

UCSF

UC San Francisco Previously Published Works

Title

DSFworld: A flexible and precise tool to analyze differential scanning fluorimetry data.

Permalink

<https://escholarship.org/uc/item/9hk68764>

Journal

Protein Science, 33(6)

Authors

Wu, Taiasean

Gale-Day, Zachary

Gestwicki, Jason

Publication Date

2024-06-01

DOI

10.1002/pro.5022

Peer reviewed

DSFworld: A flexible and precise tool to analyze differential scanning fluorimetry data

Taiasean Wu | Zachary J. Gale-Day | Jason E. Gestwicki 

Department of Pharmaceutical Chemistry, Chemistry & Chemical Biology Program and the Institute for Neurodegenerative Diseases, University of California San Francisco, San Francisco, California, USA

Correspondence

Jason E. Gestwicki, Sandler Neuroscience Center, Room 311, 675 Nelson Rising Lane, San Francisco, CA 94158, USA.
Email: jason.gestwicki@ucsf.edu

Funding information

National Institutes of Health, Grant/Award Numbers: GM141299, AR081704; National Science Foundation, Grant/Award Number: 1000259744

Review Editor: Nir Ben-Tal

Abstract

Differential scanning fluorimetry (DSF) is a method to determine the apparent melting temperature (T_{ma}) of a purified protein. In DSF, the raw unfolding curves from which T_{ma} is calculated vary widely in shape and complexity. However, the tools available for calculating T_{ma} are only compatible with the simplest of DSF curves, hindering many otherwise straightforward applications of the technology. To overcome this limitation, we designed new mathematical models for T_{ma} calculation that accommodate common forms of variation in DSF curves, including the number of transitions, the presence of high initial signal, and temperature-dependent signal decay. When tested these models against DSFbase, an open-source database of 6235 raw, real-life DSF curves, these models outperformed the existing standard approaches of sigmoid fitting and maximum of the first derivative. To make these models accessible, we created an open-source software and website, DSFworld (<https://gestwickilab.shinyapps.io/dsfworld/>). In addition to these improved fitting capabilities, DSFworld also includes features that overcome the practical limitations of many analysis workflows, including automatic reformatting of raw data exported from common qPCR instruments, labeling of data based on experimental variables, and flexible interactive plotting. We hope that DSFworld will enable more streamlined and accurate calculation of T_{ma} values for DSF experiments.

KEYWORDS

biological software, curve fitting, protein stability, thermal shift assay, thermoflour

1 | INTRODUCTION

Differential scanning fluorimetry (DSF) is an *in vitro* method to determine the apparent melting temperature (T_{ma}) of a purified protein. In DSF, a protein solution, typically at low micromolar concentrations, is combined

with a reporter dye, typically Sypro Orange, and heated through unfolding in a standard qPCR instrument (Pantoliano et al., 2001). The reporter dye fluoresces in the presence of unfolded protein, producing a fluorescent unfolding curve from which T_{ma} can be calculated (Semisotnov et al., 1991; Wu, Yu, et al., 2023). Common

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

DSF applications include the assessment of ligand binding (Huynh & Partch, 2015; Ravalin et al., 2019) and optimization of buffers for shelf-stability or structural characterization (Boivin et al., 2013; Chari et al., 2015; Reinhard et al., 2013; Ristic et al., 2015). The theory (Gao et al., 2020; Wu et al., 2020), applications (Garlick & Mapp, 2020; Scott et al., 2016; Simeonov, 2013), and protocols (Huynh & Partch, 2015; Wu, Hornsby, et al., Wu, Hornsby, et al., 2023b) for DSF have all been extensively reviewed.

However, despite the widespread use of DSF, the calculation of Tma remains a major challenge. Tma is defined as the midpoint of the unfolding curve and is traditionally modeled as a sigmoidal transition. However, in reality, DSF curves are not strictly sigmoidal; rather, they vary in shape and complexity, often including multiple transitions and/or temperature-dependent effects (Wu et al. see companion paper) that are not readily ascribed to traditional two-state unfolding. Several software packages for the analysis of DSF data have been published, but these approaches either do not address such non-canonical aspects of DSF data (Martin-Malpartida et al., 2022; Phillips & Hernandez de la Peña, 2011; Sun et al., 2020; Wang et al., 2012), reject curves outside the most straightforward archetype (Rosa et al., 2015), or coerce data into a simpler form by truncation or masking of less-sigmoidal data points (Lee et al., 2019; Schulz et al., 2013). Thus, most modern DSF analysis workflows rely on manual pre-processing steps to make curves amenable to standard sigmoid fitting. In addition to the challenges these manual steps can pose for efficiency and reproducibility, both pre-processing and subsequent sigmoid fitting perform inconsistently between different applications or even on different curves within the same experiment. As a result, DSF analysis workflows can be fragile and unreliable, ultimately limiting the uses of DSF technology as a whole.

We envisioned that an improved DSF analysis tool would incorporate a broader and more well-informed range of mathematical fitting models, accounting for the non-canonical features seen in real-world DSF data, such as multiple transitions, high initial fluorescence and temperature-dependent fluorescence decay. Towards that goal, we developed four models using a subset of 347 curves available in DSFbase, an open-source database of raw DSF data (Wu et al., n.d see companion paper). Then, the performance of the final models was tested using 5749 entries in DSFbase representing legitimate DSF curves (canonical and noncanonical subsets). The 1023 entries included to represent potentially uninterpretable (speculative subset) or entirely artifactual (errata subset) data were excluded from this analysis. To our knowledge, DSFworld is the only DSF analysis tool to be

developed and tested on a diverse set of DSF data, as previous software typically reports the use of 1–2 proteins in development and testing, while DSFbase includes data from a large, diverse set of >50 proteins. We named this analysis tool DSFworld, which is freely available online (<https://gestwickilab.shinyapps.io/dsfworld/>). Since its initial pre-publication launch (Wu, Hornsby, et al., 2023b), DSFworld has averaged over 1000 h of use per month.

In addition to expanded and improved curve fitting capabilities, DSFworld also supports practical steps of data analysis. These applications were identified in our own workflows as being gaps in current methods, including automatic reformatting of exported files from several common qPCR software (Biogen, Roche, qTower, Viia), labeling of data with experimental conditions, and interactive plotting of results. We hope that DSFworld can support efficient and consistent analysis of Tma data for a broad range of DSF experiments.

2 | RESULTS

Creation of DSFworld models. To determine the common points of variation between DSF data, we began by assembling a panel of 347 raw DSF results for 35 proteins that vary in molecular weight, biological activity, fold, and oligomeric state. In addition, these results included experimental variation in buffers, pH, concentrations of known ligands, SYPRO Orange concentrations, and heating rates. From these 347 DSF curves, four archetypes were visually identified (Figure 1a): a single transition with low initial fluorescence (Model 1), a single transition with high initial, decaying fluorescence (Model 2), two transitions with low initial fluorescence (Model 3), and two transitions with high initial, decaying fluorescence (Model 4). These archetypes were then mathematically defined as follows:

$$\text{RFU}(T) = \text{Sig}_1(T) \quad (\text{Model 1})$$

$$\text{RFU}(T) = \text{Sig}_1(T) + \text{Id}(T) \quad (\text{Model 2})$$

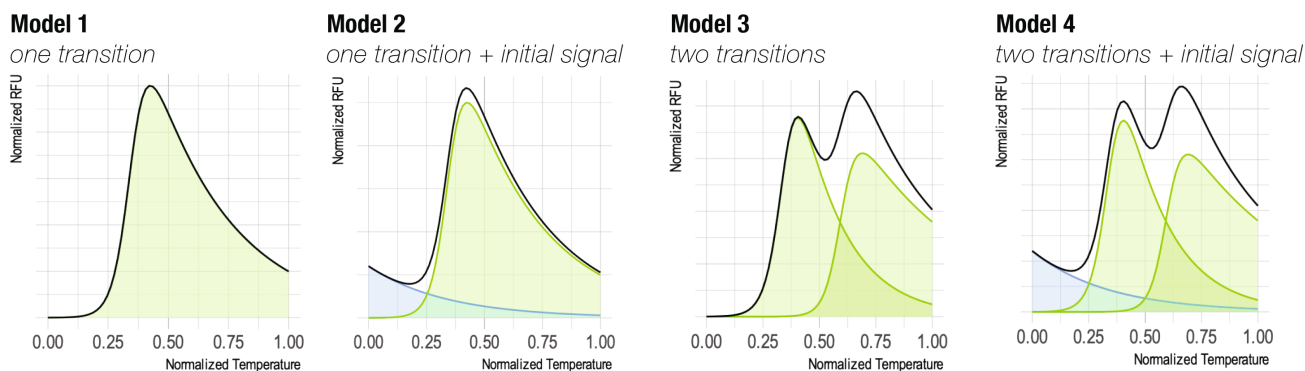
$$\text{RFU}(T) = \text{Sig}_1(T) + \text{Sig}_2(T) \quad (\text{Model 3})$$

$$\text{RFU}(T) = \text{Sig}_1(T) + \text{Sig}_2(T) + \text{Id}(T) \quad (\text{Model 4})$$

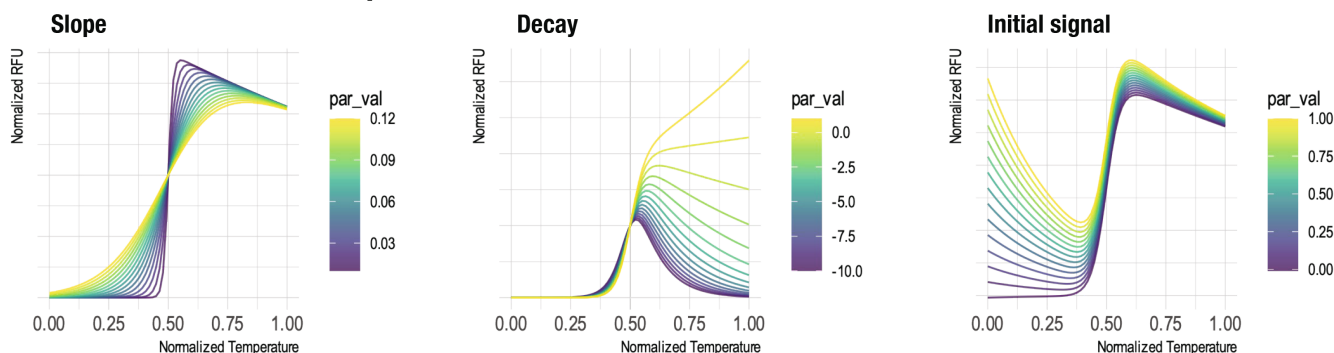
where the general form of the decaying sigmoids, $\text{Sig}_i(T)$, is:

$$\text{Sig}_i(T) = \frac{A_i}{1 + e^{(T_{\text{ma}_i} - T)/\text{scal}_i}} \times e^{d \times (T - T_{\text{ma}_i})}$$

(a) Components of DSFworld models



(b) Variation of Model 2 parameters



(c) Variation of Model 4 parameters

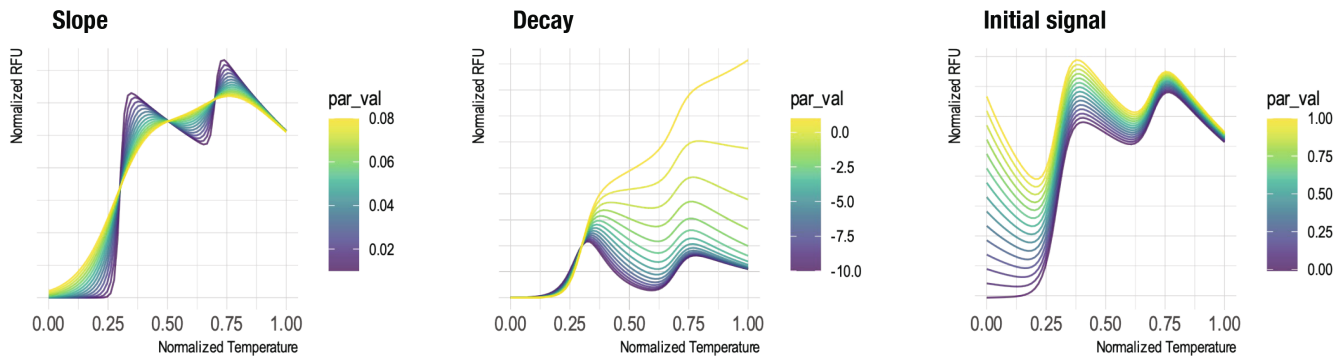


FIGURE 1 Summary of DSFworld models. Visual guide to the four DSFworld models, and the impact of changing key parameters. (a) Visual description of the separable components in each of the four DSFworld models (green: transition(s), blue: initial signal). The components are added together to produce the final model (black lines). (b and c) Impact of changing specific parameters on model results. Curves were generated by varying individual parameter values to either DSFworld Model 2 (b) or Model 4 (c). Varied parameters shown are left: slope (scal1), middle: Temperature-dependent signal decay (d1), and right: initial signal (id_d1, with id_b1 = -3).

- $\text{Sig}_i(T)$ is the RFU value at temperature T ;
- A_i is the scaling factor for the final sigmoid;
- Tma_i is the Tma_i ;
- scal_i controls the slope of the transition;
- d is the magnitude of the temperature-dependent RFU decay.

And where the general form of the initial decaying fluorescence, $\text{Id}(T)$, is:

$$\text{Id}(T) = C \times e^{(T \times \text{id})}$$

- $\text{Id}(T)$ is the RFU of the initial fluorescence at temperature T ;
- C is the starting value of the initial fluorescence;
- id defines the rate and linearity of the decay from C .

The final fitting procedure is described in detail in Appendix S1. Briefly, (1) both raw RFU and measured

temperatures are normalized to a range of 0–1. This normalization step minimizes inconsistencies in fitting introduced by arbitrary differences in measured temperature ranges and magnitudes of RFUs reported from different qPCR instruments. (2) From the normalized raw data, first- and second-order derivatives are calculated with respect to temperature using a Savitsky-Golay filter, and a smoothed, interpolated version of the first- and second-derivative data are then calculated using a Loess filter of span 0.1. The use of a Savitsky-Golay filter offers robust smoothing and derivative calculation across a wide range of noise levels, which can vary between experiments and instruments. At this step, T_{ma} values are calculated as the maximum of the smoothed, interpolated first derivative. (3) Starting parameters are estimated for each curve (see Appendix S1) based on local maxima and minima in the first and second derivative data for T_{ma}, and the magnitude of the initial fluorescence is estimated as the first relative fluorescence unit (RFU) value in the dataset. (4) Models are fit to each input curve, and the final model parameters are then used to generate curves for each individual component of the full fit (i.e., initial signal, each individual transition). (5) Finally, T_{ma} is calculated by taking the maximum of the first derivative of each isolated sigmoid component. This approach is analogous to separating a complex melting transition into its most likely individual unfolding transitions, followed by using the currently accepted methods of T_{ma} calculation. This way, the DSFworld analysis leverages the temperature-dependent fluorescence decays necessary for robust fitting of complex transitions while using practices consistent with existing DSF data analysis approaches. To visualize the impact of individual fitting parameters on curve shape, we varied the slope of the melting transition, the rate of decay of the fluorescence curve and the magnitude of the initial fluorescence and then plotted the corresponding results for Model 2 (Figure 1b) and Model 4 (Figure 1c). These results illustrate how the models incorporate the non-canonical portions of the curves into the overall fitting while also maintaining sigmoidal fits for the T_{ma} values.

The script and all associated functions to implement and modify this analysis outside of DSFworld are available on GitHub at <https://github.com/gestwicki-lab/dsfworld>. For more details, caveats and other considerations, see Appendix S1.

2.1 | Performance of DSFworld models with DSFbase

To test the performance of the models with a larger, diverse set of real-world DSF curves, each was then used to calculate T_{mas} for the data within DSFbase (Wu et al.,

see companion paper). DSFbase is a recently developed, open-access dataset containing 6235 raw DSF curves from over 50 different proteins across 79 different experiments, which vary in buffer, heating rate and other experimental variables. Buffer is a particularly important variable as some, such as Tris, are temperature dependent. DSFbase is intended to support the development of improved DSF analyses and interpretations, and it offers a highly diverse collection of raw DSF unfolding curves. In particular, DSFbase includes two subsets: “canonical” results (4144 entries), which are described as single-transitions and “non-canonical” results (1605 entries), which contain features such as multiple transitions and/or high initial signal. To test performance head-to-head, the 5749 raw DSF curves in DSFbase were fit using the four new models and the results compared to the use of the traditional sigmoid model. The data was truncated at the highest RFU value in the traditional sigmoidal model, consistent with standard practice.

We found that the traditional sigmoid fit failed in 17.2% of the tested cases, while the DSFworld models achieved a complete success rate (0% failure) (Figure 2a). In addition, the DSFworld models produced lower total residuals and had better model quality by Bayesian information criterion (BIC) (Figure 2b). This outperformance is particularly notable, as the sigmoid fitting procedure is given a considerable advantage by the truncation of the raw data to exclude the data points least amenable to sigmoid fitting. These results suggest that the DSFworld models are not only more robust, but also better describe DSF data.

As expected, fitting performance with DSFbase was readily optimized by choosing the best DSFworld model for each curve using the BIC. Among the options, Models 2 and 4 were selected as the superior model most frequently (41.2% and 31.4% of entities respectively), followed closely by Model 3 (22.3%) (Figure 2c). Comparatively, Model 1 was selected rarely (5.1%). As expected, Model 1 performed better when using the canonical subset of the DSFbase data (Figure S1), yet the overall findings suggest that incorporating non-canonical features, especially high initial fluorescence, improves a vast majority of DSF analysis workflows. To assess redundancy among the DSFworld models, the mean residual was calculated for each experiment when the data was systematically fit using only Model 1, Models 1 and 2, Models 1–3, or all four models. We found that the mean residual of all fits decreased with each additional model (Figure 2d), showing that the ability to choose among all four models is important for achieving the best possible fits. Representative fits are shown for each Model to illustrate the goodness of fit (Figure 2e).

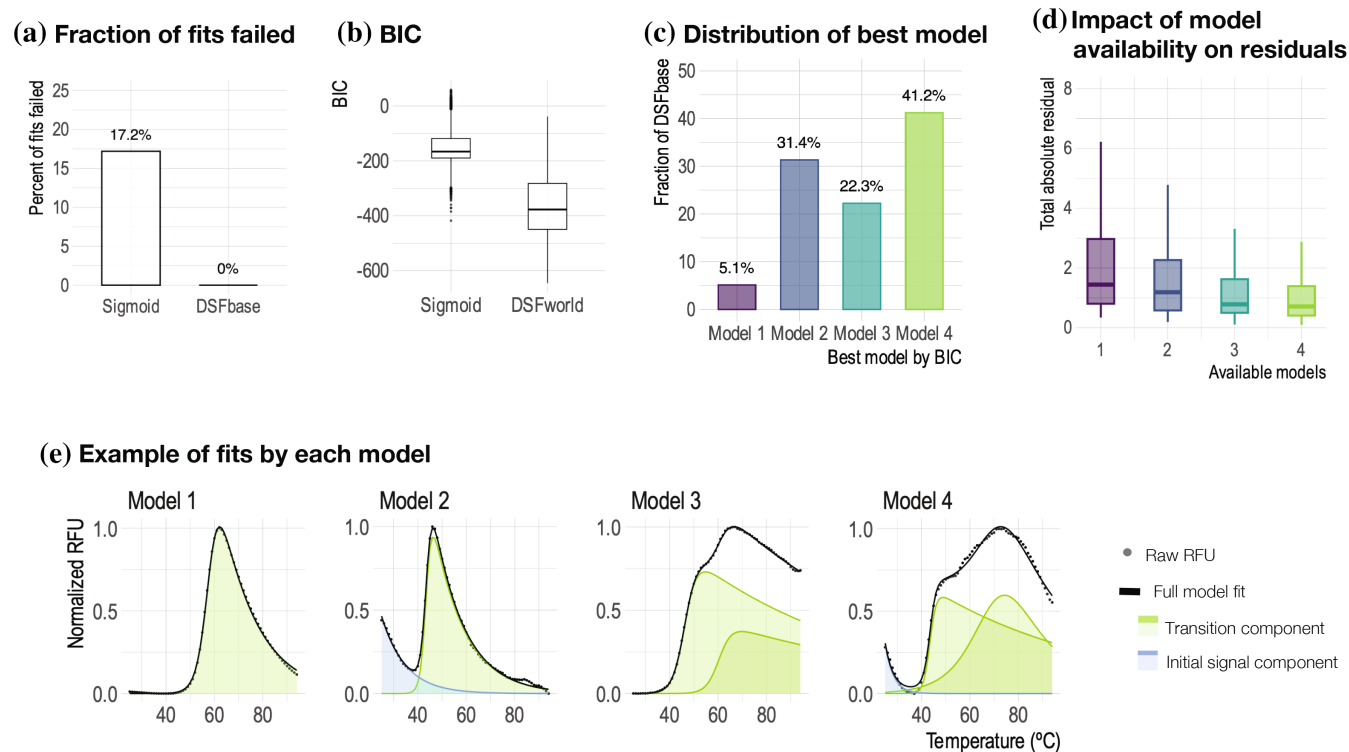


FIGURE 2 Performance of DSFworld models with DSFbase. (a) Fraction of results in the canonical and non-canonical sections of DSFbase that could not be fit with the traditional sigmoid approach vs. the four models in DSFworld. (b) Comparison of the BIC for all successful model fits using either the sigmoid model, or the best DSFworld model for each entry, as determined by BIC. (c) Fraction of the fitted datasets for which each DSFworld model was selected using BIC. (d) The total residual for final model fits of the best DSFworld model to each dataset, with an increasing number of DSFworld models available to choose between. (e) Representative examples of data best fit by each of the four models in DSFworld.

2.2 | Example 1: Model 2 improves screening data quality

One benefit of the DSFworld models is that they maintain fit quality even when the input raw data has normal fluctuations. For example, an increase in initial RFU is often observed upon the addition of small molecules or other additives (Wu, Hornsby, et al., 2023a; Wu, Hornsby, et al., 2023b). This is a major problem in the use of DSF for high throughput chemical screening, because each well can have a different amount of compound-induced fluorescence. Moreover, even stochastic variations in the initial fluorescence can lead to inaccurate hit-picking and poor statistical performance across many wells. Here, we show how Model 2 offers more robust T_{ma} calculation when this type of feature emerges. Specifically, we use a DSF dataset (Gahbauer et al., 2023) from a screen of 640 purine-like small molecules binding to the purified SARS-CoV2 protein: nsp3 macrodomain1 (nsp3 mac1). As described in Section 4, this screen was performed in 384-well plates using a single concentration (50 μ M) of each of the putative ligands, with some wells dedicated to

positive (ADP-ribose) and negative controls (DMSO alone).

From the raw screening data, T_{ma} values were calculated by four different methods (Figure 3a): (i) truncation of input data at the maximum value, followed by fitting to a standard sigmoid, (ii) maximum of the first derivative, (iii) DSFworld Model 1 (single transition with no initial fluorescence) and (iv) DSFworld Model 2 (single transition with initial fluorescence). The first two methods are pre-existing approaches to T_{ma} calculation, which allowed us to test their performance against those models introduced in DSFworld. Nsp3 mac1 produces canonical, single-transition DSF data, so if curve shape was perfectly conserved across all conditions, these four approaches to T_{ma} calculation could be used largely interchangeably. However, in practice, DSF curve shapes vary slightly between replicates, and often more dramatically between conditions, such as with the addition of a ligand. Even for the negative and positive controls (see Section 4), consisting of 32 wells of DMSO and 32 wells of 300 μ M ADP-ribose, use of the DSFworld Models 1 and 2 outperformed both sigmoid fitting and dRFU, as

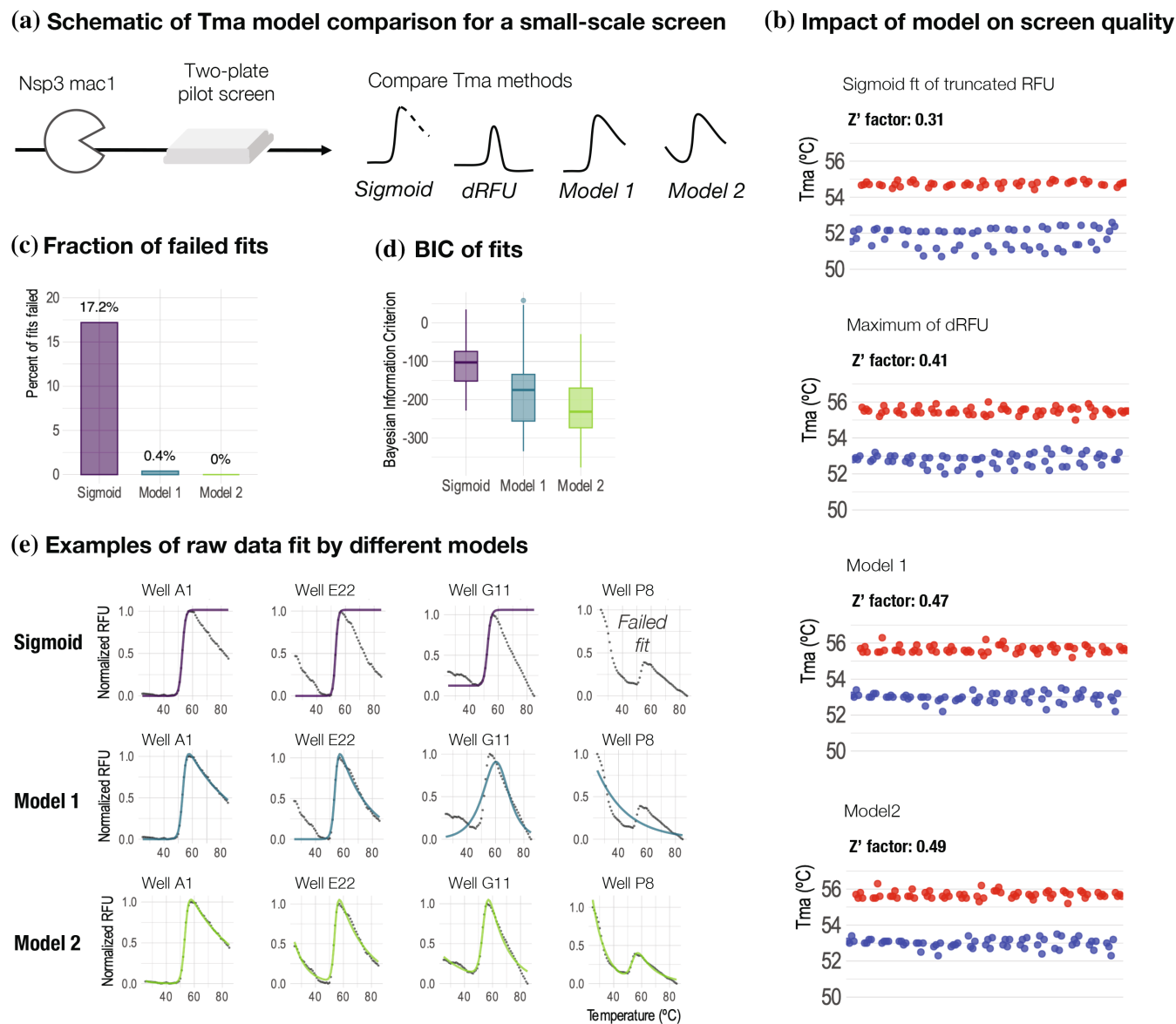


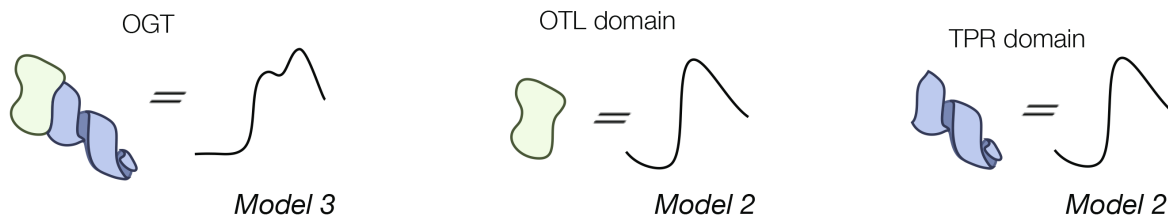
FIGURE 3 Use of Model 2 improves the quality of chemical screening data analysis. (a) Following a 640-compound pilot DSF pilot screen, Tmas were calculated using four different methods and compared: sigmoid fitting following raw data truncation, maximum of dRFU, and DSFworld Models 1 and 2. (b) Tmas for negative (blue points) and positive (red points) controls (DMSO, or 300 μ M ADP ribose), and associated Z-prime factors, for each of the four Tma calculation methods. Model 2 gives the best Z-prime score. Each point represents a single Tma value calculated by one of the four methods. (c) The fraction of input data for which each fitting model failed. (d) Of the successful fits for each model, the quality of the associated fits, by Bayesian information criterion (BIC). Lower values indicate better models. Box plot values: Sigmoid: lower whisker (LW)—228, lower hinge (LH)—152, median (M)—103, upper hinge (UH)—74, 35, upper whisker (UW)—35; Model 1: LW—334, LH—256, M—175, UH—134, UW—47; Model 2: LW—377, LH—273, M—231, UH—170, UW—29. (e) Examples of raw RFU data fit by each of the three fitting methods. Individual points: single RFU measurements. Solid lines: resulting fits from each of the three models. Sigmoid (purple, top); Model 1 (teal, middle), and Model 2 (green, bottom).

measured by Z prime factor, with DSFworld Model 2 offering the highest Z prime factor of 0.49 (Figures 3b and S1).

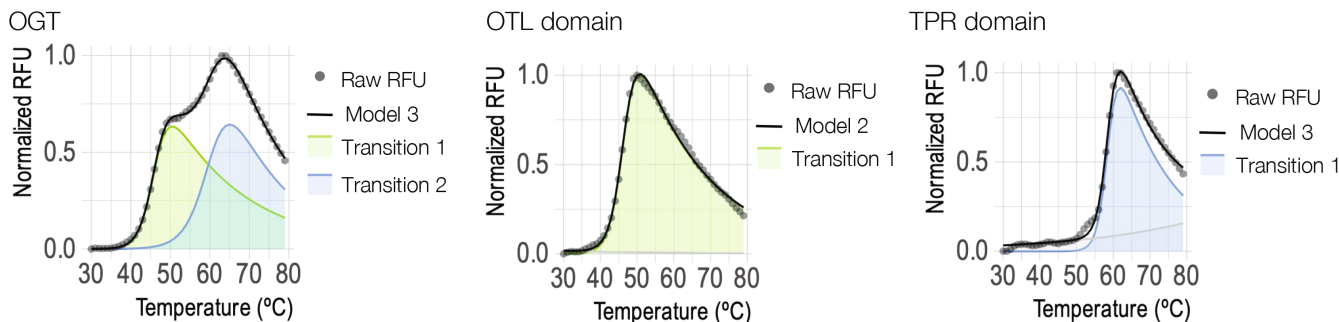
The improvement in performance from using Model 2 became more pronounced for the wells containing screening compounds. While the sigmoid fitting failed for

17.2% of these wells, Model 1 failed for only 0.4%, and Model 2 fitting succeeded for all datasets (Figure 3c). Tmas calculated using sigmoid fitting for the test compounds were more prone to report negative thermal shifts and outliers than dRFU, Model 1, or Model 2, with Model 2 reporting the least noisy Tma data between all methods

(a) Schematic of T_{ma} calculation using construct-appropriate models



(b) Example fits to double or single transition DSF data



(c) Fitting of multiple transitions enables comparison to isolated domains

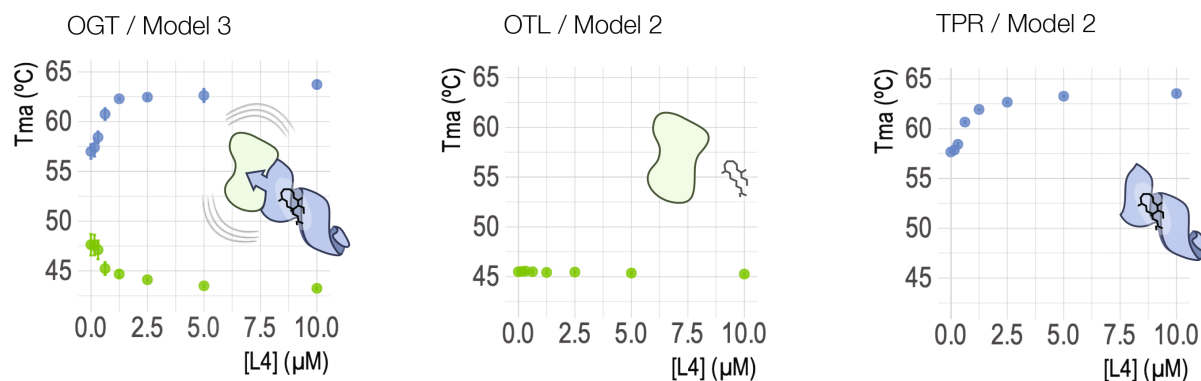


FIGURE 4 Use of Models 2 and 3 enable investigation of interdomain allostery of OGT. (a) Diagram of the models fit to each construct in the study. (b) Example raw data for the three constructs, OGT, OTL domain, and TPR domain. Points represent single RFU measurements, and lines represent single fit predictions corresponding to the displayed raw data. (c) Δ T_{mas} resulting from the addition of a TPR binding allosteric ligand. Use of Model 3 reveals simultaneous opposing shifts in the OTL- and TPR-associated transitions in the two-transition OGT curve. Use of Model 2 reveals a similar thermal shift in the TPR construct, and no thermal shift in the OTL construct alone. Points represent mean T_{mas} from three technical replicates, error bars display \pm standard deviation. Some error bars are smaller than the data points.

(Figure S2). Furthermore, model selection using BIC shows that Model 2 performs the best on the dataset as a whole (Figure 3d). Examples of four representative curves demonstrate how those with initial signal are fit more poorly by sigmoid fitting than Model 1 (Figure 3e). Moreover, these examples show how Model 2 most accurately fits wells with high initial fluorescence. This example, using real-world DSF data from a typical high throughput chemical screen, shows how DSFworld produces more robust analysis in practice.

2.3 | Example 2: Model 3 enables observation of interdomain allostery

Many proteins undergo multiple melting transitions, which complicates traditional DSF analyses, often obscuring or distorting T_{ma} values. For this reason, a major benefit of the DSFworld models is the ability to fit both single and double-transition data, allowing T_{mas} to be calculated from both types of curves without disrupting the analysis pipeline. We illustrate the importance of

fitting both single and double transitions here, using analyses of a previously reported DSF dataset (Wu, Yu, et al., 2023). Briefly, O-GlcNAc Transferase (OGT) is an enzyme that is composed of two domains, an N-terminal domain containing tetratricopeptide repeats (TPR domain), and a C-terminal catalytic domain (Figure 4a). There is evidence of inter-domain allostery in this system, such that binding of ligands, such as L4, to the TPR domain inhibits catalytic activity (Alteen et al., 2022). This allostery is often studied by comparing the full-length OGT to truncated domains: the isolated TPR domain and a construct that contains the catalytic domain and a subset of the TPR repeats (9–13.5), termed the OTL domain (Figure 4a).

Using DSFworld, Models 2 and 3 can be readily used to determine the Tma values of OGT and its two isolated domains. Specifically, Tmas can be calculated for both transitions of full-length OGT using Model 3 and single-transition Tmas can be determined for each domain using Model 2 (Figure 4b). OGT's first transition aligns with the Tma of the isolated OTL domain ($45.3 \pm 0.3^\circ\text{C}$ vs. $45.5 \pm 0.5^\circ\text{C}$, $p\text{-value} = 0.45$), while OGT's second transition aligns with the Tma of the isolated TPR domain ($57 \pm 1.2^\circ\text{C}$ vs. $57.7 \pm 0.1^\circ\text{C}$, $p\text{-value} = 0.45$). It should be noted that fitting OGT to a single-transition model, such as a traditional sigmoid or Model 2, returns poor quality fits and an aberrantly elevated, single Tma (Figure S3), exemplifying how use of an appropriate model is important for robust analysis.

As mentioned above, the L4 ligand binds only the TPR domain, but produces associated changes in the structure and activity of the catalytic domain through an allosteric mechanism (Alteen et al., 2022). DSFworld allows for the experimental exploration of this allostery. Specifically, treatment with L4 caused a thermal upshift which saturated at $5.8 \pm 0.4^\circ\text{C}$ in both the second transition of full-length OGT and the isolated TPR domain (Figure 4c, blue). In contrast, treatment with L4 did not impact the Tma of the isolated OTL domain ($\Delta\text{Tma} = 0.2 \pm 0.5^\circ\text{C}$), but it did cause a dose-dependent, thermal downshift in the first transition of full-length OGT ($\Delta\text{Tma} = -1.6 \pm 0.6^\circ\text{C}$) (Figure 4c, green). The ability to monitor allostery in OGT using a combination of single- and double-transition models illustrates how the ability to seamlessly accommodate a range of data complexities can enable studies previously inaccessible by DSF.

2.4 | Follow-along instructions

This section walks through the complete analysis of a DSF experiment using the DSFworld website (<https://gestwickilab.shinyapps.io/dsfworld/>). It is meant to demonstrate the workflow and key features of DSFworld and illustrate how DSFworld could fit into an existing

experimental workflow. To follow along with this section, the dataset can be downloaded from the “upload data” tab of DSFworld.

First, access DSFworld by navigating to its associated website <https://gestwickilab.shinyapps.io/dsfworld/>, and clicking “to data analysis” in the top bar. Begin the analysis by uploading raw fluorescence versus temperature data in the ‘Uploads’ tab (Figure 5a). To upload a dataset, click “Browse” and select the raw data file, or drag and drop the file into the “Uploads” bar. The data will appear as uploaded in a table to the right of the gray uploads bar. The sample dataset is already in the standard, readable format, with Temperature in the first column, and fluorescence data for each well in the following columns. Information on automated file reformatting is provided in the later section titled “Uploads details”.

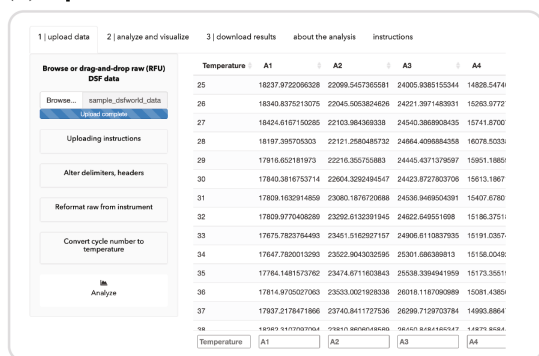
Click “Analyze” at the bottom of the gray uploads bar. The page will proceed to the plotting and Tma calculation window. Similar to the uploads screen, options for analyses appear in a gray sidebar on the left of the screen. A simple plot of the raw data will appear automatically on the right-hand side of the analysis tab.

It is possible to proceed directly to Tma calculation at this stage. However, it is often helpful to first visualize the raw data and assign experimental conditions to wells. Click “Set plate layout and replicates”, and then “Upload layout”. The creation and use of layouts, including their manual editing under ‘Method 2—edit manually’, is described in detail in the instructions tab of the DSFworld website. In this example, we use a sample plate layout which can be downloaded by clicking “Download example layout” in the gray sidebar. This layout defines three experimental variables: protein identity, compound, and compound concentration. Upload the layout by clicking “Browse...” or dragging and dropping the layout file directly into the layout uploads bar.

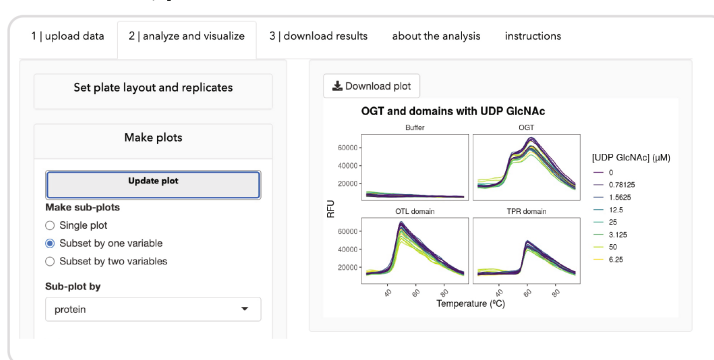
Minimize the plate layout options by clicking the “Set plate layout and replicates” heading. Below “Set plate layout and replicates, click “Make plots”. The experimental conditions uploaded in the previous step are now available for use in plotting (as well as later data analysis). In a standard plotting workflow, modify the plotting options to make an informative visualization, click “Update plot”, and the uploaded plot will appear at the right of the screen, where it can be downloaded by clicking “Download plot.” For example, here, click “Subset by one variable”, and select “protein” under “Sub-plot by”. Select “compound_concentration” for “Color” and vary line types by “protein”. Under “Edit plot labels”, update the Plot title to “Dose response by protein”, The legend title to “Compound conc. (μM)”, and the line type legend to “Protein” (Figure 5b).

Minimize the plotting options by clicking the “Make plots” heading. Below the “Make plots” heading, click

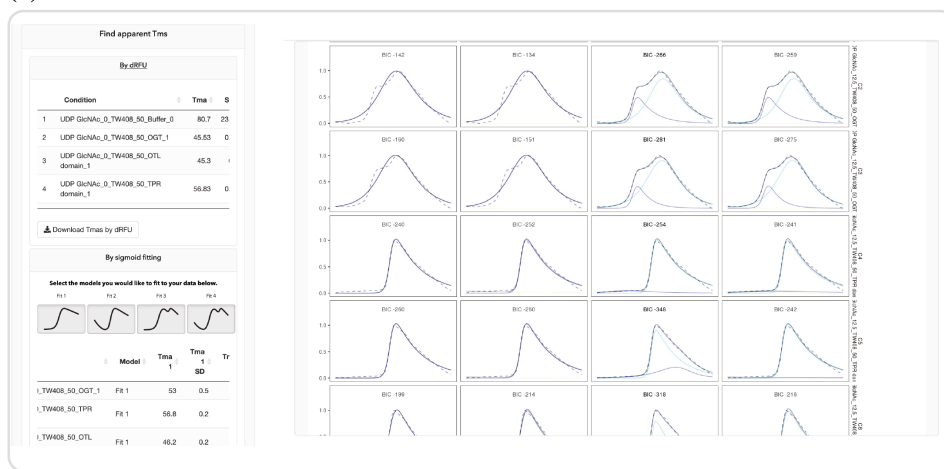
(a) Uploaded raw data



(b) Labeled, plotted data



(c) Calculated Tmas



(c) Results for download

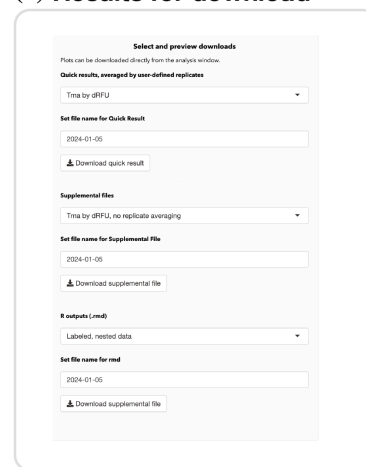


FIGURE 5 Screenshots of a typical analysis using the DSFworld website. (a) Uploaded data is displayed as formatted in a table to the right of the gray uploads options bar. (b) If an experimental layout is uploaded, raw data can be plotted interactively based on experimental variables defined in the layout. (c) Methods for Tma calculation, and resulting Tmas, appear in the gray sidebar. Fits can be plotted, and the best model for each curve can be manually updated by clicking the plot. (d) Both raw data and Tmas can be downloaded in various forms under the downloads tab.

“Find apparent Tmas”. Tmas by dRFU have already been calculated automatically. These Tmas are displayed in a table in the left sidebar and can be downloaded by clicking “Download Tmas by dRFU”. Because a layout was uploaded in this example, the Tmas are averaged by experimental replicates in this table, where replicates are defined as wells with identical values for all experimental variables. If needed, non-averaged values can be downloaded in the “3|download results” tab. In this example, the dataset includes common types of DSF data better analyzed by model fitting, namely curves containing either multiple transitions, or high initial fluorescence, or both (Figure 5c). Fit the results to all available models by clicking the associated buttons.

To visualize the resulting fits, and select the best model for each dataset, click “Display/Update fit plot.” On the right-hand side of the screen, a new plot is displayed, containing sub-plots where each row contains a different raw curve from the uploaded dataset, and each column contains fitting results from a different model.

The Bayesian Information Criterion (BIC) for each model is displayed at the top of each sub-plot, and the model with the lowest BIC is automatically selected for each dataset. Double-click a sub-plot to set that model as the chosen model for a given dataset. Display only the selected fits for each raw curve by clicking “Plot selected fits” in the gray analysis sidebar. Click “Download plot”. Download Tmas from fits by clicking “Download Tmas from fits”, below the table displaying Tmas.

At the top of the screen, click “3|download results” (Figure 5d). Under “Quick results, averaged by user-defined replicates”, select “Tma by best fit”, set a file name for the download in the text box immediately below, and then click “Download quick result.” It is often helpful to have a copy of non-replicate averaged Tmas and raw data. To get these files, under “Supplemental files” in the same gray sidebar of the downloads screen, select and download “Tma by best fit, no replicate averaging”, and then “Reformatted raw data.” If any additional DSFworld analyses or plots are desired at a later

date, this “reformatted raw data” can be uploaded and analyzed by DSFworld with no additional reformatting.

3 | CONCLUSIONS AND DISCUSSION

Here, we present DSFworld, a web-accessible software for the analysis and visualization of DSF data. The goal of DSFworld is to overcome the barriers imposed on DSF applications by the limited scope and flexibility of existing DSF data analysis methods. To achieve this goal, DSFworld introduces four updated models for fitting DSF data, accounting for common aspects of curve variation, such as high starting fluorescence and/or multiple transitions. To our knowledge, these models are the first to be designed from and tested on a large, diversified set of real-world DSF data, and the first to be compatible with a wide scope of possible DSF results. Finally, to address the labor-intensive manual processing steps that often plague DSF analysis workflows, DSFworld also supports data uploading, formatting, and visualization. DSFworld is accessible as an open-source web interface: <https://gestwickilab.shinyapps.io/dsfworld/>. It is our hope that DSFworld can offer a more reliable experience of DSF data analysis, in both data compatibility and ease of use.

4 | METHODS

4.1 | DSFbase data

The DSFbase dataset is described in a companion paper (Wu et al., n.d).

4.2 | Pilot screen of the Nsp3 macrodomain 1

Compound screens were conducted in final conditions of 2.5 μM nsp3 macrodomain 1, and either 50 μM test compound, 300 μM ADP-ribose (positive control), or 0 μM compound (negative control). The final buffer included 5 \times (10 μM) SYPRO Orange (Thermo Fisher Scientific, S6650) in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.01% Triton X-100 and 3% DMSO. Screens were performed using the following protocol: To 50 mL of screening buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, and 0.01% Triton X-100), 10 μL of SYPRO Orange (Thermo Fisher Scientific, S6650) was added to a final concentration of 5 \times (10 μM). From this buffer, a positive control solution was prepared as a 2 \times concentrate by bringing 60 μL of a 10 mM ADP ribose solution in DMSO to 1 mL

volume. For negative controls and test compounds, 10 mL of 3% DMSO buffer was prepared. Purified Nsp3 macrodomain1 (P43 construct (Schuller et al., 2021)) was prepared as a 2 \times concentrate by diluting 1054 μL of protein solution to 7 mL volume in DMSO-free buffer to a final concentration of 5 μM protein. Then, protein solution (5 μL) was added to each well of two 384-well white qPCR plates (Axygen, PCR-384-LC480WNFBC), followed by 5 μL of positive control solution in columns 23 and 24, or 5 μL of compound-free solution in columns 1–22. Compounds from two plates of a purine nucleoside-like library were added to columns 3–22 by pinning in the UCSF Small Molecule Discovery Center. In an Analytik Jena qTOWER 384G quantitative PCR instrument, each plate was continuously heated from 25 to 94°C at a rate of 1°C/min, and fluorescence was measured at each degree in the TAMRA channel (535/580 nm). The Z' factor was calculated by the standard formula: $Z' = 1 - 3(\sigma_{\text{pos}} + \sigma_{\text{neg}})/|\mu_{\text{pos}} - \mu_{\text{neg}}|$, where σ_{pos} and σ_{neg} refer to the standard deviation of the positive and negative controls respectively, while μ_{pos} and μ_{neg} refer to the means.

4.3 | Binding of ligand L4 to OGT

The experiments performed using OGT, OTL, and TPR with addition of L4 were previously reported and the full experimental conditions can be found there (Wu, Yu, et al., 2023). Statistical tests of difference between apo Tmas of OGT and the isolated domains use a Welch two sample *t*-test performed in R using the ‘t.test()’ function. Statistical tests of correlations between Tmas and L4 concentration use a Pearson’s correlation performed in R using the ‘cor.test()’ function, with method = “pearson”.

AUTHOR CONTRIBUTIONS

Taiasean Wu: Conceptualization; software; formal analysis; writing – original draft; writing – review and editing; data curation; methodology; visualization. **Zachary J. Gale-Day:** Software; writing – review and editing; methodology. **Jason E. Gestwicki:** Funding acquisition; writing – review and editing; project administration.

ACKNOWLEDGMENTS

The authors would like to acknowledge those who provided user feedback on early versions of DSFworld, especially Ziyang Zhang, Douglas Wassarman, Jack Stevenson, and Sarah Williams (UCSF). For productive scientific conversations that guided this work, we thank James Fraser, Matt Jacobson, Kangway Chuang, and Daniel Elnatan (UCSF). We thank David Voadlo and Matthew Alteen (Simon Fraser U.) for providing the OGT protein. The nsp3 mac1 work was conducted in collaboration with the Quantitative Bioscience Institute’s Coronavirus Research Group (QCRG),

especially the research groups of James Fraser, Brian Shoichet, Adam Renslo and Alan Ashworth (UCSF). This work was also supported by grants from the NIH GM141299 (to J.E.G.) and AR081704 (to Z.J.G-D.) and the NSF 1000259744 (to T.W.).

ORCID

Jason E. Gestwicki  <https://orcid.org/0000-0002-6125-3154>

REFERENCES

- Alteen MG, Peacock H, Meek RW, Busmann JA, Zhu S, Davies GJ, et al. Potent de novo macrocyclic peptides that inhibit O-GlcNAc transferase through an allosteric mechanism. *Angew Chem*. 2022;62:e202215671.
- Boivin S, Kozak S, Meijers R. Optimization of protein purification and characterization using thermofluor screens. *Protein Expr Purif*. 2013;91(2):192–206.
- Chari A, Haselbach D, Kirves J-M, Ohmer J, Paknia E, Fischer N, et al. ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nat Methods*. 2015;12(9):859–65.
- Gahbauer S, Correy GJ, Schuller M, Ferla MP, Doruk YU, Rachman M, et al. Structure-based inhibitor optimization for the Nsp3 macrodomain of SARS-CoV-2. *Proc Natl Acad Sci U S A*. 2023;120:e2212931120.
- Gao K, Oerlemans R, Groves MR. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys Rev*. 2020;12(1):85–104.
- Garlick JM, Mapp AK. Selective modulation of dynamic protein complexes. *Cell Chem Biol*. 2020;27(8):986–97.
- Huynh K, Partch CL. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protocols Protein Sci*. 2015;79:28.9.1–28.9.14.
- Lee P-H, Huang XX, Teh BT, Ng L-M. TSA-CRAFT: a free software for automatic and robust thermal shift assay data analysis. *SLAS Discovery Adv Life Sci R&D*. 2019;24(5):606–12.
- Martin-Malpartida P, Hausvik E, Underhaug J, Torner C, Martinez A, Macias MJ. HTSDSF explorer, a novel tool to analyze high-throughput DSF screenings. *J Mol Biol*. 2022;434(11):167372.
- Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen*. 2001;6(6):429–40.
- Phillips K, Hernandez de la Peña A. The combined use of the thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Curr Protocols Mol Biol*. 2011;94:Unit10.28. <https://doi.org/10.1002/0471142727.mb1028s94>
- Ravalin M, Theofilas P, Basu K, Opoku-Nsiah KA, Assimon VA, Medina-Cleghorn D, et al. Specificity for latent C termini links the E3 ubiquitin ligase CHIP to caspases. *Nat Chem Biol*. 2019;15(8):786–94.
- Reinhard L, Mayerhofer H, Geerlof A, Mueller-Dieckmann J, Weiss MS. Optimization of protein buffer cocktails using Thermofluor. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2013;69(Pt 2):209–14.
- Ristic M, Rosa N, Seabrook SA, Newman J. Formulation screening by differential scanning fluorimetry: how often does it work? *Acta Crystallogr. F Struct Biol Crystall Commun*. 2015;71(Pt 10):1359–64.
- Rosa N, Ristic M, Seabrook SA, Lovell D, Lucent D, Newman J. Meltdown: a tool to help in the interpretation of thermal melt curves acquired by differential scanning fluorimetry. *J Biomol Screen*. 2015;20(7):898–905.
- Schuller M, Correy GJ, Gahbauer S, Fearon D, Wu T, Diaz RE, et al. Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci Adv*. 2021;7(16):eabf8711. <https://doi.org/10.1126/sciadv.abf8711>
- Schulz MN, Landström J, Hubbard RE. MTSA—a Matlab program to fit thermal shift data. *Anal Biochem*. 2013;433(1):43–7.
- Scott DE, Spry C, Abell C. Differential scanning fluorimetry as part of a biophysical screening cascade. Fragment-based drug discovery lessons and outlook. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2016. p. 139–72.
- 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(1):119–28.
- Simeonov A. Recent developments in the use of differential scanning fluorimetry in protein and small molecule discovery and characterization. *Expert Opin Drug Discovery*. 2013;8(9):1071–82.
- Sun C, Li Y, Yates EA, Fernig DG. SimpleDSFviewer: a tool to analyze and view differential scanning fluorimetry data for characterizing protein thermal stability and interactions. *Protein Sci*. 2020;29(1):19–27.
- Wang CK, Weeratunga SK, Pacheco CM, Hofmann A. DMAN: a Java tool for analysis of multi-well differential scanning fluorimetry experiments. *Bioinformatics*. 2012;28(3):439–40.
- Wu T, Gestwicki JE. DSFbase: open source database of raw DSF data. *Protein Sci*. n.d.
- Wu T, Hornsby M, Zhu L, Yu JC, Shokat KM, Gestwicki JE. Protocol for performing and optimizing differential scanning fluorimetry experiments. *STAR Protocols*. 2023b;4(4):102688.
- Wu T, Joshua Y, Gale-Day Z, Woo A, Suresh A, Hornsby M, et al. Three essential resources to improve differential scanning fluorimetry (DSF) experiments. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.03.22.002543>
- Wu T, Yu JC, Suresh A, Gale-Day ZJ, Alteen MG, Woo AS, et al. Conformationally responsive dyes enable protein-adaptive differential scanning fluorimetry. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.01.23.525251>

SUPPORTING INFORMATION

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

How to cite this article: Wu T, Gale-Day ZJ, Gestwicki JE. DSFworld: A flexible and precise tool to analyze differential scanning fluorimetry data. *Protein Science*. 2024;33(6):e5022. <https://doi.org/10.1002/pro.5022>