

# Computational and synthetic approaches for the discovery of HIV-1 integrase inhibitors

Maria Letizia Barreca, Laura De Luca, Stefania Ferro, Angela Rao,  
Anna-Maria Monforte, and Alba Chimirri\*

*Dipartimento Farmaco-Chimico University of Messina, Viale Annunziata 98168, Messina, Italy*  
*E-mail: [chimirri@pharma.unime.it](mailto:chimirri@pharma.unime.it)*

---

## Abstract

In recent years our research group has been engaged in the structure-function study of IN enzyme and in the development of new HIV-1 IN-inhibitors. We first developed three-dimensional hypothetical pharmacophore models for the binding of IN inhibitors to the enzyme. In particular we focused our attention on  $\beta$ -diketo acid (DKA) derivatives which represent the major leads in the development of anti-HIV-1 IN drugs, considering that the only two IN inhibitors undergoing clinical trials belong to this family.

The resulting pharmacophore models allowed the discovery of new potential IN inhibitors, both through rational design and virtual screening. Biological testing showed that our strategy was successful in searching for new structural leads as HIV-1 IN inhibitors. In addition we built a plausible model of the full-length HIV-1 integrase dimer complexed with viral DNA on which molecular dynamics simulation studies were carried out.

**Keywords:** HIV-1 integrase, full-length IN–DNA complex, dynamics simulation; 3D-pharmacophore model, IN-inhibitors, DKA, microwave-assisted synthesis

---

## Contents

1. Pharmacophore modeling and discovery of HIV-1 Integrase inhibitors
2. Analysis and molecular dynamic studies of the full-length integrase-DNA complex

## Introduction

The fight against HIV-1 infection has been considerably boosted by the discovery of two important class of antiviral drugs, i.e. the specific inhibitors of the viral enzymes reverse transcriptase (RT) and protease (PR).<sup>1,2</sup> However, the current standard antiretroviral therapy has

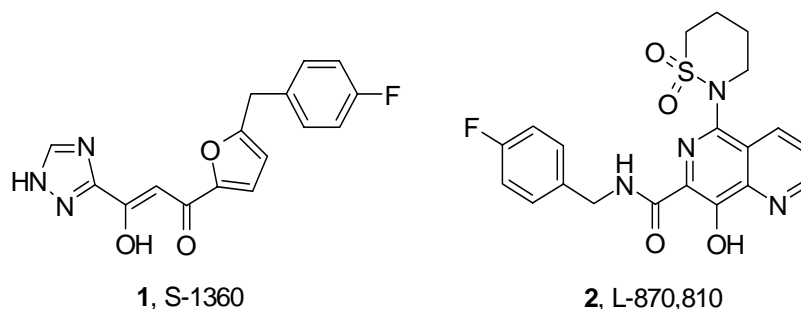
raised several problems in terms of poor tolerability and development of multidrug resistance, thus emphasizing the demand of new agents directed against alternative sites in the viral life cycle.

Integrase (IN) has thus emerged as a promising target for the development of new generation of inhibitors to be used in the anti-HIV therapy, because it plays a key role in stable infection and a known functional analogue is lacking in the human host cells. Moreover, although a wide variety of compounds have been reported as IN inhibitors, drugs active against this enzyme have not as yet been approved by the FDA.

Integrase inserts a double stranded DNA copy of the viral RNA genome into the chromosomes of an infected cell through two separate reactions; in the “3’-processing” step, IN removes two nucleotides from each 3’- end of viral cDNA, while in the “strand transfer” reaction, the two newly processed 3’-viral DNA ends are inserted into the host cell DNA. For the integration reaction, no source of energy (e.g. no ATP) is needed and only divalent cations such as  $Mn^{2+}$  or  $Mg^{2+}$  are required for the catalytic activity.

In recent years our research was addressed to the development of new HIV-1 IN-inhibitors and the structure-function study of IN enzyme.

To date,  $\beta$ -diketo acid (DKAs) and their derivatives represent the major leads in the development of anti-HIV-1 IN drugs, considering that the only two IN inhibitors undergoing clinical trials, Shionogi/Glaxo-SmithKline’s S-1360<sup>3</sup> and Merck’s L-870-810<sup>4</sup> (Figure 1, **1** and **2**), belong to this family.



**Figure 1.** Two integrase inhibitors currently in clinical trials.

These compounds have been shown to selectively inhibit the strand transfer step by sequestering the divalent cations bound in the active site of the enzyme,<sup>5</sup> and to block HIV-1 replication in infected cells.

For the above reasons, our idea was to generate 3D pharmacophore models for the binding of DKA class (Integrase Strand Transfer Inhibitors, INSTI) to the enzyme.<sup>6,7</sup>

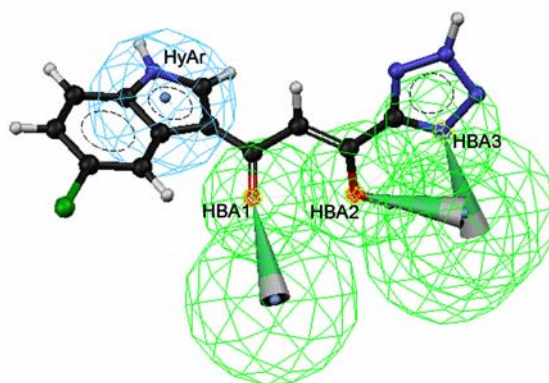
In addition, in order to study the action mechanism of IN inhibitors and considering the absence of information about the complete three-dimensional structure of HIV-1 integrase, we used a modified approach for DNA docking to obtain a model of the full-length integrase-DNA complex.<sup>8</sup> We also carried out a molecular dynamic (MD) simulation of the IN-DNA complex in

order to gain information about the enzyme motion and to explore the dynamic behaviour of a surface loop thought essential in the catalytic mechanism of IN.<sup>9</sup>

## 1. Pharmacophore modeling and discovery of HIV-1 Integrase inhibitors

Our first aim was to build a simple 3D pharmacophore model representing the distinguishing chemical features of Integrase Strand Transfer Inhibitors (INSTI) and to obtain a useful tool for further discovery of novel IN inhibitors.

Using the 5CITEP, 3 atomic coordinates available from X-ray crystallography data (PDB code 1QS4), we developed a 3D model (Figure 2) which was consistent with the proposed mechanism of action for this family of IN inhibitors.



**Figure 2.** The four-point hypothetical model for DKA analog IN inhibitors built using the Catalyst program (HBA, green: hydrogen-bond acceptor features; HyAr, cyan: hydrophobic aromatic function). The vectors represent the putative interactions between the DKA analogs and the metal ions. The structure of 5CITEP is superimposed on the model.

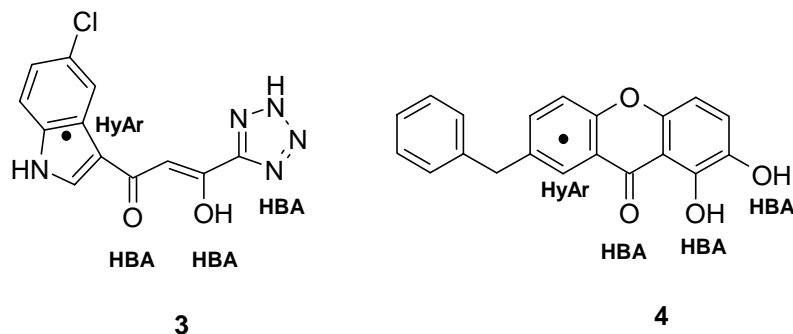
This four-point 3D model was created by assembling three hydrogen-bond acceptors, mapped over the ketoenol moiety (HBA1 and HBA2) and the N1 atom of the tetrazole ring (HBA3), and one hydrophobic-aromatic group (HyAr), mapped over the indole centroid.

To restrict the 3D spatial arrangement for the next database search, we added several distance constraints among HBAs-HBAs and HBAs-HyAr functions, according to the corresponding distances calculated in 5CITEP and in another DKA analogue, 4 (Figure 3 and Table 1).

In fact, the interaction between the ligand 5CITEP and the metal ions in the IN active site is accounted for by our hypothesis of the mapping of HBA sites described by the ketoenol functionality and the N-1 atom of the tetrazole moiety. Moreover, a hydrophobic aromatic region

(HyAr) was positioned over the indole ring since it was repeatedly reported that an aromatic moiety seems to be an essential structural element for anti-IN activity.<sup>10,11</sup>

While 5CITEP is a perfect point of reference for distances since we obtained its structure from crystallographic data, molecule **4** was chosen among the most potent DKA analogue IN inhibitors because the 1,3-diketo acid moiety is incorporated in a highly rigid system,<sup>12</sup> thus providing useful and unambiguous geometric information. The latter compound was constructed using standard bond lengths and angles from Sybyl<sup>13</sup> fragment library and fully optimized by the semiempirical quantum mechanical method AM1.

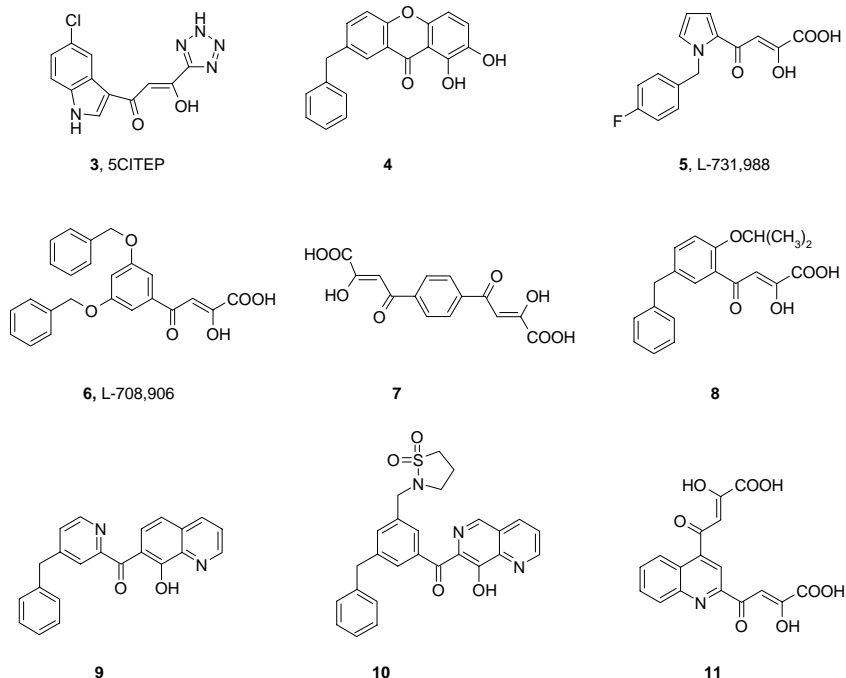


**Figure 3.** Mapping of chemical functions over 5CITEP (left) and compound **4** (right).

**Table 1.** Calculated and specified distance in 5CITEP, **4**, and our pharmacophore model

molecules	Distance (Å)			
	HBA1-HBA2	HBA2- HBA3	HBA1- HBA3	HBA1-HyAr
<b>3</b> , 5CITEP	2.70	2.99	5.38	3.71
<b>4</b>	2.79	2.70	5.25	3.69
3Dmodel	2.5-3.0	2.5-3.2	5.0-5.6	3.5-4.2

A prediction set was later used in order to validate the reliability of the postulated 3D pharmacophore hypothesis; this included the only two integrase inhibitors currently in human studies, S-1360<sup>3</sup> and L-870,810<sup>4</sup> (Figure 1), as well as eight molecules among the most representative DKA-based agents so far reported (Figure 4, **3-11**).<sup>12,14-16</sup>



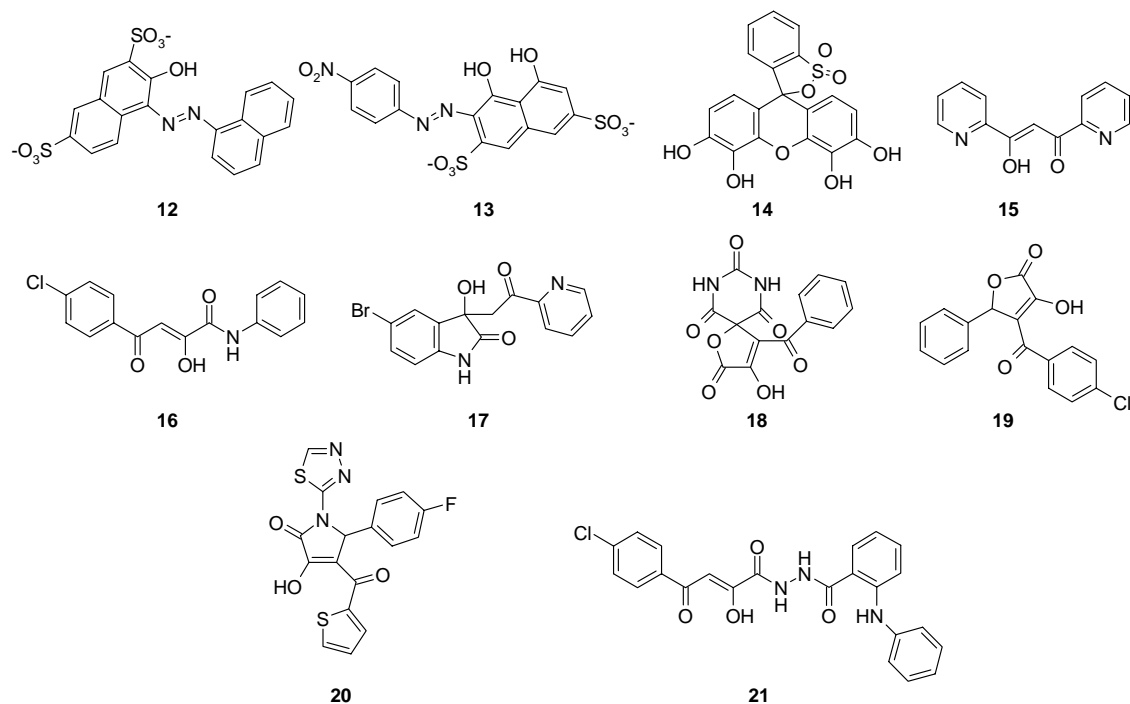
**Figure 4.** Chemical structures of the most representative DKA analogue integrase inhibitors.

All compounds of the prediction set were found to fit the pattern of the four structural features in at least one conformation thus suggesting that this hypothetical model might be a useful tool for the discovery of potential DKA-based lead compounds.

The next step was thus to use the putative pharmacophore model as a search query to identify structural templates from 3D small molecule databases. About 4000 compounds that contained, in some conformation, the specified 3D location of chemical functions were found.

A subset of these structures was then chosen by removing compounds that did not satisfy the well-known Lipinski rules describing properties of drug-like compounds.<sup>17</sup> The remaining molecules were overlaid with the pharmacophore by using the Best Fit option, and the top 100 hits were visually reviewed.

Finally, a total of 10 compounds (**12-21**, Figure 5) were selected for assaying in a stepwise fashion on the basis of (1) their fit value, (2) log *P*, (3) chemical diversity, and (4) availability and cost.



**Figure 5.** Ten compounds retrieved from the CAP2002 database and assayed for IN inhibition.

The biological results, namely, the inhibition of HIV-1 integrase activity in an enzymatic assay, suggest that our approach might be effective in identifying possible IN inhibitor lead candidates for further development (Table 2).

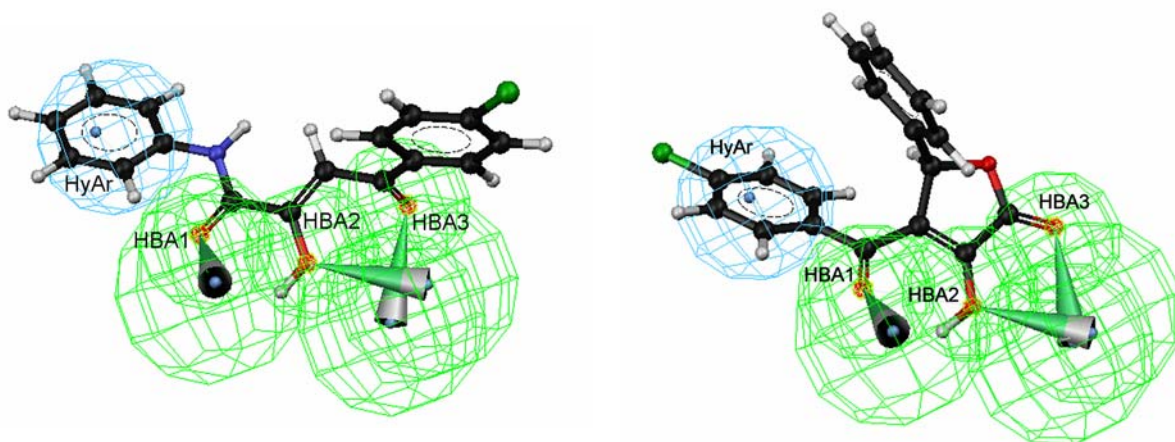
**Table 2.** Inhibition of HIV-1 integrase activity and Fit values

Compound	Overall integration IC <sub>50</sub> ( $\mu$ M) <sup>a</sup>	Fit value
<b>12</b>	83.6	3.05
<b>13</b>	97	3.12
<b>14</b>	2.8	3.68
<b>15</b>	61.4	2.34
<b>16</b>	1.9	3.85
<b>17</b>	> 100	2.08
<b>18</b>	> 100	3.14
<b>19</b>	0.9	3.80
<b>20</b>	>100	2.97
<b>21</b>	21.2	3.60

<sup>a</sup> Concentration of inhibitor that inhibits the overall integration in the oligonucleotide-based assay by 50%. Median values of three separate experiments are shown.

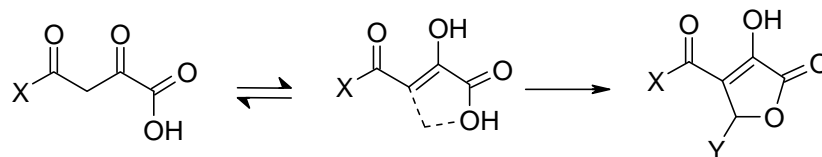
In fact, out of the 10 compounds tested, seven inhibited the HIV-1 integrase enzymatic activity. Of these, three (**14**, **16**, and **19**) showed potency below 10 $\mu$ M; in particular, compounds **16** and **19** were the most active IN inhibitors with a 50% inhibitory concentration (IC<sub>50</sub>) in the overall integration assay of 1.9 and 0.9 $\mu$ M, respectively. The superimposition of the 10 molecules against the hypothetical hypothesis revealed that the four features of the pharmacophore were well matched by the chemical groups of the molecules.

As an example, Figure 6 illustrates the alignment of compounds **16** and **19** onto the plausible 3D pharmacophore model for DKA analogues.



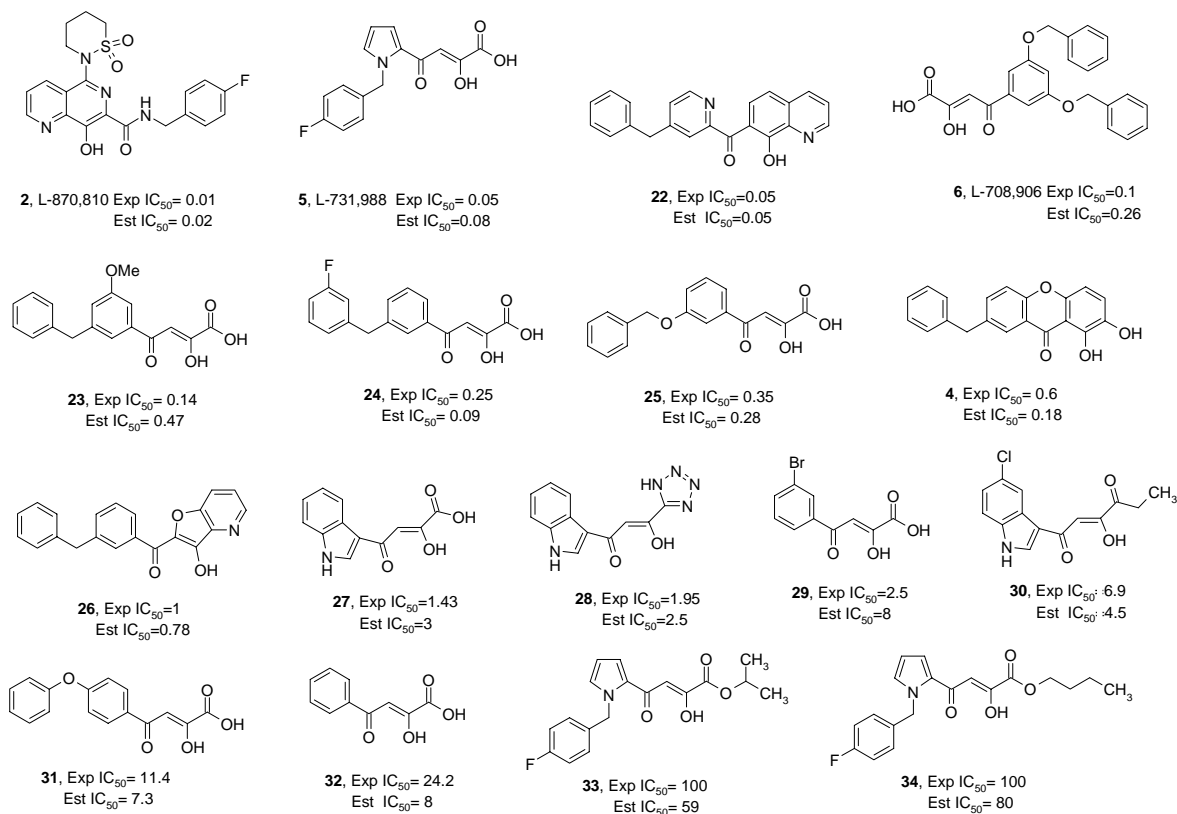
**Figure 6.** Compounds **16** (left) and **19** (right) mapped to the proposed pharmacophore model for HIV-1 IN inhibition (HBA, green, hydrogen-bond acceptor; HyAr, cyan, hydrophobic aromatic).

In particular, the most potent IN inhibitor **19**, 4-(4-chlorobenzoyl)-3-hydroxy-5-phenyl-5*H*-furan-2-one, belongs to a class of compounds which is different from previously described DKA analogue IN inhibitors. However, **19** interestingly incorporates the 1,3-diketo acid motif into a 4-keto-3-hydroxy-furan-2-one ring (Figure 7); in fact, considering that the 1,3-diketo acid moiety enolizes at the  $\alpha$ -position, this molecule can be considered a “closed-form” of the above-mentioned chemical functionality and a new bioactive scaffold for further optimization.



**Figure 7.** Structural analogy between the diketo acid motif and the 4-keto-3-hydroxy-furan-2-one system.

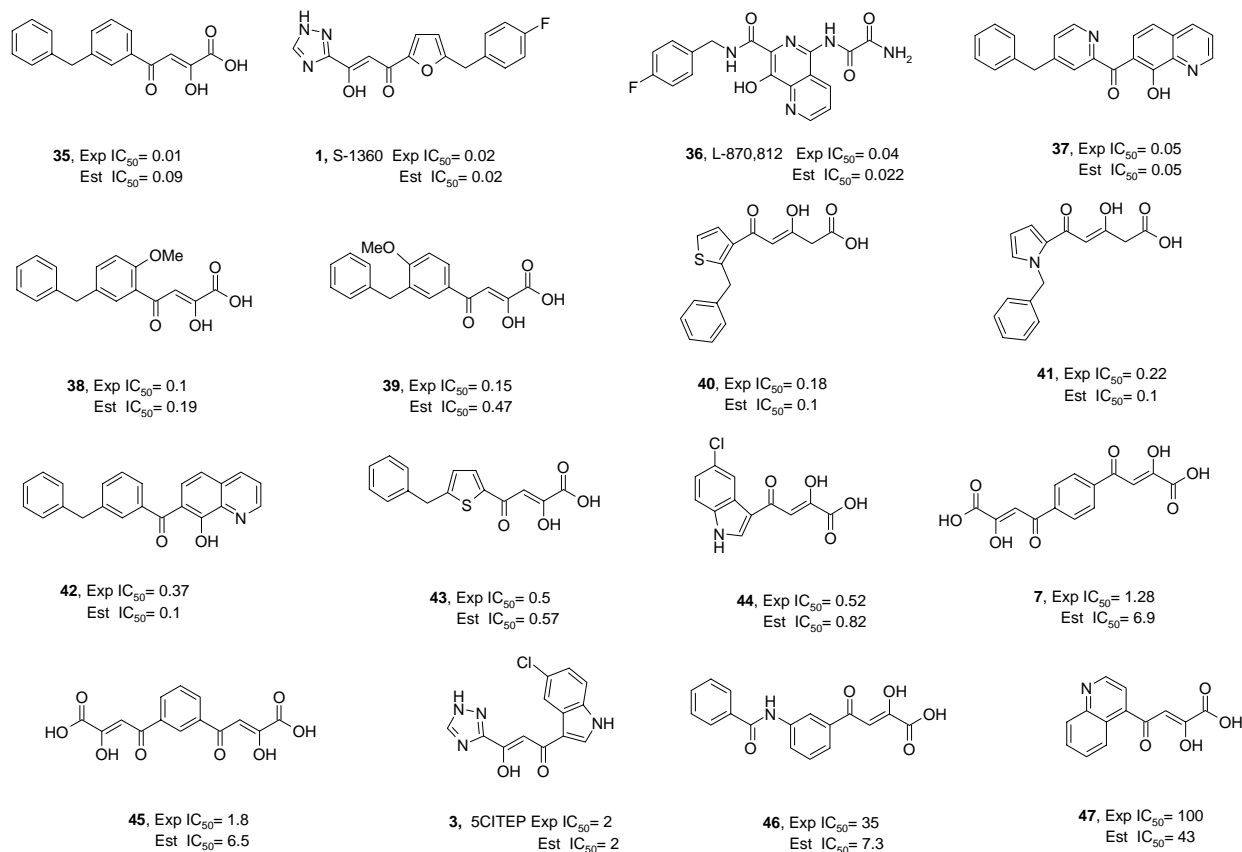
We successively extended our investigation to the development of 3D QSAR models for the same family of IN inhibitors, considering that the only two previously reported pharmacophores for DKA-like derivatives were “qualitative” hypotheses<sup>6,18</sup> (generated without the use of activity data), whereas for the generation of our “quantitative predictive” model, the biological activities were taken into account.<sup>7</sup>



**Figure 8.** Chemical structures of the 17 compounds of TS with their experimental (Exp IC<sub>50</sub>) and estimated (Est IC<sub>50</sub>) IN strand transfer inhibitory activities, both expressed in μM. The molecules are placed in order of decreasing activity.

We used a data set consisting of 33 molecules acting as IN strand-transfer-selective inhibitors and including the most representative DKAs and DKA-like derivatives reported in the literature.<sup>5,10,12,15,19,20</sup> Among them, 17 compounds were selected as the training set (TS), and the other 16 were used as the prediction set (PS). The structures and IN inhibitory activity values of both TS and PS molecules are reported in Figure 8 and Figure 9, respectively.





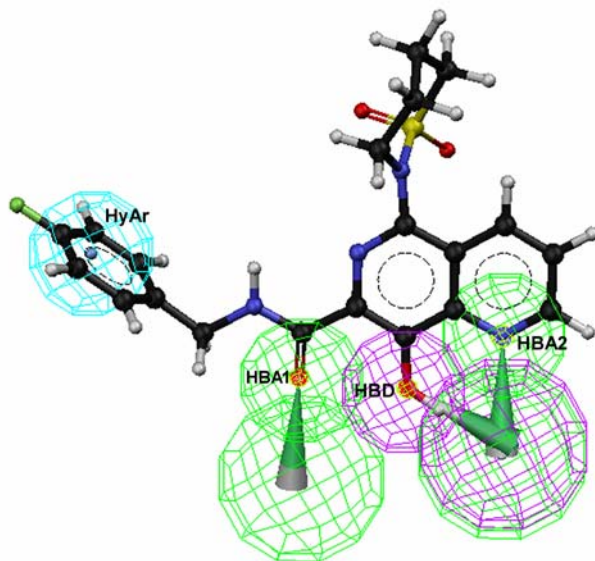
**Figure 9.** Chemical structures of the 16 compounds of PS with their experimental (Exp IC<sub>50</sub>) and estimated (Est IC<sub>50</sub>) IN strand transfer inhibitory activities, both expressed in μM. The molecules are placed in order of decreasing activity.

The range of in vitro IN inhibitory activity, expressed as IC<sub>50</sub> for strand transfer inhibition, spanned 5 orders of magnitude (0.01-100 μM), making this a good data set for HypoGen module.

Using the selected TS molecules, HypoGen algorithm constructed the 10 simplest hypotheses that showed the best correlation between estimated and measured activities. The top ranked pharmacophore model had the best predictive potentiality and statistical significance.

The 3D hypothesis consisted of one hydrophobic aromatic region (HYAr). Two hydrogen-bond acceptor (HBA1 and HBA2), and one hydrogen-bond donor (HBD) sites in a specific three-dimensional orientation;

Using the most active molecule (**2**) of the TS, which is also one of the two IN-DKA-like inhibitors in clinical development, a flexible fit of this molecule to Hypo1 is shown in Figure 10.

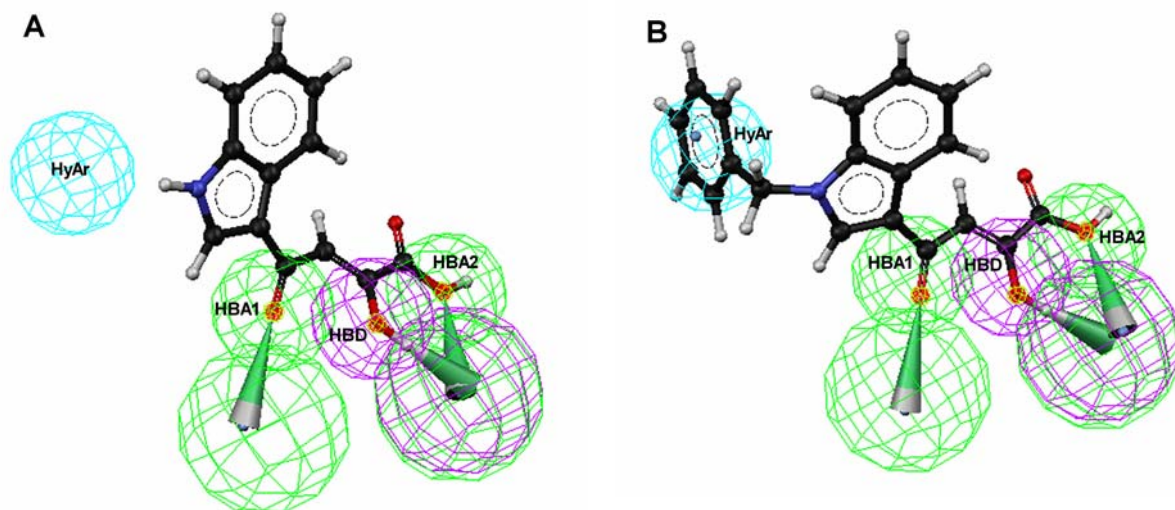


**Figure 10.** The top scoring HypoGen pharmacophore Hypo1 is mapped to the most active compound in the training set (**2**, L-870,810) (HBD, hydrogen bond donor, purple; HBA, hydrogen bond acceptor, green; HYAr, hydrophobic region, cyan).

We also used Hypo1 to perform a regression analysis with the PS of 16 compounds (Figure 9) in order to check the predictive power of this model. Linear regression of the predicted activities for PS IN inhibitors versus the experimental ones gave a fairly good correlation coefficient of 0.85, confirming the validity of the most statistically significant HypoGen hypothesis in predicting the IN inhibitory activity of DKAs and DKA like derivatives.

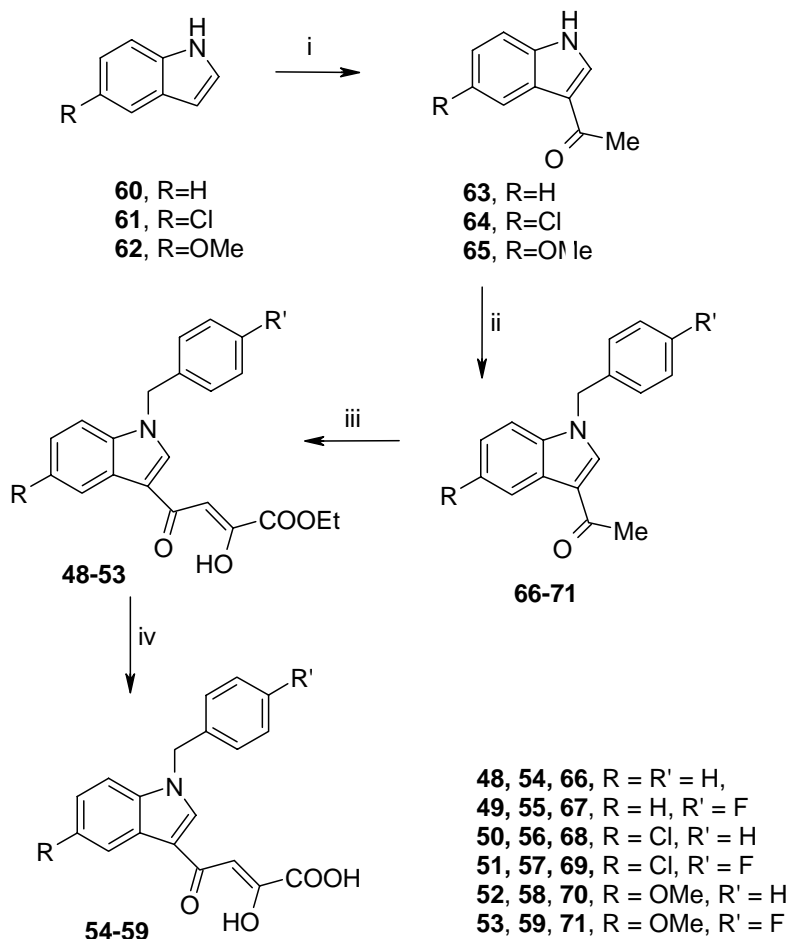
It is also worth noting that the type and number of features encoded in the automatically generated hypothesis were in full agreement with our previously manually developed ligand-based pharmacophore model.<sup>6,7</sup>

However, the spatial location of HYAr differed a little in the two hypotheses. In particular, the HypoGen runs pointed out that to have compounds with very high IN inhibitory activity, it is necessary to have one hydrophobic feature well separated by the DKA motif.



**Figure 11.** Compound **27** (A) and **54** (B) aligned on the lowest cost pharmacophore model generated for DKA inhibitors (HBD, hydrogen bond donor, purple; HBA, hydrogen bond acceptor, green; HYAr, hydrophobic region, cyan).

The most active compounds in the data set assumed a conformation that allowed proper mapping of the HYAr feature of the generated hypothesis, whereas most of the less active compounds (**27-34**, **3**, **7**, **45-47**) were unable to map this feature. Among the compounds that failed to map the HYAr feature, compound **27** drew our attention; the alignment of this IN inhibitor on Hypo1 (Figure 11A) clearly showed that its IN inhibitory potency ( $IC_{50}=1.4$ ) might be improved by further analogue synthesis.



**Scheme 1. Reagents and conditions:** i) AcCl, Et<sub>2</sub>AlCl, CH<sub>2</sub>Cl<sub>2</sub>, 0°C, 2h. ii) benzyl or 4-fluorobenzyl bromide, NaH, DMF, 0°C, 30 min; iii) diethyl oxalate, dry C<sub>2</sub>H<sub>5</sub>ONa, THF, two separated steps in the same conditions: 50°C, 2 min, 250 W, 300 psi; iv) 2N NaOH, MeOH, rt, 1.5 h.

In fact, since **27** mapped to only three of the four features of the lowest-cost Catalyst-generated DKA hypothesis, we thought that the introduction of functionality in this ligand, which might interact with the fourth feature of the hypothesis (i.e. HYAr), could provide enhanced IN inhibitory activity to the compound.

This idea was also supported by comparison of the activity data of compound **32** (IC<sub>50</sub> =24.2) and its benzyl derivative **35** (IC<sub>50</sub>=0.01). The designed *N*-benzyl derivative of compound **27** (compound **54**, Scheme 1) and the corresponding conformational models were thus edited within Catalyst. The predicted IC<sub>50</sub> value for **54** was 0.02 μM, and its mapping onto the top-ranked hypothesis is represented in Figure 11B.

The promising molecular modeling results prompted us to plan the synthesis of a series of new 1H-indole derivatives **48-59** (Scheme 1) bearing a benzyl or 4-fluorobenzyl substituent at

N-1, seeing that an overview of the most active DKA analogues highlighted the presence of a fluorine atom on the benzyl moiety.

Furthermore, a chlorine atom or a methoxy group was introduced on the benzene-fused ring of some of the newly designed benzylindole derivatives (i.e. **50-53**, **56-59**) based on the observation that chloro-substituted compound **44** ( $IC_{50}=0.52$ ) was 2-fold more potent than **27** ( $IC_{50}=1.4$ ), and that a methoxy group was present in potent IN inhibitors such as **23**, **38**, and **39**. The solventless microwave-assisted synthesis was used in many steps (Scheme 1) thus strikingly reducing the reaction times and obtaining almost quantitative yields.

**Table 3.** Inhibition of over-all integration, 3'-processing and strand transfer as  $IC_{50}(\mu M)^a$

Compound	Over-all integration	3'-Processing	Strand transfer
<b>48</b>	$0.76 \pm 0.10$	$>286.2$	$5.40 \pm 4.16$
<b>49</b>	$1.20 \pm 0.0$	$>272.20$	$1.50 \pm 0.57$
<b>50</b>	$0.22 \pm 0.04$	$10.19 \pm 5.83$	$0.03 \pm 0.01$
<b>51</b>	$0.39 \pm 0.14$	$11.47 \pm 2.5$	$0.07 \pm 0.05$
<b>52</b>	$0.15 \pm 0.05$	$162.5 \pm 2.36$	$0.68 \pm 0.42$
<b>53</b>	$0.22 \pm 0.08$	$130.7 \pm 0.0$	$0.80 \pm 0.7$
<b>54</b>	$0.02 \pm 0.0006$	$8.28 \pm 2.21$	$0.03 \pm 0.0006$
<b>55</b>	$0.002 \pm 0.001$	$5.3 \pm 1.25$	$0.015 \pm 0.003$
<b>56</b>	$0.010 \pm 0.003$	$0.97 \pm 0.18$	$0.10 \pm 0.01$
<b>57</b>	$0.20 \pm 0.08$	$1.59 \pm 1.08$	$0.01 \pm 0.001$
<b>58</b>	$0.017 \pm 0.006$	$2.16 \pm 0.68$	$0.021 \pm 0.007$
<b>59</b>	$0.019 \pm 0.008$	$2.71 \pm 0.0$	$0.004 \pm 0.001$
<b>L-870,810</b>	$0.00050 \pm 0.00028$	$0.12 \pm 0.03$	$0.0025 \pm 0.0007$

<sup>a</sup>Concentration required to inhibit by 50% the *in vitro* integrase activity assays

All the synthesized compounds were tested in IN inhibition assays.<sup>21,22</sup> The results showed that the diketo acid derivatives were generally more potent than the corresponding esters and that most of the tested compounds showed 50-100 fold selectivity toward inhibition of strand transfer in comparison to 3'-processing (Table 3).

Our lead compound (**54**) had a strand transfer inhibitory activity of  $0.03 \mu M$ ; the introduction of a chlorine atom on the indole ring (**56**) led to a significant decrease of anti-IN activity, while the presence of a methoxy group in the same position (**58**) did not influence the potency.

The best activity was displayed when a *p*-fluorine atom was present on the benzyl moiety; in fact compound **55** was 2-fold more potent compared to the unsubstituted parent **54**, and fluoro derivatives **57** and **59** were the most active compounds of the series with  $IC_{50}=0.01 \mu M$  and  $IC_{50}=0.004 \mu M$ , respectively.

In particular the contemporary presence of a fluorine atom on the benzyl moiety and a methoxy group on the indole system increased the IN inhibitor activity, and compound **59** showed potency comparable to that of L-870,810, one of the two IN inhibitors in clinical trials. The relevance of our modeling assumptions is also documented by the measured strand transfer inhibitory activity of **54** ( $IC_{50}=0.03 \mu M$ ), which compared well with the predicted value of  $0.02 \mu M$ .

## 2. Analysis and molecular dynamic studies of the full-length integrase-DNA complex

As above-mentioned our research was also addressed to the structure-function study of HIV-1 IN enzyme. In particular we used a modified approach for DNA docking to obtain a model of the full-length integrase-DNA complex<sup>8</sup> on which molecular dynamics studies were carried out.<sup>9</sup>

Retroviral IN is a 32-kDa enzyme composed of one polypeptide chain that folds into three distinct functional domains: the N-terminal domain (residues 1–50), the catalytic core domain (residues 50–212), and the C-terminal domain (residues 212–288). The amino-terminal domain includes a conserved “HH–CC” motif that binds a  $Zn^{2+}$  ion and promotes enzyme multimerization.<sup>23,24</sup> The catalytic domain is composed of a mixed  $\alpha$ -helix and  $\beta$ -sheet motif and contains an absolutely conserved D,D-35-E motif characterized by the three acid residues, D64, D116, and E152.<sup>25-27</sup> The C-domain has been showed to have a non-specific but strong DNA binding activity similar to that of the full-length IN.<sup>25,28</sup>

All three domains bind DNA and each isolated domain forms a homodimer in solution. Even though all three domains are required for full catalytic activity, site-directed mutagenesis experiments have shown that the central core domain is sufficient to promote a reverse integration reaction in vitro, known as “disintegration”, indicating that this region contains the enzymatic catalytic centre.<sup>29,30</sup>

The structures of the three separate domains had been solved by X-ray crystallography or NMR spectroscopy.<sup>31-40</sup> Two HIV-1-IN two-domain structures had been also solved by X-ray crystallography: catalytic core and C-terminal domains<sup>41</sup> and catalytic core and N-terminal domains (residues 1–56)<sup>42</sup> whereas the complete three-dimensional bioactive structure of HIV-1 integrase was still unknown. Due to this absence of complete structural information and since the elucidation of the HIV-1 IN/DNA complex is essential to enabling the target-based drug design, we decided to develop a model of the full-length protein complexed with viral DNA using an automated docking procedure, which was revised in order to satisfy our requirements.

A model of the full-length HIV-1 IN dimer (chains A and B both consisting of residues 1–270) was constructed assembling the experimentally determined structures of the single domains. In order to have a good starting model for further rational drug design, we used the only available IN core domain (residues 56–209) complexed with an inhibitor as the initial structure for our modelling studies.



**Figure 12.** Model obtained by docking the full-length integrase to the viral DNA. Monomers A and B are shown in violet and yellow, respectively, while DNA is shown in green. The three-active site residues D64, D116, and E152 are shown in blue. Amino acids establishing hydrogen bond and salt-bridge interactions with the DNA are labeled and shown as spheres. The conserved adenosine shows the proper orientation of the 3' OH group. The two Mg<sup>2+</sup> metal ions are shown as grey spheres. This figure was prepared using the program PyMOL<sup>44</sup>

This structure presented only one Mg<sup>2+</sup> ion, so the second Mg<sup>2+</sup> ion was placed in each core domain, i.e., chains A and B, in the same relative position according to the two-metal structure of the Avian Sarcoma Virus integrase (PDB code 1VSH), a high structural homolog to HIV-1 IN.<sup>43</sup>

The structure of the C-terminal domain dimer was later extracted by a fragment of HIV IN comprising the catalytic core and C-terminal domains (PDB code 1EX4).<sup>41</sup> In order to obtain the right orientation between our above-described dimer and the C-domains, the conserved catalytic

core domains in the 1EX4 structure were superimposed to our isolated core domains resulting in a plausible two-domain 56–270 dimer.

The N-terminal domains were finally added using the 1K6Y structure<sup>42</sup> that was resolved as a tetramer of catalytic core plus N-terminal domains, giving the nearly identical AB and CD dimers. We used the AB dimer that was connected with our two-domain integrase model, matching common atoms in the catalytic core domains to obtain the full structure of integrase.

Later, the program ESCHER was modified in order to tackle the automated prediction of DNA–protein complexes, and used to dock this IN dimer to the viral DNA.

The best scored docking result suggested that the viral DNA interacted with all three integrase domains in our model and in particular the conserved 3'-CA end at the viral DNA was placed close to the residues of the catalytic site D64, D116, and E152, with the 3'-OH group pointing toward the three acidic residues (Figure 12).

Our IN–DNA complex was thus validated on the basis of known experimental data such as photo-crosslinking and site-directed mutagenesis studies<sup>38,45-48</sup> that revealed the amino acids in close contact with DNA and critical for IN function.

The obtained model has been deposited into Protein Data Bank (PDB ID code is 1WKN)<sup>9</sup> and was used to perform MD simulation studies in order to gain further insights into the dynamics and conformational characteristics of the IN/DNA complex.

Even if several molecular dynamics studies of HIV-1 IN had already been published,<sup>49-54</sup> no attempts to perform MD simulations of IN/DNA complex had been reported.

It is worth underlying that in our structure, looking at the catalytic core domains both of subunits A and B, only the catalytic core of subunit B made direct contacts with the viral DNA.

The substantial difference between the two catalytic sites was fundamental in our study, as the main aim of the present work was to compare the dynamical motion of a surface loop when the viral DNA interacts (subunit B) or does not interact (subunit A) with the core domain. In fact, even if it has been reported that this loop has an important role in the catalytic activity of IN, the bioactive conformation adopted during the integration process is still unknown.

The MD simulation was carried out in 3 phases: initial period of heating from 0 to 300 K over 3000 iterations (3 ps, i.e., 1 K/10 iterations), equilibration period of 150 ps and the production phase of simulation of 1.5 ns. Only the frames memorized during this third phase were considered, with a frame stored each 1000 iterations (1.0 ps), yielding 1500 frames.

The simulation was carried out using NAMD2.51<sup>55</sup> implemented on a Dual-Athlon PC. The atom types were assigned using force field CHARMM v22 and the atomic charges according the Gasteiger–Marsili method. The trajectory obtained was analyzed using VEGA.<sup>56</sup>

We have focused our attention on the dynamical behaviour of the surface loop comprising residues 140–149 to gain insights into the conformational changes of this region in the two subunits. Table 4 reports the distances between the middle point of the segment connecting the Mg<sup>2+</sup> ions (that remain fixed due to the applied constraints) and the C $\alpha$  atoms of loop 140–149 for both subunits.



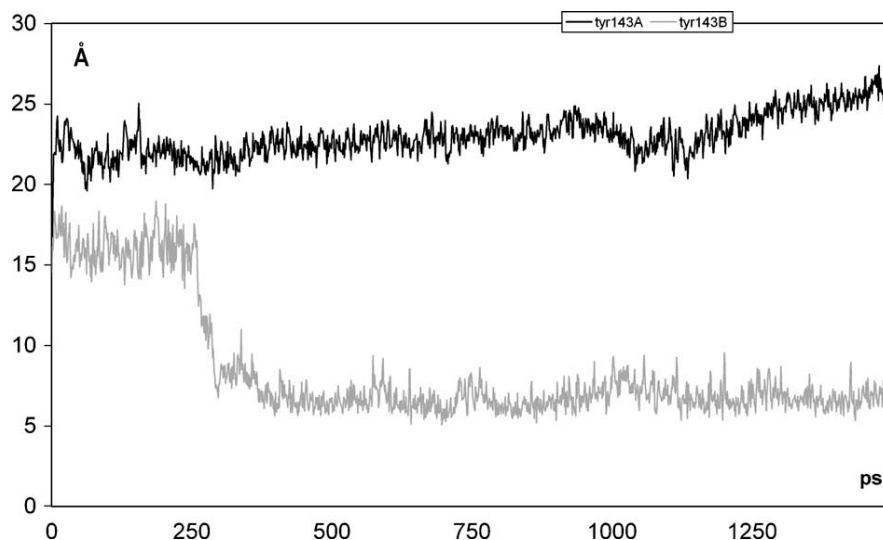
**Table 4.** Distance between Mg<sup>2+</sup> ions and C $\alpha$  for loop residues in both subunits and relative differences

Residue	Subunit A <sup>a</sup>	Subunit B <sup>a</sup>	Difference <sup>b</sup>
Gly140	9.75 to 21.97	8.55 to 17.99	3.87
	15.34 ( $\pm$ 2.67)	11.47 ( $\pm$ 1.83)	
Ile141	12.63 to 24.54	8.77 to 17.22	6.76
	17.97 ( $\pm$ 2.36)	11.21 ( $\pm$ 1.79)	
Pro142	15.04 to 24.11	5.98 to 20.23	7.62
	18.32 ( $\pm$ 1.86)	10.70 ( $\pm$ 3.65)	
Tyr143	16.11 to 26.90	6.80 to 18.00	11.20
	21.28 ( $\pm$ 1.83)	10.08 ( $\pm$ 2.92)	
Asn144	17.53 to 25.42	9.81 to 19.07	7.75
	20.69 ( $\pm$ 1.59)	12.94 ( $\pm$ 2.09)	
Pro145	16.39 to 24.51	11.25 to 19.83	5.61
	20.26 ( $\pm$ 1.43)	14.65 ( $\pm$ 1.57)	
Gln146	12.69 to 20.88	12.61 to 17.05	2.48
	16.78 ( $\pm$ 1.41)	14.30 ( $\pm$ 0.64)	
Ser147	9.95 to 19.22	10.09 to 14.41	1.94
	14.19 ( $\pm$ 1.61)	12.25 ( $\pm$ 0.64)	
Gln148	6.57 to 16.31	6.59 to 10.65	2.20
	10.75 ( $\pm$ 1.84)	8.54 ( $\pm$ 0.63)	
Gly149	7.39 to 13.92	5.89 to 11.73	2.03
	10.08 ( $\pm$ 1.09)	8.05 ( $\pm$ 0.77)	

<sup>a</sup> In each cell the first line reports minimum and maximum values, while the second line is the average  $\pm$  the standard deviation; <sup>b</sup> this column reports the differences between the averages in the two subunits.

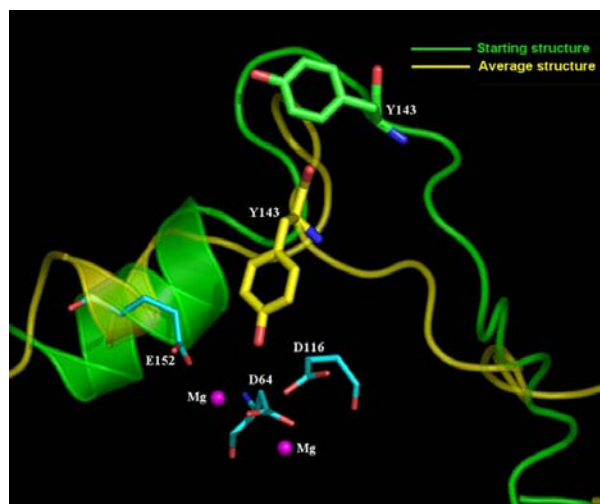
The corresponding differences showed a markedly different dynamical behaviour of the residues in the loop in chains A and B, indicating that the loop of subunit B approaches the Mg<sup>2+</sup> ions, while such a conformational shift is not seen in subunit A where the loop region displays very low mobility during the entire molecular dynamics simulation. Tyr143 is the residue that shows the maximum approaching to metal ions: as seen in Figure 13, the distance between Tyr143 atom OH and the middle point of the segment connecting Mg<sup>2+</sup> ions always appears greater than 17 Å in subunit A, while in subunit B this distance decreases up to 7 Å.

A detailed analysis of the different behaviour of Tyr143 in the two subunits allowed us to shed light on the marked approach of this residue to metal ions, as seen only in subunit B, which is due to both a different conformational shift of backbone atoms of the surface loop and a different conformational profile of the Tyr143 side chain.



**Figure 13.** Dynamic profile of interatomic distance between Tyr143 atom OH and the middle point of the segment connecting  $Mg^{2+}$  ions over the trajectory (gray line, B subunit; black line, A subunit).

The conformational changes of the loop and in particular of Tyr143 orientation between the starting structure (i.e., after the equilibration phase) and average structure from the MD trajectory are pointed out in Figure 14, where only subunit B is shown.



**Figure 14.** Superimposed starting (green) and average (yellow) structures from the MD simulation of 1WKN. This figure was prepared using the program PyMOL.<sup>44</sup>

In the equilibrated position the Tyr143 side-chain points far away from the catalytic site; conversely, the average structure shows the rearrangement of the loop with Tyr143 pointing toward the three conserved residues and the two metal ions in the active site.

Our observation that Tyr143 has been substantially deformed in the presence of viral DNA (i.e., subunit B) supports the hypothesis that its conformation and mobility are important for the catalytic activity of IN and that this residue might be a central player in the mechanism of the integration process.

## Acknowledgements

Financial support for this research by Fondo Ateneo di Ricerca (2002, Messina, Italy), MIUR (COFIN2004, Roma, Italy), and the TRIoH project (LSHB-CT-2003-503480) is gratefully acknowledged.

## References

1. Imamichi, T. *Curr Pharm Des* **2004**, *10*, 4039.
2. Anthony, N. J. *Curr Top Med Chem* **2004**, *4*, 979.
3. Billich, A. *Curr Opin Investig Drugs* **2003**, *4*, 206.
4. Pais, G. C.; Burke, T. R., Jr. *Drugs Future* **2002**, *27*, 1101.
5. Grobler, J. A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, J.; Hazuda, D. J. *Proc Natl Acad Sci U S A* **2002**, *99*, 6661.
6. Barreca, M. L.; Rao, A.; De Luca, L.; Zappala, M.; Gurnari, C.; Monforte, P.; De Clercq, E.; Van Maele, B.; Debyser, Z.; Witvrouw, M.; Briggs, J. M.; Chimirri, A. *J Chem Inf Comput Sci* **2004**, *44*, 1450.
7. Barreca, M. L.; Ferro, S.; Rao, A.; De Luca, L.; Zappalà, M.; Monforte, A. M.; Debyser, Z.; Witvrouw, M.; Chimirri, A. *J Med Chem* **2005**, *Asap article*.
8. De Luca, L.; Pedretti, A.; Vistoli, G.; Barreca, M. L.; Villa, L.; Monforte, P.; Chimirri, A. *Biochem Biophys Res Commun* **2003**, *310*, 1083.
9. De Luca, L.; Vistoli, G.; Pedretti, A.; Barreca, M. L.; Chimirri, A. *Biochem Biophys Res Commun* **2005**, *336*, 1010.
10. Pais, G. C.; Zhang, X.; Marchand, C.; Neamati, N.; Cowansage, K.; Svarovskaia, E. S.; Pathak, V. K.; Tang, Y.; Nicklaus, M.; Pommier, Y.; Burke, T. R., Jr. *J Med Chem* **2002**, *45*, 3184.
11. Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W.; Pommier, Y. *J Med Chem* **1997**, *40*, 920.
12. Zhuang, L.; Wai, J. S.; Embrey, M. W.; Fisher, T. E.; Egbertson, M. S.; Payne, L. S.; Guare, J. P., Jr.; Vacca, J. P.; Hazuda, D. J.; Felock, P. J.; Wolfe, A. L.; Stillmock, K. A.; Witmer, M. V.; Moyer, G.; Schleif, W. A.; Gabryelski, L. J.; Leonard, Y. M.; Lynch, J. J., Jr.; Michelson, S. R.; Young, S. D. *J Med Chem* **2003**, *46*, 453.

13. Sybyl 6.9; Tripos Associate Inc.: St. Louis, M., 2003.
14. Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. *Science* **2000**, *287*, 646.
15. Wai, J. S.; Egbertson, M. S.; Payne, L. S.; Fisher, T. E.; Embrey, M. W.; Tran, L. O.; Melamed, J. Y.; Langford, H. M.; Guare, J. P., Jr.; Zhuang, L.; Grey, V. E.; Vacca, J. P.; Holloway, M. K.; Naylor-Olsen, A. M.; Hazuda, D. J.; Felock, P. J.; Wolfe, A. L.; Stillmock, K. A.; Schleif, W. A.; Gabryelski, L. J.; Young, S. D. *J Med Chem* **2000**, *43*, 4923.
16. Dayam, R.; Neamati, N. *Curr Pharm Des* **2003**, *9*, 1789.
17. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv Drug Deliv Rev* **2001**, *46*, 3.
18. Dayam, R.; Sanchez, T.; Clement, O.; Shoemaker, R.; Sei, S.; Neamati, N. *J Med Chem* **2005**, *48*, 111.
19. Johnson, A. A.; Marchand, C.; Pommier, Y. *Curr Top Med Chem* **2004**, *4*, 1059.
20. Hazuda, D. J.; Young, S. D.; Guare, J. P.; Anthony, N. J.; Gomez, R. P.; Wai, J. S.; Vacca, J. P.; Handt, L.; Motzel, S. L.; Klein, H. J.; Dornadula, G.; Danovich, R. M.; Witmer, M. V.; Wilson, K. A.; Tussey, L.; Schleif, W. A.; Gabryelski, L. S.; Jin, L.; Miller, M. D.; Casimiro, D. R.; Emini, E. A.; Shiver, J. W. *Science* **2004**, *305*, 528.
21. Debyser, Z.; Cherepanov, P.; Pluymers, W.; De Clercq, E. *Methods Mol Biol* **2001**, *160*, 139.
22. Witvrouw, M.; Van Maele, B.; Vercammen, J.; Hantson, A.; Engelborghs, Y.; De Clercq, E.; Pannecouque, C.; Debyser, Z. *Curr Drug Metab* **2004**, *5*, 291.
23. Zheng, R.; Jenkins, T. M.; Craigie, R. *Proc Natl Acad Sci U S A* **1996**, *93*, 13659.
24. Lee, S. P.; Xiao, J.; Knutson, J. R.; Lewis, M. S.; Han, M. K. *Biochemistry* **1997**, *36*, 173.
25. Engelman, A.; Craigie, R. *J Virol* **1992**, *66*, 6361.
26. J. Kulkosky, K. S. J., R.A. Katz, J.P. Mack and A.M. Skalka. *Mol. Cell. Biol* **1992**, *12*, 2331.
27. Polard, P.; Chandler, M. *Mol Microbiol* **1995**, *15*, 13.
28. Vink, C.; Oude Groeneger, A. M.; Plasterk, R. H. *Nucleic Acids Res* **1993**, *21*, 1419.
29. Chow, S. A.; Vincent, K. A.; Ellison, V.; Brown, P. O. *Science* **1992**, *255*, 723.
30. Bushman, F. D.; Engelman, A.; Palmer, I.; Wingfield, P.; Craigie, R. *Proc Natl Acad Sci U S A* **1993**, *90*, 3428.
31. Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, *266*, 1981.
32. Bujacz, G.; Alexandratos, J.; Qing, Z. L.; Clement-Mella, C.; Wlodawer, A. *FEBS Lett* **1996**, *398*, 175.
33. Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. *J Mol Biol* **1998**, *282*, 359.
34. Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc Natl Acad Sci U S A* **1999**, *96*, 13040.
35. Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. *Biochemistry* **1999**, *38*, 8892.

36. Eijkelenboom, A. P.; Lutzke, R. A.; Boelens, R.; Plasterk, R. H.; Kaptein, R.; Hard, K. *Nat Struct Biol* **1995**, *2*, 807.
37. Eijkelenboom, A. P.; Sprangers, R.; Hard, K.; Puras Lutzke, R. A.; Plasterk, R. H.; Boelens, R.; Kaptein, R. *Proteins* **1999**, *36*, 556.
38. Lodi, P. J.; Ernst, J. A.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1995**, *34*, 9826.
39. Cai, M.; Huang, Y.; Caffrey, M.; Zheng, R.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Protein Sci* **1998**, *7*, 2669.
40. Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Nat Struct Biol* **1997**, *4*, 567.
41. Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc Natl Acad Sci U S A* **2000**, *97*, 8233.
42. Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. *Embo J* **2001**, *20*, 7333.
43. Bujacz, G.; Jaskolski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. *Structure* **1996**, *4*, 89.
44. W.L. DeLano, T. P. M. G. S. o. W. W. W. A. f. h. w. p. o.
45. Heuer, T. S.; Brown, P. O. *Biochemistry* **1998**, *37*, 6667.
46. Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. *Embo J* **1997**, *16*, 6849.
47. Gerton, J. L.; Ohgi, S.; Olsen, M.; DeRisi, J.; Brown, P. O. *J Virol* **1998**, *72*, 5046.
48. Lutzke, R. A.; Plasterk, R. H. *J Virol* **1998**, *72*, 4841.
49. Brigo, A.; Lee, K. W.; Fogolari, F.; Mustata, G. I.; Briggs, J. M. *Proteins* **2005**, *59*, 723.
50. Brigo, A.; Lee, K. W.; Iurcu Mustata, G.; Briggs, J. M. *Biophys J* **2005**, *88*, 3072.
51. Lee, M. C.; Deng, J.; Briggs, J. M.; Duan, Y. *Biophys J* **2005**, *88*, 3133.
52. Barreca, M. L.; Lee, K. W.; Chimirri, A.; Briggs, J. M. *Biophys J* **2003**, *84*, 1450.
53. Laboulais, C.; Deprez, E.; Leh, H.; Mouscadet, J. F.; Brochon, J. C.; Le Bret, M. *Biophys J* **2001**, *81*, 473.
54. Weber, W.; Demirdjian, H.; Lins, R. D.; Briggs, J. M.; Ferreira, R.; McCammon, J. A. *J Biomol Struct Dyn* **1998**, *16*, 733.
55. Kalé, L.; Bhandarkar, R.; Brunner, M.; Gursoy, R.; Krawetz, A.; Phillips, N.; Shinozaki, J.; Varadarajan, A.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 286.
56. Pedretti, A.; Villa, L.; Vistoli, G. *J Mol Graph Model* **2002**, *21*, 47.