Size-Dependent Relationships Between Protein Stability and Thermal Unfolding Temperature Have Important Implications for Analysis of Protein Energetics and High-Throughput Assays of Protein-Ligand Interactions

Matthew D. Watson[1], *Jeremy Monroe*[1] *and Daniel P. Raleigh*[1,2,3*]

[1]Department of Chemistry and [2]Graduate Program in Biochemistry & Structural Biology, Stony Brook University, Stony Brook, New York 11794-3400, United States

[3]Research Department of Structural and Molecular Biology, University College London, Gower Street, London, WC1E 6BT, United Kingdom

ABSTRACT

Changes in protein stability are commonly reported as changes in the melting temperature, ΔT_M, or as changes in unfolding free energy at a particular temperature, $\Delta\Delta G°$. Using data for 866 mutants from 16 proteins we examine the relationship between $\Delta\Delta G°$ and ΔT_M. A linear relationship is observed for each protein. The slopes of the plots of ΔT_M *vs* $\Delta\Delta G°$ for different proteins scale as N^{-1} where N is the number of residues in the protein. Thus, a given change in $\Delta G°$ causes a much larger change in T_M for a small protein relative to the effect observed for a large protein. The analysis suggests that reasonable estimates of $\Delta\Delta G°$ for a mutant can be obtained by interpolating measured values of T_M. The relationship between $\Delta\Delta G°$ and ΔT_M has implications for the design and interpretation of high-throughput assays of protein-ligand binding. So-called thermal shift assays rely upon the increase in stability which results from ligand binding to the folded state. Quantitative relationships are derived which show that the observed thermal shift, ΔT_M scales as N^{-1}. Hence thermal shift assays are considerably less sensitive for ligand binding to larger proteins.

INTRODUCTION

Protein stability can be defined in a variety of ways: through the unfolding free energy at a given temperature, $\Delta G°$, *via* the transition midpoint for thermal unfolding, T_M, by the resistance to proteolysis, or by *in vivo* stability. The most rigorous definition, from a thermodynamic perspective is to define $\Delta G°$ as a function of temperature, and if needed pressure. The Gibbs-Helmholtz equation is commonly used to define thermodynamic stability, but requires determination of T_M, ΔH_M, the enthalpy of unfolding at T_M, and $\Delta C_p°$, the change in heat capacity between the unfolded and folded states. In its simplest form $\Delta C_p°$ is assumed to be independent of temperature. These parameters are often not reported for proteins and instead values of T_M and $\Delta G°$ at one temperature are given. In some cases only T_M values are reported for mutants and it would be useful to be able to estimate changes in $\Delta G°$ from changes in T_M with high reliability.

Prior analysis has suggested an interesting relationship between changes in stability, $\Delta\Delta G°$ and changes in T_M, ΔT_M. Rees and Robertson showed that these quantities are related to each other, with the effect varying inversely with size of the protein.[1] Thus a small protein is predicted to exhibit a larger change in $\Delta G°$ for a given change in T_M and vice versa. Rees and Robertson examined mutational data for three proteins of different sizes to verify this relationship. Since their initial work considerably more data has been generated on the effects of mutations on both T_M and $\Delta G°$. In addition, thermal shift assays have become a popular tool for high-throughput screening of protein ligand interactions, and the scaling of $\Delta\Delta G°$ and ΔT_M have important implications for their interpretation and design. Briefly, thermal shift assays monitor an increase in T_M for a protein in the presence of a ligand which binds preferentially to the folded state; the increase in T_M results from the increase in $\Delta G°$ caused by ligand binding.[2-4] The method is attractive for its broad applicability and amenability to high-throughput screening, especially fluorescence-based thermal

shift assays, which use nonspecific, extrinsic dyes to detect protein unfolding. Recently, methods for cellular thermal shift assays have been developed, offering the promise of high-throughput *in vivo* drug screening.[5-8] The largest thermal shifts are known to be observed at significant molar excess of ligand. It is well understood how ligand binding affects $\Delta G°$, however the quantitative relationship between changes in T_M and ligand affinity is not well understood. In addition, sensitivity is also expected to depend on protein size, although this relationship has not been examined.

The prediction that changes in T_M are related to changes in ΔG in a way which scales as N^{-1}, where N is the number of residues has interesting implications for these types of experiments, since it will impact the sensitivity of the response to ligand binding. Here we analyze stability data, $\Delta G°$ and T_M for 866 mutants of 16 proteins and demonstrate a striking correlation between $\Delta\Delta G°$ and T_M. Plotting the experimental data reveals that for each protein there is a linear relationship between $\Delta\Delta G°$ and ΔT_M with the slope varying as N^{-1} where N is the number of residues in the protein. We examine the implications for thermal shift assays of ligand binding and show that small proteins are expected to have significantly larger shifts in T_M upon ligand binding than larger proteins. The analysis also illustrates how changes in T_M can be used to estimate changes in $\Delta G°$ with reasonable accuracy.

MATERIALS AND METHODS

Data was gathered from the ProTherm database as well as literature sources. For this analysis, selection criteria required that each entry have data for T_M, and $\Delta G°$ or sufficient data to calculate these values. All proteins in the dataset exhibit thermodynamically two-state and reversible unfolding. Only stability data at a single pH was included for each set of mutants.

RESULTS AND DISCUSSION

The temperature-dependent protein stability curve is described by the modified Gibbs-Helmholtz equation with the assumption that ΔC_p° is independent of T (1)

$$\Delta G(T) = \Delta H_M \left(1 - \frac{T}{T_M}\right) + \Delta C_p^\circ \left(T - T_M - T \ln\left(\frac{T}{T_M}\right)\right) \quad (1)$$

where T_M is the midpoint of the thermal unfolding transition, ΔH_M is the change in enthalpy upon unfolding at T_M and ΔC_p° is the difference in constant pressure heat capacity between the unfolded and folded states. Near the stability range of proteins, equation (1) is well approximated by a quadratic equation.[9-10] From this, a relationship can be derived to determine the maximum stability for a protein

$$\Delta G_{max} \cong \left(\frac{{\Delta H_M}^2}{2T_M \Delta C_p^\circ}\right) \quad (2)$$

where ΔG_{max} is the stability of a protein at the temperature of maximum stability (T_{max}), given by equation (3)

$$T_{max} \cong T_M - \left(\frac{\Delta H_M}{\Delta C_p^\circ}\right) \quad (3)$$

Rees and Robertson derived the following relationship between N, T_M and ΔG_{max} for any protein by substituting parameters for the chain-length dependent ΔH and ΔC_p° values of proteins determined by Robertson and Murphy into equation (2).[10-11] ΔH and ΔC_p° are approximately linear functions of N.

$$\Delta G_{max} \cong \left(\frac{0.0290N}{T_M}\right)(T_M - 282.6)^2 \text{ kJ/mol} \quad (4)$$

Rees and Robertson further demonstrate that equation (4) can be differentiated to determine the chain-length dependent relationship between ΔG_{max} and T_M, equation (5)

$$\frac{1}{N}\frac{d\Delta G_{max}}{dT_M} = 0.029 - \frac{2316}{{T_M}^2} \quad (5)$$

They examined three proteins in detail and demonstrated a linear relationship between $\Delta\Delta G_{max}$ and ΔT_M that scales as N^{-1} and predicted a value of 0.009 kJ mol^{-1} K^{-1} res^{-1} for the ratio of $\Delta\Delta G_{max}$ and ΔT_M for a protein with a T_M of 340 K. We collected thermodynamic data from the literature and a new dataset of 866 mutants of 16 proteins was compiled for which T_M and $\Delta G°$ data was available. Most of the available mutational stability data was reported at 298 K rather than at T_{max}, so $\Delta G°$ was plotted instead of ΔG_{max}. This is a reasonable substitution, since $T_{max} \approx 283$ K, independent of T_M, as noted by Rees and Robertson and the slope of the $\Delta G°$ *vs* T plot is small near T_{max} since the slope is 0 at T_{max}. In either case, the model assumes that changes in T_M are the result of changes in ΔG rather than changes in T_{max} or $\Delta C_p°$. This assumption is justified by Rees and Robertson, who observe that mutations tend to "pull up" or "push down" the protein stability curve, rather than shifting or broadening it, as well as by extensive work showing that $\Delta C_p°$ is strongly correlated to chain length.[11-13] Assuming that the change in stability results from a vertical shift of the Gibbs-Helmholtz plot, $\Delta G°$ at any temperature can be used since $\Delta\Delta G$ will be independent of the choice of T. In any case, any error due to the choice of 298 K as the reference temperature is expected to be small.

The proteins studied range in size from Trp cage at 20 residues to T4 lysozyme with 164 resides and include all-α, all-β and mixed α-β folds. Several of the domains contain disulfides while others do not (Table 1). Data for at least 9 mutants are available for each protein. The mutations include truncation of sidechains, the introduction or removal of charged residues and mutants which create cavities in the interior of proteins. These data were fitted to a linear equation to extract the value of the slope of a plot ΔT_M *vs* $\Delta\Delta G°$ for each protein. Representative plots for 8 proteins are shown in Figure 1 and the remainder are shown in supporting information (Figure S1). A strong linear correlation is observed for each protein, R^2 is greater than 0.85 for 10 of the 16 proteins and above

0.72 for five other proteins. The R^2 value for RNase A is the lowest at 0.6389. The p values for all but one protein are less than 7×10^{-8} and the other still has a p value less than 2×10^{-3}. Values of the slopes of ΔT_M *vs* $\Delta\Delta G$ for the different proteins scale as N^{-1}. Values of $\Delta T_M/\Delta\Delta G°$ were plotted against N^{-1} and fitted to a linear equation with a fixed intercept at the origin (Figure 2). The fit has R^2 and p values of 0.8726 and 2.34×10^{-11}, respectively. The relationship derived from the experimental data between chain length the change in T_M for a given stability change at 298 K is described by equation (6)

$$\frac{\Delta T_M}{\Delta\Delta G°} = \frac{90.72}{N} \text{ mol K kJ}^{-1}$$

$$\frac{1}{N}\frac{\Delta\Delta G°}{\Delta T_M} = 0.011 \text{ kJ mol}^{-1}\text{K}^{-1}\text{res}^{-1} \quad (6)$$

The value of 0.011 kJ mol^{-1} K^{-1} res^{-1} is in reasonable agreement with the value of 0.009 kJ mol^{-1} K^{-1} res^{-1} originally estimated by Rees and Robertson. The inverse correlation between chain length and the effect of stabilization at 298 K on T_M is apparent, demonstrating that the thermal stability of small proteins is more sensitive to changes in stability than in larger proteins. This is a consequence of the broad stability curves of small proteins resulting from the small $\Delta C_p°$ values of these systems. It should be noted that equation (6) is derived from direct fitting of the experimental data, and therefore is not dependent on the relationship between N, ΔG_{max}, ΔH_M and $\Delta C_p°$.

Equation (6) does more than predict ΔT_M for mutant proteins, the equation is an empirical relationship between stability and T_M. Thus, it can also be employed to predict changes in T_M observed upon the binding of a ligand to the folded or unfolded state of a protein of a given length, provided that the effects of ligand binding on stability mimic the thermodynamic effects of mutation. This is most likely the case for ligands that do not require a significant protein

conformational change to bind or lead to a large change in the heat capacity of the system. Many druglike ligands are small molecules which bind to sites which are accessible to solvent and equation (6) is likely to apply to this important class. The $\Delta G°$ of folding in the absence of ligand is given by equation (7)

$$\Delta G° = -RT \ln \frac{[F]}{[U]} \tag{7}$$

where R is the gas constant, $[F]$ is the equilibrium concentration of folded protein and $[U]$ is the equilibrium concentration of unfolded protein. In the presence of a single ligand that binds exclusively to the folded state, equation (7) becomes equation (8)[14-15]

$$\Delta G_L° = \Delta G° - RT \ln \left(1 + \frac{[L]}{K_D}\right) \tag{8}$$

where $\Delta G_L°$ is the Gibbs free energy of folding at 298 K in the presence of ligand, $[L]$ is the concentration of free ligand and K_D is the dissociation constant for ligand binding to the folded state. The change in free energy, $\Delta\Delta G°$ upon ligand binding is:

$$\Delta\Delta G_L° = -RT \ln \left(1 + \frac{[L]}{K_D}\right) \tag{9}$$

where $\Delta\Delta G_L°$ is the change in the Gibbs free energy of folding at 298 K upon the binding of ligand, which is related to ΔT_M by the linear relationship determined from the mutational dataset analysis (equation 6). A somewhat different equation applies when a ligand binds to both the folded and unfolded state, but the scaling of $\Delta\Delta G°$ and ΔT_M have the same implications and here we focus on the case where ligand binds only to the folded state. It is common in ligand binding assays to use a large molar excess of ligand such that the total ligand concentration $[L]_T$ is much greater than the total protein concentration $[P]_T$. In this limiting case $[L] \approx [L]_T$ and the curve can be approximated by a linear equation. However, it is straightforward to derive the quadratic solution for $[L]$ which is valid at all $[L]_T$ and $[P]_T$ (10)

$$[L] = \left(\frac{-([P]_T - [L]_T + K_D + K_D K_U) + \sqrt{([P]_T - [L]_T + K_D + K_D K_U)^2 - 4(-K_D - K_D K_U)}}{2}\right) \quad (10)$$

where K_U is the protein unfolding equilibrium constant in the absence of ligand. The value of K_U can be determined from the chain-length dependent stability equations parameterized by Robertson and Murphy or Sawle and Ghosh.[11, 16] However, this term is usually small compared to the other terms and can generally be excluded except in cases for proteins that are unstable in the absence of ligand (large K_U) or very weakly binding ligands (large K_D).

Equations (6), (9) and (10) can be combined by replacing $[L]$ in equation (9) with the expression in equation (10) and using equation (6) to express $\Delta\Delta G°$ to calculate the expected thermal shift for any N, $[P]_T$, $[L]_T$ and K_D. Representative plots are shown in Figure 3. Plotting $[L]_T/K_D$ normalizes $[L]_T$ relative to K_D. Several important considerations for thermal shift assays can be inferred from these plots. Firstly, thermal shifts which are much lower than the maximum achievable value are predicted when $[L]_T < [P]_T$, therefore it is preferable for the ligand to be in molar excess for the thermal shift to be most clearly observed; of course, significant thermal shifts can be observed for strongly binding ligands even at lower total ligand concentrations. When $[L]_T$ is normalized by K_D, it is apparent that the magnitude of the thermal shift is determined by the ratio of $[L]_T/K_D$. For example, when $[L]_T/K_D = 100$, $\Delta T_M = 1$ °C for a 100-residue protein for any K_D, provided that $[L]_T$ is at least tenfold higher than $[P]_T$. Consequently, thermal shift assays are not particularly sensitive to μM or weaker binding without employing high ligand concentrations.

Conversely, it is apparent from the sharp increase in the thermal shift where $[L]_T = [P]_T$ (indicated by red arrows in Figure 3) that ΔT_M is highly sensitive to the relative concentrations of protein and ligand when $[L]_T$ and $[P]_T$ are similar and larger than K_D (see for example Figure 3C, 3D). Consequently, significant loss of precision is expected to arise due to any error in the concentrations of protein and ligand under these conditions. This loss of precision is more likely

to be a problem in strongly binding systems because detection limits will necessitate protein concentrations much higher than K_D.

Most importantly, the analysis presented here predicts an inverse correlation between protein size and the magnitude of the expected thermal shift. In general, the analysis predicts that a protein will exhibit half the thermal shift experienced by a protein that is half as large. The model is validated by thermal shift assays from the literature, which report similar thermal shifts to those predicted.[17-18] Plots of thermal shift *versus* chain length at varying values of $[L]_T$ and $[P]_T$ also illustrate these predictions (Figure 4). Comparing plots of equal $[P]_T/[L]_T$, it is interesting to note that increasing $[P]_T$ and $[L]_T$ in tandem results in higher predicted thermal shifts due to the shift in the binding equilibrium, although this effect is much weaker than the effects of N, $[L]_T$ and K_D. It is important to note that the thermal shift model is based upon the chain-length dependence of stability and the relationship between ΔT_M and $\Delta\Delta G°$ for single domain proteins. The constant in equation (6) was derived considering the data sets available to us, as more data becomes available the constant may be revised, but the general conclusions are robust.

CONCLUSIONS

A dataset of mutational stability data for 866 mutants of 16 proteins was analyzed to derive a relationship between changes in $\Delta G°$ and T_M as a function of chain length. The strong linear correlation for each protein and the excellent correlation found between the slopes of the individual plots and N^{-1} indicates that changes in Gibbs free energy can be predicted from ΔT_M for a given protein provided that the parameters in equation (6) are known. From a practical perspective, $\Delta G°$ and T_M could be obtained for a subset of mutants and used to confirm linearity of the $\Delta G°$ *vs* T_M plot, allowing $\Delta G°$ to be estimated for other mutants. The data analyzed here for a range of

structurally distinct proteins provides confidence in the use of linear interpolation to estimate $\Delta\Delta G°$ values from changes in T_M.

The relationship between ΔT_M and $\Delta\Delta G°$ was used to model thermal stability changes upon ligand binding. The model predicts several important considerations for the design of thermal shift assays. Firstly, that thermal shift assays are less sensitive to μM or weaker binding affinities. Secondly, the full magnitude of the thermal shift is not observed unless the ligand is in molar excess; of course, in cases of strong binding affinities significant shifts can be observed even when the thermal shift is attenuated by the excess protein. In addition, thermal shift assays in more strongly binding systems are expected to be more prone to low precision when the concentrations of protein and ligand are very similar and larger than K_D due to the sensitivity of ΔT_M to the protein-to-ligand concentration ratio when the ratio is close to 1. A key observation is that there is a strong inverse correlation between chain length and the magnitude of the thermal shift, with the thermal shift being halved for every doubling of the protein length. Thus, thermal shift assays will be less sensitive for large proteins. There are some caveats to these conclusions; large, multidomain proteins may exhibit much larger thermal shifts than predicted if the assay monitors only the unfolding of a ligand binding domain. In addition, experimental thermal shift assays may actually be monitoring the conversion of monomeric folded proteins into an aggregated oligomeric unfolded form. In this case the equilibrium is more complicated. In some cases, thermal shift assays might be following the transition from the native state to a molten globule intermediate. The same general conclusions are expected if the thermodynamics of the native to molten globule transition scale with N in the same manner as the native to unfolded transition does since the predicted dependence of the thermal shift upon the size of a protein is based upon the relationship between the thermodynamics of protein unfolding and size.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge.

Thermodynamic stability data for mutants included in the analysis (PDF)

AUTHOR INFORMATION

Corresponding Author

*daniel.raleigh@stonybrook.edu; Tel 631 632 9547

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ABBREVIATIONS

T, temperature; T_M, midpoint of protein thermal unfolding transition; $\Delta G°$, Gibbs free energy change upon unfolding under standard conditions; N, number of residues; ΔH_M, enthalpy change upon unfolding at T_M; $\Delta C_p°$, constant pressure heat capacity change upon unfolding; $\Delta G(T)$, temperature-dependent Gibbs free energy change upon unfolding; T_{max}, temperature of maximum stability; ΔG_{max}, Gibbs free energy change at T_{max}; R, gas constant; [F], equilibrium concentration of folded protein; [U], equilibrium concentration of unfolded protein; $\Delta G_L°$, Gibbs free energy change upon folding in the presence of ligand at 298 K; K_D, dissociation constant; [L], equilibrium concentration of free ligand; $\Delta\Delta G_L°$, change in the Gibbs free energy change upon folding upon the binding of ligand at 298 K; $[L]_T$, total equilibrium concentration of ligand; $[P]_T$, total equilibrium concentration of protein; K_U, equilibrium constant of protein unfolding.

ACKNOWLEDGMENTS

We dedicate this work to Professor Ken Dill, on the occasion of his 70th birthday. This work was supported by NSF MCB-1330259 grant to DPR.

REFERENCES

(1) Rees, D. C.; Robertson, A. D., Some thermodynamic implications for the thermostability of proteins. *Protein Sci* **2001,** *10* (6), 1187-1194.

(2) Matulis, D.; Kranz, J. K.; Salemme, F. R.; Todd, M. J., Thermodynamic stability of carbonic anhydrase: Measurements of binding affinity and stoichiometry using thermofluor. *Biochemistry* **2005,** *44* (13), 5258-5266.

(3) Brandts, J. F.; Lin, L. N., Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry* **1990,** *29* (29), 6927-6940.

(4) Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P., *et al.*, High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **2001,** *6* (6), 429-440.

(5) Martinez Molina, D.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P., Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **2013,** *341* (6141), 84-87.

(6) Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundback, T.; Nordlund, P.; Martinez Molina, D., The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat Protoc* **2014,** *9* (9), 2100-2122.

(7) Jensen, A. J.; Martinez Molina, D.; Lundback, T., Cetsa: A target engagement assay with potential to transform drug discovery. *Future Med Chem* **2015,** *7* (8), 975-978.

(8) Alshareef, A.; Zhang, H. F.; Huang, Y. H.; Wu, C.; Zhang, J. D.; Wang, P.; El-Sehemy, A.; Fares, M.; Lai, R., The use of cellular thermal shift assay (cetsa) to study crizotinib resistance in alk-expressing human cancers. *Sci Rep* **2016,** *6*, 33710.

(9) Zipp, A.; Kauzmann, W., Pressure denaturation of metmyoglobin. *Biochemistry* **1973,** *12* (21), 4217-4228.

(10) Stowell, M. H.; Rees, D. C., Structure and stability of membrane proteins. *Adv Protein Chem* **1995,** *46*, 279-311.

(11) Robertson, A. D.; Murphy, K. P., Protein structure and the energetics of protein stability. *Chem Rev* **1997,** *97* (5), 1251-1268.

(12) Myers, J. K.; Pace, C. N.; Scholtz, J. M., Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **1995,** *4* (10), 2138-2148.

(13) Spolar, R. S.; Livingstone, J. R.; Record, M. T., Jr., Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **1992,** *31* (16), 3947-3955.

(14) Becktel, W. J.; Schellman, J. A., Protein stability curves. *Biopolymers* **1987,** *26* (11), 1859-1877.

(15) Creighton, T. E., *Proteins : Structures and molecular properties*.

(16) Sawle, L.; Ghosh, K., How do thermophilic proteins and proteomes withstand high temperature? *Biophys J* **2011,** *101* (1), 217-227.

(17) Giuliani, S. E.; Frank, A. M.; Collart, F. R., Functional assignment of solute-binding proteins of abc transporters using a fluorescence-based thermal shift assay. *Biochemistry* **2008,** *47* (52), 13974-13984.

(18) Krishna, S. N.; Luan, C. H.; Mishra, R. K.; Xu, L.; Scheidt, K. A.; Anderson, W. F.; Bergan, R. C., A fluorescence-based thermal shift assay identifies inhibitors of mitogen activated protein kinase kinase 4. *Plos One* **2013,** *8* (12), e81504.

(19) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H., Folding dynamics and pathways of the trp-cage miniproteins. *Biochemistry* **2014,** *53* (38), 6011-6021.

(20) Williams, D. V.; Byrne, A.; Stewart, J.; Andersen, N. H., Optimal salt bridge for trp-cage stabilization. *Biochemistry* **2011,** *50* (7), 1143-1152.

(21) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H., The trp-cage: Optimizing the stability of a globular miniprotein. *Protein Eng Des Sel* **2008,** *21* (3), 171-185.

(22) Rodriguez-Granillo, A.; Annavarapu, S.; Zhang, L.; Koder, R. L.; Nanda, V., Computational design of thermostabilizing d-amino acid substitutions. *J Am Chem Soc* **2011,** *133* (46), 18750-18759.

(23) Williams, D. V.; Barua, B.; Andersen, N. H., Hyperstable miniproteins: Additive effects of d- and l-ala mutations. *Org Biomol Chem* **2008,** *6* (23), 4287-4289.

(24) Jager, M.; Dendle, M.; Kelly, J. W., Sequence determinants of thermodynamic stability in a ww domain--an all-beta-sheet protein. *Protein Sci* **2009,** *18* (8), 1806-1813.

(25) McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S., A thermostable 35-residue subdomain within villin headpiece. *J Mol Biol* **1996,** *260* (2), 126-134.

(26) Kubelka, J.; Eaton, W. A.; Hofrichter, J., Experimental tests of villin subdomain folding simulations. *J Mol Biol* **2003,** *329* (4), 625-630.

(27) Kreitler, D. F.; Mortenson, D. E.; Forest, K. T.; Gellman, S. H., Effects of single alpha-to-beta residue replacements on structure and stability in a small protein: Insights from quasiracemic crystallization. *J Am Chem Soc* **2016,** *138* (20), 6498-6505.

(28) Hsu, W. L.; Shih, T. C.; Horng, J. C., Folding stability modulation of the villin headpiece helical subdomain by 4-fluorophenylalanine and 4-methylphenylalanine. *Biopolymers* **2015,** *103* (11), 627-637.

(29) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R., High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc Natl Acad Sci U S A* **2005,** *102* (21), 7517-7522.

(30) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J., Sub-microsecond protein folding. *J Mol Biol* **2006,** *359* (3), 546-553.

(31) Xiao, S.; Bi, Y.; Shan, B.; Raleigh, D. P., Analysis of core packing in a cooperatively folded miniature protein: The ultrafast folding villin headpiece helical subdomain. *Biochemistry* **2009,** *48* (21), 4607-4616.

(32) Zheng, T. Y.; Lin, Y. J.; Horng, J. C., Thermodynamic consequences of incorporating 4-substituted proline derivatives into a small helical protein. *Biochemistry* **2010,** *49* (19), 4255-4263.

(33) Bi, Y.; Cho, J. H.; Kim, E. Y.; Shan, B.; Schindelin, H.; Raleigh, D. P., Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain. *Biochemistry* **2007,** *46* (25), 7497-7505.

(34) Xiao, S. Protein folding and stability: Distinguishing folded from unfolded state effects. Stony Brook University, Stony Brook, NY, 2012.

(35) Peran, I. Dynamics of protein unfolded states. Stony Brook University, Stony Brook, NY, 2017.

(36) Merkel, J. S.; Sturtevant, J. M.; Regan, L., Sidechain interactions in parallel beta sheets: The energetics of cross-strand pairings. *Structure* **1999,** *7* (11), 1333-1343.

(37) Yu, M. H.; Weissman, J. S.; Kim, P. S., Contribution of individual side-chains to the stability of bpti examined by alanine-scanning mutagenesis. *J Mol Biol* **1995,** *249* (2), 388-397.

(38) Liu, Y.; Breslauer, K.; Anderson, S., "Designing out" disulfide bonds: Thermodynamic properties of 30-51 cystine substitution mutants of bovine pancreatic trypsin inhibitor. *Biochemistry* **1997,** *36* (18), 5323-5335.

(39) Jackson, S. E.; Moracci, M.; elMasry, N.; Johnson, C. M.; Fersht, A. R., Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **1993,** *32* (42), 11259-11269.

(40) de Prat Gay, G.; Johnson, C. M.; Fersht, A. R., Contribution of a proline residue and a salt bridge to the stability of a type i reverse turn in chymotrypsin inhibitor-2. *Protein Eng* **1994,** *7* (1), 103-108.

(41) elMasry, N. F.; Fersht, A. R., Mutational analysis of the n-capping box of the alpha-helix of chymotrypsin inhibitor 2. *Protein Eng* **1994,** *7* (6), 777-782.

(42) Gribenko, A. V.; Makhatadze, G. I., Role of the charge-charge interactions in defining stability and halophilicity of the cspb proteins. *J Mol Biol* **2007,** *366* (3), 842-856.

(43) Pace, C. N.; Horn, G.; Hebert, E. J.; Bechert, J.; Shaw, K.; Urbanikova, L.; Scholtz, J. M.; Sevcik, J., Tyrosine hydrogen bonds make a large contribution to protein stability. *J Mol Biol* **2001,** *312* (2), 393-404.

(44) Hebert, E. J.; Giletto, A.; Sevcik, J.; Urbanikova, L.; Wilson, K. S.; Dauter, Z.; Pace, C. N., Contribution of a conserved asparagine to the conformational stability of ribonucleases sa, ba, and t1. *Biochemistry* **1998,** *37* (46), 16192-16200.

(45) Grimsley, G. R.; Shaw, K. L.; Fee, L. R.; Alston, R. W.; Huyghues-Despointes, B. M.; Thurlkill, R. L.; Scholtz, J. M.; Pace, C. N., Increasing protein stability by altering long-range coulombic interactions. *Protein Sci* **1999,** *8* (9), 1843-1849.

(46) Takano, K.; Scholtz, J. M.; Sacchettini, J. C.; Pace, C. N., The contribution of polar group burial to protein stability is strongly context-dependent. *J Biol Chem* **2003,** *278* (34), 31790-31795.

(47) Yakovlev, G. I.; Mitkevich, V. A.; Shaw, K. L.; Trevino, S.; Newsom, S.; Pace, C. N.; Makarov, A. A., Contribution of active site residues to the activity and thermal stability of ribonuclease sa. *Protein Sci* **2003,** *12* (10), 2367-2373.

(48) Trevino, S. R.; Gokulan, K.; Newsom, S.; Thurlkill, R. L.; Shaw, K. L.; Mitkevich, V. A.; Makarov, A. A.; Sacchettini, J. C.; Scholtz, J. M.; Pace, C. N., Asp79 makes a large, unfavorable contribution to the stability of rnase sa. *J Mol Biol* **2005,** *354* (4), 967-978.

(49) Fu, H.; Grimsley, G.; Scholtz, J. M.; Pace, C. N., Increasing protein stability: Importance of deltac(p) and the denatured state. *Protein Sci* **2010,** *19* (5), 1044-1052.

(50) Lim, W. A.; Farruggio, D. C.; Sauer, R. T., Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* **1992,** *31* (17), 4324-4333.

(51) Hecht, M. H.; Sturtevant, J. M.; Sauer, R. T., Stabilization of lambda repressor against thermal denaturation by site-directed gly----ala changes in alpha-helix 3. *Proteins* **1986,** *1* (1), 43-46.

(52) Hecht, M. H.; Sturtevant, J. M.; Sauer, R. T., Effect of single amino acid replacements on the thermal stability of the nh2-terminal domain of phage lambda repressor. *Proc Natl Acad Sci U S A* **1984,** *81* (18), 5685-5689.

(53) Reidhaar-Olson, J. F.; Parsell, D. A.; Sauer, R. T., An essential proline in lambda repressor is required for resistance to intracellular proteolysis. *Biochemistry* **1990,** *29* (33), 7563-7571.

(54) Stearman, R. S.; Frankel, A. D.; Freire, E.; Liu, B. S.; Pabo, C. O., Combining thermostable mutations increases the stability of lambda repressor. *Biochemistry* **1988,** *27* (19), 7571-7574.

(55) Yang, W. Y.; Gruebele, M., Rate-temperature relationships in lambda-repressor fragment lambda 6-85 folding. *Biochemistry* **2004,** *43* (41), 13018-13025.

(56) Shirley, B. A.; Stanssens, P.; Steyaert, J.; Pace, C. N., Conformational stability and activity of ribonuclease t1 and mutants. Gln25----lys, glu58----ala, and the double mutant. *J Biol Chem* **1989,** *264* (20), 11621-11625.

(57) Shirley, B. A.; Stanssens, P.; Hahn, U.; Pace, C. N., Contribution of hydrogen bonding to the conformational stability of ribonuclease t1. *Biochemistry* **1992,** *31* (3), 725-732.

(58) Fabian, H.; Schultz, C.; Backmann, J.; Hahn, U.; Saenger, W.; Mantsch, H. H.; Naumann, D., Impact of point mutations on the structure and thermal stability of ribonuclease t1 in aqueous solution probed by fourier transform infrared spectroscopy. *Biochemistry* **1994,** *33* (35), 10725-10730.

(59) Giletto, A.; Pace, C. N., Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease t1. *Biochemistry* **1999,** *38* (40), 13379-13384.

(60) Schindler, T.; Mayr, L. M.; Landt, O.; Hahn, U.; Schmid, F. X., The role of a trans-proline in the folding mechanism of ribonuclease t1. *Eur J Biochem* **1996,** *241* (2), 516-524.

(61) Schubert, W. D.; Schluckebier, G.; Backmann, J.; Granzin, J.; Kisker, C.; Choe, H. W.; Hahn, U.; Pfeil, W.; Saenger, W., X-ray crystallographic and calorimetric studies of the effects of the mutation trp59-->tyr in ribonuclease t1. *Eur J Biochem* **1994,** *220* (2), 527-534.

(62) Kellis, J. T., Jr.; Nyberg, K.; Fersht, A. R., Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* **1989,** *28* (11), 4914-4922.

(63) Axe, D. D.; Foster, N. W.; Fersht, A. R., An irregular beta-bulge common to a group of bacterial rnases is an important determinant of stability and function in barnase. *J Mol Biol* **1999,** *286* (5), 1471-1485.

(64) Vu, N. D.; Feng, H.; Bai, Y., The folding pathway of barnase: The rate-limiting transition state and a hidden intermediate under native conditions. *Biochemistry* **2004,** *43* (12), 3346-3356.

(65) Eberhardt, E. S.; Wittmayer, P. K.; Templer, B. M.; Raines, R. T., Contribution of a tyrosine side chain to ribonuclease a catalysis and stability. *Protein Sci* **1996,** *5* (8), 1697-1703.

(66) Catanzano, F.; Graziano, G.; Capasso, S.; Barone, G., Thermodynamic analysis of the effect of selective monodeamidation at asparagine 67 in ribonuclease a. *Protein Sci* **1997,** *6* (8), 1682-1693.

(67) Salamanca, S.; Villegas, V.; Vendrell, J.; Li, L.; Aviles, F. X.; Chang, J. Y., The unfolding pathway of leech carboxypeptidase inhibitor. *J Biol Chem* **2002,** *277* (20), 17538-17543.

(68) Kadonosono, T.; Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T., Minimization of cavity size ensures protein stability and folding: Structures of phe46-replaced bovine pancreatic rnase a. *Biochemistry* **2003,** *42* (36), 10651-10658.

(69) Johnson, R. J.; Lin, S. R.; Raines, R. T., Genetic selection reveals the role of a buried, conserved polar residue. *Protein Sci* **2007,** *16* (8), 1609-1616.

(70) Pecher, P.; Arnold, U., The effect of additional disulfide bonds on the stability and folding of ribonuclease a. *Biophys Chem* **2009,** *141* (1), 21-28.

(71) Gotte, G.; Donadelli, M.; Laurents, D. V.; Vottariello, F.; Morbio, M.; Libonati, M., Increase of rnase a n-terminus polarity or c-terminus apolarity changes the two domains' propensity to swap and form the two dimeric conformers of the protein. *Biochemistry* **2006,** *45* (36), 10795-10806.

(72) Eftink, M. R.; Ghiron, C. A.; Kautz, R. A.; Fox, R. O., Fluorescence and conformational stability studies of staphylococcus nuclease and its mutants, including the less stable nuclease-concanavalin a hybrids. *Biochemistry* **1991,** *30* (5), 1193-1199.

(73) Carra, J. H.; Anderson, E. A.; Privalov, P. L., Thermodynamics of staphylococcal nuclease denaturation. Ii. The a-state. *Protein Sci* **1994,** *3* (6), 952-959.

(74) Shortle, D.; Meeker, A. K.; Freire, E., Stability mutants of staphylococcal nuclease: Large compensating enthalpy-entropy changes for the reversible denaturation reaction. *Biochemistry* **1988,** *27* (13), 4761-4768.

(75) Carra, J. H.; Privalov, P. L., Energetics of denaturation and m values of staphylococcal nuclease mutants. *Biochemistry* **1995,** *34* (6), 2034-2041.

(76) Gittis, A. G.; Stites, W. E.; Lattman, E. E., The phase transition between a compact denatured state and a random coil state in staphylococcal nuclease is first-order. *J Mol Biol* **1993,** *232* (3), 718-724.

(77) Hinck, A. P.; Truckses, D. M.; Markley, J. L., Engineered disulfide bonds in staphylococcal nuclease: Effects on the stability and conformation of the folded protein. *Biochemistry* **1996,** *35* (32), 10328-10338.

(78) Byrne, M. P.; Stites, W. E., Chemically crosslinked protein dimers: Stability and denaturation effects. *Protein Sci* **1995,** *4* (12), 2545-2558.

(79) Chen, J.; Lu, Z.; Sakon, J.; Stites, W. E., Increasing the thermostability of staphylococcal nuclease: Implications for the origin of protein thermostability. *J Mol Biol* **2000,** *303* (2), 125-130.

(80) Leung, K. W.; Liaw, Y. C.; Chan, S. C.; Lo, H. Y.; Musayev, F. N.; Chen, J. Z.; Fang, H. J.; Chen, H. M., Significance of local electrostatic interactions in staphylococcal nuclease studied by site-directed mutagenesis. *J Biol Chem* **2001,** *276* (49), 46039-46045.

(81) Maki, K.; Cheng, H.; Dolgikh, D. A.; Shastry, M. C.; Roder, H., Early events during folding of wild-type staphylococcal nuclease and a single-tryptophan variant studied by ultrarapid mixing. *J Mol Biol* **2004,** *338* (2), 383-400.

(82) Baase, W. A.; Liu, L. J.; Tronrud, D. E.; Matthews, B. W., Lessons from the lysozyme of phage t4. *Protein Science* **2010,** *19* (4), 631-641.

FIGURES

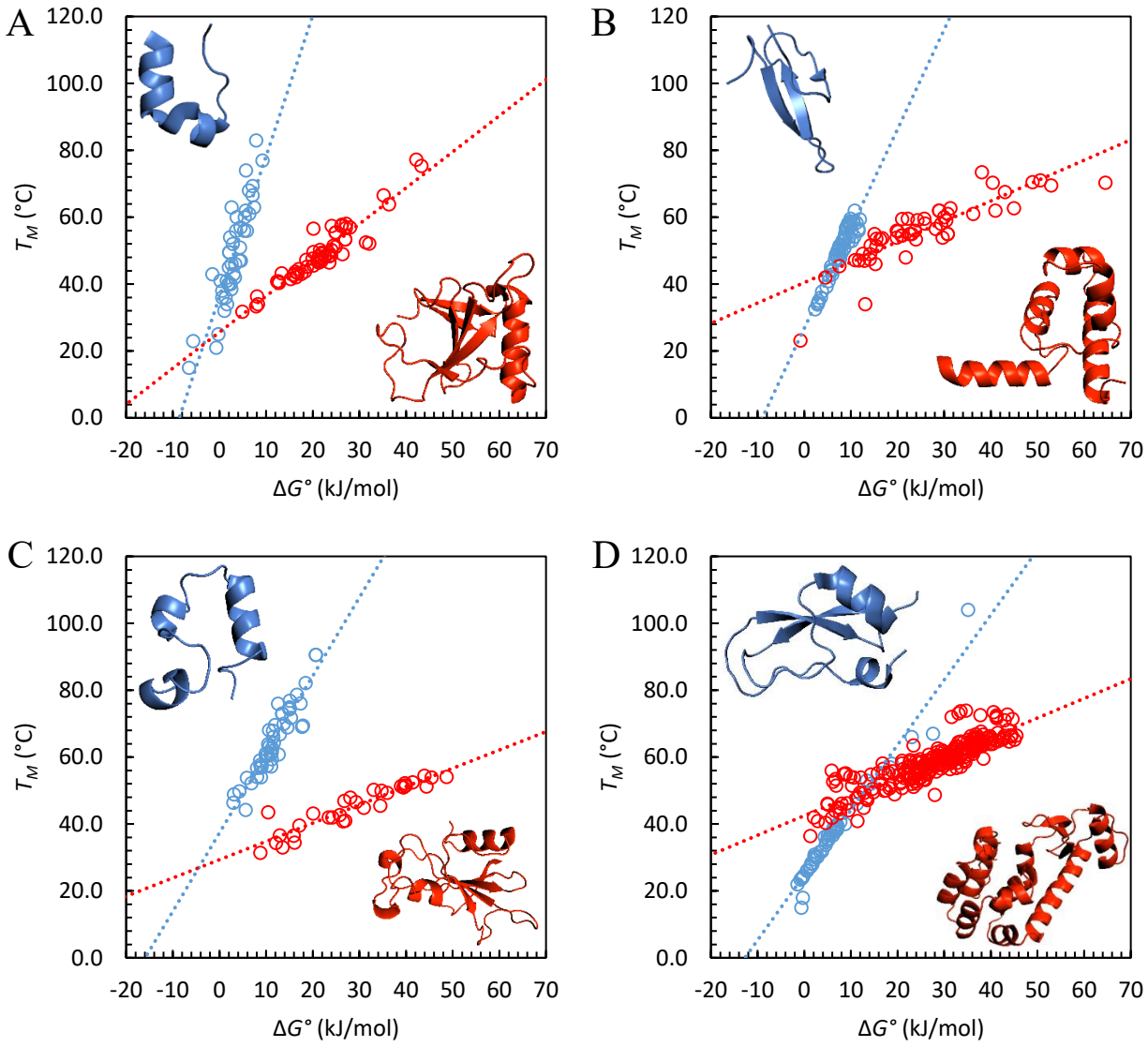


Figure 1. There is a strong linear correlation between $\Delta G°$ and T_M for mutants of single domain proteins. **(A)** Trp-cage (blue, 20 residues) and RNase Sa (red, 96 residues). **(B)** hPin1 WW domain (blue, 34 residues) and lambda repressor (red, 102 residues). **(C)** HP36 (blue, 36 residues) and barnase (red, 110 residues). **(D)** BPTI (blue, 58 residues) and T4 lysozyme (red, 164 residues). Plots for an additional 8 proteins are given in the supporting information (Figure S1).

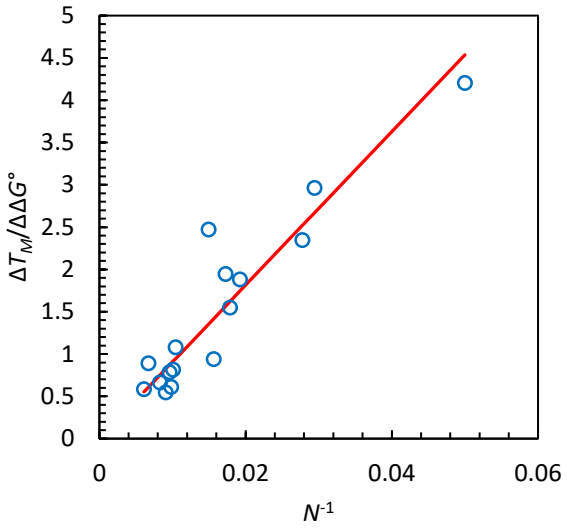


Figure 2. The slope of T_M *vs.* $\Delta G°$ plots correlates linearly with N^{-1} ($R^2 = 0.8726$, $p = 2.34\times10^{-11}$).

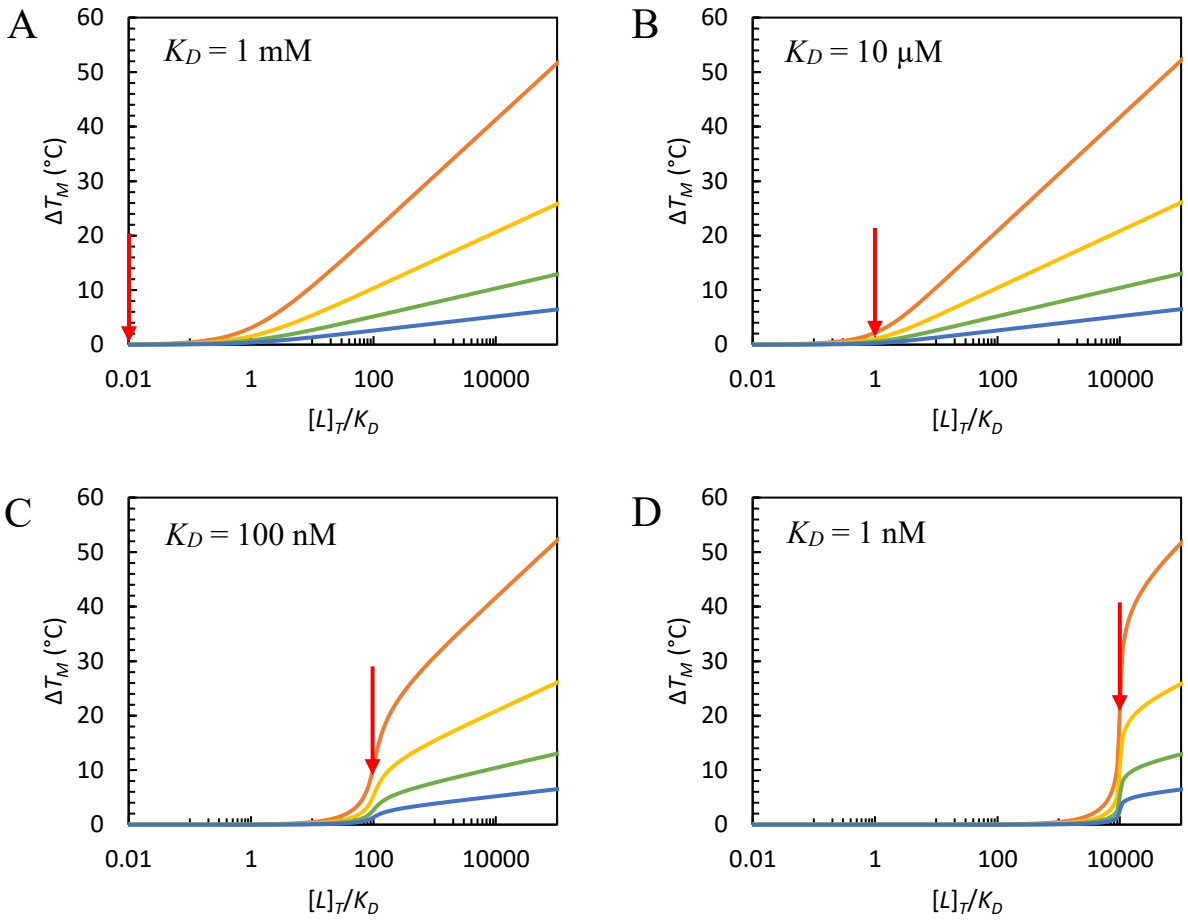


Figure 3. The thermal shift resulting from binding of a ligand with a K_D of **(A)** 1 mM **(B)** 10 µM **(C)** 100 nM and **(D)** 1 nM. Individual curves represent proteins of length 50 (orange), 100 (yellow), 200 (green) and 400 (blue). Red arrows indicate $[L]_T$ = 10 µM, which also corresponds to the protein concentration used to generate the plots. This concentration was chosen to agree with typical experimental methods from the literature.[17-18]

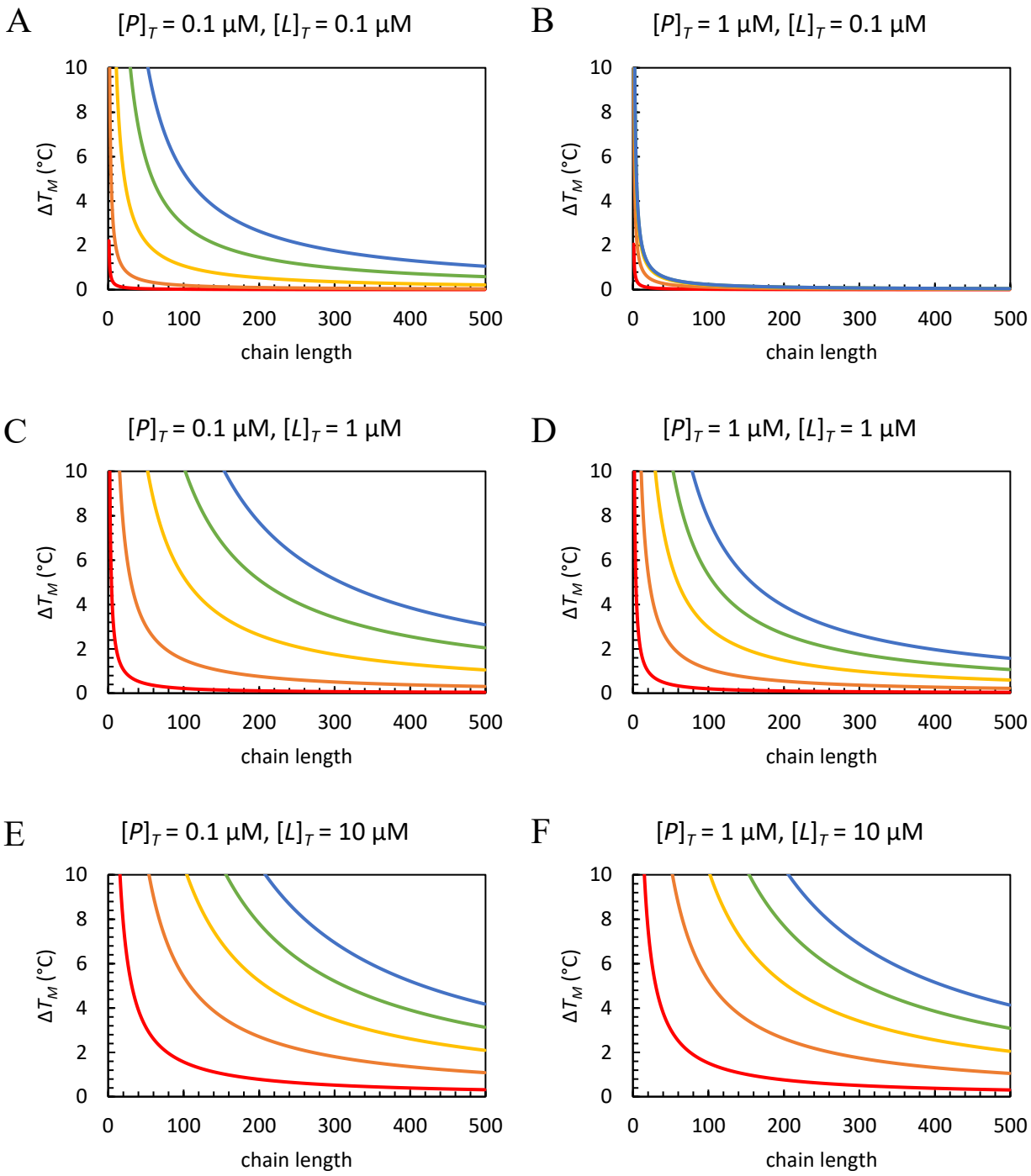


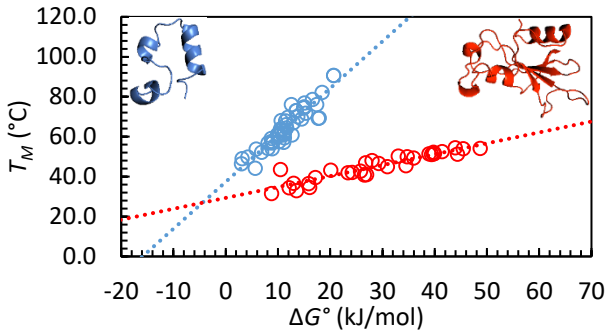
Figure 4. Chain length dependence of the thermal shift for a ligand with a K_D of 10 µM (red), 1 µM (orange), 100 nM (yellow), 10 nM (green) and 1 nM (blue). Curves were calculated at total protein concentrations of 0.1 and 1 µM and total ligand concentrations of 0.1, 1 and 10 µM.

TABLES

Table 1. Summary of proteins in the mutational dataset.

Protein	N	$\Delta T_M/\Delta\Delta G°$ (K mol/kJ)	Number of Data Points	R^2	p value	PDB
Trp-Cage[19-23]	20	4.2023	44	0.7782	2.53×10^{-15}	2JOF
hPin1 WW[24]	34	2.9628	56	0.9187	1.18×10^{-31}	2NC3
HP36[25-34]	36	2.3459	49	0.8604	1.01×10^{-21}	1VII
NTL9[35]	52	1.8823	9	0.7847	1.53×10^{-3}	2HBB
B1 domain of Protein G[36]	56	1.5507	19	0.8652	8.13×10^{-9}	2GB1
BPTI[37-38]	58	1.9451	51	0.9573	3.28×10^{-35}	5PTI
CI2[39-41]	64	0.9397	25	0.7275	6.12×10^{-8}	1COA
CspB[42]	67	2.4721	60	0.9445	4.04×10^{-38}	1CSP
RNase Sa[43-49]	96	1.0783	60	0.8826	1.17×10^{-28}	1RGG
RNase Sa3[43]	99	0.8162	9	0.9997	1.73×10^{-13}	1MGR
λ Repressor[50-55]	102	0.6100	53	0.7420	1.27×10^{-16}	1LMB
RNase T1[44-45, 56-61]	104	0.7865	50	0.8545	1.14×10^{-21}	9RNT
Barnase[62-64]	110	0.5466	32	0.8646	1.45×10^{-14}	1BNI
RNase A[65-71]	121	0.6645	31	0.6389	6.97×10^{-8}	2E3W
Staph. Nuclease[72-81]	149	0.8905	76	0.7727	1.64×10^{-25}	1STN
T4 Lysozyme[82]	164	0.5843	242	0.7752	9.69×10^{-80}	2LZM

TOC GRAPHIC


120.0
100.0
80.0
60.0
40.0
20.0
0.0
T_M (°C)
-20 -10 0 10 20 30 40 50 60 70
ΔG° (kJ/mol)