J. Mol. Biol. (1995) **253**, 333–346

JMB

High-resolution Structure of the Catalytic Domain of Avian Sarcoma Virus Integrase

Grzegorz Bujacz[1], Mariusz Jaskólski[1,2], Jerry Alexandratos[1] Alexander Wlodawer[1]*, George Merkel[3], Richard A. Katz[3] and Anna Marie Skalka[3]

[1]*Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center ABL-Basic Research Program Frederick, MD 21702, USA*

[2]*Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań Poland*

[3]*Institute for Cancer Research Fox Chase Cancer Center Philadelphia, PA 19110, USA*

**Corresponding author*

Retroviral integrase (IN) functions to insert retroviral DNA into the host cell chromosome in a highly coordinated manner. IN catalyzes two biochemically separable reactions: processing of the viral DNA ends and joining of these ends to the host DNA. Previous studies suggested that these two reactions are chemically similar and are carried out by a single active site that is characterized by a highly conserved constellation of carboxylate residues, the D,D(35)E motif. We report here the crystal structure of the isolated catalytic domain of avian sarcoma virus (ASV) IN, solved using multiwavelength anomalous diffraction data for a selenomethionine derivative and refined at 1.7 Å resolution. The protein is a crystallographic dimer with each monomer featuring a five-stranded mixed β-sheet region surrounded by five α-helices. Based on the general fold and the arrangement of catalytic carboxylate residues, it is apparent that ASV IN is a member of a superfamily of proteins that also includes two types of nucleases, RuvC and RNase H. The general fold and the dimer interface are similar to those of the analogous domain of HIV-1 IN, whose crystal structure has been determined at 2.5 Å resolution. However, the ASV IN structure is more complete in that all three critical carboxylic acids, Asp64, Asp121 and Glu157, are ordered. The ordered active site and the considerably higher resolution of the present structure are all important to an understanding of the mechanism of retroviral DNA integration, as well as for designing antiviral agents that may be effective against HIV.

Keywords: integrase; AIDS; X-ray structure; multiwavelength anomalous diffraction; structure comparison

Introduction

Retroviruses contain an RNA genome that is reverse transcribed into DNA upon infection of the host cell. The DNA copy becomes integrated into the host chromosome through a series of well-defined cutting and pasting reactions that are similar to those seen in other types of recombination and DNA transposition events. Since the integration step is not reversible, the viral DNA is passed to all descendants of the infected cell. The integration step is therefore a key target when considering therapeutic intervention for pathogenic human retroviruses such as HIV-1, which is the causative agent of AIDS.

Extensive genetic and biochemical studies have demonstrated that integration is catalyzed by a retroviral protein, the integrase (IN: for reviews, see Katz & Skalka, 1994; Goff, 1992; Whitcomb & Hughes, 1992). IN, as well as reverse transcriptase (RT), are encoded in the retroviral *pol* gene and both of these enzymes are assembled into the viral particle as part of the Gag-Pol precursor polypeptide. After budding from the host cell, mature RT and IN are produced by proteolytic processing carried out by a viral protease. Upon infection of a new host cell, RT copies the viral RNA genome to produce a blunt-ended, double-stranded DNA. The integration reaction then proceeds through two biochemically and temporally separable steps (Craigie *et al*., 1990; Katz *et al*., 1990). First, IN introduces nicks, usually two base-pairs from the 3'-ends of both viral strands (the ''processing'' reaction), immediately downstream of the highly conserved CA dinucleotides (Roth *et al*., 1989; Katzman *et al*., 1989). The second reaction (the ''joining'' reaction; Craigie *et al*., 1990; Katz *et al*., 1990) involves direct attack by the newly

Abbreviations used: IN, integrase; ASV, avian sarcoma virus; RT, reverse transcriptase; MAD, multiwavelength anomalous diffraction.

0022–2836/95/420333–14 $12.00/0

© 1995 Academic Press Limited

created 3'-OH groups of the viral DNA on host DNA phosphate groups (Engleman *et al*., 1991) that are offset by four to six base-pairs. This coupled cleavage-ligation reaction results in the covalent joining of the 3'-ends of both viral DNA strands to host DNA. The joining to offset phosphate groups on the two strands of host DNA produces single-stranded gaps in the insertion site, which are subsequently repaired to produce the characteristic target site duplication seen flanking retroviral insertions. The processing and joining reactions can be reproduced *in vitro* using short oligodeoxynucleotide duplex substrates that represent either end of the viral DNA; the only cofactor required is a metal ion, usually Mn^{2+} or Mg^{2+}. The current model for IN function states that: (1) IN acts as a multimer (Jones *et al*., 1992; Engelman *et al*.,1993; van Gent *et al*., 1993), likely to coordinate the insertion of the two viral ends at the host target DNA site, (2) IN recognizes specific sequence and structural features in the viral DNA ends, as well as nearly random target sequences in host DNA. (3) IN contains a single active site that carries out both the processing and joining reactions.

IN is composed of three domains that have been defined by amino acid alignments, amino acid replacement studies, and deletion mutagenesis (Figure 1A). The prototype avian sarcoma virus (ASV) IN is 286 amino acid residues in length. An N-terminal zinc finger-like domain, residues approximately 1 to 50, is characterized by an HHCC motif that is conserved also in retrotransposons. The C-terminal domain is less well conserved, but ASV IN (Khan *et al*., 1991; Mumm & Grandgenett, 1991), as well as HIV-1 IN (Woerner *et al*., 1992; Vink *et al*.,

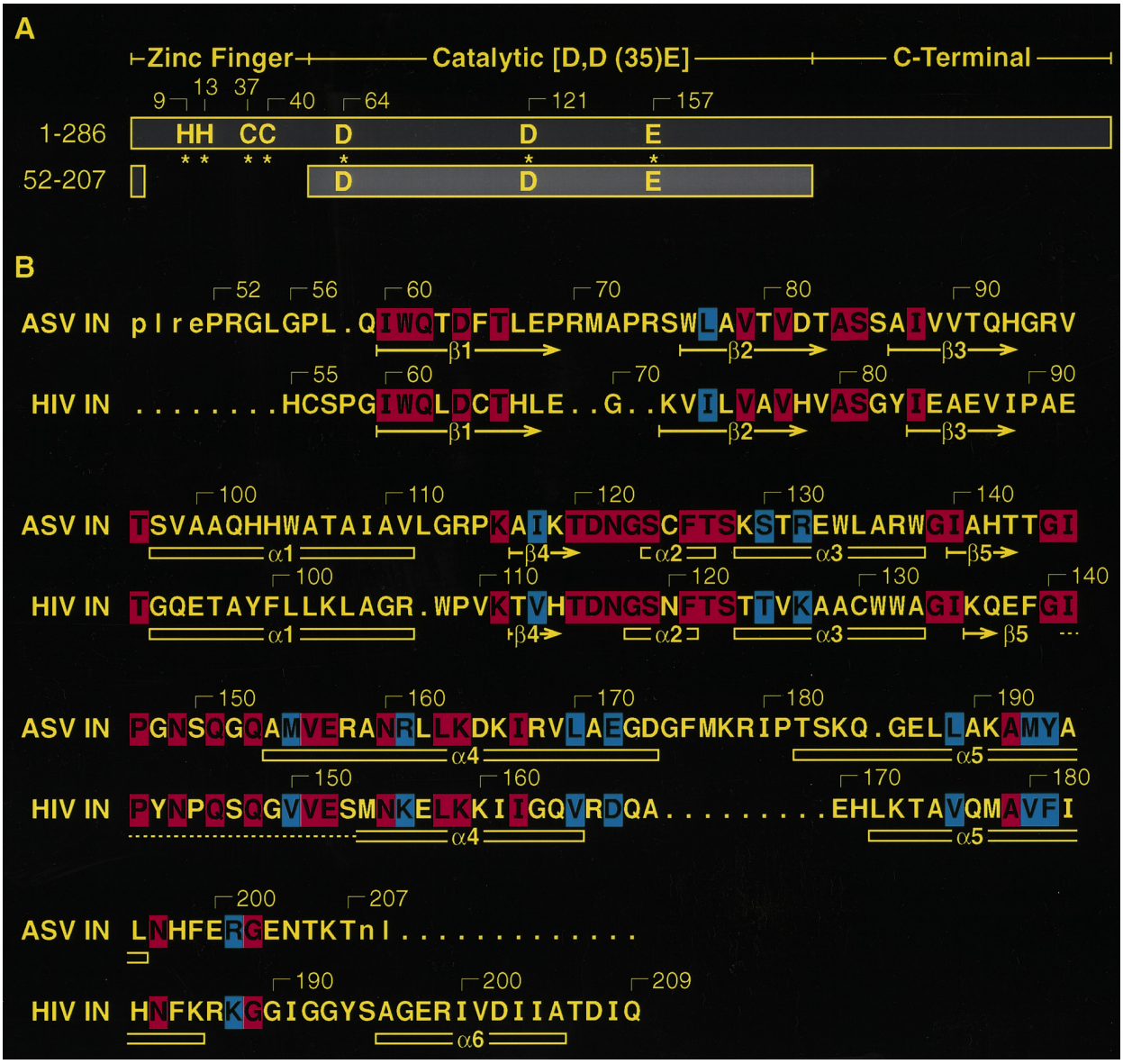


Figure 1. Primary structure of retroviral integrase. A, Representation of ASV IN showing the three functional domains and indicating the fragment included in the domain analyzed in this investigation. B, Alignment of the amino acid sequences in the ASV and HIV-1 IN catalytic domains, with the elements of secondary structure indicated. Residues shown in red are identical, those shown in blue are of similar type. The sequence with a broken underline was not observed in the crystal structure of HIV-1 IN.

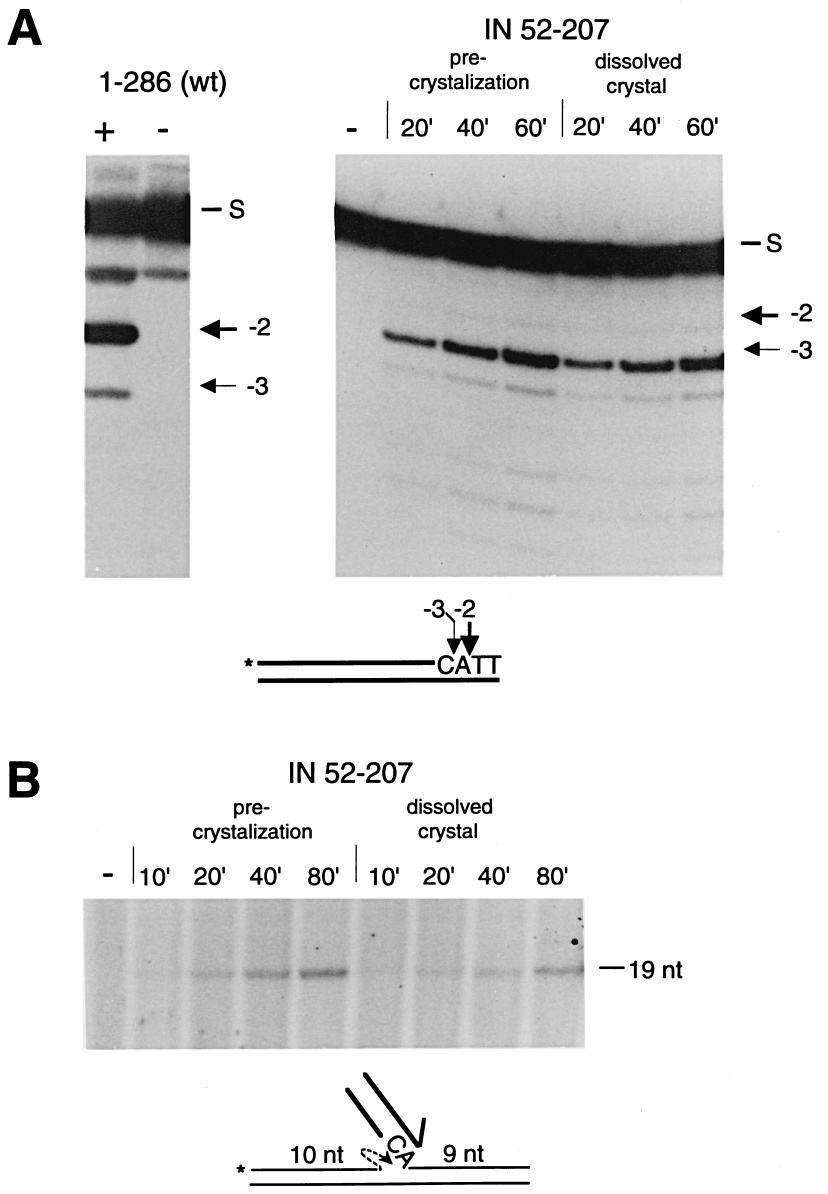


Figure 2. Activities of the 52-207 fragment from dissolved crystals. A, Left, processing activity of full-length (wt) IN on the substrate illustrated below (see Kulkosky *et al*. 1995). Substrate is an 18 base-pair duplex whose sequence corresponds to that of the U3 end of ASV DNA. The 5'-end of the plus strand of the substrate duplex is labeled with ^{32}P (asterisk). IN nicks primarily between the indicated A and T, and this corresponds to the processing activity (–2 nick). The site of the secondary (–3) cleavage is shown. Products were analyzed on a urea/polyacrylamide sequencing gel. The radioactively labeled, substrate (plus) strand is indicated (S), along with products derived from the indicated nicking sites. Right, the activity of the original IN fragment 52-207 fragment and of the protein from a dissolved crystal. Equal amounts of protein were used in these assays. Only the secondary nicking activity at the –3 site is retained by IN fragment 52-207 (Kulkosky *et al*. 1995). The – indicates incubation with no enzyme. B, Cleavage-ligation activity of pre-crystallization protein and protein from dissolved crystals, using equal amounts of the enzyme. This ''disintegration'' substrate diagramed below (Chow *et al*. 1992). Heavy lines denote viral sequences, thinner lines indicate target sequences and the asterisk indicates ^{32}P label. The IN fragment 52-207 catalyzes a reaction that releases the viral sequences and seals the target strand to produce a 19 nt product. The reaction conditions were similar to those described by Kulkosky *et al*. (1995). The reaction mixture contained 1 μM IN fragment 52-207, 8 μM substrate DNA, 50 mM Tris-HCl (pH 8.4), 200 μg/ml bovine serum albumin, 50 mM NaCl, 4% glycerol and 3 mM $MnCl_2$. The reactions were carried out at 30°C for the indicated times. The product strand was detected on a urea/polyacrylamide sequencing gel.

1993), contain a non-specific DNA binding activity in this region. The catalytic domain (residues ca 50 to 200) is defined by a constellation of carboxylate residues, the D,D(35)E motif (Asp64, Asp121, Glu157 in ASV). The conservation of the D,D(35)E array in retroviral and retrotransposon integrases, as well as in some bacterial transposases, suggested that these residues might be involved in the DNA processing and joining activities (Kulkosky *et al*., 1992, and references therein). Mutagenesis studies showed that the D,D(35)E carboxylate residues are all essential (Kulkosky, 1992; Engelman & Craigie, 1992). We have suggested that their general role is to coordinate the required metal ion(s) (Kulkosky *et al*., 1992), in a manner similar to analogous carboxylate residues in the active sites of RNase H (Davies *et al*., 1991) and the exonuclease domain of *E. coli* DNA polymerase I (Steitz, 1993).

Although the isolated catalytic D,D(35)E domain is defective in processing and joining activities (Figure 2), both the HIV-1 (Bushman *et al*., 1993; Vink *et al*., 1993) and ASV IN (Bushman & Wang, 1994; Kulkosky *et al*., 1995) domains display a coupled cleavage-ligation activity, referred to as ''DNA splicing'' or ''disintegration'', on specialized DNA substrates where a 3'-hydroxyl group is positioned adjacent to a target phosphate group (Figure 2B, diagram: and see Chow *et al*., 1992). The activity on these substrates apparently involves a direct nucleophilic attack by the 3'-hydroxyl group on the adjacent internal 5'-phosphate group, resulting in coupled cleavage-ligation. This reaction is chemically

similar to both the processing and the joining reactions. In processing, a water molecule is a nucleophile that attacks the viral DNA end, while in joining the new 3'-hydroxyl group is the nucleophile attacking the host DNA. It is generally believed that the IN active site prepares viral and host DNA phosphate groups for attack, since there is little selectivity for the attacking group *in vitro*. The specialized substrates, containing reactive groups that are juxtaposed, may reduce the need for the N and C-terminal IN domains, as the function of these domains may be to help guide the viral DNA ends and/or host DNA substrates into the required position(s). In addition to cleavage-ligation activity, the catalytic domain of ASV IN displays an intrinsic non-specific DNA nicking activity with preference for –3 cleavage, one nucleotide upstream of the processing site (Figure 2A, diagram: and see Kulkosky *et al*., 1995). Taken together, these observations indicate that the central domain includes the catalytic site and has some capacity for binding the required metal ion(s).

Here, we describe a high-resolution crystal structure of the ASV IN catalytic domain (residues 52 to 207). The general fold, as well as the juxtaposition of the conserved carboxylate residues, is similar to that found in nucleases, RNase H and RuvC, and in HIV-1 IN. In contrast to a recently reported structure of the catalytic domain of HIV-1 IN (Dyda *et al*., 1994), all three conserved carboxylate residues are ordered in the ASV IN structure. Finally, a comparison of the related nucleases and IN structures reveals important similarities, as well as some unexpected differences.

Results and Discussion

Preparation, activities, and crystallization of the catalytic domain of ASV IN

Many previous attempts in these laboratories to crystallize the intact, bacterially produced ASV IN have been unsuccessful. One reason for this may be that the C and/or N-terminal domains are flexible, reflecting their possible roles in aligning substrate DNAs for the joining reaction. Also, it is difficult to achieve high starting concentrations of the full-length protein due to its tendency to aggregate. We therefore expressed in bacteria several terminal-deleted versions of the protein that retained the catalytic domain (Kulkosky *et al*., 1995) and screened for high level of expression and increased solubility. The 52-207 fragment was expressed more efficiently and showed significantly increased solubility compared with the full-length protein. Large crystals of the catalytic domain were obtained at pH 7.5 using as precipitant either 20% (w/v) polyethylene glycol (PEG) 4000 or 2 M ammonium sulfate. The crystals were tetragonal, $P4_32_12$, with a = 66.5 Å, c = 81.0 Å, and contained one protein molecule in the asymmetric unit. They diffracted to high resolution and were stable in the X-ray beam.

Figure 2 shows that the two enzymatic activities of the 52-207 fragment could be recovered when the crystals were dissolved. Figure 2A compares the activities of the 52-207 fragment with the full-length protein using an 18-mer DNA duplex whose sequence corresponds to the ''U3'' viral DNA end. As has been shown (Katzman *et al*., 1989), the full-length IN nicks primarily following the conserved CA at the ''–2'' site (Figure 2A, left). The 52-207 domain produces a –3 nick in a reaction that is one of the activities of the full-length protein (Kulkosky *et al*., 1995), and this activity is maintained after crystals have been dissolved (Figure 2A, right). Figure 2B shows that the cleavage-ligation activity is retained also, as indicated by the 19-nucleotide product formed from the ''disintegration'' substrate.

Structure solution

The crystal structure of the ASV IN catalytic domain was solved initially at 2.2 Å resolution with phases calculated on the basis of multiwavelength anomalous diffraction (MAD) data for a selenomethionine (Se-Met) derivative, and this data set was subsequently refined at 1.95 Å (Table 1). The final electron density map was of superb quality, as illustrated in Figure 3A, which shows the area of the active site of the enzyme, with the superimposed electron density. The protein chain could be traced without interruptions, although the two N-terminal and eight C-terminal residues for which no density could be seen were considered disordered under the conditions of the experiment. Following the initial chain tracing, we refined the structure using several more data sets, all of them measured at room temperature with conventional CuKα X-ray sources. These included a native data set for an ammonium sulfate crystal extending to 1.7 Å resolution, a data set for a crystal grown from citrate buffer (without Hepes), and for a selenomethionine derivative prior to MAD experiments. The data collection statistics are summarized in Table 2, while the final refinement statistics are listed in Table 1. All structures are characterized by low R-factors and good geometry, indicating their high quality.

Table 1. Refinement statistics

	SER	SERT	AMS	CIT
Resolution (Å)	1.95	2.10	1.70	2.20
No. reflections	10,188	8479	15,858	7828
No. atoms				
Protein	1183	1172	1258	1120
Water	187	74	181	121
Hepes	15	15	15	0
Other	4	4	0	0
R-factor	0.139	0.148	0.154	0.143
R-free	0.206	0.227	0.200	0.231
r.m.s. deviations from ideality				
Bond (Å)	0.013	0.013	0.013	0.013
Angle distances (Å)	0.047	0.045	0.034	0.045
Planes (Å)	0.016	0.016	0.011	0.015
Chiral volumes (Å^3)	0.168	0.161	0.136	0.158

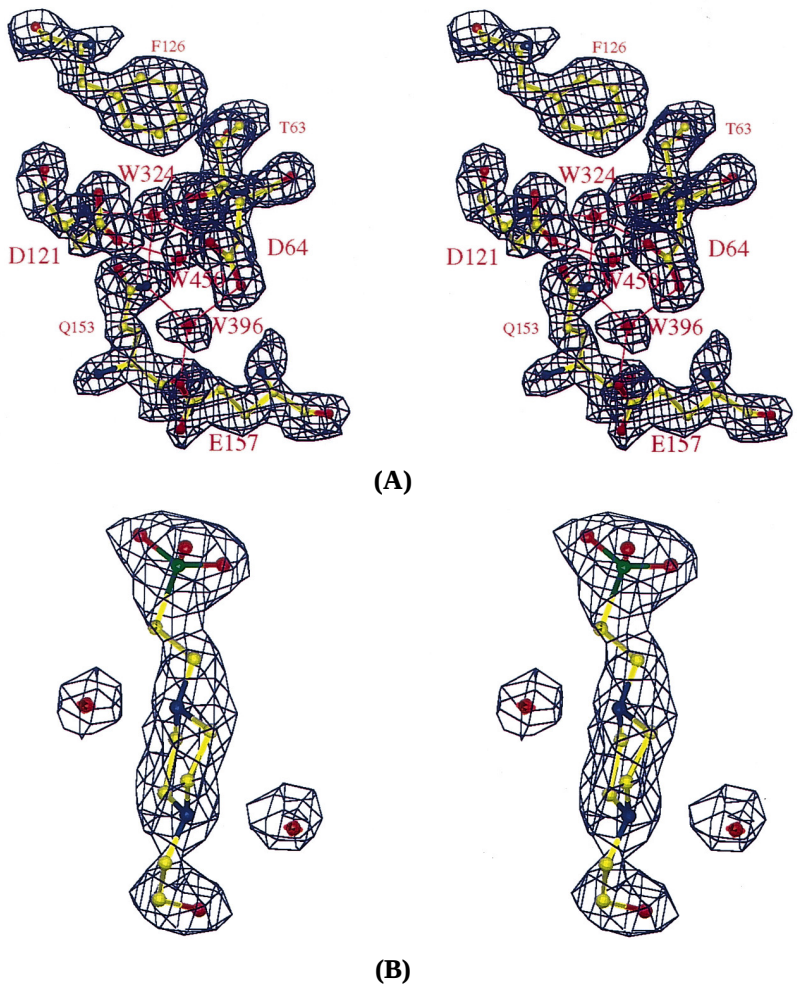


Figure 3. Stereoviews of electron density maps generated using O (Jones & Kjeldgaard, 1994). A, Active site of ASV IN in the low-temperature Se-Met structure, and the corresponding final $2F_o - F_c$ electron density map (contoured at the 1 σ level). Wat324 interacts with main-chain atoms, 121 N and 63 O, as well as with the side-chains of Asp64 and Gln153. Wat450 bridges the carboxylate groups of Asp64 and Asp121. Wat396 is a link between the DD and E components of the active site. B, The Hepes molecule and two associated water molecules from the refined low-temperature Se-Met model (SER) superimposed on the original 2.2 Å multiple isomorphous replacement map contoured at the 1.4 σ level. This map illustrates the excellent quality of the experimental phases.

Protein fold and dimer architecture

Topologically, the catalytic domain of ASV IN is a five-stranded mixed β-sheet flanked by five α-helices. The main part of the β-sheet, formed by three antiparallel β-strands, β1, β2 and β3, is extended by two short parallel β-strands, β4 and β5 (Figure 4A). On one side of the β-sheet there are three short, largely parallel helices, α1, α2 and α3. The very short helix α2 is only four residues long and is separated

Table 2. Summary of data collection and processing statistics

Protein buffer	Precipitant	Temp. (°C)	Ident.	$R(I)$[a]	Resol. (Å)	Compl. (%)	Compl. 20th shell (%)	No. refl.	No. refl. $F \geqslant 2\sigma F$	λ (Å)
Wild-type (Hepes)	PEG / IPR[b]	20	PEG	0.077	2.00	97.3	97.9	12,521	11,100	1.5418
Se-Met (Hepes)	PEG / IPR	20	SERT	0.093	2.10	97.0	97.2	10,722	9,344	1.5418
Se-Met (Hepes)	PEG / IPR	–165	SER	0.065	1.95[c]	83.7	55.0	12,157	11,427	0.9464 (remote)
Se-Met (Hepes)	PEG / IPR	–165	SEP	0.054[d]	2.20	91.3[d]	69.7[d]	15,452[d]	14,952[d]	0.9790 (peak)
Se-Met (Hepes)	PEG / IPR	–165	SEE	0.061	2.20	94.7	85.2	8,842	8,563	0.9792 (edge)
Wild-type (citrate)	PEG / IPR	20	CIT	0.062	2.20	99.0	99.1	9,620	8,729	1.5418
Wild-type (Hepes)	$(NH_4)_2SO_4$	20	AMS	0.052	1.70	92.4	95.2	19,011	16,929	1.5418

[a] $R(I) = \Sigma|I - \langle I \rangle| / \Sigma I$, Friedel pairs usually merged.
[b] PEG, polyethylene glycol; IPR, isopropanol.
[c] In the phase determination process (MAD), only data to 2.2 Å were used ($R(I) = 0.063$, completeness 94.2%).
[d] Friedel pairs kept separate.

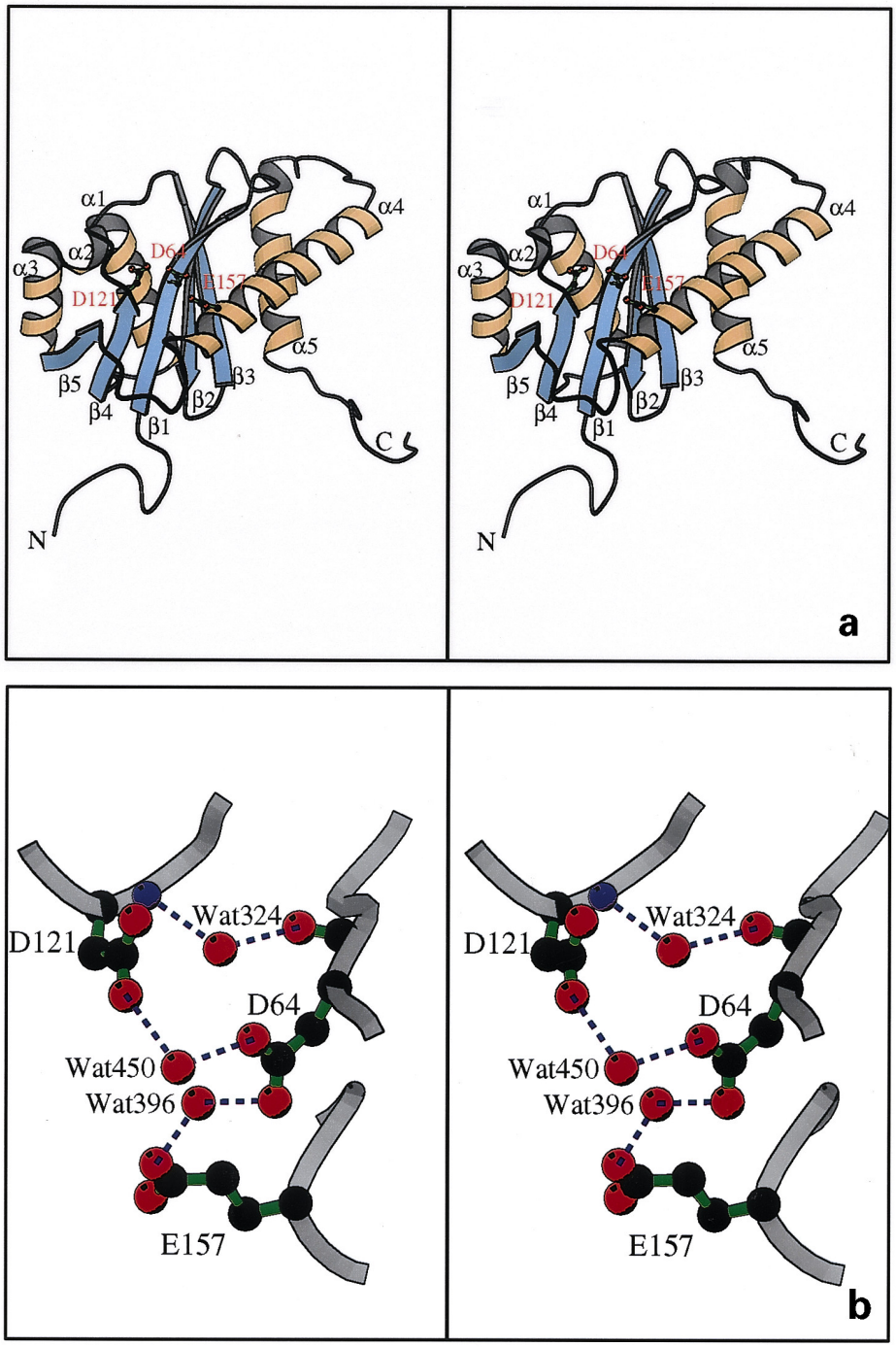


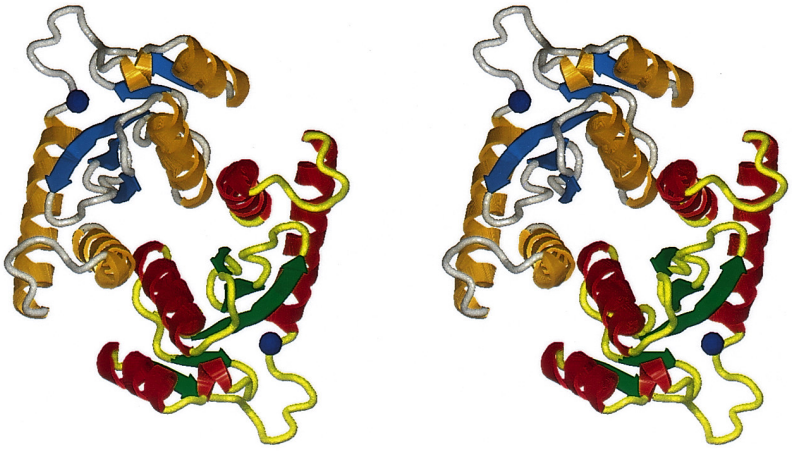
Figure 4. Stereoviews of the monomer, and the active site of ASV IN generated using MOLSCRIPT (Kraulis, 1991). a, A stereo view of the monomer of the ASV IN core domain, showing the secondary structure elements, as well as the location of the active site. b, Part of the active site of ASV IN, showing the side-chains of Asp121, Asp64 and Glu157. The water molecules present in the active site, and their hydrogen bonds to the residues above, are shown.

from α3 by just one residue (Ser128). The α3 helix is also separated from the β5 strand by a single residue (Gly139). The opposite side of the β-sheet is covered by two long antiparallel helices, α4 and α5. There is a long ten-residue loop between β5 and α4, which has somewhat different conformations in crystals grown from PEG and ammonium sulfate. There is no significant difference between the molecular models derived from crystals grown from PEG and ammonium sulfate outside of the tip of the β5-α4 loop. This loop extends out of the compact shape of the molecule and must be quite flexible, as evidenced by the high temperature factors at its tip. In the PEG structures this loop makes two turns, while in the ammonium sulfate structure it has one smooth turn, which is stabilized by interactions with the same loop from a symmetry-related molecule. The N and C termini are disordered in the PEG structures. Even in

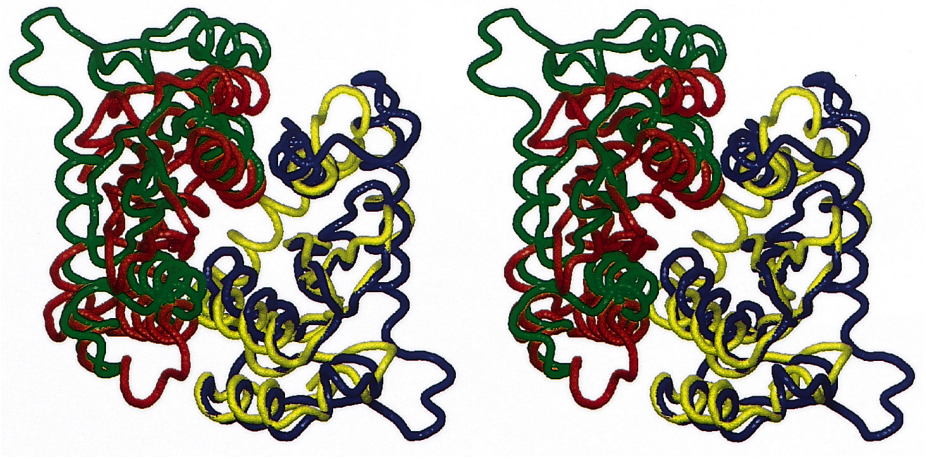
the low-temperature (SER) structure (Table 2), two N-terminal and eight C-terminal residues are missing from the model. The crystal grown from ammonium sulfate seems to have a more stable conformation of the termini. Supported by the quality and resolution (1.7 Å) of this data set, we modeled both ends completely. Very high *B*-factors and some breaks in electron density indicate disorder in these regions.

The dimer is created around the crystallographic 2-fold axis with an interface between the α1 helix in one monomer and the α5 helix in the complementary molecule (Figure 5A). The interactions between the helices are predominantly hydrophobic. Residues from the β3 strand, which is located deeper within the monomer, contribute to the middle part of the dimer interface. This location forms a cavity with some hydrophilic, positively charged residues, filled by a few water molecules. The dimer cavity has its smaller diameter of about 5 Å opening on both ends. An ion pair between Glu187 of one monomer and His103 of the other is found at the outer edges of the helices. Another intermolecular ion pair is made between Glu133 and Arg179. The dimer is further stabilized by several polar interactions, including hydrogen bonds between Glu187 and Trp134, as well as His198 and the main-chain carbonyl group of Ala110. Experimental evidence (unpublished results) indicates that the 52-207 ASV IN catalytic domain is capable of dimerization in solution although the apparent K_d of this fragment is significantly higher than that of the full-length protein.

Recent analyses (unpublished) have indicated that full-length ASV IN containing an S85G mutation is defective in dimerization. Previously, we showed that a variety of substitutions at this highly conserved position, including S85G, resulted in a severe reduction in both processing and joining activity (Katz *et al*., 1992). Only an S85T mutant retained the catalytic activity (although at a decreased level) and was not defective in viral replication. Ser85, and its



(A)



(B)

Figure 5. Stereoviews of the dimers of the catalytic domain of IN generated using O (Jones & Kjeldgaard, 1994). A, A dimer of the ASV IN core domain viewed along its 2-fold axis, with Wat450 positions marked as blue spheres. B, Superposition of the main chains of ASV and HIV-1 IN core domains. The superposition is based on the best fit between the C^{α} atoms of the blue (ASV) and gold (HIV-1) chains. The green (ASV) and red (HIV-1) subunits (generated by the corresponding molecular dyads) show considerable deviations.

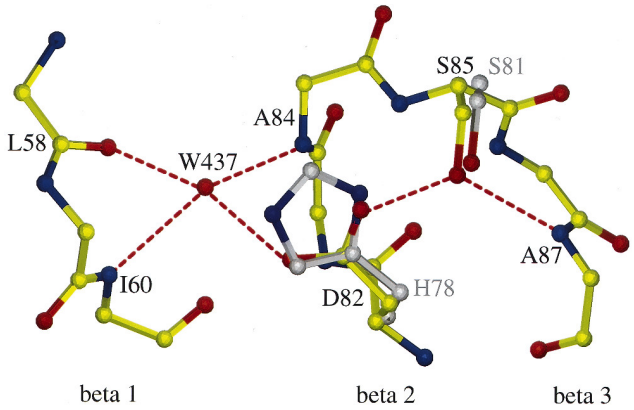


Figure 6. A fragment of the β-sheet that contains the conserved Ser85 residue. Only main-chain atoms, along with two key side-chains, Ser85 and Asp82, are shown. The hydrogen bonds (broken lines) formed by Ser-85 ''staple'' the tight β2/β3 turn near its tip. An identical system of hydrogen bonds is present in HIV-1 IN and involves the conserved serine, as well as a non-conserved His78 residue. The respective side-chains of the HIV-1 IN core domain (Dyda *et al*. 1994), shown in gray, correspond to an alignment of the two proteins based on all C^{α} atoms. A structurally conserved water molecule (Wat437) is positioned between β2 and β1, and mediates their interactions through a system of hydrogen bonds that also involves Asp82. The Figure was drawn with O (Jones & Kjeldgaard, 1994).

HIV-1 IN analog Ser81, are located at the turn between strands β2 and β3. Although in a solvent-accessible region, Ser85 has its hydroxyl group turned toward the β-sheet to a position where it forms two hydrogen bonds: with the main-chain N atom of Ala87 in β3 and with $O^{\delta 2}$ of the side-chain of Asp82 in β2 (Figure 6). These two hydrogen bonds, the latter of which is particularly strong (O . . . O, 2.4 Å), seem to be important for the integrity of the tight β2/β3 turn. Interestingly, while the Ser position and its main-chain contact are preserved in HIV-1 IN, the Asp residue is replaced by a very different His residue. However, the orientation of this His78 side-chain places its $N^{\delta 1}$ atom at the position of the $O^{\delta 2}$ of Asp82 in ASV IN, where it can form an analogous hydrogen bond with the conserved serine (Figure 6). The second carboxylate O atom of Asp82 ($O^{\delta 1}$) is involved in a network of hydrogen bonds that extend its stabilizing effect from β2 to β1. The key node of that network is Wat437, which forms two hydrogen bonds with main-chain atoms in β1, one hydrogen bond with a main-chain N atom in β2 (Ala84) and one hydrogen bond with $O^{\delta 1}$ of Asp82. (A similar, although not identical, hydrogen-bond scheme between β1 and β2 exists also in HIV-1 IN and includes a water molecule that has a structurally conserved position in both proteins.) The consequences of mutating Ser85 (to a residue other than threonine) could be devastating for the integrity of the β2/β3 turn and through freeing the Asp82 side-chain, for the cohesion of β1 in the β-sheet. Since the protein fold of integrase is organized around the dominating β-sheet, any changes in the mutual orientation of the individual β strands could lead to structural distortions that might cripple dimer formation. In summary, the highly conserved Ser85 residue plays a structural rather than catalytic role (Katz *et al*., 1992). As can be predicted from the structure, substitution of the adjacent Ser86 with alanine or glycine has little or no effect on activity (Katz *et al*., 1992).

The active site

The active site as observed in the low-temperature structure of ASV IN is shown in Figures 3A and 4B. All three acidic residues characteristic for this class of enzymes are seen clearly in the electron density. The side-chains of Asp64, Asp121 and Glu157 interact with each other through water molecules. Wat324 appears to be of particular importance, since it is present in all structures that we have investigated, has a relatively low temperature factor, and roughly tetrahedral coordination. This water molecule interacts directly with the protein in the vicinity of all three active site residues. In addition to interacting with the main-chain polar groups (63 O and 121 N), it is involved in hydrogen-bonded interactions with N^{ϵ} of Gln153, and with an O^{δ} of Asp64. The carboxylate groups of the active site residues interact through two other water molecules, Wat396 and Wat450. These water molecules are less tightly bound than Wat324, as evidenced by their higher temperature factors. We can expect that some of these molecules will be repositioned or replaced by divalent cations. The environment of the active site consists of a shallow and long depression on the surface, rather than a more distinct cavity observed in many other enzymes. This is, however, not unexpected, in view of the very large size of integrase substrates, which need to access the catalytic center. Even though the residues in the active site area are exposed to solvent, they are not involved in any way in intermolecular contacts. We can, therefore, be certain that the organization of the active site is not influenced by crystal packing.

Comparison with RNase H and RuvC

Integrase is a member of a family of enzymes which share certain functional aspects and topology (Dyda *et al*., 1994; Ariyoshi *et al*., 1994; Yang & Steitz, 1995). The prototypical member of this class is *Escherichia coli* RNase H and the closely related RNase H domain of HIV-1 RT (Katayanagi *et al*., 1990; Yang *et al*., 1990a; Davies *et al*., 1991). Another related enzyme is the Holliday junction-specific endonuclease RuvC (Ariyoshi *et al*., 1994). All of these enzymes are characterized by a similar α/β fold, and by a cluster of at least three highly conserved acidic residues in their active sites.

We have compared the structure of ASV IN with the structures of the enzymes mentioned above. For these purposes, we used the published structures of RNase H (*E. coli*, Katayanagi *et al*., 1993, PDB designation 1RDD; HIV-1, Davies *et al*., 1991, PDB designation 1HRH). For RuvC, we arbitrarily chose the coordinates of molecule 1 out of the four present in the file with PDB designation 1HRJ (Ariyoshi *et al*., 1994).

Superposition of the C^{α} coordinates of ASV IN and HIV-1 RNase H yields an r.m.s. deviation of 2.4 Å for 50 atom pairs, found in all five β strands, as well as in helices 1, 3 and 4. For the 49 atoms compared between ASV IN and *E. coli* RNase H, the deviation is 2.85 Å, with much poorer alignment particularly for helix 4, which has a considerably different angle with respect to the rest of the structure. As many as 63 atoms were used in the comparison with RuvC. In that case, there is a good match of atomic positions in helix 4, but only with the helices running in opposite directions. This is another indication that the topological and even structural similarity among these enzymes has some limitations.

As noticed by Yang & Steitz (1995), the similarity of the cluster of acidic residues forming the active sites of the nucleases is striking. With the alignment of the secondary structure elements as described above, the best agreement in the active site is for Asp64 (Table 3). In all cases, the placement and direction of this carboxylate residue is very similar. ASV IN Asp121 is close to its equivalent in HIV-1 RNase H and *E. coli* RNase H, while the side-chain of the equivalent residue in RuvC is much more distant. The agreement for the third residue of the cluster, ASV IN Glu157, is quite good, although the match in RuvC is poorer. The other acidic residues in this region (Table 3) do not have their counterparts in ASV IN.

The picture of the active sites of these enzymes emerging from these comparisons shows both similarities and differences. While the three most highly conserved acidic residues are present in similar locations in all of them, the exact relationships between them are not strictly preserved. We describe below significant differences in the orientation and relative location of these residues in the very closely related ASV and HIV-1 IN proteins. More detailed investigations will be necessary to understand the significance of these differences and delineate the range of variations to be expected in this class of enzymes.

Table 3. Structurally conserved residues in the active sites of nucleases with RNase H fold

IN	ASV	Asp64	Asp121	Glu157	—	Gln153
IN	HIV-1	Asp64	Asp116	Glu152	—	Gln148
RNase H	*E. coli*	Asp10	Asp70	Asp134	Glu48	Asn130
RNase H	HIV-1	Asp43	Asp98	Asp149	Glu78	Asn145
RuvC	*E. coli*	Asp7	Glu66	Asp141	—	—

The alignment is based on three-dimensional structures of these enzymes. Gln148 and Glu152, not observed in HIV-1 IN structure, have been aligned by analogy.

Comparison with HIV-1 IN

A comparison of the sequences of the catalytic domains of ASV IN (residues 52 to 207) and HIV-1 IN (residues 50 to 212) reveals a moderate 24% identity, and additional 8% similarity (Figure 1B). Structural alignment of the crystallographic models of the monomers carried out in ALIGN (Satow *et al*., 1986) superimposed the conserved secondary structure elements with an r.m.s. deviation of 1.4 Å between 107 C^{α} pairs (blue and gold molecules in Figure 5B). However, there are several points to be noted here. (1) A segment of as many as 13 residues (containing the third active site residue, Glu152) is missing from the HIV-1 IN model (Dyda *et al*., 1994). Tentatively, the authors assigned an internal loop to this region, which is positioned between β5 and α4. In the ASV IN structure, however, the corresponding sequence (residues 146–158) covers the β5-α4 loop (where for residues 149–150 the temperature factors exceed 70 $Å^2$ and are the highest in the non-terminal part of the SER model) and a very well defined helical segment (154 to 158), which forms the N-terminal end of the long α4 helix and which contains the catalytic Glu157 as well as Met155 (perfectly detectable, through its Se-Met analog, even by direct methods based on CuKα data, see Methods). (2) Except for the disordered residues 209 to 212, the C-terminal end of the HIV-1 protein is visible and forms a well-defined helix, α6. The region corresponding to α6 in HIV-1 IN is not included in the ASV IN 52-207 fragment (Figure 1). In the ASV protein, very little density is seen beyond helix α5 (182 to 196) and only an irregular random coil could be built from the last 11 residues even in the best resolved ammonium sulfate maps.

The general architecture of the ASV IN and HIV-1 IN dimers is basically the same. When the complete dimers are regenerated from the aligned monomers by their corresponding 2-fold axes, the symmetry-related monomers are not similarly superimposed (green and red molecules in Figure 5B). Even though the corresponding secondary structure elements can be recognized, some of them are systematically shifted by as much as 6 Å. Overall, the symmetry-related subunits have a relative 13.3° rotation and 4.8 Å translation.

The molecular surface area that is buried in the ASV IN core domain on dimerization is 766 $Å^2$ per monomer (32% hydrophilic, 68% hydrophobic). This number is significantly smaller than the corresponding value of 1395 $Å^2$ calculated for the HIV-1 IN core domain dimer (QUANTA ver. 3.3, standard probe radius of 1.4 Å). This observation is in keeping with our preliminary estimates of K_d from equilibrium sedimentation experiments, which indicate that association of the ASV IN core fragment is weaker than that of the HIV-1 IN core fragment (unpublished results).

In the area of the F185K mutation in the HIV-1 IN catalytic domain that was necessary to render it soluble (Dyda *et al*., 1994), the main chains of the two proteins align very well and the side-chain of the

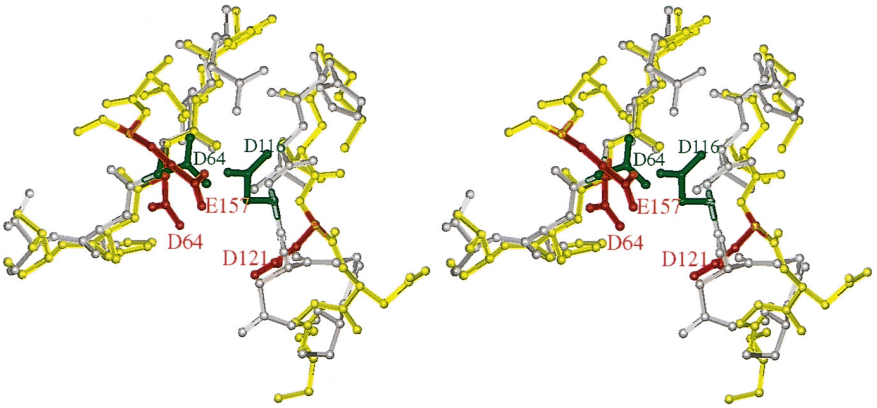


Figure 7. Superposition of the active sites of ASV (yellow and red) and HIV-1 (grey and green) IN. While very good agreement of the main chains can be observed in the strand including Asp64, the torsion angles of these residues differ, pointing them in opposite directions. For Asp121, large deviations can be seen for both the main-chain and the side-chain, and no equivalent of Glu157 can be seen in the structure of HIV-1 IN. The Figures were drawn with O (Jones & Kjeldgaard, 1994).

mutated Lys185 is placed in the general direction of the corresponding ASV IN residue His198 (Figure 1B). These corresponding residues in both structures use their side-chain N^+-H groups to form hydrogen bonds across the dimer to main-chain carbonyl groups that are shifted by one residue in the alignment of Figure 1B. In view of this comparison, it is unlikely that the F185K mutation in HIV-1 IN could have introduced any serious bias into the dimer structure as an analogous residue with an analogous function occurs naturally in ASV IN.

The monomer-monomer alignment of the HIV-1 and ASV IN catalytic domains reveals a puzzling picture in the active site. The β strands containing Asp64 has nearly identical main-chain trace in both models. In spite of this, the critical aspartate residues point in completely opposite directions in the two structures (Figure 7). On the other hand, the main chains embedding ASV IN Asp121 and its HIV-1 counterpart (Asp116) follow significantly different paths. These deviations become magnified at the aspartate sites as they again have different orientations. In fact, the carboxylate group of HIV-1 IN Asp116 is positioned closer to ASV IN Glu157 than to its Asp121 counterpart (Figure 7). It is unclear why two proteins, closely related in sequence, structure and function should have their absolutely conserved catalytic residues entirely out of alignment. We are confident that the organization observed in ASV IN is not an artifact, as it was confirmed in a number of well-refined, high-resolution structures.

The two active sites of the ASV IN dimer, as defined by the position of the water molecule between Asp64 and Asp121, are about 43 Å apart (Figure 5A). The active sites in the HIV-1 IN dimer are separated by 35 Å (Dyda *et al*., 1994). This measurement spans the carboxylate O atoms in the two Asp64 residues that are closest to Asp116; the corresponding Asp64 atoms in ASV IN are separated by 38 Å. At first glance, the separation of active sites in both HIV-1 and ASV dimers appears inconsistent with the length of the staggered cut introduced by IN during the insertion of a pair of viral DNA ends into target DNA. The respective 5 bp and 6 bp staggers, inferred from the duplication of the host sequences flanking the integrated HIV-1 and ASV proviral DNA, span only 16 to 20 Å on a *B*-DNA helix. It is possible that the dimer interface observed here is not related to selection of staggered phosphate groups. However, in recent experiments, we have examined a DNA substrate in which the two viral DNA ends are juxtaposed using a flexible single-stranded DNA tether that could be adjusted to different lengths (unpublished results). In this way, the coordinated processing of the two HIV-1 and ASV DNA ends could be examined. We observed that there is an optimum spacing between the two target phosphate groups (adjacent to the conserved CA; see Figure 2A) which, allowing for torsional and rotational flexibility, can be accommodated in the 35 to 40 Å active site separation seen in the crystal. We also propose that joining of one viral end to one strand of target DNA would torsionally relieve the target DNA to allow similar 35 to 40 Å positioning of the two target phosphate groups on host DNA. Furthermore, we note that the ~3 Å difference in separation between the ASV and HIV-1 IN active sites is compatible with the single nucleotide difference in the optimal length of the tether in the two-ended substrates of coordinated processing reactions (unpublished results). We also note that the single base-pair difference in the length of staggered cuts produced during joining *in vivo* is again consistent with the difference observed in the two dimers. Thus, the differences uncovered by our comparison of the ASV and HIV-1 IN catalytic domain dimers actually support the biological relevance of both. It seems likely that the dimer organization observed in the catalytic crystals is present in the multimeric structure of the full-length proteins in solution.

Summary and Conclusions

The catalytic domain of ASV IN was expressed as a soluble protein, and was crystallized. The protein recovered from dissolved crystals has unchanged enzymatic properties. The structure was solved using multiwavelength anomalous diffraction and refined at 1.7 Å resolution. The fold of the protein is similar to that of other nucleases, such as RNase H and RuvC, as well as the analogous catalytic domain of HIV-1 IN. The similarity of the quaternary structure of the two IN catalytic domains strongly supports the relevance of the observed crystallographic dimers to their enzymatic function. The observation that all residues comprising the predicted active site of ASV IN are fully ordered indicates that the partial disorder observed in the related domain of HIV-1 IN is not an intrinsic feature of this class of enzymes. Unexpectedly, we observed large differences in the conformation of the residues forming the active site. These structural differences may reflect biologically relevant differences between the two proteins. However, we also must question whether both structures correspond to the active enzyme. It may be that the structure of ASV IN could be useful, or perhaps essential, for a structure-based search for AIDS drugs that target IN. Studies aimed at improving our understanding of the precise roles of the catalytic residues are in progress.

Methods

Expression, purification and crystallization of the ASV IN catalytic domain

The expression strategy, purification from *E. coli* and activities of the purified ASV IN 52-207 fragment have been described (Kulkosky *et al*., 1995). For MAD experiments, selenomethionine substitution was carried out as described (Hendrickson *et al*., 1990; Yang *et al*., 1990b) using the *E. coli* methionine auxotrophic strain DL41 (Hendrickson *et al*., 1990), kindly provided by G. D. Markham. The extent of substitution was determined by electrospray ionization mass spectrometry experiments performed at the University of Michigan Protein and Carbohydrate Structure Facility by R. Ogorzalek Loo.

Even with a properly designed soluble core fragment of the ASV IN, crystallization was not immediate. The storage buffer normally contains 50 mM Hepes, 0.5 M NaCl, 1% (v/v) thiodiglycol, 0.1 mM EDTA and 40% (v/v) glycerol. Removing all glycerol caused aggregation and precipitation of the protein at lower than the target crystallization concentration. Dialysis into a 5% glycerol storage buffer retained full protein solubility at 4°C. The first crystals were produced using the hanging drop vapor diffusion method and two different types of precipitants from Hampton Research Crystal Screen I. Crystallization conditions were optimized for solutions containing 2 M ammonium sulfate, or 20% (w/v) PEG 4000 with 20% or 10% (v/v) isopropanol, and for different buffering agents. During optimization of the PEG conditions, we were able to grow crystals at isopropanol concentration from 6 to 12% and pH from 5.6 to 8.5. Optimization of the ammonium sulfate conditions required only adjustment of the protein to well ratio. Finally, we found an optimum concentration of protein between 6 and 7.5 mg/ml, using 100 mM Hepes at pH 7.5, 10% isopropanol, 20% PEG 4000, or 100 mM Hepes, 2 M ammonium sulfate, 2% PEG 400. The optimum protein to well ratio is 7 µl:4 µl, yielding 0.3 to 0.5 mm tetragonal bipyramidal crystals in two to three days for PEG and in seven to ten days for ammonium sulfate. Crystals of the Se-Met derivative were grown in an analogous manner and were handled without any attempt to maintain oxygen-free conditions.

Data collection using conventional X-ray sources

Diffraction data for the crystals PEG, SERT, CIT and AMS (Table 2) were collected at room temperature on a MAR 300 mm image plate detector, using graphite-monochromated CuKα radiation generated from a Rigaku Ru200 rotating anode operated at 50 kV and 100 mA. The diffraction data were processed with Denzo and scaled with Scalepack (Otwinowski, 1992). Data were also collected for a number of crystals soaked in heavy-atom solutions, but none of them provided useful isomorphous derivatives. The only truly isomorphous derivative was obtained by crystallizing the enzyme from selenomethionyl preparations.

Determination of the Se positions

The native and Se-Met crystals grown from PEG/isopropanol were highly isomorphous. The corresponding room-temperature CuKα diffraction data (PEG and SERT) scaled with $R(I)$ of 0.083 to 0.120 (overall 0.100) throughout the 20.0 to 3.0 Å range. After normalization (unpublished results), the two data sets were converted into a ΔE corresponding to the difference structure. In all, 189 ΔE values ($\geqslant 1.76$) in the resolution range 20 to 3.7 Å were used to locate the four Se atoms by means of the minimal function (Hauptman & Hahn, 1993; Weeks *et al*., 1993, 1994; De Titta *et al*., 1994) and a program (Langs *et al*., 1995) especially adapted for this purpose. The positions and occupancies of the Se atoms determined by direct methods agree remarkably well with those obtained by MAD techniques and in phase-model refinements (see below). However, an isomorphous difference Patterson map calculated with the Se-Met/native CuKα data would be impossible to deconvolute without the results from direct methods.

MAD data collection

Multiwavelength diffraction data were collected at line F2 of the CHESS synchrotron in Cornell using a single crystal (0.22 mm × 0.15 mm) grown from the Se-Met derivative in isopropanol/PEG/Hepes solution. The crystal had been transferred from the mother liquor through a layer of a cryoprotectant solution (mother liquor with isopropanol concentration increased to 15%), suspended in a fiber loop (Mitchell & Garman, 1994) affixed to the tip of a syringe needle, shock-frozen in liquid nitrogen and instantaneously transferred to the goniostat and placed in a stream of cold nitrogen gas (−165°C) flowing coaxially with the spindle axis.

Prior to the diffraction experiments, an X-ray fluorescence spectrum was recorded 90° away from the direct beam for a crystalline sample (0.8 mm^3) sealed in a glass capillary. The spectrum had a strong peak near the Se *K* absorption edge from which the three energies for the multiwavelength diffraction experiments were selected (λe = 0.9792 Å (inflection point), λp = 0.9790 Å (absorption peak), λr = 0.9464 Å (high-energy remote point)).

Table 4. Phase refinement statistics

	Resolution (Å)	R_{Cullis}	Figure of merit	Phasing power	
				Range	Overall
Isomorphous data	20.0–2.2	0.287	0.593	8.10–2.38	4.90
Anomalous data	20.0–2.2		0.582	9.28–3.00	5.86

Overall figure of merit 0.851.

The diffraction data were collected on a CCD detector using the oscillation method (oscillation range 1.75°, oscillation time ten seconds). Images with Friedel pairs were recorded close together in time using the inverse beam geometry (Hendrickson, 1991). At each wavelength, 51 images were recorded with ''F+'' reflections and 51 with ''F–'' reflections. The images were corrected for background (using a reference ''beamless'' exposure) and integrated with Denzo (Otwinowski, 1992) to produce three data sets, one for each wavelength. Each of the data sets was subsequently scaled using Scalepack (Otwinowski, 1992). In the scaling process, which included partially recorded reflections, the Friedel pairs measured at λp were kept separate, but were merged together at λe and λr for use in isomorphous replacement (Table 2). The integration and scaling procedures did not indicate any signs of crystal deterioration during the data collection. Postrefinement of the isotropic mosaicity parameter gave a value that was close to 1.0° and remarkably stable for all data.

Solution of the phase problem

An anomalous difference Patterson map calculated for SEP contained all the interatomic vectors corresponding to the four Se sites obtained from direct methods. An automatic interpretation of this map (program PASOL, M. Jaskólski, unpublished) including peaks down to 0.01*P(000) produced just the four Se sites, in order consistent with the direct-methods solution. An analysis of the differences between $^{\lambda r}I(hkl)$ and $^{\lambda e}I(hkl)$ (for nomenclature, see Hendrickson, 1991) in SER and SEE gave an $R(I)$ value of 0.066, which was higher than the individual $R_{merge}(I)$ values (Table 2) and much higher than $R_{Bijvoet}(I)$ calculated between $^{\lambda}I(hkl)$ and $^{\lambda}I(\overline{hkl})$ in each data set (about 0.035). A difference Patterson map calculated for SER and SEE confirmed the four Se sites. An automatic interpretation of all peaks within 0.01*P(000) carried out in PASOL found the four Se atoms (again, in order consistent with the direct-methods solution), plus one low-height spurious solution.

The parameters of the four Se atoms were refined in a phase refinement process based on the SER-SEE (isomorphous replacement, Se K absorption edge *versus* remote point scattering) and SEP (anomalous scattering) data. Since the sample and Se foil fluorescence spectra showed only a small shift (less than 2 eV), the Se anomalous scattering factors were roughly estimated from selenium dispersion curves (–2.7, 3.4 e (remote point), –7.0, 5.0 e (peak), –9.0, 3.0 (inflection point), for f' and f", respectively). The atomic parameters derived from the two contributing data sets were practically identical. The refinement converged with excellent statistics, summarized in Table 4.

The phase refinement was followed by solvent flattening, which assumed 0.42 solvent content. The solvent flattening calculations were performed in the two space group enantiomorphs leading to an unambiguous assignment for the space group as $P4_32_12$. The final statistics are as follows (values in parentheses for $P4_12_12$): map-inversion R-factor 0.189 (0.360), correlation coefficient 0.976 (0.912), mean figure of merit 0.943 (0.909). All Fourier calculations, the phase refinement and solvent flattening were done using the PHASES package (Furey & Swaminathan, 1990).

Model building and refinement

The electron density maps were calculated with the programs X-plor (Brünger, 1992) and PROTEIN (Steigeman, 1974), and were interpreted using FRODO (Jones, 1985) and O (Jones & Kjeldgaard, 1994). The model underwent one round of refinement by simulated annealing in X-plor (Brünger, 1992) followed by restrained structure-factor least-squares refinement in PROLSQ (Hendrickson, 1985) running in a VMS shell for automatic handling and submission of multiple cycles and for free R-factor calculations (M. Jaskólski, unpublished). The PROLSQ refinement was run in several rounds, which were followed by manual examination and adjustments of the model and by automatic interpretation (based on stereochemical criteria) and inclusion of difference-Fourier peaks as water molecules (M. Jaskólski, unpublished). In addition to protein atoms and water molecules,

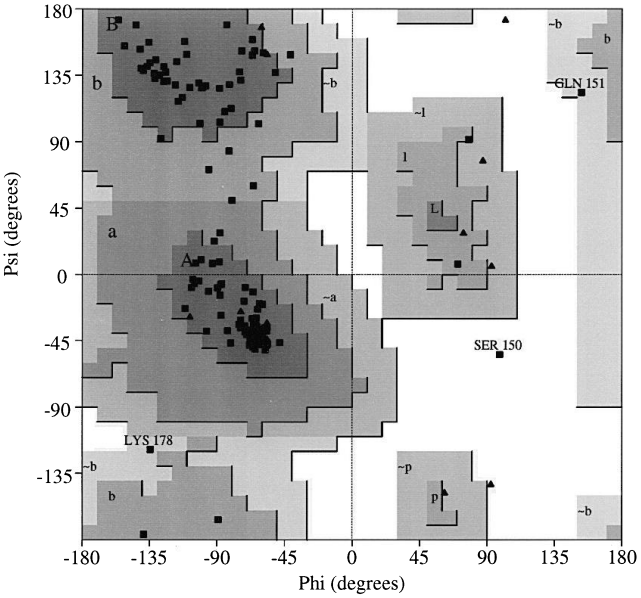


Figure 8. Ramachandran plot for the low-temperature Se-Met structure of ASV IN at 1.95 Å resolution. All residues other than glycine are marked as squares, glycine residues are marked as triangles. Three residues deviating most from the most favored regions (one of them in a disallowed region) are explicitly marked. This Figure was prepared with the program PROCHECK (Laskowski *et al*. 1994).

the final model also includes a well-ordered Hepes molecule (Figure 3B) and an isopropanol molecule that cocrystallized with the protein. The final refinement (Table 1) was based on 6.0 to 1.95 Å diffraction data (plus a limited number of reflections with *d*-spacings to 1.82 Å). The reason for limiting the low resolution end was a relatively high *R*-factor in that range that was attributed to thermal diffuse scattering (TDS) effects clearly seen in the low-resolution synchrotron data. The Ramachandran plot for the final structure is shown in Figure 8. The only pair of torsion angles in a disallowed region is for Ser150, which, together with Gln151, is in a poorly ordered region. More than 90% of the residues are in the most favored regions of the plot, in line with what would be expected for well-refined structures at this resolution.

Refinements using data sets collected with conventional X-ray source

The structure of the Se-Met protein at low temperature (SER), with a Hepes molecule but without water molecules, was used as the initial model for the ammonium sulfate structure (AMS). After the first round of refinement the loop region Thr144 through Gly152 was found in negative density, but the nearby positive electron density indicated a different conformation for this loop. During the refinement, the missing C and N-terminal residues were also modeled and 181 water molecules were added. The program FRODO was used to rebuild the model, and PROCHECK (Laskowski *et al*., 1994) was used to assess the quality of the structure during the refinement. All other structures listed in Table 1 were refined in an analogous fashion.

The coordinates have been deposited with the Protein Data Bank (accession numbers: 1ASU, 1ASV, and 1ASW for AMS, CIT, and SER, respectively).

Acknowledgements

We acknowledge Drs David Davies and Fred Dyda (NIH) for sharing with us the results of their studies of HIV-1 IN prior to publication, and for constructive discussions. We are indebted to Dr David A. Langs (Hauptman-Woodward Medical Research Institute, Buffalo) for making a developmental version of the minimal function program available to us and for valuable discussions, to Dr Gwyndaf Evans (EMBL, Hamburg) for his help in the interpretation of X-ray fluorescence scans, and to Dr Wei Yang (Yale) for the discussions of MAD phasing. Several of our Fox Chase colleagues have provided encouragement and helpful advice throughout this project. B. Mueller helped with functional analyses, G. Kukolj with Se-Met protein expression, M. Andrake with structural comparisons, and E. Assante-Appiah critiqued the manuscript. We thank R. Ogorzalek Loo at the University of Michigan protein and Carbohydrate Structure Facility for carrying out electrospray ionization mass spectrometry experiments. We acknowledged the Frederick Biomedical Supercomputer Center for the allocation of significant computing resources, and the CHESS facility for enabling us to collect MAD data. This research was sponsored in part by the National Cancer Institute, DHHS, under contract NO1-CO-46000 with ABL. Other support includes National Institutes of Health grants CA47486, CA06927, a grant for infectious disease research from Bristol-Myers Squibb Foundation, and an appropriation from the Commonwealth of Pennsylvania. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H. & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell*, **78**, 1063–1072.

Brünger, A. (1992). *X-PLOR: a System for X-ray crystallography and NMR*, Yale University Press, New Haven.

Bushman, F. D. & Wang, B. (1994). Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. *J. Virol.* **68**, 2215–2223.

Bushman, F. D, Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993). Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl Acad. Sci. USA*, **90**, 3428–3432.

Chow, S. A., Vincent, K. A., Ellison, V. & Brown, P. O. (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **225**, 723–726.

Craigie, R., Fujiwara, T. & Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell*, **62**, 829–837.

Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88–95.

DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). Structure solution by minimal-function phase refinement and Fourier filtering. I. Theoretical basis. *Acta Crystallog. sect. A*, **50**, 203–210.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.

Engelman, A. & Craigie, R. (1992). Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J. Virol.* **66**, 6361–6369.

Engelman, A., Mizuuchi, K. & Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211–1221.

Engelman, A., Bushman, F. D. & Craigie, R. (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12**, 3269–3275.

Furey, W. & Swaminathan, S. (1990). Phases—A program package for the processing and analysis of diffraction data for macromolecules. American Crystallographic Association Meeting, Program Abstract 18, 73.

Goff, S. P. (1992). Genetics of retroviral integration. *Annu. Rev. Genet.* **26**, 527–544.

Hauptman, H. A. & Hahn, F. (1993). Phasing macromolecular structures *via* structure-invariant algebra. *Acta Crystallog. sect. D* **49**, 3–8.

Hendrickson, W. A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.*, **115**, 252–270.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.

Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M. & Skalka, A. M. (1992). Retroviral integrase functions as a multimer and can turn over catalytically. *J. Biol. Chem.* **276**, 16037–16040.

Jones, T. A. (1985). Interactive computer graphics: FRODO. *Methods Enzymol.* **115**, 157–171.

Jones, T. A. & Kjeldgaard, M. (1994). *O—The Manual*. Uppsala University Press, Uppsala.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli*. *Nature*, **347**, 306–309.

Katayanagi, K., Okumura, M. & Morikawa, K. (1993). Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for single Mg^{2+}-binding site. *Proteins: Struct. Funct. Genet.* **17**, 337–346.

Katz, R. A. & Skalka, A. M. (1994). The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133–173.

Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell*, **63**, 87–95.

Katz, R. A., Mack, J. P., Merkel, G., Kulkosky, J., Ge, Z., Leis, J. & Skalka, A. M. (1992). Requirement for a conserved serine in both processing and joining activities of retroviral integrase. *Proc. Natl Acad. Sci. USA*, **89**, 6741–6745.

Katzman, M., Katz, R. A., Skalka, A. M. & Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the *in vivo* sites of integration. *J. Virol.* **63**, 5319–5327.

Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J. & Skalka, A. M. (1991). Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucl. Acids Res.* **19**, 851–860.

Krauls, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. & Skalka, A. M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* **12**, 2331–2338.

Kulkosky, J., Katz. R. A., Merkel, G. & Skalka, A. M. (1995). Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase. *Virology*, **206**, 448–456.

Langs, D. A., Guo, D. & Hauptman, H. A. (1995). Use of ''Random-Atom'' phasing models to determine macromolecular heavy atom positions. *Acta Crystallog.* In the press.

Laskowski, R. A., MacArthur, M. W., Smith, D. K., Jones, D. T., Hutchinson, E. G., Morris, A. L., Naylor, D., Moss, D. S. & Thornton, J. M. (1994). PROCHECK v.3.0. University College, London.

Mitchell, E. P. & Garman, E. F. (1994). Flash freezing of protein crystals: investigation of mosaic spread and diffraction limit with variation of cryoprotectant concentration. *J. Appl. Crystallog.* **27**, 1070–1074.

Mumm, S. R. & Grandgenett, D. P. (1991). Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis. *J. Virol.* **65**, 1160–1167.

Otwinowski, Z. (1992). *An Oscillation Data processing Suite for Macromolecular Crystallography* Yale University, New Haven.

Roth, M. J., Schwartzberg, P. L. & Goff, S. P. (1989). Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell*, **58**, 47–54.

Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603: an X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593–604.

Steigemann, W. (1974). Die Entwicklung und Anwendung von Rechenverfahren und Rechenprogrammen zur Strukturanalyse von Proteinen am Beispiel des Trypsin-Trypsininhibitor Komplexes, des freien Inhibitors und der L-Asparaginase. PhD. thesis, Technische Universität, Müchen, Germany.

Steitz, T. A. (1993). DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38.

van Gent, D. C., Vink, C., Groeneger, A. A. M. O. & Plasterk, R. H. (1993). Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* **12**, 3261–3267.

Vink, C., Groeneger, A. A. M. O. & Plasterk, R. H. (1993). Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucl. Acids Res.* **21**, 1419–1425.

Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). Applications of the minimal principle to peptide structures. *Acta Crystallog. sect. D*, **49**, 179–181.

Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). Structure solution by minimal-function phase refinement and Fourier filtering. II. Implementation and applications. *Acta Crystallog. sect. A*, **50**, 210–220.

Whitcomb, J. M. & Hughes, S. H. (1992). Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell Biol.* **8**, 275–306.

Woerner, A. M., Klutch, M., Levin, J. G. & Marcus-Sekura, C. J. (1992). Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. *AIDS Res. Hum. Retrovir.* **8**, 297–304.

Yang, W. & Steitz, T. A. (1995). Recombining the structures of HIV integrase, RuvC and RNase. *H. Structure*, **3**, 131–134.

Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990a). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science*, **249**, 1398–1405.

Yang, W., Hendrickson, W. A., Kalman, E. T. & Crouch, R. J. (1990b). Expression, purification and crystallization of natural and selenomethionyl recombinant ribonuclease H from *Escherichia coli*. *J. Biol. Chem.* **265**, 13553–13559.

Edited by I. A. Wilson

(Received 16 June 1995; accepted 31 July 1995)