PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Computational studies of the interaction between the HIV-1 integrase tetramer and the cofactor LEDGF/p75: Insights from molecular dynamics simulations and the informational spectrum method

Cristina Tintori,[1] Nevena Veljkovic,[2] Veljko Veljkovic,[2] and Maurizio Botta[1]*

[1] Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Alcide de Gasperi 2, I-53100 Siena, Italy

[2] Center for Multidisciplinary Research and Engineering, Institute of Nuclear Sciences VINCA, P.O. Box 522, 11001 Belgrade, Serbia

ABSTRACT

A crystal structure of the integrase binding domain (IBD) of the lens epithelium-derived growth factor (LEDGF/p75) in complex with the dimer of the HIV-1 integrase (IN) catalytic core domain (CCD) provides useful information that might help in the understanding of essential protein-protein contacts in HIV-1. However, mutagenic studies indicated that interactions between the full-length proteins were more extensive than the contacts observed in the co-crystal structure of the isolated domains. On the other hand, the biochemical characterization of the interaction between full-length IN and LEDGF/p75 has recently proved that LEDGF/p75 promotes IN tetramerization with two LEDGF/p75 IBD molecules bound to the IN tetramer. This experimental evidence suggests that to obtain a complete structural description of the interactions between the two proteins, the full-length tetrameric structure of IN should be considered. Our aim was to obtain a detailed picture of HIV-1 IN interactions with cellular co-factors that was of general interest, particularly for the development of small molecule IN inhibitors, which mimic the IBD of LEDGF/p75. To this end, we performed bioinformatics analyses to identify protein sequence domains involved in long-range recognition. Subsequently, we applied molecular dynamics techniques to investigate the detailed interactions between the complete tetrameric form of IN and two molecules of the IBD of LEDGF/p75. Our dynamic picture is in agreement with experimental data and, thereby, provides new details of the IN-LEDGF/p75 interaction.

Proteins 2010; 78:3396–3408.
© 2010 Wiley-Liss, Inc.

Key words: hot spots; IN Oligomerization; IN tetramer; protein-protein interactions.

INTRODUCTION

HIV-1 IN is an essential enzyme for the integration of viral DNA into the host cell genome. The process of DNA incorporation occurs in two spatially and temporally distinct steps known as 3′ processing and strand transfer. The 3′ processing occurs in the cytoplasm, where IN binds the viral DNA and then removes a dinucleotide adjacent to a conserved CA sequence from the 3′ end of each strand.[1,2] Here, IN is supposed to be in equilibrium between its monomeric, dimeric, tetrameric, and high order oligomeric states. Only the dimer is able to bind the viral DNA, and, thus, it represents the active form of the enzyme.[3,4] The protein-DNA complex is then transported into the nucleus by several cofactor such as Transportin-SR2 (TR2), recently discovered by Christ *et al.*[5]

After nuclear import, probably two DNA-bound dimers approach each other in the presence of the cellular protein LEDGF/p75 and form a tetramer. Subsequently, integration proceeds to the strand-transfer step (Fig. 1). The free IN tetramer presumably does not bind DNA directly, and tetramerization occurs only by the interaction between two DNA-bound dimers.[3,4] HIV-1 IN activity requires binding to the LEDGF/p75, which is known to activate IN in cell studies and *in vitro* by tethering it to the host chromosomes.[6] Although low concentrations of LEDGF/p75 stimulate IN binding to DNA as well as its enzymatic activity,[7] overexpression of the LEDGF/p75 IBD inhibits HIV-1 replication and blocks nuclear import of IN.[8] It has been recently hypothesized that the inhibition of IN by overexpression of LEDGF/p75 could be explained by a shiftide mechanism.

Grant sponsor: European Union—“THINC”, Collaborative Project, grant number HEALTH-2007-2.3.2-1. Grant sponsor: European Union—“CHAARM”, Collaborative Project, grant number HEALTH-F3-2009-242135. Grant sponsor: Ministero dell’Istruzione, dell’Università e della Ricerca—Prin 2008 research project, grant number 2008CE75SA_004. Grant sponsor: MSTD Republic of Serbia, grant number 143001.
*Correspondence to: Maurizio Botta, Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Alcide de Gasperi, 2, I-53100 Siena, Italy. E-mail: botta@unisi.it.
Received 26 May 2010; Revised 22 July 2010; Accepted 3 August 2010
Published online 17 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22847

© 2010 Wiley-Liss, Inc.

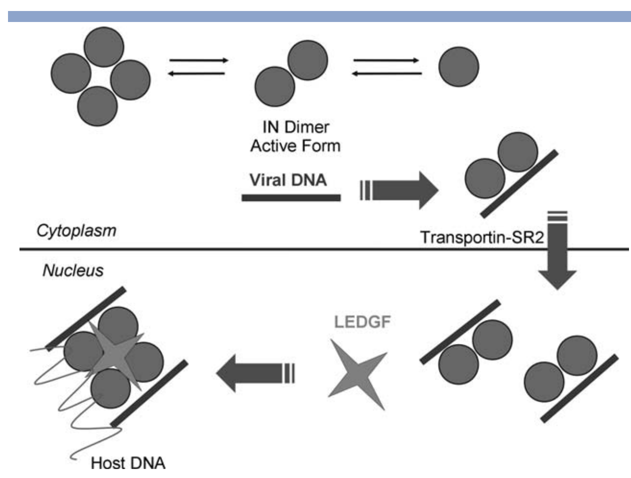


Figure 1
Hypothetical oligomerization IN equilibrium.

Briefly, this hypothetical model suggests that the IBD of LEDGF/p75 binds the IN tetramer in the cytoplasm, which shifts the IN oligomerization equilibrium from the active dimer toward the inactive tetramer.[9–11]

A crystal structure of the integrase binding domain (IBD) of the lens epithelium-derived growth factor (LEDGF/p75) in complex with the dimer of the HIV-1 integrase (IN) catalytic core domain (CCD) provides useful information that might help in the understanding of essential protein-protein contacts in HIV-1.[12] The availability of such crystal structure has made possible the identification of new inhibitors able to disrupt the interaction between LEDGF/p75 and IN.[13–15] In this regard, a series of derivatives (LEDGINs) that act as potent inhibitors of the LEDGF/p75-IN interaction and of HIV-1 replication has been recently disclosed.[15]

Despite the great impact of this crystallographic structure on the development of new anti-HIV-1 inhibitors, mutagenic studies indicated that interactions between the full-length LEDGF/p75 and IN proteins were more extensive than the contacts that were observed in the co-crystal structure of the isolated domains. In particular, Maertens *et al.*[16] have shown that not only is the CCD of IN involved in the interaction with LEDGF/p75, but the N-terminal zinc binding domain also plays an important role. In fact, both domains are essential for nuclear localization of IN as well as for the association of IN with condensed chromosomes during mitosis. Moreover, biochemical characterizations of the interactions between full-length proteins have recently shown that LEDGF/p75 could promote IN tetramerization with two LEDGF/p75 IBD molecules bound to the IN tetramer.[17] In light of these results, a three-dimensional model of the LEDGF/p75 IBD-IN/tetramer complex provides a detailed description of the contacts between the two interacting macromolecules and could act as the starting point for the design of new HIV-1 inhibitors.

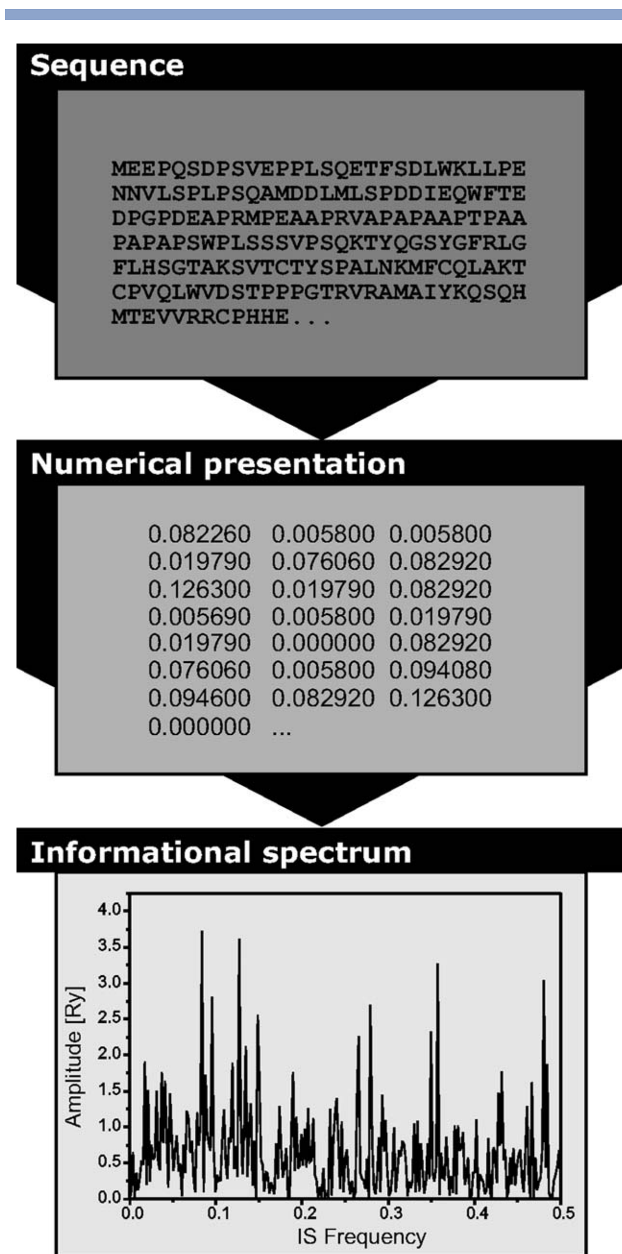
The informational spectrum method[18] (ISM), is a sequence analysis method that allows one to determine nonobvious physicochemical patterns that correlate with protein interaction domains. The ISM represents a bioinformatics concept that is useful when trying to explore the long-range properties of biological macromolecules. Interactions between biological molecules depend not only on short range interactions but also on interactions governed by long-range attractive forces which are efficient at a distance longer than one linear dimension of the interacting macromolecules (for a detailed review of the physical basis of this concept see Ref. 19). Briefly, the concordance of long-range interaction (LRI) properties between interacting molecules results in an acceleration of biochemical processes by an increase of the number of productive collisions in comparison with the accidental encounter of molecules. The ISM,[18] is a sequence analysis method that allows one to determine nonobvious physicochemical patterns that correlate with long-range properties (Scheme 1).

ISM has been used successfully several times as an initial step in the identification of partner protein interactors and interacting protein domains.[20–25] In the present work, we used the ISM to predict HIV-1 IN domains that were potentially involved in an interaction with LEDGF/p75. On the other hand, molecular dynamics techniques[26] together with the molecular mechanics-generalized Born surface area (MM/GBSA)[27] approach were applied to investigate the detailed interactions between the complete tetrameric form of IN and two molecules of LEDGF/p75 IBD. As a result, we identified amino acids of both interacting proteins that make strong and recurring interactions in the protein complex. To study the role of LEDGF/p75 in the stabilization of the IN, molecular dynamics (MD) simulations were also performed on the free tetrameric form of integrase (unbound with LEDGF/p75). By comparing the two trajectory profiles, it was possible to show that while the IN-LEDGF/p75 complex is stable during the time of simulation, the IN tetramer had very different behavior when it was free. These results are in agreement with biochemical experiments.[17] Our results provide insight into the molecular basis for the recognition between IN and LEDGF/p75, and they will be useful for the design of new anti-HIV-1 inhibitors.

METHODS

Informational spectrum method

Understanding how molecules recognize each other is very important to protein engineering and drug design. Success of this recognition mainly depends on structural complementarities between interacting molecules. Under this assumption, one only needs to consider short-range interactions (SRIs) between a drug and its target or between two proteins. However, it is well-established that while SRIs do play key roles in the recognition of two



Scheme 1

The ISM procedure. The first phase involves the transformation of the amino acid sequence symbolized in alphabetic code into a numerical sequence. This is done by presenting each residue in the sequence by the value of chemical descriptor, the electron-ion interaction potential (EIIP) which corresponds to the average energy states of all valence electrons in a particular amino acid. In the second phase, the numerical sequence is transformed into the virtual frequency domain to create an ISM spectrum by using discreet Fourier transformation.

molecules, LRIs are equally important. The ISM is a bioinformatics approach that allows for the consideration of the long-range properties of the interacting molecules by examining patterns of electron-ion interaction potential (EIIP), the amino acid parameter representing the main energy term of the valence electrons. It has been shown that EIIP is an essential physical parameter that determines the long-range properties of biological molecules and is most suitable for analysis of protein-protein interactions in this context. The ISM procedure encompasses two stages. The first involves the transformation of the amino acid sequence into a numerical sequence by assigning to each amino acid the value of the EIIP, which was calculated using the general pseudopotential model. By using the discrete Fourier transform, the numerical sequence was then transformed into the frequency domain to create an informational spectrum. The absolute value of the complex Fourier transform defines the amplitude spectrum that is relevant for protein analyses. Because the average distance between residues in a protein chain is 3.8 Å, it is assumed that the points in each derived series are equidistant, and, thus, the distance is arbitrarily set equal to 1. The maximum frequency in the spectrum is $F = 1/2d = 0.5$.

According to the ISM, the IS spectra of interacting proteins share common frequency components. To identify the common spectral characteristic(s) of the two sequences, mathematical filtering is performed by multiplication of the conjugate complex Fourier transform by the Fourier transform of the target signal. The result of the multiplication is the cross-spectral (CS) function. The prominent peaks in this function denote the common frequency components of the analyzed proteins. The numerical series that are derived from the analyzed sequences are normalized to zero mean and are zero padded to produce a vector that is equal in length to the smallest power of two greater than (or equal to) the largest domain in the data set.

Here, we have applied the ISM to identify important informational characteristic of the interaction between LEDGF/p75 and HIV-1 IN and to predict the protein domains potentially involved in interaction. Protein sequences LEDGF/p75 (O75475) and HIV-1 IN (P03366) were retrieved from UniProtKB/Swiss-Prot.

Computational peptide mapping

Peptide mapping was used to define linear protein fragments that contribute the most to the amplitude at the characteristic frequency and are, therefore, responsible for interactions described by that particular spectral characteristic. The sequences of HIV-1 IN and LEDGF/p75 were scanned by the ISM algorithm as overlapping polypeptides with a one-residue shift, which led to the identification of all deletions that decrease the amplitude at the corresponding characteristic frequency. Vanegas *et al.*[6] have shown that the 67-residues domain are necessary and sufficient to mediate the IN interaction, therefore we decided to screen the protein sequences with 50–70 residues windows.

Molecular systems

In light of the fact that the structure of full-length HIV-1 IN has yet to be determined, a whole IN tetramer

was constructed by assembling the available structures in the Protein Data Bank (PDB), as already reported by other authors.[28]

In detail, the three-dimensional structure 1K6Y[29] was used as the initial template. First, the CCDs of 1K6Y were completed by superimposing chain C of the 1BL3 structure.[30] Then, the CCD of PDB structure 1EX4[31] was superimposed with 1K6Y to add the C-terminal domains. Next, in the N-terminal domains, the lacking nine residues 47–55 were correctly placed through alignment with the 1WJD[32] structure. The resulting tetrameric structure (named the free IN tetramer) was used as the template for the alignment of the CCD-IBD (PDB code 2B4J)[12] to place the two LEDGF/p75 molecules in the correct position within the complex of the IN tetramer (named the IN tetramer-LEDGF/p75 complex). The two assemblies, visualized in Figure 2, were utilized as the starting point for MD simulations.

MD simulations

Molecular dynamic simulations have been performed to optimize the protein systems and to calculate the free energy of binding through use of the MM/GBSA. The AMBER 10 suite of programs and the parm94 force field were used.[33] A nonbonded approach was used for the catalytic zinc atoms. Zinc was assigned a formal charge of +2.0, a van der Waal's radius 1.10 Å and a well depth ε = 0.25 kcal/mol according to previously derived parameters.[34] An appropriate number of counter-ions (36 Cl^- ions for free tetrameric IN system and 40 Cl^- ions for the tetramer IN-LEGDF/p75 IBD system) were added to neutralize the system, and each of the two starting structures was placed in a rectangular parallelepiped box of TIP3P water molecules. The distance between the box walls and the protein was set to 8 Å. Both MD runs were set up with the same protocol widely validated in previous studies on similar protein systems.[35,36] Before MD simulations, two stages of energy minimization were performed to remove bad contacts. In the first stage, we kept the protein fixed with a constraint of 500 kcal/mol, and we minimized the positions of the water molecules. Then, in the second stage, we minimized the entire system, applying a constraint of 20 kcal/mol on the α carbon. The two minimization stages consisted of 5000 steps in which the first 1000 were steepest descent, and the last 4000 were conjugate gradient. MD trajectories were run using the minimized structure as a starting input. Constant volume simulations were performed for 1000 ps, during which time the temperature was raised from 0 to 300 K using the Langevin dynamics method. Then, 3000 ps of constant-pressure MD simulations were performed at 300 K in three steps of 1000 ps each. During the three periods of this second stage, the α carbons were blocked with harmonic force constants of 10, 5, and 1 kcal/mol·Å, respectively. Finally, a 24 ns MD simulation without restraint was run at a constant temperature of 300 K and a constant pressure of 1 atm. During the simulations, the Particle Mesh Ewald method was employed to calculate the long-range electrostatic interactions, while the SHAKE method was applied to constrain all covalent bonds involving hydrogen atoms. A 10-Å cutoff value was used for the nonbonded interactions, and a time step of 2 fs was used for the simulations.

Free energy decomposition

The MM-GBSA approach, which was implemented in the AMBER program,[33] was chosen to decompose the binding free energy between the full-length tetramer of

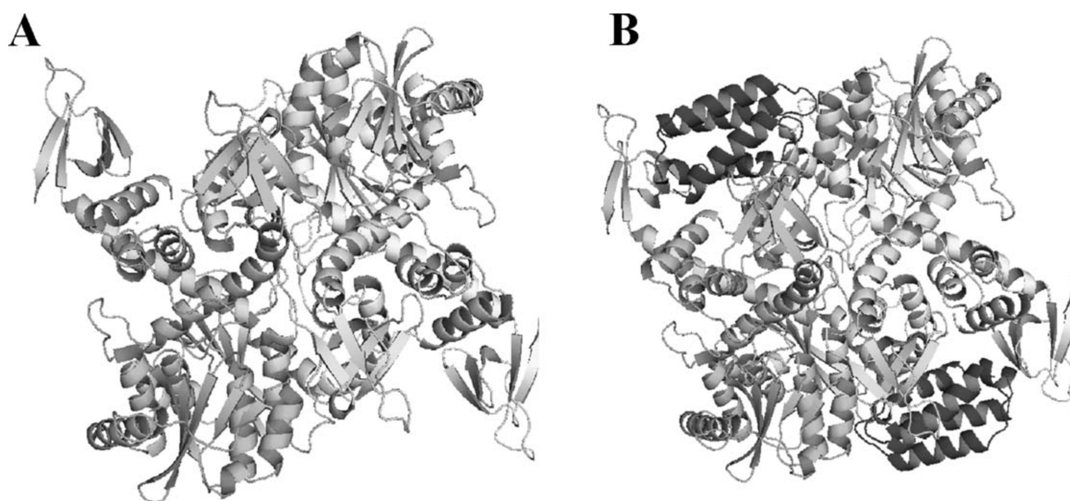


Figure 2
(A) Free IN tetramer and (B) IN tetramer-LEDGF/p75 complex.

HIV-1 IN and two molecules of LEDGF/p75 into individual amino acid contributions to map the interactions that have a dominant role in the complex formation. The MM-GBSA utilizes the structural information derived from MD simulations to develop estimates of binding free energies in a posteriori analysis of the trajectories of each structure followed by energy component averaging. The purpose of this analysis is to generate structures with explicit solvent and transfer them to continuum solvent free energy evaluations thus rendering the free energy problem computationally tractable. Even if several approximations are introduced in the MM/GBSA method that preclude calculations of absolute values of the binding free energies, this methodology can be applied successfully to identify the energy "hot spot" residues in protein-protein complexes.[37,38]

The last 12 ns of IN tetramer-LEDGF/p75 IBD complex simulation was regarded as stable and was used to extract 100 snapshots from each trajectory with intervals of 120 ps between them. Briefly, in the MM-GBSA approach, the protein-protein binding free energy ($\Delta G_{binding}$) for each snapshot is estimated using Eq. (1):

$$\Delta G_{binding} = [G_{complex}] - [G_{protein1}] - [G_{protein2}] \quad (1)$$

where $G_{complex}$, $G_{protein1}$, and $G_{protein2}$ are the free energies of the complex, protein1, and protein2, respectively. Such energies were calculated based on an average taken over snapshots from the MD trajectories. Each binding free energy term in Eq. (1) is calculated using Eq. (2):

$$G = E_{gas} + G_{solvation} - TS \quad (2)$$

where E_{gas} is the gas phase energy and is calculated by Eq. (3) with dielectric constant $\varepsilon = 1.0$ for electrostatic interaction:

$$E_{gas} = E_{int} + E_{vdw} + E_{electrostatic} \quad (3)$$

where $G_{solvation}$ is the solvation free energy and is computed with Eq. (4):

$$G_{solvation} = G_{GB} + G_{nonpolar} \quad (4)$$

The polar contribution (G_{GB}) was calculated with the modified GB model described by Onufriev *et al.* using $\varepsilon_w = 80.0$ and $\varepsilon_p = 1.0$. The nonpolar contribution ($G_{nonpolar}$) due to cavity formation and van der Waals interactions between the solute and the solvent was estimated using Eq. (5):

$$G_{nonpolar} = \gamma^* \text{ SASA} + b \quad (5)$$

Solvent-accessible surface area (SASA) was determined by recursively approximating a sphere around an each atom, starting from an icosahedra (ICOSA method).

The surface tension proportionality constant, γ, and the free energy of nonpolar solvation for a point solute B were set to 0.0072 kcal/(mol Å^2) and 0.00 kcal/mol, respectively. The radius of the probe sphere used to calculate the SASA was set to 0.0. As reported by other authors, the contribution of the entropy term is negligible. Thus, the entropy term was not calculated in this work.

Finally, binding free energy decomposition per residue was used to evaluate the contribution of each residue at an atomic level. The contribution of each residue can be expressed by Eq. (6):

$$\Delta G_{res-binding} = \Delta E_{res-int} + \Delta E_{res-vdw} + \Delta E_{res-electrostatic} + \Delta G_{res-GB} + \Delta G_{res-nonpolar} - T\Delta S_{res} \quad (6)$$

The terms in Eq. (6) are the same as described above.

Each atom's contribution to the terms in Eq. (6) was first calculated. Then, by summing these atomic contributions over the atom of a given residue, its contribution to the binding free energy, $\Delta G_{res\text{-}binding}$, was obtained.

Analysis of MD trajectories

The module Ptraj implemented in AMBER 10[33] was used to analyze trajectories. In particular, the root mean square deviations and the root mean square fluctuations (RMSFs) of the α carbons of each residue were calculated based on data from the production stage. During the same time of simulation, the hydrogen bonds between IN and LEDGF/p75 were detected when the acceptor-donor atom distance was lower than 3.5 Å and the acceptor-H-donor angle was more that 120°. Ptraj was also used to monitor the distance between atoms involved in salt-bridges.

RESULTS

Informational spectrum method: LEDGF/p75 domains in interaction with HIV-1 IN

ISM analyses was performed to predict IN domains that were potentially involved in an interaction between HIV-1 IN and cellular LEDGF/p75. ISs for both proteins were calculated [Figs. 3(A, B)], and common frequencies were determined by CS function analysis [Fig. 3(C)]. In general, the presence of common frequency components in the two ISs was indicative of a mutual interaction. These frequencies are denoted as characteristic frequencies and are determined by mathematical filtering using a CS. The dominant peak in the HIV-1 IN-LEDGF/p75 CS corresponds to the frequency $F(0.4209)$.

Further, the computational mapping based on the ISM algorithm was applied to identify the protein domains that contribute the most to the amplitude at the IS frequency $F(0.4209)$. The decrease in amplitude of the IS in

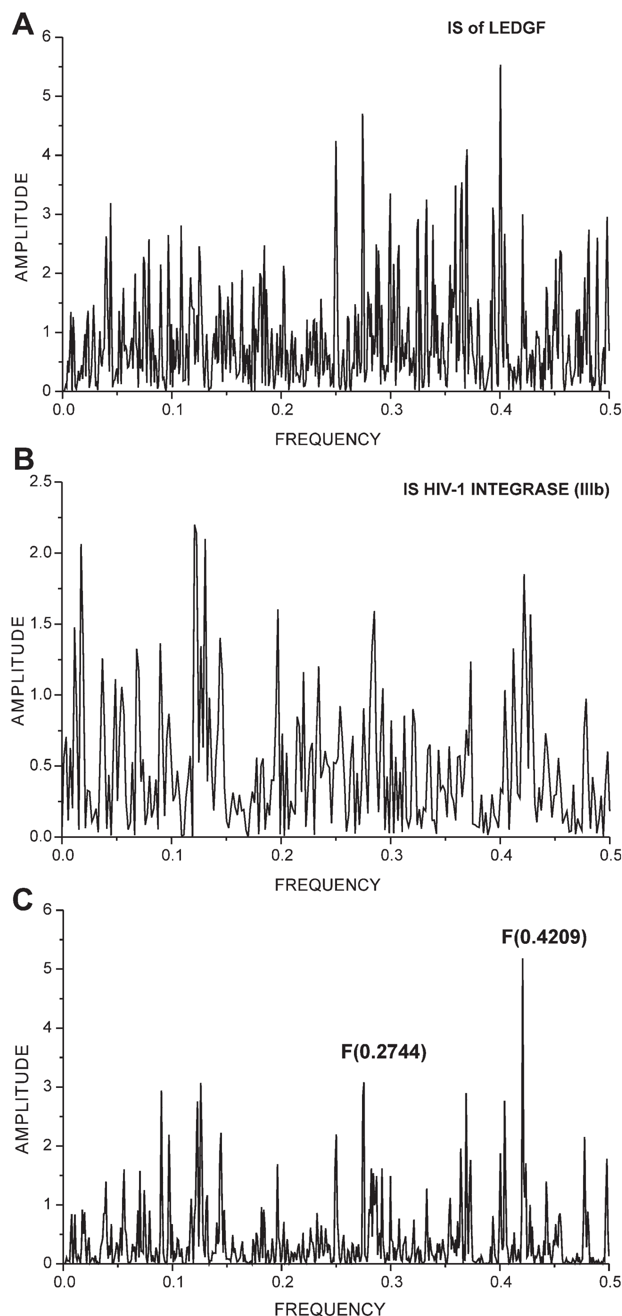


Figure 3

(A) IS for LEDGF/p75, (B) IS for IN, and (C) CS between LEDGF/p75 and IN.

both proteins at $F(0.4209)$ was calculated after performing theoretical deletions of different lengths within the primary sequence window of size 50 to 70 residues. The protein domain in which deletion decreased the amplitude the most was considered to be the peptide contributing the most to the intensity of the virtual signal at the predefined frequencies. In LEDGF/p75, this domain encompassed residues 381 to 430 (Fig. 4), the sequence that contains residues already found to be important for binding to IN by means of crystallographic studies. In Figure 4, two additional low peaks around positions 80 and 450 that are outside the IBD can be observed. According to the ISM concept, these domains potentially cooperate with key identified domains in interactions characterized by the IS frequency 0.4209. It could be assumed that although these domains are not necessary for the direct interaction with HIV-1 IN, they contribute to the long-range recognition process.

With regard to the IN sequence, two different important regions were detected (Fig. 5). The most important region involved the residues 159–222, which is also in agreement with the crystallographic data.[12] The second region refers to the N-terminal domain of IN (residue 1–64) suggesting that this region could also be involved in an interaction with LEDGF/p75.

Analysis of MD simulations

The last 24 ns of the trajectories of the simulations without restraints were regarded as the production stage and saved for analysis. The root mean square deviations (RMSDs) of the α carbons were calculated throughout the simulations with respect to the starting structure for both the free tetrameric IN system and the tetramer IN-LEGDF/p75 IBD complex. The RMSDs of the protein Cαs allowed us to check whether the protein reached the equilibration state during the MD simulations. Results are graphically represented in Figure 6. After the first 12 ns, the structure of the tetramer IN-LEGDF/p75 IBD complex was fairly stable during the remaining MD simulation. A different profile was found for free tetrameric IN system that shows an unstable profile. These results are in perfect agreement with biochemical experiments demonstrating that the LEDGF/p75 binds and stabilizes the IN tetramer.

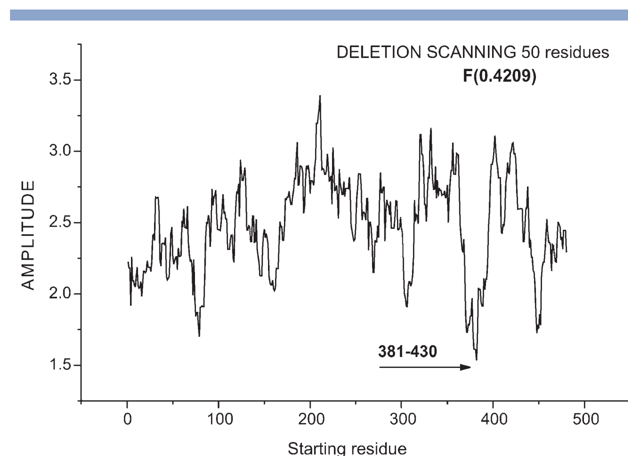


Figure 4

Identification of the region in LEDGF/p75 interacting with IN.

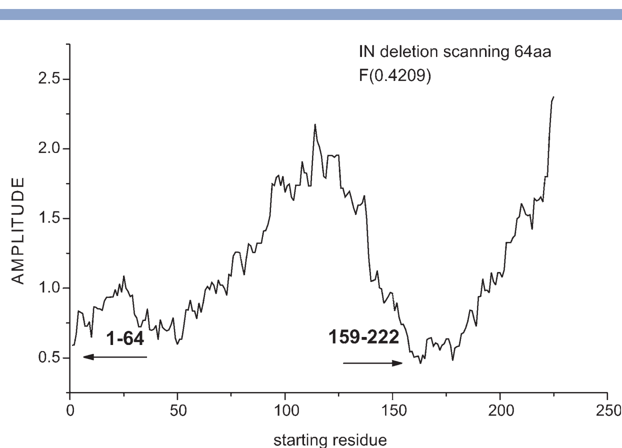


Figure 5

Identification of the region in IN interacting with LEDGF/p75.

Although the representations reported in Figure 6 were useful in analyzing the dynamics of the whole protein structures, we were also interested in determining which parts of the protein were most flexible. To this end, the RMSFs of the α carbons of each residues were calculated for both systems during the last 8 ns of simulation. Results are graphically shown in Figure 7.

The time necessary to reach a stable RMSD for both systems depends upon the dynamical behaviors of the C-terminal domains. However, it is interesting to note that the C-terminal domains of monomer B and monomer D, which directly interact with LEGDF, showed a lower flexibility in the tetramer IN-LEGDF/p75 IBD complex with respect to the free tetramer. This suggests that the two molecules of LEDGF/p75 could stabilize the tetramer by reducing its C-terminals fluctuations.

Analysis of the tetramer IN-LEGDF/IBD complex

Hot spot analysis

The MM-GBSA approach was used to identify the amino acids that make significant energetic contributions to the binding free energy in the IN tetramer-LEDGF/p75 IBD complex. A total of 100 snapshots were taken from the last 12 ns of the MD trajectories. In detail, a binding free energy decomposition at the atomic level was made to evaluate the contribution of each residue to the complex formation. Table I shows the residue "hot spots" that have a significant favorable contribution to the binding free energy ($|\Delta G_{\text{bind}}^{\text{res}}| > 1$ kcal/mol). It is of interest that we studied the binding of two LEDGF/p75 IBD by taking advantage of the fact that IN is a tetramer with two independent binding sites. Most of the residues that we have identified belong to the CCDs of the IN tetramer (Ala128, Gln168, Ala169 Glu170, His171, Thr174, Thr124, Thr125, Trp131, and Trp132). These residues match those already identified by Zhao *et al.*,[39] who performed molecular dynamic simulation of a complex between the CCDs of dimeric IN with one molecule of LEDGF/p75 IBD. However, several novel hot spots were found in both amino-terminal (Leu2, Asp3, Asp6, Gln9, Glu10, Glu11, and Glu13) and carboxy-terminal domains (Trp243, Glu246, Ala248, Val250, Ile257, Val259, Arg263, and Asp270). Concerning the hot spots identified within the LEDGF/p75 IBD subunits, the residues Lys364, Ile365, Asp366, Asn367, Leu368, Phe406, and Val408 have already been reported to be important for the complex formation. In addition to these residues, we have also identified the amino-acids Lys392, Thr398, Arg404, Arg405, Lys407, and Glu414 in LEDGF/p75 as important for its interaction with IN.

These results are in perfect agreement with all reported site-directed mutagenic studies, which indicate that residues Ile365, Asp366, and Phe406 in LEDGF/p75 and the amino acids Ala129, Trp132, Arg166, and Gln168 in IN are crucial for the interactions between the two proteins.[40–45]

Hydrogen bonds analysis

Hydrogen-bonding interactions between tetrameric IN and the two molecules LEDGF/p75-1 and LEDGF/p75-2 were determined and shown in Tables II and III, respectively. The hydrogen-bond interactions were reported when the occupancy was more than 20% in the investigated time period (the last 12 ns of simulation). Stable hydrogen bonds were found at the interface of the CCD of IN and LEDGF/p75 IBD for both LEDGF/p75-1 and LEDGF/p75-2. Particularly, amino acids Glu170, His171, and Thr174 in IN established hydrogen-bond contacts with Asp366 in LEDG, and residues Gln168 and Glu170 interacted with Ile365 and Asn367, respectively. Moreover, the indole NH of Trp131 made a hydrogen-bond interaction with the carbonyl backbone of Arg405.

In addition to these already reported crucial interactions, the tetramer IN-LEDGF/p75 IBD complex was stabilized by a network of hydrogen bonds involving

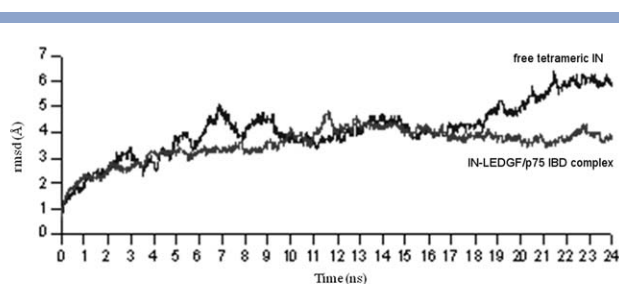


Figure 6

Time evolution of the root mean square deviations (RMSD calculated on Cα) of free tetrameric IN and tetrameric IN-LEDGF/p75 IBD complex from the corresponding initial structure.

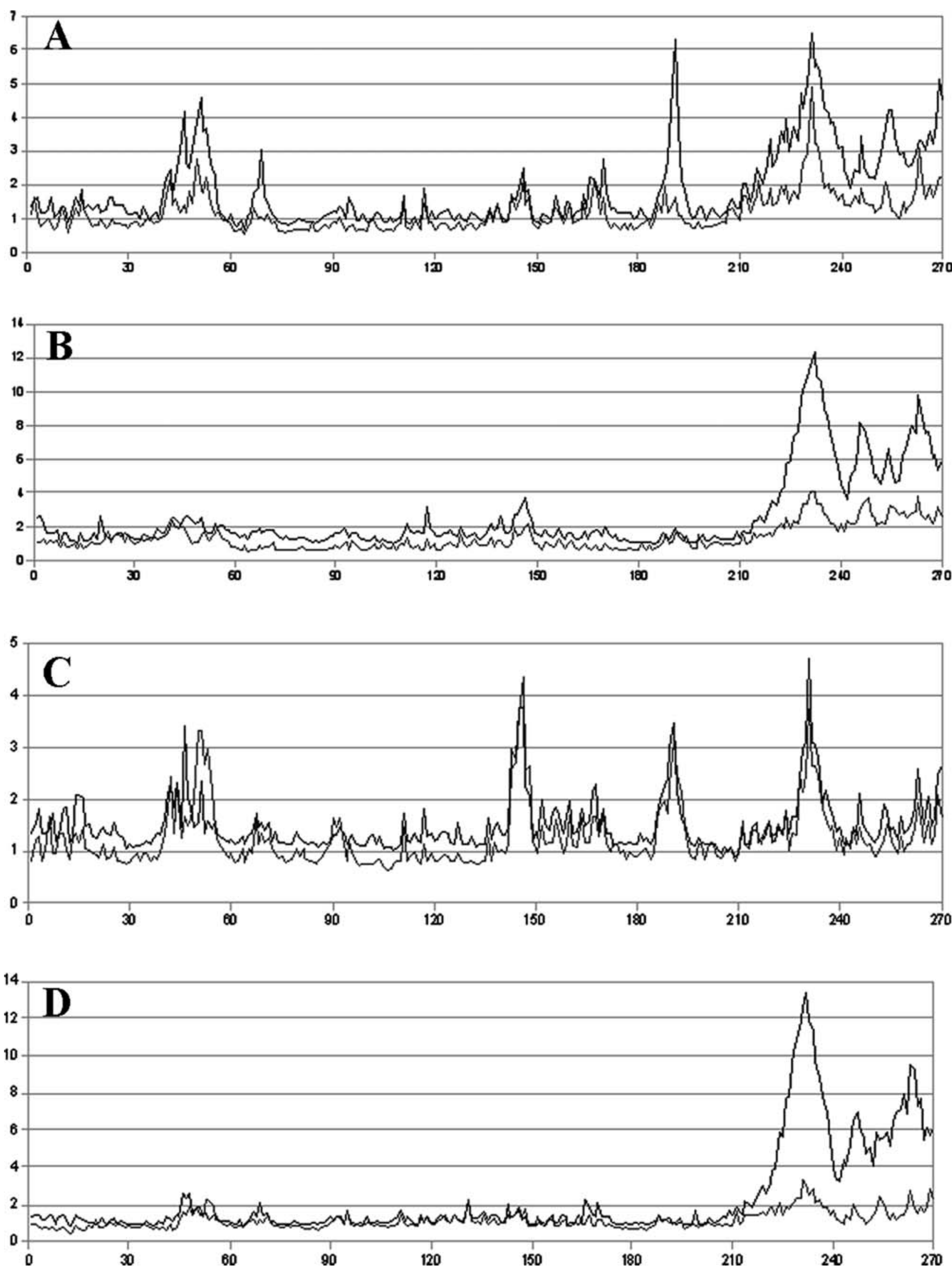


Figure 7

RMSF for each residues belong to chain A (A), chain B (B), chain C (C), and chain D (D)in IN. The number of residues is reported in abscissa while the RMSF expressed in Å is reported in ordinate.

several residues of both the amino-terminal (Gln9, Asp3, Leu2, Asp6, and Glu11) and carboxy-terminal (Glu246, Trp243, Lys392, Arg263, and Lys244) domains of IN.

In detail, LEDGF/p75-1 established hydrogen-bond contacts with Gln9, Asp3, Leu2, Asp6, Glu246, Trp243, Lys392, Arg263, and Lys244 by means of residues Thr398, Arg404, Gln391, Gln388, Arg405, Tyr420, Ile257, Glu414, Lys392, Lys401, and Lys407. On the other hand, residues Arg405, Lys407, and Lys401 of LEDGF/p75-2 were involved in hydrogen bond interactions with amino acids Glu11, Glu246, and Asp6 of IN. Some differences were observed in the pattern of the interactions of the two LEDGF/p75 IBD molecules. However, we identified many common residues in both LEDGF/p75-1 and LEDGF/p75-2 that are crucial for their interactions with IN that include Asp366, Ile365, Asn367, Arg404, Arg405, Lys401, and Lys407. These consensus interactions may lie at the basis of the recognition process between tetrameric IN and LEDGF/p75 IBD.

Table I
Hot-Spots Identified by Binding Free Energy Decomposition Per Residue

Residue IN-chain A	Residue IN-chain B	Residue IN-chain C	Residue IN-chain D	Residue LEDGF/p75-1	Residue LEDGF/p75-2
131 (−6.61)	170 (−4.66)	131 (−7.12)	170 (−5.11)	365 (−6.04)	365 (−5.97)
125 (−3.77)	169 (−4.07)	125 (−2.94)	3 (−4.25)	366 (−5.29)	405 (−5.23)
128 (−2.49)	6 (−1.99)	246 (−2.77)	6 (−3.82)	398 (−5.09)	366 (−4.61)
95 (−2.12)	257 (−1.58)	128 (−2.32)	243 (−3.13)	392 (−3.39)	364 (−3.80)
124 (−1.87)	171 (−1.53)	263 (−1.70)	13 (−2.91)	367 (−3.14)	406 (−2.87)
132 (−1.31)	174 (−1.32)	2 (−1.61)	257 (−2.73)	408 (−3.02)	408 (−2.39)
270 (−1.18)	11 (−1.17)	3 (−1.56)	9 (−2.33)	406 (−3.0)	368 (−2.11)
124 (−1.17)	259 (−1.0)	132 (−1.51)	259 (−2.27)	364 (−2.43)	367 (−1.99)
	15 (−1.0)	124 (−1.32)	171 (−1.97)	368 (−2.17)	414 (−1.41)
		95 (−1.25)	168 (−1.66)	414 (−1.84)	378 (−1.35)
		270 (−1.18)	10 (−1.45)	404 (−1.57)	369 (−1.31)
			256 (−1.32)	378 (−1.53)	358 (−1.23)
			174 (−1.31)	369 (−1.44)	379 (−1.23)
			250 (−1.25)	349 (−1.43)	375 (−1.23)
			248 (−1.09)	379 (−1.36)	349 (−1.12)
				375 (−1.35)	426 (−1.02)
				358 (−1.30)	

Table II
Hydrogen Bonds Between LEDGF/p75-1 and Tetramer IN with Occupancy >20%

IN		LEDGF/p75-1 IBD			
Residue	Group	Residue	Group	Distance (Å)	Occupancy (%)
E170 (chain B)	H	D366	OD1	2.79	99.97
T174 (chain B)	HG1	D366	OD2	2.78	96.53
Q168 (chain B)	O	I365	H	3.067	95.17
Q9 (chain D)	HE22	T398	OG1	2.99	95.07
D3 (chain D)	OD2	R404	HH12	2.95	90.34
D3 (chain D)	OD1	R404	HH22	3.02	82.31
D3 (chain D)	OD1	R404	HH12	3.04	79.84
E246 (chain C)	OE2	R405	HH12	2.88	72.54
Q95 (chain A)	HE22	D366	O	2.96	70.14
D3 (chain D)	OD2	R404	HH22	3.1	67.88
H171 (chain B)	H	D366	OD2	3.23	66.91
L2 (chain C)	O	Q391	HE21	3.05	63.78
W243 (chain D)	HE1	Q388	OE1	2.88	63.58
E246 (chain C)	OE2	R405	HH22	2.97	55.25
E170 (chain B)	OE2	N367	HD22	2.94	50.02
D3 (chain D)	OD2	Y420	HH	2.91	38.95
W131 (chain A)	HE1	R405	O	3.14	37.89
E170 (chain B)	OE1	N367	HD22	2.92	36.69
I257 (chain A)	O	K392	HZ2	2.84	35.89
R263 (chain A)	HH22	E414	OE2	3	35.45
K244 (chain A)	O	Q388	HE21	3.12	34.76
E170 (chain A)	H	D366	OD2	3.319	33.76
R263 (chain A)	HH12	E414	OE1	2.93	33.49
R263 (chain A)	HH22	E414	OE1	2.93	33.49
D6 (chain A)	OD1	K401	HZ1	2.81	32.12
R263 (chain A)	HH12	E414	OE2	3	31.99
D6 (chain A)	OD1	K401	HZ2	2.81	30.02
E246 (chain C)	OE2	K407	HZ2	2.91	29.99
E246 (chain C)	OE2	K407	HZ1	2.89	29.66
D6 (chian D)	OD1	K401	HZ3	2.81	28.19
Q9 (chain D)	OE1	K401	HZ3	2.89	26.92
Q9 (chain D)	OE1	K401	HZ1	2.89	25.26
Q9 (chain D)	OE1	K401	HZ2	2.90	25.19
E246 (chain C)	OE1	K407	HZ2	2.99	25.09
E246 (chain C)	OE1	R405	HH12	2.90	20.56

Electrostatic interactions

The interactions between the amino-terminal and the carboxy-terminal residues in IN and LEDGF/p75 IBD were almost exclusively electrostatic in nature. In particular, two well-defined salt bridges existed between Lys401 and Lys407 in both LEDGF/p75 molecules with Asp6 and Glu246 in IN, respectively. These electrostatic interactions were investigated during the MD simulations by monitoring the distance between the NE atoms of the two lysines and the CD atoms of residues Asp6 and Glu246 for both LEDGF/p75-1 and LEDGF/p75-2. The two salt bridges were stable during the last 12 ns of simulation for both LEDGF/p75 IBD molecules.

Table III
Hydrogen Bonds Between LEDGF/p75-2 and Tetramer IN with Occupancy >20%

IN		LEDGF/p75-2 IBD			
Residue	Group	Residue	Group	Distance (Å)	Occupancy (%)
E170 (chain D)	H	D366	OD1	2.836	98.27
Q168 (chain D)	O	I365	H	2.966	97.77
E11 (chain B)	OE2	R405	HH21	2.86	92.84
H171 (chain D)	H	D366	OD2	3.092	86.57
E11 (chain B)	OE1	R405	HE	2.88	88.04
E170 (chain D)	H	D366	OD2	3.141	78.41
E11 (chain B)	OE1	R405	HH21	3.1	77.14
T174 (chainD)	HG1	D366	OD2	2.74	74.08
E11 (chain B)	OE2	R405	HE	3.1	29.32
E246 (chain A)	OE2	K407	HZ1	2.92	25.29
E246 (chain A)	OE1	K407	HZ1	2.92	24.89
E170 (chain C)	OE1	N367	HD22	2.97	23.29
E170 (chain C)	OE2	N367	HD22	2.97	20.06
W131 (chain C)	HE1	R405	O	3.053	71.64
Q95 (chain C)	HE22	D366	O	2.922	57.48
D6 (chain B)	O	K401	HZ2	2.967	22.19
D6 (chain B)	O	K401	HZ3	2.958	21.66

Table IV
Residues Involved in Hydrophobic Interactions

LEDGF/p75-1	LEDGF/p75-2	IN
Ile365	Ile365	Ala98, Leu102, Ala128, Ala129, Thr125, Trp132, Ala169, Thr174, Met178
Ile403	Ile403	Trp131
Phe406	Phe406	Trp131, Ala128
Val408	Val408	Thr124, Ala128, Lys127
Val370	Val370	Thr124
Leu368	Leu368	Thr124, Ala128, Thr125
Lys364	Lys364	Ala169

Hydrophobic interactions

Several hydrophobic residues were located at the tetramer IN/LEDGF/p75 IBD binding interface, as reported in Table IV. Ile365 made several van der Waals contacts with Ala98, Leu102, Ala128, Ala129, Thr125, Trp132, Ala169, Thr174, and Met178. Phe406 mainly interacted with the indole ring of Trp131 by a π-π interaction. Furthermore, Leu368 and Val408 interacted with Thr124, Thr125, Ala128, and Lys127 through positive van der Waals contacts, and Ile403 made favorable hydrophobic interactions with Trp131. Positive van der Waals contacts were also found between the aliphatic portion of the Lys364 side chain and Ala169. Only weak hydrophobic interactions between the amino-terminal or CTDs of IN and LEDGF/p75 were observed.

CONCLUSIONS

From the first article published by Cherepanov *et al.* in 2003,[46] many research groups have published interesting studies on the role of LEDGF/p75 in HIV-1 integration in the last years.

Moreover, several studies have been conducted to establish which residues of IN and LEDGF/p75 are responsible for the interactions between the two proteins. A major advance was achieved when the co-crystal structure of the LEDGF/p75 and IN core domain was determined. However, it has been shown that not only is the central core domain of IN involved in the interaction with LEDGF/p75 but also the N-terminal zinc binding domain plays an important role.[16] Furthermore, biochemical characterizations of the interactions between full-length IN and LEDGF/p75 have recently demonstrated that LEDGF/p75 promotes IN tetramerization with two LEDGF/p75 IBD molecules bound to the IN tetramer.[17] On the other hand, it has been shown previously that concordance in EIIP patterns is an important condition for targeting between interacting proteins and that the sequence domains that contribute the most to this concordance are involved in important protein-protein interactions.[20–25] The ISM is a bioinformatics tool that we used to analyze the EIIP sequence similarity patterns. Specifically, ISM analyses were performed to predict domains that were potentially involved in the interaction of HIV-1 IN with cellular LEDGF/p75. The HIV-1 IN domains (core and N-terminal domains) involved in this interaction that were found in experimental studies are the same domains that were found from our bioinformatics analyses. In LEDGF/p75, the important domain encompasses residues 381 to 430, the sequence that contains residues also already identified as important for binding to IN by means of crystallographic studies. This result confirms that the EIIP pattern is an accurate indicator of HIV-1 IN-LEDGF/p75 recognition and interaction being in agreement in different aspects with crystallographic data,[12] mutagenesis studies,[36] and domain deletion studies.[6] In a recent work, we have successfully applied EIIP as a descriptor for virtual filtering of small molecules that act as IN inhibitors.[47]

After considering the results from this ISM analysis as well as the experimental studies described above, we decided to build a model of interactions between the tetrameric form of IN and two molecules of LEDGF/p75. Such a model was obtained by assembling all available crystallographic structures. Moreover, a detailed dynamics picture of the interactions between the two proteins was derived by molecular dynamic simulations. Several hot spots were identified by the decomposition of the binding free energy on a per-residue basis. Most of these hot spots belong to the CCDs of IN, in agreement with previously results reported by Zhao *et al.*[39] These hot spot residues were composed of Gln168, Glu170, Thr174, Thr125, Trp131, and His171 in IN, which interact with amino acids Ile365, Asp366, Phe406, and Val408 in LEDGF/p75. In addition to these already reported crucial interactions, the tetramer IN-LEDGF/p75 IBD complex was stabilized by a network of interactions involving several residues of both the amino-terminal (Gln9, Asp3, Leu2, Asp6, and Glu11) and the carboxy-terminal (Glu246, Trp243, Lys392, Arg263, and Lys244) domains of IN. Although the CCD of IN is characterized by strong van der Waals contacts and stable hydrogen bonds, LEDGF/p75 binds the amino-terminal and the carboxy-terminal domains of IN mainly with electrostatic interactions. On the whole, our results suggest that the CCD of IN is essential for the interaction with LEDGF/p75. We also show that some residues at the amino-terminal and carboxy-terminal domains could interact with LEDGF/p75 contributing to the complex formation. These computational predictions could be a helpful starting point for future site-directed mutagenesis studies. Particularly because our results are in agreement with the available experimental studies. In fact, the involvement of the N-terminal domain of IN in the interaction with LEDGF/p75 has been proved by Maertens *et al.*[16] Concerning the involvement of the C-terminal domain of IN in the interaction with LEDGF/p75, a study has been recently published in which the authors observed that the presence of LEDGF/p75 IBD in infected cells prevents

access to the IN C-terminus because it is masked in the presence of LEDGF/p75 IBD.[48]

Furthermore, to study the role of LEDGF/p75 in the stabilization of the IN, MD simulations were also performed on the free tetrameric form of integrase (unbounded with LEDGF/p75). By comparing the two trajectory profiles, it was possible to observe that while the IN-LEDGF/p75 complex was stable during the time of simulation, a different behavior was found for the free IN tetramer. In detail, the C-terminal domains of monomer B and monomer D, which directly interact with LEGDF, had a lower flexibility in the tetramer IN-LEGDF/IBD complex with respect to the free tetramer. Thus suggests that the two molecules of LEDGF/p75 could stabilize the tetramer by reducing its C-terminal fluctuations.

It is important to emphasize that the residues recognized by MD analyses in both proteins are immediately adjacent or involved in domains identified by ISM analyses. These findings are in accordance with the concept of complementary analyses of protein-protein interactions involving LRIs and SRIs that was the basis of this work. Namely, the goal of ISM is to identify parts of the sequence that contribute the most to the recognition and targeting of the proteins at distances longer than one linear dimension of the biomolecules, LRI.[19] These targeting sequences are already shown to be close in structure or partially overlapping with the domains involved in SRIs.[49,50] Complementary to ISM, MD studies aimed at studying in detail the SRI between the interacting proteins (such as hydrogen bond contacts, electrostatic interactions, van der Waals interactions) were performed. We assumed that these two kinds of analyses would enable us to make a broader image of the LEDGF/p75-HIV-1 IN interaction that could be possible by use of either one alone. As a result, this study provides a detailed dynamic picture of IN-LEDGF/p75 that is in many elements in agreement with available experimental data and, thereby, provides a new understanding of IN-LEDGF/p75 interactions. Moreover, the model is able to explain the role of LEDGF/p75 in the stabilization of the tetramer IN and is, therefore, of general interest for the development of new anti-HIV-1 inhibitors.

Our analysis agrees with Benkhelifa-Ziyyat *et al.*[48] who suggested that two types of IN-LEDGF/p75 IBD complexes may form: one in which the C-terminus is accessible and another one in which the C-terminus is masked. These two states were proposed to depend on the experimental system used during their study. In this regard, our model could represents the mode of binding of LEDGF/p75 to IN in cytoplasm and, it is complementary to the recently reported electron microscopy study by Michel *et al.*[51] These authors have obtained a three-dimensional model of HIV-1 integrase with LEDGF/p75 and DNA by fitting the known atomic structures onto negatively stained electron microscopy (EM) maps, yielding a clear determination of the interactions between IN, DNA and LEDGF/p75 in the nucleus. Remarkably, two interesting articles have been recently published by Hare *et al.*[52,53] showing that the role of HIV-1 IN N-terminus in interaction with LEDGF/p75 remains still uncertain. These authors determined the crystallographic structures of the maedi visna virus (MVV) IN tetramer in open and closed conformations.[52] These structures disagree in the locations of the amino-terminal domains with the previously determined HIV-1 IN tetramer,[29] maybe due to the fact that the overall sequence identity between MVV IN and HIV-1 IN is less than 30%, but the reasons for these differences could be multiple. Similarly, the crystal structure of HIV-2 IN (CCD and NTD) in complex with LEDGF/p75 reported by the same authors[53] shows differences in the relative position of NTD versus CCD with respect to the HIV-1 IN crystal structures reported so far. The study by Hare *et al.*[53] explains the role of the HIV-2 NTD in the interaction with LEDGF/p75 but the role of HIV-1 NTD has to be further investigated as well as the role of CTD in the interaction with LEDGF/p75 which has been just studied. Many experimental investigations are still needed to clarify these unresolved aspects.

REFERENCES

1. Esposito D, Craigie R. HIV integrase structure and function. Adv Virus Res 1999;52:319–333.
2. Assante-Appiah E, Skalka AM. HIV-1 integrase: structural organization, conformational changes, and catalysis. Adv Virus Res 1999;52:351–369.
3. Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, Parissi V. HIV-1 integrase crosslinked oligomers are active in vitro. Nucleic Acids Res 2005;33:977–986.
4. Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet JF, Brochon JC, Deprez E. Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. J Biol Chem 2006;281:22707–22719.
5. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, Arosio D, Emiliani S, Rain JC, Benarous R, Cereseto A, Debyser Z. Transportin-SR2 imports HIV into the nucleus. Curr Biol 2008;16:1192–1202.
6. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, Poeschla E. Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. J Cell Sci 2005;118:1733–1743.
7. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F, Debyser Z. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J Biol Chem 2005;280:17841–17847.
8. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, Vercammen J, Engelborghs Y, Christ F, Debyser Z. Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. J Virol 2006;80:11498–11509.
9. Hayouka Z, Rosenbluh J, Levin A, Loya S, Lebendiker M, Veprintsev D, Kotler M, Hizi A, Loyter A, Friedler A. Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. Proc Natl Acad Sci USA 2007;104:8316–8321.
10. Hayouka Z, Rosenbluh J, Levin A, Maes M, Loyter A, Friedler A. Peptides derived from HIV-1 Rev inhibit HIV-1 integrase in a shiftide mechanism. Biopolymers 2008;90:481–487.

11. Armon-Omer A, Levin A, Hayouka Z, Butz K, Hoppe-Seyler F, Loya S, Hizi A, Friedler A, Loyter A. Correlation between shiftide activity and HIV-1 integrase inhibition by a peptide selected from a combinatorial library. J Mol Biol 2008;376:971–982.
12. Cherepanov P, Ambrosio ALB, Rahman S, Ellenberger T, Engelman A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci USA 2005;102:17308–17313.
13. Du L, Zhao Y, Chen J, Yang L, Zheng Y, Tang Y, Shen X, Jiang H. D77, one benzoic acid derivative, functions as a novel anti-HIV-1 inhibitor targeting the interaction between integrase and cellular LEDGF/p75. Biochem Biophys Res Commun 2008;375:139–144.
14. De Luca L, Barreca ML, Ferro S, Christ F, Iraci N, Gitto R, Monforte AM, Debyser Z, Chimirri A. Pharmacophore-based discovery of small-molecule inhibitors of protein-protein interactions between HIV-1 integrase and cellular cofactor LEDGF/p75. ChemMedChem 2009;4:1311–1316.
15. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV, De Maeyer M, Chaltin P, Debyser Z. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol 2010;6:442–448.
16. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, Engelborghs Y. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem 2003;278:33528–33539.
17. McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, Kvaratskhelia M. Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. J Biol Chem 2008;283:31802–31812.
18. Veljkovic V, Cosic I, Dimitrijevic B, Lalovic D. Is it possible to analyze DNA and protein sequences by the methods of digital signal processing? IEEE Trans Biomed Eng 1985;BME-32:337-341.
19. Veljkovic V, Veljkovic N, Esté JA, Dietrich U. Application of the EIIP/ISM bioinformatics concept in development of new drugs. Curr Med Chem 2007;14:441–453.
20. Veljkovic N, Glisic S, Prljic J, Perovic V, Botta M, Veljkovic V. Discovery of new therapeutic targets by the informational spectrum method. Curr Protein Pept Sci 2008;9:493–506.
21. Mancini M, Veljkovic N, Corradi V, Zuffa E, Corrado P, Pagnotta E, Martinelli G, Barbieri E, Santucci MA. 14-3-3 ligand prevents nuclear import of c-ABL protein in chronic myeloid leukemia. Traffic 2009;10:637–647.
22. Glisic S, Arrigo P, Alavantic D, Perovic V, Prljic J, Veljkovic N. Lipoprotein lipase: a bioinformatics criterion for assessment of mutations as a risk factor for cardiovascular disease. Proteins 2008;70:855–862.
23. Chen K, Huzil JT, Freedman H, Ramachandran P, Antoniou A, Tuszynski JA, Kurgan L. Identification of tubulin drug binding sites and prediction of relative differences in binding affinities to tubulin isotypes using digital signal processing. J Mol Graph Model 2008;27:497–505.
24. Illingworth CJ, Parkes KE, Snell CR, Mullineaux PM, Reynolds CA. Criteria for confirming sequence periodicity identified by Fourier transform analysis: application to GCR2, a candidate plant GPCR? Biophys Chem 2008;133:28–35.
25. Pirogova E, Simon GP, Cosic I. Investigation of the applicability of dielectric relaxation properties of amino acid solutions within the resonant recognition model. IEEE Trans Nanobioscience 2003;2:63–69.
26. Karplus M, Petsko GA. Molecular dynamics simulations in biology. Nature 1990;347:631–639.
27. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham T. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
28. Wang LD, Liu CL, Chen WZ, Wang CX. Constructing HIV-1 integrase tetramer and exploring influences of metal ions on forming integrase-DNA complex. Biochem Biophys Res Commun 2005;337:313–319.
29. Wang JY, Ling H, Yang W, Craigie R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. EMBO J 2001;20:7333–7343.
30. Maignan S, Guilloteau JP, Zhou-Liu Q, Clément-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. J Mol Biol 1998;282:359–368.
31. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc Natl Acad Sci USA 2000;97:8233–8238.
32. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol 1997;4:567–577.
33. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA. Amber 10, San Francisco, CA: University of California; 2008.
34. Stote RH, Karplus M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. Proteins 1995;23:12–31.
35. Tuccinardi T, Calderone V, Rapposelli S, Martinelli A. Proposal of a new binding orientation for non-peptide AT1 antagonists: homology modeling, docking and three-dimensional quantitative structure-activity relationship analysis. J Med Chem 2006;49:4305–4316.
36. Tuccinardi T, Manetti F, Schenone S, Martinelli A, Botta M. Construction and validation of a RET TK catalytic domain by homology modeling. J Chem Inf Model 2007;47:644–655.
37. Baris I, Tuncel A, Ozber N, Keskin O, Kavakli IH. Investigation of the interaction between the large and small subunits of potato ADP-glucose pyrophosphorylase. PLoS Comput Biol 2009;5:e1000546.
38. Liu H, Yao X. Molecular basis of the interaction for an essential subunit PA-PB1 in influenza virus RNA polymerase: insights from molecular dynamics simulation and free energy calculation. Mol Pharm 2010;7:75–85.
39. Zhao Y, Li W, Zeng J, Liu G, Tang Y. Insights into the interactions between HIV-1 integrase and human LEDGF/p75 by molecular dynamics simulation and free energy calculation. Proteins 2008;72:635–645.
40. Al-Mawsawi LQ, Neamati N. Blocking interactions between HIV-1 integrase and cellular cofactors: an emerging anti-retroviral strategy. Trends Pharmacol Sci 2007;28:526–535.
41. Cherepanov P. Sun ZY, Rahman S, Maertens G, Wagner G, Engelman A. Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. Nat Struct Mol Biol 2005;12:526–532.
42. Rahman S, Lu R, Vandegraaff N, Cherepanov P, Engelman A. Structure-based mutagenesis of the integrase-LEDGF/p75 interface uncouples a strict correlation between in vitro protein binding and HIV-1 fitness. Virology 2007;357:79–90.
43. Busschots ,K. Voet A, De Maeyer M, Rain JC, Emiliani S, Benarous R, Desender L, Debyser Z, Christ F. Identification of the LEDGF/p75 binding site in HIV-1 integrase. J Mol Biol 2007;365:1480–1492.
44. Turlure F, Devroe E, Silver PA, Engelman A. Human cell proteins and human immunodeficiency virus DNA integration. Front Biosci 2004;9:3187–3208.
45. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempé D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, Christ F, Rain JC, Dargemont C, Debyser Z, Benarous R. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J Biol Chem 2005;280:25517–25523.

46. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E, Debyser Z. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem 2003;278:372–381.
47. Tintori C, Manetti F, Veljkovic N, Perovic V, Vercammen J, Hayes S, Massa S, Witvrouw M, Debyser Z, Veljkovic V, Botta M. Novel virtual screening protocol based on the combined use of molecular modeling and electron-ion interaction potential techniques to design HIV-1 integrase inhibitors. J Chem Inf Model 2007;47:1536–1544.
48. Benkhelifa-Ziyyat S, Bucher S, Zanta-Boussif MA, Pasquet J, Danos O. Changes in the accessibility of the HIV-1 Integrase C-terminus in the presence of cellular proteins. Retrovirology 2010;7:27–36.
49. Cosic I, Hodder AN, Aguilar MI, Hearn MT. Resonant recognition model and protein topography. Model studies with myoglobin, hemoglobin and lysozyme. Eur J Biochem 1991;198:113–119.
50. Veljkovic V, Veljkovic N, Muller CP, Müller S, Glisic S, Perovic V, Köhler H. Characterization of conserved properties of hemagglutinin of H5N1 and human influenza viruses: possible consequences for therapy and infection control. BMC Struct Biol 2009;9:21.
51. Michel F, Crucifix C, Granger F, Eiler S, Mouscadet JF, Korolev S, Agapkina J, Ziganshin R, Gottikh M, Nazabal A, Emiliani S, Benarous R, Moras D, Schultz P, Ruff M. Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. EMBO J 2009;28:980–991.
52. Hare S, Di Nunzio F, Labeja A, Wang J, Engelman A, Cherepanov P. Structural basis for functional tetramerization of lentiviral integrase. PLoS Pathog 2009;5:e1000515.
53. Hare S, Shun MC, Gupta SS, Valkov E, Engelman A, Cherepanov P. A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. PloS Pathog 2009;5:e1000259.