The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant

Grzegorz Bujacz[a,b], Jerry Alexandratos[a], Qing Zhou-Liu[c], Christine Clément-Mella[c], Alexander Wlodawer[a,*]

[a] *Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, MD 21702, USA*
[b] *Faculty of Food Chemistry and Biotechnology, Technical University of Łódź, 90924 Łódź, Poland*
[c] *Service de Biochimie, Rhône-Poulenc Rorer SA, 94403 Vitry sur Seine, France*

Received 20 September 1996; revised version received 18 October 1996

Abstract We solved the structure and traced the complete active site of the catalytic domain of the human immunodeficiency virus type 1 integrase (HIV-1 IN) with the F185H mutation. The only previously available crystal structure, the F185K mutant of this domain, lacks one of the catalytically important residues, E152, located in a stretch of 12 disordered residues [Dyda et al. (1994) Science 266, 1981–1986]. It is clear, however, that the active site of HIV-1 IN observed in either structure cannot correspond to that of the functional enzyme, since the cluster of three conserved carboxylic acids does not create a proper metal-binding site. The conformation of the loop was compared with two different conformations found in the catalytic domain of the related avian sarcoma virus integrase [Bujacz et al. (1995) J. Mol. Biol. 253, 333–346]. Flexibility of the active site region of integrases may be required in order for the enzyme to assume a functional conformation in the presence of substrate and/or cofactors.

Key words: Integrase; Active site; Disorder; Metal binding

1. Introduction

The retroviral integrase (IN) is one of only three virus-encoded enzymes, essential for the retroviral replication cycle. Because IN catalyzes the insertion of viral DNA into many sites on host DNA [1–3], this enzyme is an attractive target for inhibition of lentiviruses such as human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS). Analysis of the primary structure of IN coupled with deletion mutagenesis has indicated that this enzyme, consisting of about 290 amino acids, contains three structurally important domains [3,4]. Although the structure of complete IN is not yet available, crystal structures of the catalytic domains of both HIV-1 IN [5] and avian sarcoma virus (ASV) IN [6] have been solved. In addition, the NMR structure of the DNA-binding C-terminal domain of HIV-1 IN is also available [7,8].

A major difference between the structures of HIV-1 IN and ASV IN is in the conformation of the active site and the presence of metal ligands. In ASV IN, the active site is completely ordered, and a divalent cation (Mg^{2+} or Mn^{2+}) can be placed and removed [9]. However, a loop containing residues G146–G152 is found in different conformations in crystals grown under different conditions (polyethylene glycol (PEG) and ammonium sulfate). Although this flexible loop is adjacent to E157, one of the residues required for the enzymatic activity, the conformation of the helical segment containing E157 is very similar in both cases. In the structure of HIV-1 IN [5], however, the corresponding loop (residues I141–S153) is totally disordered, and the equivalent residue in the active site, E152, was not observed.

A comparison of the structures of these two catalytic domains indicated that the interactions in the vicinity of K185, a residue mutated in HIV-1 IN for solubility reasons, were similar to those near the corresponding H198 in ASV IN [10]. This similarity suggested to us that using a histidine rather than a lysine residue at 185 would be a more rational way of making a soluble catalytic domain of HIV-1 IN, which would be even more likely to show interactions similar to those found in the ASV IN. We present here the results of a crystallographic study of the F185H mutant of HIV-1 IN at 2.6 Å resolution. In addition to confirming the intermolecular interactions involving this mutation, we were able to provide an unambiguous tracing of the complete chain in the active site.

2. Materials and methods

Expression, purification, and crystallization of the soluble construct of the catalytic core domain of HIV-1 IN have been performed essentially as described previously [5,11]. The protein was stored at −20°C, in a buffer containing 50 mM HEPES, pH 7.5, 1 M NaCl, 0.7% CHAPS, 5 mM dithiothreitol, 5 mM β-mercaptoethanol, and 20% glycerol. Just before crystallization experiments were set up, the solution was loaded onto a Mono S HR10/10 column (Pharmacia) and eluted with a linear gradient of 0–1 M NaCl, in 50 mM MES, pH 5.5, 1 mM EDTA, 5 mM β-mercaptoethanol, and 10% glycerol. The purified protein was dialyzed in a simplified buffer containing 50 mM MES, pH 5.5, 50 mM NaCl, and 5 mM dithiothreitol and concentrated to 10 mg/ml. Crystals were grown using either the sitting drop vapor diffusion method or the dialysis button method from protein solutions at concentrations from 5 to 8 mg/ml mixed 1:1 with precipitation buffer containing 0.1 M sodium cacodylate, pH 6.5, 0.4 M ammonium sulfate and 3–9% PEG 8000. Crystals grew to their final size of $0.35 \times 0.35 \times 0.35$ mm^3 in several days at 19°C.

Diffraction data were collected at −165°C on an RAXIS-IIC image plate detector, using mirror-monochromated CuKα radiation. The crystal was transferred from the mother liquor through a layer of glycerol-enriched (15%) mother liquor onto a nylon loop, then shock-frozen in a nitrogen gas stream. Diffraction intensities were processed with DENZO and scaled with SCALEPACK [12]. The isomorphous structure of the F185K mutant [5] was used directly as the initial model. Data with $F > 2\sigma(F)$ were used in refinement with a low-resolution cut-off of 10 Å. The structure refinements were carried out using the program PROLSQ [13]. No data were reserved for *R*-free calculations since the two structures were very similar, and we saw no

*Corresponding author. Fax: (1) (301) 846-6128.
E-mail: wlodawer@ncifcrf.gov

Abbreviations: AIDS, acquired immunodeficiency syndrome; ASV, avian sarcoma virus; HIV-1, human immunodeficiency virus type 1; IN, integrase; PEG, polyethylene glycol

PII S0014-5793(96)01236-7

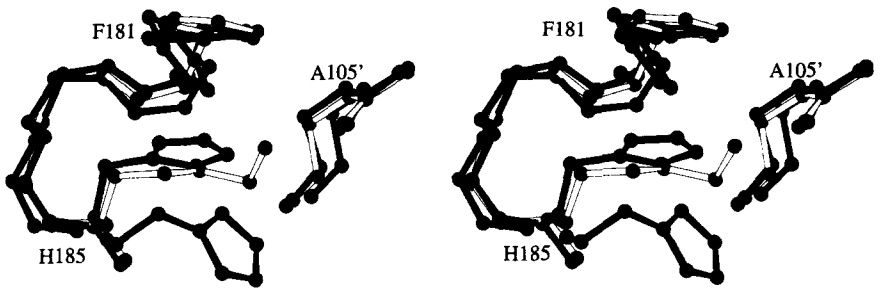

Fig. 1. Stereo view of intermolecular interactions on the dimer interface in the vicinity of the mutated residue, H185 (black) and K185 (white), superimposed on the equivalent region in ASV IN (grey). This figure, as well as Fig. 2, was prepared with the program MOLSCRIPT [15].

compelling reason to decrease the number of observations used in the refinement. The final conventional R factor was 0.200 (Table 1). The coordinates have been deposited with the Brookhaven Protein Data Bank (code 2itg).

3. Results and discussion

As expected, the structure of the F185H mutant of the catalytic domain of HIV-1 IN is very similar to the previously published structure of the F185K mutant. The catalytic domain of HIV-1 IN is formed by a five-stranded mixed β-sheet flanked by six α-helices. The main part of the β-sheet, formed by three antiparallel β-strands (β1, β2, and β3), is extended by two short parallel β-strands (β4 and β5). On one side of the β-sheet are three short helices (α1, α2, and α3). The opposite side of the β-sheet is covered by three long antiparallel helices (α4, α5, and α6). The long loop between β5 and α4 has a very extended conformation. The dimer is created around the crystallographic 2-fold axis. The interface is formed mainly by the α1 helix of one monomer and the α5 helix of the other monomer, utilizing predominantly hydrophobic interactions. Residues from the β3 strand, which is located deeper within the monomer, contribute to the central part of the dimer interface.

The F185H mutant of HIV-1 IN was engineered to provide rational interactions between molecules of the dimer in the region where systematic mutagenesis of hydrophobic residues established that the F185K mutation produced a soluble IN core [5]. Since histidine was found in the structurally equivalent location in ASV IN, we assumed that placing histidine at this site instead of lysine might make the interactions in the two enzymes more similar. However, the observed pattern of hydrogen bonds did not exactly mimic the arrangement seen in ASV IN. The ASV IN histidine creates a good hydrogen bond with a main-chain carbonyl oxygen from the neighboring molecule, although shifted one residue down the chain. This distance, 3.3 Å for NE2-194 to O-110, is similar to that observed in the K185 HIV-1 IN construct. On the other hand, both HIV-1 IN structures show the side chain of residue 185 making contact with the same main chain O-105 (Fig. 1). The hydrogen bond distance between NE2-185 and O-105 of the neighboring monomer in the dimer is 4.31 Å, greater than that observed in the previously published HIV-1 IN structure in which the NZ-185 to O-105 distance is 2.6 Å. In spite of the reduced strength of the hydrogen bond that was expected to stabilize this region, the area near the F185H mutation site and the dimer interface as a whole are essentially unchanged. It is, therefore, extremely unlikely that the F185H mutation played any role in stabilizing the extended conformation of the active site loop.

The interaction between the mutated H185 and the neighboring monomer is probably more similar to what may occur in this location with the native phenylalanine than with the lysine found in the other soluble form of HIV-1 IN. In addition to the decreased hydrogen bond interaction, we observed stacking interactions between the aromatic rings of H185 and F181 in the same α-helix. The distance between CE1-185 and CZ-181 is 3.14 Å. This stacking interaction may also stabilize the native protein. A stacking interaction is not observed between analogous residues H198 and Y194 in ASV IN, which may be due to the slightly different histidine side-chain hydrogen bond conformation. Also, Y194 is more hydrophilic and therefore may not require the same stabilization given by π-stacking.

Comparison of the loops in different structures that include the third active site residue, E152 (E157 in ASV IN), indicated that the last residue where ASV and HIV-1 IN superimpose is Q137 (H142 in ASV IN). This is also the point where the structures of F185K and F185H diverge. Although the main chain of E138 is visible in the former construct, the side chain is truncated. At the other end of the loop, the α-4 helix starts at A154 in both HIV-1 IN structures, with good alignment of main and side chains. Even for this structure with a fully visible loop, the HIV-1 IN α-4 helix is more than one full

Table 1
Data collection and refinement statistics for HIV IN (F185H)

Statistics	
Unit cell parameters	$a=b=72.40$ Å, $c=65.56$ Å
Space group	$P3_121$
R merge (%)	8.5
No. of reflections	
Total ($0\sigma(F)$ cut-off)	14 682
Unique	5 012
Used for refinement ($2\sigma(F)$ cut-off)	4 400
Completeness of data (%)	79.8
No. of atoms	
Protein	1 234
Water	58
Resolution range (Å)	10.0–2.6
R factor	0.200
Mean B factor	34.16
r.m.s. deviations from ideality	
Bond (Å)	0.011
Angle distances (Å)	0.032
Plane restraints (Å)	0.013
Chiral volumes (Å^3)	0.135

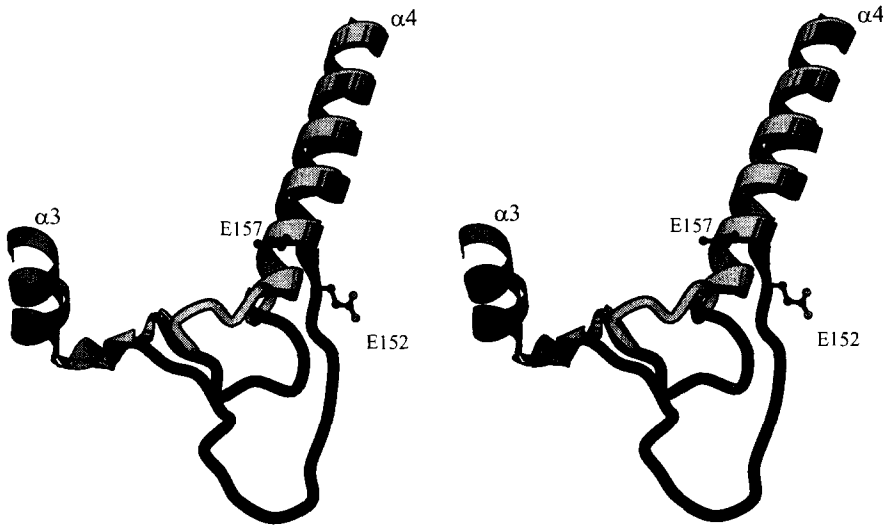

Fig. 2. Stereo view of residues 124–165 of HIV-1 IN (black), superimposed on the equivalent region (residues 129–174) of the ASV IN crystallized from PEG (light grey) and from ammonium sulfate (dark grey).

turn (5 residues) shorter than any comparable α-4 helix in the ASV IN structures.

The active site loop of HIV-1 IN is stabilized by three contacts between side chain residues and a symmetry-mate molecule which is not part of the dimer. The major role in stabilization is played by a hydrogen bond between Y143 OH and E157 OE2, with a distance of 2.97 Å. The tyrosine side chain was already clearly visible in electron density maps after the first run of refinement. Additional stabilization is provided by weak intermolecular hydrogen bonds involving side chains of two glutamine residues, Q146 and Q148, with distances of 3.4–3.6 Å. None of these residues are near the F185H mutation or immediately adjacent to the active site, nor are they involved in formation of the dimer interface. One possible explanation of these observations is that the differences in freezing techniques employed in by us and in previous work [5] may be responsible for allowing stabilization of the flexible loop.

The visible section of the active site loop in HIV-1 IN is more similar to that of ASV IN grown in ammonium sulfate than to the crystal form grown in PEG (Fig. 2). The loop is in an even more extended conformation in HIV-1 IN, with residues 146–153 essentially linear. The main chain can be traced completely in the electron density map (Fig. 3), and all side chains are visible except the side chains of residues Q146 and Q148, which have poor electron density. These loops have comparable temperature factors in ASV IN and HIV-1 IN. The average *B* factor of the HIV-1 IN structure is 34.2 Å^2, with the loop in the high 60's. The average temperature factor of ASV IN crystallized from ammonium sulfate is 26 Å^2, with loop *B* factors in the 70's. In both HIV-1 IN constructs, the C- and N-termini are flexible, but in the struc-

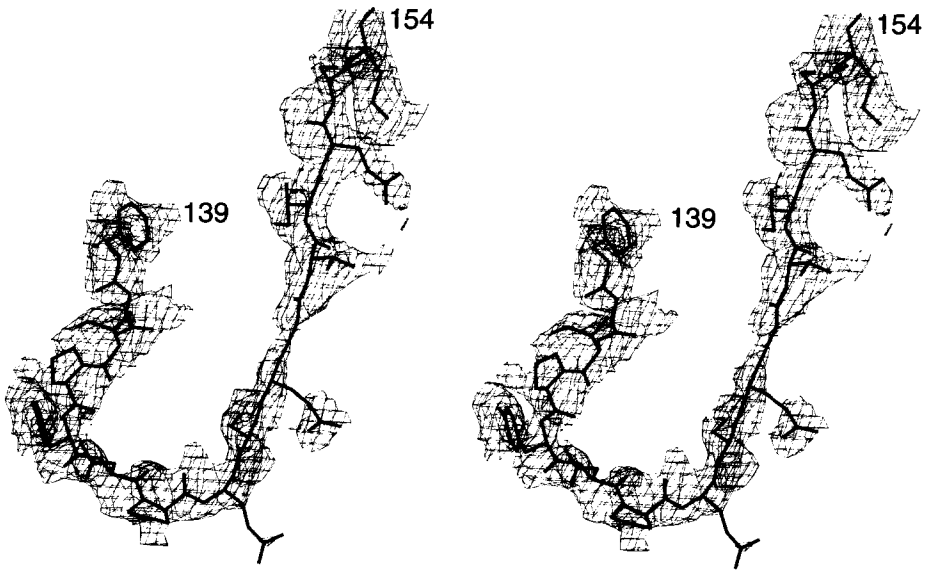

Fig. 3. Stereo view of an electron density map covering residues F139–M154 of the catalytic domain of HIV-1 IN. This $2F_o-F_c$ map, contoured at the 0.9σ level, was calculated at 2.6 Å resolution after refinement of the model. The density corresponding to the loop is very clear. This figure was prepared with the program CHAIN [16].

ture presented here we were able to place four additional residues on the N-terminus and one residue on the C-terminus.

Despite being able to see all three residues crucial for the enzymatic activity of integrase, we suspect that the conformation of the active site does not correspond to that present in the active enzyme. The carboxylate groups of D64, D116, and E152 are not very close to each other, and in fact the glutamate points completely away from the other two groups, quite different from the structures of ASV IN with cations bound [9]. Although an HIV-1 IN crystal structure with bound metal cations is not yet available, antibody binding experiments show that HIV-1 IN undergoes significant rearrangement upon binding of these essential cofactors (Asante-Appiah, E. and Skalka, A.M., personal communication). A similar situation has been previously reported for the catalytic core of the bacteriophage Mu transposase, in which the third required residue is quite far from the other two [14]. Indeed, disordering of the active site might be a specific control mechanism for these enzymes, and the active site of HIV-1 IN that we observed most likely differs from that expected in the presence of substrates and cofactors.

Acknowledgements: We thank J. Becquart, J.D. Guitton, J.Y. Crenne, and P.F. Berne for constructive discussions and helpful advice throughout this project, and A. Arthur for editorial comments. Research sponsored in part by the National Cancer Institute, DHHS, under contract with ABL. The project was also carried out as part of the BioAvenir program supported by Rhône-Poulenc with the participation of the Ministère de l'Enseignement Supérieur et de la Recherche and the Ministère de l'Industrie. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

[1] Katz, R.A. and Skalka, A.M. (1994) Annu. Rev. Biochem. 63, 133–173.
[2] Goff, S.P. (1992) Annu. Rev. Genet. 26, 527–544.
[3] Vink, C., Groeneger, O.A.M. and Plasterk, R.H. (1993) Nucleic Acids Res. 21, 1419–1425.
[4] Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J. and Skalka, A.M. (1991) Nucleic Acids Res. 19, 851–860.
[5] Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Science 266, 1981–1986.
[6] Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1995) J. Mol. Biol. 253, 333–346.
[7] Eijkelenboom, A.P.A.M., Lutzke, R.A.P., Boelens, R., Plasterk, R.H.A., Kaptein, R. and Hard, K. (1995) Nat. Struct. Biol. 2, 807–810.
[8] Lodi, P.J., Ernst, J., Kuszewski, J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1995) Biochemistry 34, 9826.
[9] Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1996) Structure 4, 89–96.
[10] Jenkins, T.M., Hickman, A.B., Dyda, F., Ghirlando, R., Davies, D.R. and Craigie, R. (1995) Proc. Natl. Acad. Sci. USA 92, 6057–6061.
[11] Bushman, F.D., Engelman, A., Palmer, I. and Craigie, R. (1993) Proc. Natl. Acad. Sci. USA 90, 3428–3432.
[12] Otwinowski, Z. (1992) An Oscillation Data Processing Suite for Macromolecular Crystallography, Yale University, New Haven.
[13] Hendrickson, W.A. (1985) Methods Enzymol. 115, 252–270.
[14] Rice, P. and Mizuuchi, K. (1995) Cell 82, 209–220.
[15] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[16] Sack, J.S. (1988) J. Mol. Graphics 6, 244–245.