Structural integrity of the ribonuclease H domain in HIV-1 reverse transcriptase: Structural Integrity of the HIV-1 RNase H domain

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · JUNE 2015

Impact Factor: 2.63 · DOI: 10.1002/prot.24843

6 AUTHORS, INCLUDING:

Justin Spiriti
University of Pittsburgh

Michael Parniak
University of Pittsburgh

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Structural integrity of the ribonuclease H domain in HIV-1 reverse transcriptase

Ryan L. Slack,[1] Justin Spiriti,[2] Jinwoo Ahn,[1] Michael A. Parniak,[3] Daniel M. Zuckerman,[2*] and Rieko Ishima[1*]

[1] Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260
[2] Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260
[3] Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260

ABSTRACT

The mature form of reverse transcriptase (RT) is a heterodimer comprising the intact 66-kDa subunit (p66) and a smaller 51-kDa subunit (p51) that is generated by removal of most of the RNase H (RNH) domain from a p66 subunit by proteolytic cleavage between residues 440 and 441. Viral infectivity is eliminated by mutations such as F440A and E438N in the proteolytic cleavage sequence, while normal processing and virus infectivity are restored by a compensatory mutation, T477A, that is located more than 10 Å away from the processing site. The molecular basis for this compensatory effect has remained unclear. We therefore investigated structural characteristics of RNH mutants using computational and experimental approaches. Our Nuclear Magnetic Resonance and Differential Scanning Fluorimetry results show that both F440A and E438N mutations disrupt RNH folding. Addition of the T477A mutation restores correct folding of the RNH domain despite the presence of the F440A or E438N mutations. Molecular dynamics simulations suggest that the T477A mutation affects the processing site by altering relative orientations of secondary structure elements. Predictions of sequence tolerance suggest that phenylalanine and tyrosine are structurally preferred at residues 440 and 441, respectively, which are the P1 and P1' substrate residues known to require bulky side chains for substrate specificity. Interestingly, our study demonstrates that the processing site residues, which are critical for protease substrate specificity and must be exposed to the solvent for efficient processing, also function to maintain proper RNH folding in the p66/p51 heterodimer.

Proteins 2015; 83:1526–1538.

Key words: NMR; enzyme; protein; HIV; proteolysis; maturation; virus.

INTRODUCTION

The gene for HIV-1 reverse transcriptase (RT) encodes a 66 kDa protein, but mature HIV-1 reverse transcriptase is a heterodimeric protein comprised of 66 kDa and 51 kDa subunits. RT is initially translated as part of a much larger 160 kDa Gag-Pol polyprotein which is then processed by HIV-1 protease in a still poorly understood manner to yield the mature RT p66/p51 heterodimer. The smaller p51 subunit is generated by removal of most of the ribonuclease H (RNH) domain from a p66 subunit [Fig. 1(A)].[1–8] Structures of both the RT heterodimer as well as the isolated RNH domain indicate that the p51-RNH cleavage site is located within the folded RNH domain, sequestered into the core β-sheet, and thus likely inaccessible to the protease [Fig. 1(B)].[9–14]

A previous study introduced mutations within and surrounding the p51–RNH cleavage site expecting to find a relative accumulation of p66 subunits due to deceased processing efficiency of these cleavage-site mutants.[15] Instead, these mutations resulted in dramatic phenotypic alterations characterized by reduced viral infectivity, a significant reduction in virion RT p66 subunits, and a concomitant increase in the relative number of p51 subunits and fragments smaller than p51, suggesting

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: National Institutes of Health; Grant numbers: P50GM082251, P50GM103368, and P41GM103712; Grant sponsor: National Science Foundation; Grant numbers: MCB-1119091 and CNS-1229064.
Ryan L. Slack and Justin Spiriti contributed equally to this work
*Correspondence to: Rieko Ishima, Room 1037, Biomedical Science Tower 3, 3501 5th Ave., Pittsburgh, PA 15260. E-mail: ishima@pitt.edu and Daniel M. Zuckerman, Room 3079, Biomedical Science Tower 3, 3501 5th Ave., Pittsburgh, PA 15260. E-mail: ddmmzz@pitt.edu
Received 12 December 2014; Revised 14 May 2015; Accepted 3 June 2015
Published online 9 June 2015 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24843

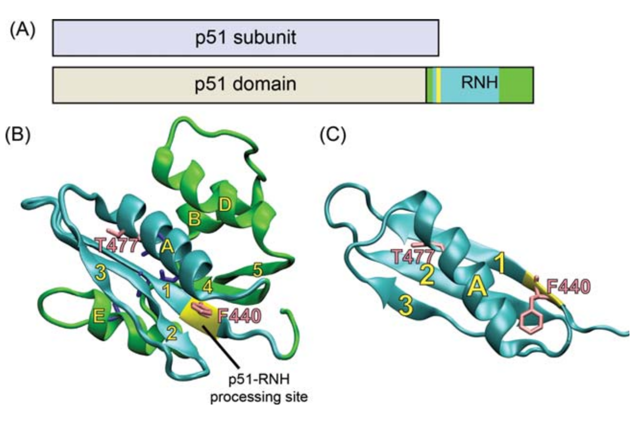

Figure 1

(A) Domain organization of RT, illustrating p66 (below) and p51 (above); the location of the protease processing site in p66 is indicated in yellow. Ribbon representation of the structures of (**B**) the RNH domain and (**C**) the part of the RNH domain, highlighting α-helix A and the β-sheet that includes β-strand 1, 2, and 3. In (B) and (C), the p51-RNH processing site is shown by yellow ribbon, and side chains of F440 and T477 are shown by pink color sticks.

unregulated degradation of RT during proteolytic maturation.[15] Degradation of RT upon mutation of the processing site has been further demonstrated recently.[16] These findings are not easily explained by changes in the side-chain volumes or charges because severe reduction of viral infectivity was observed for mutations both to hydrophobic residues, F440A and F440V, and to a hydrophilic residue, E438N. Further study revealed that an additional mutation (T477A) that arose during continued passage of the cleavage mutants rescued the p66 processing defects of these cleavage-site mutants and significantly restored viral infectivity.[17] This revertant mutation site is located >10 Å away from the processing site: T477 is on α-helix A which interfaces with the core β-sheet containing the processing site [Fig. 1(C)]. The molecular mechanism for the compensatory impact of the T477A mutation was thus unclear.

In the present study the structural basis for the differences in proteolytic stability of the processing-site mutants and the revertants were studied using nuclear magnetic resonance (NMR), differential scanning fluorimetry (DSF), molecular dynamics (MD) simulations, and computational predictions of residues tolerated at the mutation sites. In particular, we characterized conformations of two isolated RNH domain mutants that contain either the F440A or E438N processing defect mutation (denoted RNH_{F440A} and RNH_{E438N}, respectively), and those with the additional T477A rescue mutation (denoted $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$, respectively). Comparison of the ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectra of the mutant RNHs with wild-type (WT) shows that RNH_{F440A} and RNH_{E438N} are unfolded in solution but those with the compensatory T477A are not. The stark difference between the RNH_{F440A} and RNH_{E438N} and those with the compensatory T477A was also observed in the DSF study. Consistent with this observation, 200 ns MD simulations exhibit wider structural variations for RNH_{F440A} and RNH_{E438N} compared with those of $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$, respectively. Structural ensembles obtained by the MD simulations for T477A mutants all exhibit a slight increase in the relative orientation of the α-helix A against the core β-sheet, compared with the WT. Predictions of sequence tolerance using Rosetta-Backrub[18,19] suggest that phenylanaine and tyrosine are structurally preferred for residues 440 and 441, respectively. Our results suggest that specificity as a protease substrate is coupled to the structural requirement to maintain the RNH fold.

METHODS

Sample preparation

Isolated WT RNH domain constructs were prepared by expressing the domain in *Escherichia coli*. In brief, the cDNA encoding RT residues 427 to 560 was inserted into the pE-SUMO vector (LifeSensors, Malvern, PA) with a six histidine tag (His_{6}-) at the N-terminus of the SUMO-fusion construct. As opposed to our previous WT RNH construct containing three additional amino acid residues S-E-L at the N-terminus of native RNH,[20] the current construct encodes only native amino acid sequence of RNH after removal of SUMO. Mutations, E438N, E438N/T477A, F440A and F440A/T477A were introduced to the WT construct using QuickChange kits (Stratagene, La Jolla, CA) and verified by DNA sequencing. All the proteins were expressed in *E. coli* Rosetta 2 (DE3) cells. Cell cultures were grown at 37°C to an OD of 1.0, induced with IPTG, and grown at 16°C for an additional 18 h. Isotopic labeling was achieved by growing cultures in modified minimal media containing ^{15}N NH_4Cl as the sole nitrogen source using the published protocol.[21] Cells were harvested by centrifugation, resuspended in 50 m*M* Tris buffer, pH 7.5, and lysed by microfluidation. The His_6-SUMO-fusion RNHs were isolated from the cell lysate using a HisTrap HP columns (GE Healthcare, Piscataway, NJ) with a linear gradient of 0.02*M* to 0.5*M* Imidazole, followed by gel filtration on a Superdex75 26/60 column (GE Healthcare, Piscataway, NJ). The N-terminal His_6-SUMO fusion was then removed by digestion with histidine tagged ubiquitin-like-protein specific protease (ULP1). Finally, the RNH was separated from His-tagged proteins using a HisTrap column (GE Healthcare, Piscataway, NJ), and polished with a Superdex75 26/60 column (GE Healthcare, Piscataway, NJ) equilibrated with a buffer containing 25 m*M* sodium phosphate, 100 m*M* NaCl, and 3 m*M* NaN_3, at pH 7.0. For the purification of RNH_{F440A} and RNH_{E438N}, the proteins were further purified using HiTrap Q HP column (GE Healthcare, Piscataway, NJ) to remove fragmented products. Purity of the proteins at the final step of purification were confirmed by running 20% acrylamide gels in both SDS denatured and non-denaturing (native) conditions (PhastGel system, GE Healthcare, Piscataway, NJ). Protein samples were stored at $-80°C$.

NMR experiments

All NMR experiments were performed at a protein concentration of ~200 μ*M* in NMR buffer (25 m*M* sodium phosphate, 100 m*M* NaCl, pH 7.0) and supplemented with 10% D_2O. All ^{1}H-^{15}N HSQC spectra were recorded at 293 K on Bruker 600 AVANCE spectrometers, equipped with a 5-mm triple-resonance, *z*-axis gradient cryogenic probe. All data were processed with NMRPipe and analyzed with CCPNNMR analysis.[22,23]

Light scattering measurements

Size-exclusion multiangle light scattering (SEC-MALS) measurements were collected at room temperature using an analytical Superdex 75 HR 10/30 column (GE Healthcare, Piscataway, NJ) with in-line multiangle light scattering (HELEOS, Wyatt Technology), UV (Agilent 1100, Agilent Technology), and refractive index (OptilabrEX, Wyatt Technology) detectors. Protein samples (protein concentration: 100 μ*M*, sample volume: 100 μL, in NMR Buffer) were injected into the column, pre-equilibrated with sterile-filtered and degassed NMR buffer. Molecular masses of the eluted proteins were analyzed using ASTRA software, version 5.3.4 (Wyatt Technologies).

Differential scanning fluorimetry

Thermal stability of the WT RNH, RNH_{F440A}, RNH_{E438N}, $RNH_{F440A/T477A}$, and $RNH_{E438N/T477A}$ were monitored by differential scanning fluorimetry. A fluorescence microplate reader (FluoDia T70, Photon Technology International, Edison, NJ) was used to measure binding of hydrophobic dye SYPRO Orange to the unfolded fraction of the protein (Life Technologies, Carlsbad, CA).[24,25] DSF samples were prepared at a protein concentration of 5 μ*M* in a solution containing 25 m*M* sodium phosphate, 100 m*M* NaCl, and 5 × SYPRO Orange at pH 5.0, 6.0, 7.0, 8.0, or 9.0. Sample volume of 25 μL per well was loaded into 96-well PCR plates (Bio-Rad, Hercules, CA). Plates were heated from 25 to 75°C in increments of 0.5°C. Fluorescence intensity was measured using excitation/emission wavelengths of 465 and 590 nm, respectively. Fluorescence data were analyzed in MATLAB (The Mathworks Inc., Natick, MA), and melting temperatures of the proteins determined from the maximum of the first derivative of normalized fluorescence intensity signals, as described by Niesen *et al.*[26] All assays were performed in triplicate.

Molecular dynamics simulations

Wild-type and mutant RNH systems were prepared using the CHARMM package[27,28] and simulated with the NAMD package.[29,30] Initial structures for RNH_{WT}, RNH_{F440A}, RNH_{E438N}, RNH_{T477A}, $RNH_{E438N/T477A}$, and $RNH_{F440A/T477A}$ were generated using CHARMM based on residues 427 to 556 from the crystal structure of the WT RNH domain (PDB ID: 1DLO[31]), together with additional residues having the sequence SEF at the N-terminus and RKVL at the C-terminus in order to match the sequence of the constructs on which NMR studies had been performed. The coordinates of these residues, as well as any atoms not present in the original crystal structure (including hydrogen atoms), were generated using the internal coordinate facility in CHARMM; the additional residues were initially assumed to have extended configurations. The resulting structure was then energy-minimized using harmonic restraints together with the CHARMM36 force field[32,33] and GBMV solvation model.[34,35] After minimization, each structure was surrounded with TIP3 water[36] in a rhombic dodecahedral box, allowing a 12 Å margin on all sides of the protein. A total of 21 Na^{+} and 18 Cl^{-} ions were added to each system and placed using the SOLVATE program,[37] bringing the salt concentration to approximately 100 m*M* while neutralizing the charge. The water and ions in each system were then energy-minimized while keeping the protein fixed.

The systems were then simulated using NAMD.[29,30] Each system was heated to 293 K over 1.5 ns with harmonic restraints of 1.0 kcal/(mol $Å^2$) on each non-hydrogen atom in the protein. The harmonic restraints were then gradually relaxed while equilibrating the system for an additional 1 ns.

Production simulations were carried out for 200 ns for each system with NAMD[29,30] using the CHARMM36 force field and TIP3 water model. We employed a 2 fs time step, using the SHAKE[38] and SETTLE[39] algorithms to constrain all bonds involving hydrogen in the protein and water respectively to their equilibrium values. Periodic boundary conditions were used; long range electrostatics was treated with the particle mesh Ewald method, and a switching function between 8 and 12 Å was applied to the van der Waals interactions. Constant temperature was maintained using Langevin dynamics with a damping coefficient of 5 ps^{-1}, and constant pressure was maintained using a Langevin piston with an oscillation time of 100 fs and a decay time constant of 50 fs. Frames were recorded every 1 ps.

Additional weighted ensemble simulations were performed as a check on the MD results: see Supporting Information.

MD simulation analysis

The trajectories were analyzed using CHARMM[27,28] in order to better understand the effect of the mutations

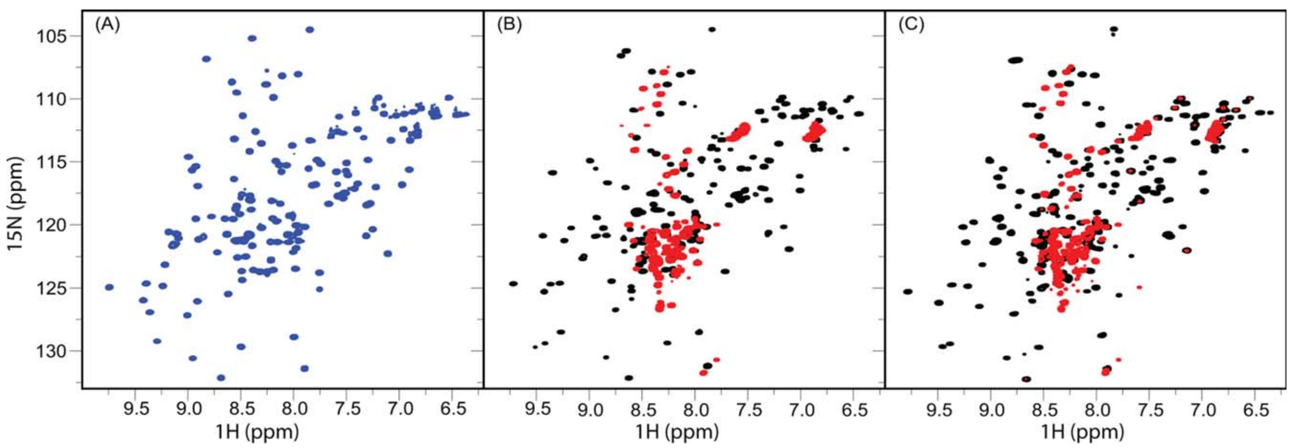

Figure 2

^{1}H-^{15}N HSQC spectra of the RNH wild-type (WT) and mutants. (**A**) The RNH WT spectrum exhibits well disperse and sharp cross-peaks, characteristic of a well-folded protein in solution (blue). (**B**) Superimposition of the RNH_{F440A} mutant (red) and the $RNH_{F440A/T477A}$ mutant (black). (**C**) Superimposition of the RNH_{E438N} mutant (red) and the $RNH_{E438N/T477A}$ mutant (black). All the spectra were obtained on a Bruker AVANCE 600 Spectrometer at 20°C.

on the structure and dynamics of the RNase H domain. The backbone RMSD relative to the starting structure (involving N, C^{α}, and C atoms of residues 427–556) was calculated for each frame in the trajectory. The overall sampling quality of the simulation was evaluated by comparing the distribution of this backbone RMSD in the first half of the trajectory to that in the second.

The structure of the protein near the mutated residues was analyzed in greater detail. In order to determine the effect of mutations on the packing of nearby residues, the number of atoms within 4 Å around each residue was counted in each frame. The effect of the mutations on the local hydrogen bonding network was investigated by calculating the fraction of time individual bonds were present. In these calculations, a hydrogen bond was defined to be present if the hydrogen-acceptor distance was less than 2.4 Å and the donor-hydrogen-acceptor angle was greater than 150°.

In order to characterize the relationship of α-helix A to the first three β-sheet strands (1, 2, 3) for each frame, the helical axis of helix A was determined by applying the algorithm of Aqvist[40] to the α-carbons of residues 474 to 488. This helical axis was then represented in a coordinate system defined by the principal axes of the moment of inertia of the backbone atoms of residues 439 to 446, 453 to 459, and 467 to 469. The helical tilt angle θ was then defined to be the angle between the helical axis and the plane formed by the two principal axes with the smaller moments. The angle ϕ was defined to be the polar angle of the projection of the helical axis on this plane, relative to the principal axis with the smallest moment. For each trajectory, the average and standard deviation of all observables except the hydrogen-bonding fractions was calculated.

Prediction of sequence tolerance to maintain structure at the mutation sites

The structurally preferred amino acid type, at residues 438, 440, 441, and 447, was examined using RosettaBackrub.[18,19,41,42] This server-based software uses flexible backbone modeling and a sequence tolerance protocol to predict amino acid substitutions which preserve near-native folding stability of the protein. We input the RNH domain structure and specified an ensemble size of 100 structures to score structural stability of each mutation. This procedure was performed using the WT and the RNH_{T477A} coordinates, which were the same as those used for the MD simulation. To check sensitivity to the choice of structure, the calculations were repeated using the RNH WT and RNH_{T477A} coordinates obtained from the MD simulations at 100 ns time point.

RESULTS

NMR experiments of the processing site mutants and the revertant

To clarify the effect of the processing-site mutants and the revertant on the RNH structure, ^{1}H-^{15}N HSQC spectra of RNH_{F440A}, RNH_{E438N}, $RNH_{F440A/T477A}$, and $RNH_{E438N/T477A}$, as well as the WT, were recorded (Fig. 2). WT spectrum exhibits folded RNH signals similar to those published previously.[14,20,43,44] In contrast to the WT spectrum, most of the resonances in the ^{1}H-^{15}N HSQC spectra of RNH_{F440A} and RNH_{E438N} were observed at the random coil region, that is, at a narrow ^{1}H chemical shift ranges (8.0–8.5 ppm), indicating that these proteins are disordered, most likely unfolded, in solution [red spectra in Fig. 2(B,C)]. The observed disordered spectral feature of RNH_{F440A} and RNH_{E438N} is not due to unfolding during the expression and purification of the mutants because refolding experiments at various conditions, such as in low salt condition or by using denaturants, did not change the results at pH 7. Instead, because changing the sample condition from pH 7 to 8 increased the folded signals in the ^{1}H-^{15}N HSQC spectra of RNH_{F440A} and RNH_{E438N} (Supporting Information Fig. S1), we believe that charge effects, such as salt-bridges or hydrogen bonds, may contribute to the folding of these mutants.

^{1}H-^{15}N HSQC spectra with the revertant, that is, $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$, showed dispersed NMR signals that are similar to those of the WT [black spectra in Fig. 2(B,C)]. For example, signals above 9 ppm ^{1}H chemical shifts are not observed in the spectra of the RNH_{F440A} and RNH_{E438N} [red color spectra in Fig. 2(B,C)], but are detected in the spectra of $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ similar to that of the WT [black color spectra in Fig. 2(B,C)]. These results indicate that restored folding was the major factor in reactivating the infectivity. However, comparison of the NMR spectra between the $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ [black color spectra in Fig. 2(B,C)] shows that signals from the unfolded fraction still remains in the $RNH_{F440A/T477A}$ spectrum whereas unfolding signals are less significant in the $RNH_{E438N/T477A}$ spectrum. Since viral infectivity of $RNH_{F440A/T477A}$ is approximately 20% greater than $RNH_{E438N/T477A}$,[15,17] the difference in the infectivity is not determined only by the folding ratio but also by a structural factor that affects the RNH activity of the mutants.

The profiles of SDS gel electrophoresis demonstrate single bands for all the RNH mutants, indicative of a singular molecular weight species [Fig. 3(A)]. This result also confirms that the observed random coil chemical shifts of RNH_{F440A} and RNH_{E438N} are not due to proteolytic fragments by *E. coli* enzymes. In contrast to the SDS gel profiles, native gel electrophoresis profiles of $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ show migration patterns with two distinct bands which stem from the monomer and dimer species, demonstrating that the double mutants have dimerization characteristics predominantly similar to that of the WT [Fig. 3(B)].[20] Although the band positions in the native gel electrophoresis for $RNH_{E438N/T477A}$ significantly differ from those of the WT and $RNH_{F440A/T477A}$ [Fig. 3(B)], monomer and dimer molecular masses were confirmed using SEC-MALS [Fig. 3(C)]. In both single mutants, RNH_{F440A} and RNH_{E438N}, a diffuse band was observed, probably due to the surface charge variations in the unfolded protein [Fig. 3(B)]. Overall, these gel profiles of the WT and the mutant RNHs as well as SEC-MALS elution profiles support the NMR observations: F440A and E438N mutations reduce the stability of the RNH folding at pH 7, while T477A rescues the folding.

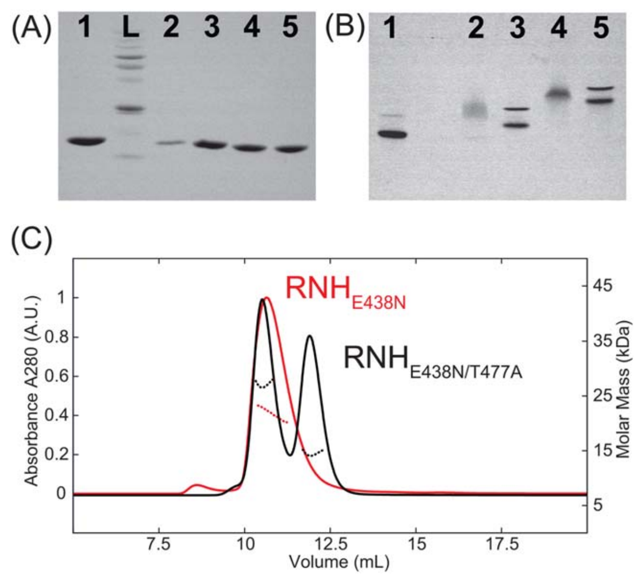

Figure 3

Gel electrophoresis profiles in (**A**) denatured and (**B**) native conditions, and (**C**) SEC-MALS UV profiles (solid) and molecular mass profiles (dashed) of RNH_{E438N} (red) and $RNH_{E438N/T477A}$ (black). In (A), L indicates a molecular weight size marker. In (A) and (B), lane 1: WT RNH, lane 2: RNH_{F440A}, lane 3: $RNH_{F440A/T477A}$, lane 4: RNH_{E438N}, lane 5: $RNH_{E438N/T477A}$. In (C) the average molecular mass of the eluted RNH_{E438N} peak was determined to be 22.1 kDa, and those of the eluted $RNH_{E438N/T477A}$ peaks at 10.5 and 12.1 mL were determined to be 14.88 kDa and 27.25 kDa, respectively.

Differential scanning fluorimetry (DSF) of the processing site mutants and the revertant

To expand on our NMR observations, we used DSF to determine the thermal stability of RNH_{F440A}, RNH_{E438N}, $RNH_{F440A/T477A}$, $RNH_{E438N/T477A}$, and WT RNH over a range of different buffer pH values[45,46] (Table I). The WT exhibited a maximum melting temperature (T_m) of 55.1 ± 2.4°C, at neutral pH, and slightly lower T_m values in alkaline buffer conditions. In all of the pH conditions used for this experiment, RNH_{F440A} and RNH_{E438N} showed high fluorescence intensity throughout the examined temperatures, and the T_m values could not be determined[26] (Supporting Information Fig. S2). Table I shows that T_m values for $RNH_{F440A/T477A}$ were lower than that of the WT at all pH conditions tested. T_m values for $RNH_{E438N/T477A}$ were similar to those of $RNH_{F440A/T477A}$ at neutral or alkaline pH conditions, but could not be obtained at pH 5 and 6, showing a similar profile to those of RNH_{F440A} and RNH_{E438N}. Overall, although the NMR data for $RNH_{F440A/T477A}$ and the $RNH_{E438N/T477A}$ showed a profile similar to that of WT, the thermal stability is low, which may explain the residual unfolded signals in their NMR spectra.

Table I
Melting Temperature of the WT RNH and the Mutants, at Different pHs, Determined by Differential Scanning Fluorimetry[a]

	Melting temperature, T_m (°C)				
pH	5.0	6.0	7.0	8.0	9.0
WT	53.2 ± 0.4	54.3 ± 2.3	55.1 ± 2.4	53.8 ± 1.6	53.8 ± 0.8
F440A	NA	NA	NA	NA	NA
F440A/T477A	47.9 ± 2.4	42.7 ± 2.2	44.6 ± 1.5	43.3 ± 0.8	46.8 ± 2.9
E438N	NA	NA	NA	NA	NA
E438N/T477A	NA	NA	43.8 ± 0.5	43.8 ± 0.5	47.5 ± 2.1

[a]NA indicates that the T_m could not be determined.

Conformational ensembles obtained by MD simulations

As seen above, $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ contain a small unfolded population and a significant dimer population, respectively (Figs. 2 and 3). Thus, it is impossible to determine unambiguous high-resolution structures of these mutants by NMR. To gain atomic level information that could help explain the experimentally observed characteristics of these RNH mutants, MD simulations for 200 ns were performed for RNH_{F440A}, RNH_{E438N}, $RNH_{F440A/T477A}$, $RNH_{E438N/T477A}$, and RNH_{T477A} as well as the WT. MD simulation data has systematic inaccuracies in force fields and limited trajectory times compared with biological timescales.[47–50] In the present study, for example, the trajectories are not long enough for the unstable mutant RNH_{F440A} to unfold, as described below. On the other hand, MD simulation is advantageous for relative comparison of protein dynamics *on the timescale of the trajectories* (200 ns, in the present study), which in turn provides clues to the physical reasons for the cross-talk between the processing site and the T477 site.

To obtain an overview of the conformational fluctuations observed during the simulation, the root-mean-square deviation (RMSD) of the backbone was calculated against the wild-type crystal structure for all frames. The distributions of this RMSD in the first and second half of each trajectory were compared in order to assess the quality of sampling in each simulation (Fig. 4, solid lines vs. symbols). Ideally, if each trajectory were of sufficient length, these distributions would be the same to within statistical error. Indeed, the RMSD was distributed from 1 to 2 Å in the WT with a 68% overlap between the 0 to 100 ns and 100 to 200 ns periods. Interestingly, the RMSD distribution was narrowed, with about 80% overlap, in RNH_{F477A}, compared with the WT [Fig. 4(A)].

A similar but more pronounced effect by the T477A mutation was observed in the F440A and E438N mutants: the RMSD distributions of the two time periods differ in RNH_{F440A} and RNH_{E438N} whereas those of the two time periods are almost identical to each other in their T477A mutants [Fig. 4(B,C)], implying a high degree of stability. Since RNH_{F440A} and RNH_{E438N} are

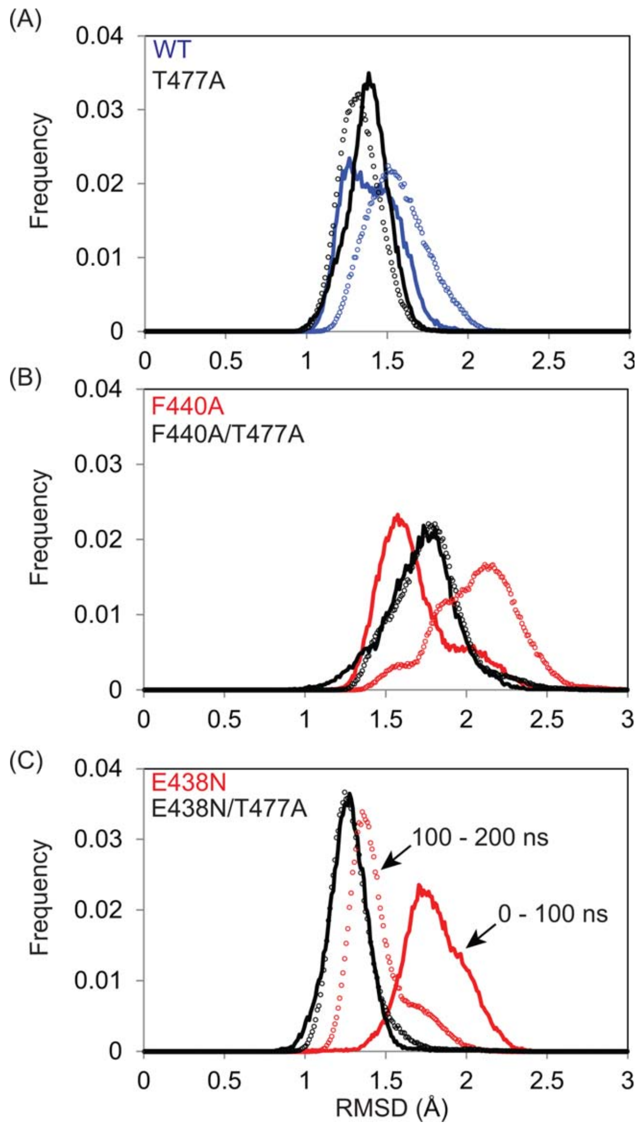

Figure 4
Global assessment of sampling and structural diversity via MD simulations. Histograms of the backbone RMSD (Å) of structures obtained in the 1 to 100 ns (solid line) and of 100 to 200 ns (symbols) MD simulations for (**A**) WT, (**B**) RNH_{F440A}, and (**C**) RNH_{E438N} (blue line, for WT; red lines, F440A and E438N mutants without T477A; black lines, with T477 mutation).

Table II
Average and Standard Deviation of Number of Nonhydrogen Atoms in Protein Within 4 Å Around Residues 438, 440, 477

Simulation	Res. 438	Res. 440	Res. 477
WT	39.5 ± 2.6	48.3 ± 3.2	40.5 ± 2.3
T477A	40.2 ± 2.5	48.3 ± 3.1	33.3 ± 2.3
F440A	42.0 ± 3.4	29.1 ± 1.9	40.7 ± 2.4
F440A/T477A	40.5 ± 2.6	28.7 ± 2.0	34.4 ± 2.5
E438N	31.8 ± 3.0	43.2 ± 3.2	40.0 ± 2.5
E438N/T477A	31.6 ± 2.7	42.9 ± 2.9	33.7 ± 2.3

mostly unfolded in solution (Fig. 2), the 200 ns simulations likely do not reflect the entire conformational ensembles of the mutants, but instead reflect conformational fluctuation around the initial folded RNH structure. Larger and longer-lived fluctuations are observed for RNH_{F440A} and RNH_{E438N} compared with the other trajectories, consistent with their structures being experimentally less stable. In addition, the RMSD comparisons demonstrate that the structures of the RNHs with the revertant mutation, T477A, are more structurally ordered on the timescale of the simulation, compared with those without the T477A mutation. Because 200 ns of MD simulation is not sufficient to characterize a protein exhibiting significant fluctuations (for example, unfolding), we also performed Weighted Ensemble (WE) simulations, which validate the observation that RNH_{F440A} is more unstable than $RNH_{F440A/F477A}$ or WT (Supporting Information Fig. S3).

Conformational characteristics observed by MD simulations

In order to investigate the changes in the conformational ensemble caused by the mutations, characteristics of the local structures around the processing site and residue 477 were compared in the various trajectories. As expected from the volume changes of the side chains, the average number of protein atoms which surround the mutation sites decreased upon mutations from 39.5 ± 2.6 for the WT compared with 31.8 ± 3.0 for the E438N mutation; from 48.3 ± 3.2 for the WT compared with 29.1 ± 1.9 for the F440A mutation; and 40.5 ± 2.3 for the WT compared with 33.3 ± 2.3 for the T477A mutation (Table II). Interestingly, the reduction in the number of residues surrounding the processing sites remains even in $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$, which exhibit stable conformations in the 200 ns simulations. These simulation results suggest that T477A mutation does not counteract the loss of side chain packing at the processing site residues due to the F440A and E438N mutations.

More quantitative conformational changes upon mutations were investigated by monitoring the stability of the hydrogen bond network around residue E438 [Fig. 5(A)]. When residue 438 is a glutamate, i.e., in the WT, the O^{ε} atoms from the carboxyl group form hydrogen bonds with the H^{ε} and H^{η} atoms from the guanidinum group of R463. At the same time, the carboxyl group also forms a hydrogen bond with H^{γ} from the hydroxyl group of T459 (Table III). When residue 438 is mutated to asparagine, it becomes neutral, and this hydrogen bond network is eliminated; instead, R463 faces outward toward the solvent. Because of this change of the R463 side chain orientation, the stability of the hydrogen bond between the D488 backbone carbonyl and the R463 side chain is also reduced upon the E438N mutation. In short, the E438N mutation induces the loss of the local hydrogen bond network.

Similar to the E438N mutation, significant reduction of the average number of surrounding atoms (from 48.3 ± 3.2 for the WT compared with 29.1 ± 1.9 for the RNH_{F440A}) occurs upon the F440A mutation. The F440A mutation slightly reduces the occupancy of the hydrogen bond from the D488 backbone and increases the hydrogen bond occupancy from the D460 and R461 side

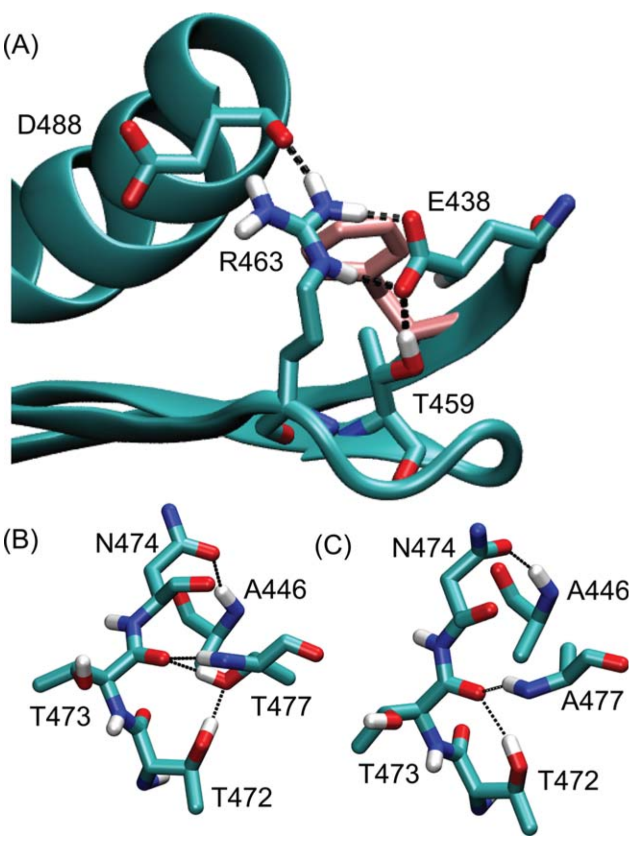

Figure 5
Hydrogen bonding networks observed in MD simulations. (**A**) Side chain orientations in the WT RNH around (A) residue 440 (pink), and residue 477 in (**B**) the WT and (**C**) the RNH_{T477A}. In (A), dashed lines indicate salt bridge network that involves E438. In (B) and (C), black lines indicate hydrogen bonds that are frequently observed in the WT RNH and the RNH_{T477A}, respectively (see Table IV).

Table III
Average Number of Selected Hydrogen Bonds, Around the Processing Site, Observed in each Simulation[a]

Simulation	E438 O^{ε}-R463 H^{ε}, H^{η}	D488 O-R463 H^{ε}, H^{η}	D460 O^{δ}-R461 H^{ε}, H^{η}	D460 O^{δ}-T439 HN	res. 438 $O^{\delta}/O^{\varepsilon}$-T459 H^{γ}
WT	1.92	0.40	0.67	0.66	0.95
T477A	1.91	0.43	0.79	0.86	0.95
F440A	1.75	0.22	1.37	0.92	0.95
F440A/T477A	1.85	0.63	0.46	0.41	0.94
E438N	0.00	0.00	1.42	0.88	0.00
E438N/T477A	0.00	0.00	1.52	0.86	0.00

[a]Hydrogen bond criteria are hydrogen-acceptor distance <2.4 Å and donor-hydrogen-acceptor angle >150°.

chains. Although the F440A mutation did not exhibit such drastic changes in the hydrogen bond network, the mutation results in larger backbone RMSD deviation compared with those of the WT and the E438N mutation (Fig. 4). Overall, reduction of the Phe to Ala side chain probably affects packing of the protein core, including hydrogen bonding networks and hydrophobic interactions.

The region around residue 477 is well folded with different hydrogen bond interactions occurring in RNH with and without the T477A mutation [Fig. 5(B,C)]. When residue 477 is a threonine, the hydroxyl side chain of T477 forms a hydrogen bond with the T473 backbone carbonyl oxygen or with T472 hydroxyl group, depending on the protonation of the hydroxyl side chain (Table IV). These hydrogen bonds, along with the hydrogen bond from N474 to A446, help to maintain the position of the N-terminal end of helix 1 relative to the first three β-strands. By contrast, the mutation of residue 477 to alanine does not allow formation of such inter helix-loop hydrogen bonds. Instead, the loop region is stabilized by forming two different hydrogen bonds, one between N477 side chain OG1 (in the loop) and A466 backbone NH (in the β-1 strand) and another between T472 hydroxyl side chain (in the loop) and T473 backbone carbonyl [Table IV and Fig. 5(C)].

Crosstalk between the processing site and the residue 477

The MD trajectories suggest a hypothesis for the mechanism of crosstalk between the region around the residue 477 and the somewhat distant processing site. The simulation results show a tendency for the helical tilt angle, θ, of RNH_{T477A} to exceed that of WT, in contrast to negligible observed changes in ϕ. The reduced size of A477 compared with T477 leads to the movement of the N-terminal end of the helix toward the beta strands, thereby increasing θ by ~3° (Table V, Fig. 6). The θ angles of $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ are similar to that of RNH_{T477A}, reflecting similar movement at the N-terminal end of the helix A (Table V, and Fig. 6). The number of residues surrounding the processing site near the C-terminus of helix A of $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ is smaller than that of WT (Table II). Thus, on the whole, this simple "levering" picture appears to explain the crosstalk between the distant sites.

Sequence tolerance of the processing site mutations

Preference of amino acid sequences on the RNH structure was systematically evaluated for the processing-site residues using the Rosettabackrub software by randomly sampling of any of the 20 amino acids except Cys and repacking of residues within 4 Å of the newly designed residue for energy minimization.[18,19,41,42] First, amino acid preferences at residues 438, 440, and 441 were calculated on the structural platforms of the WT and RNH_{T477A}. For the residue 440, the calculated frequency of the preferred amino acid demonstrates that Phe and Tyr residues are strongly preferred to maintain the WT-like structure, in both WT and RNH_{T477A} [Fig. 7(A)]. Similarly, Glu or Asp residue is preferred for the residue 438 [Fig. 7(B)], indicating that these processing site residues are favored to maintain the RNH folding.

Table IV
Fraction of the Selected Hydrogen Bonds, Around Residue 477, Observed in each Simulation[a]

Simulation	N474 O $^{\delta}$-A446 HN-	T473 O- T472 H^{γ}-	T473 O- T477 HN-	T477 O^{γ}-T472 H^{γ}-	T473 O-T477 H^{γ}
WT	0.08	0.01	0.60	0.26	0.71
T477A	0.43	0.39	0.51	0.00	0.00
F440A	0.22	0.02	0.61	0.19	0.69
F440A/T477A	0.58	0.56	0.56	0.00	0.00
E438N	0.10	0.01	0.61	0.19	0.71
E438N/T477A	0.42	0.46	0.46	0.00	0.00

[a]Hydrogen bond criteria are hydrogen-acceptor distance less than 2.4 Å and donor-hydrogen-acceptor angle greater than 150°.

Table V
Axis of Helix A (C^{α} Atoms) Relative to Inertia Axes of Backbone Atoms of β Strands 1–3[a]

Simulation	θ (°)	ϕ (°)
WT	7.7 ± 1.9	41.5 ± 1.8
T477A	10.5 ± 1.6	41.1 ± 1.8
F440A	6.0 ± 2.3	40.5 ± 2.2
F440A/T477A	10.2 ± 1.7	40.7 ± 1.6
E438N	7.2 ± 2.0	40.0 ± 1.8
E438N/T477A	10.0 ± 1.6	40.0 ± 1.9

[a]The indicated error bars represent the standard deviation over each trajectory. Since there is a slow conformational change, we were unable to obtain a reliable estimate of the standard error from the block averaging approach.[49]

Since residue 440 is the P1 site as a substrate for the HIV-1 protease, for comparison, sequence tolerance was also tested for residue 441, which corresponds to the P1' site as a substrate for the HIV-1 protease. The calculation indicates that Tyr and Phe are structurally preferred for the P1' site [Fig. 7(C)]. Since the hydroxyl group of Tyr side chain of the residue 441 forms a hydrogen bond with the backbone carbonyl of Lys 287 in the p51 subunit, Tyr must be preferred in the actual RNH domain in RT than Phe. Our observation suggests that although P1 and P1' sites are typically exposed to the solvent for protease cleavage, those of the p51-RNH processing site, F440 and Y441, contribute to structural stability to maintain the protein core (discussed later).

Interestingly, Ala is preferred for the residue 477 compared with Thr [Fig. 7(D)], which is consistent to the narrower RMSD distribution in the RNH_{T477A} compared with the WT [Fig. 4(A)], and consistent with the above observation that the processing site mutations cause folding defects in the WT but less in the T477A mutants

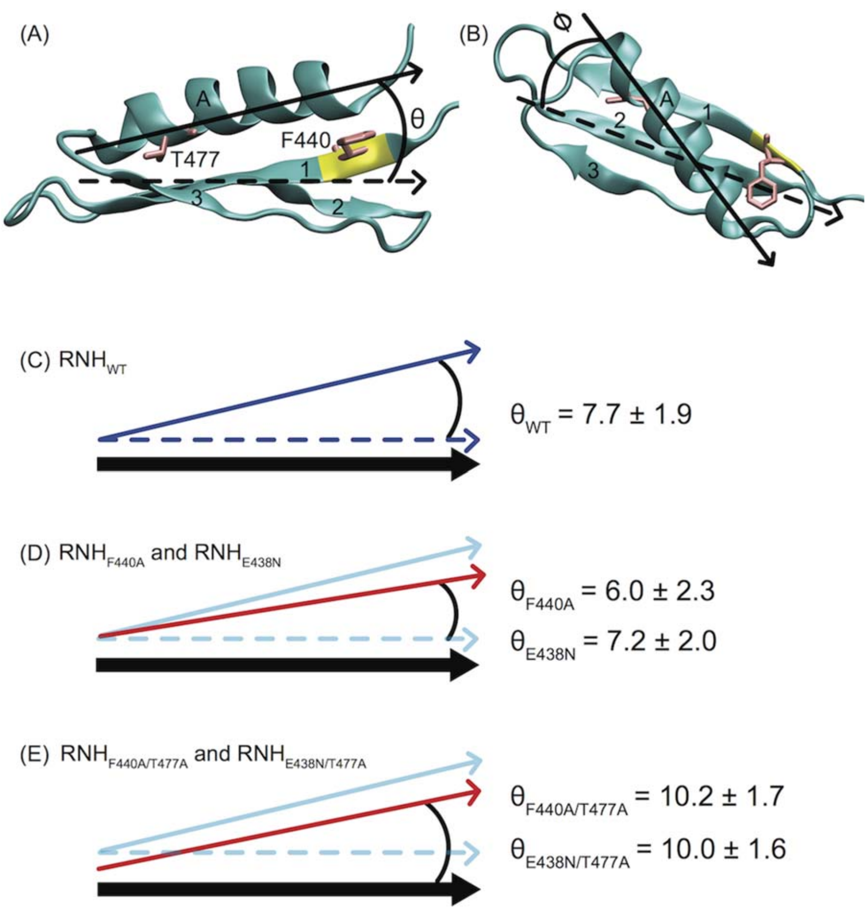

Figure 6

Structural geometry for investigating crosstalk between sites 440 and 477 via MD. The orientation of α-helix A (Cα atoms) relative to the inertial axes of the beta sheet (β strands 1–3, backbone atoms listed in Table V is characterized by angles (**A**) θand (**B**) ϕ. The average θ angle resulting from simulation of (**C**) RNH_{WT} was 7.7 ± 1.9°. (**D**) The angle θ decreased slightly in the RNH_{F440A} and RNH_{E438N} mutants (red arrow) under MD simulation. (**E**) By contrast, $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ resulted in increased θ values. In (C)–(E), schematically, the thick black arrow represents the position of the β sheet. The blue solid and dashed arrows represent the position of α-helix A and the position of the β sheet inertial axis, respectively, for the RNH_{WT} simulation. In (D) and (E), the red arrows indicate the position of α-helix A for the indicated simulations. In (D), because the simulation was not converged, the change of the angle is likely a transient effect of the conformational change within the 200 ns simulation. In (E), because the size of residue 477 decreases, the starting position of the helix is drawn in the cartoon differently from that in (D).

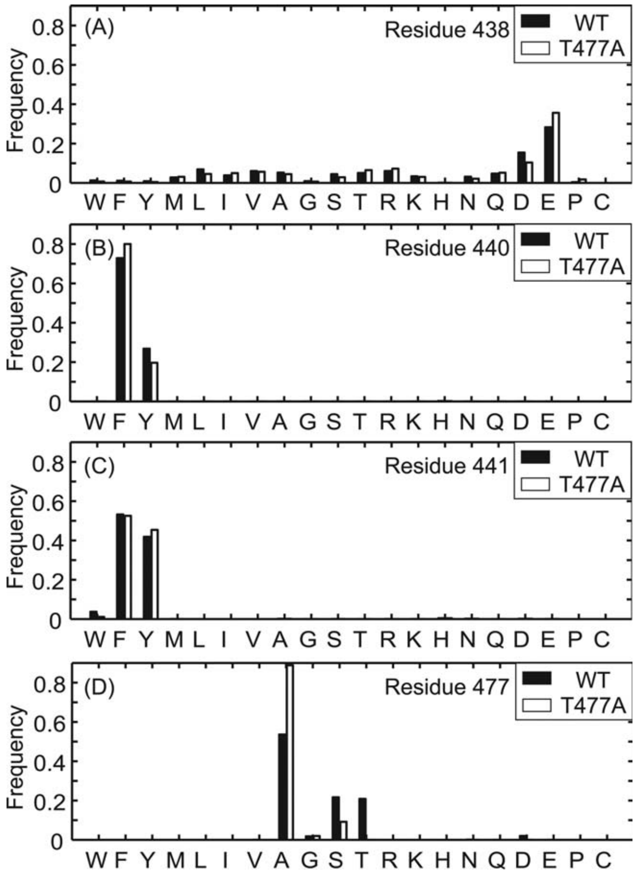

Figure 7

Prediction of sequence tolerance for the protein folding at residues (**A**) 438, (**B**) 440, (**C**) 441, and (**D**) 477 for RNH WT (filled bars) and RNH_{T477A} (open bars) coordinates (see the Materials and Methods).

(Figs. 2 and 4). Nevertheless, the sequence tolerance calculations using the RNH_{T477A} provided essentially the same results (>92%) as those using the WT for residues 438, 440, and 441 (Fig. 7). Since the results were similar even when the Rosettabackrub calculation was done using coordinates from the MD simulations at the 100 ns time point (Supporting Information Fig. S4), the observed tendencies of the residue preferences appear to be insensitive to fine details of the structure used.

DISCUSSION

The structural behavior of the p51-RNH processing site in the RNH domain is not well understood. Although the RNH domain is rigidly folded in the known crystal structures of the isolated RNH fragment and the p51/p66 RT heterodimer 9–14,[51,52] p51-RNH processing would be very inefficient if the processing site were located within a structured domain.[53] Indeed, other protease processing sites in HIV-encoded polyproteins, such as N and C-terminus of MA, CA, PR, RT (the N-terminus and the C-terminus of the p66 and p51 subunits), and IN, are exposed to solution at least in the monomer forms.[13,54–57] Thus, it has been hypothesized that the RNH domain may be unfolded or in another conformation in the pre-matured form in RT, that is, p66 monomer or p66 homodimer.[2,9,10,58,59] Alternatively, even though the isolated RNH domain is stably folded in solution,[14,43,44] the structure may have a potential plasticity to allow conformational change for the protease processing at the p51-RNH site.

To further investigate structural characteristics of the p51-RNH processing site, our study employed a combination of experimental and computational methods to enhance understanding of previously observed phenotypic changes arising from processing site mutations including the impact of the revertant mutation.[15,17] Since the isolated HIV-1 RNH domain itself is not enzymatically active, comparison of the activity cannot be examined.[60] Our experimental data show that the processing site mutations, RNH_{F440A} and RNH_{E438N}, result in substantial unfolding of the protein while the revertant, that is, $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$, are significantly folded. Simulations show that the processing site mutations cause changes in side-chain packing and the hydrogen bond network (Table II). Further, simulations of $RNH_{F440A/T477A}$ and $RNH_{E438N/T477A}$ indicate that the T477A mutation shifts the position of helix A relative to the first three β-sheets.

In HIV-1, the T477A variant is extremely common in subtypes F and G (60% and 43%, respectively), in contrast to its rarity in subtype B (1.9%).[61] According to the RNH sequence database, eight residues in the RNH domain of subtypes F and G are highly different from those in B: T477A, R463K, V466I, D471E, H/Y483Q, L491S, K/Q512R, and A534S. These mutation sites are not located between α-helix A and the β-sheet that includes β−strand 1, 2, and 3, except for R463K in the β3−strand. Indeed, residue R463 is a hydrogen bonding partner of E438. Thus, in subtypes F and G, disruption of the hydrogen bond by R463K, which likely causes similar unfolding of the protein as in the case of E438N, may be rescued by the additional T477A mutation. This evaluation of the sequence variations in the subtypes F/G provides a consistent cooperative mutation effect to the above observation.

Together with the NMR, DSF, and MD simulations, the sequence tolerance calculation demonstrates that the processing site residues are important to maintain the RNH structure, that is, buried in the protein core. However, these residues are also preferred as a protease substrate. It is known the P1 and P1' sites of the HIV-1 protease are mostly occupied by bulky side chains: the P1 site amino acids in the HIV-1 substrate are Phe, Leu, Asn, Met, Tyr, and the P1' amino acids are Phe, Pro, Leu, Tyr, Ala, Met (with frequent residues listed first).[62] The P1 and P1' sites for the p51-RNH processing are occupied by F440 and Y441, respectively. Compared with other protease cleavage sites in HIV-1 polyproteins, polymorphism at the p51-RNH site is small with high conservation of the processing site residues from E438 to V442.[42,62] Based on our analyses, this is because the residues are needed for the structural stability. Taken together, our results demonstrate that having F440 and Y441 as P1 and P1', respectively, is important for the substrate specificity as well as the structural stability of the protein core.

The structural behavior of RNH in the context of the RT dimers remains to be clarified. As a substrate cleaved by the viral protease, the RNH processing site has to be accessible to the solvent. On the other hand, as revealed above, the p51-RNH processing residues are well arranged in the protein core to maintain RNH folding. Such a coupling of opposite characteristics, substrate specificity and the structural stability, is puzzling. One possible explanation may be, as hypothesized and proposed previously, the RNH domain is unfolded or in an extended conformation in the RT precursor[2,9,10,58,59] However, if so, since the secondary and tertiary processing sites within the RNH are not protected,[1,15,53,63,64] the p51 subunit must have variation of the amino acid lengths. Based on the changes in relative fraction of the hydrogen bond network at the processing site (Tables III and IV) and our recent observation of the RNH fold in the p66/p66 homodimer,[65] we rather postulate a model in which the hydrogen-bond network may be weakened in the p66 homodimer, possibly due to fluctuation in the domain linker orientation or by the protease interaction to the linker region, increasing the population of the minor open conformation. Indeed, the hydrogen bond network observed in the simulation was not static but exhibited significant fluctuations (Table III).

CONCLUSION

Our combined NMR and computational results fill in missing pieces of the HIV RT structural story. Our data explain why p66 did not accumulate when the p51-RNH processing site was mutated, and how the revertant mutation, T477A, was able to restore RNH folding, leading to normal proteolytic processing to the p66/p51 heterodimer despite the continued presence of the p51-RNH processing site mutations. A plausible "levering" mechanism for the crosstalk between the region around T477 and the processing site has been proposed based on a total of >1 μs of all-atom MD simulation studies. Sequence tolerance calculations, as well as MD simulations and the NMR experiments, indicate that the P1 residue, F440, that is critical for substrate specificity is also important for the RNH folding; this observation is consistent with the fact that the p51-RNH processing site in the matured RT is protected within the protein core.

ACKNOWLEDGMENTS

The authors thank Maria DeLucia for technical assistance in cloning, protein expression, and purification, Martin Christen for support of the NMR data analysis, and Troy Krzysiak for guidance and assistance of the Differential Scanning Fluorimetry experiment.

REFERENCES

1. Chattopadhyay D, Evans DB, Deibel MR, Vosters AF, Eckenrode FM, Einspahr HM, Hui JO, Tomasselli AG, Zurcher-Neely HA, Heinrikson RL, Sharma SK. Purification and characterization of heterodimeric human immunodeficiency virus type 1 (HIV-1) reverse transcriptase produced by in vitro processing of p66 with recombinant HIV-1 protease. J Biol Chem 1992; 267:14227–14232.
2. Sharma SK, Fan N, Evans DB. Human immunodeficiency virus type 1 (HIV-1) recombinant reverse transcriptase. Asymmetry in p66 subunits of the p66/p66 homodimer. FEBS Lett 1994; 343:125–130.
3. Katz RA, Skalka AM. The retroviral enzymes. Ann Rev Biochem 1994; 63:133–173.
4. Divita G, Rittinger K, Geourjon C, Deleage G, Goody RS. Dimerization kinetics of HIV-1 and HIV-2 reverse transcriptase: a two step process. J Mol Biol 1995; 245:508–521.
5. Coffin JM, Hughes SH, Varmus HE. Retroviruses. Plainview, NY: Cold Spring Harbor Laboratory Press; 1997
6. Sluis-Cremer N, Arion D, Abram ME, Parniak MA, Proteolytic processing of an HIV-1 pol polyprotein precursor: insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. Int J Biochem Cell Biol 2004; 36:1836–1847.
7. Hizi A, Herschhorn A. Retroviral reverse transcriptases (other than those of HIV-1 and murine leukemia virus): a comparison of their molecular and biochemical properties. Virus Res 2008; 134:203–220.
8. Herschhorn A, Hizi A. Retroviral reverse transcriptases. Cell Mol Life Sci 2010; 67:2717–2747.
9. Davies JFn, Hostomska Z, Hostomsky Z, Jordan SR, Matthews DA. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. Science 1991; 252:88–95.
10. Jacobo-Molina A, Arnold E. HIV reverse transcriptase structure-function relationships. Biochemistry 1991; 30:6351–6356.
11. Hostomska Z, Matthews DA, Davies JFn, Nodes BR, Hostomsky Z. Proteolytic release and crystallization of the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase. J Biol Chem 1991; 266:14697–14702.
12. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 1992; 256:1783–1790.
13. Jacobo-Molina A, Ding J, Nanni RG, Clark ADJ, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P, Hizi A, Hughes SH, Arnold E. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc Natl Acad Sci USA 1993; 90: 6320–6324.
14. Pari K, Mueller GA, DeRose EF, Kirby TW, London RE. Solution structure of the RNase H domain of the HIV-1 reverse transcriptase in the presence of magnesium. Biochemistry 2003; 42:639–650.
15. Abram ME, Parniak MA. Virion instability of human immunodeficiency virus type 1 reverse transcriptase (RT) mutated in the protease cleavage site between RT p51 and the RT RNase H domain. J Virol 2005; 79:11952–11961.
16. Boso G, Örvell C, Somia NV. The nature of the N-terminal amino acid residue of HIV-1 RNase H is critical for the stability of reverse transcriptase in viral particles. J Virol 2015; 89:1286–1297.
17. Abram ME, Sarafianos SG, Parniak MA. The mutation T477A in HIV-1 reverse transcriptase (RT) restores normal proteolytic

processing of RT in virus with Gag-Pol mutated in the p51-RNH cleavage site. Retrovirology 2010; 7:6.
18. Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. J Mol Biol 2008; 380:742–756.
19. Smith CA, Kortemme T. Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible backbone design. PLoS One 2011; 6:e20451.
20. Christen MT, Menon L, Myshakina NA, Ahn J, Parniak MA, Ishima R. Structural basis of the allosteric inhibitor interaction on the HIV-1 reverse transcriptase RNase H domain. Chem Biol Drug Des 2012; 80:706–716.
21. Ahn J, Byeon IJ, Dharmasena S, Huber K, Concel J, Gronenborn AM, Sluis-Cremer N. The RNA binding protein HuR does not interact directly with HIV-1 reverse transcriptase and does not affect reverse transcription in vitro. Retrovirology 2010; 7:40.
22. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - a multidimensional spectral processing system based on unix pipes. J Biomol NMR 1995; 6:277–293.
23. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinás M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 2005; 59:687–696.
24. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P Salemme FR. High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen 2001; 6:429–440.
25. Lavinder JJ, Hari SB, Sullivan BJ, Magliery TJ. High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J Am Chem Soc 2009; 131:3794–3795.
26. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2007; 2:2212–2221.
27. Brooks BR Bruccoleri RE, Olafson BD, States DJ, Swaminathan S Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983; 4: 187–217.
28. Brooks BR, Brooks CL, MacKerell III, Nilsson ADM, Jr., Petrella L, Roux RJ, Won B, Archontis Y, Bartels G, Boresch C, Caflisch S, Caves A, Cui L, Dinner AR Q, Feig Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W York DM. Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009; 30:1545–1614.
29. Kalé LV, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys 1999; 151:283–312.
30. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005; 26:1781–1802.
31. Hsiou Y, Ding J, Das K, Clark ADJ, Hughes SH, Arnold E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. Structure 1996; 4:853–860.
32. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD, Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. J Chem Theory Comput 2012; 8:3257–3273.
33. Best RB, Mittal J, Feig M, MacKerell AD, Jr. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. Biophys J 2012; 103:1045–1051.
34. Lee MS, Salsbury FR, Brooks IIICL. Novel generalized Born methods. J Chem Phys 2002; 116:10606–10614.
35. Lee MS, Feig M, Salsbury FR, Brooks CL, III. New analytical approximation to the standard molecular volume definition and its application to generalized Born calculations. J Comput Chem 2003: 25;265–284.
36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983; 79:926–935.
37. Grubmüller H, Heymann B, Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 1996; 271:997–999.
38. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977; 23:327–341.
39. Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 1992; 13:952–962.
40. Aqvist J. A simple way to calculate the axes of an α-helix. Comput Chem 1986; 10:97–99.
41. Lauck F, Smith CA, Friedland GF, Humphris EL, Kortemme T. RosettaBackrub–a web server for flexible backbone protein structure modeling and design. Nucleic Acids Res 2010; 38:W569–W575.
42. Humphris-Narayanan E, Akiva E, Varela R, Ó Conchúir S, Kortemme T. Prediction of mutational tolerance in HIV-1 protease and reverse transcriptase using flexible backbone protein design. PLoS Comput Biol 2012; 8:e1002639.
43. Powers R, Clore GM, Stahl SJ, Wingfield PT, Gronenborn AM. Analysis of the backbone dynamics of the ribonuclease H domain of the human immunodeficiency virus reverse transcriptase using nitrogen-15 relaxation measurements. Biochemistry 1992; 31:9150–9157.
44. Yan J, Wu H, Tom T, Brodsky O, Maegley K. Targeting divalent metal ions at the active site of the HIV-1 RNase H domain: NMR studies on the interactions of divalent metal ions with RNase H and its inhibitors. Am J Anal Chem 2011; 2:639–649.
45. Chung S, Miller JT, Lapkouski M, Tian L, Yang W, Le Grice SF. Examining the role of the HIV-1 reverse transcriptase p51 subunit in positioning and hydrolysis of RNA/DNA hybrids. J Bio Chem 2013; 288:16177–16184.
46. Dunn LL, Boyer PL, Clark PK, Hughes SH. Mutations in HIV-1 reverse transcriptase cause misfolding and miscleavage by the viral protease. Virology 2013; 444:241–249.
47. Freddolino PL, Harrison CB, Liu Y, Schulten K. Challenges in protein-folding simulations. Nat Phys 2010; 6:751–758.
48. Grossfield A, Zuckerman DM. Quantifying uncertainty and sampling quality in biomolecular simulations. Annu Rep Comput Chem 2009; 5:23–48.
49. Zuckerman DM. Equilibrium sampling in biomolecular simulation. Annu Rev Biophys 2011; 40:41–62.
50. van der Vaart A. Simulation of conformational transitions. Theor Chem Acc 2006; 116:183–193.
51. Jäger J, Smerdon S, Wang J, Boisvert DC, Steitz TA. Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. Structure 1994; 2:869–876.
52. Unge T, Knight S, Bhikhabhai R, Lövgren S, Dauter Z, Wilson K, Strandberg B. 2.2 A resolution structure of the amino-terminal half of HIV-1 reverse transcriptase (fingers and palm subdomains). Structure 1994; 2:953–961.
53. Tomasselli AG, Sarcich JL, Barrett LJ, Reardon IM, Howe WJ, Evans DB, Sharma SK, Heinrikson RL. Human immunodeficiency virus type-1 reverse transcriptase and ribonuclease H as substrates of the viral protease. Protein Sci 1993; 2:2167–2176.
54. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem 1993; 62:543–585.
55. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase:

similarity to other polynucleotidyl transferases. Science 1994; 266: 1981–1986.
56. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc Natl Acad Sci USA 1996; 93:3099–3104.
57. Momany C, Kovari LC, Prongay AJ, Keller W, Gitti RK, Lee BM, Gorbalenya AE, Tong L, McClure J, Ehrlich LS, Summers MF, Carter C, Rossmann MG. Crystal structure of dimeric HIV-1 capsid protein. Nat Struct Biol 1996; 3:763–770.
58. Wang J, Smerdon SJ, Jager J, Kohlstaedt LA, Rice PA, Friedman JM, Steitz TA. Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. Proc Natl Acad Sci USA 1994; 91:7242–7246.
59. Zheng X, Pedersen LC, Gabel SA, Mueller GA, Cuneo MJ, Derose EF, Krahn JM London RE. Selective unfolding of one Ribonuclease H domain of HIV reverse transcriptase is linked to homodimer formation. Nucleic Acids Res 2014; 42:5361–5377.
60. Keck JL, Marqusee S. Substitution of a highly basic helix/loop sequence into the RNase H domain of human immunodeficiency virus reverse transcriptase restores its Mn(2+)-dependent RNase H activity. Proc Natl Acad Sci USA 1995; 92:2740–2744.
61. Rhee SY, Kantor R, Katzenstein DA, Camacho R, Morris L, Sirivichayakul S, Jorgensen L, Brigido LF, Schapiro JM, Shafer RW; Group ftiNSBH-W. HIV-1 pol mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes. AIDS 2006:20;643–651.
62. Côté HC, Brumme ZL, Harrigan PR. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. J Virol 2001; 75:589–594.
63. Schulze T, Nawrath M, Moelling K. Cleavage of the HIV-1 p66 reverse transcriptase/RNase H by the p9 protease in vitro generates active p15 RNase H. Arch Virol 1991; 118:179–188.
64. Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. 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 2005; 2:66.
65. Sharaf NG, Poliner E, Slack RL, Christen MT, Byeon IJ, Parniak MA, Gronenborn AM, Ishima R. The p66 immature precursor of HIV-1 reverse transcriptase. Proteins 2014; 82:2343–2352.