J Mol Model (2011) 17:391–399
DOI 10.1007/s00894-010-0739-z

ORIGINAL PAPER

Identification of folding preferences of cleavage junctions of HIV-1 precursor proteins for regulation of cleavability

Hirotaka Ode · Masaru Yokoyama · Tadahito Kanda · Hironori Sato

Received: 10 March 2010 / Accepted: 30 April 2010 / Published online: 18 May 2010

Abstract Human immunodeficiency virus type 1 protease (HIV-1 PR) cleaves two viral precursor proteins, Gag and Gag-Pol, at multiple sites. Although the processing proceeds in the rank order to assure effective viral replication, the molecular mechanisms by which the order is regulated are not fully understood. In this study, we used bioinformatics approaches to examine whether the folding preferences of the cleavage junctions influence their cleavabilities by HIV-1 PR. The folding of the eight-amino-acid peptides corresponding to the seven cleavage junctions of the HIV-1_{HXB2} Gag and Gag-Pol precursors were simulated in the PR-free and PR-bound states with molecular dynamics and homology modeling methods, and the relationships between the folding parameters and the reported kinetic parameters of the HIV-1_{HXB2} peptides were analyzed. We found that a folding preference for forming a dihedral angle of Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') in the range of 150 to 180 degrees in the PR-free state was positively correlated with the $1/K_m$ (R=0.95, P=0.0008) and that the dihedral angle of the O (P2)-C (P2)- C (P1)- O (P1) of the main chains in the PR-bound state was negatively correlated with k_{cat} (R=0.94, P=0.001). We further found that these two folding properties influenced the overall cleavability of the precursor protein when the sizes of the side chains at the P1 site were similar. These data suggest that the dihedral angles at the specific positions around the cleavage junctions before and after binding to PR are both critical for regulating the cleavability of precursor proteins by HIV-1 PR.

Keywords HIV-1 protease · Homology modeling · Replica exchange molecular dynamics · Substrate recognition

Introduction

Human immunodeficiency virus type 1 (HIV-1) encodes an aspartic protease (PR) in its genome. The PR is synthesized as a part of the viral precursor polyprotein Gag-Pol and cleaves two viral precursors, Gag and Gag-Pol, into multiple mature proteins, such as matrix (MA), capsid (CA), nucleocapsid (NC), PR, reverse transcriptase (RT), integrase (IN), p6, and small peptides [1–4]. The proteolytic processing proceeds in the rank order [5–15], which is important for virus replication [2, 8, 12, 16, 17].

To date, several lines of studies have reported the potential importance of structural properties around cleavage junctions for generating the substrate preference of HIV-1 PR. First, crystal structure studies have shown that the cleavage junctions (scissile amide bonds) of HIV-1 precursors have a similar extended conformation when they bind to the catalytic cleft of viral PR (Fig. 1) [18–24]. Looking more closely at the structures around the cleavage junctions, it is evident that the side chains of amino acids at the P1 and P1' positions, which flank scissile bonds, consistently protrude in the opposite directions. This common structural feature may imply that proteolytic enzymes preferentially require the structure to properly position the cleavage site for attack by either an enzyme nucleophile or a water molecule. Second, biochemical studies have reported that an artificial substrate whose

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0739-z) contains supplementary material, which is available to authorized users.

H. Ode (✉) · M. Yokoyama · T. Kanda · H. Sato
Laboratory of Viral Genomics, Pathogen Genomics Center,
National Institute of Infectious Diseases,
4-7-1 Gakuen,
MusashiMurayama-shi, Tokyo 208-0011, Japan
e-mail: odehir@nih.go.jp

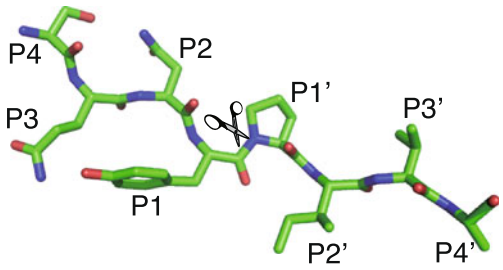

Fig. 1 Conformation of the junction of MA/CA at the binding with HIV-1 PR. The structure was drawn using PyMOL ver. 0.99 (http://www.pymol.org/) with the coordinates in a structure (PDB code: 1FJ4). The definition of positions P4, P3, P2, P1, P1', P2', P3', and P4' is shown. The scissor-shaped symbol points to the cleavage site

β-strand conformation was constrained by chemical modification showed higher binding affinity with the PR [23, 24]. This finding suggests that the extended conformation of the cleavage junction influences the binding efficiency of the substrate to the PR catalytic cleft. Thus, it is conceivable that minute differences in the conformation among the cleavage junctions may influence the efficiency of the substrate binding and cleavage.

To address more quantitatively the impacts of the structures of cleavage junctions on the cleavage reaction by HIV-1 PR, we here examined whether any dihedral angles of amino acids around the cleavage junctions could relate to the experimentally obtained kinetic parameters of the enzyme reaction. For this purpose, we simulated folding of the eight-amino-acid peptides corresponding to the seven cleavage junctions of the HIV-1_{HXB2} Gag and Gag-Pol precursors in the PR-free and PR-bound states by using replica exchange molecular dynamics (REMD) simulations [25–29] and homology modeling [30–32]. We then analyzed dihedral angles around the junctions with the models and searched for a potential relationship between the folding preferences and the reported $1/K_m$ and k_{cat} values for the 8-amino-acid-length peptides corresponding to the cleavage junctions of the HIV-1_{HXB2} strain [12, 33]. This analysis identified two dihedral angles that were significantly related to the kinetic parameters of the enzyme reaction.

Methods

Cleavage junction sequences

Eight-amino-acid peptides that could fit into the catalytic cleft of HIV-1 PR were used for the present structural study. The peptides corresponded to seven cleavage junctions between MA and CA (MA/CA), CA/p2, p2/NC, $p6^{gag\text{-}pol}$ small peptide (p6)/PR, PR/RT, RTp51/RTp66 (RT/RT'), and RT'/IN of HIV-1_{HXB2} (Table 1) [33]. Twenty-two additional eight-mer peptides, which correspond to the MA/CA, CA/p2, and p2/NC cleavage junctions of the HIV-1_{HXB2} strain and have single amino acid substitutions at the P1 position of the cleavage site [12], were also used.

Analysis of peptide folding in the PR-free state

We used replica exchange molecular dynamics (REMD) simulations [25–29] to analyze the folding preferences of substrate peptides in the PR-free state. REMD allows a search for more extensive conformations with shorter simulation times than does the classical MD [34–37]. Because we confirmed that similar data on the folding preferences were obtained with the two MD methods, we present here the results with the REMD simulations (Supplementary material (SM) Fig. S1). REMD simulations of the peptides were performed for 50 nanoseconds (ns) using the AMBER 9 software package [38, 39]. The initial structures of peptides were generated with the LEaP module. During the REMD simulation, a given peptide was folded into various conformations by exchanging semi-randomly non-interacting replicas at eight temperatures. The simulations were performed for each of the eight replicas, using the AMBER ff99SB force field [40] under the modified GB conditions developed by Onufriev *et al*. (IGB=5) [41]. The surface area calculation was also implemented in the simulations. No cut-off was applied for the long-distance interaction energy calculations. The temperature of each replica was set to the temperature at a target ratio of 0.15 [25] and regulated by Langevin dynamics with a collision frequency of 1.0. The ionic concentration was set at 0.2 mM. The shake algorithm was applied to constrain the bonds with hydrogen atoms. To analyze the folding preferences of peptides in the PR-free state, we used 60,000 trajectories in the last 30 ns of the REMD simulations. Dihedral angles of the side and main chains of neighboring amino acids of these structures were calculated with the ptraj module of AMBER 9 and in-house programs.

Analysis of peptide folding preferences in the PR-bound state

Both X-ray crystal and modeling structures were used to analyze the folding of substrate peptides in the PR-bound state. For analysis of the folding of 5 peptides corresponding to the MA/CA, CA/p2, p2/NC, RT/RT', and RT'/IN cleavage junctions, we used reported crystal structural data (PDB codes: 1KJ4, 1F7A, 1KJ7, 1KJG, and 1KJH for the peptides MA/CA, CA/p2, p2/NC, RT/RT', and RT'/IN [18, 19]). For analysis of the folding of the p6/PR and PR/RT peptides, we generated structural models because crystal structures were not available. The models

Table 1 Summary of the conformational characteristics of seven peptides representing cleavage junctions of HIV-1 precursor proteins

Cleavage site	Amino acid sequence[a]	PDB code	Angle 1[b]	Frequency (%)[c]	Angle 2[d]
MA/CA	SQNY/PIVQ	1KJ4	168.5	0.2	143.6
CA/p2	ARVL/AEAM	1F7A	168.9	5.5	155.7
p2/NC	ATIM/MQRG	1KJ7	153.2	1.1	136.2
p6/PR	SFNF/PQIT	–	168.2[e]	0.7	127.5[e]
PR/RT	TLNF/PISP	–	−176.7[e]	0.4	144.9[e]
RT/RT'	AETF/YVDG	1KJG	166.5	6.7	158.6
RT'/IN	RKIL/FLDG	1KJH	168.3	0.7	147.3

[a] A forward slash (/) indicates the cleavage site

[b] Dihedral angles of Cβ (P1) – Cα (P1) – Cα (P1') – Cβ (P1') measured with crystal structures (unmarked) [18, 19] and our homology models ([e])

[c] Frequency of dihedral angle (Cβ (P1) – Cα (P1) – Cα (P1') – Cβ (P1')) in the range of 150 to 180 degrees in 60,000 trajectories during the REMD simulations

[d] Dihedral angle of O (P2) – C (P2) – C (P1) – O (P1) measured with crystal structures (unmarked) [18, 19] or homology models ([e])

were constructed with the homology modeling method [30, 31] using the Molecular Operating Environment (MOE) 2007.09.02 (Chemical Computing Group Inc., Montreal, Quebec, Canada) (http://www.chemcomp.com/) as described previously [32]. The crystal structure of the MA/CA-peptide-PR complex (PDB codes: 1KJ4 [19]) was used as the template because this peptide has a proline at the P1' position, as seen in the p6/PR and PR/RT peptides (Table 1). Water molecules in the crystal structures were included to construct the functional conformation of the peptides for hydrolysis. N25 in the crystal structures was not changed into the aspartic acids, because the D25N substitutions hardly changed the interactions between the PR and the substrates [42, 43]. The AMBER ff99 force field [44] and the generalized Born/volume integral (GB/VI) implicit solvent model [45] were applied for the modeling. The models were thermodynamically optimized by energy minimization with the MOE package using the same force field. Physically unacceptable local structures of the optimized model were assessed and refined using programs in MOE and Verify3D (http://nihserver.mbi.ucla.edu/Verify_3D/ [46]). Dihedral angles of the side and main chains of neighboring amino acids of the crystal structures and homology models were calculated using the MOE software package.

Analysis of the relationship between folding and kinetic parameters

The statistical significance of the association between folding and kinetic parameters of the peptides was evaluated with a nonparametric rank test and simple regression analysis. Several biochemical studies have reported kinetic parameters of the HIV-1-PR-mediated cleavage reaction of peptides corresponding to the cleavage junctions of HIV-1 Gag and Gag-Pol precursor proteins [33, 42, 47–50]. In this study, we used the kinetic data reported by Maschera *et al.* [33] to search for the potential association of kinetic parameters with folding parameters. By using data from a single study rather than from multiple studies, we can avoid potential artifacts of the association analysis due to distinct experimental conditions of individual studies, such as pH, salt conditions, and the amino acid lengths of substrates [48, 49]. We chose the Maschera's data set because it provides the most comprehensive set of the K_m and k_{cat} values among the reports [33, 42, 47–50].

Y-scrambling test

Y-scrambling test were performed as described previously [51]. Briefly, we first prepared 50 data sets in which kinetic parameters were randomly shuffled to change their order in the original data set using Molegro Data Modeller 2009.2.0. (Molegro ApS, Denmark) (http://www.molegro.com/index.php). Pearson correlation R-square and Spearman rank-order correlation coefficient were determined for each data set. Scatter plot were constructed using the R-square and the Spearman coefficient to display values for the two variables (the R-square and the Spearman coefficient) for a set of data.

Prediction of changes in k_{cat}/K_m by substrate mutations

Changes in the k_{cat}/K_m values by mutations at the P1 position of the cleavage site were predicted on the basis of changes in two folding properties of the mutant peptides: the dihedral angles of Cβ (P1)-Cα (P1)-Cα (P1')-Cβ (P1') of substrate peptides in the PR-free state and the dihedral angle of the O (P2)-C (P2)-C (P1)-O (P1) of the main chains of peptides in the catalytic cleft of PR. The 22 mutant peptides used for the study correspond to MA/CA, CA/p2, and p2/NC cleavage junctions of the HIV-1_{HXB2}

strain and have single amino acid substitutions at the P1 position of the cleavage site. REMD simulations of the free peptides and homology modeling of the peptide-PR complex were performed as described above. We used three crystal structures, 1KJ4, 1F7A, and 1KJ7, as the template structures of modeling for the P1-site mutant of MA/CA, CA/p2, and p2/NC, respectively [18, 19]. The 1KJ4, 1F7A, and 1KJ7 include the structural information of the MA/CA, CA/p2, and p2/NC peptides in complex with the PR, respectively. The $1/K_m$ and k_{cat} values were estimated by extrapolating the frequency and angle values to the two regression equations obtained with the present studies on the wild-type peptides: $1/K_m = 0.9105 \times$ frequencies (%) +0.1281, $k_{cat} = -1.5379 \times$ dihedral angle (degrees) +242.23 (see "Results"). The fold increase in the k_{cat}/K_m for the mutant peptides compared to those for the wild-type peptide was calculated with the estimated $1/K_m$ and k_{cat} values.

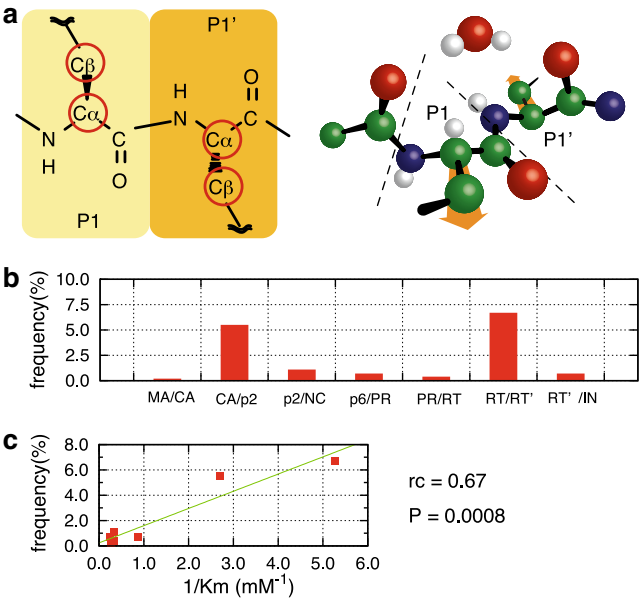

Fig. 2 Structural property correlated with $1/K_m$. **a**. Chemical structure and tertiary structure at the cleavage junction. The notable dihedral angle is highlighted with red circles (left) or orange arrows (right). **b**. The frequency of occurrence of the dihedral angle of Cβ (P1) - Cα (P1) and Cα (P1') - Cβ (P1') in the range of 150 to 180. **c**. The relationship between the frequencies shown in 2B and the reported kinetic parameter $1/K_m$ [33]. The regression coefficients are also shown

Results

Characteristics of substrate folding in the PR-bound state

The dihedral angles of side and main chains of amino acids are an important parameter to define the folding preferences of the peptides. To understand the characteristics of substrate folding in the PR-bound state, we analyzed dihedral angles in the peptide substrates corresponding to the cleavage junctions of the HIV-1_{HXB2} strain using five of the available crystal structures of substrate-PR complexes [18, 19] (Fig. 2a). The complex structures show that two side chains of neighboring amino acids around the cleavage junctions consistently protrude in the opposite directions [18, 19] (Supplementary material (SM) Table S1). For example, the dihedral angle of the Cβ (P1)-Cα (P1)- Cα (P1')-Cβ(P1') at the P1 and P1' positions of the cleavage site consistently ranged from 150 to 180 degrees among the 5 substrate peptides (Table 1, angle 1).

Folding preferences that impact the $1/K_m$

We speculated that folding preferences around the cleavage junctions could be different from junction to junction and that this difference could influence the susceptibility of the substrate to cleavage reaction. To evaluate each of the possibilities, we first analyzed the intrinsic folding preferences of cleavage junctions in the PR-free state by using REMD simulations [26–29] of eight-mer peptides that represent the cleavage junctions of HIV-1HXB2 Gag and Gag-Pol precursor proteins (see "Methods"). Individual trajectories of the REMD simulations were used to examine the dihedral angles of the side and main chains of amino acids in the peptides. We calculated the frequencies of occurrence of individual dihedral angles using 60,000 trajectories over the last 30 ns of the simulations. The distributions of the frequencies for particular angles at specific positions were quite different among the seven peptides tested (SM Fig. S1 for a dihedral angle of the Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') as an example). The differences were reproducible with repeated trials of the REMD simulations (SM Fig. S1). The percentage of cases in which the dihedral angle of the Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') was from 150 to 180 degrees ranged from 0.2% to 6.7% in the simulations (Fig. 2b, Table 1, frequency). Furthermore, the 100-ns standard MD simulations [34–37] of each peptide under the same GB conditions at 300K showed a tendency similar to that of the REMD simulations (SM Fig. S1). These results suggest that intrinsic folding preferences around the cleavage junctions of HIV-1 Gag and Gag-Pol precursors are different from junction to junction in the PR-free state.

We then examined whether any differences in the folding preferences could influence the binding ability of substrate to HIV-1 PR. To address this issue, we analyzed the relationship between preferences of various dihedral angles in the 8-mer peptides of the cleavage junctions of HIV-1_{HXB2} precursors in the PR-free state and the experimentally obtained $1/K_m$ values [33] using the same set of peptides and PR of the HIV-1_{HXB2} strain. We found that the

folding preference for forming a dihedral angle of Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') in the range of 150 to 180 degrees in the PR-free state seemed to positively correlate with the $1/K_m$ (SM Table S2). The significance of association was evaluated with a nonparametric rank test and simple regression analysis. Nonparametric analysis showed a correlation of the frequency values and the $1/K_m$ values (Spearman rank-order correlation coefficient (rs) =0.75, P=0.067, two-tailed) (Fig. 2c). Simple regression analysis showed that the frequency value was positively correlated with the $1/K_m$ value (R=0.95, P=0.0008).

In addition, we found weak associations between the percentage of cases having a Cβ (P1')-Cα (P1')- Cα (P2')-Cβ (P2') in the range of −165 to −195 degrees and the $1/K_m$ (R=0.80, P=0.03), and between the percentage of cases having a φ angle at the P2' site in the range of −120 to −150 degrees (R=0.77, P=0.04) (SM Tables S3–S5). We further analyzed the potential relationship between particular secondary structures, such as α-helices, β-sheets, or turns, and the $1/K_m$ values, using the same 60,000 trajectories of the REMD simulations (SM Table S6). We found weak associations between the percentage of cases having turn structures at the P1' site and the $1/K_m$ (R=0.75, P=0.05). Taken together, these results suggest that differences in the intrinsic folding preferences around the cleavage junction of HIV-1 Gag and Gag-Pol precursors in the PR-free state can influence the binding ability of the cleavage junction to the HIV-1 PR, and that a folding preference for forming a dihedral angle of Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') in the range of 150 to 180 degrees in the PR-free state has the most prominent impact on the binding ability.

Folding preferences that impact the k_{cat}

We next examined whether the PR catalytic reaction is influenced by the folding preferences of the peptides at the PR-bound state by analyzing the crystal structures [18, 19] and homology models of the complexes of the 8-mer substrate peptides and PR of HIV-1_{HXB2}. Because the crystal structures of the complexes were not available for all cleavage sites, we constructed complex models for the p6/PR and PR/RT peptides by homology modeling. During the analysis of various dihedral angles around the substrate cleavage junctions in the PR, we noticed an unusual folding distortion at the P1 and P2 positions, *i.e.*, the dihedral angle of O (P2)-C (P2)- C (P1)- O (P1) of the peptides (Fig. 3a) ranged from 127.5 to 158.6 degrees in all 7 peptides analyzed (Fig. 3b and Table 1, angle 2). Such a conformational distortion at the P1 and P2 positions was rarely detected with the structures of free peptides in the REMD simulations (SM Table S7), suggesting that the folding distortion is caused by interactions of the peptides with the

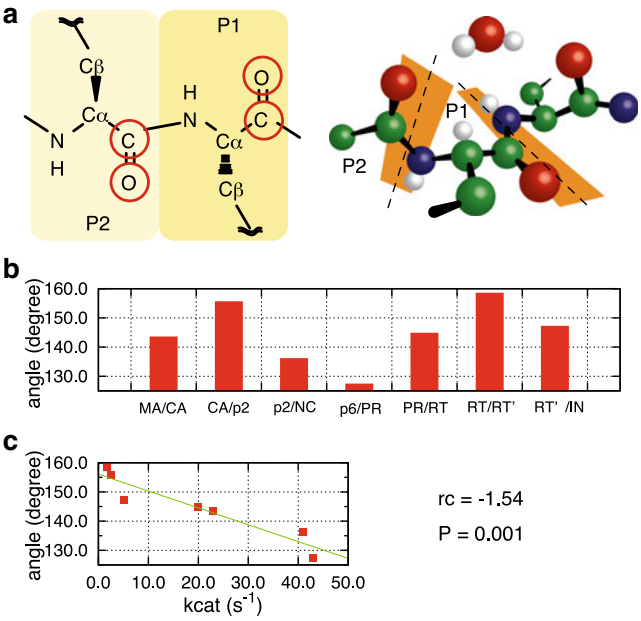

Fig. 3 Structural property correlated with k_{cat}. **a**. Chemical structure and tertiary structure at the cleavage junction. The notable dihedral angle is highlighted with red circles (left) or orange squares (right). **b**. The magnitude of the dihedral angle of O (P2) – C (P2) and C (P1) - O (P1). **c**. Relationship between the dihedral angles shown in 3B and the reported kinetic parameter k_{cat} [33]. The regression coefficients are also shown

PR. We searched for other conformational distortions with the PR-free and PR-bound structures described above. However, we could detect no significant distortions. Thus, the folding distortion appears to be specific to the P1-P2 site and PR-bound state.

The dihedral angle of the O (P2)-C (P2)- C (P1)- O (P1) of the main chains at the P1 and P2 positions in the peptide-PR complex structures was used to examine the relationship between the distorted conformation of the cleavage site and the efficiency of the cleavage reaction (SM Table S2). Nonparametric analysis showed correlation between the dihedral angles and the reported k_{cat} values obtained with the same set of peptides and HIV-1 PR [33] (Fig. 3c) (Spearman rank-order correlation coefficient (rs) = −1, P= 0.014, two-tailed). Simple regression analysis showed that the dihedral angles were inversely correlated with the k_{cat} value (R=0.94, P=0.001). We also examined the potential correlation of other dihedral angles of the peptides with the k_{cat} values (SM Tables S8–S9). Finally, we examined the potential correlation of the folding preference for particular secondary structures with the k_{cat} values, using the 60,000 trajectories of the REMD simulations (SM Tables S3–S6). However, significant correlations (P<0.05) were not detected in these analyses. Taken together, these results suggest that differences in the folding preferences around the cleavage junction of HIV-1 Gag and Gag-Pol precursors in the PR-bound state can influence the cleavage reaction

rate, and that a folding distortion at the P1 and P2 positions has the most prominent impact on the rate.

Y-scrambling test

Because we used relatively small numbers of the experimental data to search correlations with folding properties, it is possible that the correlations we have observed are simply the chance correlations [51]. To address this possibility, we performed Y-scrambling test based on the QSAR methodology [51]. Kinetic parameters were randomly shuffled, and Pearson correlations R-square and Spearman rank-order correlation coefficient were obtained with each randomized data set, as well as original non-randomized data set. The scattering data of the Pearson correlation R-square and Spearman rank-order correlation coefficients show that the correlation values from the original non-randomized data set were positioned most distantly from the origin of the scatter plot and also positioned separately from a cluster of those from the scrambled data set (SM Fig. S2). These results suggest that the original data set had more statistically significant correlation than other scrambled data set. Therefore, it is less likely that the folding properties we described were accidentally correlated with the experimental values.

Impacts of the folding preferences at cleavage junctions for overall cleavability of the precursor protein

We further examined whether the above two folding properties of the peptide are sufficient to determine the overall cleavability of precursor proteins. We estimated the k_{cat}/K_m values for an additional 22 mutant peptides on the basis of the two folding properties alone and analyzed consistency between the predicted k_{cat}/K_m and the reported *in vitro* cleavability of the Gag precursor protein [12]. Briefly, we performed REMD simulations and homology modeling using 22 eight-mer peptides, which correspond to the MA/CA, CA/p2, and p2/NC cleavage junctions of the HIV-1_{HXB2} Gag precursor protein but have single amino acid substitutions at the P1 position of the cleavage site. Then, we estimated the k_{cat}/K_m values for the peptides using the regression equations for the two correlations obtained in the above sections (see "Methods", and SM Table S10). Subsequently, we analyzed the relation between the predicted k_{cat}/K_m values and the reported data on the cleavability of the P1 mutants of the Gag precursor protein [12].

Table 2 shows the fold increases in the predicted k_{cat}/K_m values for the 22 mutant peptides compared to those for the wild-type peptides. The majority of the P1-site mutations were predicted to reduce k_{cat}/K_m values (Table 2, Predicted). In most cases, the P1-site mutations consistently attenuated the cleavage efficiencies of Gag precursors [12] (Table 2, Experiment). However, the predicted fold changes in the k_{cat}/K_m did not show statistically significant correlation with the experimental fold changes in the precursor cleavability (Table 2e and SM Table S10). These results suggest the presence of additional factors influencing the cleavability of precursor proteins.

Table 2 The predicted and experimental cleavability of P1 mutants

	Predicted (peptide)	Experiment (precursor)	
MA/CA	Activity[a]	Activity[ab]	Volume (ratio)[c]
WT(Y)	1.0	1.0	1.00
F	1.4	1.5	0.98
R	0.0[d]	≈0	0.90
I	0.0[d]	≈0	0.86
L	0.7	0.1	0.86
M	1.8[e]	0.25	0.84
T	0.4[e]	≈0	0.60
D	0.2[e]	≈0	0.57
C	3.0[e]	0.05	0.56
A	0.2[e]	≈0	0.46
G	1.9[e]	≈0	0.31
CA/p2	Activity	Activity	Volume (ratio)
Y	0.8[e]	2	1.37
F	1.0[e]	2	1.16
WT(L)	1.0	1.0	1.00
I	0.0[d]	≈0	1.00
T	0.2[e]	≈0	0.70
D	0.6[e]	≈0	0.67
C	0.9[e]	0.35	0.65
p2/NC	Activity	Activity	Volume (ratio)
W	1.0[e]	0.1	1.40
Y	0.2	0.1	1.19
F	0.8	1.0	1.17
L	0.6	0.35	1.02
I	0.0	≈0	1.02
WT(M)	1.0	1.0	1.00
C	0.6	0.15	0.67

[a] Activity is the ratio of the k_{cat}/K_m of mutants to that of the wild-type (WT).

[b] The cleavage rate was taken from a previous report. [12] "≈0" indicates that the mutant precursor was not cleaved at a rate of at least 3% to 5% of the WT rate.

[c] The ratio of the volume of amino acid to the WT one. The volume was taken from a previous report. [52]

[d] A negative value was considered to be zero.

[e] The predicted activity did not reproduce the experimental activity.

Notably, the mutations that gave inconsistent results between the prediction and experiment generally induced larger changes in the side chain size at the P1 position than the mutations that gave consistent results (Table 2, Volume [52]). When the side chain sizes of the mutant peptides were similar to those of the wild-type peptides, the predicted changes in the k_{cat}/K_m had a clear link with the changes in the cleavability obtained by the experiments. In the case of the MA/CA mutant peptides, the predicted changes in the k_{cat}/K_m by F, I, L, and R substitutions were positively correlated with the changes in the cleavability of the MA/CA junction of the precursor ($R=0.90$, $P=0.038$). Such a positive correlation was also detected with Y, F, I, and L substitutions of the p2/NC junction ($R=0.85$, $P=0.031$). Statistical evaluation of the correlation was difficult with the CA/p2 junction because there were few mutants with side chains of a size similar to those of the wild-type.

Discussion

In this study, we investigated the roles of folding preferences of the cleavage junctions of HIV-1 Gag and Gag-Pol precursors in regulating the processing by viral PR. We first demonstrated by molecular dynamics simulations that folding preferences around the cleavage junctions of the HIV-1 Gag and Gag-Pol precursors could be different from junction to junction in the PR-free state. We further showed that these differences could influence $1/K_m$ of the cleavage junction to the HIV-1 PR and that a folding preference for forming a dihedral angle of Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') in the range of 150 to 180 degrees in the PR-free state has a positive correlation with the $1/K_m$. The correlations were reproducible independent of the algorithms and conditions used to infer the folding (SM Fig. S1). These results suggest that the dihedral angles of side chains at the P1 and P1' positions of the cleavage junctions may be important for determining the binding affinity of the cleavage junctions to the HIV-1 PR, because the $1/K_m$ is an approximate indicator of the binding affinity of substrates in enzymatic reactions. The opposite orientation of side chains at the P1 and P1' positions in the PR-free state is commonly seen in the stable PR-bound state [42, 43, 53]. Therefore, it is possible that an increased probability of having the folding feature in the PR-free state increased the likelihood of forming a quasi-stable complex of substrate/water-molecule/PR for subsequent nucleophilic reaction without major conformational changes. However, it should also be noted that the $1/K_m$ may not represent the binding affinity itself. For example, the binding affinity can also be affected by structural factors other than substrate folding properties, such as interactions between substrate and the PR in the PR-bound state. Therefore, although we showed that the particular substrate folding property is correlated with the $1/K_m$, the finding may not imply that the folding property is correlated with binding affinity. Further study is necessary to clarify how the folding property influences the binding affinity itself rather than $1/K_m$.

We next demonstrated by computer-assisted approaches that a folding distortion of substrates in the PR-bound state enhances the cleavage reaction. We showed with available crystal structures of the HIV-1 substrate-PR complexes [18, 19] and our homology models that the dihedral angle of the O (P2)-C (P2)- C (P1)- O (P1) of the main chains of peptides in the catalytic cleft of PR was negatively correlated with k_{cat} for the substrate-PR complex [33]. A smaller dihedral angle at this site indicates a greater distortion of folding immediately upstream of the scissile bond, and the k_{cat} is an indicator of the turnover number of enzymatic reactions. Therefore, our data suggest that a more twisted conformation at the cleavage junction results in more efficient cleavage by HIV-1 PR. This possibility is consistent with the results of quantum mechanical studies in which twisting of the carbonyl group at the P1 residue was shown to be necessary for the access of a water molecule to the scissile bond and for the subsequent nucleophilic reaction [54, 55]. Therefore, it is conceivable that greater magnitudes of twisting of the carbonyl group at the P1 residue increase the likelihood of proper positioning of the water molecule for cleavage reaction.

We further investigated whether the above two folding properties at cleavage junctions are sufficient to determine the overall cleavability of precursor proteins. We found that the k_{cat}/K_m values that we estimated solely on the basis of data of the two folding properties explain the effects of the P1-site mutations on the cleavability of the precursor at MA/CA and p2/NC cleavage junctions only when the sizes of the side chains at the P1 site are similar. The data suggest that the two folding properties are important but insufficient determinants of the overall cleavability of precursor proteins and that the side chain sizes of amino acids around the cleavage junctions are an additional key factor that influences the precursor cleavability. The side chain sizes at the P1 site play a key role in interaction with the catalytic site of PR [18, 19]. Therefore, it is possible that the side chain sizes influence the precursor cleavability by modulating the stability of a complex formed between the substrate and PR.

In this study, we used eight-mer peptides to investigate the relationship between the folding preferences of the cleavage junctions and enzyme kinetic parameters. Therefore, it is unclear how amino acids outside of the cleavage junctions influence the cleavability of the natural substrates. As mentioned above, folding preferences of the cleavage

junctions alone were insufficient to explain the overall cleavability of natural precursor proteins observed in the cells. The accessibility of the PR active site into the cleavage junctions in the precursor proteins may be an additional factor influencing the cleavage order. Further study is necessary to clarify this issue.

Processing of precursor proteins by PR plays an essential role in producing mature infectious viruses in many viruses, as well as in the regulation of many biological events in nature. Notably, two side chains of neighboring amino acids around the cleavage junctions consistently protruded in the opposite directions in the available crystal structures of the PR-substrate complexes [23]. Therefore, it would be interesting to study whether the specific folding preferences or the dihedral angles around the cleavage junctions are critical for regulating their cleavabilities in other PR/substrate reaction systems.

Summary

We used bioinformatics approaches to examine whether the folding preferences of the cleavage junctions of HIV-1 Gag and Gag-Pol precursor proteins influence their cleavabilities by HIV-1 PR. We found that a folding preference for forming a dihedral angle of Cβ (P1)-Cα (P1)- Cα (P1')-Cβ (P1') in the range of 150 to 180 degrees in the PR-free state was positively correlated with the $1/K_m$ ($R=0.95$, $P=0.0008$) and that the dihedral angle of the O (P2)-C (P2)-C (P1)- O (P1) of the main chains in the PR-bound state was negatively correlated with k_{cat} ($R=0.94$, $P=0.001$). We further found that these two folding properties influenced the overall cleavability of the precursor protein when the sizes of the side chains at the P1 site were similar. These data suggest that the dihedral angles at the specific positions around the cleavage junctions before and after binding to PR are both critical for regulating the cleavability of precursor proteins by HIV-1 PR.

Acknowledgments This work was supported by a Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) Research Fellowships, and a Grant for HIV/AIDS Research from the Ministry of Health, Labor and Welfare of Japan.

Author contributions Designed the research: H.O. and H.S. Performed the research: H.O. Contributed analytic tools: M.Y. and T. K. Analyzed the data: H.O. Wrote the paper: H.O. and H.S.

References

1. Leis J, Baltimore D, Bishop JM et al. (1988) Standardized and simplified nomenclature for proteins common to all retroviruses. J Virol 62:1808–1809
2. Partin K, Krausslich HG, Ehrlich L et al. (1990) Mutational analysis of a native substrate of the human immunodeficiency virus type 1 proteinase. J Virol 64:3938–3947
3. Wondrak EM, Louis JM, de Rocquigny H et al. (1993) The gag precursor contains a specific HIV-1 protease cleavage site between the NC (P7) and P1 proteins. FEBS Lett 333:21–24
4. Louis JM, Dyda F, Nashed NT et al. (1998) Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. Biochemistry 37:2105–2110
5. Mervis RJ, Ahmad N, Lillehoj EP et al. (1988) The gag gene products of human immunodeficiency virus type 1: alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. J Virol 62:3993–4002
6. Gowda SD, Stein BS, Engleman EG (1989) Identification of protein intermediates in the processing of the p55 HIV-1 gag precursor in cells infected with recombinant vaccinia virus. J Biol Chem 264:8459–8462
7. Krausslich HG, Ingraham RH, Skoog MT et al. (1989) Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. Proc Natl Acad Sci USA 86:807–811
8. Pettit SC, Moody MD, Wehbie RS et al. (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. J Virol 68:8017–8027
9. Lindhofer H, von der Helm K, Nitschko H (1995) In vivo processing of Pr160gag-pol from human immunodeficiency virus type 1 (HIV) in acutely infected, cultured human T-lymphocytes. Virology 214:624–627
10. Almog N, Roller R, Arad G et al. (1996) A p6Pol-protease fusion protein is present in mature particles of human immunodeficiency virus type 1. J Virol 70:7228–7232
11. Wiegers K, Rutter G, Kottler H et al. (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. J Virol 72:2846–2854
12. Pettit SC, Henderson GJ, Schiffer CA et al. (2002) Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. J Virol 76:10226–10233
13. Pettit SC, Everitt LE, Choudhury S et al. (2004) Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. J Virol 78:8477–8485
14. Pettit SC, Clemente JC, Jeung JA et al. (2005) Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. J Virol 79:10601–10607
15. Pettit SC, Lindquist JN, Kaplan AH et al. (2005) Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. Retrovirology 2:66
16. Jupp RA, Phylip LH, Mills JS et al. (1991) Mutating P2 and P1 residues at cleavage junctions in the HIV-1 pol polyprotein. Effects on hydrolysis by HIV-1 proteinase. FEBS Lett 283:180–184
17. Tritch RJ, Cheng YE, Yin FH et al. (1991) Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein. J Virol 65:922–930
18. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2000) How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. J Mol Biol 301:1207–1220
19. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease:

analysis of crystal structures of six substrate complexes. Structure 10:369–381
20. King NM, Prabu-Jeyabalan M, Nalivaika EA et al. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease. Chem Biol 11:1333–1338
21. King NM, Prabu-Jeyabalan M, Nalivaika EA et al. (2004) Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. J Virol 78:12012–12021
22. Prabu-Jeyabalan M, Nalivaika EA, King NM et al. (2004) Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. J Virol 78:12446–12454
23. Tyndall JD, Nall T, Fairlie DP (2005) Proteases universally recognize beta strands in their active sites. Chem Rev 105:973–999
24. Fairlie DP, Tyndall JD, Reid RC et al. (2000) Conformational selection of inhibitors and substrates by proteolytic enzymes: implications for drug design and polypeptide processing. J Med Chem 43:1271–1281
25. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters 314:141–151
26. Pitera JW, Swope W (2003) Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. Proc Natl Acad Sci USA 100:7587–7592
27. Andrec M, Felts AK, Gallicchio E et al. (2005) Protein folding pathways from replica exchange simulations and a kinetic network model. Proc Natl Acad Sci USA 102:6801–6806
28. Im W, Brooks CL (2005) Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. Proc Natl Acad Sci USA 102:6771–6776
29. Seibert MM, Patriksson A, Hess B et al. (2005) Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. J Mol Biol 354:173–183
30. Marti-Renom MA, Stuart AC, Fiser A et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325
31. Baker D, Sali A (2001) Protein structure prediction and structural genomics. Science 294:93–96
32. Shirakawa K, Takaori-Kondo A, Yokoyama M et al. (2008) Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. Nat Struct Mol Biol 15:1184–1191
33. Maschera B, Darby G, Palu G et al. (1996) Human immunodeficiency virus. Mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. J Biol Chem 271:33231–33235
34. Zoete V, Michielin O, Karplus M (2002) Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. J Mol Biol 315:21–52
35. 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
36. Ode H, Neya S, Hata M et al. (2006) Computational simulations of HIV-1 proteases–multi-drug resistance due to nonactive site mutation L90M. J Am Chem Soc 128:7887–7895
37. Ode H, Matsuyama S, Hata M et al. (2007) Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. J Mol Biol 370:598–607
38. Pearlman DA, Case DA, Caldwell JW et al. (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
39. Kollman PA, Massova I, Reyes C et al. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
40. Hornak V, Abel R, Okur A et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725
41. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized Born model. Proteins 55:383–394
42. Tie Y, Boross PI, Wang YF et al. (2005) Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 angstroms resolution crystal structures of HIV-1 protease mutants with substrate analogs. FEBS J 272:5265–5277
43. Sayer JM, Liu F, Ishima R et al. (2008) Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. J Biol Chem 283:13459–13470
44. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calcluating conformational energies of organic and biological molecules? J Comput Chem 29:1693–1698
45. Labute P (2008) The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. J Comput Chem 29:1693–1698
46. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
47. Kontijevskis A, Prusis P, Petrovska R et al. (2007) A look inside HIV resistance through retroviral protease interaction maps. PLoS Comput Biol 3:e48
48. Szeltner Z, Polgar L (1996) Rate-determining steps in HIV-1 protease catalysis. The hydrolysis of the most specific substrate. J Biol Chem 271:32180–32184
49. Tozser J, Gustchina A, Weber IT et al. (1991) Studies on the role of the S4 substrate binding site of HIV proteinases. FEBS Lett 279:356–360
50. You L, Garwicz D, Rognvaldsson T (2005) Comprehensive bioinformatic analysis of the specificity of human immunodeficiency virus type 1 protease. J Virol 79:12477–12486
51. Dogra SK. "Y Scrmabling" from QSARWorld - A Strand Life Sciences Web Resource. Available from: http://www.qsarworld.com/qsar-statistics-y-scrambling.php
52. Zamyatin AA (1972) Protein volume in solution. Prog Biophys Mol Biol 24:107–123
53. Prabu-Jeyabalan M, King NM, Nalivaika EA et al. (2006) Substrate envelope and drug resistance: crystal structure of RO1 in complex with wild-type human immunodeficiency virus type 1 protease. Antimicrob Agents Chemother 50:1518–1521
54. Okimoto N, Tsukui T, Hata M et al. (1999) Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: investigation by the ab initio melecular orbital method. J Am Chem Soc 121:7349–7354
55. Okimoto N, Tsukui T, Kitayama K et al. (2000) Molecular dynamics study of HIV-1 protease-substrate complex: roles of the water molecules at the loop structures of the active site. J Am Chem Soc 122:5613–5622