



Novel Non-linear Curve Fitting to Resolve Protein Unfolding Transitions in Intrinsic Fluorescence Differential Scanning Fluorimetry

Augustijn, Dillen; Mahapatra, Sujata; Streicher, Werner; Svilenov, Hristo; Kulakova, Alina; Pohl, Christin; Rinnan, Åsmund

Published in:
European Journal of Pharmaceutics and Biopharmaceutics

Link to article, DOI:
[10.1016/j.ejpb.2019.06.001](https://doi.org/10.1016/j.ejpb.2019.06.001)

Publication date:
2019

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Augustijn, D., Mahapatra, S., Streicher, W., Svilenov, H., Kulakova, A., Pohl, C., & Rinnan, Å. (2019). Novel Non-linear Curve Fitting to Resolve Protein Unfolding Transitions in Intrinsic Fluorescence Differential Scanning Fluorimetry. *European Journal of Pharmaceutics and Biopharmaceutics*, 142, 506-517.
<https://doi.org/10.1016/j.ejpb.2019.06.001>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Research paper

Novel Non-linear Curve Fitting to Resolve Protein Unfolding Transitions in Intrinsic Fluorescence Differential Scanning Fluorimetry

Dillen Augustijn, Sujata Mahapatra, Werner Streicher, Hristo Svilenov, Alina Kulakova, Christin Pohl, Åsmund Rinnan

PII: S0939-6411(19)30309-1
DOI: <https://doi.org/10.1016/j.ejpb.2019.06.001>
Reference: EJPB 13069

To appear in: *European Journal of Pharmaceutics and Biopharmaceutics*

Received Date: 22 March 2019
Revised Date: 31 May 2019
Accepted Date: 5 June 2019

Please cite this article as: D. Augustijn, S. Mahapatra, W. Streicher, H. Svilenov, A. Kulakova, C. Pohl, a. Rinnan, Novel Non-linear Curve Fitting to Resolve Protein Unfolding Transitions in Intrinsic Fluorescence Differential Scanning Fluorimetry, *European Journal of Pharmaceutics and Biopharmaceutics* (2019), doi: <https://doi.org/10.1016/j.ejpb.2019.06.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Novel Non-linear Curve Fitting to Resolve Protein Unfolding Transitions in Intrinsic Fluorescence Differential Scanning Fluorimetry.

Dillen Augustijn^{1*}, Sujata Mahapatra^{2,3}, Werner Streicher², Hristo Svilenov⁴, Alina Kulakova^{2,3}, Christin Pohl^{2,3}, Åsmund Rinnan¹

Keywords: Differential scanning fluorimetry, Intrinsic fluorescence DSF, Multivariate data analysis, Non-linear curve fitting, thermal stability screening, protein screening, biotherapeutics.

Abstract

In biotherapeutic protein research, an estimation of the studied protein's thermal stability is one of the important steps that determine developability as a function of solvent conditions. Differential Scanning Fluorimetry (DSF) can be applied to measure thermal stability. Label-free DSF measures amino acid fluorescence as a function of temperature, where conformational changes induce observable peak deformation, yielding apparent melting temperatures. The estimation of the stability parameters can be hindered in the case of multidomain, multimeric or aggregating proteins when multiple transitions partially coincide. These overlapping protein unfolding transitions are hard to evaluate by the conventional methodology, as peak maxima are shifted by convolution. We show how non-linear curve fitting of intrinsic fluorescence DSF can deconvolute highly overlapping transitions in formulation screening in a semi-automated process. The proposed methodology relies on synchronous, constrained fits of the fluorescence intensity, ratio and their derivatives, by combining linear baselines with generalized logistic transition functions. The proposed algorithm is applied to data from three proteins; a single transition, a double separated transition and a double overlapping transition. Extracted thermal stability parameters; apparent melting temperatures $T_{m,1}$, $T_{m,2}$ and melting onset temperature T_{onset} are obtained and compared with reference software analysis. The fits show $R^2 = 0.94$ for single and $R^2 = 0.88$ for separated transitions. Obtaining values and trends for T_{onset} in a well-described and automated way, will aid protein scientist to better evaluate the thermal stability of proteins.

¹ Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.

² Novozymes A/S, Krogshøjvej 36, 2880 Bagsværd, Denmark

³ Department of Chemistry, Technical University of Denmark, 2800 Lyngby, Denmark.

⁴ Department of Pharmacy, Pharmaceutical Technology and Biopharmaceutics, Ludwig-Maximilians-Universität München, Butenandtstrasse 5-13, Munich D-81377, Germany

*Corresponding author: dillen@food.ku.dk, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.

1. Introduction

High-throughput thermal stability screening is an important, but challenging task that examines protein stability in solution using various techniques – aiding further development of biotherapeutics.[1, 2] For instance, in formulations screening, a protein is formulated under various conditions (i.e. pH salts, excipients and/or surfactants) in order to find the combination that yields high protein stability. The stability of these formulations is essential to retain the efficacy and safety of the drug after its production, fill-finish, storage, shipping and handling. One of the important characteristics for the stability of biotherapeutic proteins is their conformational stability in solution.[3] The propensity of a protein to partially unfold, even at relatively low temperatures or relatively low denaturant concentrations, indicates unfavorable biophysical properties, as protein unfolding often leads to aggregation and loss of functionality.[4] Moreover, the formation of large aggregates and subvisible particles may lead to immunogenicity of the drug substance.[5, 6] Screening for the optimal formulation involves assessment of several protein stability indicators, such as the denaturation or melting temperature, denoted as T_d or T_m , and the Gibbs free energy of unfolding, ΔG .

Differential scanning fluorimetry (DSF) can be employed to assess protein thermal stability. DSF was developed as an extrinsic dye-based method, which probes solvent-exposed hydrophobic patches during thermal unfolding. The extrinsic dye binds to these patches and can subsequently be characterized by the change in its fluorescence properties. [7, 8] Intrinsic fluorescence DSF, sometimes referred to by device manufacturers as nanoDSF™ or nDSF, is a relatively novel technique that utilizes the native protein fluorescence peak to measure protein unfolding. Thermal unfolding induces structural changes, where the solvent exposure of amino acids that are buried in the hydrophobic protein core causes detectable differences in the intrinsic protein fluorescence emission spectra. IF DSF monitors the emission intensity as a function of temperature at a set excitation wavelength. The observed native protein fluorescent spectrum consists of contributions from phenylalanine, tyrosine and tryptophan.[9] Folded proteins generally show only fluorescence in the tryptophan [290 - 450 nm] or tyrosine [280 - 340 nm] peak range. The excitation spectrum of tyrosine is similar to that of tryptophan. However, tyrosine residues have lower absorption coefficients, a lower quantum yield, and the emission at a lower wavelength causes Förster resonance energy transfer (FRET) to the tryptophan residues.[10-12] It therefore is only feasible to study tyrosine fluorescence peak changes when very few or no tryptophan residues are present. Generally, to reduce the influence of tyrosine, an excitation wavelength between 285 nm and 300 nm can be used.[13] This focusses on the well described native tryptophan fluorescence peak, and peak changes are associated with unfolding of the tryptophan-containing domains. In a typical application of nanoDSF™, the changes of the fluorescence peak are monitored by simultaneously measuring at two different emission wavelengths, namely 330 and 350 nm.[14]

Some complex proteins, such as monoclonal antibodies, have multiple domains and their unfolding can be characterized by intrinsic fluorescence DSF.[13] The advantage of this technique lies in the small quantities used and ease of sample preparation, as no dyes are added. Additionally, the potential effects of these dyes on the protein stability are nullified. The downsides of using native fluorescence are a decrease in fluorescence signal with increasing temperature, as well as an influence of solvent conditions on protein peak shape, intensity and intensity decline.[15, 16]

The analysis of thermal protein unfolding using IF DSF data is based on the temperature dependent shift in fluorescence, as characterized by the ratio between the intensities at the two emission wavelengths. The protein thermal stability is further characterized as *apparent melting temperature* T_m , at the maximum in the first derivative of the ratio.[13] However, difficulties with data analysis occur when protein unfolding steps are overlapping, and slopes in the fluorescent ratio merge. Additionally, evaluating transitions from the fluorescence intensity ratio alone can result in biases in some cases.[17] It should be noted that apparent T_m values were correlated to the true T_m values obtained by differential scanning calorimetry (DSC) [8, 18, 19]. However, in the IF DSF experiments presented here, the reversibility of the transition is not studied, therefore the term T_m is used to indicate *apparent melting temperature*. Additionally, the onset temperature of the investigated thermal protein unfolding, here denoted as T_{onset} , is of particular interest.

This study proposes a curve fitting method based on protein states and the fluorescence intensities observed to deconvolute unfolding transitions. The model is validated by comparing the non-linear fitting approach to conventional analysis reference values of both a single, and a double (baseline separated) unfolding transition. We then apply our methodology to an IgG1 type mAb, PPI-03, that shows increasing transition overlap with increasing formulation pH, as illustrated in Figure 1.

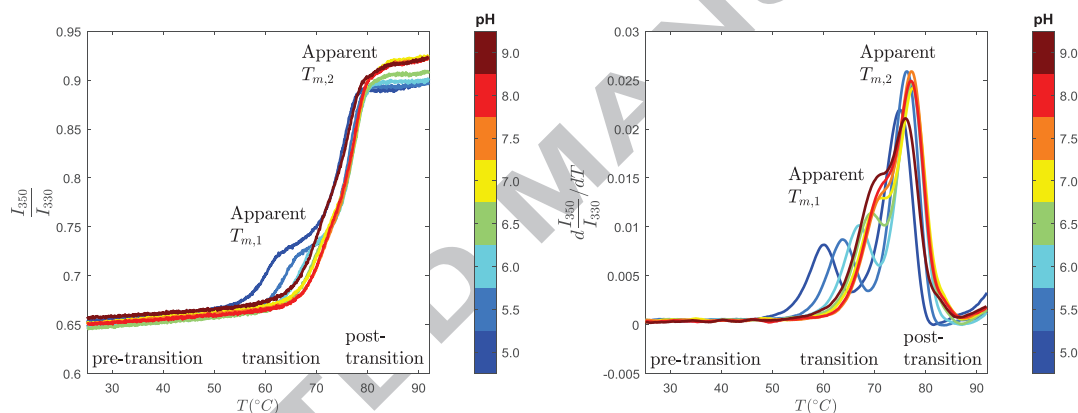


Figure 1. PPI-03 fluorescence ratio (left) and its derivative (right) for formulations at 70 mM NaCl at various pHs. The first transition shifts to a higher temperature at higher pH, indicating higher thermal stability. At high pH the two transitions overlap more strongly, making an estimation of both apparent T_m values more difficult. Indicated are the pre-transition region, the transition region containing two ratio shifts, and the post-transition region.

The non-linear curve fitting is based on simultaneous fitting of fluorescence intensities, the ratio between them and the derivative of these curves. We employ a theoretical transition model with a three-state fit and extract valued formulation characteristics (T_{onset} and T_m) for both transitions, with prospected extensions to other fluorescence-based methods.

1.1 Terms

Table 1: Throughout the paper various terms are used, for consistency the following are defined:

T	Temperature.
$j \dots J$	Protein state as folded ($j = 1$), intermediate ($j = 2$) and unfolded state ($j = 3$).
$s \dots J - 1$	Transition s , for $J - 1$ transitions.
λ	Observed emission wavelength - 330 nm or 350 nm.
$I_\lambda, I_{\lambda,j}$	Total observed fluorescence intensity I_λ , and the pure state contribution $I_{\lambda,j}$ of state j .

ϕ_j	Fraction of proteins in state j , represented for all states J , summing to 1.
$\vec{\kappa}$	Penalty vector κ , used to regulate the importance of the obtained data fits in the optimization function.
\vec{w}	Weight vector \vec{w} is used to downweigh the edges of the data, i.e. at the lowest and highest temperature.
p	Fitting parameter p , encompasses the baseline parameters $b_{\lambda,j,0...N}$ for an N^{th} order polynomial fit at emission wavelengths λ , and all transition parameters A_s , B_s and v_s .
f_λ	Fit of fluorescence intensity I_λ at emission wavelength λ .
$T_{m,s}$	Apparent melting temperature of transition s . Other conventional denotations are $T_{d,s}$ or $T_{\frac{1}{2},s}$.
SSE	Weighted sum of squared errors of prediction.
R^2	Squared correlation coefficient
RMSE	Root-mean-squared error

2. Materials and methods

2.1. Materials and experimental screening

The protein formulation screening data is a full factorial design, with 24 combinations of pH and ionic strength, as part of a large screening effort. PPI-03 and PPI-13 are monoclonal antibodies and PPI-30 is interferon alpha-2a (IFN- α 2a).

Interferon alpha-2a is a single domain protein containing 165 amino acids and two disulfide bridges.[20, 21]

Both the model monoclonal antibodies (mAb), PPI-03 and PPI-13, belong to the human IgG1 subclass. The mAbs are multidomain Y-shaped proteins composed of four polypeptide chains, which are grouped into different domains, two identical F_{ab} domains and one F_c domain. Monoclonal antibody derived products are an important factor in the treatment of immunological disorders, and have a major contribution in biotherapeutics, many of which are in the process of development. Thermal unfolding behavior interpretation of mAbs becomes complex due to its multidomain nature - although it has been observed that mAbs unfolds in multiple steps. Lowest $T_{m,1}$ peak usually corresponds to C_H2 domain followed by the F_{ab} , and sequentially unfolding of the C_H3 domain at higher temperature $T_{m,2}$. [22]

2.1.1. Bulk solutions

The PPI-03 bulk solution contained 46.4 mg/ml protein in 25 mM histidine and 205 mM (7%) sucrose at pH 6.0 and had a molecular weight of 145 kDa. The PPI-13 bulk solution contained 48.9 mg/ml of protein in 20 mM succinate, 95 mM arginine, 180 mM mannitol and 20 mM NaCl at pH 6.0 and has a molecular weight of 149 kDa. The PPI-30 bulk solution contained 1.35 mg/ml protein in 25 mM ammonium acetate and 120 mM NaCl at pH 5.0, and has a molecular weight of \sim 20 kDa. All bulk solutions were free of surfactants.

2.1.2. Formulation procedures

Each protein was formulated in 24 conditions as a full factorial design across eight different pH values (5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 9.0) and three different ionic strengths (0 mM, 70 mM and 140 mM NaCl concentration). In other words, a design with eight levels in one mode, and three in the

other. The reason for choosing such a comprehensive design was to ensure that good formulation conditions were covered for all proteins. Chemicals used for formulations are L-histidine (J.T.Baker®), L-histidine monohydrochloride (J.T.Baker®), Trizma® base (Sigma-Aldrich®), Trizma® hydrochloride (Sigma-Aldrich®) and sodium chloride (Sigma-Aldrich®). The differences in the formulations are because the measurements were performed at two different laboratories, and the samples were formulated using two procedures.

For PPI-03, and first the PPI-13 & PPI-30 lab-replicates; Slide-A-Lyzer™ dialysis cassettes (Thermo Fisher Scientific®) were used for dialyses of each protein in 10 mM histidine buffer at pH 5.5, 10 mM histidine buffer at pH 7.0 and 10 mM tris buffer at pH 8.5. A buffer exchange was performed after two and four hours, ensuring a dilution of at least 200 times the sample volume. Each dialysis was continued overnight at cold temperature (4 °C). The protein concentration after dialysis was measured using NanoDrop™ 8000 Spectrophotometer using the respective extinction coefficients at 280 nm and then a protein stock solution with a concentration of 20 mg/ml was made by dilution with the dialysis buffers. The final formulations screened were obtained by a 1:20 dilution of the protein stock solution into the respective buffers for the proteins to be tested at a uniform concentration of 1 mg/ml across all samples. The protein stock solution in 10 mM histidine at pH 5.5 served as a stock solution for all formulations in pH 5.0, 5.5 and 6.0, the protein stock solution in 10 mM histidine at pH 7.0 served as a stock solution for all formulations in pH 6.5, 7.0 and 7.5 and the protein stock solution in 10 mM tris at pH 8.5 served as a stock solution for all formulations in pH 8.0 and 9.0. The pH of the final samples was checked and adjusted where necessary. PPI-30 was up concentrated to 30 mg/ml before dialysis and then dialyzed as described.

For the second cross-lab replicate, PPI-03 and PPI-13 were dialyzed in 10 mM histidine buffer at pH 6.0 using Slide-A-Lyzer™ dialysis cassettes. Buffer exchange was performed after two and four hours, ensuring a dilution of at least 200 times the sample volume. The dialysis was continued overnight at cold temperature (4 °C). The protein concentration after dialysis was measured using NanoDrop™ 2000 Spectrophotometer using the respective extinction coefficients at 280 nm and a protein stock solution with a concentration of 20 mg/ml was prepared by dilution with the dialysis buffer. The final screening conditions were obtained by a 1:20 dilution into the respective formulation buffers to achieve a final concentration of 1 mg/ml protein. The pH of the final samples was checked and adjusted where necessary.

2.2. Intrinsic fluorescence DSF measurements

Measurements of the samples were carried out using two different but identically equipped Prometheus NT.48 (NanoTemper® Technologies GmbH) instruments for cross-lab accurate measurements, using a 1 °C/min temperature ramp between 25 and 95 °C. The observed two emission wavelengths, 330 and 350 nm, are monitored at an excitation wavelength of 280 nm. The fluorescence is recorded at an average of one measurement per 0.044 °C, or 23 measurements per 1 °C, resulting in up to 1600 data points per emission wavelength.

Each formulation was measured as a technical triplicate. Additionally, PPI-03 and PPI-13 have cross-laboratory duplicates to evaluate the formulation protocol and cross-laboratory reproducibility, resulting in a total of 144 measurements for these proteins. For PPI-03 one measurement (pH 7.5,

140 mM NaCl) is not obtained, resulting in 5 replicates. PPI-30 precipitated at pH range 5.5 to 6.5 and therefore no data is obtained within this range, and 45 measurements are obtained for PPI-30.

The data is preprocessed by calculation of the fluorescence intensity ratio, denoted $\frac{I_{350}}{I_{330}}$, and the 1st derivate of the ratio and fluorescent signals using the manufacturer's software (PR.Control v1.12.2, NanoTemper® Technologies GmbH). For each measurement six curves are obtained, as a function of temperature: the fluorescence intensity at 330 nm, the smoothed 1st derivative at 330 nm, the fluorescence intensity at 350 nm, the smoothed 1st derivative at 350 nm, the ratio and the smoothed 1st derivative of the ratio.

2.3. Reference data analysis method

The reference analysis of the data of PPI-30, PPI-03 and PPI-13 was performed using the manufacturer's software (PR.Control v1.12.2, NanoTemper® Technologies GmbH), and reference values for T_{onset} , $T_{m,1}$, and $T_{m,2}$ were determined for pH 5, 5.5, 6.0, 6.5, 7.0, 7.5, 8 and 9 at salt concentrations of 0 mM, 70 mM and 140 mM NaCl, according to the methodology described by Martin et al. (2014)[23]. A value for $T_{m,2}$ was not obtained for PPI-30, which has one discernable transition using the reference method.

3. Curve fitting

A three-state model, comprising of the native, intermediate and unfolded state, is applied and each state is expressed as a relative fraction. This model has been applied by Eftink et al.[11] and by Lazer et al.[24] to model similar data in isothermal chemical denaturation (ICD) experiments as well as by Harder et al.[25] for circular dichroism (CD) and ICD experiments. Based on the cited works, the rate constants for the transitions, k_1 and k_2 , and the relative state fractions, $\phi_{1...3}$, are defined as:



The observed emission signal intensity I_λ at wavelength λ , is described as the sum of the contributions from each state's pure signal $I_{\lambda,j}$ and its fraction.

$$I_\lambda = I_{\lambda,1} \phi_1 + I_{\lambda,2} \phi_2 + I_{\lambda,3} \phi_3 \quad (2)$$

The observed emission signal intensity can be approached by modelling the signal contributions as a function of temperature and concentration.

3.2. Baselines

A baseline is approximated for each combination of protein state and monitored wavelength. The absorption spectra of single-state tryptophan residues do not change with temperature; however, the quantum yield declines as the non-radiative return of the excited state increases. Signal intensity I is then proportional to $1/T$ (K^{-1}) according to:[16]

$$\ln \left(\frac{d \left(-\frac{1}{I} \right)}{d \left(\frac{1}{T} \right)} \right) \propto \frac{1}{T} \quad (3)$$

Which links the slope at the two emission wavelengths as a proportional relative decrease. The signal $I_{\lambda,j}$, at emission wavelength λ and state j , is modelled as a 1st order polynomial on the relatively short temperature intervals.

$$I_{\lambda,j}(T) = b_{\lambda,j,0} + b_{\lambda,j,1} T = \sum_{n=0}^1 b_{\lambda,j,n} T^n \quad (4)$$

The relative decline of the slope, or loss of signal, at any given temperature where the protein is in a single state should be equal at both emission wavelengths λ . However, at the baseline, multiple shifted fluorescent amino acid residues or experimental interference in the emission spectrum can cause small deviations, and a 1-to-1 relationship on the temperature dependent signal loss is not enforced.

3.3. Transitions

The transitions are modeled using a generalized logistic function, $S(T)$ eq.5, over interval $[0,1]$. Here the conversion S is expressed as a function of temperature T , independent of the observed emission wavelength. The curve is characterized by the transition midpoint A_s , the logistic growth rate B_s and the asymmetry factor v_s . Resulting in three parameters for each transition. When $v_s = 1$ a regular logistic function is obtained.

$$S(T) = (1 + e^{B_s(A_s - T)})^{-\frac{1}{v_s}} \quad (5)$$

Inclusion of an asymmetry parameter has the advantage that the description of the transition is flexible towards temperature induced factors that affect the shape of the transition. Introducing flexibility has drawbacks, both in optimization of the fit and uncertainty in the resulting curves. This problem is addressed using hard and soft constraints on v_s , as described in 4.2-4.3.

The fraction of the intermediate state can be expressed as $\phi_2 = S_1(T) \cdot (1 - S_2(T))$, i.e. as the amount of native state that has unfolded to the intermediate state minus the amount of intermediate state that has further unfolded. It follows that $\phi_1 = 1 - S_1(T)$ and $\phi_3 = S_1(T) \cdot S_2(T)$, such as that all the fractions $0 > \phi_{j...j} < 1$, and $\sum_j^J \phi_j = 1$ at any T , resulting in:

$$[1 - S_1] + [S_1(1 - S_2)] + [S_1S_2] = 1 \quad (6)$$

The value for $T_{m,1}$ can be obtained by equating the transition fraction S_1 to $\frac{1}{2}$, corresponding to the mid-point of the transition. This method can be generalized for any transition fraction S_s , by rearranging eq.5 into:

$$T_s(S_s) = A_s - \frac{\ln(-1 + S_s^{-v_s})}{B_s} \quad (7)$$

Eq.7 also allows for the calculation of the transition onset, $S_1 = 4\%$, or as $T_{onset,1}(0.04)$ corresponding to 4% conversion of the first transitions, analogous to the traditional methodology.

Combining the baselines and transitions allows us to formulate a fit f_λ of the observed intensity I_λ at emission wavelength λ over J states.

$$f_\lambda(p, T) = \sum_{j=1}^J \left(\phi_j \cdot \sum_n b_{\lambda,j,n} T^n \right) \quad (8)$$

For brevity, the parameter vector \vec{p} is defined to contain relevant fitting parameters, $[A, B, v, b]$. When fitting two wavelengths in a three-state model, p includes 18 parameters. The fit is guided by constraining and penalizing the parameter space.

The ratio of the fits is obtained from the combination of the fits at different wavelengths, analogue to the way it is calculated from the observed data.

$$\frac{f_{350}(p, T)}{f_{330}(p, T)} = \frac{f_{350}(p, T)}{f_{330}(p, T)} \quad (9)$$

4. Optimization algorithm

The curve fitting algorithm aims to find the best values for the fitting parameters p , such that the fitted curves most closely resemble the data. This is done by minimization of the objective function (4.1) as a function of p , where the fit is compared to the observed signals and ratio and their derivatives. Penalization of these various forms is achieved through a penalty vector κ , which regulates the emphasis during fitting. Further optimization is guided by hard (4.2) and soft constraints (4.3). These constraints aim to limit the search space or penalize off-values, with the goal of (a) finding a minimum that is physically feasible, (b) speeding up the convergence whilst (c) avoiding local minima and edge regions in the temperature domain. Practical application of the model is considered in terms of the computational implementation (4.4), model validation (4.5) and optimization of applied constraints (4.6).

4.1. Objective function

The objective function is defined as the sum of the squared distances between the fits and the data. We can calculate the *weighted sum of squared errors of prediction (SSE)* for each of the six curves obtained in IF DSF, and independently penalize the SSE in the four data forms: the fluorescence intensity spectra, the smoothed 1st derivative of these spectra, the ratio $\frac{350}{330}$ and the smoothed 1st derivative of the ratio. For each of these forms, we use one of the first four terms of the *penalty term* $\kappa_{\{1...4\}}$. A *weight vector*, \vec{w} , is applied to downscale the importance of the edge regions in the fit. In the optimization of the fits, we define a general minimization problem which includes the soft constraints.

Shape of the weighted sum of squared error

$$SSE_n = \frac{\kappa_n}{\text{var}(I_\lambda)} \sum w |I_\lambda(T) - f_\lambda(p, T)|^2 \quad (10)$$

Total optimization

$$\arg \min_p \sum SSE_n + \text{additional penalties} \quad (11)$$

The penalties κ and weights \vec{w} are to be optimized per screening. Typically, with strongly overlapping transitions, a higher penalty value κ is used for the fluorescence ratio and its derivative. Increasing κ leads to a larger SSE for that specific data form, thus putting more emphasis on a correct fit for that form. Note that it is practical to normalize the penalties by the variance of the curves, such as that they fall within the same effect-scale.

The *weight vector* \vec{w} is chosen before fitting, rescaled to an average weight of 1, and used to emphasize the data in a specific temperature domain. Suggested \vec{w} are (1) trapezoid from the edges, (2) as a trapezoid connecting the maxima of the first derivative of the ratio, or (3) flat, uniformly over the temperature domain. The edge regions contain the less important features of the data whilst also causing potential local minima. These local minima are found when fitted transitions that overlap with the edges are only partially evaluated. Downweighing the edge regions will effectively emphasize the transition-range or region of interest; this effect is stronger for the *maxima*-based trapezoid. The three used shapes of \vec{w} are plotted in Figure 2.

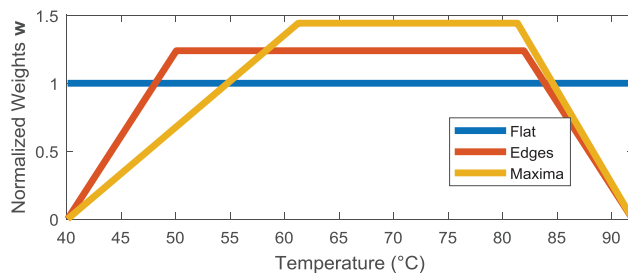


Figure 2. Example of normalized Weights \vec{w} over a chosen temperature interval 40 to 92 °C: Flat, uniform weights (blue), linearly decreasing towards the edges over 10 °C (red), or linearly decreasing towards the edges from the ratio's first derivative maxima (yellow). The weight for temperatures below 40 °C is set to zero for the Edges and Maxima trapezoids.

4.2. Hard constraints

Hard constraints allow for the inclusion of chemical and physical knowledge of our sample and instrumentation into the model. This information is used to define hard borders for the fitting parameters, in order to steer the optimization algorithm towards a meaningful local minimum and away from the edges of the data.

Various hard constraints are imposed on the system. For the baselines a positive intercept $b_{j,1} > 0$ and a negative slope $b_{j,2} < 0$ are imposed, to ensure the pure state signals are positive and decreasing. The transition midpoint A_s should fall well within the measured interval, and such that the two transitions are separated. The A_s terms are limited to $40^\circ\text{C} < A_1 < A_1 + 3^\circ\text{C} < A_2 < 85^\circ\text{C}$. The 3°C represents a flexible minimum distance between transition midpoints in order to avoid local fitting minima where only one transition is described using two transition fits. The transition width B_s , a unitless factor, is fixed according to 10 % - 90 % transition around the midpoint with a minimum width of $\sim 6^\circ\text{C}$ and a maximum width of $\sim 30^\circ\text{C}$, resulting in $0.15 < B_s < 0.6$. The asymmetry factor v_s is constrained around the symmetry, $0.9 < v_s < 1.1$, and further penalized (see section 4.3).

4.3. Penalties κ

Next to the hard constraints, which model the physical boundaries of the fit, penalties are used to guide the fit towards a suitable minimum. The first four terms of the penalties $\kappa_{\{1...4\}}$ are used for the mixing of the four data forms, and two soft constraints are added as penalties, denoted as κ_5 and κ_6 .

The asymmetry of the fit v_s is constrained to allow for small asymmetries around the value of 1, using a quadratic constraint in the form:

$$\text{Asymmetry penalty} = \kappa_5 \sum_{s=1}^{J-1} (v_s - 1)^2 \quad (12)$$

We limit the scope of asymmetry, especially with overlapping functions as the asymmetry and slope of the baselines is partially confounded, mainly as the baseline of the intermediate state becomes less defined when the transitions overlap.

Furthermore, a quadratic soft constrain is applied to the relative loss of signal between the wavelengths at each of the folding states. The protein peak of a single state should only decrease during the temperature ramp, but deformation is a sign of state change. A small baseline would offset a direct relationship in the relative slope at the two emission wavelengths. Constraining the slopes avoids local minima and expedites convergence, whilst retaining flexibility in the accommodation of sloped baselines.

$$\text{Slope penalty} = \kappa_6 \sum_{j=1}^J \left(\frac{b_{\lambda=350,j,1}}{b_{\lambda=350,j,0}} - \frac{b_{\lambda=330,j,1}}{b_{\lambda=330,j,0}} \right)^2 \quad (13)$$

The effect of the slope penalty is to strongly promote the tendency for flat baselines in the derivative of the ratio.

4.4. Algorithm implementation, parameters and initial p_0

For a single transition, 11 parameters are fitted, 3 terms are used to describe the transition and 8 baseline terms remain. For the three-state model seven parameters are added; the 3 parameters of the second transition as well as two parameters for each additional baseline. All 18 fitting parameters contained in p are fitted simultaneously.

Initial parameters p_0 are estimated within the hard constraint borders, these initial values are found to be robust to the quality of the fit. The baseline terms are set by a simple polynomial fit of the baselines at the edge regions and an average of these are used for the intermediate state. Additionally, the transitions can be established uniformly for all formulations of a protein, either by visual inspections of the first derivative of the ratios, or by setting it equidistant in the hard constraint borders of the A_s terms and initiate the remaining transition parameters using $B_s = 0.4$ and $v_s = 1$ for all transitions. An analogous approach is applied in the case of a single transition.

The MATLAB® Optimization Toolbox® is employed to find the minimum of a constrained nonlinear multivariable function using Sequential Quadratic Programming (SQP). In SQP an approximation is made of the Hessian of the cost-function using a quasi-Newton updating method, which defines a subproblem to be solved by quadratic programming. Iteratively a new Hessian is calculated and solved until convergence is obtained.[26] This results in fast, global optimization that benefits from a constrained search space to aid the optimization of the subproblems, and with the added advantage that the Jacobian and Hessian matrices do not have to be provided. Input parameters p are normalized to compensate for the different scales, e.g. the expected baseline parameters are $\sim 10^4$ times larger than B_s or v_s . Using SQP, a three-state fit is typically obtained in less than 5 seconds (CPU 4 cores, 2.90 GHz).

4.5. Validation

The validity of each fit can be assessed by visual inspection of the fits in the different forms of the curve, or by analyzing the objective error minimized in the fitting. When visually inspecting the fits, good characteristics are the consistency of parameter values p within the replicates, and preservation of trends in fitting parameters in the formulation series. A robust analysis of the objective function can be obtained from the fitting error or SSE. A standard deviation σ_{SSE} is estimated from the SSE obtained. This standard deviation is used to quantify and reject deviant fits, where an adverse local minimum is found. The optimization procedure and rejection of the incorrect fits occurred independently from the reference data, such as to mimic applicability conditions and avoid bias.

After the independent optimization, the fitted models are compared to the reference values in order to evaluate the applied methodology. We calculate the square of the Pearson correlation coefficient (R^2) and the Root Mean Square Error (RMSE). Values for R^2 will give an idea about direct comparability between the predicted and reference values, with R^2 values $[0, 1]$. We calculate the offset as the difference in the means of predicted and reference T_m and T_{offset} , to show possible biases between methodologies. This offset value ($\bar{T}_{ref} - \bar{T}_{pred}$) indicates by what measure the predicted values have to be shifted by their average. The RMSE measures the overall average

prediction error for the specific data included in the model. Visual evaluation of the results gives insight into the conservation of the trend, or correlation.

4.6. Optimization of constraints and penalties

Optimization of the penalty vector κ relies on the evaluation of the resulting fits and is adjusted in a step-wise manner. Initial penalties $\kappa_{1...4}$ are set as equal, and after inspection of the fits of a subset of samples, the penalties for data forms that show large misalignments by visual inspection should be increased. Furthermore, a high penalty for the soft-constraints $\kappa_{5...6}$, related to v_s and the relative slope of the baselines, is initially applied. These penalties are lowered to improve the fit until overall SSE no longer decreases. However, with strongly overlapping fits the asymmetry should remain constrained. The initial weight vector, \vec{w} , can be set to an edge distanced trapezoid or as a flat value of 1. Consider using larger edge regions or maxima-based weights with small baselines or artefacts in the edge regions. A 10 °C edge interval is proposed, which can be decreased if transitions occur near the temperature edges. When the baseline at either end of the temperature range shows non-linear behavior, the data's temperature range is reduced to achieve a correct linear baseline estimation. Also, limiting the temperature range reduces the amount of data put into the fitting algorithm, which greatly reduces computation time. However, when only computation time is to be considered, removing alternating data points in the baseline region and doubling the weights of the remaining points will retain baseline estimation whilst reducing the amount of evaluated data. Adverse local minima are addressed by adjusting hard constraints, typically the minimum distance between A_1 and A_2 , possibly in combination with limitations on their ranges.

5. Results

The results divided into four sections, one for each of the three proteins analyzed, according to increasing complexity of the problem (single transition, double transition and overlapped transitions), and a final section summarizing the results.

5.1. PPI-30: Single transition curves

PPI-30 formulation data is fitted on the interval $35\text{ °C} \leq T \leq 95\text{ °C}$ and a \vec{w} with edge regions of 10 °C is applied. The non-linear curve fitting parameters were calculated using a constraint vector κ which favors the ratio and its derivative.

For the two-state, single transition fit 11 parameters are obtained and the *weighted sum of squared errors of prediction* (SSE) is calculated. The SSE limit is set at $3 \cdot \sigma_{SSE}$, which excluded 0 out of 45 samples. Results for the formulations with 0 mM NaCl are consistent with a single transition fit, and the fit at pH 7.5 with 0 mM NaCl is plotted as an example (Figure 3).

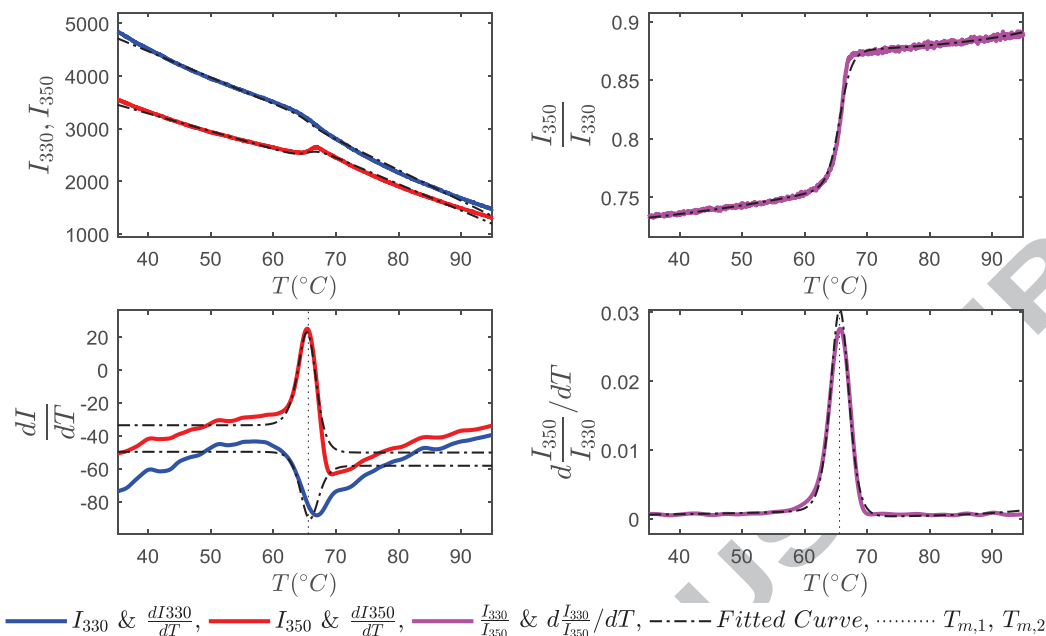


Figure 3. Two-state fit of PPI-30, at pH 7.5, 0 mM NaCl. The fit was made focusing on minimizing the errors in the ratio and the derivative of the ratio (in purple).

This two-state model shows a larger fitting error at salt concentrations 70 mM and 140 mM NaCl. Investigation of the raw data showed that a two-step transition is visible in the raw-fluorescence channels: An initial major change in fluorescence ratio followed by a large increase in both I_{350} and I_{330} (Figure 4, A). The chosen penalties, $\kappa_{1...4}$, focus on the ratio and its derivative. Only one transition is fitted, which matches the reference methodology. Samples with 0 mM NaCl show the same pattern, however, the ratio change and the increase happen concurrently.

Fitting a three-state model allows for the fitting of a second transition (Figure 4, B). These fits show definite improvements both in SSE, SSE spread, and visual inspection of the fits. Analysis of the fitting-error ($3\sigma_{SSE}$) excluded 3 of 45 fits. Both the two- and three-state models of PPI-30, formulated at pH 7.5 and 140 mM NaCl, are shown to illustrate the effect of the increased complexity of the model (Figure 4).

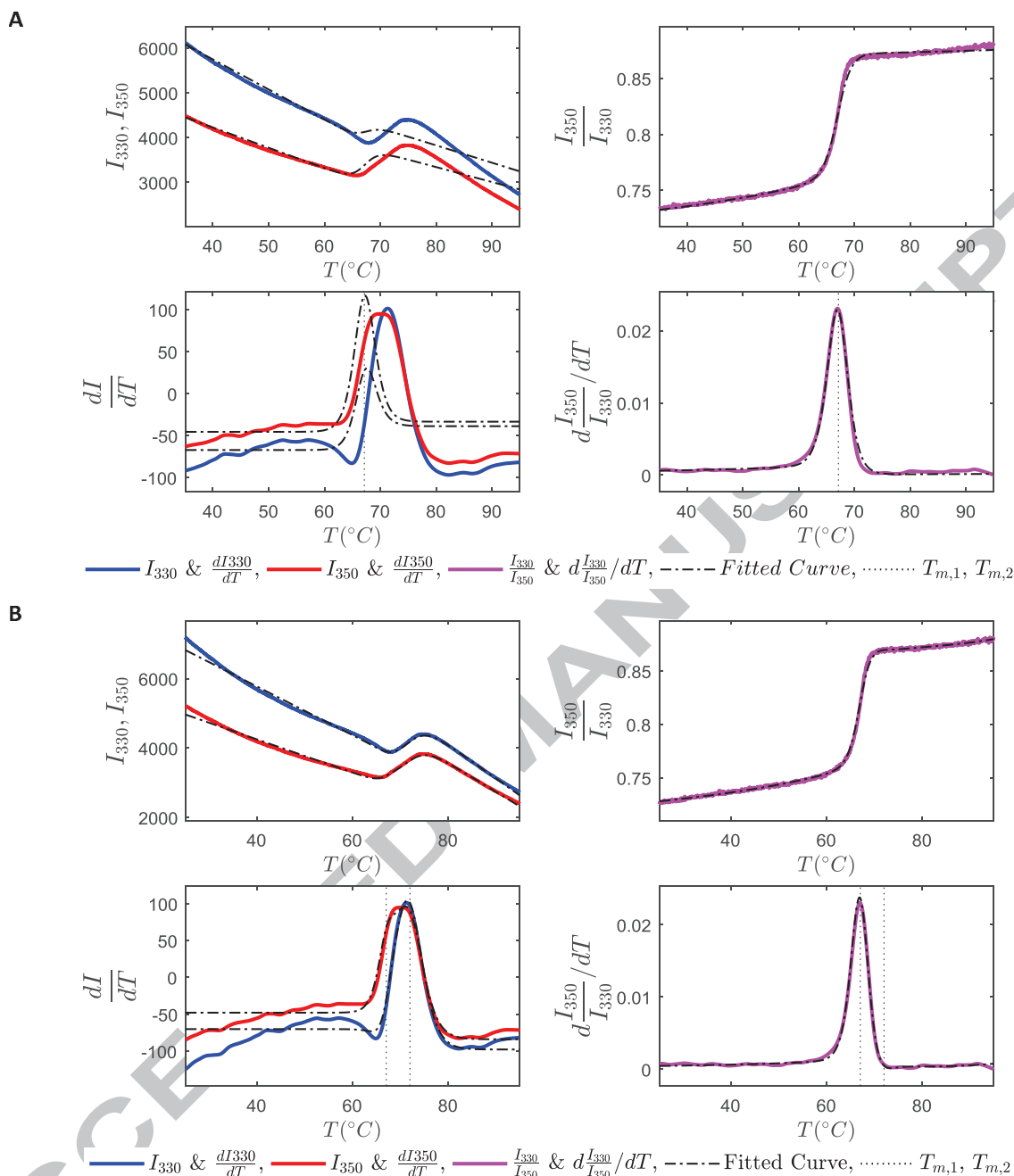


Figure 4. Two-state (A) and three-state (B) fits of PPI-30 at pH 7.5, 140 mM NaCl. The two-state model fits the ratio changes, focusing on peak deformation. The three-state model incorporates the fluorescence intensity to fit a second transition, this second transition is characterized by increased intensity without an observable shift in the ratio.

The first transition, S_1 , corresponds to the majority of the ratio change, where the peak is shifted towards a higher wavelength, resulting in increased f_{350}/f_{330} . The second transition, S_2 , is an increase in total peak intensity with only a minor peak shift. This change can only be appreciated by inspecting the two plots on the left in Figure 4B, where the change in fluorescence intensity is plotted.

An intermediate state ϕ_2 is fitted through the three-state model (Figure 5).

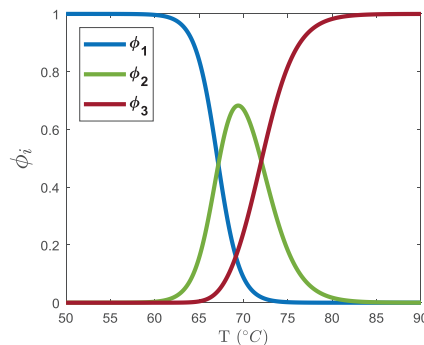


Figure 5. Extracted protein fractions $\phi_{1\dots3}$ of the fit of PPI-30 at pH 7.5 and 140 mM NaCl, corresponding to the fit depicted in Figure 4B. (print: Color)

For the three-state fit of PPI-30 the predicted $T_{m,1}$ and T_{onset} , at 4% conversion, are plotted against the reference values (Figure 6). Predicted $T_{m,2}$ values can be found in the supplementary materials, but cannot be compared directly with references values as only $T_{m,1}$ is obtained by the reference method. Increasing the ionic strength destabilizes PPI-30 at low pH, which is observed as a lower T_{onset} and $T_{m,1}$. At high pH the effect is reversed, and 70 and 140 mM added NaCl increase stability. In previous work, it was shown that pH strongly influences the thermal stability of PPI-30. Additionally, it is shown that increasing the ionic strength has an unfavourable effect on the colloidal stability of PPI-30 at low pH, causing the protein to aggregate at a lower temperature. The aggregation of the protein during thermal unfolding will cause a shift of the apparent protein melting temperature, $T_{m,1}$. [27]

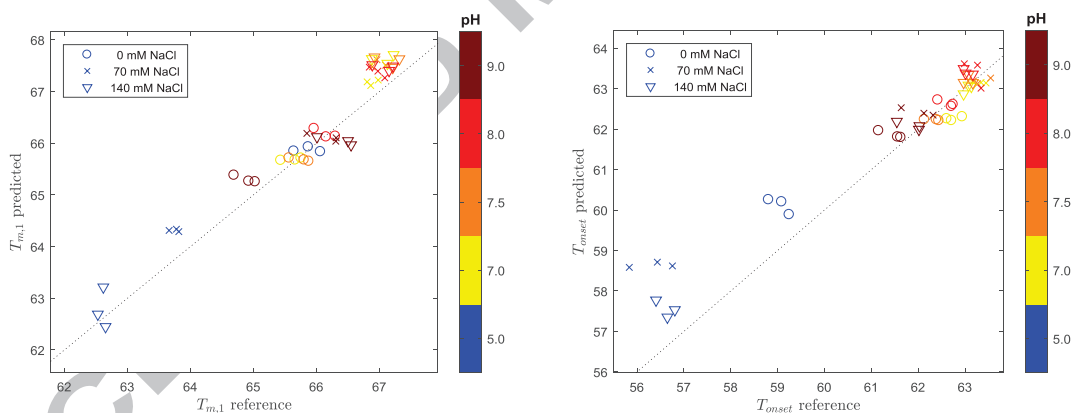


Figure 6. PPI-30 three-state fitting results, extracted T_{onset} predictions versus references values (left) and extracted $T_{m,1}$ predictions versus reference values (right).

5.2. PPI-13: Separated transition curves

Fits for 144 PPI-13 experiments were obtained using a three-state model on the interval $25\text{ °C} \leq T \leq 95\text{ °C}$. A \vec{w} with an edge region of 10 °C on both sides was used (i.e. in the $25\text{--}35\text{ °C}$ and the $85\text{--}95\text{ °C}$ range). Constraints κ were optimal with respects to the fitting error at equally mixed SSE for the various data forms, and an increased penalty on the derivative of the ratio and asymmetry factors. Example of a fitted model is depicted in Figure 7.

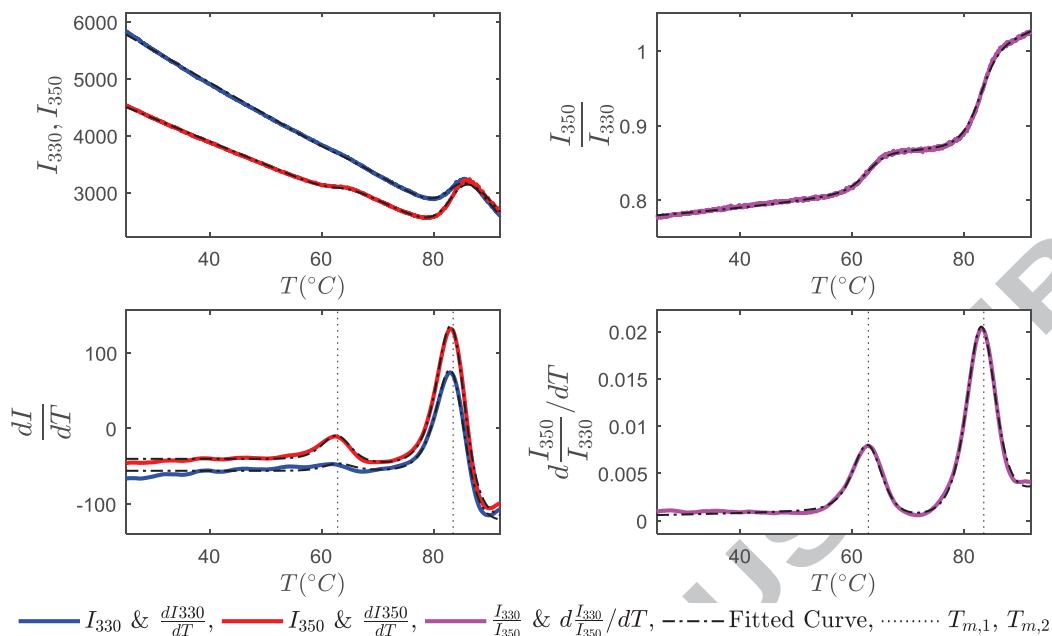


Figure 7. Fitted model of PPI-13 pH 6.0, 0 mM NaCl. The plots from left to right are fluorescence intensities, fluorescence ratio, 1st derivative of the intensity and 1st derivative of the ratio. (print: Color)

Clear trends are observed in predicted $T_{m,1}$ and $T_{m,2}$ values, with only small deviations in the cross-lab replicates (Figure 8). Analysis of the fitting-error and cut-off at $3 \cdot \sigma_{SSE}$ excluded 3 of 144 fits. At 0 mM NaCl and low pH (5.0 & 5.5), two outliers can be found. In these cases, the algorithm had problems distinguishing the first transition as there is little to no change in I_{330} (Figure 7), which, in combination with a strong baseline above $T_{m,2}$, created a local minimum that ignored $T_{m,1}$. Overall the low pH samples yielded relatively high SSE.

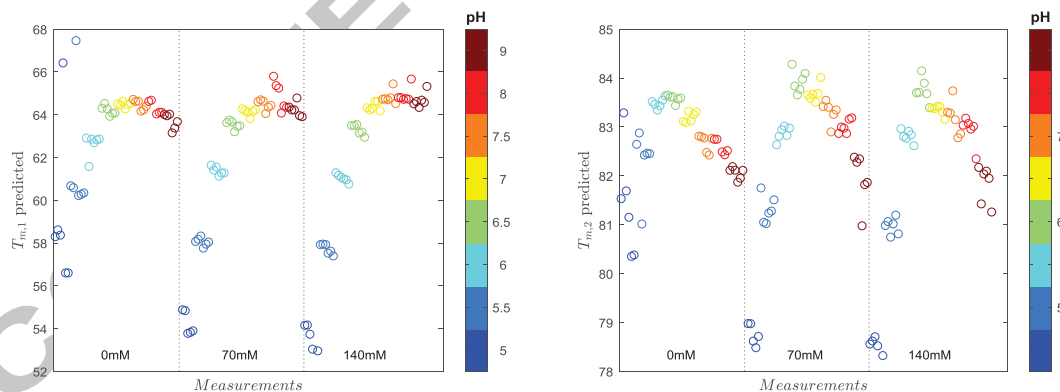


Figure 8. Trends in predicted (a) $T_{m,1}$ and (b) $T_{m,2}$ values for PPI-13. Noticeable are two outliers, at pH 5.0 and pH 5.5, showing too high $T_{m,1}$ values, confirmed by visual inspection of the fit (not shown). The two cross-lab are plotted in order, the three left-most markers correspond to the first formulation procedure, followed by three replicates from the second formulation procedure. (print: Color)

Predicted $T_{m,1}$ and $T_{m,2}$ are in good agreement with reference values (Figure 9). Extracted T_{onset} values can be found in the supplementary materials.

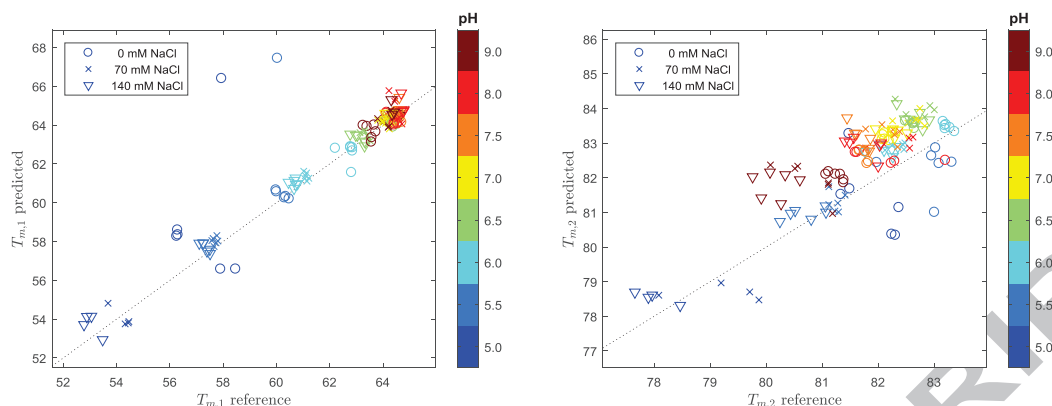


Figure 9. Predicted $T_{m,1}$ (left) and $T_{m,2}$ (right) values for the fits obtained from PP-13. The two low pH outliers also found in Figure 8 show high values for $T_{m,1}$, but fell within the SSE limit. (print: Color)

5.3. PPI-03: Overlapping transition curves

Fits for 144 PPI-03 experiments were obtained by applying the three-state model on the interval $25\text{ }^{\circ}\text{C} \leq T \leq 95\text{ }^{\circ}\text{C}$. A \vec{w} with an edge region of $10\text{ }^{\circ}\text{C}$ on both sides was used, similar to PPI-13. Constraints κ were optimal with respects to the fitting error around equally mixed SSE for the various data forms, with an increased penalty on the derivative of the ratio and asymmetry factors. An example of the fit for highly overlapping transitions, pH 7.5 with 0 mM NaCl, is shown Figure 10.

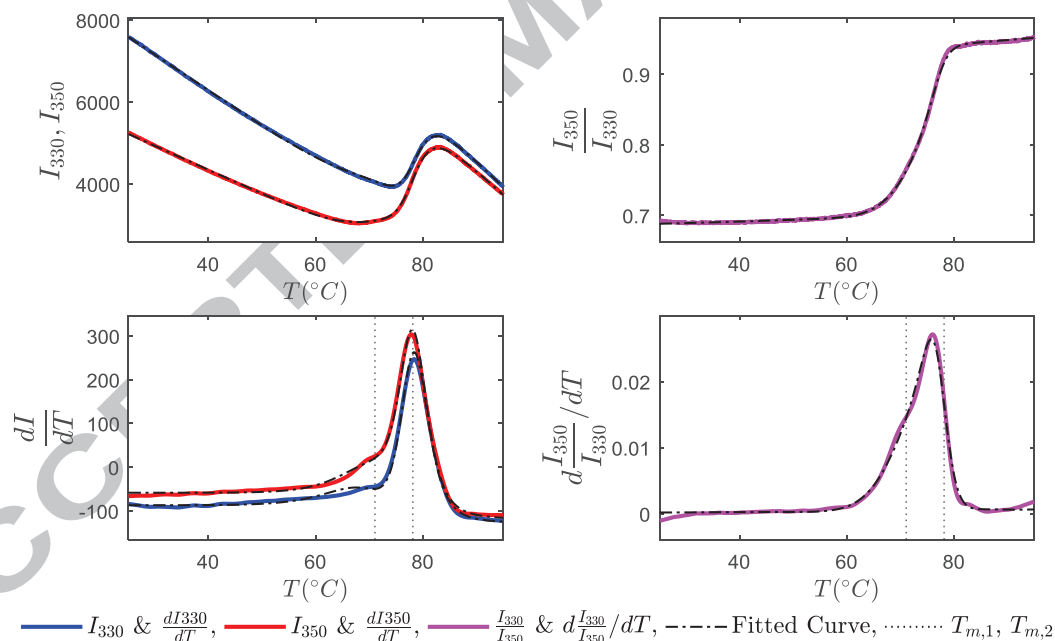


Figure 10. Example fit of overlapping of $T_{m,1}$ and $T_{m,2}$, as found in PPI-03 at pH 7.5, 0 mM NaCl. (print: Color)

Analysis of the fitting-error showed relatively high SSE values for low pH values, a cut-off at $3 \cdot \sigma_{SSE}$ is applied, which excludes 3 of 143 fits, the SSE can be found in the supplementary materials.

The remaining fitted parameters and extracted values show good agreement in experimental triplicates, cross-laboratory duplicates, and with respect to observed trends in both pH and salt concentration. Clear trends are observed in predicted $T_{m,1}$ and $T_{m,2}$ values, with only small deviations in the cross-lab replicates (Figure 11).

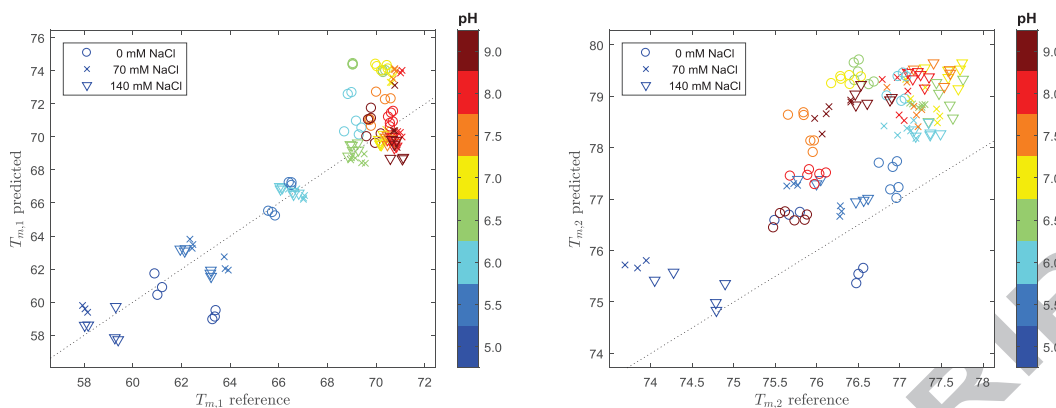


Figure 11. Predicted $T_{m,1}$ and $T_{m,2}$ values for the fits obtained from PPI-03. (print: Color)

For IgG1 antibodies, $T_{m,1}$ unfolding is strongly associated with aggregation, and T_{onset} is an important measurement as small amounts of partially denatured protein will initiate aggregation[8]. The trends in predicted T_{onset} are calculated for PPI-03 (Figure 12, left). The predicted T_{onset} at 0 mM NaCl shows a deviating pattern, with apparent miss fitting of samples at pH 6, 0 mM NaCl. This shows that the proposed method and the reference method determined different $T_{m,2}$ and T_{onset} values in this case. The applied three-state fit's deconvolution allows for the estimation of T_{onset} without an overlap effect of the second transition. This second transition overlap is stronger at high pH, as plotted for a subset of samples in Figure 1. The difference between the fitted and reference onset temperatures (Figure 12, right) therefore does not emulate that of PPI-30 and PPI-13.

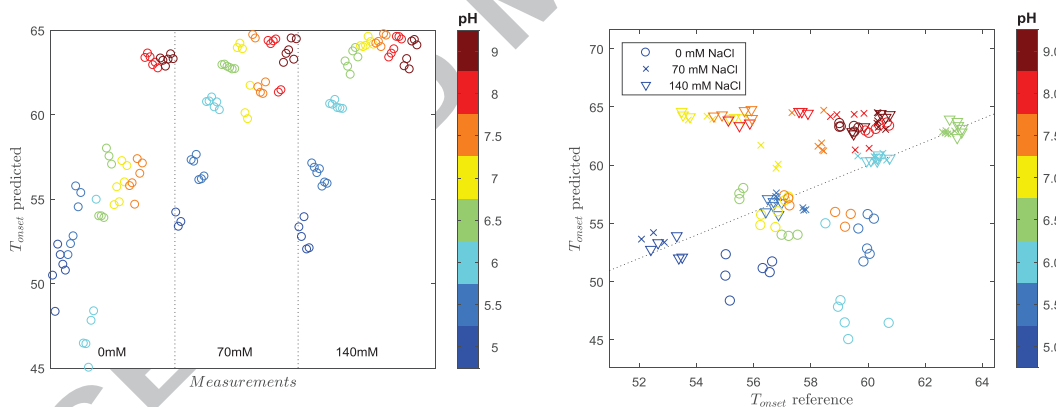


Figure 12. Predicted T_{onset} for PPI-03. (left) Trends in the formulation series are visible. The two cross-lab triplicates are plotted in order, the three left-most markers correspond to the first formulation procedure. (right) Reference method T_{onset} values are highly deviating. Note that formulation at pH 8 and 9 are in tris buffer, which would explain the deviating T_{onset} at 0 mM NaCl. The triplicate at pH 6, 0 mM NaCl is noted as an outlier, its counterpart laboratory duplicate has been filtered on the basis of high SSE values. (print: Color)

5.4. Result overview

In Table 2, we have summarized the differences between the suggested novel methodology, and the existing methodology. In other words, we are looking across all statistics, as have been shown for PPI-03 in Figure 11 and Figure 12 (right). High correlation coefficients are found for PPI-30 and PPI-13. For PPI-03, the double transition, the T_{onset} shows poor correlation (R^2) with the reference values, as the transitions highly overlap (Figure 12, right).

Table 2. Summary of the comparisons between the reference values and the fitted models.

$T_{m,1}$	$T_{m,2}$	T_{onset}
-----------	-----------	-------------

protein	Model states (J)	σ_{SSE} cutoff	R^2 (corr. coef.)	RMSE ($^{\circ}C$)	Offset	Reference SD_{pooled}	R^2 (corr. coef.)	RMSE ($^{\circ}C$)	Offset	Reference SD_{pooled}	R^2 (corr. coef.)	RMSE ($^{\circ}C$)	Offset	Reference SD_{pooled}
PPI-30	2	3	0.962	0.258	-0.020	0.18					0.964	0.691	-0.390	0.42
PPI-30	3	3	0.940	0.401	-0.233	0.18					0.952	0.779	-0.340	0.42
PPI-13	3	3	0.887	1.116	-0.273	0.28	0.650	1.046	-0.712	0.32	0.741	1.825	0.583	0.37
PPI-03	3	3	0.822	1.899	-0.465	0.49	0.521	1.846	-1.646	0.23	0.078	5.046	-1.221	0.67

6. Discussion

The proposed methodology has a large number of fitted parameters leading to a large feasible solution space for the basic fits, f_{350} and f_{330} , with multiple local minima. A conservative approach is used with regards to the asymmetry of the fit, in order to avoid fitting sloped $\frac{I_{350}}{I_{330}}$ baselines as part of the transition. However, a sloped first derivative of the ratio at baseline regions is challenging in two ways. (1) The prediction of non-pure and intermediate states can overlap with transition slopes, which hinders accurate estimation of the transition. (2) These slopes are indicative of physical changes to the sample, such as aggregation, which would also hinder conventional analysis of the data. In conventional analysis the onset transition temperature is determined by the start of the transition where the interference is minimal, rather than by the entire transition, potentially yielding better results for more asymmetric transitions. A 1-to-1 relation in the relative decline of the baselines was attempted in order to reduce the number of fitted parameters in the three-state model from 18 to 15. It was found that non-linear baselines in the fluorescence ratio $\frac{I_{350}}{I_{330}}$ could not be fitted, resulting in overall poor optimization of the fits with high values in ν_s .

The proposed method can take into account three-state unfolding transition. How an extended version of this method for a four-state unfolding pattern would behave is difficult to state. The difficulty concerns the expansion of the number of input parameters, the corresponding constraints, and the computational difficulty and evaluation thereof. However, current data analysis methods for IF DSF data struggle with overlapping transition states, indicating that it is likely that the extension of this novel methodology would be a better alternative.

In the proposed fitting model, there is no distinction between the two possible intermediate states: the order of unfolding steps is set. The distinction between proteins with different unfolded domains, that is to say with two intermediate states, is not feasible, neither by the proposed non-linear curve resolution algorithm nor by conventional analysis.

Data artefacts outside of the transition range will lead to high SSE values, even when fitted correctly, and can also create local, unexpected minima. In both cases a conservative approach would disregard this data, leading to a remeasurement of triplicates or loss of accuracy.

Optimization of the fitting penalties can be subjective between different formulation series, and should be considered when comparing different protein formulation series based on trends obtained. This has become especially clear when both the two- and three-state fit of PPI-30 are considered.

Opportunities for an analogous, multi-state, non-linear fit model lie with isothermal chemical denaturation experiments and with extrinsic dye-based DSF experiments. Currently retailed nanoDSF™ instruments can be outfitted with backscatter detectors to measure particle formation during the temperature ramp. This additional information could be incorporated into the fit, or used to evaluate at what fraction aggregation occurs.

7. Conclusion

The non-linear curve fitting of PPI-30 and PPI-13 formulation screening data yields fitted curves that closely resemble the measured data. Extracted parameters T_{onset} , $T_{m,1}$ and $T_{m,2}$ values are in agreement with reference measurements and trends in the formulation data are observed. The non-linear fits are obtained in an objective, uniform manner that yields robust results after optimization of the penalties. Assumptions in the proposed method, such as a linear baseline near the transition, have to be evaluated in order to appropriately apply the method.

The largest benefit of applying the novel method instead of the existing methodology, is for proteins with highly overlapping transitions. This can be appreciated through the evaluation of the extracted parameters $T_{m,1}$ and $T_{m,2}$ of PPI-03, which has strongly overlapping transitions. The results here cannot directly be obtained by the traditional method. Correction of the fitting of the second transition allows for correct evaluation of the T_{onset} , yielding better insight into a critical protein stability parameter, clearly showing the benefits of the proposed fitting method.

8. Acknowledgements

This study was funded as part of the EU Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie grant agreement No 675074.

9. Conflicts of Interest

The authors declare no conflicts of interest.

10. References

- [1] T.J. Kamerzell, R. Esfandiary, S.B. Joshi, C.R. Middaugh, D.B. Volkin, Protein-exciipient interactions: mechanisms and biophysical characterization applied to protein formulation development, *Advanced Drug Delivery Reviews*, 63 (2011) 1118-1159.
- [2] F. He, C.E. Woods, G.W. Becker, L.O. Narhi, V.I. Razinkov, High-throughput assessment of thermal and colloidal stability parameters for monoclonal antibody formulations, *Journal of pharmaceutical sciences*, 100 (2011) 5126-5141.
- [3] M.L. Brader, T. Estey, S. Bai, R.W. Alston, K.K. Lucas, S. Lantz, P. Landsman, K.M. Maloney, Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies, *Molecular pharmaceutics*, 12 (2015) 1005-1017.
- [4] E.Y. Chi, S. Krishnan, T.W. Randolph, J.F. Carpenter, Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation, *Pharmaceutical research*, 20 (2003) 1325-1336.
- [5] V. Filipe, A. Hawe, H. Schellekens, W. Jiskoot, Aggregation and immunogenicity of therapeutic proteins, in: W. Wang, C.J. Roberts (Eds.) *Aggregation of therapeutic proteins*, 2010, pp. 403-433.
- [6] C.J. Roberts, T.K. Das, E. Sahin, Predicting solution aggregation rates for therapeutic proteins: Approaches and challenges, *International Journal of Pharmaceutics*, 418 (2011) 318-333.
- [7] F.H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nature protocols*, 2 (2007) 2212.
- [8] D.S. Goldberg, S.M. Bishop, A.U. Shah, H.A. Sathish, Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: Role of conformational and colloidal stability, *Journal of pharmaceutical sciences*, 100 (2011) 1306-1315.

- [9] P. Garidel, M. Hegyi, S. Bassarab, M. Weichel, A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy, *Biotechnology Journal*, 3 (2008) 1201-1211.
- [10] J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Third ed., Springer Science & Business Media, 2006.
- [11] M.R. Eftink, The use of fluorescence methods to monitor unfolding transitions in proteins, *Biophysical journal*, 66 (1994) 482-501.
- [12] M.R. Eftink, Fluorescence techniques for studying protein structure, *Methods of biochemical analysis*, 35 (1991) 127-205.
- [13] D.B. Temel, P. Landsman, M.L. Brader, 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: A.L. Feig (Ed.) *Methods in Enzymology*, Elsevier, 2016, pp. 359-389.
- [14] C. Duy, J. Fitter, How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra, *Biophysical journal*, 90 (2006) 3704-3711.
- [15] S.O. Ugwu, S.P. Apte, The effect of buffers on protein conformational stability, *Pharmaceutical Technology*, 28 (2004) 86-109.
- [16] J. Gally, G. Edelman, The effect of temperature on the fluorescence of some aromatic amino acids and proteins, *Biochimica et biophysica acta*, 60 (1962) 499-509.
- [17] G. Žoldák, D. Jancura, E. Sedlák, The fluorescence intensities ratio is not a reliable parameter for evaluation of protein unfolding transitions, *Protein Science*, 26 (2017) 1236-1239.
- [18] T. Menzen, W. Friess, High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants, *Journal of pharmaceutical sciences*, 102 (2013) 415-428.
- [19] F. He, S. Hogan, R.F. Latypov, L.O. Narhi, V.I. Razinkov, High throughput thermostability screening of monoclonal antibody formulations, *Journal of pharmaceutical sciences*, 99 (2010) 1707-1720.
- [20] H. Svilenov, G. Winter, Rapid sample-saving biophysical characterisation and long-term storage stability of liquid interferon alpha2a formulations: Is there a correlation?, *International Journal of Pharmaceutics*, 562 (2019) 42-50.
- [21] A. Braun, L. Kwee, M.A. Labow, J. Alsenz, Protein Aggregates Seem to Play a Key Role Among the Parameters Influencing the Antigenicity of Interferon Alpha (IFN- α) in Normal and Transgenic Mice, *Pharmaceutical research*, 14 (1997) 1472-1478.
- [22] E. Garber, S.J. Demarest, A broad range of Fab stabilities within a host of therapeutic IgGs, *Biochemical and Biophysical Research Communications*, 355 (2007) 751-757.
- [23] L. Martin, S. Schwarz, D. Breitsprecher, Analyzing Thermal Unfolding of Proteins, *The Prometheus NT*. 48. Application Note NT-PR-001, (2014).
- [24] K.L. Lazar, T.W. Patapoff, V.K. Sharma, Cold denaturation of monoclonal antibodies, *mAbs*, 2 (2010) 42-52.
- [25] M.E. Harder, M.L. Deinzer, M.E. Leid, M.I. Schimerlik, Global analysis of three-state protein unfolding data, *Protein Science*, 13 (2004) 2207-2222.
- [26] MathWorks, *Matlab Optimization Toolbox User's Guide*, (2018).
- [27] J.M. Sanchez-Ruiz, Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry, *Biophysical journal*, 61 (1992) 921-935.

11. Supplementary Materials.

A1. MATLAB® Implementation example, three-state fit.

B1. Fitting overview PPI-30, two-state fit.

B2. Fitting overview PPI-30, three-state fit.

C1. Fitting overview PPI-13, Three-state fit.

D1. Fitting overview PPI-13, three-state fit.

ACCEPTED MANUSCRIPT

A1. MATLAB® Implementation example, three-state fit.*Data definition*

```
%          Data                                vector size.
xm; % Temperature, vector;                    [ N   x 1 ]
ym1; % I330                                  [ N   x 1 ]
ym2; % I350                                  [ N   x 1 ]
rym; % I350/I330                             [ N   x 1 ]
sym1; % 1st derivative I330                  [ N-1 x 1 ]
sym2; % 1st derivative I350                  [ N-1 x 1 ]
srym; % 1st derivative I350/I330            [ N-1 x 1 ]
```

kappa and weight vector

```
% kappa
kappa = [2 2 1 10 80 2];
% weight vector
clear a;
[~, a(1)] = min(abs(xm - min(xm + 10)));
[~, a(2)] = min(abs(xm - max(xm - 10)));
W = [linspace(0, 1, a(1)) ones(1, a(2) - a(1)) linspace(1, 0, length(xm) - a(2))];
clear a;
```

Fitting function

```
S = @(s, x) 1./(1 + exp(s(2).*(s(1) - x))).^(1 / s(3));

fm = @(p, x) sum((S(p(1:3), x).* S(p(4:6), x) * [0 -1 1] + S(p(1:3), x)* ... % Fitted Signal
[-1 1 0]+ [1 0 0]) .* [p(7) + p(8) .* x, p(9) + p(10) .* x, p(11) + p(12) .* x], 2);

fratio = @(p, x) ... % Fitted ratio
sum((S(p(1:3), x) .* S(p(4:6), x) * [0 -1 1] + S(p(1:3), x) * [-1 1 0] + [1 0 0]) .* ...
[p(13) + p(14) .* x, p(15) + p(16) .* x, p(17) + p(18) .* x], 2) ./ ...
sum((S(p(1:3), x) .* S(p(4:6), x) * [0 -1 1] + S(p(1:3), x) * [-1 1 0] + [1 0 0]) .* ...
[p(7) + p(8) .* x, p(9) + p(10) .* x, p(11) + p(12) .* x], 2);

Tm = @(s)      s(:, 1) - log(-1 +          2 .* s(:, 3)) ./ s(:, 2);
Tx = @(s, frac) s(:, 1) - log(-1 + (1 ./ frac').^ s(:, 3)) ./ s(:, 2);
```

Layout of fitting parameters in p.

```
% p = [ A1 B1 nu1  A2 B2 nu2  ...
%      [b1,1 b1,2  b2,1 b2,2  b3,1 b3,2] ... % 330 nm
%      [b1,1 b1,2  b2,1 b2,2  b3,1 b3,2] ] ; % 350 nm
```

Objective function

```
objective = @(p) sum( ...
... % 1 Raw curves
(kappa(1) ./ sum([var(ym1) var(ym2)])) .* sum(w .* ((fm(p(1:12), xm) - ym1) .^ 2 + ...
(fm(p([1:6 13:18]), xm) - ym2) .^ 2))...
+ ... % 2 Ratio
(kappa(2) ./ var(rym)) .* sum(w .* (((fratio(p, xm) - (rym)) .^ 2)) ....
```



```

+ ... % 3 Raw 1st deriv
(kappa(3) ./ sum([var(sym1) var(sym2)])) .* sum(w(2:end) .* (diff(fm(p(1:12), xm) ...
- sym1) .^ 2 + (diff(fm(p([1:6 13:18]), xm)) - sym2) .^ 2))...
+ ... % 4 Ratio 1st deriv
(kappa(4) ./ (var(srym) * 8)) .* sum(w(2:end) .* ((diff(fratio(p, xm)) - srym) .^ 2)) ...
+ ... % Asymmetry penalty.
(kappa(5) * 1e5) * ((p(3) - 1) .^ 4 + (p(6) - 1).^ 2) ...
+ ... % Slope penalty
(kappa(6) * 1e5) * (sum((p(8:2:12) ./ p(7:2:12) - p(14:2:18) ./ p(13:2:18)).^ 2));

```

Initial p0

```

p0 = [50, .4, 1, 75, .4, 1, ... % initial guess: values within search space but not critical.
fliplr(polyfit(xm(xm < 45), ym1(xm < 45), 1)), ...
(fliplr(polyfit(xm(xm < 45), ym1(xm < 45), 1)) + ...
fliplr(polyfit(xm(xm > 88), ym1(xm > 88), 1))) ./ 2, ...
fliplr(polyfit(xm(xm > 88), ym1(xm > 88), 1)), ...
fliplr(polyfit(xm(xm < 45), ym2(xm < 45), 1)), ...
(fliplr(polyfit(xm(xm < 45), ym2(xm < 45), 1)) + ...
fliplr(polyfit(xm(xm > 88), ym2(xm > 88), 1))) ./ 2, ...
fliplr(polyfit(xm(xm > 88), ym2(xm > 88), 1)),,];

```

hard constraints

```

% Dynamic Constraints: A.*p <= b
A = [1 0 0 - 1 0 0 zeros(1, length(p0) - 6)];
b = - 3;
Aeq = [];
beq = [];
% non-dynamic Constraints
% Set the LB tighter, to aid the algorithm in spacing the search grid.
lb = [45 0.15 0.9 65 0.15 0.9 1000 - 200 1000 - 200 1000 - 200 1000 - 200 1000 - 200 1000 - 200];
ub = [70 0.60 1.1 85 0.60 1.1 inf 0 inf 0 inf 0 inf 0 inf 0];
fmincon_opt = optimset(optimset('fmincon'));
fmincon_opt = optimset(fmincon_opt, 'disp', 'iter', 'LargeScale', 'on', 'MaxIter', 4000, ...
'MaxFunEvals', 4000);
fmincon_opt = optimset(fmincon_opt, 'Algorithm', 'sqp', 'ScaleProblem', 'obj-and-constr');
% For more information you can see
% https://se.mathworks.com/help/optim/ug/fmincon.html
% https://se.mathworks.com/help/optim/ug/choosing-the-algorithm.html

```

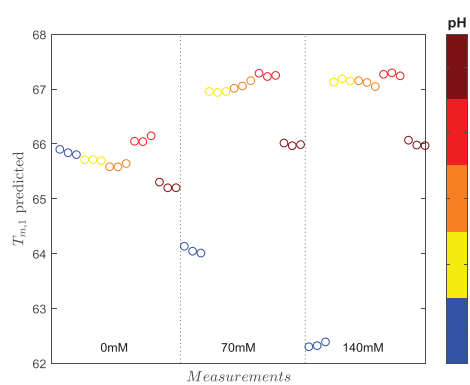
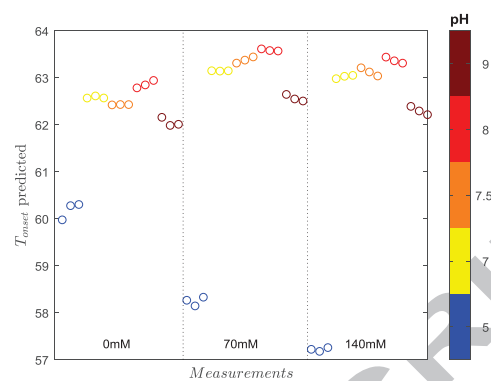
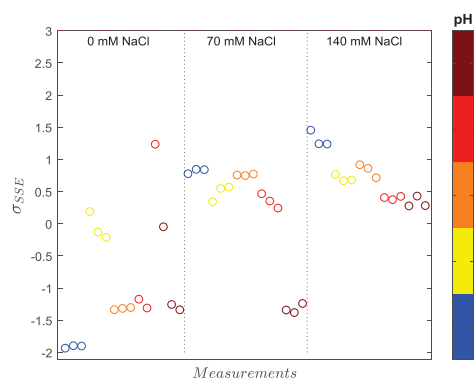
Fitting

```

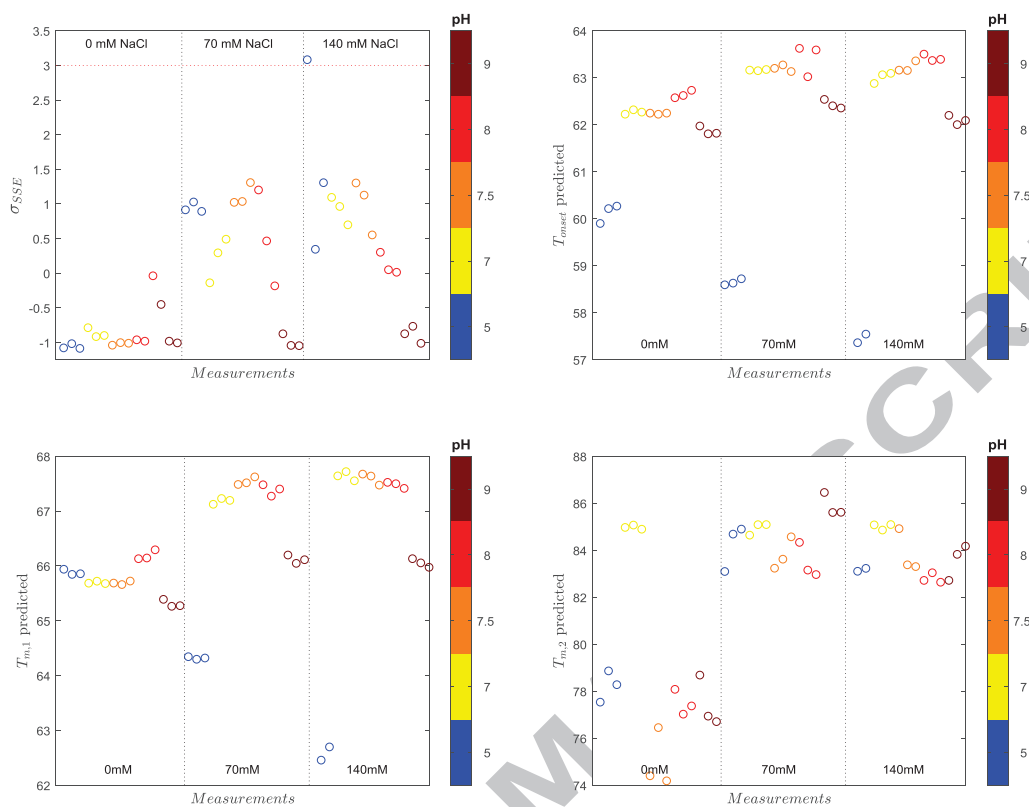
popt = fmincon(objective, p0, A, b, Aeq, beq, lb, ub, [], fmincon_opt);

```

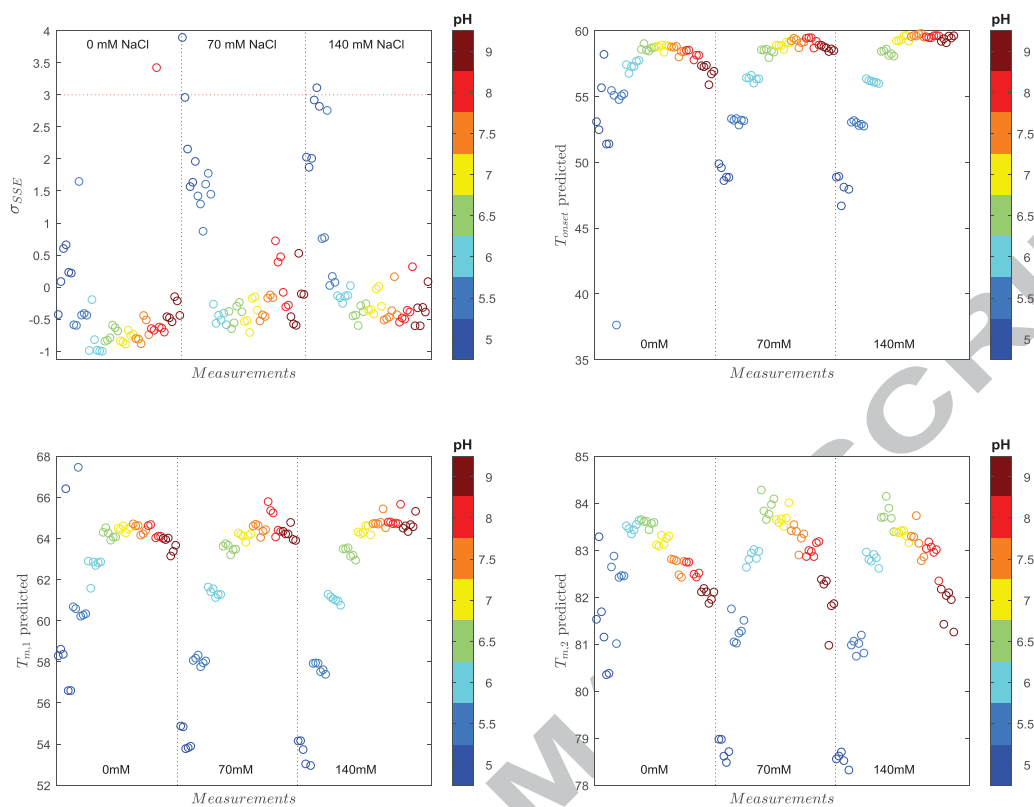
B1. Fitting overview PPI-30, two-state fit. (print: Color)



B2. Fitting overview PPI-30, three-state fit. (print: Color)

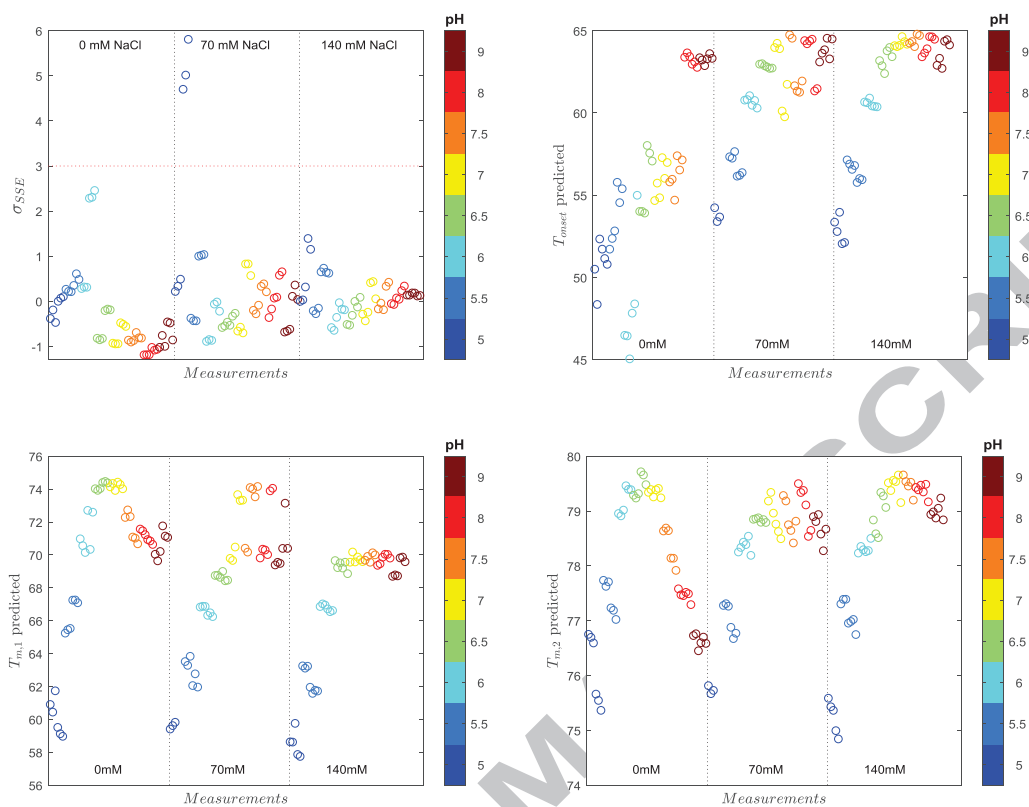


C1. Fitting overview PPI-13, Three-state fit. (print: Color)



The two cross-lab triplicates are plotted in order, the three left-most markers correspond to the first formulation procedure.

D1. Fitting overview PPI-13, three-state fit. (print: Color)



The two cross-lab triplicates are plotted in order, the three left-most markers correspond to the first formulation procedure.

