Taylor & Francis
Taylor & Francis Group

HIV-1 integrase–DNA interactions investigated by molecular modelling

C. Fenollar-Ferrer[1], V. Carnevale[1], S. Raugei and P. Carloni*

International School for Advanced Studies and CNR-INFM Democritos, Trieste, Italy

HIV-1 integrase is the viral enzyme responsible for the insertion of the viral DNA into the host cell chromosome. This process occurs through two distinct biochemical reactions: the $3'$-processing of the viral DNA and the transesterification reaction. Because experimental structural information on the reaction intermediate is not available, several molecular models have been developed. Unfortunately, none of the models of the enzyme–substrate complex is fully consistent with the available molecular biological data. We have constructed a new theoretical model based on mutagenesis experiments and cross-linking data, using a relatively accurate computational setup. The structural features of the model along with its limitations are discussed here.

Keywords: HIV-1 integrase; molecular modelling; molecular dynamics; synaptic complex; protein–DNA interactions

Introduction

HIV-1 integrase plays an essential role in the retrovirus' life cycle [11] and it is an established target of anti-AIDS therapy [1,50]. The enzyme inserts the reverse-transcribed viral DNA (vDNA, hereafter) into the host cell genome [51]. The process occurs via two nucleophilic reactions [56]: (i) a hydrolysis, performed by a Mg^{2+}-activated water molecule, of dinucleotides (conserved across the retroviruses family to which HIV-1 belong [11]) at each terminus of vDNA ($3'$-processing); and (ii) a transesterification, in which the two $3'$-hydroxyl groups resulting from step (i) perform a nucleophilic attack on the two $5'$-ends of the host DNA (hDNA, hereafter; [55]). The two $5'$-insertion sites within hDNA are located 5 bp apart [55].

The integrase can exists as monomer (IN), dimer (IN2) and dimer of dimers (IN4)[2]. IN consists of three functional domains, the N-terminal domain (NTD, residues 1–49), the core domain (CD, residues 50–186) and the carboxy-terminal domain (CTD, residues 213–288). So far, the structures of individual domains [5,9,20], as well as those of the NTD/CD and CD/CTD monomers and the CTD/CD dimer [9], have been solved by X-ray crystallography. The available structural information (Figure 1) shows that NTD consists of a bundle of orthogonal α-helices featuring an invariant His-His-Cys-Cys (HHCC) zinc-binding motif which is involved in interactions with other subunits in IN oligomers and also with vDNA ends [3]. The Zn ion promotes multimerization and enhances the enzymatic activity [10]. In contrast, CD is a α–β-double-layer sandwich. Finally, CTD is a β-roll architecture involved in non-specific DNA binding and IN oligomerization [11].

The active site – the same for both reactions – contains a Mg^{2+} ion[3] bound to a three negatively charged residues[4] [21,34,44,47] which are located within the β_1, β_4 and α_5 motifs

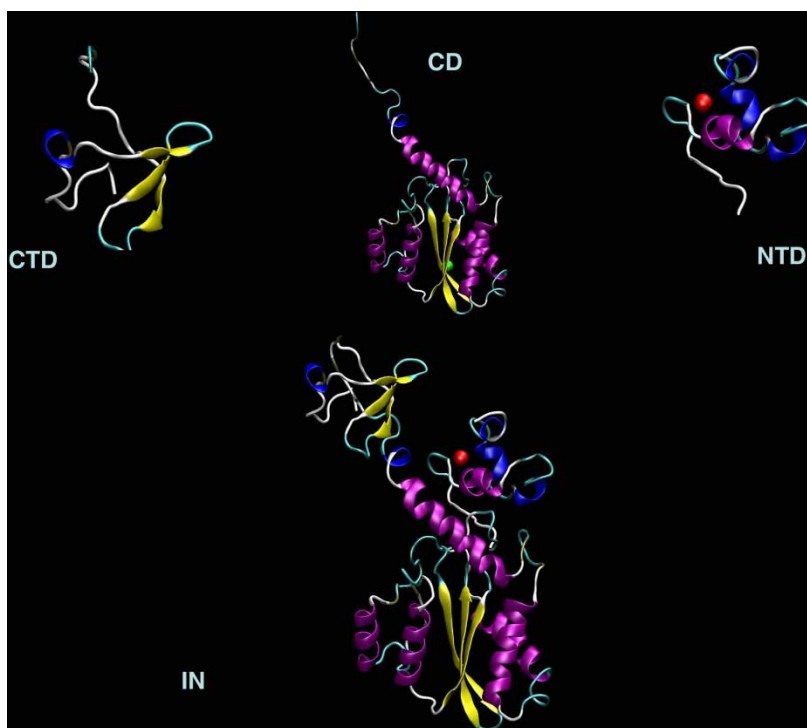---

*Corresponding author. Email: carloni@sissa.it

Figure 1. Structural domains of IN as obtained by X-ray crystallography [9,57]: NTD, CD and CTD are represented as cartoons. Available in colour online.

[11]. However, one more Mg^{2+} ion might bind in the presence of vDNA and hDNA. Indeed, two Mg^{2+} ions were found in the active site of the structurally related Tn5 transposase/DNA complex, for which the X-ray structure is available [52]. Both enzymes belong to the polynucleotidyl transferase family and they share mechanistic similarities and perform similar functions [46]. Thus, one or two Mg^{2+} ion(s) may activate the two nucleophiles in the two reaction steps – the water molecule and the $3'$-hydroxyl moiety – by binding to them [2].

The most important vDNA feature for integration is the CA/TG dinucleotide pair located 3 and 4 bp from each vDNA end within the long-terminal repeats (LTR; [30]). Although no structure of the IN/vDNA complex has been determined, it has been shown by mutagenesis that V72, S153, K160, I161, G163, V165, H171 and L172 confer IN specificity for LTR termini [8].

The function of the enzyme depends on its oligomerization state. Monomeric IN is not catalytically active [28]. IN2 is the most competent state for the $3'$-processing step *in vivo* [29] as well as *in vitro* [28]. Based on structural considerations IN4 is expected to perform the transesterification step [55]. Indeed, for step (ii) the enzyme must bind simultaneously to both double stranded vDNA and hDNA (forming the so-called synaptic complex). This can be best achieved by IN4 which provides the required 5 bp spacing between the insertion sites (~ 15 Å; [11]). In contrast, the dimer has the two active sites too far apart as seen in the X-ray structure of the NTD/CD dimer [57] in which the active sites are more than 30 Å apart. Thus, it is likely that IN2 catalyzes $3'$-processing whilst IN4 performs the transesterification step.

So far, only a model of the IN2 reaction intermediate of step (i) in which one vDNA strand was hydrolized has been reported [15]. This model, which features two Mg^{2+} ions in the active

site, is consistent with cross-linking experiments, which indicate that Y134 binds to vDNA [24]. However, this model may not be appropriate for the intermediate of step (ii) as the binding of further two IN monomers and a vDNA molecule (required for the integration reaction), is expected to cause large conformational changes.

In addition, different models of the IN4-based reaction intermediates have been reported, representing the species after $3'$-processing and either before [8,15,58], or after [27,30,33,43,59] hDNA binding (Types I and II models, respectively). Both model types contain IN4 in complex with one or two vDNA fragments while type II models also include hDNA.

As for Type I models, Chen *et al.* [8] and Wang *et al.* [58] constructed IN4 in complex with two vDNA duplexes. Both groups suggested that the dimer–dimer interface (IN2/IN2 hereafter) is stabilized by four salt bridges between E11 and K186, and D25 and K188, respectively. This proposal is consistent with *in vivo* mutagenesis experiments [3], which suggest that these residues are crucial for IN oligomerization. However, in the model of Wang *et al.*, the distance between catalytic sites is 40 Å, a value which is much larger than that required to accommodate 5 bp from hDNA [55]. In addition, the model, by lacking the C-terminus region (residue 271–288) does not include residue K273, which has been shown to bind DNA [6]. Furthermore, in this model K159, K186, K187, K188 and K211 do not interact with vDNA despite conclusive experimental evidence for such interactions [3]. Finally, this model features no Mg^{2+} ions in the active site, in spite of the fact that the ion is required for the function [11].

In comparison, the model of Chen *et al.*, is consistent with experimental results such as mutations that alter LTR specificity [8]. The model also correctly predicted that amino acids V72, S153, K160, I161, G163, V165, H171 and L172 are in contact with vDNA since experimental evidence shows that these residues determine the specificity for LTR termini [8]. This model is also fairly similar to the structure of Tn5 transposase/DNA complex [52]. The latter is considered as a good model for integrase/DNA complexes. However, although Chen's model is arguably more reliable than Wang's, it also lacks the C-terminus region (residues 271–288; [8,58]).

As for Type II models, Wielens *et al.* [59] constructed IN4 in complex with two vDNA's and one hDNA, in the presence of a Mg^{2+} ion inside the active site. This model is consistent with the data from protein foot-printing experiments in which the resistance of IN polypeptide to proteolytic attack was investigated [17]. However, as in the case of Wang's model, it does not take into account the experimentally determined interactions between residues K159, K186, K187, K188, K211 and vDNA [3].

Karki *et al.* [33] proposed several models of the IN4/vDNA/hDNA complex, in which either one or two Mg^{2+} ions are present in the active site. Although, these models are consistent with the protein foot-printing experiments [17], they cannot conclude which model represents the true synaptic complex. In fact, they point out that further experiments are necessary to exclude several degrees of freedom in the modelling. Finally, Podtelezhnikov *et al.* [43], Gao *et al.* [27] and Heuer *et al.* [30] constructed IN4 in complex with two vDNAs and one hDNA in the absence of Mg^{2+} ions. In these models, F185 is found to be at the tetramerization interface, a result which is consistent with the experimental observation of a replication defective phenotype for the F185K virus mutant [23]. Also in these models the C-terminus region is missing. In addition, in these models the distance between catalytic sites is much larger than that required to accommodate 5 bp of hDNA (it spans from 20 to 60 Å; [27,30,43]). Finally, Mg^{2+} is absent, although it is required for catalysis.

We conclude that at present Type II (or synaptic complex) models are not consistent with main structural features which can be derived by experimental studies such as the correct distance between the active sites and the interaction between K159, K186, K187, K188, K211 and vDNA.

Prompted by this fact, we have constructed a novel synaptic complex model, which attempts to overcome many of the above issues. We approach this goal by two strategies. The first is homology modelling of IN using the crystal structure of Tn5 transposase/DNA complex as template (pdb code: 1mus; [14,52]). This approach takes into account possible conformational changes caused by DNA binding as well as includes two Mg^{2+} ions. Indeed, the presence of a second Mg^{2+} ions opens the active site cavity of Tn5 transposase and deforms vDNA structure [14]. It is not known, however, whether the rearrangements of IN are the same as those observed for Tn5. The second approach is based on docking the two DNA fragments onto IN4 model, which in turn is obtained by using the available structural information on IN and IN2 [8]. This approach has the advantage that uses directly HIV-1 IN structure and provides as a bonus the structure of IN4 in the DNA-free state. However, this model would include just one Mg^{2+} ion and would not consider possible large conformational changes caused by vDNA binding. Here, we present the model based on the latter approach. The model resulting from the first approach will be subject of a future publication. We focus on IN4−v5′-TGTGGAAAATC-TCTAGCA-3′−h-5′-TTGAGCCGCAAGCGCCTCGACGCGCAGCCGAT-3′ complex (IN4−DNA hereafter); the vDNA sequences are those of the viral LTR's, while for the hDNA the same sequence as in Ref. [59] has been used. In this regard, we remark that due to the lack of specificity of IN for the target DNA, the choice of the hDNA sequence is not crucial in the modelling.

Our model includes important residues which were missing from the previous models [8,27,30,43,57], it is relaxed by MD for longer time and it uses a rather accurate parametrization for the Zn binding site. Comparison is made with the model of Chen *et al.* [8], which at the present is more reliable than Type 2 models.

Methods

Our procedure shares several similarities with that of Chen *et al.* [8]. First, we build IN, then IN4 and finally the synaptic complex.

Full-length IN. The X-ray structure of full-length IN has not been determined as yet. Here, we built it based on the available X-ray structures of the CD/CTD dimer fragment ([8]; pdb code: 1ex4) and NTD/CD monomer ([57]; pdb code: 1k6y). To this end, the CD's from the two structures were superimposed and the structure of the CTD from the CD/CTD fragment was then added to that of the NTD/CD fragment. The resulting model lacks the G47-D55 and G140-Q148 loops. These loops were constructed using the MODELLER program [25] by generating 3000 distinct loop conformations and retaining the most representative one[5] (Figure 2).

Full length IN4. Four monomers were superimposed onto NTD/CD tetramer X-ray structure (PDB id 1k6y) obtaining a dimer of dimers: hereafter subunits A/B and C/D refer to those constituting the first and the second dimer, respectively. Due to the symmetry of the model, the contacts established between A and B are equivalent to those established between C and D. The model was refined using the HADDOCK program [16,18] by performing a rigid body energy minimization, followed by a simulated annealing MD simulation using the CHARMM force field [26] in which the dynamics of the isolated protein was simulated *in vacuo*. In this MD calculations, only the torsional angles of the side chains at the interface were allowed to vary. The resulting structural model is shown in Figure 2.

IN4−vDNA. Two 5′-TGTGGAAAATCTCTAGCA-3′ oligonucleotides (double strand), representing the two termini of vDNA LTR fragments, were docked onto IN4 following the same protocol as adopted by Chen *et al.* The resulting IN4•vDNA model underwent 2 ns MD simulation in aqueous solution.[6]
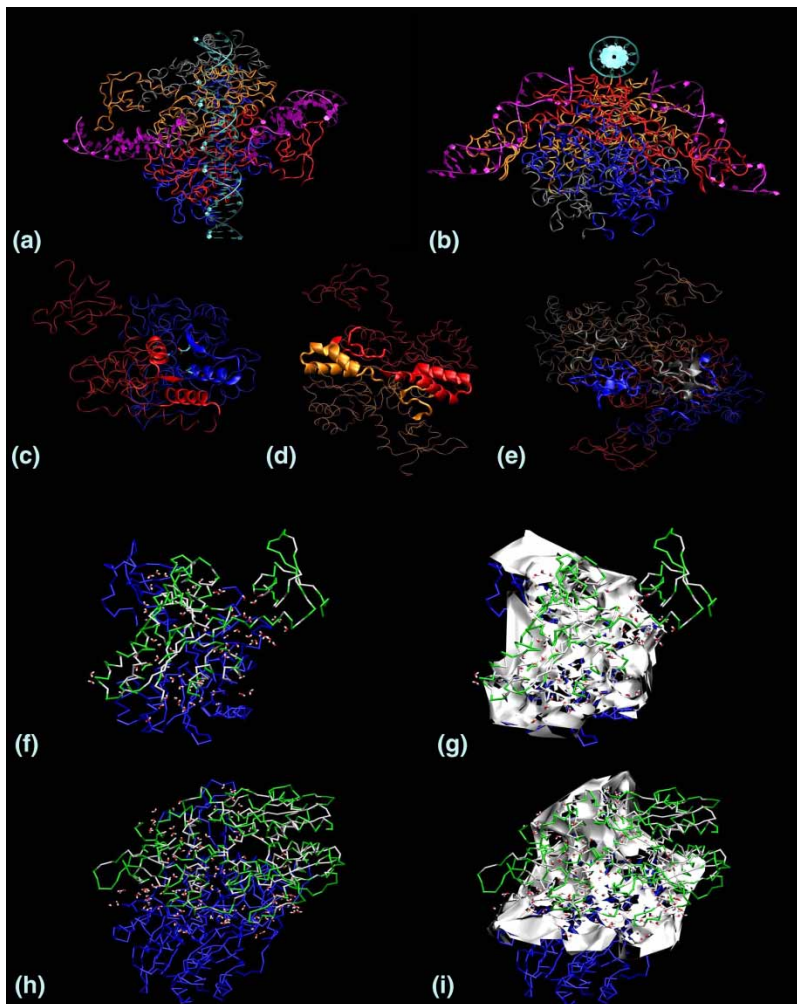
Figure 2. Integrase interactions in the synaptic complex. The four IN4 subunits are depicted by different colours: red, blue, gray and orange are used for chains A, B, C and D, respectively; purple (cyan) cartoons represent the bound viral (host) DNA. Front (a) and bottom (b) views of the whole IN4−vDNA−hDNA complex. (c)−(e) Interactions among IN subunits in the synaptic complex, protein−protein interfaces are highlighted as cartoons. (c) IN/IN interface between chains A and B featuring the α/β−α/β structural motif. (d) and (e) IN2/IN2 interactions involving chains A and D (d) and chains B and C (e); secondary structural elements defining regions 1 (d) and region 2 (e), are highlighted. (f)−(i) Hydration of IN/IN and IN2/IN2 interfaces, the protein backbone of one interacting unit is coloured in green (white) according to the hydrophilic (hydrophobic) character of each amino-acid. Water molecules located within 3 Å from both the IN units are also reported, along with the contact surface (for the sake of clarity they are separately portrayed) for IN/IN interface (f) and (g) and IN2/IN2 (h) and (i) interfaces.

The 5′-TTGAGCCGCAAGCGCCTCGACGCGCAGCCGAT-3′ hDNA (double strand) was built using the NUCGEN module of AMBER9 package assuming a B conformation. It was manually docked onto the IN4−vDNA structural model taking into account the following constraints: (i) 5 bp are located between the two non-buried IN4−vDNA active sites; and (ii) the oxygen atoms of the two phopsho-diester bonds at the cleavage site chelate the Mg^{2+} ion, according to the accepted reaction mechanism described in Ref. [2]. The resulting IN4−vDNA−hDNA

structural model (IN4–DNA complex, hereafter) was energy minimized using AMBERff99 force field [7]. Our procedure is different from the previous ones. Indeed, Chen *et al.* [8] and Karki *et al.* [33] used a different force field (in particular for the Zn fingers) and shorter simulation time-scale, whilst Wielens *et al.* did not carry any MD simulations at all [59].

Computational alanine scanning mutagenesis calculations of IN4 were performed on the FOLDX web server [49] and were used to detect 'hot spots' of protein–protein interactions, *i.e.* identify residues involved in highly stabilizing interactions at subunit interfaces.

Results and discussion

IN4

Here, we discuss the structure of IN4 in the free state and in the complex with vDNA and hDNA (*i.e.* the synaptic complex). Comparison is made with the IN4 and IN4–vDNA model of Chen *et al.* [8]. Overall, our model is similar to the latter, although it has been obtained using a different computational approach (see Methods). The RMSD of the C_α's between Chen's model and ours is only 3.8 Å – a small value considering the size of the system. However, it features a different conformation of loops G47-D55 and G140-Q148 and a bending of CTD toward the CD. This causes the formation of a deeper crevice between these two domains, which is able to accommodate nicely vDNA (Figure 2(a) and (b)).

Next, we present a structural and energetic analysis of the main feature of protein–protein interactions.

IN/IN dimerization interface

The interface is characterized by an α/β−α/β structural motif (Figure 2(c)). The buried surface (calculated for the dimer constituted by subunits A and B) is $1.23 \times 10^4 \text{ Å}^2$ (to be compared to the values $1.27 \times 10^4 \text{ Å}^2$ and $1.34 \times 10^4 \text{ Å}^2$ of the total solvent accessible area of the two monomers), and formed in large proportion (60%) by hydrophilic residues hydrated by 100 water molecules (corresponding to ∼0.01 water molecules/Å^2; Figure 2(f) and (g)).

The β-sheet in subunit A (residues 82–89), situated in the central part of the motif (Figure 1(c)), interacts with the first α-helix in subunit B; E85 and E87 interact with R107, K103 and Q99. The first α-helix in subunit A interacts with the last α-helix on subunit B; E96 interact with K173. All these residues play a significant role in subunit/subunit stabilization (Table 2), as suggested by computational alanine scanning mutagenesis calculations.

IN2/IN2 interactions

The contact surface is $3.16 \times 10^4 \text{ Å}^2$, of which 60% pertains to hydrophilic residues and ∼140 water molecules at the surface (corresponding to ∼0.005 water molecules/Å^2; Figure 2(h) and (i)). The interactions in the IN2/IN2 interface can be split into two regions as follows. *Region 1* (Figure 2(d)) encompasses the interactions between two α-helices from the CD of subunit B and two loops from the NTD of subunit D; D167, Q164, K186, R187 and I151 from subunit B interact with K42, H16, E11, C43 and Q146 from subunit D. K186 and R187 contribute to the IN2/IN2 interactions significantly, as suggested by computational mutagenesis (Table 2). This result is consistent with experimental evidence that shows that residues from 186 to 188 are important for IN oligomerization ([3]; Table 1). Residues I151 and Q146 establish an interaction that allows loops (139–152) in chains B and D get fixed and get exposed to possible DNA contacts. So, even if experiments point out these two residues as possible DNA contact residues, in our model Ile151 and Q146 are crucial because we allow 139–152 region to interact with

Table 1. Comparison with the experimental data.

Residue	Suggested role	Contact	References
1–12	DNA binding	vDNA	[30]
49–69	DNA binding	vDNA	[30]
54–57	3′-processing	vDNA	[8]
124–125	3′-processing	vDNA–hDNA	[8]
128–130	3′-processing	vDNA–hDNA	[8]
139–152	DNA binding	HDNA	[30]
148–143	vDNA binding	vDNA–hDNA	[24]
153	LTR specificity	vDNA	[8]
156–157	3′-processing	vDNA	[8]
159	vDNA binding	vDNA	[32]
160–161	LTR specificity	vDNA	[8]
163–165	LTR specificity	vDNA	[8]
186–188	Oligomerization	IN2/IN2	[3]
213–247	DNA binding	vDNA	[30]
228–231	DNA binding	vDNA	[8,38,53]
236	DNA binding	vDNA	[53,60]
242, 244, 246	DNA binding, oligomerization	IN2/IN2	[22,37,40]
247–275	DNA binding	vDNA	[30]
258	DNA binding	vDNA	[39]
263	DNA binding	vDNA, IN2/IN2	[37,38]
264	DNA binding	vDNA	[38]
269	Oligomerization	IN2/IN2	[40]

Several sequence regions and single amino-acids have been shown, by means of mutagenesis experiments and cross-linking analysis, to be crucial in one or several molecular events leading to the formation of the synaptic complex. For each of them (listed in column 1), the putative role (column 2) and the contacts established in our structural model (column 3) are reported. References in column 4 point to the papers in which corresponding experimental results were presented.

DNA, which accords with experimental data (Table 1). In addition, also E11 appears to play a significant role in subunit association.

Region 2 Figure 2(e) consists mainly of two segments from subunits A and C interacting with the NTD and CD from the other subunits. The NTD and CD of chain A interacts with the CD and NTD in chain C, respectively. K136, E138, D139, S141, K215, E246 from subunit A (or C) interact with D256, K264, S230, R231, E214 and K113 from subunit C (or A), respectively. D139, R231 and S230 play an important role in IN2/IN2 stabilization (Table 2). Furthermore,

Table 2. Changes in free energy of binding (in kcal/mol) upon mutations into alanine of the residues at the IN/IN and IN2/IN2 interfaces.

IN/IN		IN2/IN2 Region 1		IN2/IN2 Region 2	
Residue	$\Delta\Delta G$	Residue	$\Delta\Delta G$	Residue	$\Delta\Delta G$
E85	3.3	E11	1.8	D139	2.1
E87	3.1	K186	2.2	S230	1.0
E96	1.5	R187	3.1	D231	1.2
Y99	3.4				
K103	2.8				
R107	1.9				
W108	–				
K173	1.9				

Notice that $\Delta\Delta G$ greater than 0.2 kcal/mol was deemed as significant [49].

the results obtained for E246 are in accord with experiments [37,38,53]. Indeed, mutagenesis assays indicate that mutation of K264 decreases the IN activity of 50%, whereas mutation of E246 inhibits the oligomerization of IN [36,53].

Overall, our IN4 model identifies several key residues that stabilize IN4. Their key role have been investigated by computational alanine scanning mutagenesis. Free energy changes associated with IN2 and/or IN2/IN2 interface mutations[7] are consistent with available muta-genesis experiments (Table 1). Specially in the case of K186, R187 and E246, where the mutation of any of these residues to alanine decreases the activity of the enzyme by destabilizing IN oligomers. The main interactions found in our model are also present in Chen *et al.* [8] IN4 model while this is not the case for the model of Wielens *et al.* [59] as residue E246 is not involved in subunit–subunit interactions and thus, its position does not agree with experimental data (Table 1).

IN4–DNA

The synaptic complex models that have been built so far are not consistent with a variety of structural features and other experimental constraints. Our model shows that the presence of hDNA does not alter the IN4 architecture as proposed by previous Type II models. Therefore, we compare our synaptic complex model with that of Chen *et al.*, except for hDNA interactions, which are absent in the latter model.

IN4–vDNA moiety

Overall, the structure is similar to that presented by Chen *et al.* Each vDNA double strand binds one IN2 moiety and is located between CD of one chain and the CTD of the second one. Both vDNA substrates fit symmetrically inside the grooves formed by the NTD, CD and CTD (Figure 2(a) and (b)). In addition, subunits B and D (or A and C) feature an α-helix/loop structural motif that interacts with vDNA. This motif has also been found in the X-ray structure of the complex between T4 endonuclease V and a DNA strand ([54]; Figure 3(a) and (b)). This structural motif was also detected in Chen's model [8] and it could be common to other DNA binding proteins.

The location of the two vDNA is consistent with the observation that IN2 performs the $3'$-processing as it binds to one of the two LTR termini (similar to Chen *et al.*). The residues interacting with vDNA (defined here as those located at $7\,\text{Å}$ cut-off of vDNA) are: 49–69, 139–152 and 247–288. This result is consistent with the available cross-linking and
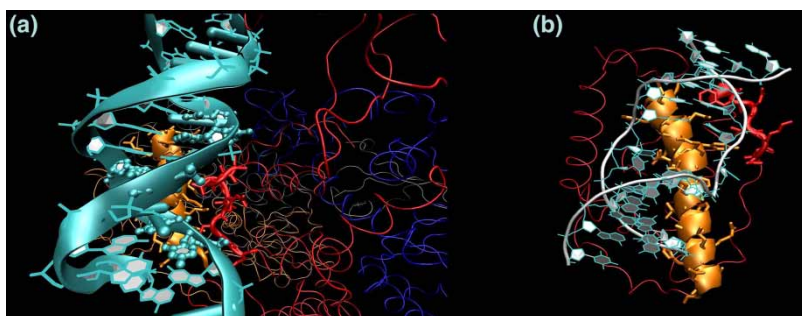

Figure 3. Structural similarity between (a) the IN4–vDNA complex (as obtained here by molecular modelling) and (b) the T4-endonuclease-V–DNA complex, as obtained by X-ray crystallography [54]. The α-helices, which contact the DNA major groove, are highlighted as orange cartoons. Available in colour online.

mutagenesis studies ([19,24,30,31,37,38,48]; as that of Chen *et al.*), which can be translated into structural features (Table 1, Figure 5(a) and (b)). Indeed, the residues experimentally shown as involved in or determining the specificity towards vDNA (Table 1) turn out to interact directly with vDNA or, at least, they stabilize residues that interact with vDNA by an additional network of hydrogen bonds (Figure 5).

IN4 interactions with hDNA

hDNA's minor and major grooves are accommodated by a 'saw tooth' like region, made up by solvent-exposed residues of B and C subunits (Figure 4).

Several features of the model are consistent with experiment: (i) the residues interacting with the DNA strands (as above defined as those within 7 Å cut-off) are: 64, 113–119, 122–125, 127, 128, 135–143, 151 and 152 (subunits B and D). All of these, except in the case of 113–119 and 127, have been shown to be involved in IN–DNA interactions (Table 1; Figure 5); (ii) the 5 bp between both catalytic sites are well accommodated [11]; (iii) the phosphate groups within the hDNA cleavage site are coordinated to the Mg^{2+} ion (distance 5.5 Å). This coordination is required for the subsequent reaction step (integration) where $3'$-ends have to be integrated into the hDNA; (iv) residues within the IN4 regions that are near both vDNA and h-DNA in our model have been found to be relevant for DNA binding (residues 54–57, 124–125 and 128–130) and (v) in our model these residues are close to the $3'$-OH of the vDNA and of the hDNA, consistently with the experimental studies in which these residues have been found to interact with DNA (Table 1). Our model places the hDNA binding region within the region predicted by Chen *et al.*, however, hDNA was not present in their model.

As for the active site, we notice that the processed $3'$-OH end of each DNA strand is located too far (11 Å) from the Mg^{2+} ion. A similar feature was also presented in Chen's model. Obviously, one could speculate that the protein may undergo a conformational transition before integration occurs so $3'$-OH ends would move closer to the Mg^{2+} ion. However, it is likely that this feature results from modelling limitations, including (but not necessarily being limited to):
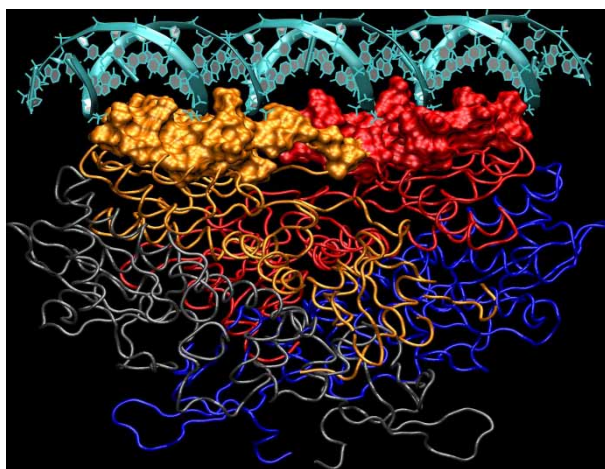

Figure 4.   Close up of the IN4–hDNA interface, as obtained here by molecular modelling. The four IN4 subunits are depicted by different colours: red, blue, gray and orange are used for subunits A, B, C and D, respectively. For the purpose of disclosing the shape complementarity between IN4 and the minor groove of the hDNA the molecular surface of the IN4 interacting region has been portrayed. Available in colour online.
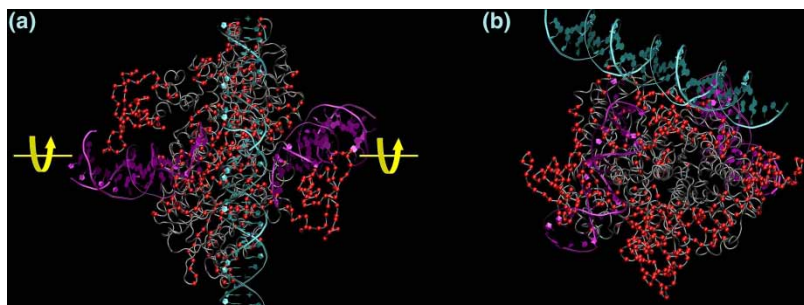
Figure 5. Functionally relevant residues for the synaptic complex formation. Front (a) and side (b) views of the whole IN4–vDNA–hDNA complex, the C_α atoms of the residues crucial for the synaptic complex formation, listed in Table 1, are represented as red spheres. Available in colour online.

(i) the limitation of the protein/DNA docking process performed here; and (ii) the fact that only one metal ion was placed in the active site. This might be suggested by the fact that the second Mg^{2+} ion opens the active site cavity in the structurally related Tn5 transposase/vDNA complex [14]; there could be a similar feature in our complex.

Final remarks

A structural model of HIV-1 integrase synaptic complex (IN4–DNA) which is consistent with all of the experimental data is still lacking. We have presented a model of IN4 in the free state and in complex with both vDNA and hDNA, which is based on several methodological improvements with respect to the previously reported models: (i) the structure is relaxed by longer MD time simulations; (ii) MD simulations are based on accurate parametrization of the Zn finger moieties in the structure (see Methods) and (iii) many protein residues absent from previous models are included. In addition, computational alanine scanning mutagenesis has been performed to identify key residues stabilizing IN/IN and IN2/IN2 interfaces. Comparison with the models of IN4 and Type I model of Chen *et al.*, revealed consistency of key structural features and better agreement with available experimental results. Previously, proposed models that include hDNA have not been considered in this comparison since the disposition of IN4 oligomer in these models does not agree with experimental studies [3,11,28,29].

As in the case of Chen *et al.* [8], our model turns out to be consistent with all mutagenesis and cross-linking data available in literature that can be translated into structural features (Table 1). In particular, it is consistent with the fact that the 3′-processing needs just one IN2 and evidences a complementarity between the CTD and CD domains. Furthermore, it features four K–E salt bridges at the IN2/IN2 interface involving residues, which were shown to be crucial for the *in vivo* IN tetramer formation [3]. The 3′-OH termini of vDNA are pointing towards the DDE motif and residues K159, K186, K187, K188 and K211 interact with vDNA forming a continuous positive surface for vDNA binding. Alanine scanning mutagenesis calculations identified key residues in stabilization of both the IN/IN and IN2/IN2 interfaces and some of these residues reduce the enzymatic activity to 50% or destabilize IN oligomerization when mutated in the protein [37]

In spite of these encouraging results, we do also find that the 3′-OH end of vDNA and the catalytic centre are far apart. The available experimental information does not allow us to conclude whether this is an inconsistency of the model or it is due to conformational changes associated with the hydrolysis of hDNA, which would bring the vDNA 3′-OH ends closer to the catalytic centre. A comparison with a model based on Tn5 transposase X-ray structure (in progress) will provide further insights into this crucial issue.

Acknowledgement

We are indebted to Vladimir Sulimov for scientific discussions.

Notes

1. Equally contributing.
2. Intriguingly, the oligomeric state is affected by the interactions with host–cell proteins. Indeed, LEDGF/p75 has been shown to shift the equilibrium towards the IN tetrameric species in *in vivo* [45].
3. Although in the X-ray structure the ion could be either Mg^{2+} or Mn^{2+}, and both are functional *in vivo*, only Mg^{2+} is believed to be present *in vivo* [4].
4. D64, D116 and E152, DDE motif.
5. The most representative conformation is here defined as the central member of the most populated cluster following the protocol of Ref. [13]. For that, structures are defined as belonging to the same cluster if their RMSD is less than 0.5 Å.
6. The detail of the simulations are as follows: the complex was solvated by 54,285 water molecules and periodic boundary conditions were applied to a rectangular unit cell. Constant temperature [41], constant pressure [42] MD simulations were performed with the GROMACS package [35] using the AMBER parm99 force field [7] for all atoms except those in the four HHCC motifs. For these, an apt parametrization protocol was adopted for atomic charges, described in Ref. [12], so as to effectively take into account the polarization of the zinc ion ligands.
7. We recall that a change of 0.2 kcal/mol is generally considered significant [49].

References

[1] M. Anker and R.B. Corales, *Raltegravir (mk-0518): A novel integrase inhibitor for the treatment of HIV infection*, Expert Opin. Investig. Drugs 17 (2008), pp. 97–103.

[2] F. Bernardi, A. Bottoni, M. De Vivo, M. Garavelli, G. Keseru, and N.-S. Gábor, *A hypothetical mechanism for HIV-1 integrase catalytic action: DFT modelling of a bio-mimetic environment*, Chem. Phys. Lett. 362 (2002), pp. 1–7.

[3] L. Berthoux, S. Sebastian, M.A. Muesing, and J. Luban, *The role of lysine 186 in HIV-1 integrase multimerization*, Virology 364 (2007), pp. 227–236.

[4] G. Bujacz, J. Alexandratos, A. Wlodawer, G. Merkel, M. Andrake, R.A. Katz, and M. Skalka, *Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity*, J. Biol. Chem. 272 (1997), pp. 18161–18168.

[5] M. Cai, R. Zheng, M. Caffrey, R. Craigie, G.M. Clore, and M. Gronenborn, *Solution structure of the N-terminal zinc binding domain of HIV-1 integrase*, Nat. Struct. Biol. 4 (1997), pp. 567–577.

[6] A. Cereseto, L. Manganaro, M.I. Gutierrez, M. Terreni, A. Fittipaldi, M. Lusic, A. Marcello, and M. Giacca, *Acetylation of HIV-1 integrase by p300 regulates viral integration*, EMBO J. 24 (2005), pp. 3070–3081.

[7] T.E. Cheatham, III, P. Cieplak, and P.A. Kollman, *A modified version of the Cornell et al., force field with improved sugar pucker phases and helical repeat*, J. Biomol. Struct. Dyn. 16 (1999), pp. 845–862.

[8] A. Chen, I.T. Weber, R.W. Harrison, and J. Leis, *Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat ends*, J. Biol. Chem. 281 (2006), pp. 4173–4182.

[9] J.C. Chen, J. Krucinski, L.J. Miercke, J.S. Finer-Moore, A.H. Tang, A.D. Leavitt, and R.M. Stroud, *Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding*, Proc. Natl. Acad. Sci. USA 97 (2000), pp. 8233–8238.

[10] T.K. Chiu and D.R. Davies, *Structure and function of HIV-1 integrase*, Curr. Top. Med. Chem. 4 (2004), pp. 965–977.

[11] R. Craigie, *HIV integrase, a brief overview from chemistry to therapeutics*, J. Biol. Chem. 276 (2001), pp. 23213–23216.

[12] M. Dal Peraro, K. Spiegel, G. Lamoureux, M.D. Vivo, W.F. DeGrado, and M.L. Klein, *Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations*, J. Struct. Biol. 157 (2007), pp. 444–453.

[13] X. Daura, K. Gademann, B. Jaun, B. Seebach, W.F. van Gunsteren, and A.E. Mark, *Peptide folding: When simulation meets experiment*, Angew. Chem. Int. Ed. 38 (1999), pp. 236–240.

[14] D.R. Davies, I.Y. Goryshin, W.S. Reznikoff, and I. Rayment, *Three-dimensional structure of the Tn5 synaptic complex transposition intermediate*, Science 289 (2000), pp. 77–85.

[15] L. De Luca, G. Vistoli, A. Pedretti, M.L. Barreca, and A. Chimirri, *Molecular dynamics studies of the full-length integrase–DNA complex*, Biochem. Biophys. Res. Commun. 336 (2005), pp. 1010–1016.

[16] S.J. de Vries, A.D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, and A.M. Bonvin, *HADDOCK vs. HADDOCK: New features and performance of HADDOCK2.0 on the capri targets*, Proteins 69 (2007), pp. 726–733.

[17] A.M. Dirac and J. Kjems, *Mapping DNA-binding sites of HIV-1 integrase by protein footprinting*, Eur. J. Biochem. 268 (2001), pp. 743–751.

[18] C. Dominguez, R. Boelens, and A.M. Bonvin, *HADDOCK: A protein–protein docking approach based on biochemical or biophysical information*, J. Am. Chem. Soc. 125 (2003), pp. 1731–1737.

[19] R.R. Drake, N. Neamati, H. Hong, A.A. Pilon, P. Sunthankar, S.D. Hume, G.W. Milne, and Y. Pommier, *Identification of a nucleotide binding site in HIV-1 integrase*, Proc. Natl. Acad. Sci. USA 95 (1998), pp. 4170–4175.

[20] A.P. Eijkelenboom, R. Sprangers, K. Hård, R.A. Puras Lutzke, R.H. Plasterk, R. Boelens, and R. Kaptein, *Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase*, Proteins 36 (1999), pp. 556–564.

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

[22] A. Engelman, G. Englund, J.M. Orenstein, M.A. Martin, and R. Craigie, *Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication*, J. Virol. 69 (1995), pp. 2729–2736.

[23] ———, *Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase*, J. Virol. 71 (1997), pp. 3507–3514.

[24] D. Esposito and R. Craigie, *Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein–DNA interaction*, EMBO J. 17 (1998), pp. 5832–5843.

[25] A. Fiser and A. Sali, *Modeller: Generation and refinement of homology-based protein structure models*, Methods Enzymol. 374 (2003), pp. 461–491.

[26] N. Foloppe and A.D. MacKerell, *All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data*, J. Comp. Chem. 21 (2000), pp. 86–104.

[27] K. Gao, S.L. Butler, and F. Bushman, *Human immunodeficiency virus type 1 integrase: Arrangement of protein domains in active cDNA complexes*, EMBO J. 20 (2001), pp. 3565–3576.

[28] E. Guiot, K. Carayon, O. Delelis, F. Simon, P. Tauc, E. Zubin, M. Gottikh, J.F. Mouscadet, J.C. Brochon, and E. Deprez, *Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity*, J. Biol. Chem. 281 (2006), pp. 22707–22719.

[29] Z. Hayouka, J. Rosenbluh, A. Levin, S. Loya, M. Lebendiker, D. Veprintsev, M. Kotler, A. Hizi, A. Loyter, and A. Friedler, *Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium*, Proc. Natl. Acad. Sci. USA 104 (2007), pp. 8316–8321.

[30] T.S. Heuer and P.O. Brown, *Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase–DNA complex*, Biochemistry 37 (1998), pp. 6667–6678.

[31] ———, *Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase–DNA complex*, Biochemistry 37 (1998), pp. 6667–6678.

[32] T.M. Jenkins, D. Esposito, A. Engelman, and R. Craigie, *Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking*, EMBO J. 16 (1997), pp. 6849–6859.

[33] R.G. Karki, Y. Tang, T.R. Burke, and M.C. Nicklaus, *Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design*, J. Comput. Aided Mol. Des. 18 (2004), pp. 739–760.

[34] J. Kulkosky, K.S. Jones, R.A. Katz, J.P. Mack, and M. Skalka, *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 (1992), pp. 2331–2338.

[35] E. Lindahl, B. Hess, and D. van der Spoel, *Gromacs 3.0: A package for molecular simulation and trajectory analysis*, J. Mol. Mod. 7 (2001), pp. 306–317.

[36] R. Lu, H.Z. Ghory, and A. Engelman, *Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase*, J. Virol. 79 (2005), pp. 10356–10368.

[37] R. Lu, H.Z. Ghory, and A. Engelman, *Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase*, J. Virol. 79 (2005), pp. 2493–2505.

[38] R.A. Lutzke, C. Vink, and R.H. Plasterk, *Characterization of the minimal DNA-binding domain of the HIV integrase protein*, Nucleic Acids Res. 22 (1994), pp. 4125–4131.

[39] T. Masuda, M.J. Kuroda, and S. Harada, *Specific and independent recognition of U3 and U5 att sites by human immunodeficiency virus type 1 integrase* in vivo, J. Virol. 72 (1998), pp. 8396–8402.

[40] J.E. Murphy and S.P. Goff, *A mutation at one end of moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase* in vivo, J. Virol. 66 (1992), pp. 5092–5095.

[41] S. Nosé, *A unified formulation of the constant temperature molecular dynamics methods*, J. Chem. Phys. 81 (1984), pp. 511–519.

[42] M. Parrinello and A.J. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*, J. Appl. Phys. 52 (1981), pp. 7182–7190.

[43] A.A. Podtelezhnikov, K. Gao, F.D. Bushman, and J.A. McCammon, *Modeling HIV-1 integrase complexes based on their hydrodynamic properties*, Biopolymers 68 (2003), pp. 110–120.

[44] P. Polard and M. Chandler, *Bacterial transposases and retroviral integrases*, Mol. Microbiol. 15 (1995), pp. 13–23.

[45] N.K. Raghavendra and A. Engelman, *Ledgf/p75 interferes with the formation of synaptic nucleoprotein complexes that catalyze full-site HIV-1 DNA integration* in vitro: *Implications for the mechanism of viral cDNA integration*, Virology 360 (2007), pp. 1–5.

[46] P.A. Rice and T.A. Baker, *Comparative architecture of transposase and integrase complexes*, Nat. Struct. Biol. 8 (2001), pp. 302–307.

[47] S.J. Rowland and K.G. Dyke, *Tn552, a novel transposable element from staphylococcus aureus*, Mol. Microbiol. 4 (1990), pp. 961–975.

[48] K. Sayasith, G. Sauvé, and J. Yelle, *Characterization of mutant HIV-1 integrase carrying amino acid changes in the catalytic domain*, Mol. Cells 10 (2000), pp. 525–532.

[49] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, *The foldx web server: An online force field*, Nucleic Acids Res. 33 (2005), pp. 382–388.

[50] E.A. Semenova, C. Marchand, and Y. Pommier, *HIV-1 integrase inhibitors: Update and perspectives*, Adv. Pharmacol. 56 (2008), pp. 199–228.

[51] S. Sierra, B. Kupfer, and R. Kaiser, *Basics of the virology of HIV-1 and its replication*, J. Clin. Virol. 34 (2005), pp. 233–244.

[52] M. Steiniger-White, I. Rayment, and W.S. Reznikoff, *Structure/function insights into Tn5 transposition*, Curr. Opin. Struct. Biol. 14 (2004), pp. 50–57.

[53] D.C. van Gent, A.A. Groeneger, and R.H. Plasterk, *Mutational analysis of the integrase protein of human immunodeficiency virus type 2*, Proc. Natl. Acad. Sci. USA 89 (1992), pp. 9598–9602.

[54] D.G. Vassylyev, T. Kashiwagi, Y. Mikami, M. Ariyoshi, S. Iwai, E. Ohtsuka, and K. Morikawa, *Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: Structural basis for damaged DNA recognition*, Cell 83 (1995), pp. 773–782.

[55] C. Vink, M. Groenink, Y. Elgersma, R.A. Fouchier, M. Tersmette, and R.H. Plasterk, *Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA*, J. Virol. 64 (1990), pp. 5626–5627.

[56] ———, *Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage*, J. Virol. 65 (1991), pp. 4636–4644.

[57] J.Y. Wang, H. Ling, W. Yang, and R. Craigie, *Structure of a two-domain fragment of HIV-1 integrase: Implications for domain organization in the intact protein*, EMBO J. 20 (2001), pp. 7333–7343.

[58] L.D. Wang, C.L. Liu, W.Z. Chen, and C.X. Wang, *Constructing HIV-1 integrase tetramer and exploring influences of metal ions on forming integrase-DNA complex*, Biochem. Biophys. Res. Commun. 337 (2005), pp. 313–319.

[59] J. Wielens, I.T. Crosby, and D.K. Chalmers, *A three-dimensional model of the human immunodeficiency virus type 1 integration complex*, J. Comput. Aided Mol. Des. 19 (2005), pp. 301–317.

[60] M. Wiskerchen and M.A. Muesing, *Human immunodeficiency virus type 1 integrase: Effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates and sustain viral propagation in primary cells*, J. Virol. 69 (1995), pp. 376–386.