, KEK	19-ID, APS	19-ID, APS
Detector	ADSC 315	ADSC 315	ADSC 315
Crystal-to-detector distance (mm)	250.00	350.00	330.00
No. of images	90	180	240
Oscillation width (°)	1.00	1.00	1.00
Wavelength (Å)	1.00	0.98	0.98 [Se peak]
Space group	$P4_32_12$	$P4_12_12$	$P4_12_12$
Unit-cell parameters (Å)	$a = b = 89.1$, $c = 147.7$	$a = b = 75.2$, $c = 96.1$	$a = b = 75.2$, $c = 96.1$
Resolution range (Å)	50.00–2.70 (2.80–2.70)	50.00–2.70 (2.80–2.70)	50.00–2.77 (2.87–2.77)
R_{merge} (%)	11.1 (43.6)	6.0 (38.2)	8.6 (39.8)
Mean $I/\sigma(I)$	19.1 (4.6)	35.6 (5.8)	52.6 (8.9)
Completeness (%)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
Multiplicity	7.1 (7.3)	9.3 (9.5)	17.9 (19.0)
Refinement			
Resolution (Å)	34.11–2.70	40.49–2.70	
No. of reflections	16941	8010	
R_{work}/R_{free} (%)	18.7/23.8	20.3/26.1	
No. of atoms	2692	1266	
Protein	2572	1199	
Water	120	78	
Mean B value (Å^2)	29.2	54.4	
R.m.s. deviations			
Bond lengths (Å)	0.01	0.01	
Bond angles (°)	1.00	1.07	
Ramachandran analysis			
Favoured region	284 [96.3%]	133 [97.1%]	
Allowed region	7 [2.4%]	4 [2.9%]	
Outliers	4 [1.4%]	0 [0.0%]	

centrifuged at 12 000*g* for 30 min at 277 K to remove nucleic acids. The supernatant was adjusted to 70% saturation with ammonium sulfate and centrifuged again. The pellet was resuspended in buffer *A* and dialyzed against the same buffer. The dialyzed sample was applied onto an equilibrated HiTrap SP column (5 ml; GE Healthcare), washed with buffer *A* and eluted with a 0.3–1 *M* NaCl gradient in buffer *A*. Fractions were analyzed by 15% SDS–PAGE and those containing pure wild-type or mutant integrase protein were pooled. For biochemical studies, purified proteins were dialyzed against buffer *B* [20 m*M* Tris–HCl pH 7.5, 1 m*M* DTT, 0.1 m*M* EDTA, 0.3 *M* NaCl, 10%(*w*/*v*) glycerol], aliquoted and stored at 203 K. For crystallization trials, purified proteins were dialyzed against buffer *C* [20 m*M* Tris–HCl pH 7.5, 300 m*M* NaCl, 5%(*w*/*v*) glycerol, 1 m*M* DTT] and concentrated to ∼10 mg ml^{-1} at 283 K. Protein concentrations were determined by the Lowry method using bovine serum albumin (BSA) as the standard.

2.3. Crystallization, data collection and structure determination

Both conventional (without protease treatment) and *in situ* proteolysis methods were employed to obtain crystals of SSV1 Int. Crystallization screening was carried out using commercially available sparse-matrix screens (Hampton Research). However, only the *in situ* proteolysis method yielded crystals. Briefly, α-chymotrypsin (Sigma) was added to the purified protein at a mass ratio of 1:1000 immediately prior to crystallization trials, as described previously (Dong *et al.*, 2007). In the optimization stage, crystallization was performed in hanging drops at 289 K by the addition of 1 µl of the protease–protein mixture to 1 µl reservoir solution [0.2 *M* NaCl or $MgCl_2$, 0.1 *M* HEPES pH 7.5, 25%(*w*/*v*) polyethylene glycol 3350]. The crystals were mounted on a nylon loop and immediately cooled in liquid nitrogen without any cryo-treatment. Diffraction data for native crystals were collected to 2.70 Å resolution on beamline 19-ID at APS. SeMet SSV1 Int was crystallized using the same *in situ* proteolysis method. Data for Se-SAD phasing were collected to 2.77 Å resolution on beamline 19-ID at APS. Two sets of data were integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Se-atom positions were determined and phases were obtained using *SOLVE* (Terwilliger & Berendzen, 1999).

The C-terminal domain of SSV1 Int (C174) was crystallized in hanging drops at 289 K, mixing an equal volume of C174 (10 mg ml^{-1}) with well solution (0.056 *M* NaH_2PO_4, 1.35 *M* K_2HPO_4 pH 8.2). Diffraction data for C174 crystals were collected to 2.70 Å resolution on beamline BL-1A of Photon Factory, KEK, Tsukuba, Japan. The structure was determined by molecular replacement using *Phaser* (McCoy *et al.*, 2007) from the *CCP*4 program suite (Winn *et al.*, 2011) with the structure of chymotrypsin-treated SSV1 Int as the search model.

The initial phases were improved with *OASIS* (He *et al.*, 2007). The model was manually improved in *Coot* (Emsley *et al.*, 2010). Refinement was carried out using *REFMAC* (Murshudov *et al.*, 2011) and *PHENIX* (Terwilliger *et al.*, 2008) alternately. Statistics of data collection and refinement are given in Table 1. The quality of the final model was validated with *MolProbity* (Chen *et al.*, 2010). The coordinates and structure factors for the C174 crystal and the chymotrypsin-treated SSV1 Int crystal have been deposited in the PDB under accession codes 3vcf and 4dks, respectively. All images were prepared using *PyMOL* (DeLano, 2002) and *ESPript* (Gouet *et al.*, 1999).

2.4. Gel filtration

Protein samples (100 µl; 1.3 mg ml^{-1} for SSV1 Int and ∼0.6 mg ml^{-1} for each truncated mutant) were loaded onto a Superdex 200 column (10/300 GL; GE Healthcare) equili-

brated in buffer *B* and run at 0.5 ml min^{-1} at 283 K on an ÄKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare). The column was calibrated with gel-filtration calibration kits (LMW and HMW; GE Healthcare) and alcohol dehydrogenase (150 kDa).

2.5. Blue native PAGE

Blue native polyacrylamide gel electrophoresis (PAGE) was performed as described previously (Schägger *et al.*, 1994). Samples (20 µg) of the proteins were dissolved in sample buffer for blue native PAGE [50 m*M* ε-aminocaproic acid, 10 m*M* bis-Tris–HCl, 5 mg ml^{-1} Serva Blue G, 300 m*M* NaCl, 7.5%(*w*/*v*) sucrose pH 7.0 at 277 K] to a final concentration of 1 mg ml^{-1} and loaded onto a blue native PAGE gel with a 4% stacking gel and a 6–20% separating gel. Electrophoresis was initially conducted at 100 V for 1 h and was subsequently performed with the current limited to 15 mA and the voltage limited to 500 V. Thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), alcohol dehydrogenase (150 kDa), conalbumin (75 kDa) and BSA (66 and 132 kDa) were used as molecular-weight markers.

2.6. Substrate preparation

Oligonucleotides were synthesized (Sangon BioTech, Shanghai, People's Republic of China) and labelled at the 5′-end with [γ-^{32}P]-ATP using polynucleotide kinase. Unincorporated nucleotides were removed by spin dialysis and the labelled oligonucleotides were annealed with a 1.5-fold excess of unlabelled complementary strand in TE buffer (10 m*M* Tris–HCl pH 8.0, 1 m*M* EDTA).

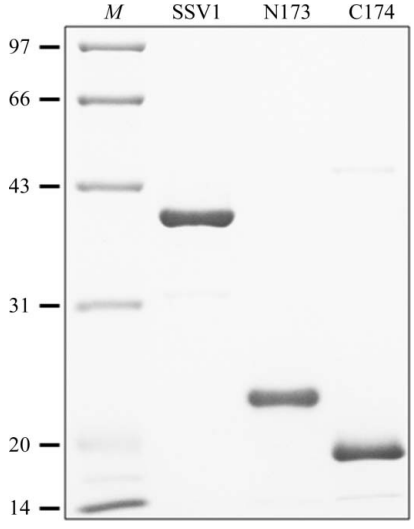

Figure 1
Analysis of purified SSV1 Int and truncated mutants by SDS–PAGE. An aliquot (3 µg) of each purified recombinant protein was subjected to 15% SDS–PAGE. The gel was stained with Coomassie Brilliant Blue R-250. Lane *M*, molecular-mass markers (labelled in kDa on the left). SSV1 Int, amino acids 1–335; N173, amino acids 1–173 ; C174, amino acids 174–335.

2.7. Electrophoretic mobility-shift assays (EMSA)

The standard EMSA reactions (20 µl) consisted of 20 m*M* bis-Tris–HCl pH 6.7, 50 µg ml^{-1} BSA, 0–150 m*M* NaCl, 1 m*M* EDTA, 6% glycerol, 2 n*M* 5′-end-labelled attachment-site DNA fragment (SSV1_A58 or SSV1_A118; Supplementary Table S1) and the indicated amount of wild-type or truncated SSV1 Int. For competitive EMSA, wild-type or truncated SSV1 Int (300 n*M*) was incubated with 2 n*M* 5′-end-labelled SSV1_A58 and increasing amounts of unlabelled SSV1_A58 or poly(dI–dC)·poly(dI–dC) in binding buffer with 150 m*M* NaCl. After incubation for 10 min at room temperature, the reaction mixtures were loaded onto a 5% nondenaturating polyacrylamide gel which had been pre-run to a constant current and electrophoresis was performed in 0.1× Tris–borate–EDTA (TBE) buffer at room temperature for 2 h at 10 V cm^{-1}. The gel was dried, exposed to X-ray film and analyzed using a PhosphorImager (Amersham Biosciences).

2.8. DNA topoisomerase activity assays

Supercoiled plasmid pBR322 DNA (0.6 µg; Fermentas) was incubated for 4 h at 338 K with the indicated amount of wild-type or truncated SSV1 Int in 50 m*M* bis-Tris–HCl pH 6.7, 50 µg ml^{-1} BSA, 5 m*M* EDTA, 150 m*M* NaCl in a total volume of 20 µl. After incubation, samples were treated with SDS (1% final concentration) and protease K (1 mg ml^{-1} final concentration) for 1 h at 323 K and mixed with a 1/10 volume of loading buffer (1% SDS, 50% glycerol, 0.05% bromophenol blue, 0.05% xylene cyanol FF). An aliquot of each reaction mixture (containing 0.3 µg pBR322) was loaded onto a 1.2% agarose gel and electrophoresis was performed in 1× Tris–acetate–EDTA (TAE) buffer at room temperature for 3.5 h at 3 V cm^{-1}. DNA was visualized by staining with ethidium bromide. As a positive control, plasmid pBR322 was incubated for 30 min at 310 K with calf thymus topoisomerase I (10 U; Takara) and the sample was treated as described above.

2.9. DNA cleavage assays

The standard cleavage reactions (30 µl) consisted of 20 m*M* bis-Tris–HCl pH 6.7, 50 µg ml^{-1} BSA, 150 m*M* NaCl, 12 n*M* 5′-end-labelled attachment-site DNA fragment (SSV1_XTB; Supplementary Table S1) and the indicated amount of wild-type or truncated SSV1 Int. Reaction mixtures were incubated for 4 h at 338 K and quenched in 40 m*M* Tris–HCl pH 6.8, 3% SDS, 8% glycerol, 250 m*M* β-mercaptoethanol and 0.005% bromophenol blue. Reaction products were analyzed by electrophoresis (12% SDS–PAGE). The gels were dried and exposed to X-ray film.

3. Results and discussion

3.1. Structure determination

In order to obtain crystals of SSV1 Int, the gene encoding the protein (gi:9625521) was cloned and overexpressed in *E. coli* and the recombinant protein was purified to homogeneity (Fig. 1). Both conventional (without protease

treatment) and *in situ* chymotrypsin proteolysis methods were used to crystallize the protein. However, crystals were only obtained using the latter method. Since SSV1 Int shares less than 30% homology at the amino-acid sequence level with the best-matching proteins in the PDB, crystallographic phases could not be resolved by molecular replacement. Therefore, we used the single-wavelength anomalous dispersion (SAD) method to solve the phase problem. SeMet SSV1 Int was crystallized using the same *in situ* proteolysis method. The resolved structure of chymotrypsin-treated SSV1 Int consisted of residues 174–237, 244–271, 274–306 and 315–334, with one protomer occupying the asymmetric unit in space group $P4_12_12$ (Table 1). The resistance of residues 174–334 to chymotrypsin treatment raised the possibility that this region represents a structurally compact domain in SSV1 Int. Indeed, sequence alignment with other tyrosine recombinases identified this region as the C-terminal catalytic domain of the protein. Unfortunately, the crystal structure lacked electron density for several important regions, including those in which the conserved Arg240 and the nucleophile Tyr314 are located. *B*-factor sharpening (DeLaBarre & Brunger, 2003; Su *et al.*, 2010) was attempted in order to unveil the missing key residues from enhanced weak electron densities, but failed to yield sufficiently convincing solutions.

The above observations prompted us to prepare two truncated mutants of SSV1 Int, *i.e.* N173 (amino acids 1–173) and C174 (amino acids 174–335) (Fig. 1), and to crystallize the two mutant proteins. Crystals were only obtained for C174. The structure of C174 was resolved in space group $P4_32_12$ by molecular replacement using the structure of chymotrypsin-treated SSV1 Int as the search model. The asymmetric unit of the C174 crystal included two protomers (Supplementary Fig. S1A). Protomer *A* of C174 contained continuous residues 176–335 (the intact C-terminal domain), whereas protomer *B* contained residues 177–238, 244–305 and 314–335. The two protomers of C174 as well as the chymotrypsin-treated SSV1 Int are essentially identical. However, it is worth noting that the two loops that contain Arg240 and Tyr314 appear to be flexible: they are visible in protomer *A* but invisible in protomer *B* and in the chymotrypsin-treated SSV1 Int (Supplementary Fig. S1B).

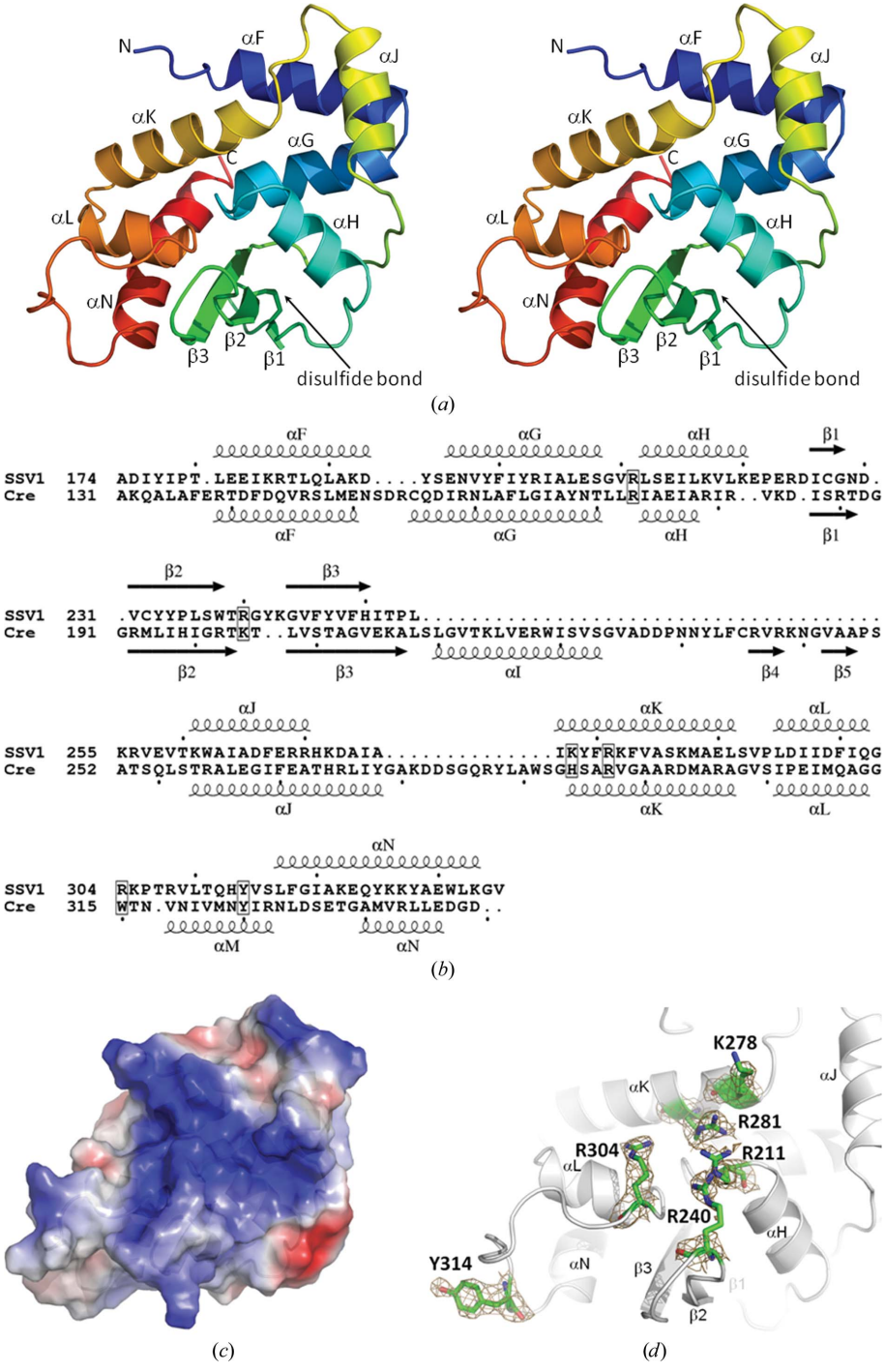

Figure 2
Overall structure of the C174 domain. (*a*) Stereoview of the C174 domain. The secondary-structure elements were numbered according to the structure superposition with Cre and the positions of the amino- and carboxyl-termini are indicated. A disulfide bond bridging the C-terminus of $\beta1$ and the N-terminus of $\beta2$ is indicated by an arrow. (*b*) Structure-based sequence alignment of C174 and Cre with their corresponding secondary-structure elements indicated. Active-site residues are framed. (*c*) The solvent-accessible surface of C174 coloured according to electrostatic potential. Blue, positively charged; red, negatively charged; white, neutral. (*d*) Experimental electron density of the active-site residues. The views in (*c*) and (*d*) are similar to that in (*a*).

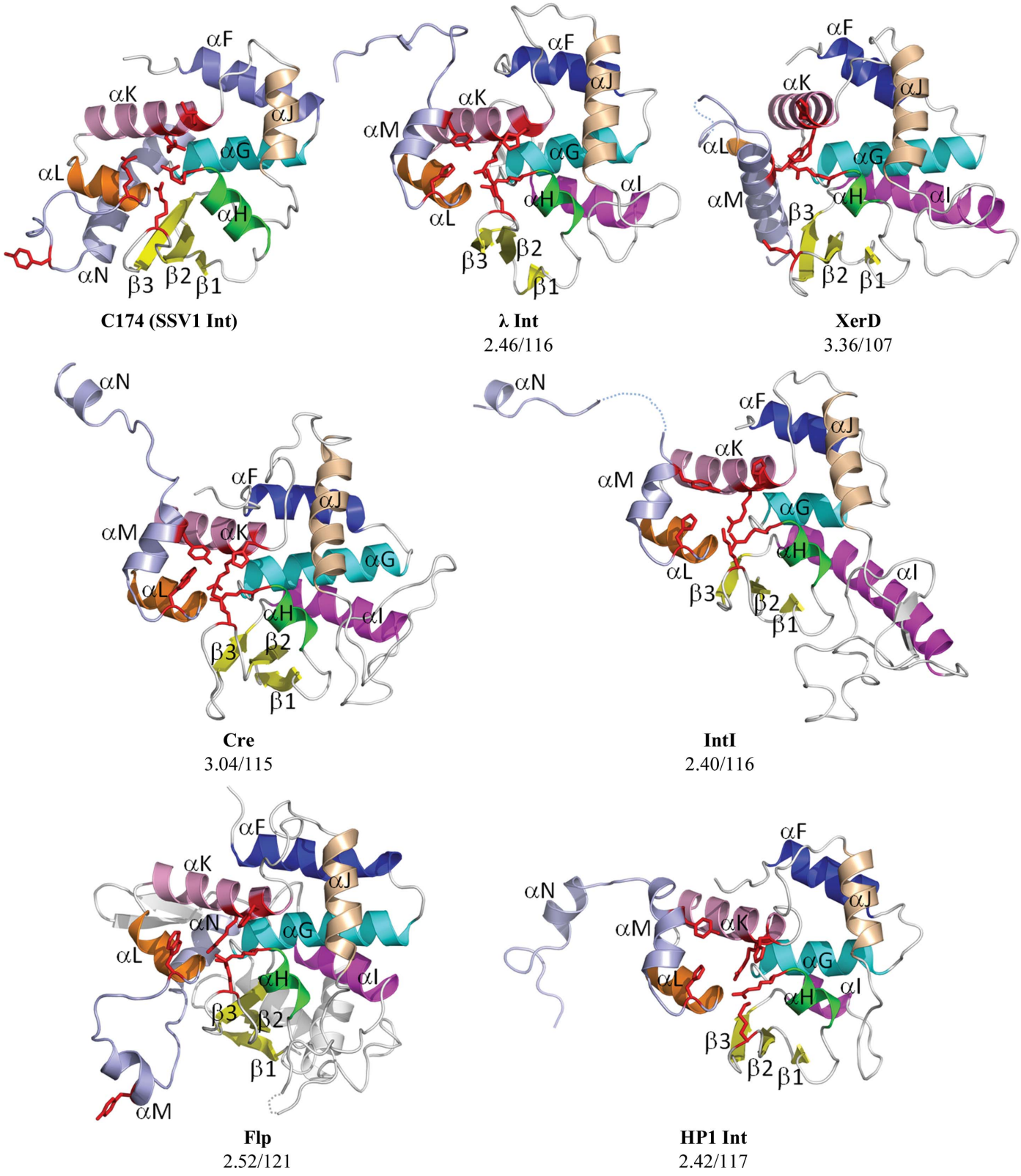

Figure 3
Structural similarity of the catalytic domains in crystallized tyrosine recombinases. Structural superposition was carried out using the *CCP*4*MG* program and the figures were produced with *PyMOL*. For clarity, each structure is shown separately. The conserved secondary-structure elements corresponding to those in Cre are coloured as follows: αF, blue; αG, cyan; αH, green; αI, magenta; αJ, wheat; αK, pink; αL, orange; αM and αN, light blue; β1–β3, yellow. Active-site residues are shown as red sticks, except for His270 in XerD, which is invisible. PDB codes: C174, 3vcf (chain *A*); λ Int, 1z1b (chain *A*; Biswas *et al.*, 2005); XerD, 1a0p (Subramanya *et al.*, 1997); Cre, 1crx (chain *B*; Guo *et al.*, 1997); IntI, 2a3v (chain *A*; MacDonald *et al.*, 2006); Flp, 1flo (chain *C*; Chen *et al.*, 2000); HP1, 1aih (chain *B*; Hickman *et al.*, 1997). The r.m.s.d. values (Å)/numbers of C^{α} atoms included in the superposition between C174 and each of the other tyrosine recombinases are indicated below the names of the corresponding proteins.

3.2. The structure of C174

The C174 domain of SSV1 Int is a compact α/β-fold comprising mainly of seven α-helices, with a three-stranded antiparallel β-sheet along one edge (Fig. 2*a*). The secondary-structural elements were numbered according to the structure superposition with Cre (Figs. 2*b* and 3). A disulfide bond bridges the C-terminus of β1 and the N-terminus of β2, a feature that has not been observed in other tyrosine recombinases (Fig. 2*a*). A similarly placed disulfide bond appears to exist in some of the SSV-type integrases that are closely related to SSV1 Int, as revealed by amino-acid sequence alignment (Supplementary Fig. S2). A highly positively charged surface forms a cleft circumscribed by αH, αK, αL and the β2–β3 hairpin. The conserved catalytic pentad is located in the cleft (Figs. 2*c* and 2*d*).

Superposition studies reveal that the fold of C174 is similar to those of the catalytic domains in all seven structurally determined tyrosine recombinases of both bacterial and eukaryotic origin (Fig. 3). Despite their extremely low sequence similarity, the structural arrangement of helices αF–αL, including the three-stranded β-sheet, in C174 bears a striking resemblance to those in the other known proteins. The remarkable conservation of the catalytic core structure in tyrosine recombinases from all three domains of life suggests the presence of a shared catalytic mechanism. However, C174 lacks a helix equivalent to αI in Cre. This helix is conserved in tyrosine recombinases from Bacteria and Eukarya. It is possible that αI serves an architectural role and is dispensable, since it is neither located in the vicinity of the target DNA in synaptic complexes nor involved in protein–protein interactions between two neighbouring recombinase monomers. The C-terminal portion of C174, which is equivalent to helices αM and αN in Cre and contains the nucleophile Tyr, differs in secondary structure and orientation from those in other tyrosine recombinases (Fig. 3). The nucleophile Tyr-containing helix αM is close to the active-site cleft in λ Int, XerD, Cre, IntI and HP1 Int, whereas helix αN (in Cre, IntI and HP1 Int) or a short β-strand (in λ Int complexed with DNA) extends from the main body in these proteins, with the exception of XerD, which has a long αM helix but lacks helix αN. In addition, helix αM stretches out and helix αN folds back in Flp. In comparison, in C174 a long and flexible loop containing the nucleophile Tyr314, which is equivalent to helix αM in Flp, extends far away from the active-site cleft and is followed by a folded-back helix.

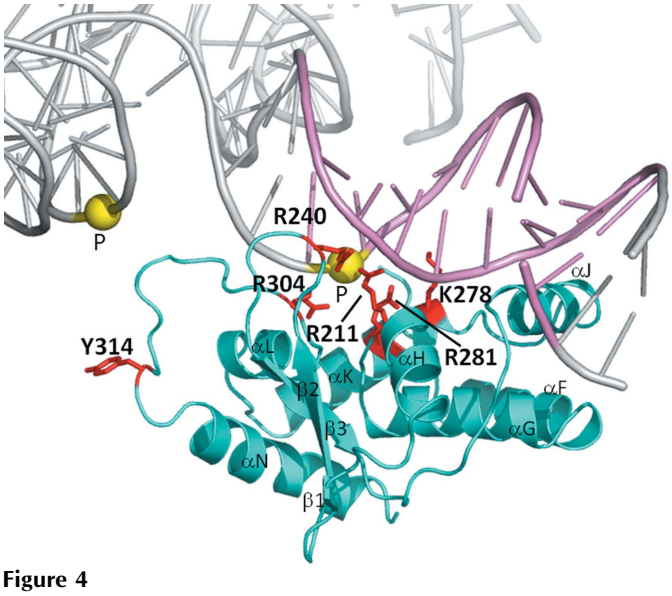

Figure 4
A model for the C174 domain of SSV1 Int interacting with DNA. Since superpositions of C174 with λ Int–DNA, Cre–DNA and Flp–DNA complexes yielded similar results, a model based on superposition with the λ Int–DNA complex (PDB entry 1z1g; Biswas *et al.*, 2005) is shown. C174 is shown in cyan, with active-site residues indicated as red sticks. The Holliday junction is shown in grey and the area predicted to be bound by C174 is shown in pink. The scissile phosphates are shown as yellow spheres.

3.3. The active site and DNA-binding region

As revealed by amino-acid sequence alignment, the nucleophile Tyr314 and the pentad Arg211, Arg240, Lys278, Arg281 and Arg304 in SSV1 Int are likely to be directly involved in catalysis. Three residues, *i.e.* Arg240 (instead of Lys243), Lys278 and Arg304, in the pentad differ from the consensus residues of the tyrosine recombinase family (R-K-H*xx*R-H/W-Y). Mutation of these residues has been shown to result in deficiency in DNA cleavage and relaxation without affecting DNA binding (Letzelter *et al.*, 2004). In C174, the pentad clusters around the basic cleft and is roughly oriented towards the centre of the cleft (Figs. 2*c* and 2*d*). As found in the other crystallized tyrosine recombinases, Arg211 is located at the N-terminus of αH, Arg240 in the middle of loop β2–β3, Lys278 and Arg281 on the same face of helix αK and Arg304 close to the C-terminus of αL in C174. Arg240 instead of Lys243 appears to correspond to the consensus residue Lys in the other tyrosine recombinases, judging from their respective distances from the centre of the cleft (Supplementary Fig. S3).

Superposition of the structure of C174 with those of other tyrosine recombinases in complex with DNA allows the identification of potential DNA contacts of the C174 domain (Fig. 4). It appears that the basic cleft on the surface of C174 binds DNA through electrostatic interactions. The catalytic pentad in the cleft is properly oriented towards the scissile phosphate. Helix αJ inserts into the major groove near the scissile phosphate. Loop β2–β3 containing Arg240 lies in the minor groove. Additionally, based on the model of the C174–DNA complex, the C174 domain is predicted to bind approximately ten base pairs of DNA, which is in agreement with the results of our EMSA assays (§3.5; Figs. 4 and 6*a* and Supplementary Fig. S4).

Intriguingly, inspection of the C174 domain structure reveals that the nucleophile Tyr314, which is located on a long and flexible loop, is at a great distance from the centre of the cleft surrounded by the catalytic pentad (Figs. 2*d* and 4). How does such a distally located nucleophile attack the scissile phosphate? Two possibilities exist. Firstly, if SSV1 Int cleaves the target DNA *in trans*, as in the case for the eukaryotic tyrosine recombinase Flp, the flexible Tyr314-containing loop

may require only minor movement to attack the scissile phosphate coordinated by the catalytic pentad of an adjacent recombinase monomer. In this scenario, helix αN folds back, stabilizing the stretched-out long loop containing Tyr314, as in Flp. Alternatively, if SSV1 Int cleaves the DNA *in cis*, as demonstrated for all characterized bacterial tyrosine recombinases, binding of the DNA may trigger inward movement of the Tyr314-bearing loop towards the other active residues of the same protomer, permitting Tyr14 to attack the scissile phosphate nearby. Cleavage in this mode probably requires large rearrangement of the C-terminal helix αN, which stretches out in almost all *cis*-cleavage complex structures to interact with the neighbouring monomer and enforce half-of-the-sites activity (Fig. 3; Biswas *et al.*, 2005; Guo *et al.*, 1997; MacDonald *et al.*, 2006; Grindley *et al.*, 2006). This 'induced-fit' phenomenon has been observed in λ Int (Kwon *et al.*, 1997; Aihara *et al.*, 2003). The *trans*-cleavage mode is supported by a previous *in vitro* complementation study, which showed that mixing an inactive SSV1 Int Y314F mutant with other inactive SSV1 Int mutants with mutations at active-site residues other than Tyr314 restored the relaxation and cleavage activities (Letzelter *et al.*, 2004). Therefore, it is highly possible that SSV1 Int cleaves the target DNA *in trans*. However, complete elimination of the possibility that SSV1 Int adopts the *cis*-cleavage mode awaits determination of the structure of a functional synaptic complex of SSV1 Int.

As found in all of the structures of tyrosine recombinase–DNA complexes, the nucleophile Tyr-containing region in an attacking protomer folds into an α-helix whether it cleaves DNA *in cis* or *in trans* (Fig. 3). Secondary-structure prediction shows that part of the Tyr314 loop (residues TQHYVS) has a strong potential to form an α-helix (data not shown). Presumably, binding of SSV1 Int to DNA triggers refolding of the Tyr314-containing region into an α-helix, orienting Tyr314 in the attacking protomer precisely towards the scissile phosphate in a *trans*-cleavage or *cis*-cleavage mode.

3.4. Oligomeric forms in solution

Protein–protein interactions play an important role in the formation of a synaptic complex between a site-specific recombinase and its target DNA sequence and in the regulation of the ensuing recombination process. To determine the ability of SSV1 Int to form oligomers, we subjected the full-length and truncated forms of the protein to gel filtration and blue native PAGE. SSV1 Int, N173 and C174 all eluted as single peaks on a Superdex 200 column, with estimated molecular masses of 83, 55 and 11 kDa, respectively (Fig. 5*a*). Based on the theoretical molecular masses of SSV1 Int (38.9 kDa), N173 (19.8 kDa) and C174 (19.1 kDa), we suggest that SSV1 Int and N173 exist as dimers and C174 as a monomer in solution. It appears that SSV1 Int dimerizes primarily through interactions in the N173 domain between two monomers. When N173 and C174 were mixed at a 1:1 molar ratio and loaded onto a Sephadex column, two peaks corresponding to those of the truncated proteins were obtained, ruling out the presence of stable physical interactions between the N173 and C174 domains.

The ability of SSV1 Int and the two truncated mutant proteins to form oligomers was also examined using blue native PAGE. As shown in Fig. 5(*b*), all three proteins appeared to oligomerize under the experimental conditions. Both full-length SSV1 Int and N173 preferentially formed even-numbered oligomers, suggesting that they readily dimerize in solution. Surprisingly, C174 formed a ladder of oligomers. An equimolar mixture of N173 and C174 resulted in the formation of oligomers that were unobserved with N173 or C174 alone, in addition to oligomers that were identical to those generated with either truncated protein. The bands that formed only in the presence of the two truncated mutants may indicate an interaction between the N- and C-terminal domains of the protein. The discrepancy between the gel-filtration results and the blue native PAGE results

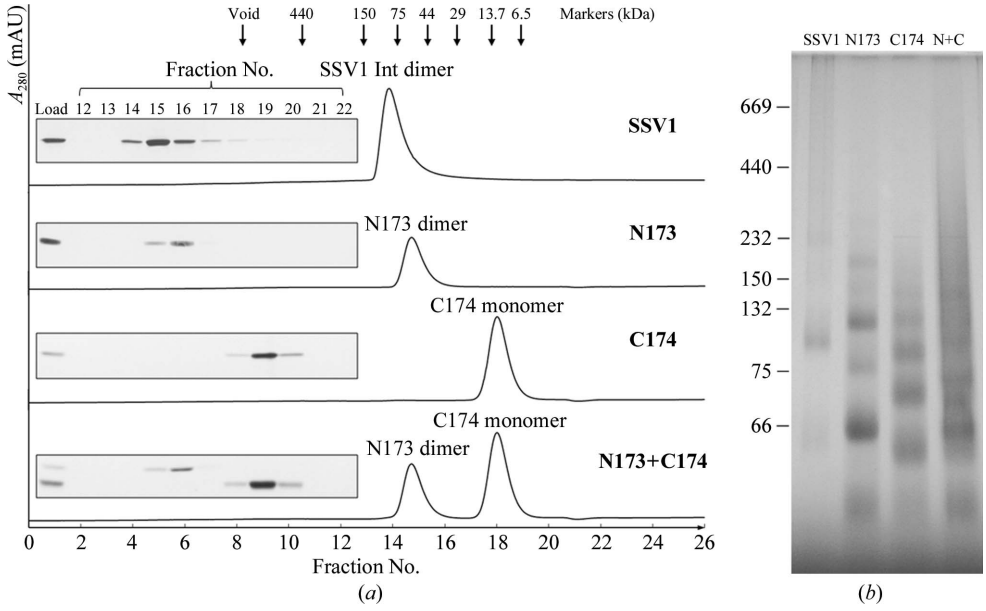

Figure 5
Oligomeric forms of SSV1 Int and the two truncated mutants. (*a*) Gel-filtration profiles. A sample (100 µl) of the indicated protein was loaded onto a Superdex 200 column (10/300 GL; GE Healthcare) and 1 ml fractions were collected. The void volume of the column and protein size markers are indicated by arrows at the top. Silver-stained SDS–PAGE gels of the fractions are shown as inserts. The oligomeric forms of the proteins, as deduced according to the calibration curve, are indicated. N173+C174 corresponds to an equimolar mixture of N173 and C174. (*b*) Blue native PAGE. Samples (20 mg) of indicated proteins were loaded onto a 6–20% gradient blue native PAGE gel. N+C corresponds to an equimolar mixture of N173 and C174. Molecular-mass markers in kDa are indicated on the left.

presumably arose from the difference in protein concentration between the two assays. Since the protein samples experienced far greater dilution during gel filtration on the Superdex column than during electrophoresis in blue native gel, they presumably existed at lower concentrations under the former conditions than under the latter conditions, given the amounts of protein used in the assays. Taken together, our data indicate that SSV1 Int dimerization involves strong protein–protein interactions between the N-terminal domains of protomers, whereas interactions between C174 molecules and between C174 and N173 were weak and were only detectable at high protein concentrations.

3.5. DNA binding and sequence specificity

DNA binding by full-length SSV1 Int and the truncated mutant proteins was studied by EMSA using a 58 bp double-stranded *att*A DNA fragment containing the core-type site (AOA′) as the labelled probe (Fig. 6*a*). The core-type site consisted of a 7 bp overlap region (O) and a pair of 10 bp imperfect inverted repeats A and A′, which were predicted to be bound by two molecules of SSV1 Int (Serre *et al.*, 2002; Letzelter *et al.*, 2004). In our preliminary experiments we tested a range of assay conditions, including temperature (277, 293 or 338 K) and voltage of electrophoresis and the NaCl concentration in the binding buffer. We found that the gel quality was significantly affected by the concentration of NaCl in the binding buffer, but not by the temperature or the voltage of electrophoresis (data not shown). Protein–DNA bands were better resolved at lower NaCl concentrations (Fig. 6*a* and Supplementary Fig. S4B). Notably, the

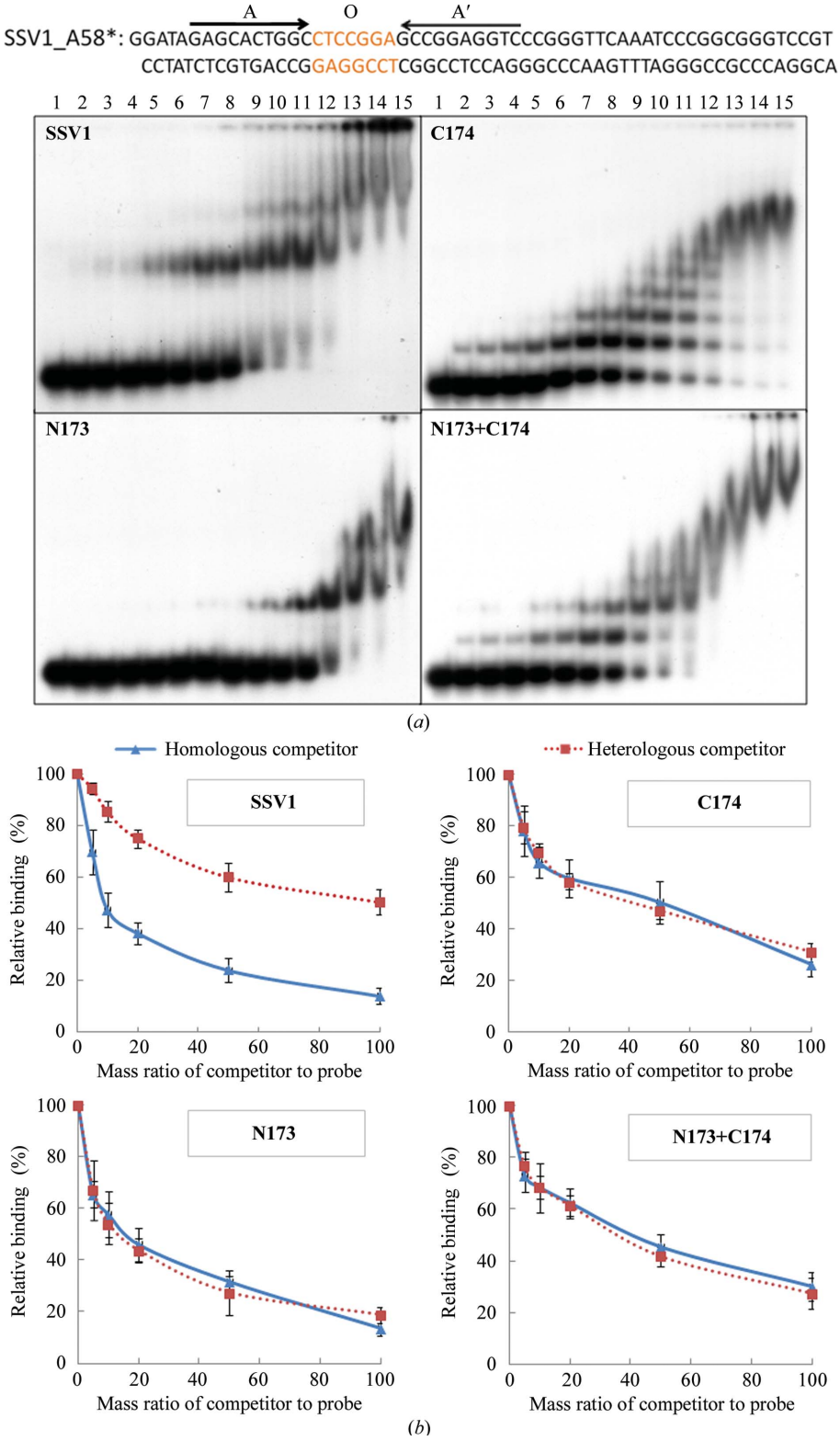

Figure 6
DNA binding and sequence specificity of SSV1 Int and its truncated mutants. (*a*) Binding of SSV1 Int and its truncated mutants to SSV1_A58 DNA. The binding reaction was carried out in the absence of NaCl. The sequence of the DNA fragment is shown at the top. The 10 bp imperfect inverted repeats (A and A′) flanking the 7 bp overlap region (O) are indicated by horizontal arrows. The protein concentrations in lanes 1–15 were 0, 2, 4, 6, 9, 12, 18, 25, 50, 75, 100, 200, 300, 400 and 500 n*M*, respectively. (*b*) Sequence specificity of the binding of SSV1 Int and its truncated mutants to SSV1_A58 DNA. Competitive EMSA assays were carried out in the presence of increased amounts of unlabelled 58-mer (homologous competitor) or poly(dI–dC)·poly(dI–dC) (heterologous competitor). The mass ratios of competitor to the radiolabelled fragment were 0, 5, 10, 20, 50 and 100, respectively. The data represent an average of at least three independent measurements. Error bars indicate standard deviations.

gel-retardation patterns were similar within the range of NaCl concentrations tested (0–150 m*M* NaCl). In the absence of competitor DNA, the full-length SSV1 Int and the two truncated mutants bound to the 58-mer with similar affinities (K_D = 1–5 × 10^{-8} M), as estimated based on the amount of protein required to retard one half of the labelled probe, when no NaCl was added to the mixtures (Fig. 6*a*). K_D values of 0.5–2 × 10^{-7} *M* (Supplementary Fig. S4B) and 2–4 × 10^{-7} *M* (data not shown) for the binding of the proteins to the 58-mer were obtained in the presence of 50 and 150 m*M* NaCl, respectively. The salt sensitivity of DNA binding by SSV1 Int and its domains is consistent with the electrostatic nature of the interactions between the DNA and the basic cleft on the surface of C174. Furthermore, C174 appeared to display a slightly higher affinity for the DNA than N173.

Interestingly, the full-length and the truncated proteins displayed different DNA-retardation patterns (Fig. 6*a*). SSV1 Int and N173 generated two retarded bands, whereas C174 formed up to six bands on the 58-mer as the protein concentration increased. Since both N173 and C174 are about half of the size of the full-length SSV1 Int, and as SSV1 Int and N173 existed stably as dimers and C174 as a monomer in solution at relatively low protein concentrations, we speculated that SSV1 Int and N173 bound to the 58-mer as dimers and C174 as monomers in a side-by-side fashion. In other words, the maximum number of retarded bands depends on the size of the DNA fragment. When a 118 bp *att*A fragment was used as a probe, the numbers of shifts were doubled (Supplementary Fig. S4A). Based on the EMSA patterns obtained on DNA fragments of different lengths, we inferred that one dimer of SSV1 Int or N173 binds 27–29 base pairs of DNA and one monomer of C174 binds ∼10 base pairs of DNA. The estimate for the binding size of C174 agrees with the prediction based on structural modelling (Fig. 4). Addition of both N173 and C174 to the binding reaction did not restore the binding pattern typical of the full-length protein (Fig. 6*a*). Instead, the two truncated proteins appeared to bind the DNA independently since the binding pattern resembled that typical of C174 alone at lower protein concentrations, in agreement with the higher affinity of C174 than that of N173 for DNA,

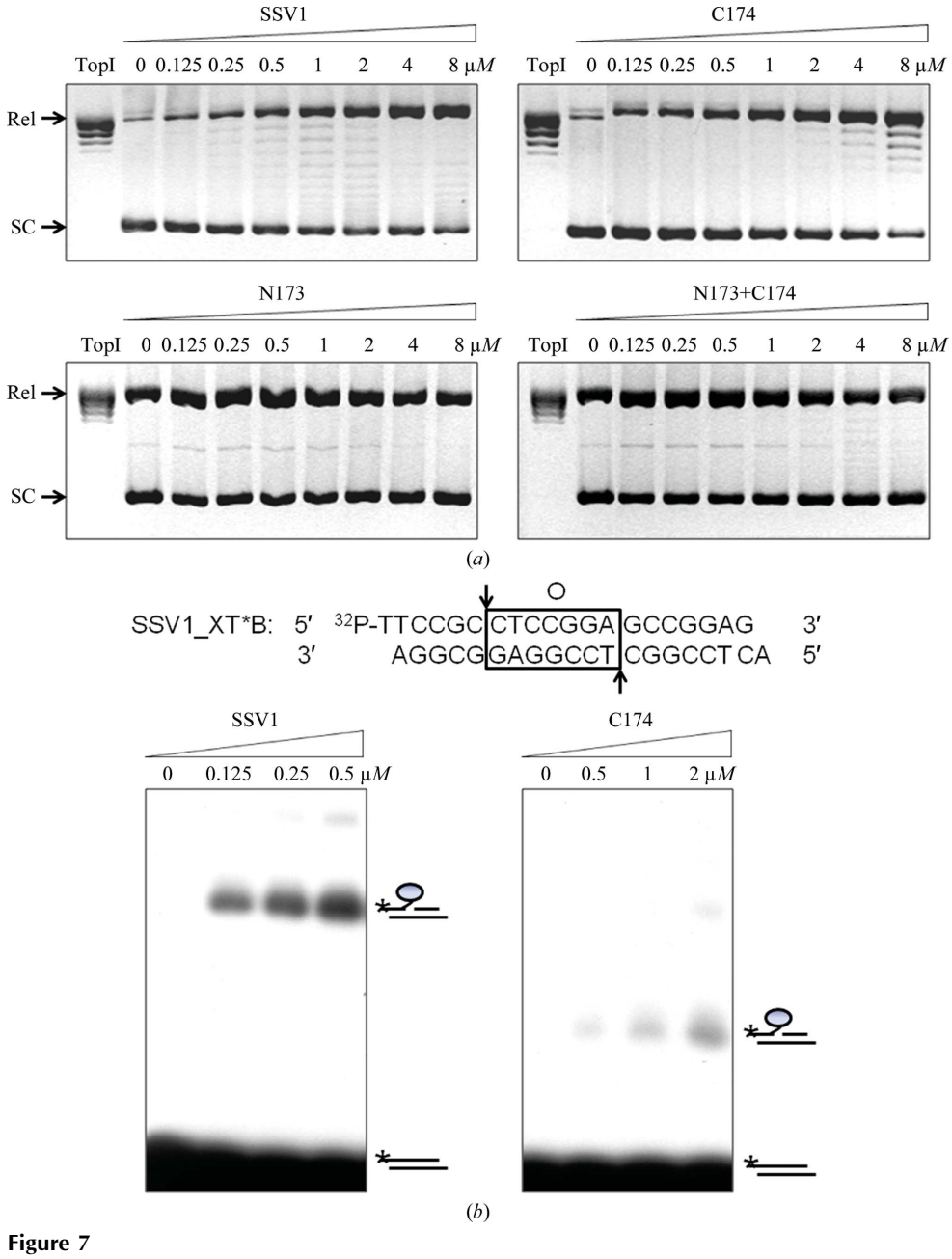

Figure 7
Activities of the C174 domain. (*a*) Topoisomerase activity assays. Supercoiled plasmid pBR322 (0.6 µg) was incubated for 4 h at 338 K with an increasing amount of the indicated protein. The samples were treated with SDS and protease K. An aliquot (0.3 µg pBR322) of each sample was electrophoresed in 1.2% agarose. As a positive control, plasmid treated with calf thymus topoisomerase I (TopI) was loaded onto the gel. N173+C174 indicates equimolar ratios of N173 and C174. Negatively supercoiled (SC) and relaxed (Rel) plasmids are indicated by arrows. (*b*) DNA cleavage assays. 5′-End-labelled DNA substrate (12 n*M*) was incubated for 4 h at 338 K with SSV1 Int or C174 domain. Samples were subjected to 12% SDS–PAGE and the gel was visualized by autoradiography. The overlap region (O) is boxed and the cleavage positions are indicated by arrows. The cartoons on the right indicate the position of the free probe and retarded covalent complexes.

and switched to a pattern that probably represents binding of both proteins at higher protein concentrations.

To examine the contribution of the N-terminal and C-terminal domains to the sequence specificity of binding of SSV1 Int to DNA, we performed competitive binding assays. Binding of the full-length and truncated proteins to the radiolabelled 58-mer was analyzed in the presence of increasing amounts of unlabelled 58-mer or poly(dI–dC)·poly(dI–dC) (Fig. 6*b*). Surprisingly, although SSV1 Int showed significant binding specificity, neither N173 nor C174 did. Furthermore, the binding specificity of the full-length protein was not restored by mixing N173 and C174. Therefore, these results suggest that the DNA-binding specificity of SSV1 Int depends on covalent linkage between its two domains. In comparison, the ability of several well studied tyrosine recombinases to recognize and bind to a specific core-type site is mainly provided by either the N-terminal or the C-terminal domain. For instance, the C-terminal domain of λ Int exhibits extremely low affinity for DNA and the central domain (corresponding to the N-terminal domain in two-domain tyrosine recombinases) is responsible for the recognition and binding of core-type DNA (Tirumalai *et al.*, 1997, 1998). The opposite was observed in Cre and Flp (Hoess *et al.*, 1990; Panigrahi & Sadowski, 1994).

3.6. Activities of C174

Many tyrosine recombinases show topoisomerase activity, probably because they are capable of strand cleavage and ligation, which are the key steps in DNA relaxation by topoisomerases (Grindley *et al.*, 2006; Sherratt & Wigley, 1998; Cheng *et al.*, 1998). As revealed by amino-acid sequence alignment and structural analysis, C174 possesses all of the catalytic residues. To determine whether C174 was active, we measured the DNA topoisomerase and cleavage activities of the protein. The full-length SSV1 Int has been shown to act as a type IB topoisomerase independent of a specific DNA site (Letzelter *et al.*, 2004). In the present study, we measured the relaxation activity of C174 on negatively supercoiled plasmid pBR322. As shown in Fig. 7(*a*), both SSV1 Int and C174 were able to relax the plasmid, whereas N173 was not. The full-length protein was around tenfold more active than the C-terminal domain, as found with λ Int (Tirumalai *et al.*, 1997). Surprisingly, the addition of N173 inhibited the topoisomerase activity of C174 (Fig. 7*a*). The mechanism for the inhibition is unclear, but it may be related to the reduced binding of C174 to the plasmid as a result of binding competition from N173 under the assay conditions.

A crucial step in DNA recombination catalyzed by tyrosine recombinase is the cleavage of single-stranded DNA and the generation of a covalently linked recombinase–DNA complex detectable by SDS–PAGE. In this study, covalent cleavage of the substrate DNA by SSV1 Int, C174 and N173 was examined as described previously (Serre *et al.*, 2002). We found that C174 was active in cleaving the target DNA (Fig. 7*b*). However, the cleavage activity of C174 was much lower than that of SSV1 Int and was not significantly affected by N173 (data not shown).

Based on these results, we conclude that the C174 domain is responsible for the catalysis of the full-length SSV1 integrase in DNA cleavage and ligation. It appears that covalent linkage between the N-terminal and C-terminal domains of the protein permits effective inter-domain communication, which is presumably essential for sequence-specific binding of the core-type site and efficient catalysis by SSV1 integrase.

We are grateful to Arthur Landy, Georgi Muskhelishvili and Marie-Claude Serre for kind suggestions and to Neil Shaw for proofreading the manuscript. This work was supported by grants 30730003, 30921065 and 31070660 from the National Natural Science Foundation of China.

References

Abremski, K. E. & Hoess, R. H. (1992). *Protein Eng.* **5**, 87–91.
Aihara, H., Kwon, H. J., Nunes-Düby, S. E., Landy, A. & Ellenberger, T. (2003). *Mol. Cell*, **12**, 187–198.
Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V. L., Pierson, L. S. III, Sternberg, N. & Leong, J. M. (1986). *EMBO J.* **5**, 433–440.
Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. & Ellenberger, T. (2005). *Nature (London)*, **435**, 1059–1066.
Cao, Y. & Hayes, F. (1999). *J. Mol. Biol.* **289**, 517–527.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. (2000). *Mol. Cell*, **6**, 885–897.
Cheng, C., Kussie, P., Pavletich, N. & Shuman, S. (1998). *Cell*, **92**, 841–850.
DeLaBarre, B. & Brunger, A. T. (2003). *Nature Struct. Biol.* **10**, 856–863.
Delano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Deneke, J., Ziegelin, G., Lurz, R. & Lanka, E. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 7721–7726.
Dong, A., Xu, X. & Edwards, A. M. (2007). *Nature Methods*, **4**, 1019–1021.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Esposito, D. & Scocca, J. J. (1997). *Nucleic Acids Res.* **25**, 3605–3614.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Grindley, N. D., Whiteson, K. L. & Rice, P. A. (2006). *Annu. Rev. Biochem.* **75**, 567–605.
Guo, F., Gopaul, D. N. & van Duyne, G. D. (1997). *Nature (London)*, **389**, 40–46.
He, Y., Yao, D.-Q., Gu, Y.-X., Lin, Z.-J., Zheng, C.-D. & Fan, H.-F. (2007). *Acta Cryst.* D**63**, 793–799.
Hickman, A. B., Waninger, S., Scocca, J. J. & Dyda, F. (1997). *Cell*, **89**, 227–237.
Hoess, R., Abremski, K., Irwin, S., Kendall, M. & Mack, A. (1990). *J. Mol. Biol.* **216**, 873–882.
Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997). *Science*, **276**, 126–131.
Letzelter, C., Duguet, M. & Serre, M.-C. (2004). *J. Biol. Chem.* **279**, 28936–28944.
MacDonald, D., Demarre, G., Bouvier, M., Mazel, D. & Gopaul, D. N. (2006). *Nature (London)*, **440**, 1157–1162.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Muskhelishvili, G. (1994). *Syst. Appl. Microbiol.* **16**, 605–608.
Muskhelishvili, G., Palm, P. & Zillig, W. (1993). *Mol. Gen. Genet.* **237**, 334–342.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Panigrahi, G. B. & Sadowski, P. D. (1994). *J. Biol. Chem.* **269**, 10940–10945.
Rajeev, L., Malanowska, K. & Gardner, J. F. (2009). *Microbiol. Mol. Biol. Rev.* **73**, 300–309.
Schägger, H., Cramer, W. A. & von Jagow, G. (1994). *Anal. Biochem.* **217**, 220–230.
Serre, M.-C., Letzelter, C., Garel, J. R. & Duguet, M. (2002). *J. Biol. Chem.* **277**, 16758–16767.
Sherratt, D. J. & Wigley, D. B. (1998). *Cell*, **93**, 149–152.
Su, J., Li, Y., Shaw, N., Zhou, W., Zhang, M., Xu, H., Wang, B.-C. & Liu, Z.-J. (2010). *Protein Cell*, **1**, 453–458.
Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J. & Wigley, D. B. (1997). *EMBO J.* **16**, 5178–5187.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Tirumalai, R. S., Healey, E. & Landy, A. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 6104–6109.
Tirumalai, R. S., Kwon, H. J., Cardente, E. H., Ellenberger, T. & Landy, A. (1998). *J. Mol. Biol.* **279**, 513–527.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.