J Mol Model (2011) 17:2693–2705
DOI 10.1007/s00894-011-0971-1

ORIGINAL PAPER

A comparative study of HIV-1 and HTLV-I protease structure and dynamics reveals a conserved residue interaction network

Pia Rücker · Anselm H. C. Horn · Heike Meiselbach · Heinrich Sticht

Received: 17 September 2010 / Accepted: 11 January 2011 / Published online: 29 January 2011

Abstract The two retroviruses human T-lymphotropic virus type I (HTLV-I) and human immunodeficiency virus type 1 (HIV-1) are the causative agents of severe and fatal diseases including adult T-cell leukemia and the acquired immune deficiency syndrome (AIDS). Both viruses code for a protease that is essential for replication and therefore represents a key target for drugs interfering with viral infection. The retroviral proteases from HIV-1 and HTLV-I share 31% sequence identity and high structural similarities. Yet, their substrate specificities and inhibition profiles differ substantially. In this study, we performed all-atom molecular dynamics (MD) simulations for both enzymes in their ligand-free states and in complex with model substrates in order to compare their dynamic behaviors and enhance our understanding of the correlation between sequence, structure, and dynamics in this protein family. We found extensive similarities in both local and overall protein dynamics, as well as in the energetics of their interactions with model substrates. Interestingly, those residues that are important for strong ligand binding are frequently not conserved in sequence, thereby offering an explanation for the differences in binding specificity. Moreover, we identified an interaction network of contacts between conserved residues that interconnects secondary structure elements and serves as a scaffold for the protein fold. This interaction network is conformationally stable over time and may provide an explanation for the highly similar dynamic behavior of the two retroviral proteases, even in the light of their rather low overall sequence identity.

Keywords Molecular dynamics simulation · Protein structure · Protein–ligand interactions · HIV · HTLV

Abbreviations

ATL	Adult T-cell leukemia
CA/NC cleavage site	Protease cleavage site in HTLV-I polyprotein precursor between the capsid and nucleocapsid proteins
HIV-PR	HIV-1 protease
HTLV-PR	HTLV-I protease
MA/CA cleavage site	Protease cleavage site in HIV-1 and HTLV-I polyprotein precursors between the matrix and capsid proteins
MD	Molecular dynamics
p2/NC cleavage site	Protease cleavage site in HIV-1 polyprotein precursor between the p2 and nucleocapsid proteins
PDB	Protein Data Bank
RMSF	Root-mean-square fluctuation
TSP/HAM	Tropical spastic paraparesis

Introduction

In 1980, human T-lymphotropic virus type I (HTLV-I) was the first human retrovirus to be isolated [1], and it was later identified as the causative agent of severe diseases such as adult T-cell leukemia (ATL) [2] and tropical spastic paraparesis (TSP/HAM) [3], as well as a variety of other

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-0971-1) contains supplementary material, which is available to authorized users.

P. Rücker · A. H. C. Horn · H. Meiselbach · H. Sticht (✉)
Bioinformatik, Institut für Biochemie and Emil-Fischer-Zentrum,
Friedrich-Alexander-Universität Erlangen-Nürnberg,
Fahrstr. 17,
91054 Erlangen, Germany
e-mail: h.sticht@biochem.uni-erlangen.de

diseases [4–7]. Ten to twenty million people are estimated to be infected with HTLV-I worldwide [8], of whom about 10% are at high risk of developing ATL or TSP/HAM [9, 10]. Human immunodeficiency virus type 1 (HIV-1), another member of the retrovirus family, was also first discovered in the early 1980s [11], and was quickly linked to another devastating disease in humans: the emerging worldwide pandemic of acquired immune deficiency syndrome (AIDS) [12].

Many antiviral drugs for the treatment of HIV-1 have been developed and released. Today, treatment options for HIV-1 patients are based on a variety of substances that use different strategies to block the viral life cycle and have dramatically improved life expectancy. The largest and most efficient group of drugs is directed against the viral protease—the enzyme that is essential for processing viral proteins from a polyprotein precursor [13]. However, the problem of the rapid emergence of resistant strains still poses a challenge to researchers and clinicians.

There is still no specific antiviral agent available for the treatment of HTLV-I infections, as most anti-HIV drugs have proven to be ineffective. ATL patients can be treated with Zidovudine (AZT), a nucleoside analog and inhibitor of HIV reverse transcriptase, but inhibition of HTLV replication has not been shown to date [14]. Available PR inhibitors also do not appear to block HTLV-I protease (HTLV-PR) in therapeutic doses [15]. Targeting HTLV-PR therefore requires the development of new drugs against HTLV-I. When the crystal structure of HTLV-PR was solved, this was an essential step towards rational drug-design approaches [16]. Despite the low sequence identity of 31% between HIV-1 protease (HIV-PR) and HTLV-PR, the folds of the two proteases are very similar (Fig. 1). Yet, similarities and differences between the two enzymes have not been sufficiently characterized, and no new inhibitors that specifically target HTLV-PR have been identified to date. In our study, we aim to compare the dynamic behavior and underlying structural features of HIV-PR and HTLV-PR in order to provide a comprehensive picture of the two retroviral proteases regarding similarities and differences. Here, we examine for the first time the dynamics of HTLV-PR and propose that a stable interaction network of conserved residues form the basis for the common structural and dynamic features in HTLV-PR and HIV-PR, and possibly other retroviral proteases.

Methods

The starting structures of the proteases in complex with substrate analogs were taken from the Protein Data Bank (PDB) [17]. The PDB entry used for HIV protease complexed with a substrate analog was 4HVP [18]. The HIV protease of the 4HVP file containing the substrate analog Ac-Thr-Ile-Nle-(CH_2-NH)-Nle-Gln-Arg (MVT-101) was modified as described previously [19] by replacing the uncleavable CH_2-NH bond between Nle and Nle by a Met–Met peptide bond, yielding a natural p2/NC cleavage site for the HIV protease (Ac-Thr-Ile-Met-Met-Gln-Arg). The PDB entry used for HTLV protease was 2B7F [18], where HTLV-PR protease is in complex with the substrate analog Ac-Ala-Pro-Gln-Val-Sta-Val-Met-His-Pro, which is based on the natural MA/CA cleavage site of HTLV-I protease. The unliganded forms of both proteases were generated by removing the ligand present in the crystal structure following an approach already performed in previous studies [19–21]. Protonation of one active site aspartate in each protease was performed based on the calculation of the aspartates' protonation states using PROPKA [22]; this protonation state is also consistent with a previous

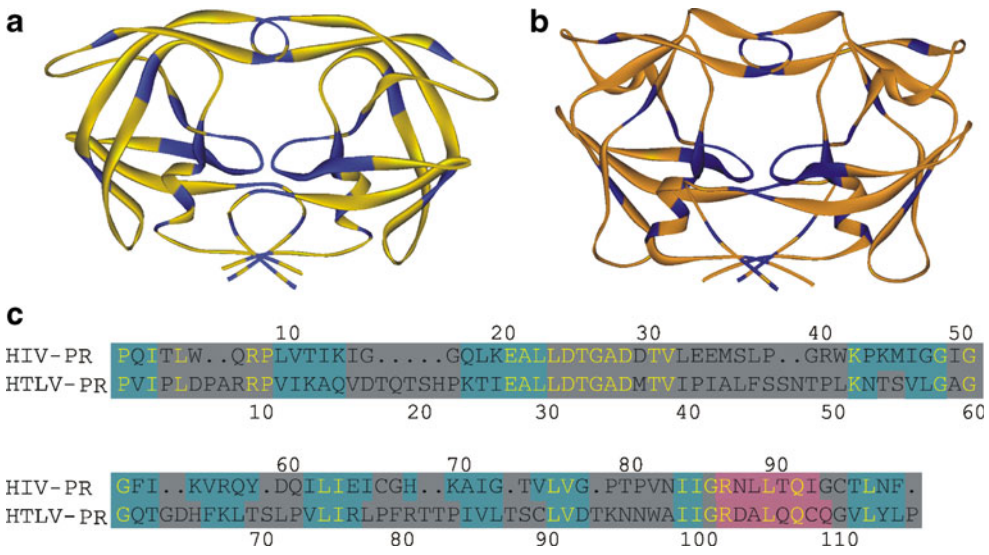

Fig. 1 Structures of HIV-PR (**a**) in *yellow* and HTLV-PR (**b**) in *orange*, both shown in ribbon representation. Positions of conserved residues are indicated by a *dark blue color*. **c** Structure-based sequence alignment of HIV-PR and HTLV-PR sequences. Conserved residues are shown in *yellow*; background color indicates secondary structure: β-sheets in *turquoise*, α-helix in *magenta*

experimental study [23]. All molecular dynamics (MD) simulations presented in this work were performed using AMBER 9 [24–26] with the parm99SB force field [27, 28] and the TIP3P water model [29]. For the γ-amino acid statine (Sta), the general AMBER force field (gaff) [30] was used. Missing parameters and partial atomic charges for statine were generated using a protocol also described in a previous work [19]. Simulations were performed in a periodic water box with at least 10 Å of solvent around every atom of the solute. An appropriate number of counterions was added to neutralize the charges of the systems, and the particle mesh Ewald summation method [31] was employed to calculate the long-range electrostatic interactions. All structures were minimized in a three-step procedure using the SANDER module of AMBER following a previously established protocol [19–21]. MD simulations were performed using the SHAKE procedure [32] to constrain all bonds involving hydrogen atoms. The integration time step of the simulation was 1 fs, and an 8.5 Å cutoff was used for the nonbonded interactions, which were updated every 15 steps. The temperature of each system was gradually heated to 298 K during the first 10 ps. Subsequently, 50 ns MD simulations were performed for data collection. The plot of the potential energy in the initial stages of the simulation is shown in Fig. S1 of the "Electronic supplementary material."

Correlation matrices of backbone C_α atoms were computed using the AMBER9 ptraj routine on 500 snapshots extracted every 100 ps from the trajectories. For visualization and for structural and energetic analysis of the trajectory data, the programs Sybyl 7.3 [33], DS ViewerPro Suite 6 [34], AMBER [26], and LIGPLOT [35] were used.

Results and discussion

Substrate-binding properties of HIV-PR and HTLV-PR

HIV-PR and HTLV-PR exhibit highly similar folds (Fig. 1), but nevertheless differ significantly in their substrate-binding properties [15, 36]. These differences are also reflected in the drug-binding profiles, as HTLV-PR cannot be inhibited by therapeutic doses of HIV-PR inhibitors such as saquinavir, indinavir, ritonavir, nelfinavir or amprenavir [15, 37].

In order to assess similarities and differences in protease–substrate interactions on the molecular level, we performed 50 ns all-atom molecular dynamics simulations in explicit solvent of HIV-1 and HTLV-I proteases in their free and substrate-bound states. From the simulation trajectories, we determined mean van der Waals interaction energies with the substrate analogs over the course of simulation for each residue of HIV-PR and HTLV-PR (Fig. 2). The main interactions occur in equivalent regions of the two proteases: The N-terminus (HIV-PR R8; HTLV-PR R10), the active site (HIV-PR L23–D30; HTLV-PR L30–D36), the flap region (HIV-PR M46–I50; HTLV-PR S55–A59), and the lateral loops of the ligand-binding site (HIV-PR P81–I84; HTLV-PR N97–I100). At a first glace, it might be surprising that the two catalytic aspartates (D32/D32′) of HTLV-PR exhibit slightly repulsive interactions with the substrate (Fig. 2). Inspection of the electrostatic interaction energy, however, reveals a large attractive electrostatic interaction of D32/D32′ with the substrate that overcompensates for the repulsive van der Waals interactions (Fig. S2).

The profile for the interaction energy (Table 1) also shows that the contributions of the residues at equivalent sequence positions are highly similar in HIV-PR and HTLV-PR. Interestingly, although several well-conserved amino acids have strong quaternary interactions with the substrates (R8/R10, A28/A35, D29/D36, G49, G58, and I84/I100), many of the amino acids at the strongest interacting sequence positions are not conserved between HIV-PR and HTLV-PR (see for example D30/M37, I47/V56, G48/L57, I50/A59, and V82/W98 in Table 1). These differences in amino acid sequence at identical positions of the ligand binding pocket allow the formation of tight interactions to different types of substrates, thereby explaining the differences in binding specificity.

Strong ligand interactions are also observed for corresponding nonconserved residues comprising the flap tips of HIV-PR and HTLV-PR. An experimental study showed that exchanging these residues in HTLV-PR with the HIV-PR sequence results in a dramatic reduction of the protease's capability to cleave the HTLV CA/NC cleavage site [38]. According to our data, the nonconserved residues HIV-PR D30/D30′ and HTLV-PR M37/M37′ also contribute strongly to the binding of the substrate analogs in both proteases. These amino acids have previously been described as possible determinants of substrate specificity [38, 39].

Our results concerning binding energies of HIV-PR and HTLV-PR with selected substrate analogs are in line with the findings of previous studies which have demonstrated that substrate specificities and inhibition profiles differ considerably between the two proteases [15, 36]. For example, it was shown earlier that a HIV p2/NC peptide, which is the basis for our model HIV-PR substrate, could not be cleaved by HTLV-PR [40].

Despite the differences described above, the average sequence identity of the substrate binding pocket (~50%) is still higher than the overall sequence identity (31%). This is mainly due to the fact that the floor of the active site (HIV-PR: L23–D29; HTLV-PR: L30–D36) is entirely conserved to ensure catalytic activity (Fig. 1).

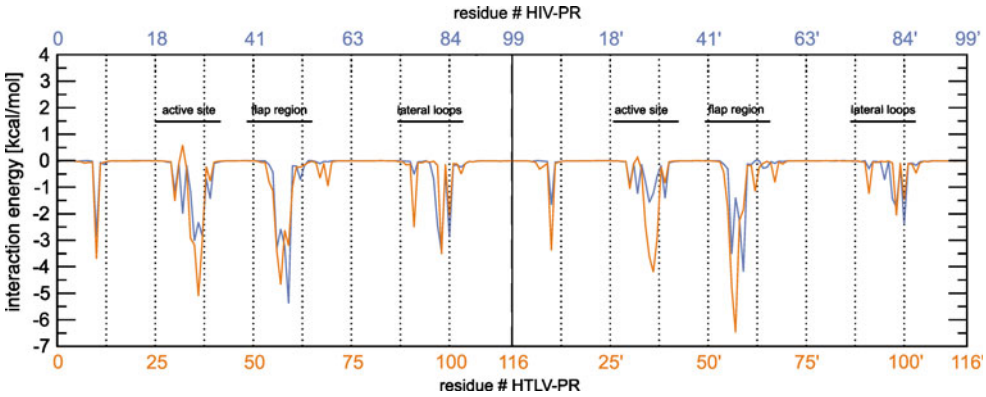

Fig. 2 Mean van der Waals interaction energies of each protease residue with the respective model substrate measured over 50 ns of simulation time. *Blue curve* HIV-PR, *orange curve* HTLV-PR. To allow a direct comparison of equivalent sequence positions, the sequences of HIV-PR and HTLV-PR have been adjusted according to the structure-based sequence alignment in Fig. 1c

In summary, we found that interactions occur at the same spatial sites and with similar binding energies (Fig. 2), but are frequently formed by different types of amino acids (Table 1). The latter observation offers an explanation for substrate specificity. Due to different side-chain properties, substrate binding can be modulated by providing a unique steric and physicochemical microenvironment, thereby determining ligand binding specificity.

Next, we addressed the question of whether the dynamic behavior is preserved between the two enzymes, since dynamic properties are likely to influence substrate binding and catalytic activity.

Dynamics of the protease flaps

With respect to protease dynamics, the flaps represent functionally important segments that allow substrate access to the active site and interact with the ligand [41]. In the unliganded HIV protease, the flaps are rather flexible [42, 43] and have been shown to open to a semi-open conformation [44, 45]. When a ligand is bound, however, the flaps of the bound protease close [41, 46] to allow for substrate cleavage. Therefore, the dynamics of the flaps play a critical role in the proteases' ability to bind and process their substrates.

Table 1 Mean van der Waals interaction energies (kcal mol^{-1}) between residues of HIV-PR /HTVLP with the substrate over the course of a 50 ns MD simulation

Subunit 1				Subunit 2			
HIV-PR		HTLV-PR		HIV-PR		HTLV-PR	
Residue	Interaction energy (kcal/mol)	Residue	Interaction energy (kcal/mol)	Residue	Interaction energy (kcal/mol)	Residue	Interaction energy (kcal/mol)
R8	−2.84	R10	−3.70	R8′	−1.67	R10′	−3.38
L23	−1.16	L30	−1.52	L23′	−0.93	L30′	−1.06
L24	−0.10	L31	−0.12	L24′	−0.10	L31′	−0.11
D25	−1.99	D32	0.59	D25′	−1.24	D32′	0.13
T26	−0.18	T33	−0.32	T26′	−0.14	T33′	−0.33
G27	−1.17	G34	−2.93	G27′	−0.80	G34′	−2.45
A28	−3.02	A35	−3.17	A28′	−1.57	A35′	−3.61
D29	−2.32	D36	−5.11	D29′	−1.23	D36′	−4,20
D30	−2.84	M37	−3.03	D30′	−0.59	M37′	−2.92
M46	−0.43	S55	−1.13	M46	−0.31	S55	−1.81
I47	−3.31	V56	−3.39	I47′	−3.51	V56′	−4.79
G48	−2.58	L57	−4.67	G48′	−1.39	L57′	−6.47
G49	−3.23	G58	−2.62	G49′	−2.28	G58′	−2.23
I50	−5.38	A59	−3.21	I50′	−4.19	A59′	−1.84
T80	−0.69	N96	−0.02	T80′	−0.71	N96′	−0.01
P81	−2.44	N97	−0.17	P81′	−1.44	N97′	−0.09
V82	−3.43	W98	−3.52	V82′	−1.76	W98′	−2.06
I84	−2.88	I100	−2.09	I84′	−2.41	I100′	−1.48

In our study we analyzed the dynamics of the protease flaps by measuring the flap–Asp distance [45] over the course of simulation to characterize the opening and closing motions of the flaps. To this end, we calculated the interatomic distances between the C_β atom of the catalytic aspartate of each subunit (D25/D25′ in HIV-PR and D32/D32′ in HTLV-PR) and the C_α atom of the flap tip residue in the respective subunit (I50/I50′ in HIV-PR and A59/A59′ in HTLV-PR) for the free and ligand-bound states of both proteases (Fig. 3a-f).

The flaps of both free proteases are rather flexible. Reversible opening motions to distances of about 20 Å are observed in the HIV-PR subunit 2 after about 15 ns and in the HTLV-PR subunit 1 after about 35 ns (Fig. 3c, d). Flap distances of 15–20 Å that are detected over the simulation time indicate that a semi-open conformation [45, 47] is temporarily adopted in both enzymes. This behavior of flaps is well described from simulations of ligand-free proteases [21, 44, 45] and NMR studies [48, 49]. Maximum distances of 20–22 Å in both proteases indicate that the flap does not open entirely. For subunits 1 and 2 of HIV-PR, distance values fluctuate around means of 14.7±1.0 Å and 13.7±1.7 Å, respectively (distance values in the starting structure are 12.8 Å and 13.0 Å, respectively). The mean distances are slightly higher in HTLV-PR: 17.0±1.9 Å and 15.5±1.7 Å for subunits 1 and 2, respectively (distances in the starting structure: 13.1 Å and 12.7 Å). The standard deviations of flap distances in HTLV-PR are in a similar range to those of HIV-PR, indicating a similar degree of flexibility. Viewed over the entire simulation and for both subunits, the flap dynamics of the unliganded proteases are surprisingly similar with respect to the degree and frequency of fluctuation.

For the substrate-bound structures, the similarities in the dynamic behavior of the flaps are even more pronounced (Fig. 3e, f). Although the mean distances are smaller in HTLV-PR (subunit 1: 9.9 Å; subunit 2: 11.4 Å) than in HIV-PR (subunit 1: 13.5 Å; subunit 2: 14.1 Å), the distance curves run almost parallel. The distance fluctuations are generally much smaller than in the ligand-free state, reflecting the proteases' decreased flexibility due to the stabilization of the flaps by interactions with the substrate ligand.

The high similarities between the dynamic behaviors of the flap structures, especially in the ligand-free state, show that the dynamics are largely conserved, despite the low overall sequence identity of 31%. In the flaps (HIV-PR 40–60

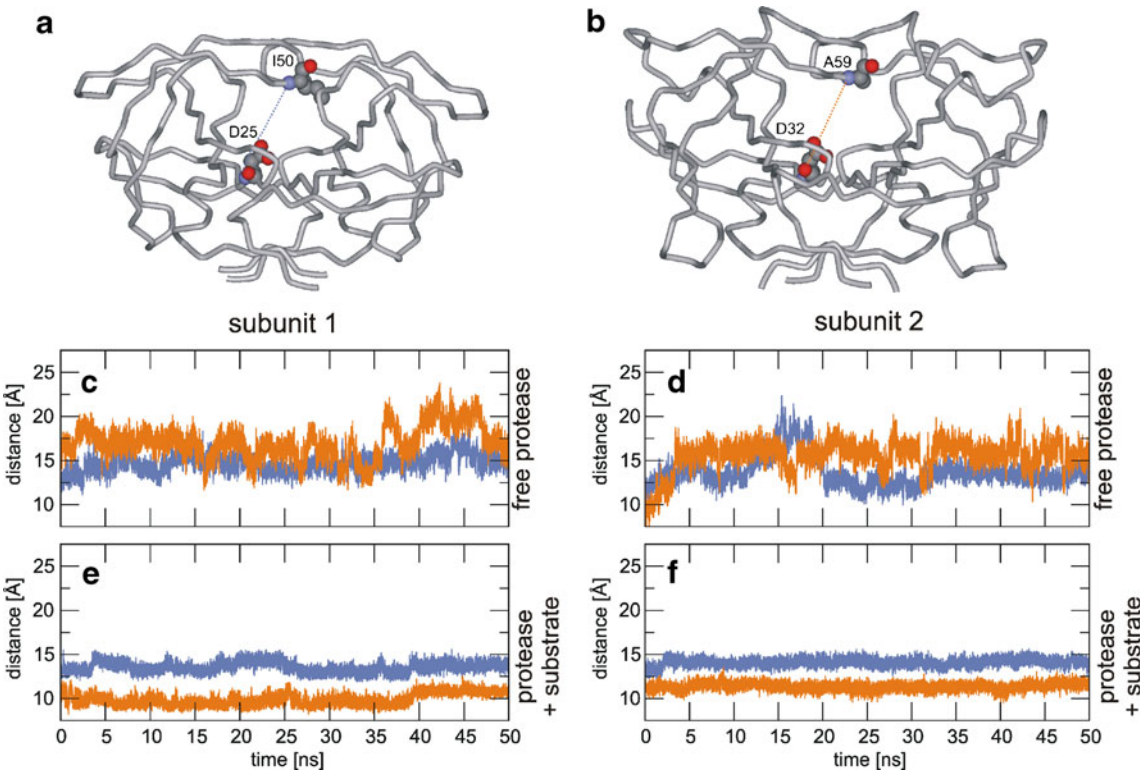

Fig. 3 Schematic backbone representations of HIV-PR (**a**) and HTLV-PR (**b**), with the residues D25 and I50 and D32 and A59, respectively, shown in *spacefill* and *standard CPK coloring*. The flap–Asp distance (drawn here only for subunit 1 in each protease) is indicated by *dotted lines* in both structures using the same colors as in graphs **c**–**f**. Interatomic distances between the C_β atom of the catalytic aspartate of each subunit (HIV-PR D25/D25′, HTLV-PR D32/D32′) and the C_α atom of the flap tip residue in the respective subunit (HIV-PR I50/I50′, HTLV-PR A59/A59′) shown over the course of simulation (**c**–**f**). **c**, **d** Unliganded proteases; **e**, **f** proteases in complex with the model substrates; *blue curve* HIV-PR; *orange curve* HTLV-PR

and HTLV-PR 49–72) alone, the sequence identity is even lower, with only four conserved residues: HIV-PR K43 (HTLV-PR K52) and the three glycines (HIV-PR G49, G51, G52; HTLV-PR G58, G60, G61) of the flap tips. Moreover, there is a two amino acid insertion (G64, D65) in the HTLV-PR flap, followed by a single serine insertion at position 71. These insertions also result in structural differences between the flaps as they result in protruding "horns" in the HTLV-PR flaps at positions 64–66. This divergence in sequence and structure between HIV-PR and HTLV-PR in the flaps makes the high degree of resemblance in the flap dynamics even more surprising.

The large similarities detected between HIV-PR and HTLV-PR are particularly interesting in light of the fact that single amino acid exchanges have been shown to substantially alter the flap flexibility of HIV-PR [21, 45, 50–53].

Although flap flexibility is a functionally important dynamic feature of retroviral proteases, it represents rather a local measure of dynamics. In order to investigate whether the global dynamics are conserved in addition to the flap dynamics, we studied the correlated motions of both enzymes.

Correlated motions

For HIV-PR and HTLV-PR, the correlated motions are shown in Fig. 4b and are compared to the pattern of interactions observed in the static crystal structure (Fig. 4a). From the distance matrix (Fig. 4a), the close contacts within the short C-terminal helix (HIV-PR N88–I93, HTLV-PR D103–Q110) and six antiparallel β-strands (compare the secondary structure in Fig. 1c) can be seen as red areas

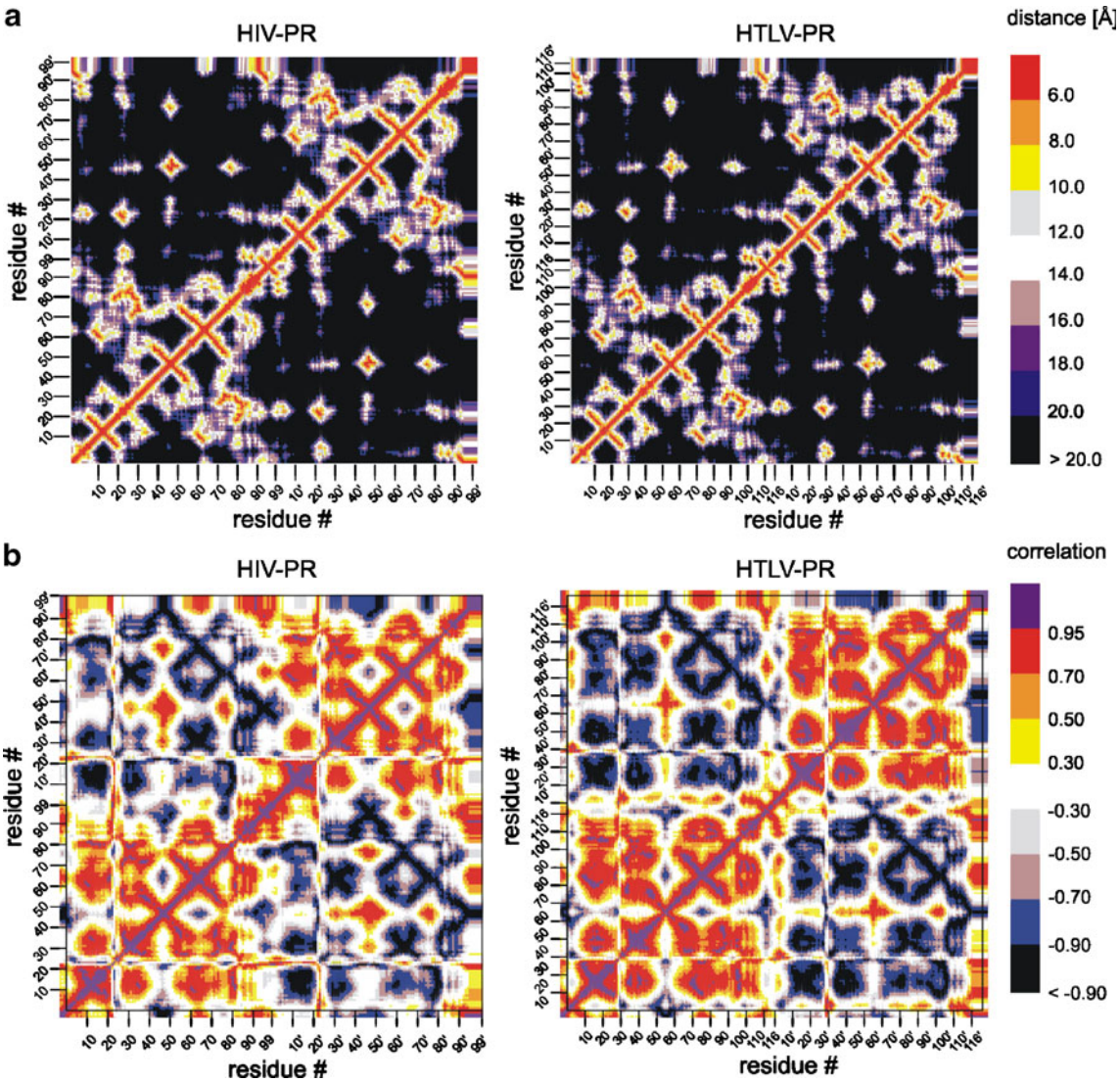

Fig. 4 **a** Distance matrices of C_α distances between all residues in the starting structures of HIV-PR (*left*) and HTLV-PR (*right*). **b** Correlation matrices showing correlated motions between C_α atoms of all residues within free HIV-PR (*left*) and within free HTLV-PR (*right*)

running parallel or orthogonal to the diagonal, respectively. Additional tight contacts are formed outside these elements of secondary structure, predominantly between the active site (HIV-PR residues 21–32, HTLV-PR residues 28–39) and a segment containing the lateral loops of the binding pocket (HIV-PR residues 76–90, HTLV-PR residues 91–106) (Fig. 4a).

Analysis of the correlated motions reveals that the strongest correlation occurs between amino acids that are adjacent in the 3D structure (Fig. 4b). We observe a strong correlation within the individual subunits in both enzymes, while motions between the subunits are anti-correlated. Interestingly, the C-termini of subunit 2 in both proteins, however, act rather as if they were dynamically detached from subunit 2 by showing strong correlations with the N-terminus, the flap region, and the lateral binding pocket loop of subunit 1. For the C-termini of subunit 1, this correlation is not observed. Correlations between the subunits are otherwise confined to the flap tips (HIV-PR G49–G51; HTLV-PR G58–G60) correlating with each other and with parts of the lateral loop of the other subunit (HIV-PR: 79–81; HTLV-PR 95–97), respectively. We find anti-correlation of the flaps (HIV-PR: 45–50; HTLV-PR: 53–58) with the lower "cheek" regions (HIV-PR: 62–68; HTLV-PR: 73–81) of the same subunit in both proteases. This anti-correlation is more pronounced in subunit 1 in both enzymes (Fig. 4b). Since the lower "cheek" region is the extension of the upper flap strand, it appears plausible that this loop performs a downward movement when the flap tips move upward during flap opening motions. Therefore, this finding supports the suggestions made in an earlier study [45]. In summary, the pattern and degree of correlated motions are highly similar between HIV-PR and HTLV-PR for the entire peptide chain.

The striking similarities of the flap motions and overall correlated motions indicate that the dynamics of this class of enzymes are in principle quite well conserved despite the low primary sequence identity of only 31% between HIV-PR and HTLV-PR. In this context, it is tempting to speculate about a common conserved mechanism that assures the proper structure and motions necessary for ligand binding and enzymatic function in both proteases, despite their low primary sequence identity. The residue positions that are invariant between the HIV-PR and HTLV-PR are particularly interesting in this context.

Conserved residues and interactions

In an attempt to identify conserved features in HIV-PR and HTLV-PR, that may be responsible for their high similarities in terms of structure and dynamics, we took a more detailed look at the 31 residues that are identical between the two proteases (Fig. 1). Evaluation of the noncovalent interactions formed by conserved residues reveals that their interaction patterns are highly similar in HIV-PR and HTLV-PR. An overlay of the inter-residue distance matrix and the interactions of conserved residues shows that not only the interactions between sequentially neighboring residues are very much alike, but the interactions of sequentially distant residues are also highly conserved (Fig. 5).

When we investigated the conserved residues' interactions in the crystal structures, we found the great majority of conserved amino acids to be localized within corresponding clusters of interactions in the protease

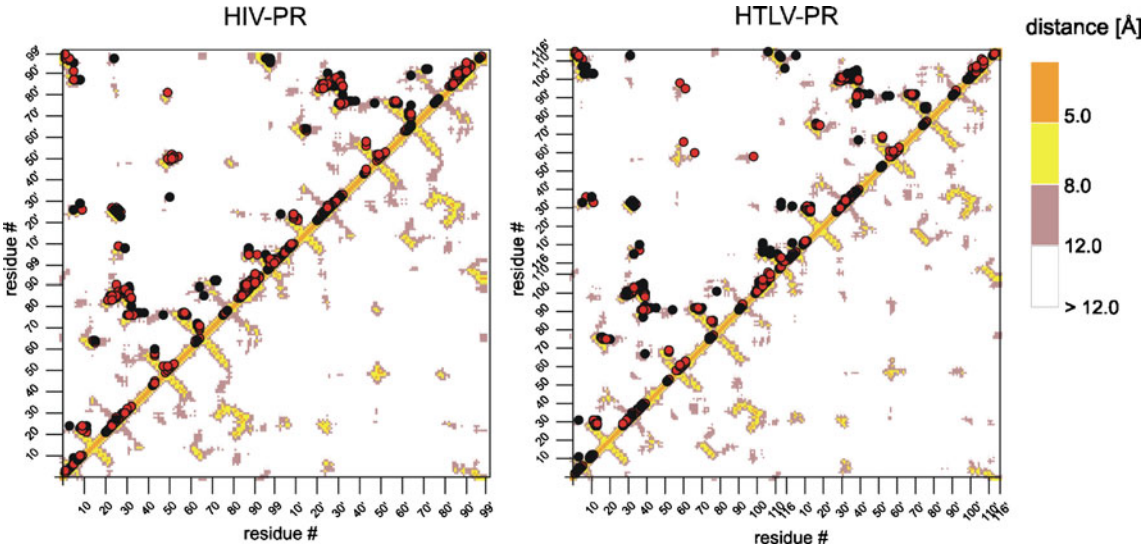

Fig. 5 Overlay of C_α distance matrices of all residues within HIV-PR (*left*) and within HTLV-PR (*right*) with interactions of conserved residues, shown for only one subunit for the sake of clarity (*upper left half of graph*). *Black and red dots* denote backbone and sidechain interactions between conserved residues, respectively

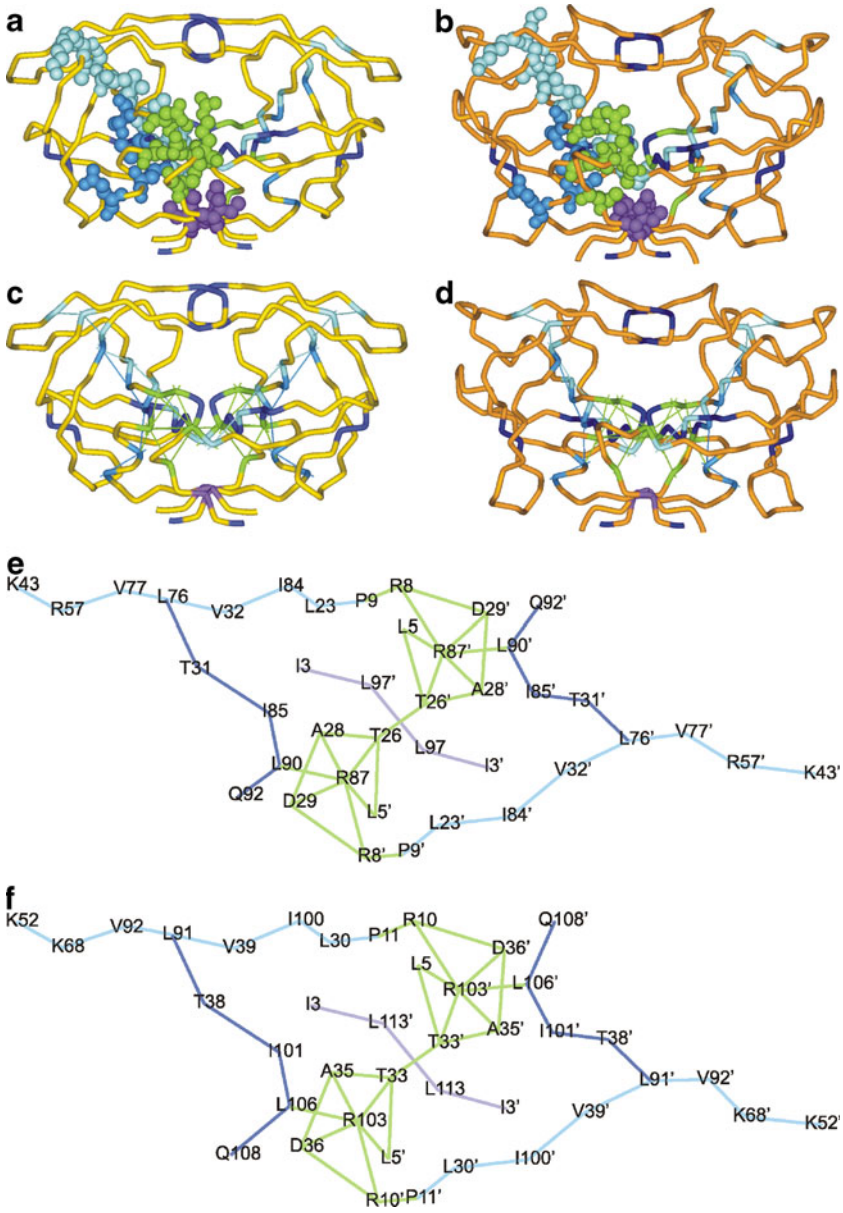

Fig. 6 Tube representations of HIV-PR (**a**) and HTLV-PR (**b**). Residues of cluster I are shown in *pale blue*, cluster II in *blue*, cluster III in *purple*, and cluster IV in *green*. Conserved residues that were not classified in one of the clusters are shown in *dark blue*. Residues of clusters I–IV are shown in *spacefill representation* for one subunit in both proteases. Tube representations of HIV-PR (**c**) and HTLV-PR (**d**): noncovalent interactions within the cluster network are indicated as *rods*, color code as in **a** and **b**. Schematic 2D projection of the interaction network of HIV-PR (**e**) and HTLV-PR (**f**); color code as in **a**–**d**

structures (Fig. 6). By evaluating the interactions between conserved residues and their localization in the 3D protease structures, we could identify four primary interaction clusters I–IV. Three of them (I–III) form almost linear axes, and one (IV) forms a central knot. The axis of cluster I (HIV-PR: K43-R57-V77-L76-V32-I84-L23-P9, HTLV-PR: K52-K68-V92-L91-V39-I100-L30-P11) runs from the lateral hinge parts of the flap β-sheet (the protease "ears")

Table 2 Interaction clusters of conserved contacts and involved residues in HIV-PR (*left*) and HTLV-PR (*right*). Sites of mutations in HIV-PR are indicated by *asterisks* (*), and sites of polymorphisms by a *superscript P* (^{P})

Cluster	Involved residues in HIV-PR	Involved residues in HTLV-PR
Cluster I	K43*, R57*, V77^{P}, L76*, V32*, I84*, L23*, P9	K52, K68, V92, L91, V39, I100, L30, P11
Cluster II	L76*, T31, I85*, L90*, Q92*	L91, T38, I101, L106, Q108
Cluster III	I3, L97′, L97, I3′	I3, L113′, L113, I3′
Cluster IV	R8*, D29′, A28′*, R87, L5, T26′, T26, L5′, R87, A28*, D29, R8′*	R10, D36′, A35′, R103, L5, T33′, T33, L5′, R103, A35, D36, R10′

all the way down to the dimerization interface, crossing four β-strands, the 80s loop and the ligand-binding site. Cluster II (HIV-PR: L76-T31-I85-L90-Q92, HTLV-PR: L91-T38-I101-L106-Q108) branches from cluster I and takes a slightly different direction but also leads to the dimerization interface. Cluster III (HIV-PR: I3-L97′-L97-I3′, HTLV-PR: I3-L113′-L113-I3′) interconnects the N- and C-termini of the two subunits, serving as a clamp that fixates the flexible peptide ends. The central knot is formed by cluster IV (HIV-PR: R8-D29′-A28′-R87-L5-T26′-T26-L5′-R87-A28-D29-R8′, HTLV-PR: R10-D36′-A35′-R103-L5-T33′-T33-L5′-R103-A35-D36-R10′), which interconnects the two subunits and stabilizes the loops containing the catalytic aspartates (HIV-PR: D25, D25′; HTLV-PR: D32, D32′). The location of the four clusters within the protease structure and the complete list of interacting residues are given in Fig. 6 and Table 2, respectively.

All of the abovementioned residues are strictly conserved. The only exception is a conservative replacement of R57 in HIV-PR by K68 in HTLV-PR. For this sequence position a R57K mutation is observed in several HIV subtypes [54], rendering the abovedescribed interaction clusters fully identical in sequence between HIV-PR and HTLV-PR. This strict conservation might also explain why approaches that focus on co-evolution of amino acids did not identify these clusters.

Only a minority of the conserved residues are not included in these interaction clusters. These are, for example, residues adjacent to the catalytic site that directly play a role in protease–ligand interactions and enzymatic function, as well as the glycines of the flap tips that are essential for flap flexibility.

A graphic presentation of the conserved interactions as rods (Fig. 6c and d) reveals that the interaction clusters, among which I, II and IV are connected and form a basket-like structure, run orthogonal to the antiparallel β-sheets, thereby interconnecting secondary structure elements. This observation implies that the interaction network serves as a scaffold that stabilizes the tertiary structure of the proteases and ensures proper folding and dynamic properties.

To address this hypothesis, we studied the interactions within the network over time. Of the 44 interactions that were analyzed, 27 proved to be stable over 90% of the simulation time in both proteases (Table 3), while three additional interactions were stable for more than 80% of the simulation. With the interaction network being conformationally stable to such an extent, we hypothesized that it may represent a stable scaffold for the protease structure and dynamics.

When we calculated the root mean square fluctuations (RMSFs) of the C_α atoms of all residues in HIV-PR and HTLV-PR (Fig. 7), we found that the RMSF values for the conserved residues in the interaction network are particularly low compared to other residues. Most of them have

Table 3 Presence of conserved noncovalent contacts in the interaction clusters over the course of a 50 ns MD simulation in percent

HIV-PR		HTLV-PR	
Contacts	Presence (%)	Contacts	Presence (%)
L76–V32	100	L91–V39	99.8
V32–I84	100	V39–I100	100
I84–L23	99.6	I100–L30	100
L23–P9	98.6	L30–P11	97.8
L76–T31	100	L91–T38	100
T31–I85	88.2	T38–I101	45.6
I85–L90	77.8	I101–L106	42.0
L90–Q92	100	L106–Q108	100
L90–R87	95.6	L106–R103	92.4
R87–A28	100	R103–A35	100
R87–D29	100	R103–D36	99.6
A28–T26	100	A35–T33	100
T26–T26′	99.8	T33–T33′	100
T26–L5′	41.8	T33–L5′	51.8
T26–R87	76.0	T33–R103	100
R87–L5′	100	R103–L5′	32.6
R87–R8′	96.2	R103–R10′	78.0
D29–R8′	100	D36–R10′	83.0
I3–L97	100	I3–L113′	100
K43–R57	96.0	K52–K68	98.2
V77–R57	100	V92–K68	100
L76–R57	99.6	L91–K68	98.6
I3′–L97	100	I3′–L113	100
L97–L97′	24.4	L113–L113′	18.2
K43′–R57′	89.4	K52′–K68′	98.8
V77′–R57′	100	V92′–K68′	100
L76′–R57′	98.6	L91′–K68′	99.4
L76′–V32′	99.2	L91′–V39′	100
V32′–I84′	100	V39′–I100′	100
I84′–L23′	99.4	I100′–L30′	97.4
L23′–P9′	97.4	L30′–P11′	98.8
V76′–T31′	100	L91′–T38′	97.0
T31′–I85′	75.6	T38′–I101′	66.2
I85′–L90′	49.8	I101′–L106′	31.6
L90′–Q92′	100	L106′–Q108′	100
L90′–R87′	91.4	L106′–R103′	89.0
R87′–A28′	100	R103′–A35′	100
R87′–D29′	97.2	R103′–D36′	97.4
A28′–T26′	96.0	A35′–T33′	99.2
T26′–L5	49.4	T33′–L5′	60.8
R87′–T26′	70.2	R103′–T33′	100
R87′–L5	99.0	R103′–L5	7.2
R87′–R8	93.6	R103′–R10	67.7
D29′–R8	38.6	D36′–R10	29.2

RMSF values that are below 1 Å and are found in regions with the lowest RMSF values. Even HIV-PR K43/K43′ and the corresponding HTLV-PR K52/K52′, which are located in the most distal position of cluster I in the flexible protease "ears," show considerably lower RMSF values than the sequentially adjacent residues. Therefore, they can be considered anchoring residues for the most lateral β-strand. These findings support the notion that the wide-stretched basket-like network represents a rather rigid connection of elements of secondary structure, thereby stabilizing the tertiary structure and determining dynamic behavior. This finding is also interesting in the light of a recent study [55] which revealed that the sequence stretches W42–P44 and K55–R57 in HIV-PR form the binding site for the allosteric 1F1 (indole-6-carboxylic acid) ligand. These two sequence stretches contain the conserved basic residues K43 and R57 of cluster I and exhibit a particularly low RMSD compared to the adjacent regions (Fig. 7). This observation might explain why the respective surface region acts as a binding site for the allosteric 1F1 fragment.

For many of the residues involved in the interaction network, mutations or polymorphisms have been described in HIV-PR (see Table 2), some of which have been linked to inhibitor resistance [54, 56–59]. There have been previous attempts to elucidate the drug resistance mechanisms of mutations that are not directly located in the active site or interact with inhibitors. In earlier studies, L90M substitution, for example, was shown to alter protease dynamics [60, 61], and like R8Q [58], to decrease protease stability. R8 has also previously been recognized as a key residue for mediating the binding specificity of HIV-1 protease [62]. Mutations of the amino acids L24, L90 and L97 have been linked to a negative impact on the enzyme's catalytic activity, although they are situated outside of the catalytic center, or their sidechains are oriented away from the ligand [57].

The observation that these hydrophobic residues are part of a conserved interaction network that runs orthogonal to the β-sheets suggests that these residues also play a role in stabilizing the proper register of the hydrogen bonds in the β-sheets, thereby preventing "hydrophobic sliding," which was suggested to be an important mechanism in mediating drug resistance of HIV-PR [63].

Together with all these findings, our results further support the hypothesis that local changes in the protease may allosterically influence the entire protein architecture and dynamic properties if they are located in crucial positions within stabilizing interaction networks. Local distortions of interaction patterns may therefore affect inhibitor binding and substrate processing not only by direct influence but also via long-range or global conformational effects.

Conclusions

In summary, our study provides the first comprehensive comparison of the sequences, structures, and dynamics of

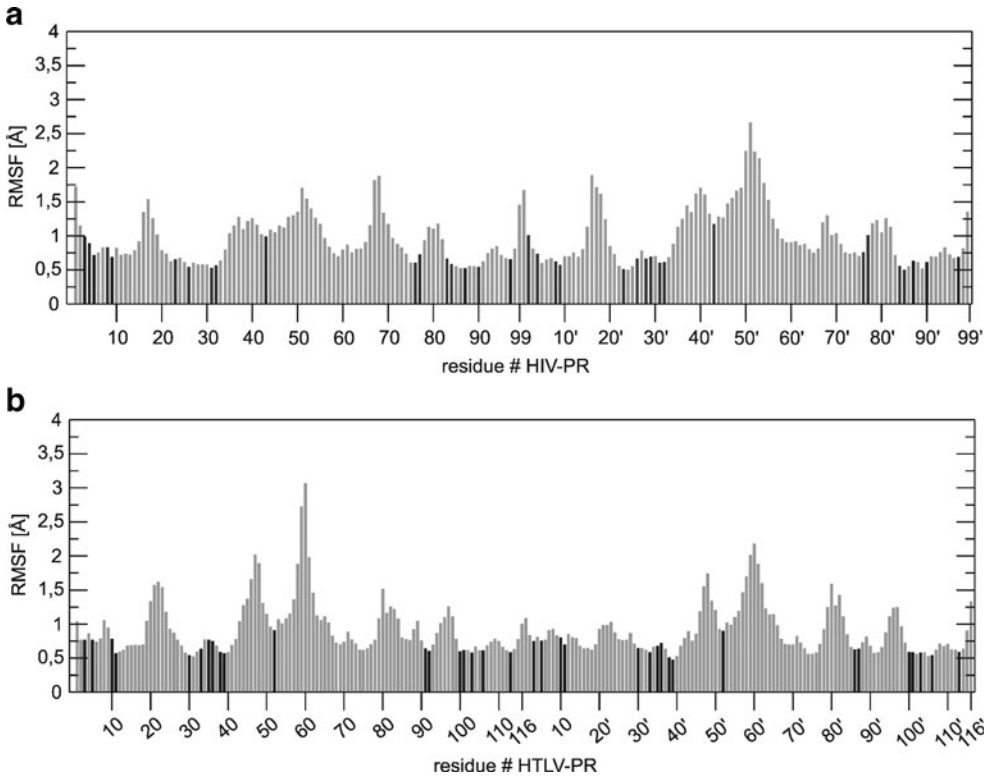

Fig. 7 Root mean square fluctuation (RMSF) over 50 ns of simulation of C_α atoms for all residues of HIV-PR (**a**) and HTLV-PR (**b**). Values for conserved residues involved in the interaction network are shown as *black bars*; all others are represented by *gray bars*

the two retroviral proteases from HIV and HTLV. The local and overall dynamics of the two proteins proved to be remarkably similar in both their free and ligand-bound states, suggesting that this feature is also a prerequisite for proper enzyme functions. An energetic analysis revealed that the roles of the individual sequence positions for substrate binding are also very similar in both proteases. Interestingly, the residues that are important for strong ligand binding are frequently not conserved in sequence, thereby offering an explanation for the differences in the binding specificities of HIV-PR and HTLV-PR.

Moreover, our study revealed that the amino acids that are identical in the two proteases play a key role in establishing nonlocal contacts that stabilize the tertiary fold. This structurally conserved network of interactions is highly similar in both proteases and is also conformationally stable over the simulation time. The noncovalent contacts formed in this network provide a basket-shaped scaffold that stabilizes the three-dimensional structure of each subunit by linking individual secondary structure elements, and interconnects the subunits in the dimeric enzymes.

This finding sheds new light on the sequence–structure–dynamics interrelationship of retroviral proteases. A basic common principle of structure stabilization guaranteeing similar dynamic properties that permit physiological enzymatic function can apparently be realized even with a low degree of sequence identity. This finding also has implications for the causality of inhibitor resistance due to mutations of HIV protease residues that do not interact directly with the ligands, but rather influence inhibitor binding via altered dynamics. Further studies of HTLV-PR and HIV-PR that also take into account the fourth dimension of dynamics should therefore provide a promising basis for future drug development.

Acknowledgments The authors thank the Regionales Rechenzentrum Erlangen (RRZE) for technical support and the Deutsche Forschungsgemeinschaft (graduate program GRK1071) for funding.

References

1. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci USA 77:7415–7419
2. Yoshida M, Seiki M, Yamaguchi K, Takatsuki K (1984) Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. Proc Natl Acad Sci USA 81:2534–2537
3. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, de The G (1985) Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet 2(8452):407–410
4. LaGrenade L, Hanchard B, Fletcher V, Cranston B, Blattner W (1990) Infective dermatitis of Jamaican children: a marker for HTLV-I infection. Lancet 336(8727):1345–1347
5. Mochizuki M, Watanabe T, Yamaguchi K, Takatsuki K, Yoshimura K, Shirao M, Nakashima S, Mori S, Araki S, Miyata N (1992) HTLV-I uveitis: a distinct clinical entity caused by HTLV-I. Jpn J Cancer Res 83:236–239
6. Morgan OS, Rodgers-Johnson P, Mora C, Char G (1989) HTLV-1 and polymyositis in Jamaica. Lancet 2(8673):1184–1187
7. Nishioka K, Maruyama I, Sato K, Kitajima I, Nakajima Y, Osame M (1989) Chronic inflammatory arthropathy associated with HTLV-I. Lancet 1(8635):441–441
8. de The G, Bomford R (1993) An HTLV-I vaccine: why, how, for whom? AIDS Res Hum Retroviruses 9:381–386
9. Gessain A (1996) Virological aspects of tropical spastic paraparesis/HTLV-I associated myelopathy and HTLV-I infection. J Neurovirol 2:299–306
10. Zucker-Franklin D, Pancake BA (1998) Human T-cell lymphotropic virus type 1 tax among American blood donors. Clin Diagn Lab Immunol 5:831–835
11. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220:868–871
12. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Bea S (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science 224:500–503
13. Kohl NE, Emmi EA, Schleif WA, Davies LJ, Heimbach JC, Dixon RAF (1988) Active human immunodeciency virus protease is required for viral infectivity. Proc Natl Acad Sci 85:686–690
14. Ishitsuka K, Tamura K (2008) Treatment of adult T-cell leukemia/lymphoma: past, present, and future. Eur J Haematol 80:185–196
15. Louis JM, Oroszlan S, Tozser J (1999) Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. J Biol Chem 274:6660–6666
16. Li M, Laco GS, Jaskolski M, Rozycki J, Alexandratos J, Wlodawer A, Gustchina A (2005) Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. Proc Natl Acad Sci USA 102:18332–18337
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
18. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246:1149–1152
19. Wartha F, Horn AHC, Meiselbach H, Sticht H (2005) Molecular dynamics simulations of HIV-1 protease suggest different mechanisms contributing to drug resistance. J Chem Theory Comput 1:315–324
20. Dirauf P, Meiselbach H, Sticht H (2010) Effects of the V82A and I54V mutations on the dynamics and ligand binding properties of HIV-1 protease. J Mol Model 16:1577–1583
21. Meiselbach H, Horn AH, Harrer T, Sticht H (2007) Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. J Mol Model 13:297–304
22. Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pK_a values for protein-ligand complexes. Proteins 73:765–783
23. Smith R, Brereton IM, Chai RY, Kent SB (1996) Ionization states of the catalytic residues in HIV-1 protease. Nat Struct Biol 3:946–950

24. Case DA, Cheatham T, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods R (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
25. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, KMerz M, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco
26. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE 3rd, DeBolt S, Ferguson D, Seibel G, Kollman P (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comp Phys Commun 91:1–41
27. Cheatham TE 3rd, Cieplak P, Kollman PA (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J Biomol Struct Dyn 16:845–862
28. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KMJ, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulaton of proteins, nucleic acids and organic molecules. J Am Chem Soc 117:5179–5197
29. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
30. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174
31. Darden TA, York DM, Pedersen LG (1993) Particle mesh Ewald. An N.log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
32. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
33. Tripos (1991–2008) Sybyl 7.3. Tripos, St. Louis
34. Accelrys (2005) DS ViewerPro Suite 6.0. Accelrys, San Diego
35. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. Protein Eng 8:127–134
36. Luukkonen BG, Tan W, Fenyo EM, Schwartz S (1995) Analysis of cross reactivity of retrovirus proteases using a vaccinia virus-T7 RNA polymerase-based expression system. J Gen Virol 76:2169–2180
37. Bagossi P, Kadas J, Miklossy G, Boross P, Weber IT, Tozser J (2004) Development of a microtiter plate fluorescent assay for inhibition studies on the HTLV-1 and HIV-1 proteinases. J Virol Methods 119:87–93
38. Kadas J, Weber IT, Bagossi P, Miklossy G, Boross P, Oroszlan S, Tozser J (2004) Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell leukemia virus type 1 protease. J Biol Chem 279:27148–27157
39. Tozser J, Zahuczky G, Bagossi P, Louis JM, Copeland TD, Oroszlan S, Harrison RW, Weber IT (2000) Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur J Biochem 267:6287–6295
40. Tozser J, Gustchina A, Weber IT, Blaha I, Wondrak EM, Oroszlan S (1991) Studies on the role of the S4 substrate binding site of HIV proteinases. FEBS Lett 279:356–360
41. Gustchina A, Weber IT (1990) Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. FEBS Lett 269:269–272
42. Furfine ES, D'Souza E, Ingold KJ, Leban JJ, Spector T, Porter DJ (1992) Two-step binding mechanism for HIV protease inhibitors. Biochemistry 31:7886–7891
43. Rodriguez EJ, Debouck C, Deckman IC, Abu-Soud H, Raushel FM, Meek TD (1993) Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry. Biochemistry 32:3557–3563
44. Hornak V, Okur A, Rizzo RC, Simmerling C (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proc Natl Acad Sci USA 103:915–920
45. Perryman AL, Lin JH, McCammon JA (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci 13:1108–1123
46. Toth G, Borics A (2006) Closing of the flaps of HIV-1 protease induced by substrate binding: a model of a flap closing mechanism in retroviral aspartic proteases. Biochemistry 45:6606–6614
47. Spinelli S, Liu QZ, Alzari PM, Hirel PH, Poljak RJ (1991) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie 73:1391–1396
48. Freedberg DI, Ishima R, Jacob J, Wang YX, Kustanovich I, Louis JM, Torchia DA (2002) Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. Protein Sci 11:221–232
49. Ishima R, Freedberg DI, Wang YX, Louis JM, Torchia DA (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. Structure 7:1047–1055
50. Galiano L, Ding F, Veloro AM, Blackburn ME, Simmerling C, Fanucci GE (2009) Drug pressure selected mutations in HIV-1 protease alter flap conformations. J Am Chem Soc 131:430–431
51. Heaslet H, Rosenfeld R, Giffin M, Lin YC, Tam K, Torbett BE, Elder JH, McRee DE, Stout CD (2007) Conformational flexibility in the flap domains of ligand-free HIV protease. Acta Crystallogr D Biol Crystallogr 63:866–875
52. Kear JL, Blackburn ME, Veloro AM, Dunn BM, Fanucci GE (2009) Subtype polymorphisms among HIV-1 protease variants confer altered flap conformations and flexibility. J Am Chem Soc 131:14650–14651
53. Seibold SA, Cukier RI (2007) A molecular dynamics study comparing a wild-type with a multiple drug resistant HIV protease: differences in flap and aspartate 25 cavity dimensions. Proteins 69:551–565
54. HIVDB Team (1998–2008) Stanford HIV Database. http://hivdb.stanford.edu/
55. Perryman AL, Zhang Q, Soutter HH, Rosenfeld R, McRee DE, Olson AJ, Elder JE, David Stout C (2010) Fragment-based screen against HIV protease. Chem Biol Drug Des 75:257–268
56. Alcaro S, Artese A, Ceccherini-Silberstein F, Ortuso F, Perno CF, Sing T, Svicher V (2009) Molecular dynamics and free energy studies on the wild-type and mutated HIV-1 protease complexed with four approved drugs: mechanism of binding and drug resistance. J Chem Inf Model 49:1751–1761
57. Kontijevskis A, Prusis P, Petrovska R, Yahorava S, Mutulis F, Mutule I, Komorowski J, Wikberg JE (2007) A look inside HIV resistance through retroviral protease interaction maps. PLoS Comput Biol 3:424–435
58. Mahalingam B, Louis JM, Reed CC, Adomat JM, Krouse J, Wang YF, Harrison RW, Weber IT (1999) Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. Eur J Biochem 263:238–245
59. Weber IT, Harrison RW (1999) Molecular mechanics analysis of drug-resistant mutants of HIV protease. Protein Eng 12:469–474

60. Kozisek M, Bray J, Rezacova P, Saskova K, Brynda J, Pokorna J, Mammano F, Rulisek L, Konvalinka J (2007) Molecular analysis of the HIV-1 resistance development: enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants. J Mol Biol 374:1005–1016
61. Ode H, Neya S, Hata M, Sugiura W, Hoshino T (2006) Computational simulations of HIV-1 proteases–multi-drug resistance due to nonactive site mutation L90M. J Am Chem Soc 128:7887–7895
62. Sherman W, Tidor B (2008) Novel method for probing the specificity binding profile of ligands: applications to HIV protease. Chem Biol Drug Des 71:387–407
63. Foulkes-Murzycki JE, Scott WR, Schiffer CA (2007) Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. Structure 15:225–233