An Unusual Helix Turn Helix Motif in the Catalytic Core of HIV-1 Integrase Binds Viral DNA and LEDGF

Hayate Merad, Horea Porumb, Loussinée Zargarian, Brigitte René, Zeina Hobaika, Richard G Maroun, Olivier Mauffret, Serge Fermandjian

To cite this version:


HAL Id: hal-00349896
https://hal.archives-ouvertes.fr/hal-00349896
Submitted on 5 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.
An Unusual Helix Turn Helix Motif in the Catalytic Core of HIV-1 Integrase Binds Viral DNA and LEDGF

Hayate Merad1, Horea Porumb1, Loussiné Zargarian1, Brigitte René1, Zeina Hobaika1, Richard G. Maroun2, Olivier Mauffret1, Serge Fermandjian1*

1 LPBA, CNRS (UMR 8113)–École Normale Supérieure de Cachan, Cachan, France, 2 Département des Sciences de la Vie et de la Terre, Faculté des Sciences, Université Saint Joseph, CST-Mar Roukos, B. P. 1514, Beyrouth, Liban

Abstract

**Background:** Integrase (IN) of the type 1 human immunodeficiency virus (HIV-1) catalyzes the integration of viral DNA into host cellular DNA. We identified a bi-helix motif (residues 149–186) in the crystal structure of the catalytic core (CC) of the IN-Phe185Lys variant that consists of the α4 and α5 helices connected by a 3 to 5-residue turn. The motif is embedded in a large array of interactions that stabilize the monomer and the dimer.

**Principal Findings:** We describe the conformational and binding properties of the corresponding synthetic peptide. This displays features of the protein motif structure thanks to the mutual intramolecular interactions of the α4 and α5 helices that maintain the fold. The main properties are the binding to: 1- the processing-attachment site at the LTR (long terminal repeat) ends of virus DNA with a Kd (dissociation constant) in the sub-micromolar range; 2- the whole IN enzyme; and 3- the IN binding domain (IBD) but not the IBD-Asp366Asn variant of LEDGF (lens epidermal derived growth factor) lacking the essential Asp366 residue. In our motif, in contrast to the conventional HTH (helix-turn-helix), it is the N terminal helix (α4) which has the role of DNA recognition helix, while the C terminal helix (α5) would rather contribute to the motif stabilization by interactions with the α4 helix.

**Conclusion:** The motif, termed HTHi (i, for inverted) emerges as a central piece of the IN structure and function. It could therefore represent an attractive target in the search for inhibitors working at the DNA-IN, IN-IN and IN-LEDGF interfaces.

Introduction

The integration of the HIV-1 genome into the host cell chromosome is mediated by the viral integrase (IN) [1–6]. The enzyme catalyzes a multi-step reaction i.e., 3' end processing and strand transfer, to integrate a linear DNA copy (cDNA) of the retroviral genome into the host cell DNA [2,7,8]. The retroviral DNA integration mimics that of insertion elements and bacteriophage Mu transposons [9–11] and bears resemblance to the RAG1/2 recombinaise [12].

The HIV-1 IN is essential for the viral life cycle and is therefore an attractive target for developing anti-HIV drugs [13,14]. The enzyme (288 amino acid residues, 32 kDa) has three well defined structural domains: an N terminal domain (residues 1 to 49), a central catalytic domain or catalytic core, CC (residues 50 to 212), and a C terminal domain (residues 213 to 288) [15–17]. Several crystal structures of the CC domain and of two-domain fragments (CC domain linked either to the C terminal domain or the N terminal domain) have been already resolved by X-ray crystallography [18–25] while the N terminal and C terminal domains have been analyzed in solution by NMR [26,27]. Each domain, taken separately, forms a dimer and this is true also true for the N-terminal-CC and the C-terminal CC bi-domains [18–29]. The CC dimer (Fig. 1a) is organized around a two fold axis with a large interface involving, in particular, helices α1 and α2 (residues 172–184) [18,30]. Other retroviral IN CC structures display the same dimer boundary, indicating that this type of interface is biologically relevant.

Actually, cross-linked dimers have been shown to be active for strand processing and single end integration [31]. Yet, a large number of data suggest that the tetramer is the form stabilizing the synaptic complexes of IN with the two viral DNA ends and appears to be the form required for the strand transfer [32–37]. Several theoretical models of the DNA-IN complexes have proven the relevance of tetramers to position the viral and cellular DNA partners at reactive distance [38–41].

The CC domain is organized in five β-strands surrounded by six α helices (α1 to α6), and contains a highly conserved catalytic D, DX4E motif embedded in a protein RNase H fold [17,20,21]. The amphipathic α4 helix, (residues 148–167), which protrudes at the protein surface, bears the catalytic residue Gln-152 and several other residues, such as Glu-148, Lys-156 and Lys-159, which have been shown to be important for the binding of IN to DNA and for virus survival. In the crystal structure of CC bound to the inhibitor
Figure 1. Identification of an “inverted” HTH motif (HTHi) at the catalytic core surface of integrase (PDB ID 1BIU [20]).

(a). Crystal structure of the catalytic core domain, associated into a dimer. (b). Representation of the HTHi motif, with the loop residues shown by van der Waals spheres. (c). The side chain residues involved in intramolecular contacts, shown by sticks and van der Waals spheres. (d). The electrostatic potential at the solvent-accessible surface; the Lys-156, Lys-159 and Lys-160 residues are shown by sticks. (e). HTHi motif of IN, superimposed onto the “classical” HTH motif of the HMG (highly mobile group) protein LEF-1 (lymphoid enhancer binding factor, PDB ID 2LEF, brown). (f). HTHi motif of IN, superimposed onto the HTHi motif of the Signal Recognition Particle (PDB ID 2FFH, green).

doi:10.1371/journal.pone.0004081.g001

5CITEP [1-(5-chloroindol-3-yl)-3-(tetrazoyl)-1, 3-propanedione enol] among the six protein-drug interactions, five involve amino acid side chains of the α5 helix [42], confirming the relevance of the α5 helix to IN function [41,43–47].

The propensity of IN to associate and form aggregates is the main barrier to the study of its structure and interactions by physical-chemical means including the widely used x-rays and NMR techniques. To date the only detailed depiction of the IN-DNA complex is provided by molecular modeling [38–41]. To overcome the difficulties inherent to the use of the entire enzyme we decided to dissect its properties and to analyze it part by part. Our laboratory has previously shown, in line with a systematic search by Li et al. [48], that peptides deriving from the α5 helix had inhibitory activity against IN [49,50]. This was also the case for peptides reproducing the sequences of the α1 and α3 helices involved in the CC dimer interface [α1:α5; and α1:α3] [30]. We have also shown that the peptide region 163–171, encompassing both the turn and the N and C terminal parts of the α4 and α5 helices, respectively, was a strong epitope and antibodies refined against this region inhibited both the 3’ processing and strand transfer reactions in in vitro assays [49]. Interestingly, the epitope region, in particular the residue Gln168 important for turn integrity, IN dimer formation and virus replication, has been shown to share hydrogen bonds with the IBD (IN binding domain) of LEDGF/p75 (lens epidermal derived growth factor), a transcriptional coactivator that is also an essential HIV integration cofactor in vivo [49,51–56]. Moreover, it has been assumed that the peptide region 161–173 superposing our epitope was involved in the nuclear import of IN [57], since a peptide reproducing this sequence caused the active nuclear import of BSA tethered to it [57,58]; the matter was somehow reassessed in so far as the fact that Val-165 and Arg-166 did not play the anticipated specific roles in the nuclear localization of HIV-1 pre-integration complex [59,60].

Actually, the α4 and α5 helices form a stable bi-helix fold at the protein surface of the CC crystal structure (Fig.1a–f). This recalls the well-known HTH (helix-turn-helix) motif of proteins specialized in DNA recognition [61]. Such motifs are generally associated with biological functions, protein structures and even evolutionary history. To assess the possible role of this bi-helix in IN we carried out an analysis of the structural and binding properties of the corresponding synthetic peptide (residues Gly-149 to Lys-186) using circular dichroism (CD) [43] and fluorescence. The peptide dealt with in this work represents the sequence found in the IN-CC Phe185Lys variant (Table 1a). It has been logically preferred to the native wt sequence (Table 1a) because it is the Phe185Lys variant that was used in the various crystallizations [20] and co-crystallizations so far reported [42,52]. The Phe185Lys mutation is not expected to affect the interactions with LTR and LEDGF IBD [51,52]. We have nevertheless performed a number of experiments (not shown here) to verify that the structure of the wt bi-helix was similar to that of the mutant.

Generally, the HTH motifs are strongly influenced by their context within the proteins and often lose their original biological functions when they are isolated from their protein environment. Here, we show that the isolated HTH motif continues to fulfill three of the major functions of the whole parental enzyme: i) it recognizes the U5LTR extremity of viral DNA at the attachment-processing site in a specific manner i.e., it preserves its viral DNA binding property; ii) it self-associates and associates to IN i.e., it reproduces some of the enzyme oligomerisation behavior, and iii) it binds to the IBD of LEDGF but not to its Asp366Asn-IBD variant lacking the essential Asp366 residue i.e., it maintains its ability to interact with specific partners.

Results

The HTHi conformation

The bi-helix motif we focus on (α5 helix-turn-α3 helix) occurs in the peptide segment from Gly-149 to Lys-186 of the IN-Phe185Lys variant. Like many classical HTH motifs, this bi-helix is exposed at the protein surface where it is largely accessible to solvent, proteins, DNA and organic ligands (Fig. 1a) [20]. The two helices are linked by a flexible turn of 3 to 5 residues bringing the amphipathic α4 helix (situated in N terminal position) onto the C terminal α5 helix through an angle of about 120° (Fig. 1b). The helices interact with each other through 16 mainly hydrophobic side chain-side chain contacts (Fig. 1c and Table 2). The juxtaposition of the helices generates a positive electrostatic potential exposed at the solvent accessible surface (Fig. 1d). The only difference between the present motif and the standard HTH motif is the inversion of position of the DNA recognition helix (α5 helix) and of the so-called stabilizing helix (α3) (Fig. 1c). Actually, the bi-helix motif of IN resembles the RNA-recognizing motif of
the SRP (Signal Recognition Particle, Fig. 1f), where the role of the helices are inverted, the N terminal one recognizing the DNA and the C-terminal one stabilizing the structure [62]. The bi-helix motif was therefore hereafter called HTHi (helix-turn-helix inverted).

The comparison of the GOR IV [63] and AGADIR [64] predictions strongly suggest that long distance interactions–like those stabilizing the protein tertiary structures–are needed besides the short distance interactions in order to keep the helix secondary structure in the HTHi motif (Fig. 2). Examination of the reported crystal structures shows that long range interactions occur in great number inside the HTHi motif itself i.e., between the a4 and the a5 helices (there are 16 such interactions, listed in Table 2), but also between the HTHi motif and other components of the protein (for example, there are 16 interactions between HTHi and the neighboring b-sheet–Table 2). Accordingly, a loss of helicity could be feared upon isolation of the motif from the protein context.

CD analysis indicates that the secondary structure of the HTHi peptide in solution is strongly concentration dependent. At 6 mM, with its two negative bands at 208 nm and 222 nm, the CD spectrum of HTHi reflects a helical content of 10–15% (full line curve in Fig. 3). Upon increasing the peptide concentration to 25 mM, the CD spectrum manifests an important change. This is illustrated by the reduction of the broad negative band at 222 nm, and the

<table>
<thead>
<tr>
<th>Table 1. Sequences of Peptides and Oligonucleotides.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peptides</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*The HTHi peptide reproduces the Gly-149 to Lys-186 sequence of HIV-1 IN Phe185Lys [20]. The loop region is underlined. The mutated residues are in bold characters. An N-terminal labeled carboxyfluorescein derivative of HTHi was also used (not shown). The peptides a4, K156 and INHS were dealt with in [30,43,50].

**The LTR34 oligonucleotides are designed to adopt a hairpin conformation, as already reported [43]. The hairpins contain a 17 bp stem, reproducing the HIV-1 U5LTR end, and a loop formed by three thymine nucleotides (TTT in bold characters). The fluorescein reporter group, f, is grafted either to the 5′ extremity (LTR34F5, CRE) or to the central T residue (f) of the loop (LTR34fm).

doi:10.1371/journal.pone.0004081.t001

<table>
<thead>
<tr>
<th>Table 2. Main interactions within HTHi (helices a4 and a5) and of HTHi with the neighbouring b-sheet in the IN catalytic core (data from PDB ID 1BIU, [20]).</th>
</tr>
</thead>
<tbody>
<tr>
<td>a4-helix</td>
</tr>
<tr>
<td>L172</td>
</tr>
<tr>
<td>A174</td>
</tr>
<tr>
<td>V176</td>
</tr>
<tr>
<td>Q177</td>
</tr>
<tr>
<td>M178</td>
</tr>
</tbody>
</table>

| a5-helix | A179 | V180 | φ | φ | φ |
| I182 | φ | φ | |
| H183 | E | φ | |
| N184 | φ | φ | |
| I186 | E | φ | |
| Q62 | φ | φ | |
| N64 | φ | E/H | |

| b-sheet | L68 | I73 | V75 | I84 | φ | φ | |

φ: hydrophobic contacts; H: hydrogen bonds; E: Electrostatic interactions.
doi:10.1371/journal.pone.0004081.t002
region of the CD spectra. Indeed, while the separate components of the bi-helix show important negative bands in the 195 nm domain of the spectrum, this feature vanishes in the spectrum of HTHi, which develops a positive band below 190 nm, compatible with a helical structure.

**Binding of the HTHi motif to viral DNA**

We carried out CD and fluorescence experiments to demonstrate that peptide HTHi binds with a good affinity to the viral LTR DNA.

To begin with, CD was used to assess the conformational changes undergone by the HTHi peptide upon interacting with LTR34 DNA (Table 1). Indeed, the latter exerts a helix stabilizing effect that is attested in the difference spectrum (i.e., peptide-DNA complex minus DNA) by the increase of the positive band at 190 nm and of the two negative bands at 208 and 222 nm, less perceptible in the spectra of unbound HTHi (Fig. 3). The HTHi motif practically doubles its helical content when binding to LTR34. This behavior is not unique: we have already shown that a peptide derived from the b-ZIP (basic-zipper) motif of the c-jun protein manifested the same helix stabilization upon binding to DNA [65].

The binding parameters were determined by fluorescence anisotropy measurements. A representative binding isotherm illustrating the titration of the oligonucleotide LTR34fm (Table 1b), by peptide HTHi is shown in Fig. 4a. The binding curve is biphasic, whereas that obtained with the structural analog LTR345 is monophasic. With LTR34fm, each plateau indicates the saturation of an interaction. The first, corresponding to the high affinity binding, has been assigned to the specific recognition by HTHi of the extremity of viral DNA [43]. The high affinity is abolished by the grafting of the fluorescein moiety at the 5’ extremity (LTR345) (Fig. 4a). The dissociation constants of peptide HTHi, determined by curve fitting, are $K_{d1} = 0.05 \text{ M}$ and $K_{d2} = 1.9 \text{ M}$ for LTR34fm and $K_{d2} = 2.9 \text{ M}$ for LTR345 (Fig. 4a). Note that the non-specific binding to the negative control sequence (CRE, cAMP response element) is within this latter range of affinities ($K_d = 6.6 \text{ M}$).

**Interaction of the HTHi motif with wt IN**

In a previous study, based on the intrinsic fluorescence properties of IN conferred by its six Trp residues, we have shown the ability of peptide INH5 (Table 1a) to interact with IN with a concentration midpoint ($C_{0.5}$) of $168 \text{ nM}$ [30]. The interaction, occurring at the $\alpha_4$:trp 5 interface, caused the dissociation of the tetramers and dimers of IN [30]. The experiments performed with HTHi under conditions similar to those we had previously used for peptide INH5 [30], show that HTHi interacts with wt IN at $100 \text{ nM}$ monomer concentration) with a $C_{0.5}$ value of $104 \text{ nM}$, while the peptide $\alpha_4$, taken alone, binds to wt IN with a $C_{0.5}$ value of $53 \text{nM}$ (Fig. 3b). Indeed, we had previously shown that peptide $\alpha_4$ interacted with IN likely by forming a coiled-coil structure with its counterpart in the protein [50,66].

**Interaction of the HTHi motif with LEDGF**

Fluorescence anisotropy was also used to assess the binding of the HTHi motif to IBD-LEDGF. A fluorescein-labeled HTHi peptide (Fig. 4c) was prepared for this purpose and titrated by both the wt and the Asp366Asn mutated LEDGF IBD protein ligands. It has been previously shown that Asp366 is essential for the binding of LEDGF to IN [52,54]. The overall pattern unambiguously shows that only the wt IBD binds HTHi. Curve analysis provides a $K_d$ of the order of $12 \text{ M}$.

We find that HTHi reproduces the properties of the CC of IN in so far as the recognition of LEDGF is
concerned. The essential role of the IBD residue Asp-366 in the interaction is confirmed.

**Discussion**

The so-called HTHi motif (α4-turn-α5) identified at the IN protein surface presents an unexpected multiplicity of functions revealed by the study of the corresponding synthetic HTHi peptide. In the latter, hydrophobic contacts between its α4 and α5 arms confer stability to the ensemble, i.e., to the tertiary structure and the helical secondary structure. The α4 helix is more basic than the α5 helix, which is consistent with its DNA binding property. However, only a stable preformed scaffold permits the proper spatial orientation of the interactive amino acid side chains and their adjustment onto their complementary nucleotides of the DNA target without large conformational change and therefore large consumption of energy. This improves the affinity and confers specificity to the binding. 

Our results strongly suggest that helix stability in the HTHi motif is essential to the specific recognition of viral DNA. In particular, the α4 peptide taken in isolation, which does not benefit of the HTHi stabilizing context, is a very poor DNA binder. The difference of DNA binding affinity between the peptides HTHi and α4 can be attributed to the larger conformational change that the latter must undergo in order to adopt the bound helical conformation.

In a previous work we pointed out the importance of the conformational entropy in the binding of the α4 helix to the viral DNA. In order to reduce this effect, we had designed an analog of the α4 peptide that had its secondary structure stabilized by helicogenic mutations (Gly149Ala, Ile161Leu, Ile162Leu and Gly163Ala). The resulting peptide, K156, was a good mimic of the α4 helix in the protein context and was better organized for binding than peptide α4. The use of peptide K156 with the oligonucleotide LTR34fm permitted the identification of a high affinity site (Kd1 = 2.1 nM), corresponding to the attachment site of the enzyme on the six outermost bases of viral DNA and of a low affinity site in the micromolar range (Kd2 = 54 μM) corresponding to the non-specific binding of IN[43]. The grafting of the bulky fluorescein at the 5′ extremity (LTR34f5; Table 1b), i.e., at a position in close proximity to the 3′-processing site, suppressed the high affinity binding mode of K156 but preserved the low affinity mode. The absence of effect on the binding to the low affinity site from the grafted fluorescein suggested that this site was distant on the DNA from the high affinity site [43].

The behavior of peptide HTHi is similar to that of peptide K156. With HTHi there are two binding modes as well, one with a low dissociation constant in the nanomolar range (Kd1 = 50 nM) and the other with a high dissociation constant in the micromolar range (Kd2 = 1.9 μM), in the case of LTR34fm; and there is only one binding mode in the micromolecular range (Kd = 2.9 μM ) in the case of LTR34f5 (Fig. 4a). The non-specific binding of peptide HTHi to the negative control sequence, CRE, is within this latter range of affinities; Kd = 6.6 μM. ; comparatively, that of peptide K156 was 16 μM [43].

Thus, the two peptide ligands, HTHi and K156, present the same DNA binding properties. The two binding modes, of which the first appears specific for the intact processing-attachment site of LTR DNA, are present in the two cases, albeit with different affinities. The role of conformation is noteworthy. Peptide K156, whose secondary structure has largely benefited from helicogenic mutations, manifests a greater affinity for viral DNA than peptide HTHi, in which the α4 helix component, certainly stabilized by its interactions with the α5 helix, is nevertheless less stable than...
cases, the fluorescein-labeled DNA was at 20 nM. The $K_d$ values
anisotropy titration of fluorescein-grafted HTHi peptide (50 nM) by
a
$6\times10^4$ respectively. 

Concerning the role of the HTHi motif is not limited to that of its $\zeta_4$ and $\zeta_5$

The role of the HTHi motif is not limited to that of its $\zeta_4$ and $\zeta_5$

Figure 4. Fluorescence. a). Interaction with DNA. Fluorescence
anisotropy titration by peptide HTHi of LTR34fm (squares, red) and
LTR34f5' (triangles, blue). Inset: titration of CRE by peptide HTHi. In all
cases, the fluorescein-labeled DNA was at 20 nM. The $K_a$ values
obtained by curve fitting (with $\pm 10\%$ uncertainty) are: 0.05 $\mu$M and
1.9 $\mu$M for LTR34fm; 2.9 $\mu$M for LTR34f5'; and 6.6 $\mu$M for CRE. b).
Interaction with wt IN. Quenching of the intrinsic fluorescence of wt
IN by peptide HTHi. Inset: Titration of wt IN by peptide $\zeta_4$. In all cases,
the wt IN monomer concentration was 100 nM. The midpoints of the
titration curves are at $104\pm8$ nM and $53\pm5$ nM for peptides HTHi and
$\zeta_4$ respectively. c). Interaction with LEDGF-IBD. Fluorescence
anisotropy titration of fluorescent-grafted HTHi peptide (50 nM) by
LEDGF-IBD (wt IBD, round symbols, and mutated IBD, squares). The $K_a$
of the wt IBD-HTHi interaction is $12.3\pm0.3$ $\mu$M, with a Hill number of
11.$\pm3$. 

doi:10.1371/journal.pone.0004081.g004

peptide K156. This is consistent with the idea of a higher pre-
formed conformation for binding of peptide K156 compared with
peptide HTHi that lowers the entropy cost of interaction of the
former. As a corollary, the isolated peptide $\zeta_4$, which mostly
displays random conformation because of the absence of either
helicogenic mutations in its sequence or helix stabilization
conferred by the interaction with $\zeta_4$, is completely unable to
specifically bind the viral DNA.

X-ray crystallographic studies of the co-crystal of CC with
5CITEP also illustrate the importance of the HTHi $\zeta_5$ helix in
the IN enzyme. Results strongly suggest that the 44 helix would be the
primary target of the DKA (diketo acid/aryl) family of inhibitors of IN
acting on the 3' processing step [14,42,67]. In the complex, the
drug interacts with five residues of the $\zeta_5$ helix and at least four of
them (Lys-156, Lys-159, Glu-148 and Glu-152) are also implicated in the
stabilization of the viral DNA-CC complex [14,43].

Concerning the role of the $\zeta_2$ helix, it is clear from the various
crystal structures that it participates to both the dimerization of the
enzyme and the stabilization of the motif [18–29]. There are
examples in the literature showing that the so-called stabilizing
helix of HTH motifs is involved in the protein dimerization.
This is the case of topoisomerase II, in which the binding site to DNA is
fashioned from the interaction of two HTH motifs [68]. The
implication of the $\zeta_2$ helix in the enzyme structure and activity has
been highlighted in our previous report showing that peptide
INH5, which comprises the $\zeta_2$ helix, inhibits IN by dissociating
the IN dimers and tetramers [30].

The role of the HTHi motif is not limited to that of its $\zeta_4$ and $\zeta_5$
helix components. At the solvent accessible surface, in the center of
the motif, the electrostatic potential is uniform and positive and it
changes sign at the turn connecting the $\zeta_4$ and $\zeta_5$ helices (Fig. 1d).
The 161–173 region is a strong epitope and also a protein binder
[49,51–56]. Antibodies enriched against this epitope inhibit both the
3' processing and the strand transfer reaction [49]. The particular variation of the electrostatic potential at the solvent
accessible surface in that region is compatible with these functions.
As a matter of fact, a peptide corresponding to the epitope region
has been shown to mediate the nuclear import of the covalently
linked serum albumin [58], although this region has been
incorrectly proposed as being a possible NLS (nuclear localization
signal) [57]. Nevertheless, the present results confirm the
propensity of the epitope region to associate with important
proteins, such as LEDGF. The co-crystal structure of CC bound to
LEDGF-IBD shows that the epitope region is recognized by LEDGF.
The formation of a complex between IN and LEDGF is
biologically relevant and is a prerequisite to the strand transfer into
the cell chromosome [51–56]. In the complex, there are numerous
stabilizing contacts, such as those involving the $\zeta_5$ helix and a
portion of the $\zeta_4$ helix, but it is the turn region of HTHi (residues
Glu-170, His-171 and Glu-168 ) that makes the major contribu-
tion, particularly in binding the hotspot residue Asp-366 and the
residue Ile-365 of LEDGF[51,52,54]. Our results are compatible
with these findings. Resembling the IN CC, the peptide HTHi
binds wt IBD, but not its Asp366Asn variant. We are thus allowed
to conclude that the pattern of interactions found in the CC-IBD
crystal is still present in the peptide HTHi-IBD complex.

All together, the results mentioned above indicate that the
HTHi motif identified in IN behaves like a multifunctional entity.
Being involved in enzyme oligomerisation (via its $\zeta_4$ helix), in LTR
end recognition (via its $\zeta_5$ helix) and in binding to LEDGF (via the
turn region, which also acts as a strong epitope), it emerges as a
central piece of both the IN structure and activity. Moreover, the
$\zeta_4$ helix of HTHi could be the target of inhibitors belonging to the
DKA family, as suggested by the crystal structure of 5 CITEP
bound to IN CC [14,42,67]. We propose that HTHi could constitute a reliable model for the study of new inhibitors acting at the IN-IN, the IN-DNA and the IN-LEDGF interfaces. This issue will be further addressed possibly by looking for peptidomimetics to the α and β helices, as well as to the epitope region–with the aim of transforming them into therapeutic agents.

**Materials and Methods**

**Peptides**

We used the synthetic peptides shown in Table 1a. They were made according to the Fmoc procedure, as already reported for peptide α4, for its “stabilized” analogue i.e., peptide K156, and for the peptide INH5, derived from the α1 helix [30,43]. Peptide α4 reproduces the α1 helix sequence of IN; peptide K156 is a variant of the α4 peptide resulting from the replacement of weakly helicogenic residues by more helicogenic ones in several biologically irrelevant positions [43]; peptide INH5 contains both the α1 helix and the turn connecting the α2 helix to the α1 helix [30]. The peptides HTHi and wt HTHi reproduce the sequences of the HTHi motifs in the IN-Phe185Lys variant and in wt IN respectively. A version of the HTHi peptide with a carboxylfluorescein at the N-terminus has also been prepared. Tyr or Trp aromatic residues, when absent in the native sequences, were purposely added in order to enable peptide quantification by UV absorption spectra. Peptide concentrations were determined using a molar absorption coefficient at 280 nm equal to 1 280 M⁻¹ cm⁻¹ for the tyrosine-containing peptides α4, INH5 and HTHi and 3600 M⁻¹ cm⁻¹ for the tryptophane-containing peptides K156 and wt IN.

**DNA oligonucleotides**

The oligonucleotides (Table 1b) were purchased from Cybergene ESGS (France) and Eurogentec (Belgium). The choice of monomolecular hairpin-forming oligonucleotides, rather than bimolecular duplex-forming ones, was motivated by the need for stability under the low concentrations inherent to the fluorescence anisotropy experiments. The fluorescence reporter group is grafted either to the central T nucleotide (LTR34f5') or to the 5' extremity (LTR34S and CRE). The CRE (cAMP responsive element) sequence was used as negative control. The version of LTR34 without fluorescein was used in CD studies.

**Proteins**

Fluorescence titrations were performed with the wild-type IN (wt IN) [69]. The presence of Trp in the enzyme (IN contains six Trp residues) but not in the peptides, was exploited when performing intrinsic fluorescence quenching titrations of the former. The wt IBD is the fragment 347–471 of LEDGF (GST removed) [33,51,52]. The mutant IBD is the Asp366Asn version of the former. Like the whole protein, these contain no Trp. The concentrations of wt IN (33.781 kDa) and of the IBD fragments (16.542 kDa) were estimated from UV absorption at 280 nm, using the molar absorbivities of 46542 M⁻¹ cm⁻¹ and 1400 M⁻¹ cm⁻¹ respectively. The proteins were used in the "reaction buffer" containing 20 mM HEPES (pH 6.8), 10 mM MgCl₂ and 10 mM DTT.

**Secondary structure predictions**

Secondary structure predictions were carried out using the AGADIR and GOR IV computer programs available on the web respectively at http://www.cmmr-heidelberg.de/serices/serrona/agadir-start.html and http://phil.univ-lyon1.fr/ (Pôle Bioinformatic Lyonnais, France) [63,64]. The AGADIR prediction considers short range interactions between residues and provides helical propensity per residue of peptides in solution, independently of tertiary structure interactions, with the possibility of selecting the pH and the temperature conditions. GOR IV uses parameters derived from crystallographic data of proteins. It thus provides more realistic structure predictions for peptide segments submitted to tertiary and quaternary structures constraints i.e., within the protein environment. Thus the two types of predictions can be used in conjunction to find the impact of the protein context on the secondary structure of its peptide elements, to identify the effect of mutations on secondary structures and to select the mutations that would reinforce the helicity of peptides, as was done in the case of peptide K156 (Table 1a) [43].

**CD Spectroscopy**

CD spectra were recorded on a Jobin-Yvon CD6 dichrograph (HORIBA Jobin-Yvon, France). Peptide concentrations varied from 5 to 25 μM in the "assay buffer", containing 40 mM sodium phosphate, 0.5 mM EDTA, pH 6.0. The samples, incubated 10 min at the chosen temperature to allow the solutions to reach their equilibrium state, were placed in thermally jacketed cells with a 1 mm path length. The spectra, recorded in 1 nm steps, were averaged over ten scans and corrected for the base line, then presented as differential molar absorptivities per residue, ∆ε (M⁻¹ cm⁻¹), for peptides, and as difference spectra (complex minus DNA), for peptide-DNA complexes. For the HTHi-LTR interaction, aliquots of peptide were added to LTR34 (10 μM) and the control spectrum of unliganded LTR was subtracted from that of the complex. Recall that in a simple random coil – α helix equilibrium, the α-helix content of peptides could be approximated by the relation: Pₐ = -[Δε222 x 100] / (Pₐ; percentage of α helix; Δε222: differential molar absorptivity per residue at 222 nm) [70].

**Modeling**

We used the Insight II® program (Accelrys Software Inc., CA, USA) to estimate distances between the various atoms in HTHi and between HTHi and the β-stands located in the proximity, using the CC IN crystal structure [PDB ID 1BIU] [20] (Fig. 1 and Table 2); Graphics was by PyMol (http://www.pymol.org) [71] with the APBS software (http://apbs.sourceforge.net/) [72].

**Fluorescence Measurements**

Intrinsic fluorescence quantum yield and fluorescence anisotropy studies were carried out with a Jobin-Yvon Fluoromax II instrument (HORIBA Jobin-Yvon, France) equipped with an ozone-free 150 W xenon lamp. The samples [800 μl] were placed at 5°C in thermally jacketed 1 cm 0.5 cm quartz cells. At least ten measurements for each titration point were recorded with an integration time of 5 sec.

The intrinsic fluorescence [73] of wt IN was measured at 100 nM monomer concentration, in the "reaction buffer" (see above). The Trp residues were excited at 295 nm and emission was recorded between 305 and 450 nm, with 2 nm and 5 nm excitation and emission slit widths, respectively. A maximal emission was obtained at 337 nm. Quenching of the fluorescence of IN following the addition of HTHi or of peptide α4 (void of Trp) was expressed as 1-F/F₀, where F₀ is the fluorescence of the enzyme in the absence of peptide.

Fluorescence anisotropy titration [74,75]. Fluorescein-labeled oligonucleotides were used at 20 nM in the "assay buffer" (see Table 1b and above). With fluorescein as fluorophore, excitation and emission were at 488 nm and 516 nm, with 4 nm and 5 nm slits, respectively. Peptides were added to DNA as successive aliquots. For each anisotropy measurement, the parallel (I HH) and the
perpendicular ($I_\perp$) intensities of the background solution (i.e., buffer and protein contributions) were subtracted from those of the sample.

The fluorescence anisotropy of the fluorescein-conjugated version of HTHi (50 nM) titrated by wt and mutated IBD in the "assay buffer" was used to evaluate the $K_i$ of the LEDGF-HTHi complex.

Acknowledgments
We thank Alan Engelman and Moped Jamal Dar (Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA) for the generous gift of the wt LEDGF-IBD and of the Asp566Asn LEDGF-IBD variant. We wish also to thank Christophe Piase and Jean-Pierre Levillain for skilled assistance in peptide synthesis and Jean-François Moussaclet for the gift of the HIV-1 integrases.

Author Contributions
Conceived and designed the experiments: HP OM SF. Performed the experiments: HM HP LZ ZH. Analyzed the data: HM HP LZ RGM OM SF. Contributed reagents/materials/analysis tools: BR. Wrote the paper: HP OM SF. As a part of PhD thesis work: HM. As Director of thesis to HM: HP. As a part of a joint PhD thesis, Paris, Beyrouth: ZH.

References


