University of California, San Francisco

Computational DSF modeling

Gestwicki Lab
March 18, 2020

This website and the associated computations DSF data models are presented in a publication currently under review (as of 2020-03-17): "Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments." What follows here is a combination of the main body text from "Section I: An improved theoretical framework for DSF" of that paper, and its supplementary information. The full code for DSFworld, as well as stand-alone scripts and modularized R Shiny web apps for its various capabilities, are freely available on GitHub, at https://github.com/gestwicki-lab/dsfworld. If you find these computational DSFworld data models useful, please spread the world by citing the paper! For more information, you can download the full paper and its supplementary information from the "About DSFworld" tab.

DSF is a simple readout of a complex phenomenon. In each DSF experiment, the measured fluorescence signal is a product of multiple molecular events, including protein unfolding, dye binding, and dye activation (e.g. an increase in quantum yield). The purpose of the following note is to elaborate on Model 2, the computational DSF data model presented on this website. Model 2, while a useful framework for DSF in our experience, is not exhaustive. It is likely that individual DSF applications may require adjustments of improvements to this framework for maximum utility. To facilitate this, the following is provided as a starting point for these adjustments.

Part I: thermodynamic models of protein thermal unfolding.

Typical DSF experiments assume a model where the protein has only two relevant states—native and unfolded—and these states are always at equilibrium [3, 5]. That is:

$$Native \leftrightharpoons Reversibly\ unfolded \qquad \text{(Model scheme 1)}$$

Through the 1970s and 1980s, the thermal unfolding of over 50 proteins was monitored by DSC [9, 2]. When considered in sum, these experiments suggested that the relationship between ΔG_{Fold} and temperature could be adequately described in terms of (1) the melting temperature of the protein $T_{1/2}$, (2) the enthalpy of unfolding ($\Delta H_{unfolding}$), and (3) the change in heat capacity ΔC_P associated with protein unfolding, using the empirically-derived relationship [9, 2, 10]:

$$\Delta G(T) = \Delta H_U \frac{T_{1/2} - T}{T_U} - \Delta C_p \left(T_{1/2} - T \left(1 - ln \frac{T}{T_{1/2}} \right) \right) \qquad (1)$$

Equation 1 relatesΔG_{Fold} to temperature, which is converted to the fraction of protein folded (X_{Folded}) and unfolded ($X_{Unfolded}$) via the Gibbs-Hemholtz equation:

$$\Delta G = -RTln(K) \qquad (2)$$

re-arranged to:

$$K = e^{\frac{-\Delta G}{RT}} \qquad (3)$$

For two-state systems like Model scheme 1, the equilibrium constant K is related to the population of the native and reversibly unfolded states:

$$X_{Folded} = 1 - \frac{K}{K+1}$$

$$X_{Reversibly\ unfolded} = \frac{K}{K+1} \qquad \text{(Model 1)}$$

Equation 1 is an established empirical relationship; equations 2 and 3 are general principles of thermodynamics. The thermodynamic model of unfolding built from them—"Model 1"—is displayed in the top panel of the interactive simulation in DSFworld. The outcomes of Model 1 respond only to changes in the thermodynamic properties of the protein–ΔH , the "true" (thermodynamic) Tm $T_{1/2}$, and ΔC_P–as expected for purely thermodynamic unfolding. $T_{1/2}$ controls the midpoint of the unfolding transition; ΔH and ΔC_P change the shape.

Part II: Incorporation of the influence of kinetics.

The influence of kinetics is then included by adapting Model 1 to contain an irreversible unfolding step. That is:

$$Native \leftrightharpoons Reversibly\ unfolded \rightarrow Irreversibly\ unfolded \quad \text{(Model scheme 2)}$$

Model scheme 2 is the simplest of the classic Lumry-Eyring models of mixed thermodynamic-kinetic unfolding. Our incorporation of kinetic influence into thermodynamic unfolding models via the inclusion of a second irreversible unfolding step is adapted directly from previous work by Jose Sanchez-Ruiz extrapolating Lumry-Eyring models to DSC [11]. Briefly, the rate of conversion from the reversibly to irreversibly unfolded population is captured by a first-order rate constant k. The Arrhenius equation—a standard formula for the temperature-dependence of reaction rates—modifies k as temperature increases in the simulation:

$$k\ (min^{-1}) = e^{-\frac{E_a}{R}\left(\frac{1}{T}-\frac{1}{T_{kin}}\right)} \tag{5}$$

Where E_a is the activation energy of irreversible unfolding, T the temperature in Kelvin, T_{kin} a benchmark for the temperature at which irreversible unfolding becomes significant (precisely, where $k = 1min^{-1}$), and R the gas constant 8.314 $\frac{J}{molK}$. The population of each state (e.g. folded, reversibly unfolded, irreversibly unfolded) is described by the two thermodynamic models from equation 3, joined now by an irreversibly unfolded state, each modified by the same temperature- and ramp rate-dependent kinetic parameter, $L(T)$:

$$\begin{aligned} X_{Folded} &= 1 - \frac{K}{K+1}L(T) \\ X_{Rev.\ unf.} &= \frac{K}{K+1}L(T) \\ X_{Irrev.\ unf.} &= 1 - L(T) \end{aligned} \tag{6}$$

Where $L(T)$ is obtained for each temperature T by integration from a low temperature where k is negligible (in this work, $T_o = 298K$), followed by division by the thermocycling ramp rate in minutes, ν, as below:

$$L(T) = 1 - e^{-M(T)} \tag{7}$$

where

$$M(T) = -\frac{1}{\nu}\int_{T_0}^{T} k\left(\frac{K}{K+1}\right)dT \tag{9}$$

As expected, in the absence of irreversible unfolding and therefore kinetic influence $k = 0$, which makes $L(T) = 1$ and $X_{Irrev.\ unf.} = 0$ for all temperatures, and equation 9 becomes equal to the thermodynamic models in Model 1.

Part III: Incorporation of dye binding and fluorescent activation.

All of the modeling presented to this point applies broadly to heat-based denaturation of proteins, predicting the relative populations of the folded, reversibly unfolded, and irreversibly unfolded states of the protein. DSF data is simulated from these relative populations by multiplying the abundance of each state (native, reversibly unfolded, irreversibly unfolded) by the extent to which it activates the dye, comprising both dye binding and quantum yield:

$$RFU_{Folded} = X_{Folded} \times Dye\ detection_{Folded}$$

$$RFU_{Rev.\ unf.} = X_{Rev.\ unf.} \times Dye\ detection_{Rev.\ unf.} \tag{10}$$

$$RFU_{Irrev.\ unf.} = X_{Irrev.\ unf.} \times Dye\ detection_{Irrev.\ unf.}$$

The extent of dye activation resulting the three states is summed to produce the total signal:

$$RFU_{all} = (RFU_{Folded} + RFU_{Rev.\ unf.} + RFU_{Irrev.\ unf.}) \times decay(T) \tag{Model 2}$$

where

$$decay(T) = 1 - D \times \left(\frac{T}{T_{final}}\right) \tag{11}$$

The final modeled data, $RFU_{observed}$, is calculated from RFU_{all} by multiplication by a linear temperature-dependent decay. This linear decay encompasses protein-independent loss of dye quantum yield and hydrophobic-driven binding at elevated temperatures.

Considerations and caveats.

- There are many theoretical models of protein unfolding [11, 1, 12]. If the system of interest follows a different model of protein unfolding, then Model 2 will need to be adjusted. For example, Model 2 uses a first-order rate constant (equation 5) to describe the kinetically-controlled irreversible step, while aggregation-based irreversible unfolding processes, and particularly amyloid formation, are widely thought to follow higher-order kinetics [4].
- The ΔH_{Unfold} of the transition from the reversibly unfolded to the irreversibly unfolded state is assumed to be 0. This is currently a widespread assumption in the literature derived from DSC data [11, 7, 8, 6]. However, this is a caveat for the application of Model 2 to systems where $\Delta H_{Irrev.unf.}$ unfolding is known to be significant.

- Model 2 does not consider the influence of dye binding on the unfolding trajectory. In reality, dye-protein interactions must occur with some favorable $\Delta G_{binding}$, though both the magnitude and protein-to-protein consistency of this $\Delta G_{binding}$ is likely variable not clear at this time. The omission of dye binding energies from Model 2 should not affect outcomes for dye binding to the irreversibly unfolded state, as this transition is dominated by the associated activation energy (which is insensitive to the energy of the irreversibly unfolded state) rather than the relative ΔG between the reversibly and irreversibly unfolded states. However, dye binding to the reversibly unfolded state would perturb the $\Delta G_{unfolding}$, likely by stabilizing the reversibly unfolded state, thereby simultaneously decreasing the protein Tm_a and increasing the activation energy of the irreversible, kinetically-controlled transition to the irreversibly unfolded state. Whether this dye-based stabilization of the reversibly unfolded state would lead to dye concentration-dependent increases or decreases in Tm_a would likely depend on the extent to which the dye also detected the irreversibly unfolded state. Whether or not reversibly unfolded states are rendered irreversible in the context of DSF due to dye binding remains an interesting and unanswered question.

References

[1] Jennifer M Andrews and Christopher J Roberts. "A LumryEyring Nucleated Polymerization Model of Protein Aggregation Kinetics: 1. Aggregation with Pre-Equilibrated Unfolding". In: *The Journal of Physical Chemistry B* 111.27 (July 2007), pp. 7897–7913.

[2] Wayne J Becktel and John A Schellman. "Protein Stability Curves". In: *Biopolymers Original Research on Biomolecules* 26.11 (1987), pp. 1859–1877.

[3] Ashwin Chari et al. "ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space". In: *Nature Methods* 12.9 (Aug. 2015), pp. 859–865.

[4] Nami Hirota, Herman Edskes, and Damien Hall. "Unified theoretical description of the kinetics of protein aggregation". In: *Biophysical Reviews* 11.2 (Mar. 2019), pp. 191–208.

[5] Mei-Chu Lo et al. "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery". In: *Analytical Biochemistry* 332.1 (Sept. 2004), pp. 153–159.

[6] Susan P Manly, Kathleen S Matthews, and Julian M Sturtevant. "Thermal denaturation of the core protein of lac repressor". In: *Biochemistry* 24.15 (1985), pp. 3842–3846.

[7] P L Privalov. "Domains in the Fibrinogen Molecule". In: *Journal of Molecular Biology* 159 (1982), pp. 665–683.

[8] P L Privalov. “Stability of proteins. Proteins which do not present a single cooperative system.” In: *Advances in protein chemistry* 35 (1982), pp. 1–104.

[9] Douglas C Rees and Andrew D Robertson. “Some thermodynamic implications for the thermostability of proteins”. In: *Protein Science* 10.6 (June 2001), pp. 1187–1194.

[10] Andrew D Robertson and Kenneth P Murphy. “Protein Structure and the Energetics of Protein Stability”. In: *Chemical Reviews* 97.5 (Aug. 1997), pp. 1251–1268.

[11] Jose M Sanchez-Ruiz. “Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry”. In: *Biophysical Journal* 61.4 (Apr. 1992), pp. 921–935.

[12] Hans-Joachim Schonfeld and Joachim Seelig. “Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy Two-state model versus sequential unfolding”. In: *QUARTERLY REVIEWS OF BIOPHYSICS* (May 2016), pp. 1–24.