

In-depth interrogation of protein thermal unfolding data with MoltenProt

Vadim Kotov^{1,2,3} | Georg Mlynek⁴ | Oliver Vesper^{1,2,3} | Marina Pletzer⁴ | Jiri Wald^{1,2,3} | Celso M. Teixeira-Duarte⁵ | Herve Celia⁶ | Maria Garcia-Alai^{1,9} | Stephan Nussberger⁷ | Susan K. Buchanan⁶ | João H. Morais-Cabral⁵ | Christian Loew^{1,9} | Kristina Djinovic-Carugo^{4,8} | Thomas C. Marlovits^{1,2,3} 

¹Centre for Structural Systems Biology (CSSB), Hamburg, Germany

²Institute for Structural and Systems Biology, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany

³German Electron Synchrotron Centre (DESY), Hamburg, Germany

⁴Department of Structural and Computational Biology, Max Perutz Labs Vienna, University of Vienna, Vienna, Austria

⁵Instituto de Investigação e Inovação em Saúde (i3S) and Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal

⁶Laboratory of Molecular Biology, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, Bethesda, Maryland

⁷Department of Biophysics, Institute of Biomaterials and Biomolecular Systems, University of Stuttgart, Stuttgart, Germany

⁸Department of Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia

⁹European Molecular Biology Laboratory (EMBL), Hamburg Unit, Hamburg, Germany

Correspondence

Thomas C. Marlovits, University Medical Center Hamburg-Eppendorf (UKE), Institute for Structural and Systems Biology, Notkestrasse 85, D-22607 Hamburg, Germany.
Email: marlovits@marlovitslab.org

Present address: Vadim Kotov, European Molecular Biology Laboratory (EMBL), Hamburg Unit, Notkestrasse 85, D-22607 Hamburg, Germany.

Funding information

Austrian Science Fund, Grant/Award Numbers: I1593, I525, P19060, P22276, W1221; Behörde für Wissenschaft, Forschung und Gleichstellung of the city of Hamburg; Boehringer Ingelheim Fonds; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: INST 152/777-1, INST 152/772-1, INST 152/774-1, INST 152/775-1, INST 152/776-1; Fundação Luso-Americana para o Desenvolvimento; Fundação para a Ciência e a Tecnologia, Grant/Award Number: SFRH/

Abstract

Protein stability is a key factor in successful structural and biochemical research. However, the approaches for systematic comparison of protein stability are limited by sample consumption or compatibility with sample buffer components. Here we describe how miniaturized measurement of intrinsic tryptophan fluorescence (NanoDSF assay) in combination with a simplified description of protein unfolding can be used to interrogate the stability of a protein sample. We demonstrate that improved protein stability measures, such as apparent Gibbs free energy of unfolding, rather than melting temperature T_m , should be used to rank the results of thermostability screens. The assay is compatible with protein samples of any composition, including protein complexes and membrane proteins. Our data analysis software, MoltenProt, provides an easy and robust way to perform characterization of multiple samples. Potential applications of MoltenProt and NanoDSF include buffer and construct optimization for X-ray crystallography and cryo-electron microscopy, screening for small-molecule binding partners and comparison of effects of point mutations.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

BD/123761/2016; National Institute of Diabetes and Digestive and Kidney Diseases; Vienna Science and Technology Fund, Grant/Award Number: LS17-008

KEYWORDS

buffer optimization, high-throughput screening, melting temperature, MoltenProt, NanoDSF, protein unfolding, thermostability

1 | INTRODUCTION

Functional, biochemical or structural analyses of proteins and protein complexes require stable and monodisperse protein samples. A widely used proxy for overall protein stability is thermostability.¹ In a typical assay (Figure 1 (a)) a protein sample is gradually heated up, and the fraction of unfolded molecules is monitored using a spectroscopic or calorimetric readout. The inflection point of the

unfolding curve, also known as the melting temperature (T_m), is determined from the experimental derivative or curve fitting. Protein samples with higher T_m tend to be more stable and perform well in biochemical and structural biology research.^{1–4} Thermostability measurements are robust, consume low amounts of protein material and do not require expensive equipment. Furthermore, resulting T_m values are highly reproducible⁵ and agree across orthogonal assays, such as differential scanning

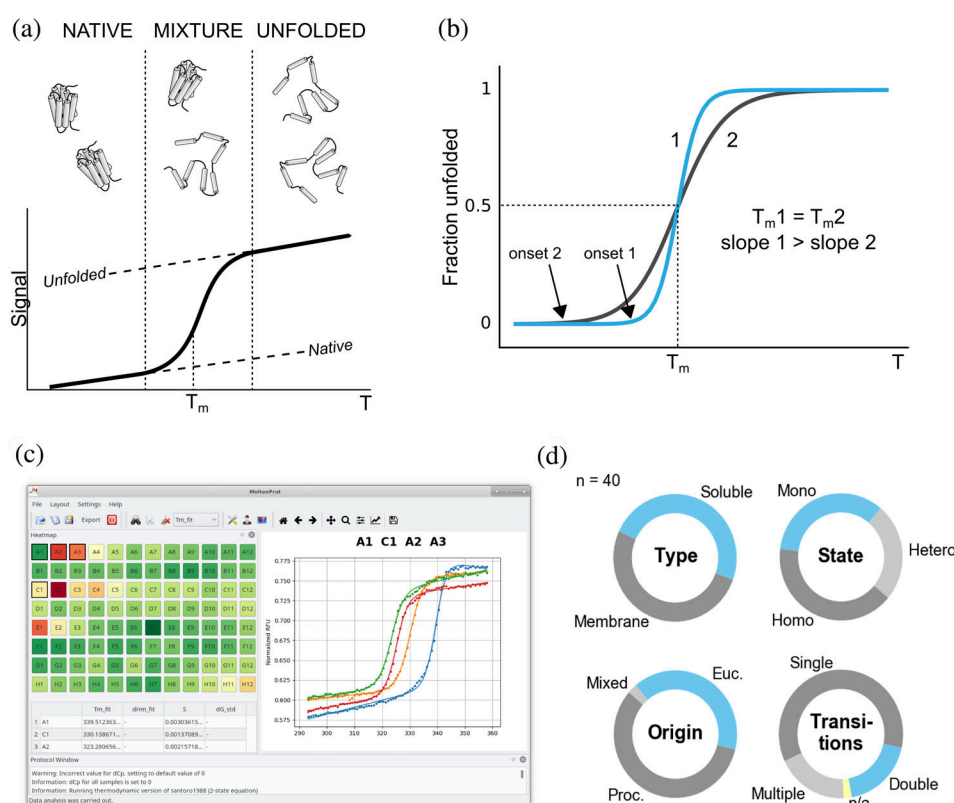


FIGURE 1 MoltenProt is a fast and efficient tool for analysis of protein unfolding data. (a) A typical thermal unfolding assay. Signal increases as a protein unfolds with increasing temperature (T). The inflection point of the curve (melting temperature T_m) indicates the temperature where 50% of protein molecules are in an unfolded state. Dashed lines indicate linearly extrapolated baselines, that is, temperature dependence of native or unfolded state fluorescence. Vertical dashed lines denote the transition region (1%–99% of protein molecules are unfolded). (b) Interpretation of slope values for thermal unfolding curves. Curves 1 and 2 have identical T_m , but curve 1 has a steeper (higher) slope. A steeper slope translates to a later onset of unfolding, that is, the temperature at which an arbitrary fraction of protein molecules becomes unfolded. Thus, the folded state of the protein in condition 1 is more resistant to heat, and it could be considered more stabilizing than condition 2. Importantly, these conditions are identical in terms of T_m . (c) A screenshot of the MoltenProt GUI. Left-hand side presents the samples in a 96-well format color-coded for a chosen fit parameter (e.g., $\Delta G_u^{\circ'}$, T_m or ΔH_m). Right-hand side displays one or more curves for comparison and assessment of fit quality. (d) Diversity of tested protein samples ($n = 40$) by type, oligomeric state, origin and the number of unfolding transitions. State indicates whether a protein sample forms oligomers (homooligomers or heterooligomers) or is a monomeric protein. Mixed origin indicates protein complexes with subcomponents from different species (e.g., antigen–antibody complex)

calorimetry (DSC)⁶ or circular dichroism measurements (CD).⁴ This makes thermostability an excellent choice for screening buffer conditions and optimization of protein constructs.^{4,7–9}

To date, high-throughput approaches for thermostability measurements, such as ThermoFluor¹⁰ or NanoDSF,^{11,12} were primarily focused on determination of T_m . This approach was criticized earlier,⁷ and the importance of the slope of the unfolding curve was highlighted (Figure 1(b)). Indeed, apart from the baselines, the simplest sigmoidal curve can be described by two values: (a) inflection point (T_m) and (b) the slope of the unfolding curve, that is, how steep or flat the transition is. Characterizing these curves with two parameters also presents a challenge on how to rank the results,¹³ since the parameters may have different weights and changes in parameters may contradict each other. To address this issue, Chari et al.⁷ proposed an empirical “hierarchical sorting” procedure, where the results are sequentially sorted and filtered by three curve parameters (quality of the fit, slope, and T_m). This approach assigns discrete numbers to final hits, so the differences between sample stabilities are qualitative and not quantitative. Furthermore, the sorting is done in a specific sequence of steps, thus assigning arbitrary weights to parameters. Finally, “hierarchical sorting” assesses $1/4$ of the data, disregarding the majority of the stability-related information and was thus recommended exclusively as a “qualitative guideline”.⁷

In this work we introduce and validate a quantitative approach for ranking results of thermal unfolding screens and describe a software package, MoltenProt, that streamlines the data analysis pipeline and makes it easy to use. In addition, we discuss several case studies to prove the utility of MoltenProt in structural and biochemical research.

1.1 | Overview of available techniques

Diverse assays are available to monitor chemical and thermal unfolding of proteins and obtain thermodynamic parameters (Table S1). DSC provides direct estimates of protein ΔC_p , ΔH_m , and T_m , however, the assay consumes high amounts of protein material (2.5–2,500 μg for a single measurement, depending on the sample) and cannot be efficiently parallelized. Furthermore, unstable buffer components, such as dithiothreitol (DTT) and tris (2-carboxyethyl)phosphine (TCEP), degrade during heating and distort the signal by affecting the buffer baseline.¹⁴ CD measurements take the advantage of chirality in amino acids to quantify the secondary structure content in proteins.¹⁵ CD is routinely used to monitor thermal and chemical unfolding of proteins,¹⁶ however, the assay requires all buffer components to be optically inactive in the ultraviolet

(UV) region.¹⁷ In particular, chloride ions, imidazole, DTT, dimethylsulfoxide (DMSO), glycerol, and several detergents absorb strongly in the far-UV region and should be avoided. Furthermore, CD was not adapted for high-throughput analysis and sample consumption can be high (2.25–2,250 μg for a single measurement, depending on the sample). High-throughput assays to assess protein thermostability include ThermoFluor^{10,18} and the 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM) dye assay.¹⁹ These assays use a fluorescent dye to report protein unfolding: in ThermoFluor, fluorescence increases when the dye binds to hydrophobic regions exposed during protein unfolding, while the CPM dye is only fluorescent upon reaction with exposed thiol groups. Protein sample is dispensed in a microplate, and efficient parallelization can be easily implemented. For both assays as low as 2–4 μg of protein per sample is sufficient, however, the increase in throughput is achieved by lower quality and information content of the signal. Previous work⁷ reported up to 30% relative error in ΔH_m estimation and 0.5 K error for T_m obtained with a modified ThermoFluor assay. SYPRO Orange, the most commonly used fluorescent dye for ThermoFluor, is not compatible with detergents and a number of other buffer components²⁰ as well as with heme proteins or proteins tagged with green fluorescent protein (GFP). Other dyes, such as 1-anilino-naphthalene-8-sulfonate (ANS) and GloMelt (manufactured by Biotium, San Francisco, California), were reported to be compatible with detergents.¹⁰ The CPM dye assay is limited to proteins with buried cysteine residues, and the nature of the signal was questioned in literature.²¹ Finally, fluorescent dyes that respond to changes in viscosity were successfully used to report protein unfolding.²²

Given the limitations outlined above, we chose measurements of intrinsic tryptophan (Trp) fluorescence in the UV range to obtain the best combination of sample consumption, reagent compatibility and data quality. This approach uses Trp residues as a built-in probe to report protein unfolding: in a hydrophobic environment Trp UV fluorescence peaks around 330 nm, and in a hydrophilic the fluorescence peak is around 350 nm. During unfolding a Trp residue would typically move from the hydrophobic core of the protein to the solvent, thus producing a red-shift in Trp fluorescence spectrum. A number of approaches to perform data reduction from a full spectrum to a single value was proposed¹²: (a) focusing on a single characteristic wavelength (e.g., fluorescence emission at 330 or 350 nm); (b) the difference between characteristic wavelengths; (c) ratio of characteristic wavelengths; (d) area under curve for the spectrum between specified wavelengths; (e) determination of the wavelength of maximum fluorescence intensity.

We chose Trp fluorescence measurements as the main readout to screen protein thermal stability for

several reasons. First, the nature of the signal is well-established and described.²³ Second, the signal readout is compatible with most common buffer components, most notably detergents. Third, protein sequences usually contain at least one Trp residue (1% estimated probability²⁴), and also other aromatic residues, for example, tyrosine (Tyr, 3% estimated probability²⁴), can report unfolding.^{25,26} Finally, recent developments in instrumentation (NanoDSF assay^{11,12}) provide high-resolution (temperature control accuracy 0.1°C) thermal unfolding curves for up to 48 samples in parallel with as low as 0.5–1 µg protein per sample. To speed up data acquisition, the fluorescence intensity is only measured at two wavelengths (330 and 350 nm with excitation at 280 nm, F330 and F350) instead of the full spectrum. Fluorescence ratio at 350 and 330 nm (F350/F330 ratio) is then used as a proxy for global changes in the Trp fluorescence spectrum. Protein samples have to be loaded into glass capillaries, however, in our experience, this does not limit the throughput and all 48 samples can be transferred to the device within 10 min.

It should be noted that any bulk spectroscopic measurement provides only a partial description of the whole protein molecule. Indeed, CD measurements report changes in the secondary structure, while Trp fluorescence reports alterations in the hydrophobic core. Thus, the parameters derived from curve analysis will characterize the respective properties of a “melting unit”,²⁷ that is, the molecular entity that elicits the observed signal. For the purposes of thermostability screening it can be assumed that overall protein stability positively correlates with the stability of a melting unit. The apparent nature of the values is indicated with a prime sign (') throughout the manuscript.

2 | RESULTS

2.1 | Ranking measures derived from common protein unfolding models

NanoDSF assay provides high resolution thermal unfolding curves with excellent throughput. However, the analysis implemented in the manufacturer's software (PR.ThermControl, NanoTemper GmbH, Munich, Germany) is limited to estimation of T_m and T_{onset} with a nondisclosed algorithm, which is presumably based on the analysis of the smoothened first derivative curve. In this section we use three models of protein unfolding (equilibrium thermodynamic, kinetic and empiric) to introduce measures of protein stability that capture all information available in the experimental data.

Protein unfolding is a complex multi-step process. Recent studies of bacteriorhodopsin, a small α -helical

membrane protein, revealed multiple intermediates of different properties and lifetime.^{28,29} A bulk spectroscopic measurement, as usually done in a thermal or chemical unfolding experiment, cannot capture this complexity, and the signal originating from the intermediate states is blended together.^{30,31} Unless there is a long-lived stable intermediate, and the transitions between states are widely separated, the unfolding curve would in many cases adopt a simple sigmoidal shape even for complex multimolecular assemblies. Thus, the problem of ranking the results of thermal unfolding screens can be often reduced to ranking of sigmoidal curves.

2.1.1 | Equilibrium thermodynamic model

Quantitative description of protein unfolding using classical thermodynamics is common and well-established.^{32–36} Core thermodynamic characteristics of a protein, including enthalpy of unfolding at T_m (ΔH_m), heat capacity change of unfolding (ΔC_p) and Gibbs free energy of unfolding (ΔG_u) can be obtained by thermal or chemical unfolding assays. The dependence of ΔG_u on temperature (T) is usually expressed as a variation of the Gibbs-Helmholtz equation:

$$\Delta G_u(T) = \Delta H_m \cdot \left(1 - \frac{T}{T_m}\right) + \Delta C_p \cdot \left(T - T_m - T \cdot \ln \frac{T}{T_m}\right)$$

ΔG_u is a particularly valuable measure, because it quantifies how thermodynamically favorable the folded state of the protein is under given conditions with higher ΔG_u corresponding to higher stability. ΔG_u calculations thus provide a natural way to combine the slope (ΔH_m) and inflection point (T_m) of a sigmoidal thermal unfolding curve into a single quantitative metric to rank the results. To ensure consistency with other sources of thermodynamic values,³⁷ which typically refer to some standardized conditions, we propose to use ΔG_u extrapolated to 298.15 K and standard pressure (100 kPa), that is, ΔG_u° .

2.1.2 | Kinetic model

Classical thermodynamic model of protein unfolding assumes that protein unfolding is a fully reversible reaction, and the equilibrium is reached at every step of a thermal unfolding experiment. In practice, proteins rarely exhibit such behavior, and a kinetic description of protein unfolding was proposed.^{38,39} In the simplest case of irreversible unfolding from native state to unfolded ($N \rightarrow U$), the fraction of state N (x_N) as a function of temperature is described as:

$$x_N(T) = \int_{T_{\min}}^{T_{\max}} \frac{-1}{v} \cdot \exp\left(\frac{-E_a}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_f}\right)\right) \cdot x_N$$

where $T_{\min}$ and $T_{\max}$ are the start and end temperatures of the measurement, v is the scan rate (degrees/min), E_a is the activation energy of unfolding, T_f is the temperature where reaction rate constant of unfolding (k_u) equals 1, R is the universal gas constant. x_N is assumed to be 1 at $T_{\min}$, that is, the protein is fully folded in the beginning of the measurement. Similar to the classical thermodynamic model described above, T_f and E_a are the core characteristics of the sigmoidal unfolding curve: T_f describes the overall positioning of the curve on the temperature axis, while E_a corresponds to the slope of the curve. The key difference is that depending on the scan rate T_f may or may not coincide with the inflection point of the sigmoidal curve (maximum of the first derivative). Since the rate constant of unfolding is inversely related to protein stability, we propose to use the negative logarithm of k_u extrapolated to 298.15 K (pk_u°) to rank the results of a thermal unfolding experiment described with this model:

$$\text{pk}_u^\circ = -\log \exp\left(\frac{-E_a}{R} \cdot \left(\frac{1}{298.15} - \frac{1}{T_f}\right)\right)$$

2.1.3 | Empirical model

The transition in thermal unfolding curves can be also described empirically by two characteristic temperatures: T_m (i.e., the mid-point of the transition) and onset temperature T_{onset} , which corresponds to the first detectable deviation of the experimental curve from the extrapolated linear baseline.^{40,41} Since T_m and T_{onset} are orthogonal characteristics of the curve, they can be combined into a single measure by computing the Euclidean distance from point with 0 K coordinates in a scatter plot of T_m and T_{onset} :

$$T_{\text{eucl}} = \sqrt{T_m^2 + T_{\text{onset}}^2}$$

In this case the sample that has the most optimal combination of both T_m and T_{onset} will have the highest T_{eucl} .

2.1.4 | Models of protein unfolding exhibit high rank-order correlation

We tested if there are differences in protein stability ranking between the three measures described above. We used a set of diverse thermal unfolding curves of the *E. coli* multidrug transporter MdfA⁴² in a variety of detergents⁴⁰ (see also “Protein stability heatmaps”

section below). The curves were fit to a classic thermodynamic, kinetic and empirical model using MoltenProt software (see next section), and the final ranking parameters ΔG_u° , pk_u° , and T_{eucl} were calculated (Figure S1a). The pair-wise scatter plots demonstrate strong linear correlation (Pearson correlation coefficient over 0.99) and suggest high rank order consistency (Kendall's τ in range 0.91–0.98). Thus, if the signal can be described with a sigmoidal curve, the same conditions will come up as the top hits in a thermal unfolding screen regardless of the choice of the protein unfolding model. To avoid redundancy, we will use ΔG_u° throughout the rest of the manuscript.

2.1.5 | Linear extrapolation of ΔG_u is sufficient for relative comparison

According to the Gibbs-Helmholtz equation, three thermodynamic parameters are required to calculate ΔG_u at a specific temperature. ΔH_m and T_m can be readily obtained from thermal unfolding assays,²⁶ while experimental estimation of ΔC_p usually requires multiple DSC runs and is not always feasible. ΔC_p primarily depends on the size of the protein,⁴³ and as the first approximation ΔC_p can be assumed independent of temperature, pH or buffer composition.³² Neglecting the contribution of ΔC_p to ΔG_u will introduce a systematic error to ΔG_u with the following dependence on T_m and reference temperature T_{ref} :

$$\text{Error}(T_{\text{ref}}, T_m) = \Delta C_p \cdot \left(T_{\text{ref}} - T_m - T_{\text{ref}} \cdot \ln \frac{T_{\text{ref}}}{T_m}\right)$$

Since $T_{\text{ref}} < T_m$, this error is always negative. Thus, assuming zero ΔC_p would produce an overestimate of ΔG_u , as demonstrated with chicken egg lysozyme in the next section. In a typical thermal unfolding screen one is not concerned with absolute values of protein stability, but rather in the relative ranking to identify the most stabilizing conditions. $\Delta \Delta G_u$ between a sample and a reference sample (e.g., wild-type protein, or original buffer condition) computed with ΔC_p being neglected will be close to the real $\Delta \Delta G_u$ as long as the difference in T_m is within 10–15 K.³² Unless otherwise stated, ΔC_p was not taken into account for ΔG_u° calculations.

2.2 | Thermodynamic values obtained with MoltenProt are reliable and agree with orthogonal assays

Application of the concepts outlined in the previous section to a high-throughput assay such as NanoDSF is

challenging, because it requires a robust nonlinear curve fitting procedure and an easy way to view the results. This prompted us to develop a software package MoltenProt⁴⁴ (Figure 1(c), Figure S1c), which we successfully used to characterize a large NanoDSF dataset.⁴⁰ Our previous work, however, did not compare the thermodynamic parameters obtained by the combination of NanoDSF and MoltenProt with other methods. To address this, we measured thermal unfolding profiles of model proteins (chicken egg lysozyme and ribonuclease A [RNase] from bovine pancreas) in the presence of increasing concentrations of chemical denaturants (urea and guanidine hydrochloride [GuHCl]). ΔH_m at zero denaturant concentration agreed well with literature data^{45,46} (Table S2). To further validate the result, we computed ΔC_p by performing a linear fit of the $\Delta H_m(T_m)$ dependence (Figure S2a) and found good agreement of obtained ΔC_p values with literature data^{45,46} (Table S2). Furthermore, in case of lysozyme we could obtain an independent estimate of $\Delta G_u^{\circ'}$ by fitting the values of F350/F330 ratio at 298 K at different denaturant concentrations (Figure S2b, Table S2). As expected, $\Delta G_u^{\circ'}$ (ΔC_p assumed zero) obtained from thermal unfolding was 17.9 kJ/mol higher than $\Delta G_u^{\circ'}$ from chemical unfolding. If ΔC_p was taken into consideration, then the $\Delta G_u^{\circ'}$ values agreed well (Table S2). These results demonstrate high consistency between thermal and chemical unfolding properties of model proteins when characterized with NanoDSF technique. A similar result was obtained previously for R16 α -spectrin domain.¹¹

We next analyzed thermal unfolding of human filamin C domain 19 (FlnC-d19, see also below) with DSC, CD, and Trp fluorescence (Figure S2c, S2d, Table S3). T_m and ΔH_m values agreed well for all assays. ΔC_p from Trp fluorescence data (4,846 J/mol/K) is similar to the DSC result (4,910 J/mol/K) and was lower than the theoretical value (5,742 J/mol/K, see Table S3).

Finally, we assessed the overall level of uncertainty in $\Delta G_u^{\circ'}$ estimation, which is derived from combined errors in ΔH_m and T_m . We measured several protein samples in replicates on different days (Table S4). Typically, standard deviation of the measurement was in the range of 0.3–0.9 K for T_m (below 0.5% relative error) and 10–50 kJ/mol for ΔH_m (typically, 3–10% relative error). This result highlights the high quality of the data obtained with Trp fluorescence. Previous work⁷ reported up to 30% relative error in ΔH_m estimation and 0.5 K error for T_m obtained with the modified Thermofluor assay.

2.3 | Case studies

2.3.1 | Ranking based exclusively on T_m may produce misleading results

We used MoltenProt to characterize thermal unfolding of a variety of proteins and protein complexes (Figure 1(d)). In many cases, the result ranking with either $\Delta G_u^{\circ'}$ or T_m gave similar results. For instance, the ExbBD complex⁴⁷ (Figure 2(a), Table S5) had a higher T_m and $\Delta G_u^{\circ'}$ at pH 5.5 compared to pH 6.9. In a number of cases, however, focusing on T_m was misleading.

KtrAB is a nonselective prokaryotic cation channel composed of a ring of four KtrA dimers that bind adenosine nucleotides (ATP, ADP, and AMP) and two KtrB subunits embedded in the membrane forming the channel⁴⁸ (Figure 2(d)). We analyzed thermostability of KtrAB in presence of two known ligands, ATP and AMP. While T_m was similar for both samples, $\Delta G_u^{\circ'}$ decreased two-fold in presence of AMP (Figure 2(c), Table S5). This agrees with biochemical data (functional activation with ATP, low affinity for AMP) and structural research (ATP-bound state of the complex produced crystals diffracting to 3.5 Å⁴⁸).

DgoT, a prokaryotic sugar transporter, was initially purified in the detergent DDM (0.03% final concentration),⁴⁹ and then the detergent was exchanged to 1.2% CHAPS. We observed a 15.6 K increase in T_m (from 321.3 K in DDM to 337.1 K in CHAPS), whereas $\Delta G_u^{\circ'}$ decreased two-fold (Figure 2(e), Table S5). The increase in T_m is somewhat unexpected, because CHAPS is a zwitter-ionic detergent, and this type of detergents is generally more harsh compared to nonionic ones, including DDM.⁵⁰ Furthermore, the protein could only be crystallized in maltoside detergents (DM, UDM, DDM, Figure 2(f)).

Taken together, these results indicate that T_m must be used with great care when selecting hits in protein thermal unfolding screens. Such screens are usually performed at early stages of a project, so following up a false-positive hit can result in significant loss of time and resources. This can be easily avoided by combining inflection point and slope of the sigmoidal curve in a simple quantitative measure, such as $\Delta G_u^{\circ'}$.

2.3.2 | Functional insights obtained from $\Delta G_u^{\circ'}$

Measuring protein characteristics obtained with thermal unfolding assays can provide useful insights about their functional interactions. ExbBD is part of the Ton complex

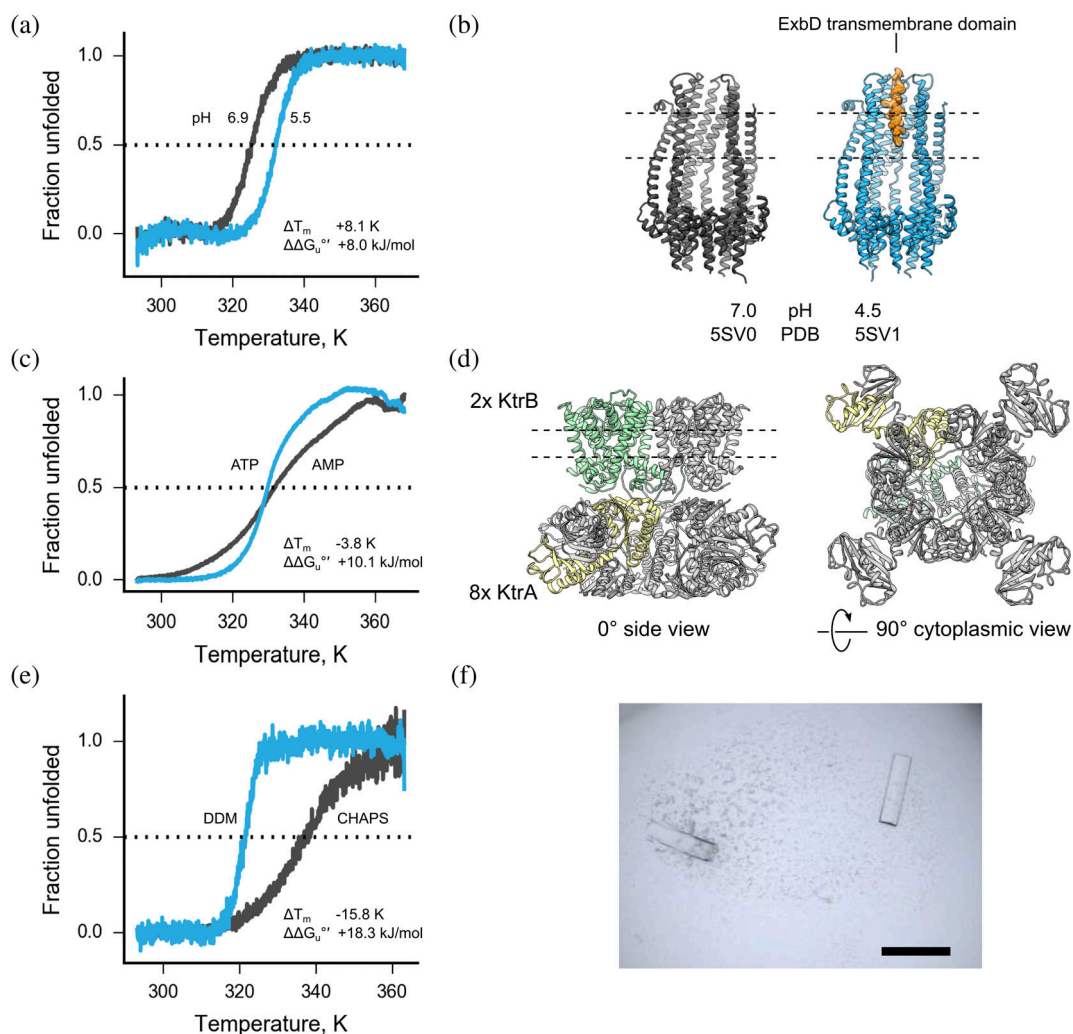


FIGURE 2 ΔG_u is a superior thermostability metric compared to T_m . (a) Representative thermal unfolding curves of ExbBD complex at pH 5.5 (blue) or pH 6.9 (gray). Raw fluorescence readings are provided in Figure S2e. (b) Crystal structure of ExbBD at pH 4.5 (blue) and pH 7.0 (gray). The transmembrane domain of ExbD (orange) is only visible in crystals formed at pH 4.5, which have higher thermostability. Two ExbB subunits were removed from display to visualize the lumen of the channel. (c) Representative thermal unfolding curves of KtrAB complex in presence of ATP (blue) or AMP (gray). Raw fluorescence readings are provided in Figure S2f. (d) Crystal structure of KtrAB in complex with ATP (PDB 4J7C). Two KtrB subunits are buried in the membrane and perform cation transport; nucleotide-bound KtrA forms a tetramer of dimers. The approximate position of the lipid bilayer is shown with dashed lines. One KtrB subunit is shown in green and one KtrA subunit is shown in yellow. (e) Representative thermal unfolding curves of DgoT in presence of DDM (blue) or CHAPS (gray). Raw fluorescence readings are provided in Figure S2g. (f) Crystals of the transporter DgoT in purified in 0.03% DDM and crystallized in 0.03 M $MgCl_2$, 0.1 M MES pH 6.5, 28% polyethyleneglycol-400 (PEG-400). Scale bar 200 μm . In thermal unfolding curves the intersection with the horizontal dashed line at 0.5 fraction unfolded corresponds to T_m . ΔT_m and $\Delta \Delta G_u$ are computed by subtraction of the value for a less stable state (gray unfolding curve) from the value of a more stable state (blue unfolding curve). Specific values as well as information on the replicates are provided in Table S5

embedded in inner membranes of Gram-negative bacteria and plays an important role in nutrient uptake.⁴⁷ Conductivity measurements indicated that the complex is fully functional at pH above 6.5, while lower pH values decrease conductivity.⁴⁷ We compared thermostability of ExbBD at different pH values using 100 mM broad-range buffer system SPG (succinate/phosphate/glycine mixed in molar ratio 1:4:3), and observed higher stability at

lower pH values (Figure 2(a), Table S5). This indicates that the closed state of ExbBD is thermodynamically more stable, which might be attributed to the stabilization of the ExbD transmembrane domains in the channel formed by the ExbB pentamer (Figure 2(b)).

RuvB is an ATPase that resolves the Holiday junction.^{51,52} Similar to many ATPases, it assembles to a functional hexamer upon addition of ATP or its

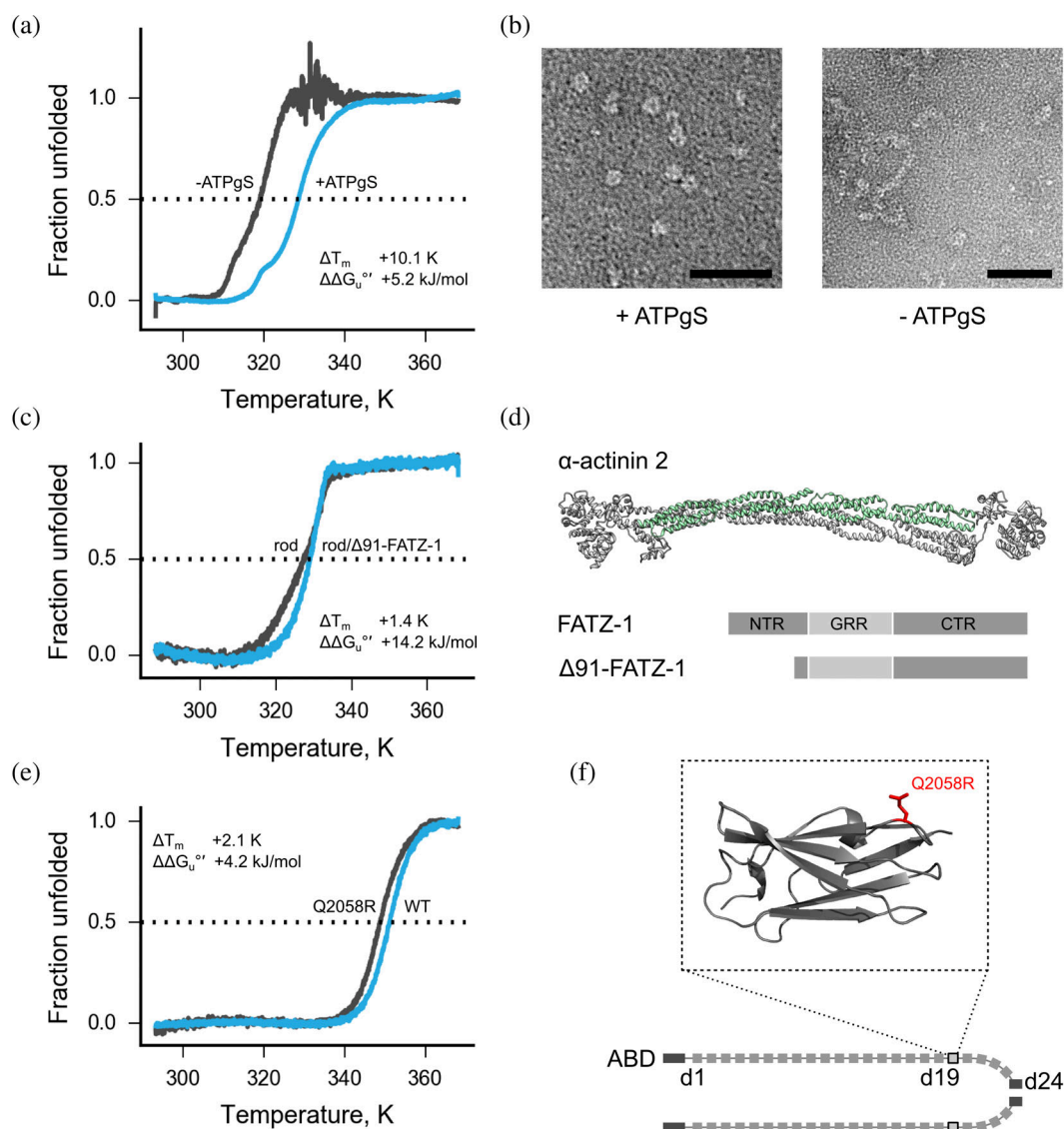


FIGURE 3 Functional insights obtained by ΔG_u^{or} analysis. (a) Representative thermal unfolding curves of RuvB in presence or absence of ATPgS (blue and gray). Raw fluorescence readings are provided in Figure S3a. (b) RuvB in presence or absence of ATPgS visualized with negative staining electron microscopy (representative images, scale bar 50 nm). Full-size micrographs are provided in Figure S3b. (c) Representative thermal unfolding curves of α -actinin-2 rod domain (gray) and its complex with the intrinsically disordered protein $\Delta 91$ -FATZ-1 (blue). Raw fluorescence readings are provided in Figure S3c. (d) Crystal structure of α -actinin-2 dimer (PDB 4D1E) and a schematic representation of FATZ-1. Rod domain (residues 274–746) of one α -actinin-2 subunits is shown in green. $\Delta 91$ -FATZ-1 construct contains a deletion of 91 N-terminal residues. NTR, N-terminal region; CTR, C-terminal region; GRR, glycine-rich region. (e) Representative thermal unfolding curves of wild type FlnC-d19 (blue) and cardiomyopathy causing mutation Q2058R (gray). Raw fluorescence readings are provided in Figure S3e. (f) Schematic diagram of FlnC and a model of FlnC-19 bearing a cardiomyopathy-causing mutation Q2058R (inset, residues 2036 to 2,130; Q2058R is highlighted in red). The model is based on a partially refined crystal structure of FlnC domains 18 and 19 (Mlynek et al., manuscript in preparation). In thermal unfolding curves (a, c, e) the intersection with the horizontal dashed line at 0.5 fraction unfolded corresponds to T_m . ΔT_m and $\Delta \Delta G_u^{\text{or}}$ are computed by subtraction of the value for a less stable state (gray unfolding curve) from the value of a more stable state (blue unfolding curve). Specific values as well as information on the replicates are provided in Table S5

analogs. We compared thermostability of RuvB from *Streptococcus thermophilus* in presence and absence of 1 mM adenosine 5'-(γ -thio)-triphosphate (ATPgS), a slowly hydrolysable ATP analog. The enzyme was heavily destabilized without its ligand and produced a weak signal in the thermal unfolding assay, whereas a

ligand-bound state was more stable (Figure 3(a)). These results agree well with electron microscopy analysis of the sample, where ATPgS-free protein showed mainly aggregated particles, while ATPgS-bound RuvB demonstrated a homogeneous population of ring-like particles (Figure 3(b)).

FATZ-1 (calsarcin-2/myozenin-1) is an intrinsically-disordered protein found in the Z-discs of skeletal muscles, the boundaries between two adjacent sarcomeres. FATZ-1 binds to the major Z-disc protein α -actinin-2, which functions as a homodimer.^{53–56} We generated a complex of FATZ-1 lacking the first 91 residues ($\Delta 91$ -FATZ-1) with the rod domain of α -actinin-2 (Figure 3(d)). The complex consists of two $\Delta 91$ -FATZ-1 molecules bound to one rod domain dimer, as demonstrated by size-exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS, Figure S3d). In thermal unfolding assays $\Delta G_{u^{\circ}}$ of the complex was almost two-fold higher than $\Delta G_{u^{\circ}}$ of the rod domain alone (Figure 3(c), Table S5) confirming the functional interaction between the proteins.

Filamin C (FlnC) is located at premyofibrils, myofibrillar Z-discs and myofibrillar attachment sites of striated muscle cells, where it is involved in mechanical stabilization, mechanosensation and intracellular signaling.^{57–59} FlnC is a homodimer, where each subunit consists of an N-terminal actin-binding domain (ABD) followed by 24 immunoglobulin-like (Ig-like) domains (Figure 3(f)). Mutations in FlnC give rise to skeletal muscle diseases and cardiomyopathies.⁶⁰ We investigated stability and fold integrity of FlnC mutant Q2058R implicated in cardiomyopathy located in the Ig-like domain 19 of FlnC (FlnC-d19, Figure 3(f), inset). $\Delta G_{u^{\circ}}$ of wild-type FlnC-d19 was 53.0 kJ/mol, while in the case of the Q2058R mutant it decreased by 4.1 kJ/mol, suggesting that a single mutation can destabilize FlnC-d19.

2.3.3 | Protein stability heatmaps

Minimal sample consumption by modern equipment for thermal unfolding measurements combined with MoltenProt analysis is an excellent tool for sampling and visualizing protein stability as a function of various conditions. Figure 4(a) represents a heatmap view of stability of the TOM core complex, a protein precursor entry gate in eukaryotic mitochondria,⁶¹ in a variety of buffer systems and pH values. The optimal pH value for the TOM core complex was specific to the buffer system used. Remarkably, the stability of the complex was independent of pH in phosphate buffer, which could be attributed to the salting-in effect of phosphate ions.^{62,63} In most other buffer systems the optimal pH range for the TOM core complex is narrow with an overall trend towards moderately acidic pH values.

Due to the broad reagent compatibility, Trp fluorescence is widely used to assess stability of membrane proteins.^{49,64,65} Figure 4(c) presents the $\Delta G_{u^{\circ}}$ heatmap of MdfA in a variety of detergents. We identified two

important trends: (a) in agreement with a broad body of evidence,⁶⁶ detergents with longer alkyl chain lengths increase the stability of MdfA within a single detergent family; (b) detergent headgroup chemistry defines the stabilizing properties of the detergent. In particular, neopentyl-glycol detergents,^{67,68} albeit with relatively short alkyl chains, promote stability of MdfA, whereas zwitterionic fos-choline detergents are harsh towards the protein regardless of the chain length. We observed a similar trend by testing additional eight membrane proteins.⁴⁰ This result suggests that in order to identify a suitable detergent for a particular protein it is sufficient to screen a limited number of common detergent families and then optimize the alkyl chain length within the best-ranking detergent family. If detergents with shorter alkyl chains are not available, organic solvents known to decrease micelle size (e.g., ethanol or heptanetriol⁶⁹) can be tested.

2.3.4 | Signal detection in lieu of Trp residues

Trp belongs to one of the most rare residues found in protein sequences.²⁴ Furthermore, only the Trp residues that change their environment (hydrophobic to hydrophilic, but also vice versa) over the course of an unfolding experiment will provide a usable signal. Tyr and phenylalanine (Phe) residues are also capable of reporting protein unfolding, as demonstrated previously^{25,26} and also in this work with RNase (Tables S2 and S5). To gain further insights in the assay requirements we measured thermal unfolding of ubiquitin (Table S5), which contains only two Phe and one Tyr residue. The protein had to be used in concentrations above 0.5 mg/ml in order to obtain sufficient fluorescence counts. In agreement with the published data,⁷⁰ the thermostability of ubiquitin was very high. The full unfolding curve could not be recorded in the available temperature range of the device (20–95°C). Due to the absence of the posttransition baseline T_m was estimated from the peak of the derivative curve and was 365 K. Thus, even a protein with very few fluorescent residues can be characterized with NanoDSF. It should be noted, however, that according to the Fourier transform infrared spectroscopy analysis of ubiquitin unfolding, the molecule does not undergo significant unfolding at the temperatures above 85°C, but rather forms aggregates via intermolecular β -sheets.⁷⁰

3 | DISCUSSION

The ultimate objective of a thermal unfolding screen is to boil down a large set of conditions (typically, 100–300) to

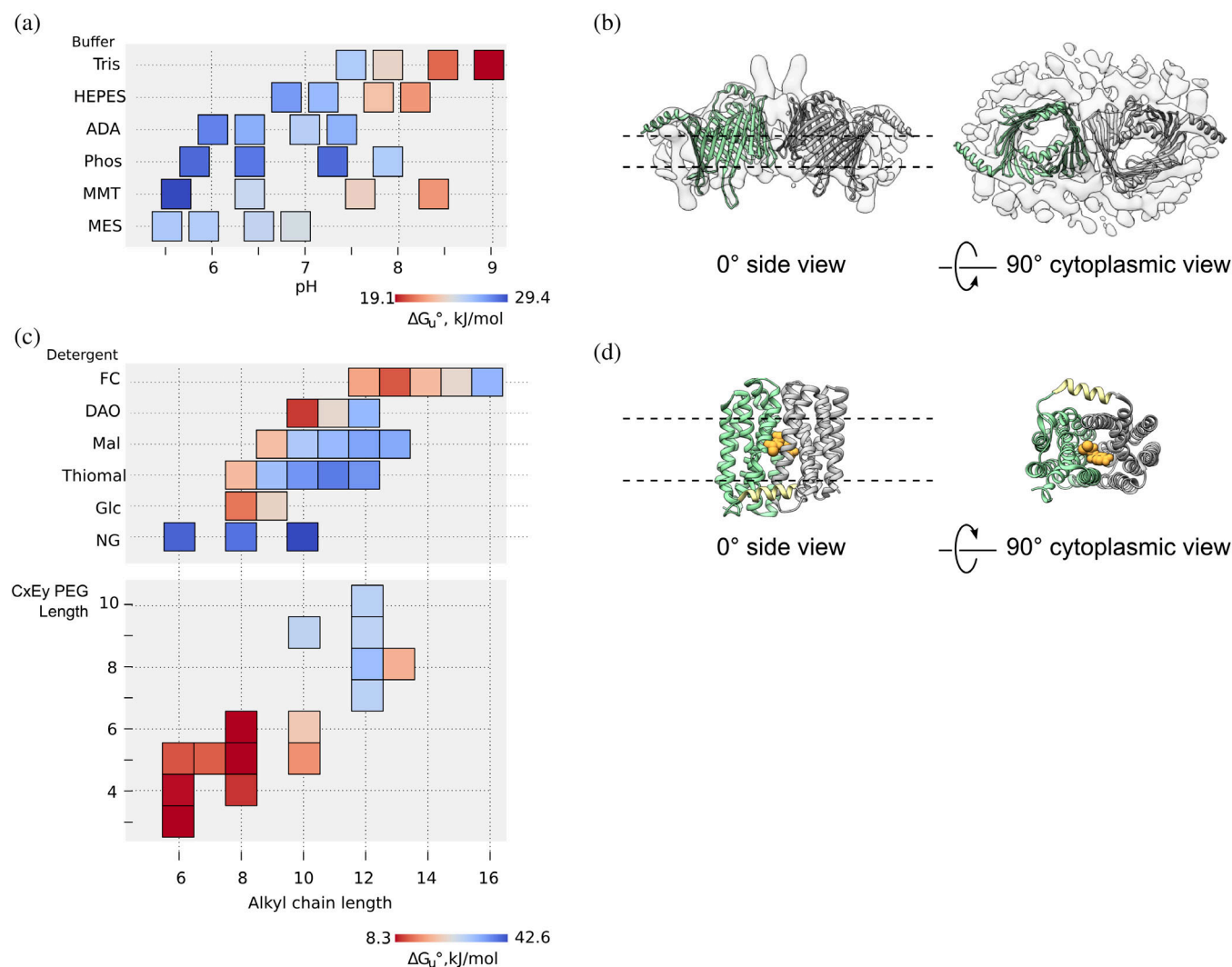


FIGURE 4 Exemplary $\Delta G_u'$ heatmaps. (a) Stability of the TOM core complex as a function of pH and buffer system. Each square in the heatmap represents an experimental measurement of $\Delta G_u'$ ($n = 2$). The middle point of each square is positioned on either categorical axis (buffer system) or numeric axis (pH). Indicated buffer system was added to a final concentration of 100 mM to the original buffer system (20 mM HEPES, pH 7.2). (b) Structure of the TOM core complex. Two copies of Tom40 model (PDB 5O8O, green and gray) were placed in the 6.8 Å electron density map (EMDB 3761). The approximate position of the lipid bilayer is shown with dashed lines. (c) Stability of MdfA as a function of detergent chemistry. Each square in a heatmap represents an experimental measurement of $\Delta G_u'$ ($n = 2$). The middle point of each square is positioned on either categorical axis (detergent family) or numeric axis (alkyl chain length). Lower heatmap shows the heatmap for alkyl ether detergents (CxEy), where the length of alkyl chain (Cx) is on X-axis and the length of the hydrophilic PEG component (Ey) is on Y-axis. FC, fos-choline; DAO, dimethylamineoxide; Mal, maltoside; Thiomal, thiomaltoside; Glc, glucoside; NG, maltose neopentyl-glycol. (d) Crystal structure of MdfA in complex with chloramphenicol (PDB 4ZOW). The protein contains two pseudo-symmetric domains (green and gray) and the amphipatic interdomain linker (yellow). Bound chloramphenicol is shown in orange; the approximate position of the lipid bilayer is shown with dashed lines

a smaller set for in-depth characterization with orthogonal low-throughput techniques. Thus, in many cases the simplicity of analysis is preferred over accuracy, and T_m is so far the most widely used readout for thermal unfolding assays. While being a reliable indicator of protein stability with proved empirical performance, T_m only partially describes the unfolding process, and a situation where T_m -based ranking produces a misleading result can be easily modeled.

In this work we introduce an improved framework for quantitative characterization of protein stability from thermal unfolding assays. We derive new measures from classic thermodynamic, kinetic and empirical models of protein unfolding that incorporate all characteristics of a sigmoidal curve (inflection point and slope). We show that these measures can be used interchangeably and help avoid the pitfalls that are possible when samples are ranked exclusively with melting temperature T_m .

Application of outlined concepts into practice, particularly, with the purpose of high-throughput analysis, may be challenging in terms of data analysis and management. To aid this task we designed an open source package MoltenProt, which aims to minimize human intervention and present the user with a minimal set of decision points. Currently MoltenProt supports Trp fluorescence data from Prometheus NT.48 as well as plain text formats (e.g., for analysis of CD data). The open source nature of the code allows easy tailoring and extension of MoltenProt for custom applications, for example, new types of readouts. We believe that analysis of label-free thermal and chemical unfolding data with MoltenProt will help biochemists, biophysicists and structural biologists to systematically assess the stability of their samples and serve as a guide to better structures and faster results. Potential applications include buffer and construct optimization for structural and biochemical studies as well as validation of existing buffer formulations, screening for small-molecule binding partners and comparison of effects of point mutations.

4 | MATERIALS AND METHODS

4.1 | Protein samples

Chicken egg lysozyme and RNase A were purchased from Sigma-Aldrich (Saint Louis, Missouri). Lysozyme was dissolved in 50 mM sodium phosphate buffer, pH 7.5, 150 mM NaCl at concentration 20 mg/ml. RNase A was dissolved in 10 mM Tris-HCl, pH 7.5, 15 mM NaCl at concentration 20 mg/ml.

DgoT was expressed and purified as described previously.⁴⁹ Protein crystals formed at DgoT concentration 5 mg/ml in 0.03 M MgCl₂, 0.1 M MES, pH 6.5, 28% v/v PEG-400 using the vapor diffusion technique (ratio 1:1) at 20°C.

Purification and detergent screening of MdfA was described previously.⁴⁰

ExbBD was purified as described previously.⁴⁷ The buffer system/pH were changed by adding a new buffer system to final concentration 100 mM.

KtrAB was purified and assembled as described previously.⁴⁸ The nucleotides were added to KtrA at the affinity chromatography step (5 mM ATP or 50 mM AMP). SEC of KtrA was performed in absence of any nucleotides; pooled fractions were supplemented with either 1 mM ATP or 10 μ M AMP and dialyzed against 50 mM Tris-HCl, pH 7.5, 150 mM KCl, 5 mM DTT with either 1 mM ATP or 10 μ M AMP. KtrA-ATP or KtrA-AMP was added to KtrB at the affinity chromatography step of KtrB

purification in buffer supplemented with 1 mM ATP or 10 μ M AMP. The complex was dialyzed and further purified by SEC in 20 mM Tris-HCl, pH 8.0, 120 mM NaCl, 30 mM KCl, 5 mM DTT, 1.5 mM CYMAL-6 without any nucleotide added.

Coding sequence of RuvB from *S. thermophilus* was cloned into pET-52b(+) vector and purified using nickel affinity chromatography followed by tag cleavage with tobacco etch virus protease and size-exclusion chromatography at Superdex-200 column (GE Healthcare, Chicago, Illinois) equilibrated with 100 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.5 mM DTT, 15% w/v glycerol.

TOM core complex was purified as described previously.⁶¹ The pH/buffer system screen was prepared in-house (stock concentration 0.5 M). Three microliter of the screen were added to 12 microliter of 0.25 mg/ml TOM core complex in 20 mM HEPES, pH 7.2, 2% v/v DMSO, 350 mM KCl, 0.1% w/v DDM. In addition, a triplicate of the sample in the original buffer was measured to assess the uncertainty in curve parameter estimation and a water control to assess the destabilizing effect of dilution, where an equal volume of ultra-pure water was added instead of the screen.

FlnC-d19, Δ 91-FATZ-1, and α -actinin-2 rod were prepared using the platform and protocols described previously.⁷¹ The detailed purification protocols for individual samples will be published elsewhere (Mlynek et al., manuscript in preparation). Protein concentration was determined from absorbance at 280 nm. Where applicable, extinction coefficients were calculated from the primary amino acid sequence using ProtParam.⁷²

Ubiquitin with a C-terminal His-tag (sequence MAS-MTGGQQMGRGSMQIFVKTLTGKITLEVEPSDTIENV-KAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYNIQK-ESTLHLVLRRLRGKKLAAALEHHHHHHH) was cloned in pET21a vector and purified in 10 mM sodium/potassium phosphate buffer, 137 mM NaCl, 2.7 mM KCl using Ni-affinity chromatography followed by size-exclusion chromatography on Superdex-75 column (GE Healthcare, Chicago, Illinois).

4.2 | Trp fluorescence measurements (NanoDSF assay)

Prometheus NT.48 (NanoTemper GmbH, Munich, Germany) was used to run all Trp fluorescence measurements. Standard-grade glass capillaries were filled with 10–15 μ l of the sample, excitation light was preadjusted to get fluorescence readings above 2000 arbitrary units for F330 and F350, and samples were measured in temperature range 20–95°C or 15–95°C with temperature

slope of 1°C/min. Up to 48 samples could be measured simultaneously. Prior to the measurements samples were centrifuged for 10 min at 16000g at 4°C to remove any large aggregates.

4.3 | CD spectroscopy

Electronic circular dichroism was measured using Chirascan (Applied Photophysics, Leatherhead, UK). Prior to the measurements the xenon arc lamp, monochromator and sample chamber were flushed with nitrogen. Temperature was increased stepwise in range 20–95°C with slope of 1°C/min using a Peltier element. UV–vis absorbance and CD were measured at path length 0.5 mm, with 1 nm spectral bandwidth and 7 s scan time. FlnC-d19 was measured at concentration 0.59 mg/ml in 20 mM phosphate buffer, 150 mM sodium fluoride, pH 7.4.

CD at 208 and 218 nm as a function of temperature was processed in MoltenProt as described below.

4.4 | DSC

To achieve identical buffer composition for reference and sample cells every sample was buffer exchanged using a Superdex-200 5/150 GL or Superdex-200 10/300 Increase column (GE Healthcare, Chicago, Illinois) connected to an HPLC 1260 Infinity with fraction collector (Agilent Technologies, Santa Clara, CA).

DSC experiments were performed with the PEAQ Differential Scanning Calorimeter Automated (Malvern Panalytical, Malvern, UK). Experiments were performed under increased pressure (~62 psi) to prevent the solutions from boiling. At the beginning of an experiment, at least four buffer runs (i.e., buffer in both sample and reference cells) were performed to establish the thermal history of the cells and to collect optimal buffer scans for buffer subtraction.

A temperature slope of 1°C/min was used for all experiments. Upon scanning from 20°C to 130°C and reheating again, we did not encounter any protein refolding. Therefore, the reheated run (i.e., rescan) was taken for buffer subtraction. Data analysis was performed with the PEAQ-DSC analysis software (Malvern Panalytical, Malvern, UK). The baseline was fitted with the spline baseline correction model to account for differences in the heat capacities of the folded and unfolded states of the protein. Concentration normalization was performed and transitions were fitted with a two-state thermal unfolding model.

4.5 | Characterization of lysozyme and RNase a unfolding

Eight molar stock of GuHCl was prepared by mixing 7.64 g of GuHCl with 4.21 ml assay buffer (50 mM sodium phosphate buffer, pH 7.5, 150 mM NaCl). pH of GuHCl solution was further adjusted to pH 7.5 using 1 M Tris, pH 8.0. Nine molar stock of urea was prepared freshly by mixing 5.41 g urea with 5.9 ml 1x assay buffer. One microliter of concentrated protein stock (final concentration 0.1 mg/ml for lysozyme and 0.5 mg/ml for RNase A) was added to 40 µl of series of denaturant concentrations and the mixture was incubated for 1 hr and 16 hr at 25°C.

Thermal unfolding data were processed in MoltenProt as described below to obtain ΔH_m and T_m values at different denaturant concentrations. Scatter plots of ΔH_m as a function of T_m demonstrate a clear linear dependence (Figure S2a). According to the definition of ΔC_p :

$$\left(\frac{\delta\Delta H}{\delta T}\right)_p = \Delta C_p$$

the slope of this line corresponds to the ΔC_p value.

Readings of fluorescence at the start of thermal unfolding measurement as a function of denaturant concentration were fit to the equation described by Santoro and Bolen⁷³ to calculate m (changes in ΔG_u for each 1 M of denaturant) and $\Delta G_u^{\circ'}$ at zero denaturant concentration.

Curve fitting was performed using `scipy.curve_fit` function.

4.6 | Electron microscopy

Four microliter of RuvB in 20 mM Tris, pH 8.0, 20 mM NaCl, 10 mM MgCl₂, 2% glycerol, 0.5 mM DTT with or without 1 mM ATPγS was applied onto home-made glow-discharged carbon-coated Cu/Pd grids and incubated for 30 s. The grids were stained with 2% uranyl acetate for 30–40 s, blotted and visualized in Talos L120C (FEI Company, Hillsboro, Oregon) microscope using TEM Imaging and Analysis Software (FEI Company, version 4.15).

4.7 | SEC-MALLS

Analytical SEC was performed using HPLC 1260 Infinity (Agilent Technologies, Santa Clara, California), linked to a Superdex-200 10/300 Increase column (GE Healthcare,

Chicago, Illinois). Column was preequilibrated with buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 50 mM arginine, 50 mM glutamic acid, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM EDTA, 1 mM β -mercaptoethanol). Hundred microliter of sample was injected at concentration between 2 and 4 mg/ml and the separation was performed at 20°C using a flow rate of 0.5 ml/min.

Online MALLS detection was performed with a miniDawn Treos detector (Wyatt Technology, Santa Barbara, California) using a laser emitting at 690 nm. Protein concentration was measured by refractive index using a Shodex RI-101 (Showa Denko, Tokyo, Japan) using a typical refractive index increment (dn/dc) of 0.186 ml/g.⁷⁴ Average molecular weight was calculated with Astra software (Wyatt Technology, Santa Barbara, California).

4.8 | High-throughput analysis of thermal unfolding data

Thermal unfolding and aggregation curves were read by MoltenProt directly from spreadsheet files generated by Prometheus NT.48 control software (PR.ThermControl, NanoTemper GmbH, Munich, Germany). Derivatives of experimental curves were computed with Savitzky-Golay filter (as implemented in `scipy.signal.savgol_filter`) using 4-th order polynomial and window size of 10 K. Derivative curves were only used to provide an initial estimate for T_m .

In most cases Prometheus NT.48 produces high quality data, which can be used directly for curve fitting and analysis. For suboptimal curves MoltenProt provides preprocessing steps to “rescue” the data. These include removal of data points in the start or end of curves, curve smoothing with median filter and data binning to suppress noise and expose trends.

Nonlinear fitting requires estimation of starting values for all parameters, and good starting values increase the convergence of the algorithm and may also prevent reaching the wrong local minimum. To estimate starting values for pre- and posttransition baseline parameters MoltenProt runs linear fitting of 10 K stretches in the beginning and the end of the thermal unfolding curve. Initial values of the baselines are also used to determine if the curve is S-shaped (signal increases with temperature) or Z-shaped (signal decreases with temperature). This information is important for obtaining an initial value for T_m : if the signal increases with temperature, then the temperature corresponding to the local maximum of the derivative is taken as an initial value. Alternatively, the local minimum of the derivative curve is used. Obtaining robust initial values for ΔH_m is not straightforward, so the initial value was always set to 100,000 J/mol.

Curve fitting routine supports setting parameter bounds (as part of Trust Region Reflective algorithm implemented in `scipy.curve_fit`). For T_m the bounds were set to the temperature range of the experiment. ΔH_m was restricted to 60–4,000 kJ/mol. The bounds for baseline parameters were not set.

Several values were used to characterize the fitting results and filter out suboptimal unfolding curves (Figure S1c). The quality of the fit was quantified with standard error of the estimate S :

$$S = \sqrt{\frac{\sum (F_{\text{exp}} - F_{\text{fit}})^2}{N - n}}$$

Where F_{exp} is experimental signal at given datapoint, F_{fit} respective fit value at given datapoint, N is total number of datapoints and n is number of fit parameters.

The standard deviation of all fit parameters was computed as a square root of diagonal values from fit parameter covariance matrix reported by `scipy.curve_fit` function. The curves with standard deviation for T_m above 0.5 K were considered unreliable fits and were discarded.

Finally, to compare the height of unfolding transitions between different samples and readouts we introduced baseline separation factor (BS-factor):

$$BS\text{-factor} = 1 - \frac{6 \cdot S}{|k_U \cdot T_m + b_U - k_N \cdot T_m - b_N|}$$

Where S is standard error of estimate, k_N , b_N – slope and intercept of the pretransition baseline (temperature dependence of native state fluorescence), k_U and b_U – slope and intercept of the posttransition baseline (temperature dependence of unfolded state fluorescence) as obtained from nonlinear curve-fitting and T_m is melting temperature. BS-factor metric relates the uncertainty of the fit, expressed as standard error of the estimate, and the distance between the pre- and posttransition baselines at T_m (Figure S1b) and is conceptually similar to Z-factor commonly applied to assess quality of high-throughput screens.⁷⁵ If the baselines are too close to each other, then their variability bands overlap, and the unfolding transition cannot be well distinguished. BS-factor is dimensionless and can be used to compare the quality of different readouts. In general, curves with BS-factor from 0.5 to 1 are considered excellently separated, while negative BS-factor indicates an unreliable curve.

For the majority of the data the F350/F330 ratio was used (Table S5). In some cases the BS-factor for F350/F330 ratio data was suboptimal, and the fluorescence at individual fluorescence wavelength was used. Shape of unfolding curves is similar between F330 and F350 readouts, so only F330 was used for analysis.

For the purposes of visualization we computed the baseline-corrected experimental curves, that is, the fraction unfolded versus temperature⁷³:

$$K_u(T) = \frac{f_{unf}}{f_{native}} = \frac{k_N \cdot T + b_N - \text{Signal}(T)}{\text{Signal}(T) - k_U \cdot T - b_U}$$
$$f_{unf} = \frac{K_u(T)}{1 + K_u(T)} = \frac{\text{Signal}(T) - k_N \cdot T - b_N}{k_U \cdot T + b_U - k_N \cdot T - b_N}$$

where f_{unf} is fraction of U state, f_{native} is fraction of N state, K_u is equilibrium constant of unfolding and other variables as described above.

The results of individual experiments can be visualized in an interactive HTML report or using a graphical user interface (GUI, Figure 1(c)). MoltenProt can export the results for further analysis and manipulation in a spreadsheet (XLSX) and comma-separated value (CSV) format.

In addition to the data produced by PR.ThermControl, MoltenProt supports import of plain CSV files with one or more samples. CD data were converted to this format and analyzed as described above.

MoltenProt is written in Python v.3.8 using the following modules: pandas,⁷⁶ scipy,⁷⁷ numpy,⁷⁸ matplotlib,⁷⁹ PyQt5 (Riverbank Computing).

4.9 | Code availability

MoltenProt is open source/libre software licensed under GNU General Public License (version 3 or above). The source code is supplied in the Supplementary material and also available via <http://www.marlovitslab.org>.

4.10 | Additional software

Molecular visualization was performed with UCSF Chimera.⁸⁰ Additional plots were made with seaborn.⁸¹

ACKNOWLEDGEMENTS

VK thanks Boehringer Ingelheim Fonds for generous support through the PhD fellowship, Martin Krzywinski for help and advice in figure design and Oliver Gajsek for technical assistance. The authors thank Werner Kühlbrandt, Franck Duong, Yazan Abbas and Walter Keller for providing additional test samples. GM thanks Irene Schaffner and Jakob Wallner for technical support and Antonio Sponga for SEC-MALLS measurements. We acknowledge technical support by the SPC facility at EMBL Hamburg. This project was supported by funds

available to TCM furnished by the Behörde für Wissenschaft, Forschung und Gleichstellung of the city of Hamburg at the Institute for Structural and Systems Biology of the University Medical Center Hamburg-Eppendorf (UKE). This project was supported by EQ-BOKU VIBT GmbH and the BOKU Core Facility for Biomolecular & Cellular Analysis. KDC research was supported by a Marie Curie Initial Training Network: MUZIC (N°238423), Austrian Science Fund (FWF) Projects I525 and I1593, P22276, P19060 and W1221, Federal Ministry of Economy, Family and Youth through the initiative “Laura Bassi Centres of Expertise”, funding the Centre of Optimized Structural Studies (COSS), N°253275, by the Wellcome Trust Collaborative Award (201543/Z/16), COST action BM1405—nonglobular proteins—from sequence to structure, function and application in molecular physiopathology (NGP-NET), WWTF (Vienna Science and Technology Fund) Chemical Biology project LS17-008, by the Universität Wien Research Platform Comammox, Christian Doppler Laboratory for High-Content Structural Biology and Biotechnology, and the Austrian-Slovak Interreg Project B301 StruBioMol. GM was supported by COSS and Comammox. Part of this work was performed at the EM facility at CSSB, supported by UHH and DFG grant numbers (INST 152/772-1|152/774-1|152/775-1|152/776-1|152-777-1 FUGG). SKB and HC were supported by the Intramural Research program of NIH, National Institute of Diabetes & Digestive & Kidney Diseases (NIDDK). JHMC was awarded a FLAD Life Science 2020 by Fundação Luso-Americana para o Desenvolvimento; CMTD was supported by a FCT fellowship (SFRH/BD/123761/2016). Open access funding enabled and organized by Projekt DEAL.

SIGNIFICANCE

In this work we present MoltenProt, an open source software tool for robust analysis of high-throughput protein thermal unfolding data. MoltenProt will be useful in diverse biochemical and structural biology applications, including buffer and construct optimization, functional analysis and small-molecule screening.

AUTHOR CONTRIBUTIONS

Vadim Kotov: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; validation; visualization; writing-original draft; writing-review and editing. **Georg Mlynek:** Formal analysis; investigation; writing-review and editing. **Oliver Vesper:** Investigation; writing-review and editing. **Marina Pletzer:** Investigation; writing-review and editing. **Jiri Wald:** Investigation; writing-review and editing. **Celso M. Teixeira-Duarte:** Investigation; resources; writing-review and editing. **Herve Celia:** Investigation; resources;

writing-review and editing. **Maria Garcia-Alai:** Investigation; methodology; resources; writing-review and editing. **Stephan Nussberger:** Resources; writing-review and editing. **Susan K. Buchanan:** Resources; writing-review and editing. **João H. Morais-Cabral:** Resources; writing-review and editing. **Christian Loew:** Investigation; methodology; resources; writing-review and editing. **Kristina Djinojic-Carugo:** Methodology; resources; writing-review and editing. **Thomas C. Marlovits:** Conceptualization; formal analysis; funding acquisition; methodology; writing-original draft; writing-review and editing.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

ORCID

Thomas C. Marlovits  <https://orcid.org/0000-0001-8106-4038>

REFERENCES

- Deller MC, Kong L, Rupp B (2016) Protein stability: A crystallographer's perspective. *Acta Cryst F* 72:72–95.
- Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem*. 2006;357:289–298.
- Dupeux F, Röwer M, Seroul G, Blot D, Márquez JA. A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Cryst*. 2011;D67:915–919.
- Vedadi M, Arrowsmith CH, Allali-Hassani A, Senisterra G, Wasney GA. Biophysical characterization of recombinant proteins: A key to higher structural genomics success. *J Struct Biol*. 2010;172:107–119.
- Kumar MDS, Bava KA, Gromiha MM, et al. ProTherm and ProNIT: Thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res*. 2006;34:D204–D206.
- Johnson CM. Differential scanning calorimetry as a tool for protein folding and stability. *Arch Biochem Biophys*. 2013;531:100–109.
- Chari A, Haselbach D, Kirves J-M, et al. ProteoPlex: Stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nat Methods*. 2015;12:859–865.
- Boivin S, Kozak S, Meijers R. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr Purif*. 2013;91:192–206.
- Gessmann D, Mager F, Naveed H, et al. Improving the resistance of a eukaryotic β -barrel protein to thermal and chemical perturbation. *J Mol Biol*. 2011;413:150–161.
- Pantoliano MW, Petrella EC, Kwasnoski JD, et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen*. 2001;6:429–440.
- Alexander CG, Wanner R, Johnson CM, et al. Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings. *Biochim Biophys Acta*. 2014;1844:2241–2250.
- Temel DB, Landsman P, Brader ML Chapter fourteen - orthogonal methods for characterizing the unfolding of therapeutic monoclonal antibodies: Differential scanning calorimetry, isothermal chemical denaturation, and intrinsic fluorescence with concomitant static light scattering. In: Feig AL, editor. *Methods in enzymology*. Vol. 567. Calorimetry, USA: Academic Press; 2016. pp. 359–389. Available from: <http://www.sciencedirect.com/science/article/pii/S0076687915004966>
- Abraham Y, Zhang X, Parker CN. Multiparametric analysis of screening data: Growing beyond the single dimension to infinity and beyond. *J Biomol Screen*. 2014;19:628–639.
- Stevens R, Stevens L, Price NC. The stabilities of various thiol compounds used in protein purifications. *Biochem Educ*. 1983;11:70–70.
- Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. 2006;1:2876–2890.
- Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc*. 2006;1:2527–2535.
- Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta Proteins Proteomics*. 2005;1751:119–139.
- Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas' AF, Gilmanshin RI. Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*. 1991;31:119–128.
- Alexandrov AI, Mileni M, Chien EYT, Hanson MA, Stevens RC. Microscale fluorescent thermal stability assay for membrane proteins. *Structure*. 2008;16:351–359.
- Kroeger T, Frieg B, Zhang T, et al. EDTA aggregates induce SYPRO orange-based fluorescence in thermal shift assay. *PLoS One*. 2017;12:e0177024.
- Wang Z, Ye C, Zhang X, Wei Y. Cysteine residue is not essential for CPM protein thermal-stability assay. *Anal Bioanal Chem*. 2015;407:3683–3691.
- Ablinger E, Leitgeb S, Zimmer A. Differential scanning fluorescence approach using a fluorescent molecular rotor to detect thermostability of proteins in surfactant-containing formulations. *Int J Pharm*. 2013;441:255–260.
- Vivian JT, Callis PR. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J*. 2001;80:2093–2109.
- Shen S, Kai B, Ruan J, Torin Huzil J, Carpenter E, Tuszynski JA. Probabilistic analysis of the frequencies of amino acid pairs within characterized protein sequences. *Phys A*. 2006;370:651–662.
- Misetic V, Reiners O, Krauss U, Jaeger K-E (2016) NanoDSF thermal unfolding analysis of proteins without tryptophan residues (Application Note NT-PR-007). Available from: <https://resources.nanotempertech.com/application-notes/application-note-nt-pr-007-unfolding-without-tryptophan>
- Pace CN, Laurents DV. A new method for determining the heat capacity change for protein folding. *Biochemistry*. 1989;28:2520–2525.
- Finkelstein AV, Ptitsyn O. Protein physics: A course of lectures. Lecture 17, Netherlands: Elsevier, 2016.

28. Jacobson DR, Uyetake L, Perkins TT. Membrane-protein unfolding intermediates detected with enhanced precision using a zigzag force ramp. *Biophys J*. 2020;118:667–675.
29. Yu H, Siewny MGW, Edwards DT, Sanders AW, Perkins TT. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science*. 2017;355:945–950.
30. Bakk A. Is it always possible to distinguish two- and three-state systems by evaluating the van't Hoff enthalpy? *Phys Biol*. 2004;1:152–158.
31. Mayne L, Englander SW. Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. *Protein Sci*. 2000;9:1873–1877.
32. Becktel WJ, Schellman JA. Protein stability curves. *Biopolymers*. 1987;26:1859–1877.
33. Bedouelle H. Principles and equations for measuring and interpreting protein stability: From monomer to tetramer. *Biochimie*. 2016;121:29–37.
34. Privalov PL, Khechinashvili NN. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. *J Mol Biol*. 1974;86:665–684.
35. Clarke J, Fersht AR. Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry*. 1993;32:4322–4329.
36. Cooper A. Thermodynamics of protein folding and stability. *Protein Compr Treatise*. 1999;2:217–270.
37. Wagman DD, editor. The NBS tables of chemical thermodynamic properties : Selected values for inorganic and C₁ and C₂ organic substances in SI units. Washington: American Chemical Society and American Institute of Physics for the National Bureau of Standards, 1982.
38. Mazurenko S, Kunka A, Beerens K, Johnson CM, Damborsky J, Prokop Z. Exploration of protein unfolding by modelling calorimetry data from reheating. *Sci Rep*. 2017;7:16321.
39. Sanchez-Ruiz JM. Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophys J*. 1992;61:921–935.
40. Kotov V, Bartels K, Veith K, et al. High-throughput stability screening for detergent-solubilized membrane proteins. *Sci Rep*. 2019;9:10379.
41. Lepock JR, Frey HE, Inniss WE. Thermal analysis of bacteria by differential scanning calorimetry: Relationship of protein denaturation in situ to maximum growth temperature. *Biochim Biophys Acta*. 1990;1055:19–26.
42. Heng J, Zhao Y, Liu M, et al. Substrate-bound structure of the *E. coli* multidrug resistance transporter MdfA. *Cell Res*. 2015;25:1060–1073.
43. Robertson AD, Murphy KP. Protein structure and the energetics of protein stability. *Chem Rev*. 1997;97:1251–1268.
44. Kotov V, Vesper O, Alai MG, Loew C, Marlovits TC. Moltenprot: A high-throughput analysis platform to assess thermodynamic stability of membrane proteins and complexes. *Biophys J*. 2019;116:191a.
45. Pfeil W. Protein Stability and Folding: A Collection of Thermodynamic Data. Berlin Heidelberg: Springer-Verlag; 1998. Available from: <http://www.springer.com/de/book/9783642637162>
46. Gromiha MM, Sarai A. Thermodynamic database for proteins: Features and applications. *Methods Mol Biol*. 2010;609:97–112.
47. Celia H, Noinaj N, Zakharov SD, et al. Structural insight into the role of the ton complex in energy transduction. *Nature*. 2016;538:60–65.
48. Vieira-Pires RS, Szollosi A, Morais-Cabral JH. The structure of the KtrAB potassium transporter. *Nature*. 2013;496:323–328.
49. Veith K, Martinez Molledo M, Almeida Hernandez Y, et al. Lipid-like peptides can stabilize integral membrane proteins for biophysical and structural studies. *Chembiochem*. 2017;18:1735–1742.
50. Yang Z, Wang C, Zhou Q, et al. Membrane protein stability can be compromised by detergent interactions with the extramembranous soluble domains. *Protein Sci*. 2014;23:769–789.
51. Tsaneva IR, Müller B, West SC. RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity in vitro. *Proc Natl Acad Sci U S A*. 1993;90:1315–1319.
52. Adams DE, West SC. Unwinding of closed circular DNA by the *Escherichia coli* RuvA and RuvB recombination/repair proteins. *J Mol Biol*. 1995;247:404–417.
53. Faulkner G, Pallavicini A, Comelli A, et al. FATZ, a filamin-, actinin-, and telethonin-binding protein of the Z-disc of skeletal muscle. *J Biol Chem*. 2000;275:41234–41242.
54. Frey N, Richardson JA, Olson EN. Calsarcins, a novel family of sarcomeric calcineurin-binding proteins. *Proc Natl Acad Sci U S A*. 2000;97:14632–14637.
55. Takada F, Vander Woude DL, Tong HQ, et al. Myozenin: An alpha-actinin- and gamma-filamin-binding protein of skeletal muscle Z lines. *Proc Natl Acad Sci U S A*. 2001;98:1595–1600.
56. Gontier Y, Taivainen A, Fontao L, et al. The Z-disc proteins myotilin and FATZ-1 interact with each other and are connected to the sarcolemma via muscle-specific filamins. *J Cell Sci*. 2005;118:3739–3749.
57. Djinojic-Carugo K, Carugo O. Structural portrait of filamin interaction mechanisms. *Curr Protein Pept Sci*. 2010;11:639–650.
58. Mao Z, Nakamura F. Structure and function of filamin C in the muscle Z-disc. *Int J Mol Sci*. 2020;21:2696.
59. Razinia Z, Mäkelä T, Ylännä J, Calderwood DA. Filamins in mechanosensing and signaling. *Annu Rev Biophys*. 2012;41:227–246.
60. Leber Y, Ruparel AA, Kirfel G, et al. Filamin C is a highly dynamic protein associated with fast repair of myofibrillar microdamage. *Hum Mol Genet*. 2016;25:2776–2788.
61. Bausewein T, Mills DJ, Langer JD, Nitschke B, Nussberger S, Kühlbrandt W. Cryo-EM structure of the TOM core complex from *Neurospora crassa*. *Cell*. 2017;170:693–700.
62. Mazzini V, Craig VSJ. What is the fundamental ion-specific series for anions and cations? Ion specificity in standard partial molar volumes of electrolytes and electrostriction in water and non-aqueous solvents. *Chem Sci*. 2017;8:7052–7065.
63. Okur HI, Hladílková J, Rembert KB, et al. Beyond the Hofmeister series: Ion-specific effects on proteins and their biological functions. *J Phys Chem B*. 2017;121:1997–2014.
64. Majd H, King MS, Palmer SM, et al. Screening of candidate substrates and coupling ions of transporters by thermostability shift assays. *Elife*. 2018;7:e38821.
65. Martinez Molledo M, Quistgaard EM, Löw C. Tripeptide binding in a proton-dependent oligopeptide transporter. *FEBS Lett*. 2018;592:3239–3247.

66. Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim Biophys Acta Biomembr*. 2004;1666:105–117.
67. Chae PS, Rasmussen SGF, Rana RR, et al. Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat Methods*. 2010;7:1003–1008.
68. Chae PS, Rasmussen SGF, Rana RR, et al. A new class of amphiphiles bearing rigid hydrophobic groups for solubilization and stabilization of membrane proteins. *Chem A Eur J*. 2012;18:9485–9490.
69. Timmins PA, Hauk J, Wacker T, Welte W. The influence of heptane-1,2,3-triol on the size and shape of LDAO micelles implications for the crystallisation of membrane proteins. *FEBS Lett*. 1991;280:115–120.
70. Herberhold H, Winter R. Temperature- and pressure-induced unfolding and refolding of ubiquitin: A static and kinetic Fourier transform infrared spectroscopy study. *Biochemistry*. 2002;41:2396–2401.
71. Mlynek G, Lehner A, Neuhold J, et al. The Center for Optimized Structural Studies (COSS) platform for automation in cloning, expression, and purification of single proteins and protein-protein complexes. *Amino Acids*. 2014;46:1565–1582.
72. Gasteiger E, Hoogland C, Gattiker A, et al. Analysis tools on the ExPASy server. In: Walker JM, editor. *The proteomics protocols handbook*. Totowa, NJ: Humana Press, 2005; p. 571–607. Available from: <https://doi.org/10.1385/1-59259-890-0:571>.
73. Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry*. 1988;27:8063–8068.
74. Wang T, Lucey JA. Use of multi-angle laser light scattering and size-exclusion chromatography to characterize the molecular weight and types of aggregates present in commercial whey protein products. *J Dairy Sci*. 2003;86:3090–3101.
75. Zhang J-H, Chung TDY, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen*. 1999;4:67–73.
76. McKinney W. Data structures for statistical computing in python. *Proceedings of the 9th python in science conference*, 2010; p. 51–56. <http://conference.scipy.org/proceedings/scipy2010/mckinney.html>
77. Jones E, Oliphant T, Peterson P, others SciPy: Open source scientific tools for Python. 2001. Available from: <http://www.scipy.org/>
78. Oliphant TE. *Guide to NumPy*. 2nd ed. USA: CreateSpace Independent Publishing Platform, 2015.
79. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. 2007;9:90–95.
80. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–1612.
81. Waskom, Botvinnik, Joel Ostblom, Maoz Gelbart, Saulius Lukauskas, Paul Hobson, David C Gemperline, Tom Augspurger, Yaroslav Halchenko, John B. Cole, et al. Mwaskom/seaborn: v0.10.1 (April 2020). Zenodo; 2020. Available from: <https://zenodo.org/record/3767070#.XtuDBZ9fiu1>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Kotov V, Mlynek G, Vesper O, et al. In-depth interrogation of protein thermal unfolding data with MoltenProt. *Protein Science*. 2021;30:201–217. <https://doi.org/10.1002/pro.3986>