Supplemental information.
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*], Rieko Ishima[1*]

Departments of Structural Biology[1], and Computational and System Biology[2], Microbiology and Molecular Genetics[3], University of Pittsburgh School of Medicine, Pittsburgh, PA 15260.

[†] Ryan L. Slack and Justin Spiriti contributed equally to this work,

[*] Corresponding authors:
Rieko Ishima, Room 1037, Biomedical Science Tower 3, 3501 Fifth Avenue, Pittsburgh, Pennsylvania 15260; Tel: 412-648-9056; Fax: 412-648-9008; Email: ishima@pitt.edu

Daniel M. Zuckerman, Room 3079, Biomedical Science Tower 3, 3501 Fifth Avenue, Pittsburgh, Pennsylvania 15260, United States; Email: ddmmzz@pitt.edu

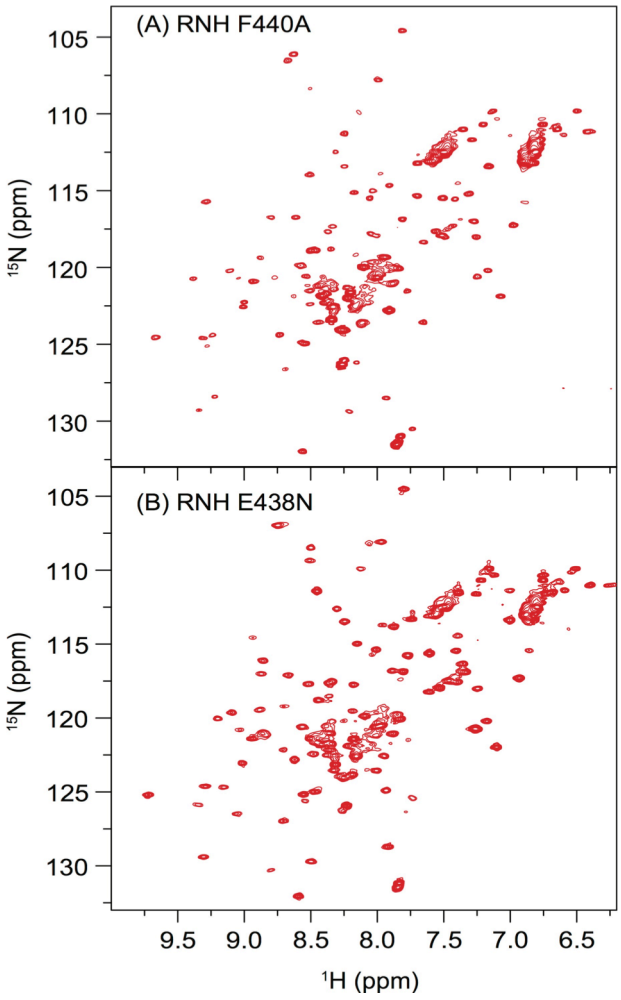

Figure S1. ^{1}H-^{15}N HSQC spectra of (A) the RNH_{F440A} mutant and (B) the RNH_{E438N} mutant, recorded at pH 8.

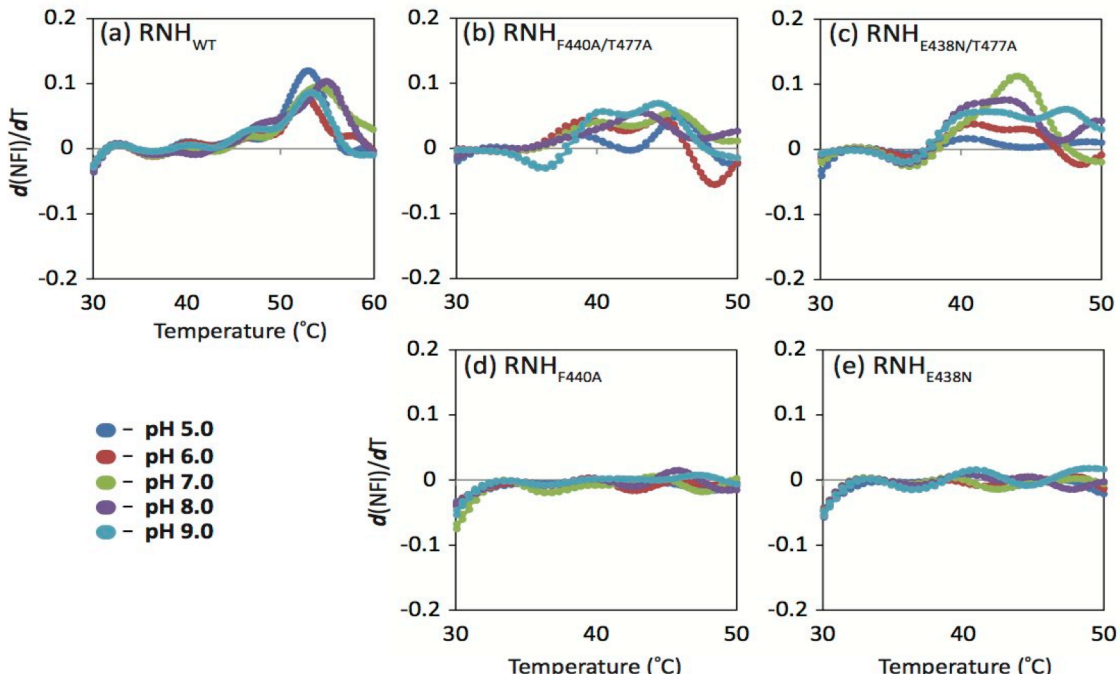

Figure S2. The first derivative of normalized fluorescence intensity signals obtained by Differential Scanning Fluorimetry, *d(NFI)/dT*, as a function of temperature for (A) WT RNH, (B) $RNH_{F440A/T477A}$, (C) $RNH_{E438N/T477A}$, (D) RNH_{F440A}, and (E) RNH_{E438N} mutants at pH 5.0, 6.0, 7.0, 8.0, and 9.0. The peak of the *d(NFI)/dT* was clearly observed for WT RNH, $RNH_{F440A/T477A}$, and $RNH_{E438N/T477A}$. However, no change of the derivative was observed for RNH_{F440A} and RNH_{E438N}.

Weighted Ensemble Simulations

Because of the challenges of sampling RNH using traditional MD simulations, we also performed Weighted Ensemble (WE) simulations [1-5]. WE is a specialized technique capable of precisely characterizing low-probability events occurring on short timescales. Two independent WE simulations were performed for WT and $RNH_{F440A/T477A}$, while six were performed for the RNH_{F440A} to examine the 1 ns timescale in Figure S3. The WE simulations unambiguously show a wider probability distribution for RNH_{F440A}, compared to those of the WT and $RNH_{F440A/T477A}$.

Methodologically, WE is a method for orchestrating an ensemble of many MD trajectories with statistical weights, subject to occasional replication and pruning (every 10 ps in the present study) to ensure coverage of a pre-chosen "progress coordinate" (RMSD, in this case). WE is an unbiased approach with an established statistical basis that is ideal for modern parallel computing environments [3]. All the simulations were started from the same equilibrated structures used to start the MD simulations. These simulations were used to calculate the probability of a substantial conformational change taking place within a short timescale. Figure S3 shows probability distributions averaged over the time window from 0.5 to 1 ns.

1. Huber, G.A. and Kim, S. (1996) Weighted-ensemble Brownian dynamics simulations for protein association reactions. Biophys J 70: 97.

2. Zhang, B.W., Jasnow, D., and Zuckerman, D.M. (2007) Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. Proc Natl Acad Sci U S A, 104: 18043-8.

3. Zhang, B.W., Jasnow, D., and Zuckerman, D.M. (2010) The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. J Chem Phys, 132: 054107.

4. Bhatt, D. and Zuckerman, D.M. (2010) Heterogeneous path ensembles for conformational transitions in semi-atomistic models of adenylate kinase. J Chem Theory Comput, 6: 3527-38.

5. Bhatt, D., Zhang, B.W., and Zuckerman, D.M. (2010) Steady-state simulations using weighted ensemble path sampling. J Chem Phys, 133: 014110.

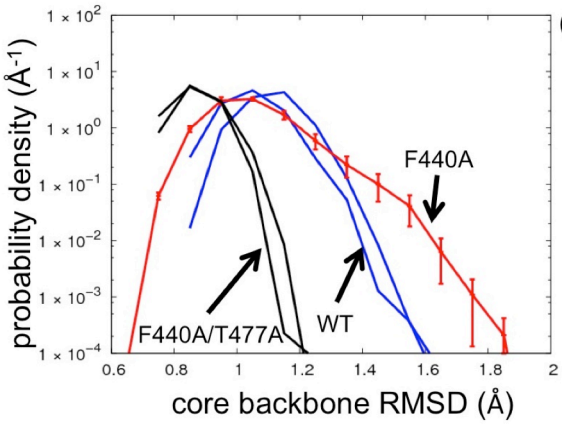

Figure S3. Variations in the short-time conformational distribution of WT and mutants based on weighted ensemble (WE) simulations. Probability distributions of the backbone RMSD (involving N, Cα, and C atoms of residues 427-556) with respect to the starting structure for WT (blue), $RNH_{F440A/T477A}$ (black) and RNH_{F440A} (red) are shown based the 0.5 – 1.0 ns time interval following initiation from identical starting structures. For the RNH_{F440A} simulations, the mean probability and the standard error are shown.

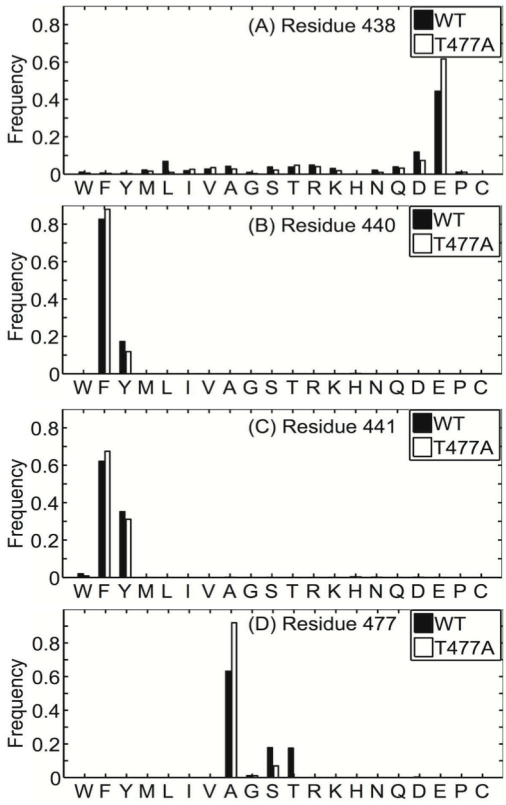

Figure S4. Prediction of sequence tolerance for the protein folding at residues (A) 438, (B) 440, (C) 441 and (D) 477 for RNH WT (filled bar) and RNHT477A (open bar) , calculated using the structural coordinates at a 100 ns simulation point.