Published in final edited form as:

J Chem Theory Comput. 2016 June 14; 12(6): 2489–2492. doi:10.1021/acs.jctc.6b00120.

Thermostability of Enzymes from Molecular Dynamics Simulations

Tim Zeiske[†], Kate A. Stafford[†], and Arthur G. Palmer III^{†,*}

[†]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

Abstract

Thermodynamic stability is a central requirement for protein function and one goal of protein engineering is improvement of stability, particularly for applications in biotechnology. Herein, molecular dynamics simulations are used to predict *in vitro* thermostability of members of the bacterial Ribonuclease HI (RNase H) family of endonucleases. The temperature dependence of the generalized order parameter, S , for four RNase H homologs, from psychrotrophic, mesophilic and thermophilic organisms, is highly correlated with experimentally determined melting temperatures and with calculated free energies of folding at the midpoint temperature of the simulations. This study provides an approach for *in silico* mutational screens to improve thermostability of biologically and industrially relevant enzymes.

Graphical abstract

Understanding protein stability, and more specifically thermostability, has long been of interest in structural biology and biophysics, but also in biotechnology.¹ Thus, thermostable enzymes may be useful catalysts for industrial processes run at relatively high temperatures,

Corresponding Author: agp6@cumc.columbia.edu.

ASSOCIATED CONTENT

Supporting Information

Additional details and supporting figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

The authors declare no competing financial interests.

which has many advantages, such as increased reaction rate, increased solubility of reactants, and reduced contaminating microbial growth^{1–3}.

The unfolding, or melting temperature, T_m is a metric for the thermostability of a protein. T_m is the temperature at which the Gibbs free energy G of the folded and unfolded states, and thus their populations, are equal. While attempts have been made to determine T_m by molecular dynamics simulations, computational limitations usually limit the study of complete unfolding processes to very small and fast-folding proteins.^{4,5} Computational tools to predict protein stability fall into two main categories. The energy of unfolding (ΔG) is studied using physical, statistical or empirical potentials in the first instance, or using machine-learning methods trained on datasets of experimental unfolding energies in the second^{6–8}. Promising results have been reported recently by applying a Monte Carlo simulation approach to the bacterial enzyme dihydrofolate reductase.⁹ A high-throughput approach has been developed for finding multiple stabilizing mutations of a protein.¹⁰

Ribonuclease HI (RNase H; EC 3.1.26.4) enzymes, have been studied extensively to shed light on thermostability, because highly conserved structural homologs exist in both psychrotrophic and thermophilic organisms.^{11–24} These enzymes nonspecifically cleave the RNA strand of RNA:DNA hybrid substrates,²⁵ and have been implicated in many biological processes, including removal of R-loops, removal of Okazaki fragments, synthesis of multicopy single-stranded DNA, and removal of ribonucleotides misincorporated into the genome.^{26,13}

To study the influence of temperature on the dynamical behavior of RNase H variants *in silico*, we conducted Molecular Dynamics (MD) simulations for RNase H enzymes from four different organisms, as shown in Table 1.¹³ The set is composed of proteins from psychrotrophic (soRNH), mesophilic (ecRNH), moderately thermophilic (ctRNH), and thermophilic (ttRNH), bacteria. The protein ctRNH, despite its origins in the proteome of a moderate thermophile, has a melting temperature T_m that is similar to that of the mesophile protein ecRNH.

The four proteins were simulated for 100 ns at 273, 300 and 340 K. The preparation of the simulations has been described previously.^{12,13,27,28} Each trajectory was divided into 10 ns blocks to reflect global tumbling time^{12,29} and the squares of the generalized order parameters (S^2) for the backbone NH bond vector (henceforth simply “order parameters”) were calculated for each block and then averaged over all blocks. The order parameters describe the orientational fluctuations of the backbone NH vectors and can be measured experimentally by NMR spectroscopy.³⁰ Order parameters for all proteins at the three studied temperatures are shown in the Supporting Figure S1. All order parameters were scaled by $\xi = (1.02/1.04)^6 \approx 0.89$ to account for zero point vibrational motions of the NH bond vectors.^{30,31}

The temperature dependence of S^2 is described by the dimensionless parameter Λ , which is directly related to the temperature-dependent effective potential for the fluctuations of the NH bond vector, and together with S^2 provides information about the contributions of these fluctuations to heat capacity:^{35–37}

$$\Lambda = \frac{d \ln (1 - S)}{d \ln T}$$

The Λ values determined from experimental order parameters for *E. coli* RNase H and the villin headpiece helical subdomain HP36 exceed simulated values by a factor of about 2^{35,36}. Even though agreement between experimental and simulated order parameters is typically very good ($R = 0.7 - 0.9$)¹³, small systematic differences in the order parameter can have a larger influence on Λ , as described previously³⁵. This discrepancy does not affect the approach described herein as long as comparisons are between simulations using identical protocols.

Supporting Figure S2 shows plots of $\ln(1-S)$ versus $\ln T$ for select residues in *E. coli* RNase H. Linear regression was used to extract Λ values as the slope of these plots for each residue. The quality of the fit was assessed by the correlation coefficient R . Incorporating simulations at additional temperatures has only minor effects on resulting Λ values (Supporting Figure S3).

Figure 1 shows Λ and R values as a function of sequence for the four studied proteins. For most residues the linear regression yielded R values close to unity. The majority of the low R values correspond to low Λ values and thus low temperature dependence. Accordingly, the number of residues with R values smaller than 0.8 is much smaller for ecRNH (15 out of 149) than for ttRNH (38 out of 140). The distributions of Λ for the different proteins are compared in Figure 2. A plot of Λ averaged over all residues for each organism against experimental melting temperature T_m reveals a homomorphic relationship between the two, with a lower average Λ value corresponding to a higher melting temperature (Fig. 2). A similar correlation is observed between Λ and ΔG for folding calculated at 304 K, the midpoint of the investigated temperature range, using the Gibbs-Helmholtz equation (not shown).

The linear relationship of all four data points is statistically highly significant ($p < 0.01$ for weighted least squares). Removing Λ values of residues with R values < 0.8 increases the average Λ values in a similar fashion for all proteins, without affecting the qualitative relationship between average Λ and T_m , the slope of the fitted line is -0.008 or -0.009 K^{-1} with R values of 0.98 or 0.99, if these residues are excluded or included, respectively.

To study the effect of a single thermostabilizing mutation on Λ values, we conducted another set of MD simulations on an *E. coli* variant with a thermostabilizing Glycine insertion at position 80b (mimicking this position in ttRNH); the simulated order parameters are shown in Supporting Figure 4a. This mutant has been crystallized (PDB: 1GOA) and thermodynamically characterized.^{38,39} The insertion has a small thermostabilizing effect and increases T_m by 1.2 K.^{38,39} Supplementary Figure 4b and 4c show that the correlation between Λ values for ecRNH and the iG80b mutant are extremely high for the first approximately 60 residues, while they differ quite significantly around and following (in sequence) the insertion. Thus the effects of the insertion are not restricted locally but lead to changes across the protein, presumably through changes in packing (ttRNH has a reduced

solvent accessible surface area compared to ecRNH – *ca.* 8800 Å² vs *ca.* 9500 Å² –, the iG80b mutant lies in between with *ca.* 9200 Å²). The average value for Λ decreases slightly from 0.995 to 0.965, agreeing with the increase in T_m of 1.2 K. Supplementary Figure S4d shows that including this fifth datapoint improves the statistical significance of the weighted linear regression further ($p < 0.001$).

A plot of Λ values averaged separately for α -helices, β -sheets and loops, shows that no particular secondary structure element dominates the trend versus T_m (Supporting Figure S5). Loop backbone order parameters have on average a greater temperature dependence than secondary structure order parameters. Interestingly, ctRNH has loop Λ values that are closer to those of ttRNH, which could reflect a larger contribution of loop dynamics to heat capacity. C_p of ctRNH has been shown to be closer to that of ttRNH, while the overall melting temperature is closer to that of ecRNH³³.

A plot of the Λ values of the four wild-type proteins and the iG80b mutant, accounting for insertions and deletions, (Supporting Figure S6) shows regions with strong differences in Λ values especially for the psychrotrophic (soRNH) and thermophilic (ttRNH) proteins. Figure 3 shows the structures of soRNH and ttRNH with residues with particularly high or low Λ values highlighted. These regions include helix α_B , the handle region and the β_5/α_E -loop, which have all been implicated in substrate binding.^{11,12,40} Many of these residues have been implicated in thermostability as well.^{40,41} As one example, the octapeptide LKKAFTEG in helix α_B , of ttRNH (comprising the iG80b insertion) has an average Λ value of 0.4 ± 0.2 in ttRNH and an average of 2.2 ± 0.4 for the corresponding heptapeptide MRQGIMT in soRNH. The substitution of the ttRNH octapeptide into the corresponding sequence of ecRNH leads to a ~ 5 K increase in T_m at pH 5.5.⁴¹ The observation that many of the residues with extreme values of Λ are important for both substrate binding and thermostability may reflect evolutionary trade-offs between activity and stability.

We have shown, at least within one homologous family of enzymes including proteins from psychrotrophic, mesophilic and thermophilic organisms, that a linear homomorphic relationship exists between experimentally determined melting temperature and the average of the simulated parameter Λ , describing the temperature dependence of backbone conformational fluctuations of amide bond vectors. This result suggests that Λ is a good proxy for thermostability within a family of homologous proteins. Lastly, this study suggests the possibility for *in silico* identification of thermostabilizing mutations within a protein family, without the need to simulate full unfolding events, as an adjunct to other computational approaches for protein design.¹⁰

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by an NSF graduate research fellowship (K.A.S.) and NIH grant GM50291 (A.G.P.). We thank the Center for Computational Biology and Bioinformatics (C2B2) for computational resources.

REFERENCES

1. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. *Nature*. 2012; 485:185–194. [PubMed: 22575958]
2. Zamost BL, Nielsen HK, Starnes RL. Thermostable enzymes for industrial applications. *J. Ind. Microbiol.* 1991; 8:71–81.
3. Kristjansson JK. Thermophilic organisms as sources of thermostable enzymes. *Trends Biotechnol.* 1989; 7:349–353.
4. Yang JS, Wallin S, Shakhnovich EI. Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:895–900. [PubMed: 18195374]
5. Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:5915–5920. [PubMed: 23503848]
6. Khan S, Vihinen M. Performance of protein stability predictors. *Hum. Mutat.* 2010; 31:675–684. [PubMed: 20232415]
7. Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng. Des. Sel.* 2009; 22:553–560. [PubMed: 19561092]
8. Thiltgen G, Goldstein RA. Assessing predictors of changes in protein stability upon mutation using self-consistency. *PloS One*. 2012; 7:e46084. [PubMed: 23144695]
9. Tian J, Woodard JC, Whitney A, Shakhnovich EI. Thermal stabilization of dihydrofolate reductase using Monte Carlo unfolding simulations and its functional consequences. *PLoS Comput. Biol.* 2015; 11:e1004207. [PubMed: 25905910]
10. Wijma HJ, Floor RJ, Jekel PA, Baker D, Marrink SJ, Janssen DB. Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* 2014; 27:49–58. [PubMed: 24402331]
11. Palmer AG. Enzyme dynamics from NMR spectroscopy. *Acc. Chem. Res.* 2015; 48:457–465. [PubMed: 25574774]
12. Stafford KA, Trbovic N, Butterwick JA, Abel R, Friesner RA, Palmer AG. Conformational preferences underlying reduced activity of a thermophilic ribonuclease H. *J. Mol. Biol.* 2015; 427:853–866. [PubMed: 25550198]
13. Stafford KA, Robustelli P, Palmer AG. Thermal adaptation of conformational dynamics in ribonuclease H. *PLoS Comput. Biol.* 2013; 9:e1003218. [PubMed: 24098095]
14. Butterwick JA, Palmer AG. An inserted Gly residue fine tunes dynamics between mesophilic and thermophilic ribonucleases H. *Protein Sci.* 2006; 15:2697–2707. [PubMed: 17088323]
15. Butterwick JA, Loria JP, Astrof NS, Kroenke CD, Cole R, Rance M, Palmer AG. Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J. Mol. Biol.* 2004; 339:855–871. [PubMed: 15165855]
16. Kanaya, S. Ribonucleases H. Crouch, R.; Toulmé, J., editors. INSERM; 1998. p. 1-38.
17. Kroenke CD, Loria JP, Lee LK, Rance M, Palmer AG. Longitudinal and transverse ^1H – ^{15}N dipolar/ ^{15}N chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J. Am. Chem. Soc.* 1998; 120:7905–7915.
18. Kroenke CD, Rance M, Palmer AG. Variability of the ^{15}N chemical shift anisotropy in *Escherichia coli* ribonuclease H in solution. *J. Am. Chem. Soc.* 1999; 121:10119–10125.
19. Mandel AM, Akke M, Palmer AG. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 1995; 246:144–163. [PubMed: 7531772]
20. Mandel AM, Akke M, Palmer AG. Dynamics of ribonuclease H: temperature dependence of motions on multiple time scales. *Biochemistry*. 1996; 35:16009–16023. [PubMed: 8973171]
21. Hollien J, Marqusee S. Structural distribution of stability in a thermophilic enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:13674–13678. [PubMed: 10570131]
22. Hollien J, Marqusee S. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry*. 1999; 38:3831–3836. [PubMed: 10090773]

23. Hollien J, Marqusee S. Structural distribution of stability in a thermophilic enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:13674–13678. [PubMed: 10570131]
24. Kanaya S, Itaya M. Expression, purification, and characterization of a recombinant ribonuclease H from *Thermus thermophilus* HB8. *J. Biol. Chem.* 1992; 267:10184–10192. [PubMed: 1315754]
25. Hostomsky, Z.; Hostomska, Z.; Mathews, D. Nucleases. Linn, S.; Roberts, R., editors. Cold Spring Harbor Laboratory Press; 1993. p. 341-76.
26. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* 2009; 276:1494–1505. [PubMed: 19228196]
27. Stafford KA, Palmer AG. Evidence from molecular dynamics simulations of conformational preorganization in the ribonuclease H active site. *F1000 Research.* 2014; 3:67. [PubMed: 25075292]
28. Stafford KA, Ferrage F, Cho J-H, Palmer AG. Side chain dynamics of carboxyl and carbonyl groups in the catalytic function of *Escherichia coli* ribonuclease H. *J. Am. Chem. Soc.* 2013; 135:18024–18027. [PubMed: 24219366]
29. Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, Arkin IT, Jensen MØ, Xu H, Trbovic N, Friesner RA, Palmer AG, Shaw DE. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J. Phys. Chem. B.* 2008; 112:6155–6158. [PubMed: 18311962]
30. Zeiske T, Stafford KA, Friesner RA, Palmer AG. Starting-structure dependence of nanosecond timescale intersubstate transitions and reproducibility of MD-derived order parameters. *Proteins.* 2013; 81:499–509. [PubMed: 23161667]
31. Case DA. Calculations of NMR dipolar coupling strengths in model peptides. *J. Biomol. NMR.* 1999; 15:95–102. [PubMed: 10605083]
32. Tadokoro T, You DJ, Abe Y, Chon H, Matsumura H, Koga Y, Takano K, Kanaya S. Structural, thermodynamic, and mutational analyses of a psychrotrophic RNase HI. *Biochemistry.* 2007; 46:7460–7468. [PubMed: 17536836]
33. Ratcliff K, Corn J, Marqusee S. Structure, stability, and folding of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: Comparison with thermophilic and mesophilic homologues. *Biochemistry.* 2009; 8:5890–5898. [PubMed: 19408959]
34. Hollien J, Marqusee S. A Thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry.* 1999; 38:3831–3836. [PubMed: 10090773]
35. Vugmeyster L, Trott O, James McKnight C, Raleigh DP, Palmer AG. Temperature-dependent dynamics of the villin headpiece helical subdomain, an unusually small thermostable protein. *J. Mol. Biol.* 2002; 320:841–854. [PubMed: 12095260]
36. Massi F, Palmer AG. Temperature dependence of NMR order parameters and protein dynamics. *J. Am. Chem. Soc.* 2003; 125:11158–11159. [PubMed: 16220912]
37. Johnson E, Palmer AG, Rance M. Temperature dependence of the NMR generalized order parameter. *Proteins.* 2007; 66:796–803. [PubMed: 17173286]
38. Ishikawa K, Nakamura H, Morikawa K, Kimura S, Kanaya S. Cooperative stabilization of *Escherichia coli* ribonuclease HI by insertion of Gly-80b and Gly-77-->Ala substitution. *Biochemistry.* 1993; 32:7136–7142. [PubMed: 8393706]
39. Ishikawa K, Kimura S, Kanaya S, Morikawa K, Nakamura H. Structural study of mutants of *Escherichia coli* ribonuclease HI with enhanced thermostability. *Protein Eng.* 1993; 6:85–91. [PubMed: 8381958]
40. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell.* 2007; 28:264–276. [PubMed: 17964265]
41. Kimura S, Nakamura H, Hashimoto T, Oobatake M, Kanaya S. Stabilization of *Escherichia coli* ribonuclease HI by strategic replacement of amino acid residues with those from the thermophilic counterpart. *J. Biol. Chem.* 1992; 267:21535–21542. [PubMed: 1328237]

Figure 1.

Λ and R values for ecRNH (black), ctRNH (green), soRNH (blue), ttRNH (red). Prolines do not have an NH bond vector and were omitted from the sequences for this figure. Green and red bars represent sheets and helices, respectively.

Figure 2. Histograms of Λ values for the four organisms with experimentally determined T_m values: soRNH (blue), ctRNH (green), ecRNH (black), ttRNH (red). The inset shows the homomorphic relationship between average Λ values and experimentally determined T_m .

Figure 3. Structures of soRNH (left) and ttRNH (right). soRNH: Residues with Λ values over 1.2 in green, residues with Λ values over 2 in blue. ttRNH: Residues with Λ values under 0.5 in orange, residues with Λ values under 0.3 in red. The iG80b in ttRNH insertion is shown as purple spheres.

Table 1

Ribonuclease H Enzymes.

Protein	PDB ID	Source Organism	T_m (°C)
soRNH	2E4L	<i>Shewanella oneidensis</i>	53.2 ³²
ctRNH	3H08	<i>Chlorobium tepidum</i>	68.5 ³³
ecRNH	2RN2	<i>Escherichia coli</i>	70.7 ³²
ttRNH	1RIL	<i>Thermus thermophilus</i>	89 ³⁴