Journal of Computer-Aided Molecular Design (2005) 19: 301–317
DOI 10.1007/s10822-005-5256-2

A three-dimensional model of the human immunodeficiency virus type 1 integration complex

Jerome Wielens*, Ian T. Crosby & David K. Chalmers
Department of Medicinal Chemistry, Monash University, 381 Royal Parade, 3052 , Parkville, Vic., Australia

Received 10 November 2004; accepted in revised form 7 April 2005

Key words: amino acid conservation, GRID, HIV-1 integrase, integration complex, molecular modelling

Summary

While the general features of HIV-1 integrase function are understood, there is still uncertainty about the composition of the integration complex and how integrase interacts with viral and host DNA. We propose an improved model of the integration complex based on current experimental evidence including a comparison with the homologous Tn5 transposase containing bound DNA and an analysis of DNA binding sites using Goodford's GRID. Our model comprises a pair of integrase dimers, two strands of DNA to represent the viral DNA ends and a strand of bent DNA representing the host chromosome. In our model, the terminal four base pairs of each of the viral DNA strands interact with the integrase dimer providing the active site, while bases one turn away interact with a flexible loop (residues 186–194) on the second integrase dimer. We propose that residues E152, Q148 and K156 are involved in the specific recognition of the conserved CA dinucleotide and that the active site mobile loop (residues 140–149) stabilises the integration complex by acting as a barrier to separate the two viral DNA ends. In addition, the residues responsible for DNA binding in our model show a high level of amino acid conservation.

Introduction

HIV-1 integrase (IN) catalyses the incorporation of the HIV-1 genome into the chromosome of the host cell. This is an essential step in viral replication and for the establishment of HIV infection [1, 2]. Although there are a variety of *in vitro* assays of IN function [3] and a large number of partial IN crystal structures, there is to date no detailed information about the structure of the integration complex or how IN interacts with DNA. In this situation, molecular modelling approaches can be used to aggregate the available information to produce a self-consistent model and provide insight into what the integration complex may encompass. This process has been performed by several research groups who have previously produced models of the HIV-1 integration complex [4–7] or who have predicted IN–DNA contacts [8–14]. In the absence of more definitive structural data, these models provide our best insight into the understanding of this complex enzyme.

IN catalyses two reactions during the integration process: 3′-processing and strand transfer [15–17]. 3′-Processing removes the terminal two nucleotides from both 3′-ends of the viral DNA, adjacent to a conserved CA dinucleotide. 3′-Processing occurs in the cytoplasm of the infected cell shortly after reverse transcription [1]. In contrast, strand transfer occurs in the nucleus of the host cell and involves the direct nucleophilic attack of the 3′-hydroxyl group of the recessed 3′-viral DNA ends on the phosphodiester backbone of the host chromosome. The resulting DNA intermediate, in which the 3′-ends of the viral DNA are covalently linked to the host DNA and the 5′-ends of the viral DNA are flanked by short gaps, is thought to be repaired by cellular

*To whom correspondence should be addressed. Fax +61-3-99039674; E-mail: jerome.wielens@vcp.monash.edu.au

enzymes [18, 19]. A number of simple bioassays have been developed to simulate the 3′-processing and strand transfer activities of IN [3, 16, 20–22]. In addition, IN can also catalyse the reverse of the strand transfer reaction, called disintegration, *in vitro* [23].

IN comprises three functional domains connected by short flexible linkers; the N-terminal domain, the catalytic or core domain and the C-terminal domain. The structure of the core domain has been determined by X-ray crystallography [24–30] and structures of the C- and N-terminal domains have been solved by NMR spectroscopy [31–35]. More recently several structures of two domain fragments have been reported [6, 9, 10].

The N-terminal domain (residues 1–45) contains a conserved HHCC motif, which binds Zn^{2+}. This domain has been shown to interact with DNA as part of the IN complex [36, 37] but does not exhibit DNA binding properties when isolated in solution [38, 39]. In HIV-1 IN, a functional N-terminal domain is required for 3′-processing and strand transfer activity *in vitro* [40] however, addition of IN with a functional N-terminal domain in *trans* to an IN mutant missing this domain can restore strand transfer and 3′-processing activity *in vitro* [41, 42]. The N-terminal domain is not required for disintegration [39] and has been implicated in protein–protein interactions and multimerisation of IN [42].

The core domain (residues 55–212) has an RNaseH-like fold and contains an essential DDE motif comprising D64, D116 and E152. The DDE motif forms a divalent cation-binding site, which is universally conserved among the retroviral integrases, a number of bacteriophage transposases, RNaseH and the Holliday junction RuvC resolvases [25, 43]. Mutation of any of these residues severely diminishes or abolishes all catalytic activities of IN both *in vitro* [44–46] and *in vivo* [47,48].

The C-terminal domain (residues 220–270) is the least conserved domain among the retroviral integrases [38, 49, 50]. The isolated C-terminal domain has been shown to bind DNA in a non-specific manner [38, 49–51] and also dimerises when isolated in solution [52]. In addition, this domain has been implicated in protein–protein interactions including interactions with reverse transcriptase [53, 54].

Although there is no reported crystal structure of a complex between DNA and IN, a number of crystal structures of functionally similar proteins with DNA bound in the active site are available [55–57]. Of these, Tn5 transposase is the most closely related. This enzyme rearranges transposable elements in the host DNA using the 3′-processing and strand transfer reactions similar to IN [43, 58–61]. Like IN, Tn5 transposase contains a conserved DDE motif and requires a Mg^{2+} or Mn^{2+} ion for catalytic activity but in contrast, Tn5 transposase forms a hairpin intermediate in the transposon end after 3′-processing, whereas IN does not [58, 60]. Despite this, the similarities between the two reaction mechanisms are clear. A structure-based sequence alignment of Tn5 transposase and the core domain of HIV-1 IN reveals that, in addition to the DDE motif, these enzymes share a number of functionally important active site residues which interact with the terminal nucleotides of the 3′-end of the Tn5 transposon DNA. Based on the homology between Tn5 transposase and IN, we propose that the binding orientation of the viral DNA in the IN active site will be similar to the transposon DNA in the Tn5 transposase crystal structure.

The model that we present here represents an improved description of the integration complex that utilises the Tn5 transposase crystal structure as the template to position the viral DNA into the IN active site. It has been refined using a number of constraints from the research of Heuer and Brown [4, 36], Esposito and Craigie [37], Gao et al. [5] and Jenkins et al. [62] and from our current understanding of the integration mechanism. Further to this, we have used Goodford's GRID [63] to determine favourable interaction sites for a phosphate group around the IN domains and have used these data to position the DNA strands. The completed model is supported by the pattern of amino acid conservation found in a diverse set of 189 HIV-1 IN amino acid sequences. The picture of the integration process captured by the model is that of the strand transfer reaction just prior to integration of the recessed viral DNA ends into the host DNA. An enhanced understanding of IN function improves our ability to develop new anti-IN drugs which are needed to complement the existing anti-HIV therapies.

Methods

Model components

Modelling of the integration complex was performed using InsightII (Accelrys Inc., San Diego, CA, USA). The integration complex was assembled manually in an iterative process; using residues 1–46 of chain A of 1WJA [34], residues 50–209 of chains A and C of 1BL3 [30] and residues 220–270 of chain A of 1IHV [31], to represent the three IN domains. The IN core domain dimer was generated by replacing the coordinates of chain B of 1BL3 with those of chain C. The viral DNA ends were modelled using the coordinates of the Tn5 transposon DNA strand from the 1MUH crystal structure [57]. After removal of the DNA hairpin, two nucleotides were appended to the 5′-end of the extracted transposon DNA to represent the 5′-overhang of the recessed viral DNA strand not present in this structure. Each nucleotide was then mutated using the biopolymer module of InsightII to correspond to the terminal bases of the U5 and U3 HIV-1 long terminal repeat (LTR) ends. The sequences and numbering of the two viral LTRs and the host DNA substrate are shown in Figure 1. The terminal three nucleotides on the 5′-end of each viral DNA strand were positioned manually to remove steric clashes in the completed model. Bases 15–20 of the Tn5 transposon structure were retained unchanged in the modelled viral DNA strands to extend the viral DNA past position 14. The coordinates for the host DNA were extracted from the Heuer and Brown model of the integration complex [4]. The HIV-1 IN core domain (chain C of 1BL3) and the Tn5 transposase structures were superimposed using the Dali method [64]. With the exception of the modifications to the viral LTRs described above, the coordinates of the structures used in preparing our model of the integration complex were not altered because we do not believe that this will produce a significantly improved model. This results in some minor steric clashes between IN and the viral DNA strands. The atomic coordinates of this model have been deposited with the Protein databank (PDB ID code 1ZA9).

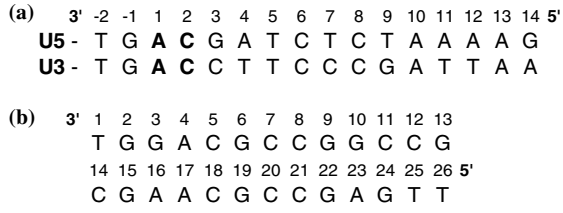

Figure 1. (a) Sequence of the U5 and U3 double stranded HIV-1 viral DNA substrates. The conserved CA is highlighted and the terminal 16 bases of the LTR are numbered. (b) The sequence of the double stranded 26-mer host DNA substrate.

GRID methods

Binding affinity maps for a phosphate probe were generated using the GRID 19 software suite (Molecular Discovery Ltd, London, UK) for each of the three IN domains. The Mg^{2+} ion and the four crystallographic water molecules coordinated to it were retained in chain C of the 1BL3 structure, as was the Zn^{2+} ion in chain A of 1WJA. All other water molecules were removed. Hydrogen atoms, charges and bonding information for each IN domain were taken from the GRUB.dat parameter file. The water molecules coordinated to the Mg^{2+} ion were treated as Type 95 extended water molecules. This allows free rotation of hydrogen atoms around the fixed oxygen atom during the GRID calculations, except where sterically constrained by other atoms. The Mg^{2+} ion was assigned atom type 112, which allows coordination to oxygen and nitrogen atoms.

Spacing between grid points was 0.5 Å (NPLA = 2). At each point on the grid, the non-bonded interaction energy between the phosphate probe and the IN receptor was calculated, with distance dependent cut-offs for calculating hydrogen bonding (FARH) and Lennard-Jones interactions (FARR) set to 5 and 8 Å respectively.

Amino acid conservation in HIV-1 IN

The 2001 compendium from the HIV Sequence Database [65] (http://www.hiv.lanl.gov) was used as the source of protein sequences for HIV-1 isolates. Amino acids encoded by the *pol* gene in this dataset were edited to include only residues in HIV-1 IN. All consensus sequences present in this data set were removed. A total of 189 sequences remained and were used in this study.

Table 1. Scoring criteria and colour scheme for Conscore.

Score	Criteria	Colour
10	Invariant	Green
9	100% Conserved	Blue
8	>95% Conserved	Red
7	>90% Conserved	
6	> 85% Conserved	
5	>80% Conserved	Pink
4	>75% Conserved	
3	>70% Conserved	
2	>60% Conserved	
1	>50% Conserved	Grey
0	<50% Conserved	White

Conscore

A simple colouring scheme (Conscore) was used to visualise amino acid conservation in the HIV-1 IN isolates (Table 1). Conservative substitution was defined as being in one of the following seven groups: aliphatic hydrophobic (alanine, valine, leucine and isoleucine), aromatic hydrophobic (tryptophan, tyrosine and phenylalanine), acidic (glutamic acid and aspartic acid), basic (arginine, histidine and lysine), neutral polar (serine, methionine, threonine, cysteine, asparagine and glutamine), glycine and proline. The calculated scores were entered into the temperature factor column of each pdb file and coloured using the 'colour by property spectrum' function in InsightII.

Nomenclature of the integration complex model

We define the IN dimer catalysing integration of the U5 viral LTR as Dimer A and the dimer catalysing the integration of the U3 viral LTR as Dimer B. The two monomer units of Dimer A are named A and A′, where Monomer A contains the catalytic active site interacting with the viral and host DNA strands. The active site of Monomer A′ is not involved in the integration reaction. The same nomenclature is used for the IN monomers in Dimer B i.e. B and B′, where Monomer B provides the catalytic active site for integration. The numbering of the viral DNA strands is shown in Figure 1. When referring to the complementary base at each nucleotide position a prime (′) is used. We describe domains from the same IN monomer to be 'in *cis*' to each other whereas domains from other monomers are 'in *trans*' to the first monomer. For example, the C-terminal domain of Monomer A is in *cis* to the core domain of Monomer A but the C-terminal domain of Monomer A is in *trans* to the core domains of Monomers A′, B and B′. For clarity, the active site flexible loop (residues 140–149), commonly called the 'mobile loop', will be referred to as Loop140 and the flexible loop (residues 186–196) will be called Loop186.

Results

The model of the HIV-1 integration complex was built using the structures of the core, C-terminal and N-terminal domains (see Methods) and was assembled using a number of experimentally derived constraints. The complete model is shown in Figure 2 and contains two IN dimers (grey/black and yellow/green), two 20-mer double strands of DNA that model the sequence of the U5 (light green) and U3 (blue) viral LTR ends and a 26-mer double strand of bent DNA to represent the host DNA (orange). Residues within 5 Å of the DNA strands are listed in Table 2.

The segments of the model were arranged to satisfy the following constraints:

(1) *The viral DNA ends bind in the IN active site*. For IN to catalyse the removal of two base pairs from the 3′-end of the viral DNA and then mediate the integration of the recessed 3′-ends into the host DNA, these components must be close to the active site residues.

(2) *The integration sites of the two HIV-1 viral DNA ends into the host chromosome are five base pairs apart*. The five base pair spacing between the insertion sites of the two viral DNA ends is constant and studies have shown that this feature is independent of cell type [15].

(3) *The host DNA is bent and integration occurs across a major groove*. Several groups have shown that the favoured site of integration of the two viral DNA ends is across a highly bent [66–69] and slightly unwound [70] major groove, i.e. DNA wrapped around a nucleosome. Heuer and Brown [4] modelled this feature into their host DNA

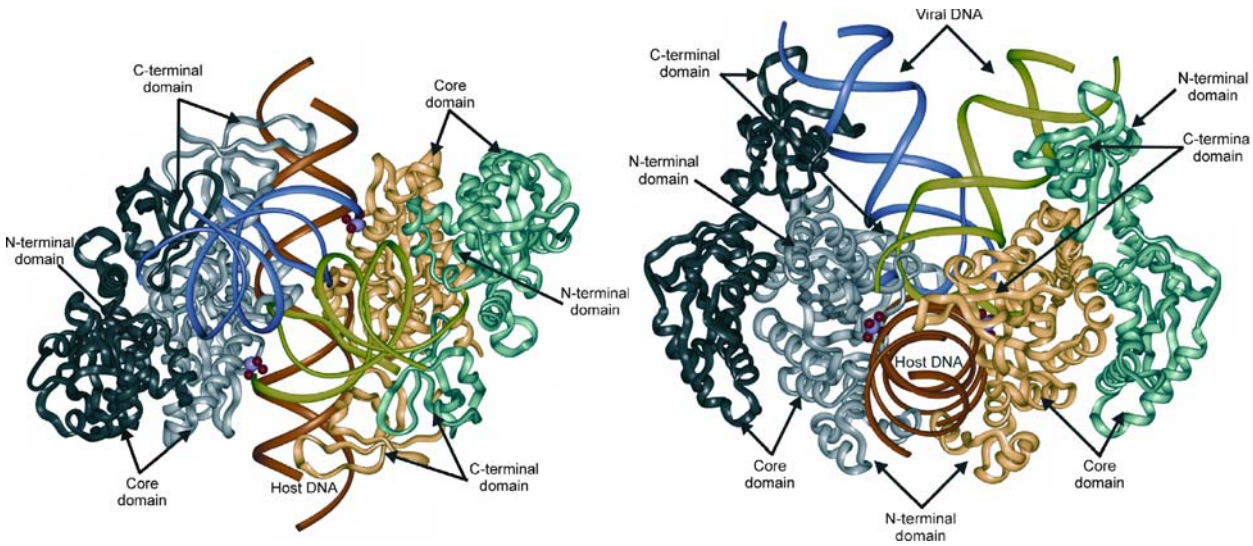

Figure 2. Two views of our proposed model of the HIV-1 integration complex, where two dimers of HIV-1 IN catalyse the integration of the U5 and U3 viral DNA ends in the host DNA. Monomer A (grey) provides the active site for integration of the U5 LTR (light green) into the host DNA (orange) and Monomer B (yellow) provides the active site for integration of the U3 viral DNA end (blue). The C- and N-terminal domains of Monomers A′ (black) and B′ (green) provide further stabilisation of the integration complex in *trans* to the catalytic active sites.

Table 2. HIV-1 IN residues within 5 Å of (A) the U5 viral DNA LTR, (B) the U3 viral DNA LTR and (C) the host DNA.

A	Residues
Monomer A	53, 64–68, 73, 116, 141–161
Monomer A'	No interactions
Monomer B	44, 45, 47, 50–60, 62, 79, 80, 111, 112, 114, 142–150, 157, 160, 187–190, 207, 220, 241–251, 257–259, 261–263
Monomer B′	13–15, 17–27, 230, 242–250, 257, 259–264
B	Residues
Monomer A	44, 45, 47, 50–60, 62, 79, 80, 111, 112, 114, 142–150, 157, 160, 187–190, 207, 220, 241–251, 257–259, 261–263
Monomer A′	13–15, 17–27, 230, 242–250, 257, 259–264
Monomer B	53, 64–68, 73, 116, 141–161
Monomer B′	No interactions
C	Residues
Monomer A	1, 2, 5, 9, 35, 39–41, 92, 114, 116–120, 138–142, 229–231, 256–259
Monomer A′	No interactions
Monomer B	1, 2, 5, 9, 35, 39–41, 92, 114, 116–120, 138–142, 229–231, 256–259
Monomer B′	No interactions

substrate in the generation of their integration complex model.

(4) *IN acts as a dimer*. Crystal structures of IN show the core domain as a dimer in all cases. This is also supported by other experimental evidence demonstrating IN self-association *in vitro* [40, 71–75]. The N- and C-terminal domains are not assumed to be dimers. A recent crystal structure of HIV-1 IN [10] showed that the two C-terminal domains of an IN dimer do not necessarily dimerise. Other studies have proposed that the N-terminal domain is involved in IN multimerisation [40, 76, 77].

(5) *Minimally, two IN dimers are required for integration*. In a single IN dimer, the two active sites are greater than 35 Å apart, which is inconsistent with the observed five base

pair spacing between the insertion sites of the two viral DNA ends (equivalent to ~18 Å in B-form DNA). *Trans*-complementation studies have shown that two separately inactive IN proteins, one lacking the N-terminal domain and the other lacking the C-terminal domain can restore IN activity when combined *in vitro* [78–80]. IN was also found to predominantly exist as a tetramer in solution in gel filtration experiments, when purified in the absence of detergent and in the presence of Zn^{2+} [40].

(6) *DNA binding to IN is predominantly through interactions with the phosphodiester backbone*. IN binds DNA in a sequence specific manner by recognising the conserved CA dinucleotide of the viral DNA end and non-specifically when binding host DNA. The non-specific binding ability of IN suggests that many of the interactions between the DNA strands and IN are with the phosphodiester backbone.

(7) *Experimental constraints*: (7a) *Photocross-linking*. Heuer and Brown performed a photocross-linking study to map interactions between HIV-1 IN and its DNA substrates using azidophenacyl (AZP) modified DNA substrates and the more sensitive 5-iododeoxycytosine (iodo) photocross-linking agent in the context of the disintegration reaction [4, 36]. Jenkins et al. proposed that the N7 nitrogen of the conserved adenosine is close to the NZ nitrogen of K159 [62]. Gao et al. showed that E246 is close to the nucleotide at position 5 of the U5 viral DNA in *trans* to the active site and S230 is close to position 7 after the 3′-processing step [5]. Esposito and Craigie [37] also performed a series of photocross-linking studies between viral DNA and HIV-1 IN. Their experiments were conducted in the context of the 3′-processing reaction with Mg^{2+} as the metal cofactor. These researchers found that residues 139–152 interact with the terminal 5′-adenosine of the viral DNA and that residues 51–64 interact with the penultimate cytosine of the 5′-viral DNA end. The conserved adenosine at the 3′-end of the viral DNA interacts with K159 and residues 247–270 interact with a thymine seven bases from the 3′-end of the viral DNA. (7b) *Mutational analysis*. In addition to the photocross-linking experiments, Esposito and Craigie [37] studied the effects of base substitution at each point in the terminal 14 nucleotides of the U5 viral LTR. These studies were conducted with the 3′-processing assay in the presence of either Mg^{2+} or Mn^{2+} to determine the different effect of these metals on the 3′-processing assay and viral DNA binding. They found that the terminal six base pairs of the viral LTR are important for IN recognition, as are nucleotides one turn away (nucleotides 8–11).

(8) *Homology with Tn5 transposase*. Tn5 transposase and IN are both members of the superfamily of polynucleotidyl transferases [25, 81]. While these two enzymes have low overall primary sequence identity (~7%), they have a similar function and share a common RNaseH-like fold in their catalytic domains. A superimposition of the core domains of Tn5 transposase (residues 70–372) and IN (residues 50–212) shows that 103 α-carbon atoms overlap between the two structures with a RMSD of 3.7 Å (Figure 3). A structure-based sequence alignment of the Tn5 transposase core domain and the HIV-1 IN core domain (Table 3) revealed that a number of residues were conserved or conservatively substituted in the overlapping region. Of interest in this alignment are residues T99, H329, K330 and K333 in Tn5 transposase, which are equivalent to T66, N155, K156 and K159 in HIV-1 IN respectively. These residues form the phosphate-binding pocket that recognises the penultimate phosphate group at the 3′-end of the Tn5 DNA in the 1MUH crystal structure. In addition, K330 and K333 are involved in specific interactions with the three terminal nucleotides of the 3′-transposon end and the equivalent residues have been shown to be important for recognition of the two viral DNA ends in HIV-1 IN [62, 82, 83].

In addition to the constraints described above, several other pieces of evidence supporting the proposed model were considered. Dirac and Kjems [84] investigated which residues were protected from proteolytic attack upon formation of the integration complex. They found that K34,

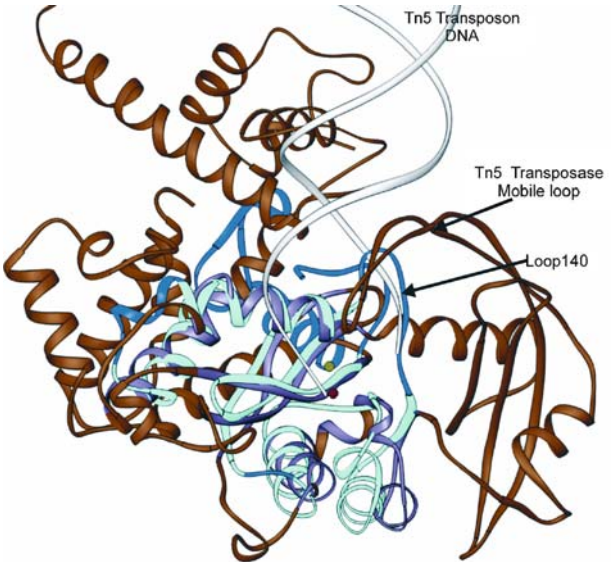

Figure 3. Superimposition of the Tn5 transposase/DNA complex (orange) and the core domain of HIV-1 IN (blue). Regions in Tn5 transposase and IN with similar topology are shown in pink and light blue respectively. The Tn5 transposon DNA is white and the Mn^{2+} of Tn5 transposase (green) and the Mg^{2+} ion (red) are shown as spheres.

Table 3. Structure-based sequence alignment of the HIV-1 IN core domain (1BL3 C [30]) and the core domain of Tn5 transposase (1MUH [57])[a].

1BL3 C	G59	I60	W61	Q62	L63	D64	C65	T66	H67	L68	E69	G70	K71	V72	I73
1MUH	L92	L93	A94	I95	E96	D97	T98	T99	S100	L101	S102	G123	W124	W125	V126
1BL3 C	L74	V75	A76	V77	H78	V79	S81	G82	Y83	I84	E85	A86	E87	V88	I89
1MUH	H127	S128	V129	L130	L131	L132	R138	T139	G141	L142	L143	H144	Q145	E146	W147
1BL3 C	P90	T93	G94	E96	T97	A98	Y99	F100	L101	L102	K103	L104	A105	G106	R107
1MUH	W148	E161	G163	K164	W165	L166	A167	A168	A169	A170	T171	S172	R173	L174	R175
1BL3 C	W108	P109	V110	K111	T112	V113	H114	T115	D116	N117	G118	F121	T122	S123	T124
1MUH	M176	M179	M180	V183	I184	A185	V186	C187	D188	R189	E190	A191	D192	I193	H194
1BL3 C	T125	V126	K127	A128	A129	C130	W131	W132	A133	G134	I135	K136	Q137	E138	F139
1MUH	A195	Y196	L197	Q198	D199	K200	L201	A202	H203	N204	E205	R206	F207	V208	V209
1BL3 C	G149	V150	I151	E152	S153	M154	N155	K156	E157	L158	K159	K160	I161	I162	G163
1MUH	W323	R324	I325	E326	E327	F328	H329	K330	A331	W332	K333	T334	A336	A338	G337
1BL3 C	E170	L172	K173	T174	A175	V176	Q177	M178	A179	V180	F181	I182	H183		
1MUH	N348	L349	E350	R351	M352	V353	S354	I355	L356	S357	F358	V359	A360		

[a]Invariant residues are shaded and residues found to be conservatively substituted by Conscore are outlined.

K111, E138, K185, K186, K188, D207, E246 and K273 were strongly protected from proteolytic attack, K136, K156, K159, K160 and K215 were weakly protected, and K258 showed variable results. Wang et al. [6] found an ordered phosphate ion in their crystal structure of IN (1K6Y). This phosphate ion is present in all four IN monomers in the asymmetric unit and interacts with H67, K159 and T66.

Model construction

Position of the U5 viral LTR

A preliminary model of the integration complex was prepared from a superimposition of the Tn5 transposase (1MUH) and HIV-1 IN core domain (1BL3 C) crystal structures (Figure 3). The modelled U5 viral DNA strand was then

superimposed onto the Tn5 transposon. This positioned the 3′-hydroxy group of the conserved adenosine close to the active site Mg^{2+} ion. The viral DNA strands were modelled directly from the coordinates of the Tn5 transposase/transposon crystal structure because the terminal four DNA base pairs in this crystal structure deviated from canonical B-form DNA. Only base pairs at positions 1 to 4 of the U5 viral LTR interact with Monomer A (Figure 2). The terminal three nucleotides of the 5′-end were manually positioned so that the phosphate groups of these bases overlapped GRID phosphate hotspots and agreed with the photocross-linking data of Heuer and Brown [4, 36] and Esposito and Craigie [37]. This results in the terminal three bases of the 5′-end flipping out of the DNA helix so that they can interact with residues in the core domain of Monomer B.

Position of the host DNA strand

To satisfy *Constraint 3*, the host DNA was positioned so that the zenith of the phosphodiester backbone flanking the bent major groove was close to the 3′-hydroxyl of the U5 viral DNA end and the active site Mg^{2+} ion (Figure 4). The orientation of the host DNA strand was then optimised using the phosphate GRID maps contoured to −10 kcal mol^{-1} (Figure 4). In the final position, the host DNA strand interacts with residues 92, 114, 116–120 and 138–142 of the core domain of Monomer A. In addition, a Connolly surface [85] mapped over Monomer A shows good shape complementarity with the architecture of the major groove. The position of the host DNA is consistent with the photocross-linking data for residues 49–69 and 139–152 (*Constraint 7a*) and is supported by the observation that mutation of either N120 or S119 (Figure 4, white arrow) alters the preferred integration site on the host DNA [46, 86, 87]. Appa et al. [88] also showed that residues in α2 (residues 120–123) altered the preferred integration site on the host DNA using a series of HIV-1/FIV IN chimeras.

Position of Dimer B and the U3 viral DNA strand

The Tn5 transposase template was used to position the U3 viral LTR in the Dimer B active site. Dimer B and the U3 viral LTR were then located in the model so that they were related to Dimer A and the U5 viral LTR by an approximate two-fold axis of symmetry around the host DNA strand and were consistent with *Constraints 1, 2, 3* and *5*. Two points suggest symmetry of the integration com-

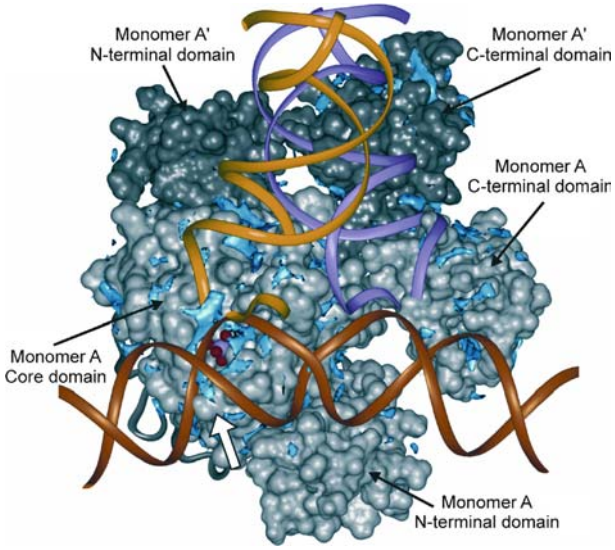

Figure 4. Results of the GRID phosphate affinity probe mapped onto Dimer A of the integration complex model contoured to −10 kcal mol^{-1}. The phosphate affinity probe (blue) is mapped on to Dimer A (grey/black). The host DNA and the U5 and U3 viral DNA strands are shown in orange, yellow and purple respectively. The white arrow indicates the position of S119 and N120. Dimer B is not shown for clarity.

plex: (1) the terminal four base pairs of the U5 and U3 viral LTRs, which constitute the major recognition sequences, are the same and (2) both ends of the viral DNA undergo the 3′-processing and strand transfer steps.

In the final orientation, the integration site of the 3′-hydroxyl of the U3 viral DNA end is positioned five base pairs from the integration site of the 3′-hydroxyl of the U5 viral DNA end, on the opposite strand of the host DNA across the bent major groove. Like Dimer A, there is good shape complementarity between a Connolly surface over Dimer B and the contour of the host DNA. An interesting feature of the model is the interwoven architecture formed between the two IN dimers and the two modelled viral DNA strands (Figure 2). The natural contour of the viral DNA strands means that the small flexible loop of Monomer B (residues 186-196, Loop186) is located close to the minor groove of the U5 viral DNA strand near positions 8–11. It is proposed that Loop186 will bind into the minor groove of the U5 viral LTR and therefore account for the specificity of nucleotides observed at these positions. Moreover, the *trans* relationship of this interaction agrees with the photocross-linking data (*Constraint 7a*). Similarly, Loop186 of Monomer A interacts with the minor groove of nucleotides 8 to 11 in the U3 viral LTR.

Positioning the C- and N-terminal domains

The C-terminal domains were arranged in the model to satisfy the photocross-linking data (*Constraint 7a*) and were optimised using the phosphate affinity maps generated with GRID (*Constraint 6*). In all cases, interactions between the C-terminal domains and the viral DNA strands occur in *trans* to the IN monomer providing the catalytic active site. Interestingly, the positions of the C-terminal domains in Monomers A and B are similar to the position of the C-terminal domain in the SIV IN crystal structure (1C6V) with respect to the core domain and the positions of the C-terminal domains in Monomers A′ and B′ are similar to the position of the C-terminal domain in the HIV-1 IN two domain crystal structure (1EX4) with respect to the core domain (data not shown). The use of the 1IHV structure to represent the C-terminal domain in this model means that coordinates for the flexible linker region (residues 210–219) between the core and C-terminal domains are not included in the current model. Despite this, the estimated position of the linker residues should satisfy the photocross-linking data. Similarly, residues 270–288, for which no structural data is available, are expected to satisfy the photocross-linking data in the current model.

The N-terminal domains of Monomers A and B were modelled to satisfy the photocross-linking data (*Constraint 7a*) and were placed near the integration sites of the two viral DNA ends so that residues 1–11 of Monomer A were near host DNA nucleotides 9 and 9′ and residues 1–11 of Monomer B were near host DNA nucleotides 14 and 14′. These positions differ from the 1K6Y crystal structure but could easily be accommodated by the flexible linker that joins the N-terminal and core domains. The contacts between the N-terminal domains of the two IN dimers are consistent with data showing that the N-terminal domain enhances multimerisation and activity of IN and stabilises the binding of the host DNA in the integration complex [40]. The interactions between the N-terminal domains and the host DNA in the model may also explain why a smaller host DNA substrate is required for the 3′-processing and strand transfer assays in the presence of a functional N-terminal domain compared to its absence [89, 90].

The N-terminal domains of Monomers A′ and B′ were arranged to best interact with the viral DNA without overlapping other parts of the model. These arrangements are speculative but are included for completeness of the model. The role of these two domains in the integration complex is unknown. These domains could be involved in interactions with other proteins in the integration complex like viral protein R, reverse transcriptase, barrier to autointegration factor or high mobility group protein A1 or may just provide structural support like the core domains of Monomers A′ and B′.

Discussion

The major difficulty in assembling the model of the integration complex is locating the viral and host DNA. Complementation and photocross-linking studies provide approximate constraints but these are not of sufficient resolution to propose detailed IN–DNA interactions, especially with respect to

the specific binding of the conserved CA dinucleotide in the IN active site. The best current source of information to locate the DNA strands is the crystal structure of Tn5 transposase/DNA and its structural and functional homology with IN. The core domains of IN and Tn5 transposase have a similar protein fold and several conserved active site residues but there are also some significant differences between these enzymes. For example, the homology between Tn5 transposase and IN does not extend to the region around the IN active site mobile loop (residues 140–149, Loop140). While both enzymes have flexible loops that flank the active site and interact with the viral/transposon DNA, the loop of Tn5 transposase is much larger and binds over the transposon strand (Figure 3). Loop140 is not large enough to do this. In addition, the Tn5 transposase active site contains a bound Mn^{2+} ion, which is coordinated between E326 and D97 (equivalent to E152 and D64 of HIV-1 IN) in contrast to IN, which binds the catalytic ion between D97 and D188 (equivalent to D64 and D116) (Figure 3). The differences between these active site residues and the mobile loops flanking these active sites can be attributed to differences in mechanism. The Tn5 transposase forms a hairpin intermediate between the 3′-processing and strand transfer steps, whereas IN does not. Despite these differences, the common active site residues T66, N155, K156 and K159 (T97, H329, K330 and K333 in Tn5 transposase) in these two enzymes suggests that there is sufficient similarity to use the Tn5 transposon as a model for the viral DNA in the integration complex.

The symmetry of the model means that interactions between the U3 viral LTR and IN mirror the interactions between the U5 viral LTR and IN (Table 4). This discussion describes only the interactions between IN and the U5 viral LTR.

Conserved CA and 5′-overhang

The conserved adenosine is located in the active site pocket, where the 3′-OH of the recessed U5 viral DNA is within hydrogen bonding distance of a water molecule coordinated to the Mg^{2+} ion (Figure 5a). The 5′-phosphate group of the adenosine forms hydrogen bonds with N155 and K159 and T66. The position of this phosphate group corresponds to a region of high phosphate affinity found using the GRID phosphate probe and occupies a similar position to the phosphate ion observed in the 1K6Y crystal structure [6]. A unique feature of an adenosine is the amino group on the purine ring. We propose that this amino group forms hydrogen bonds with the sidechain carboxylate of E152 thus accounting for specific

Table 4. Residues in HIV-1 IN that are (A) conserved or (B) conservatively substituted according to Conscore from 189 HIV-1 patient isolate sequences.

A	Residues
N-terminal domain	F1, L2, A8, H12, H16,[a] W19, P29, K34, C40, C43,[b] G47, E48
Core domain	H51, G52, Q53, P58, W61, Q62, D64, C65, H67, I73, V75, V77, E87, V88, A98, L102, A105, W108, P109, V110, H114, D116,[c] N117, G118, F121, A128, C130, W131, W132, G140, Y143, N144, P145, Q146, Q148, S149, E152,[d] L158, Q164, Q168, A169, E170, L172, A175, V180
C-terminal domain	F223, V225, Y227, R228, P233, W235, K236, G237, A239, W243, K244, G245, E246, G247, A248, K258, P261, R262, K266, Y271, K273, A281, Q284, E286
B	Residues
N-terminal domain	I5, D25, L28, I31,[e] V32, E35, I36
Core domain	V54, E69, H78, V79, Y83, E85, T93, I113, K127, V150, V151, I162, V165, D167, K173, I182, K188, I204, I217, I220
C-terminal domain	Y227, S230, R231, V249, V250, I251, D253, D256, I257, V259, V260, R263, I267, I268

[a]Substituted with an arginine in isolate A2D.KR.97.97KR004.
[b]Substituted to a glycine in isolates A2.CD.-.97CDKTB48 and A2G.CD.97.97CDKP58.
[c]Substituted to a alanine in the BF.VE.-.V62 isolate.
[d]Substituted to an alanine, a lysine or an aspartic acid in the B.AU.96.MBCC98, 06_cpx.SN.97.97SE1078 and B.AU.-.VH isolates respectively.
[e]52% isoleucine and 48% valine.

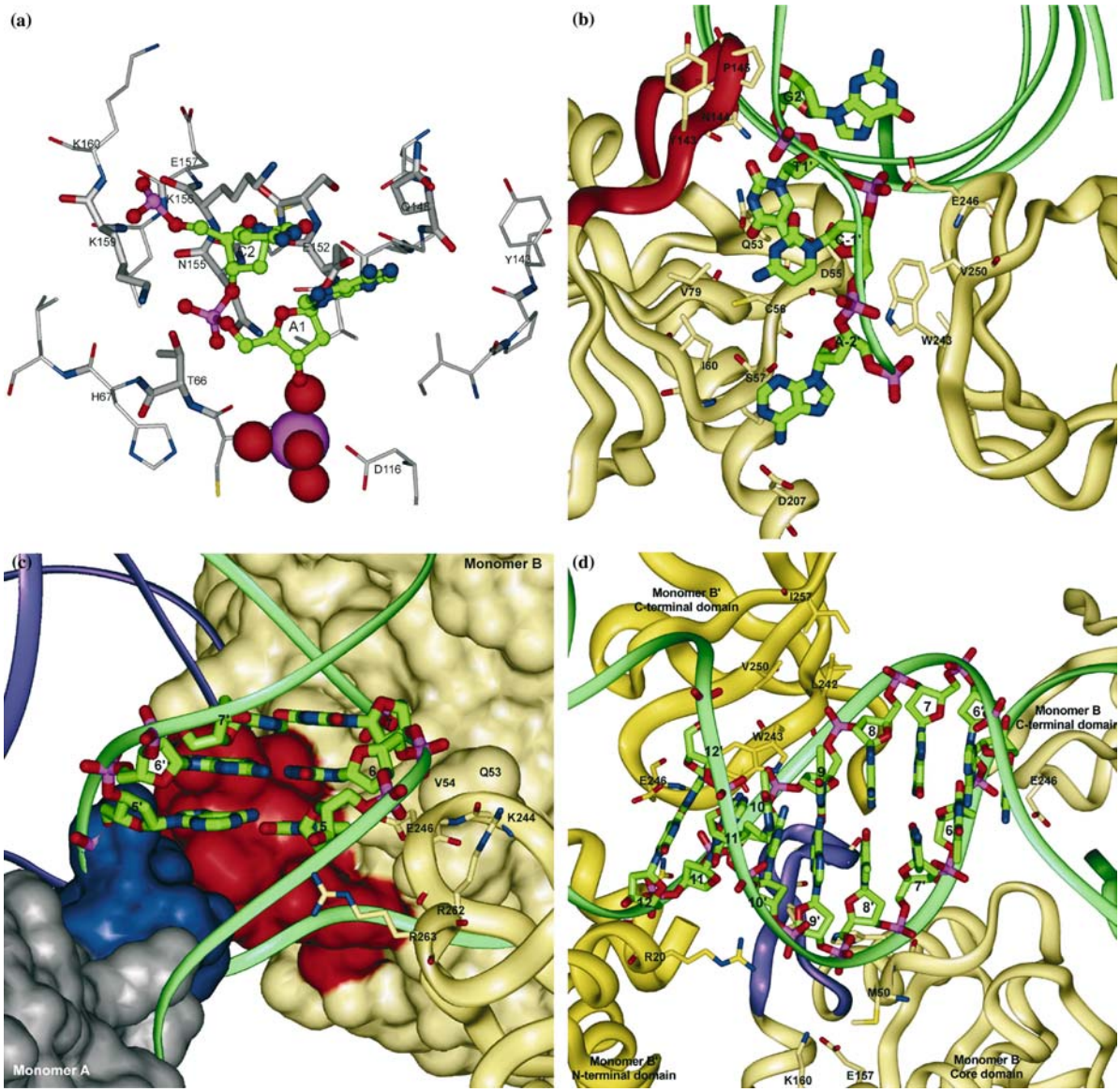

Figure 5. Interactions between the U5 viral LTR and IN. (a) Residues in Monomer A (grey) that interact with the conserved CA dinucleotide of the U5 viral LTR. The CA dinucleotide is shown in ball and stick (green), with residues contacting these bases are shown in bold. The Mg^{2+} and the four waters coordinated to it are shown as spheres. (b) Interactions between bases at positions −2′ to 2′ and Monomer B. Residues within 5 Å of these bases are shown as sticks, with other residues in the core and C-terminal domains of Monomer B (lemon) shown in ribbon. The mobile loop (residues 140–149) is red and the U5 viral DNA strand is green. Interactions with the host DNA and Monomer A are not shown for clarity. (c) Nucleotides at positions 5, 6 and 7 interacting with IN. The core domains of Monomers A (grey) and B (lemon) are represented by a Connolly surface. The active site mobile loops (residues 140–149, Loop140) of these two domains are coloured blue and red respectively. The C-terminal domain of Monomer B is shown in ribbon and residues interacting with the nucleotides at positions 5, 6 and 7 (green stick) are labelled. The U3 viral DNA strand (purple) is close to nucleotides 5′, 6′ and 7′ of the U5 viral LTR. D) Interactions between bases at positions 6 to 12 and IN. Interactions occur in *trans* to the Monomer A active site (not shown). The small flexible loop (residues 184–194, Loop186, purple) in Monomer B (lemon) binds into the minor groove of the U5 viral DNA between bases 8 and 11 and accounts for the observed specificity for these bases *in vitro* [37]. The N- and C-terminal domains of Monomer B′ (yellow) also interact with these bases.

recognition of this base. A similar suggestion was made by Heuer and Brown [4]. This proposal conflicts with a second metal ion binding site in IN, as the carboxylate group of E152 would not be available to bind the second metal ion. In the Tn5 transposase crystal structure, the terminal 3′-nucleotide of the transposon, which interacts with the Mn^{2+} ion, is guanosine and not adenosine; therefore, conclusions about the role of E152 in specifically binding the unique NH_2 cannot be drawn from the template structure.

The conserved cytosine (position 2) is situated in the active site, adjacent to K156, where the C4 carbonyl group of this base points towards the Q148 sidechain (Figure 5a). Carbon atoms in the K156 sidechain form hydrophobic interactions with the deoxyribose ring of the cytosine as well as with the guanosine at position 3. We expect that the sidechain nitrogen of K156 will hydrogen bond to the C2 carbonyl of the conserved cytosine. This proposal is supported by mutational data, which shows HIV-1 IN activity being significantly affected by mutations at this position, especially to acidic residues [62]. Q148 and S153 are also located near the conserved cytosine and either residue could potentially interact with this base. In addition, Q148 could also make specific interactions with the amino group at C2 of the guanosine (position 2′) as well, thus accounting for specific recognition of this base pair.

K156 is not absolutely conserved among the lentiviral integrases but is replaced by a histidine in HIV-2 IN and simian immunodeficiency virus (SIV) IN, an asparagine in equine infectious anaemia virus (EIAV) IN and feline immunodeficiency virus (FIV) IN and an arginine in avian sarcoma virus (ASV) IN. Mutation of K156 to a glutamic acid abolished HIV-1 replication in tissue culture [62, 91] but substitution to an alanine retained 60–84% replication activity [92]. The model can be used to explain these observations. Substitution of K156 with a small hydrophobic residue would retain non-specific interactions with the deoxyribose group but substitution with an acidic residue would cause repulsion of the cytosine base, thus affecting viral DNA binding. In contrast, polar residues at this position would still act as hydrogen bond donors thus stabilising interactions with the carbonyl group but we expect with a lower affinity compared to basic residues. A Q148N mutation maintained 3′-processing activity but was defective for strand transfer and mutation of Q148 to a leucine abolished 3′-processing and strand transfer activity *in vitro* [46, 93].

The guanosine at position 2′ maintains its hydrogen bonding pattern with the conserved cytosine and, as previously mentioned, hydrogen bonding interactions between the Q148 sidechain and the amino group at the C2 position of the guanosine would account for the specific recognition of this base by HIV-1 IN *in vivo*. In contrast, the conserved thymine at position 1′ is flipped out of the DNA helix, where it interacts with residues in Monomer B (Figure 5b). In this orientation, the thymine can bind specifically to Q53 and N144 of Monomer B. This base is also close to Y143. Esposito and Craigie mapped the GT dinucleotide (positions 2′ and 1′ respectively) to Y143 and Q148, which is consistent with their location in our model.

The terminal four bases of the 5′-end of the U5 viral DNA strand are shown in Figure 5b. The cytosine at position −1′ is close to D55 and C56 in Monomer B. The adenosine at position −2′ is close to S57 and D207 of Monomer B where the amino group on the purine ring could form hydrogen bonds with D207 and/or S57. Alternatively, this base could stack with W243 in the C-terminal domain of Monomer B. Mutation of the cytosine to an adenosine did not have any affect on the 3′-processing activity of IN but substitution with a thymine or a guanosine abolished all 3′-processing activity *in vitro* [37].

Nucleotides at positions 3 and 4

Our model places, the phosphodiester backbone of nucleotides 3 and 4 close to K156, K160 and E157, and nucleotides 3′ and 4′ near residues 146–151 of Monomer A and residues 256–262 of Monomer B. This location is consistent with the photocrosslinking data [36]. Residues 256–262 in Monomer B are suitably positioned to recognise bases at positions 3′ and 4′ but specific interactions between IN and bases at positions 3 and 4 were not observed. Residues 256–262 are highly conserved in the HIV-1 isolates. Esposito and Craigie showed that the wild type sequence of the terminal six nucleotides (i.e. positions -2 to 4) of the HIV-1 viral LTR are moderately important for 3′-processing activity [37], which is consistent with our model.

Nucleotides at positions 5, 6 and 7

The interactions between nucleotides at positions 5, 6 and 7 of the U5 viral LTR, Monomers A and B, and the U3 viral LTR are shown in Figure 5c. Contacts between these nucleotides and IN are with the phosphodiester backbone and R262, R263, K244, G247, Q53 and V54 of Monomer B. The E246 sidechain of Monomer B is positioned near the adenosine at position 5′ and satisfies the disulfide-bridge cross-linking experiments of Gao et al. [5] and the photocross-linking data [37]. The complementary nucleotides 5′, 6′ and 7′ are close to the phosphodiester backbone of the U3 viral LTR. The interactions of nucleotides at positions 5, 6 and 7 are also consistent with the mutational data of Esposito and Craigie [37], who found that substitutions in these nucleotides had no significant effect on 3′-processing activity.

Nucleotides at positions 8–14

Esposito and Craigie [37] reported that some substitutions in nucleotides at positions 8–11 in the U5 viral LTR significantly reduce the 3′-processing activity of HIV-1 IN. A small flexible loop comprising residues 186–196 (Loop186) of Monomer B is located close to these nucleotides (Figure 5d, purple). We propose that this loop binds into the minor groove of the viral DNA and makes specific interactions with these nucleotides. Residues in Loop186 are well conserved in HIV-1 IN and could account for the base specificity observed for these positions. Residues 246–250 and residue 20 of Monomer B′ interact with the phosphodiester backbone of the complementary nucleotides at positions 8′–11′. In addition, W243 of Monomer B′ is positioned suitably to insert between nucleotide bases at position 10′ and 11′. In contrast, the mutational studies showed no significant effect on 3′-processing activity with base substitutions at positions 12–14 [37]. Interactions between IN and these nucleotides comprise basic residues in the C-terminal domain of Monomer B′ and the phosphodiester backbone of the viral LTR in the model.

The host DNA

The preferred binding site of the integration complex on the host chromosome is determined by structural features [66–70] and not by recognition of a particular nucleotide sequence [69]. In our proposed model, residues in all three IN domains of Monomers A and B (Table 2) interact with the host DNA, predominantly through contacts with the phosphodiester backbone (Figure 4).

Proposed function of Loop140 (active site mobile loop)

It is well established that Loop140 is necessary for efficient IN function [94] but its exact role in the integration process is not known. Loop140 is flexible [95–98] and is poorly resolved in most crystal structures of the IN core domain. In cases where the electron density for this region has been traced, the reduced flexibility of the Loop140 is due to crystal contacts between adjacent IN dimers. For example, in the 1BL3 crystal structure the position of Loop140 in chain C is influenced by crystal contacts with the chain A/B dimer. Greenwald et al. [94] showed that mutating the glycine residues in Loop140 was detrimental to 3′-processing and strand transfer activity *in vitro* but not DNA binding. With the exception of P145 and Q148, mutation of other residues in Loop140 have not significantly reduced 3′-processing, strand transfer or disintegration activity *in vitro* [46, 83, 91, 93, 99, 100].

In the model, Loop140 of Monomers A and B are positioned across the major groove of the host DNA where they each interact with all three DNA strands, including the conserved CA/GT and the 5′-overhang (Figure 5c). Further to this, these loops interact with each other through contacts involving residues Y143, P145 and Q146 of both monomers. The experimentally observed five base pair spacing between the two integration sites of the viral DNA ends suggests that the viral DNA strands must be close to one another in the integration complex. We propose that Loop140 in both Monomers A and B act as barriers to help separate these negatively charged strands and therefore stabilise the integration complex. This proposal represents a unique role for Loop140 in integration and stabilisation of the integration complex.

Agreement with protein footprinting

The final model was compared to the findings of Dirac and Kjems [84], who investigated the

resistance of residues in HIV-1 IN to proteolytic attack upon DNA binding. Although these researchers used Mn^{2+} rather than the more specific Mg^{2+} metal cofactor in their experiments and observed similar footprinting patterns with and without a specific viral DNA substrate, there is good correlation between their findings and our model. The residues in HIV-1 IN that were strongly protected against proteolytic attack (K34, K111, E138, K185, K186, K188, D207 and E246) are buried in the integration complex model, except for E138 which is buried in Monomers A and B but exposed in Monomers A′ and B′. K273 is not resolved in our model but is expected to be buried through contacts with the core domain. Further to this, there was good correlation between residues that were weakly protected from proteolytic attack upon DNA binding and IN complex formation (K136, K156, K159, K160 and K215) and their location in our model. In this case, K136, K156, K159, K160 and K215 in Monomers A and B are involved in viral DNA binding but, in contrast, are exposed to the solvent in Monomers A′ and B′. Dirac and Kjems reported variable results for the susceptibility of K258 to proteolytic attack between repeated experiments [84]. In the current model, K258 of Monomers A′ and B′ interacts with the phosphodiester backbone of the viral DNA and K258 of Monomers A and B interacts with the host DNA phosphodiester backbone. In all cases, this residue is partially solvent exposed.

Amino acid conservation in HIV-1 IN and DNA interactions

One mechanism that allows HIV to evade an effective immune response is the high error rate in proviral DNA synthesis by reverse transcriptase. This leads to a large diversity in the genomic sequences of HIV isolated from infected individuals [101]. To compare the variability of amino acid residues in HIV-1 IN to the proposed binding positions of the viral and host DNA strands in the integration complex model, we scored amino acid conservation from 189 HIV-1 infected patient samples using Conscore and mapped the results onto the surface of the IN dimers in our model. The results for the U5 viral LTR and IN are shown in Figure 6.

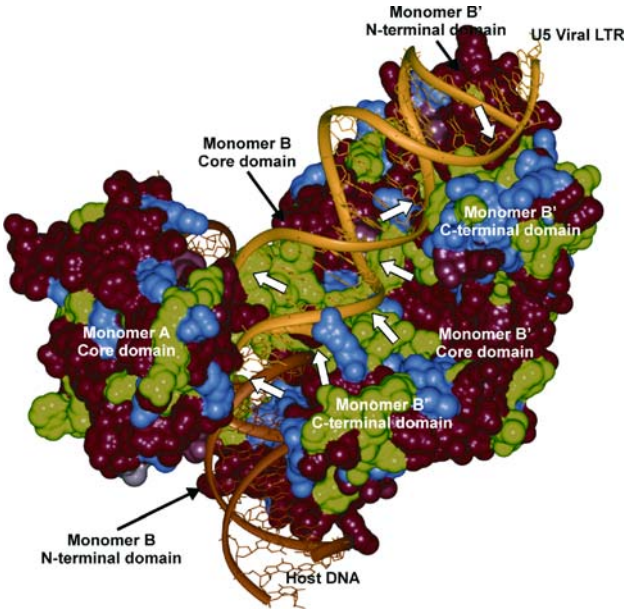

Figure 6. Visualisation of amino acid conservation from 189 HIV-1 isolates sequences using Conscore and mapped onto IN. Invariant residues are green and conservatively substituted residues are blue. Other residues are coloured from red to white in a sliding scale from 95% conservative substitution (red) to less than 50% conservative substitution (white). The U5 viral LTR is yellow and the host DNA is orange. The U3 viral LTR and other IN domains are not shown for clarity. White arrows indicate regions of good agreement between DNA interactions and high amino acid conservation.

The level of amino acid conservation in the HIV-1 IN sequences studied is high, with residues having an average Conscore value of 8 (i.e. >95% conservation). There is a high level of amino acid conservation between residues that are in contact with the viral and/or host DNA strands (Figure 6, white arrows). Amino acids in these regions are generally invariant or are completely conserved in the HIV-1 sequences. The highest concentration of invariant residues was in the IN active site, including residues in Loop140. This further supports proposals that the conserved CA dinucleotide of the viral LTR ends bind specifically to this region.

Comparison with other published IN models

A number of models of the integration complex [4–7] or of predicted IN–DNA contacts [8–14] have been described. Of these models, the coordinates of the Heuer and Brown [4], Gao et al. [5] and Podtelezhnikov et al. [7] models are available. Each of these models was built using a similar approach to that used in preparing our model (see *Constraints 1–5 and 7*). However, while each model proposes that two active sites are involved in the integration of the viral DNA ends into the host DNA, the gross geometry and the placement of the host and viral DNA strands differ in each model. The Heuer and Brown model was the first published and presents an integration complex containing four IN dimers. Like our model, it uses an interwoven DNA-protein architecture where only the terminal bases of the viral DNA interact with the monomer providing the active site. Bases further away interact with an adjacent monomer. This model does not contain positions for the N-terminal domains, which were not available at the time. The Gao et al. model is based on the 1EX4 crystal structure, which contains the C-terminal and core domains [10] and on site-specific cross linking. Like our model, this model posits that a pair of IN dimers is minimally involved in the integration complex but Gao et al. propose that the two IN dimers only interact through the N-terminal domains. The Podtelezhnikov et al. model [7] is based on the synaptic Tn5 transposase complex proposed by Davies et al. [57], rather than the orientation of the Tn5 transposon as in our model. This model contains an IN tetramer where the two IN dimers interact along a large interface on the same face of the host DNA. In our model, contacts between the N-terminal domains and the core domains of Dimers A and B occur on the opposite face of the target DNA (Figure 2).

Our model of the integration complex, based on the Tn5 transposase/transposon crystal structure, is consistent with the available experimental data and shows good complementarity between IN–DNA interactions and the observed patterns of amino acid conservation in HIV-1 isolates. Several inferences can be made from the model: E152 is involved in the specific recognition of the conserved adenosine of the viral DNA ends and that Q148 and K156 are important for recognition of the conserved cytosine and its complementary guanosine base. We propose that Loop186 binds into the minor groove one turn away from the conserved CA nucleotides in *trans* to the catalytic active site and that Loop140 of Monomers A and B are each involved in binding to the host DNA and both viral DNA ends. This model clarifies the role of the core domain in host DNA target site selection.

Acknowledgments

J.W. was supported by an Australian Postgraduate Award provided by the Australian Government. The authors thank Amrad Corporation Ltd for supporting the research and David Rhodes for his critical reading of this manuscript.

References

1. Brown, P.O., In Retroviruses, Coffin J.M., Hughes S.H. and Varmus H.E. (Eds.), Integration, Cold Spring harbour Press, Plainview, NY, 1997, 161 pp.
2. Hindmarsh, P. and Leis, J., Microbiol. Mol. Biol. Rev., 63 (1999) 836.
3. Chow, S.A., Methods, 12 (1997) 306.
4. Heuer, T.S. and Brown, P.O., Biochemistry, 37 (1998) 6667.
5. Gao, K., Butler, S.L. and Bushman, F., EMBO J., 20 (2001) 3565.
6. Wang, J.Y., Ling, H., Yang, W. and Craigie, R., EMBO J., 20 (2001) 7333.
7. Podtelezhnikov, A.A., Gao, K., Bushman, F.D. and McCammon, J.A., Biopolymers, 68 (2003) 110.
8. Yang, Z.N., Mueser, T.C., Bushman, F.D. and Hyde, C.C., J. Mol. Biol., 296 (2000) 535.
9. Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V.,

Emini, E.A., Hazuda, D. and Kuo, L.C., J. Mol. Biol., 296 (2000) 521.
10. Chen, J.C., Krucinski, J., Miercke, L.J., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D. and Stroud, R.M., Proc. Natl. Acad. Sci. USA., 97 (2000) 8233.
11. Perryman, A.L. and McCammon, J.A., J. Med. Chem., 45 (2002) 5624.
12. De Luca, L., Pedretti, A., Vistoli, G., Letizia Barreca, M., Villa, L., Monforte, P. and Chimirri, A., Biochem. Biophys. Res. Commun., 310 (2003) 1083.
13. Chiu, T.K. and Davies, D.R., Curr. Top. Med. Chem., 4 (2004) 965.
14. Marchand, C., Johnson, A.A., Karki, R.G., Pais, G.C., Zhang, X., Cowansage, K., Patel, T.A., Nicklaus, M.C., Burke, T.R. Jr. and Pommier, Y., Mol. Pharmacol., 64 (2003) 600.
15. Bushman, F.D., Fujiwara, T., Craigie, R., Science (Washington, DC), 249 (1990) 1555.
16. Bushman, F.D. and Craigie, R., Proc. Natl. Acad. Sci. USA., 88 (1991) 1339.
17. Engelman, A., Mizuuchi, K. and Craigie, R., Cell, 67 (1991) 1211.
18. Acel, A., Udashkin, B.E., Wainberg, M.A. and Faust, E.A., J. Virol., 72 (1998) 2062.
19. Brin, E., Yi, J., Skalka, A.M. and Leis, J., J.Biol. Chem., 275 (2000) 39287.
20. Engelman, A. and Craigie, R., J. Virol., 69 (1995) 5908.
21. Hansen, M.S., Smith, G.J., Kafri, T., Molteni, V., Siegel, J.S. and Bushman, F.D., Nat. Biotechnol., 17 (1999) 578.
22. Hwang, Y., Rhodes, D. and Bushman, F., Nucl. Acids Res., 28 (2000) 4884.
23. Chow, S.A., Vincent, K.A., Ellison, V. and Brown, P.O., Science (Washington, D.C.), 255 (1992) 723.
24. Bujacz, G., Jaskoski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M., J. Mol. Biol., 253 (1995) 333.
25. Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R., Science (Washington, D.C.), 266 (1994) 1981.
26. Goldgur, Y., Dyda, F., Hickman, A.B., Jenkins, T.M., Craigie, R. and Davies, D.R., Proc. Natl. Acad. Sci. USA., 95 (1998) 9150.
27. Goldgur, Y., Craigie, R., Cohen, G.H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. and Davies, D.R., Proc. Natl. Acad. Sci. USA., 96 (1999) 13040.
28. Lubkowski, J., Yang, F., Alexandratos, J., Wlodawer, A., Zhao, H., Burke, T.R. Jr., Neamati, N., Pommier, Y., Merkel, G. and Skalka, A.M., Proc. Natl. Acad. Sci. USA., 95 (1998) 4831.
29. Lubkowski, J., Dauter, Z., Yang, F., Alexandratos, J., Merkel, G., Skalka, A.M. and Wlodawer, A., Biochemistry, 38 (1999) 13512.
30. Maignan, S., Guilloteau, J.P., Zhou-Liu, Q., Clement-Mella, C. and Mikol, V., J. Mol. Biol., 282 (1998) 359.
31. Lodi, P.J., Ernst, J.A., Kuszewski, J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M. and Gronenborn, A.M., Biochemistry, 34 (1995) 9826.
32. Eijkelenboom, A., Sprangers, R., Hard, K., Lutzke, R.A.P., Plasterk, R.H.A., Boelens, R. and Kaptein, R., Proteins: Struct., Funct., Genet., 36 (1999) 556.
33. Eijkelenboom, A., Vandenent, F.M.I., Vos, A., Doreleijers, J.F., Hard, K., Tullius, T.D., Plasterk, R.H.A., Kaptein, R. and Boelens, R., Curr. Biol., 7 (1997) 739.
34. Cai, M.L., Zheng, R.L., Caffrey, M., Craigie, R., Clore, G.M. and Gronenborn, A.M., Nat. Struct. Biol., 4 (1997) 839.
35. Cai, M.L., Huang, Y., Caffrey, M., Zheng, R.L., Craigie, R., Clore, G.M. and Gronenborn, A.M., Protein Sci., 7 (1998) 2669.
36. Heuer, T.S. and Brown, P.O., Biochemistry, 36 (1997) 10655.
37. Esposito, D. and Craigie, R., EMBO J., 17 (1998) 5832.
38. Vink, C., Oude Groeneger, A.M. and Plasterk, R.H., Nucl. Acids Res., 21 (1993) 1419.
39. Bushman, F.D., Engelman, A., Palmer, I., Wingfield, P. and Craigie, R., Proc. Natl. Acad. Sci. USA., 90 (1993) 3428.
40. Zheng, R., Jenkins, T.M. and Craigie, R., Proc. Natl. Acad. Sci. USA., 93 (1996) 13659.
41. van den Ent, F.M., Vos, A. and Plasterk, R.H., J. Virol., 73 (1999) 3176.
42. Ellison, V., Gerton, J., Vincent, K.A. and Brown, P.O., J. Biol. Chem., 270 (1995) 3320.
43. Rice, P.A. and Baker, T.A., Nat. Struct. Biol., 8 (2001) 302.
44. Engelman, A. and Craigie, R., J. Virol., 66 (1992) 6361.
45. Leavitt, A.D., Shiue, L. and Varmus, H.E., J. Biol. Chem., 268 (1993) 2113.
46. Van Gent, D.C., Oude Groeneger, A.A.M. and Plasterk, R.H.A., Proc. Natl. Acad. Sci. USA., 89 (1992) 9598.
47. Cannon, P.M., Wilson, W., Byles, E., Kingsman, S.M. and Kingsman, A.J., J. Virol., 68 (1994) 4768.
48. Engelman, A., Englund, G., Orenstein, J.M., Martin, M.A. and Craigie, R., J. Virol., 69 (1995) 2729.
49. Engelman, A., Hickman, A.B. and Craigie, R., J. Virol., 68 (1994) 5911.
50. Lutzke, R.A., Vink, C. and Plasterk, R.H., Nucl. Acids Res., 22 (1994) 4125.
51. Mumm, S.R. and Grandgenett, D.P., J. Virol., 65 (1991) 1160.
52. Eijkelenboom, A., Lutzke, R.A.P., Boelens, R., Plasterk, R.H.A., Kaptein, R. and Hard, K., Nat. Struct. Biol., 2 (1995) 807.
53. Zhu, K., Dobard, C. and Chow, S.A., J. Virol., 78 (2004) 5045.
54. Hehl, E.A., Joshi, P., Kalpana, G.V. and Prasad, V.R., J. Virol., 78 (2004) 5056.
55. Mack, J.P.G. and Theochem, J., Mol. Struct., 423 (1998) 41.
56. Davies, D.R., Braam, L.M., Reznikoff, W.S. and Rayment, I., J. Biol. Chem., 274 (1999) 11904.
57. Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. and Rayment, I., Science (Washington, D.C.), 289 (2000) 77.
58. Haren, L., Ton-Hoang, B. and Chandler, M., Annu. Rev. Microbiol., 53 (1999) 245.
59. Reznikoff, W.S., Mol. Microbiol., 47 (2003) 1199.
60. Rice, P., Craigie, R. and Davies, D.R., Curr. Opin. Struct. Biol., 6 (1996) 76.
61. Williams, T.L. and Baker, T.A., Science (Washington, D.C.), 289 (2000) 73.
62. Jenkins, T.M., Esposito, D., Engelman, A. and Craigie, R., EMBO J., 16 (1997) 6849.
63. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
64. Holm, L. and Sander, C., Nucl. Acids Res., 27 (1999) 244.
65. Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Wolinsky, S. and Korber, B. (Eds.), HIV Sequence Compendium 2001: Theoretical Biology and

Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM (2001).
66. Pruss, D., Bushman, F.D. and Wolffe, A.P., Proc. Natl. Acad. Sci. USA., 91 (1994) 5913.
67. Pruss, D., Reeves, R., Bushman, F.D. and Wolffe, A.P., J. Biol. Chem., 269 (1994) 25031.
68. Pryciak, P.M. and Varmus, H.E., Cell, 69 (1992) 769.
69. Bor, Y.C., Miller, M.D., Bushman, F.D. and Orgel, L.E., Virology, 222 (1996) 283.
70. Katz, R.A., DiCandeloro, P., Kukolj, G. and Skalka, A.M., J. Biol. Chem., 276 (2001) 34213.
71. Jenkins, T.M., Engelman, A., Ghirlando, R. and Craigie, R., J. Biol. Chem., 271 (1996) 7712.
72. Andrake, M.D. and Skalka, A.M., J. Biol. Chem., 270 (1995) 29299.
73. Deprez, E., Tauc, P., Leh, H., Mouscadet, J.F., Auclair, C. and Brochon, J.C., Biochemistry, 39 (2000) 9275.
74. Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J.C., LeCam, E., Coulaud, D., Auclair, C. and Mouscadet, J.F., Biochemistry, 39 (2000) 9285.
75. Asante-Appiah, E. and Skalka, A.M., Adv. Virus Res., 52 (1999) 351.
76. Asante-Appiah, E. and Skalka, A.M., J. Biol. Chem., 272 (1997) 16196.
77. Vink, C., Lutzke, R.A. and Plasterk, R.H., Nucl. Acids Res., 22 (1994) 4103.
78. van Gent, D.C., Vink, C., Groeneger, A.A. and Plasterk, R.H., EMBO J., 12 (1993) 3261.
79. Engelman, A., Bushman, F.D. and Craigie, R., EMBO J., 12 (1993) 3269.
80. Bao, K.K., Wang, H., Miller, J.K., Erie, D.A., Skalka, A.M. and Wong, I., J. Biol. Chem., 278 (2003) 1323.
81. Yang, W. and Steitz, T.A., Structure, 3 (1995) 131.
82. Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J. and Skalka, A.M., Nucl. Acids Res., 19 (1991) 851.
83. Sayasith, K., Sauve, G. and Yelle, J., Mol. Cells, 10 (2000) 525.
84. Dirac, A.M. and Kjems, J., Eur. J. Biochem., 268 (2001) 743.
85. Connolly, M.L., Science (Washington, D.C.), 221 (1983) 709.
86. Engelman, A., Liu, Y., Chen, H.M., Farzan, M. and Dyda, F., J. Virol., 71 (1997) 3507.
87. Harper, A.L., Skinner, L.M., Sudol, M. and Katzman, M., J. Virol., 75 (2001) 7756.
88. Appa, R.S., Shin, C.G., Lee, P. and Chow, S.A., J. Biol. Chem., 276 (2001) 45848.
89. Vincent, K.A., Ellison, V., Chow, S.A. and Brown, P.O., J. Virol., 67 (1993) 425.
90. Chow, S.A. and Brown, P.O., J. Virol., 68 (1994) 7869.
91. Engelman, A., Adv. Virus Res., 52 (1999) 411.
92. Wiskerchen, M. and Muesing, M.A., J. Virol., 69 (1995) 376.
93. Gerton, J.L., Ohgi, S., Olsen, M., Derisi, J. and Brown, P.O., J. Virol., 72 (1998) 5046.
94. Greenwald, J., Le, V., Butler, S.L., Bushman, F.D. and Choe, S., Biochemistry, 38 (1999) 8892.
95. Barreca, M.L., Lee, K.W., Chimirri, A. and Briggs, J.M., Biophys. J., 84 (2003) 1450.
96. Lins, R.D., Briggs, J.M., Straatsma, T.P., Carlson, H.A., Greenwald, J., Choe, S. and McCammon, J.A., Biophys. J., 76 (1999) 2999.
97. Weber, W., Demirdjian, H., Lins, R.D., Briggs, J.M., Ferreira, R. and McCammon, J.A., J. Biomol. Struct. Dyn., 16 (1998) 733.
98. Ni, H., Sotriffer, C.A. and McCammon, J.A., J. Med. Chem., 44 (2001) 3043.
99. Pommier, Y., Pilon, A.A., Bajaj, K., Mazumder, A. and Neamati, N., Antiviral Chem. Chemother., 8 (1997) 463.
100. van den Ent, F.M., Vos, A. and Plasterk, R.H., J. Virol., 72 (1998) 3916.
101. McGrath, K.M., Hoffman, N.G., Resch, W., Nelson, J.A. and Swanstrom, R., Virus Res., 76 (2001) 137.