**Journal of
Biomolecular Screening**

Meltdown – a tool to help in the interpretation of thermal melt curves acquired by Differential Scanning Fluorimetry

Journal:	*Journal of Biomolecular Screening*
Manuscript ID:	JBSC-14-0215.R1
Manuscript Type:	Application Note
Date Submitted by the Author:	n/a
Complete List of Authors:	Rosa, Nicholas; Commonwealth Scientific and Industrial Research Organisation (CSIRO), Manufacturing Flagship Ristic, Marko; Commonwealth Scientific and Industrial Research Organisation (CSIRO), Manufacturing Flagship Seabrook, Shane; CSIRO, Collaborative Crystallisation Centre Lovell, David; Queensland University of Technology, Science and Engineering Faculty Lucent, Del; Wilkes University, Engineering and Physics Newman, Janet; Commonwealth Scientific and Industrial Research Organisation (CSIRO), Manufacturing Flagship
Keywords:	Database and data management, Sample preparation, Statistical analyses, Automation or robotics
Abstract:	The output of a Differential Scanning Fluorimetry (DSF) assay is a series of melt curves, which need to be interpreted in order to get value from the assay. An application that translates raw thermal melt curve data into more easily assimilated knowledge is described. This program, called 'Meltdown', performs four main activities: control checks, curve normalization, outlier rejection, and melt temperature (Tm) estimation, and performs optimally in the presence of triplicate (or higher) sample data. The final output is a report that summarizes the results of a DSF experiment. The goal of Meltdown is not to replace human analysis of the raw fluorescence data, but to provide a meaningful and comprehensive interpretation of the data to make this useful experimental technique accessible to inexperienced users, as well as providing a starting point for detailed analyses by more experienced users.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.
supplementary info-2015.7z

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

For Peer Review

Meltdown – a tool to help in the interpretation of thermal melt curves acquired by Differential Scanning Fluorimetry

Nicholas Rosa[1]

Marko Ristic[1]

Shane A. Seabrook[1]

David Lovell[2]

Del Lucent[3]

Janet Newman[1]*

The first two authors contributed equally to this work.

1. Manufacturing Flagship, CSIRO, 343 Royal Parade, Parkville, VIC, 3052, Australia

2. Digital Productivity Flagship, CSIRO, North Road, Acton, ACT, 2601, Australia

3. Department of Physics, Wilkes University, 141 S. Main St, Wilkes-Barre, PA, 18766, United States

* Corresponding author email: janet.newman@csiro.au

Abstract

The output of a Differential Scanning Fluorimetry (DSF) assay is a series of melt curves, which need to be interpreted in order to get value from the assay. An application that translates raw thermal melt curve data into more easily assimilated knowledge is described. This program, called 'Meltdown', performs four main activities: control checks, curve normalization, outlier rejection, and melt temperature (T_m) estimation, and performs optimally in the presence of triplicate (or higher) sample data. The final output is a report that summarizes the results of a DSF experiment. The goal of Meltdown is not to replace human analysis of the raw fluorescence data, but to provide a meaningful and comprehensive interpretation of the data to make this useful experimental technique accessible to inexperienced users, as well as providing a starting point for detailed analyses by more experienced users.

Introduction

Thermal Shift Analyses may be performed in many ways; one of the more popular techniques uses a Real-Time Polymerase Chain Reaction (RT-PCR) machine and this version (known as Thermofluor[1] or more generally as Differential Scanning Fluorimetry[2] (DSF)) is becoming widely used in structural biology and biophysics, driven in part by its simplicity, low cost and the wide availability of the hardware used in the assay. Initially, these miniaturized thermal analyses were used to investigate ligand binding to a protein target[1,3], but DSF has been adopted as a general method to assess relative protein stability[4–6].

In the DSF experiment, an RT-PCR machine is used to measure the fluorescence of an array containing different protein/fluorescent dye mixtures as they are heated. Although a dye-free DSF system has been described[7], most DSF experiments use an exogenous dye. There are a number of dyes that can be used[8], but all have the property of preferring a hydrophobic to a hydrophilic environment; thus these dyes bind to the hydrophobic core of an unfolding protein in aqueous solution. Furthermore, the fluorescence from these dyes is quenched in an aqueous environment. Many DSF experiments start at moderate temperature (around room temperature, or 20°C), and measure the fluorescence response from this point up to high temperature: 80°C or more. At the beginning of an ideal experiment, the protein is well folded and there is

little interaction between the dye and the protein; the dye is in the bulk solution and its fluorescence is quenched in this aqueous environment. As the dye/protein mixture is heated, the protein unfolds, exposing its hydrophobic core to which dye binds. The dye bound to the hydrophobic environment of the unfolding protein fluoresces. As the protein continues to unfold the fluorescence from bound dye increases. At some point the unfolded protein chains aggregate, excluding dye in the process. The excluded dye is returned to the surrounding aqueous environment, or is simply quenched at higher temperatures, and the fluorescence signal decreases. The temperature *vs* fluorescence plot from this ideal experiment has a flat pre-transition region, a steep sigmoidal unfolding region and an aggregation region - Figure 1-A. The value most often reported from a DSF experiment is the temperature of the midpoint of the sigmoidal transition region of the fluorescence curve, which is the temperature of hydrophobic exposure – usually reported as the melt temperature (T_m)[9]. Other attributes of the curves also contain information – the steepness of the sigmoidal transition, for example, or the flatness of the pre-transition baseline.

The individual trials in a DSF experiment may be used to probe the stability of a protein under different conditions – in the presence of small molecules[2,10] or under different pH or salt conditions[11,12]; a higher T_m is associated with increased stability. The raw fluorescence *vs.* temperature plots obtained from RT-PCR machines need to be interpreted to extract the stability information, and there are tools to help in the interpretation of the raw curves. Earlier applications tended to be spreadsheet based, and required a significant effort to use[10]. More recently developed tools such as DMAN[13], MTSA[14] and ThermoQ[15] overcome many of the difficulties with the spreadsheet analysis tools, but these are general applications for displaying curves and extracting specific parameters, rather than offering an interpretation for the overall experiment. In general, the currently available tools aid the experienced user, rather than helping a less experienced user to interpret a DSF result. The knowledge that is required to interpret a bank of DSF experiments includes (a) an understanding that some curves are outliers or are otherwise unlikely to give sensible results; (b) recognizing when a curve might be an outlier and (c) recognizing that some curve shapes do not provide any information, or may provide ambiguous results. In earlier work, we showed that replication and the inclusion of suitable controls is essential for reasonable interpretation of a DSF experiment[16]. Building on that, we set out to build a robust, extensible analysis tool that would be easily accessible, and which would use

any replication and some basic knowledge of the experiment to create a report to help pilot both less and more experienced users through a DSF experiment.

For structural biology, in particular crystallography, the production of protein and the production of crystals from the protein are limiting - as both steps are unreliable and expensive; DSF experiments may ameliorate some of the limitations of these steps[11,17]. Because of the importance of the formulation of protein used in all biophysical analyses, we implemented a standard buffer screen[16] (Buffer Screen 9, BS9) as part of the suite of offerings through the Collaborative Crystallisation Centre (C3, http://crystal.csiro.au). BS9 is the ninth iteration of our in-house formulation screen and it captures our experience that controls, replication and consistent layout are all critical for reliable downstream interpretability. The Meltdown analysis tool was initially developed for the interpretation and display of BS9 results, but has been extended and is appropriate for analyzing many DSF experiments, particularly ones where replication has been used.

Materials and Methods

Input into Meltdown

The machines that can generate the fluorescence readouts which are the basis for a Meltdown analysis tend to be plate-based devices, and thus produce 96 or 384 fluorescence curves simultaneously; however Meltdown has no inbuilt limitations to the number of curves that can be analyzed at once.

Meltdown requires two input files: a text file containing the raw fluorescence values and a text file containing the contents of each well (the 'contents map' file). The contents map is used to group replicates within a DSF experiment, and to provide information about how the results of the experiment should be presented. In C3, the export option of the BioRad CFX Manager analysis package (Version 3.1 or above) is used to obtain the raw data as a text file (*i.e.* tab separated .txt file). Only one of the exported files is used in the analysis – the file that contains the fluorescence reading for each well at each temperature point (the BioRad CFX manager software calls this file "*Runname* - Melt Curve RFU Results_FRET.txt"). This file is arranged so that each row is a temperature point, and each column is a well (or a position in the experimental plate). Text files from

other systems can be used, but must follow the same arrangement of rows and columns. The first column gives the temperature at which the fluorescence was measured - this column must have exactly the string "Temperature" (no quotes) as its header. The first row gives the positional identifiers of the plate well (except the first column, which must contain the string "Temperature", as described above). Meltdown has no intrinsic limits on either the number of wells that can be analyzed or on the temperature step size or range. The contents map file is also a tab delimited file; an example is provided along with the Meltdown code. The order of the information in the contents map file is unimportant; however the header information may not vary. The contents map file contains columns including a positional identifier and a number of other columns for content description. The contents map file also contains a column that allows a user to enter a buffer temperature dependence term, which is included in the Meltdown analysis if provided. Replicates are identified by having the same string in the first 'Condition Variable' column of the contents map file. Standard controls 'Lysozyme', 'No Dye', 'No Protein', 'Protein as supplied' are recognized by Meltdown, but other controls may also be defined in the contents map file. Table 1 gives a more extensive description of the contents map file structure.

1.1. Meltdown Analysis

The Meltdown analysis of an experiment considers replicates of the experimental trials, and applies analytical techniques to determine curve outliers and curves unsuitable for T_m calculation. The analysis is conducted as follows:

1. All curves are normalized such that the area under the melt curve integrates to unity, and the factor used in normalization is retained for later use.

2. Curves are identified as being above background by comparing the normalization constant of the curve with that of the "no protein" curves, if available. Curves are considered to have a signal above background if their normalization factor is at least 15% greater than the average normalization factor found in the "no protein" controls of the same run. Each curve is also checked for saturation – curves that have a flat top are identified by finding the temperature which gives the greatest fluorescence response – curves which have that same fluorescence value for ten or more consecutive temperature steps are considered to be overloaded. Curves are only considered valid and used for further analysis if they fulfil both the requirement for being above background and are not overloaded.

3. Outliers amongst a replicate set are located through a graph-based analysis. The Aitchison distance[18] between each pair of the normalized replicates (that is, the sum of differences squared of the natural logarithm) is used to generate a full connected graph where each node is a normalized replicate and each edge is weighted by the distance: if an edge distance is above an experimentally derived threshold, that edge in the graph is removed.

4. After processing all distances, a replicate is retained if it belongs to the largest connected component of the resulting graph. Non-retained curves are the discarded outliers. If this process returns two equally large connected components, the one with a smaller sum of distances is selected.

The threshold cutoff (point 3 above) was derived from the spread obtained when 168 lysozyme curves (the lysozyme positive control curves from 56 different runs of BS9 performed in C3) were normalized and overlaid in the same manner, Figure 1-B.

After selection of valid curves and removing outliers (Figure 1-C) the remaining replicates are tested for monotonicity and used to estimate a melt temperature (T_m):

5. If the differences between each point of a moving window of five consecutive points in a melt curve are all negative, the replicate is considered monotonic and is not further analysed, Figure 1-D. This analysis is made more robust by softening the requirement for negative decrements by a "noise factor" derived from the normalization constant. Furthermore, within the series of consecutive points, a single point may show a positive difference; however this invokes a penalty and requires that the string of consecutive points be longer to fulfil the requirement of monotonicity.

6. The negative first derivative of each remaining (i.e. valid, non-monotonic, non-outlier) melt curve is calculated, and used to estimate if there are single or multiple transitions. If multiple minima are found, the melt curve is considered "complex", which is the term we use for curves which do not adopt the canonical melt profile shown in Figure 1-A.

7. The T_m of the selected curves is estimated in two ways – first by using a quadratic fit to the data around the global minimum of the first derivative curve (this value is used as the T_m in subsequent analyses), and second by finding the temperature associated with the midpoint in the fluorescence response between the high point and the low point of the melt curve. If the melt temperatures estimated by the two different methods differ by more than 5°C then the curve is

considered "complex". The low point of the graph is the lowest point found starting from the left hand side of the graph, and the high point is the highest point of the melt curve after the low point.

8. Curves where the minimum of the inverse first derivative is within a small (empirically derived) distance from 0 are considered to be "shallow".

9. The T_m values of the individual melt curves within a replicate set are used to find an average T_m and an estimate of the variation in the T_m. If only one curve of a replicate set remains, no estimate of the variation is made. If a buffer pH temperature dependence value is given in the contents map file, then the estimated pH at the T_m is also provided.

10. A report is generated which presents an overall summary graph of the T_m *vs* content (from the content map), and an estimate of the robustness of the overall analysis. This estimation considers the number of curves that could be used to generate a T_m as a percentage of the total number of experimental (non-control) melt curves, the average estimation of error in the T_m for all replicate sets and the unfolding behaviour of the protein in its original formulation if identified in the contents map file. Along with the summary graph, the superposed normalized curves from the 'Protein as supplied' wells are shown, Figure 2. The T_m values which are potentially less reliable - that is, derived from a complex curve, a shallow curve and/or from a single melt curve - are shown on the summary graph with a diamond shaped symbol, rather than the default solid circle symbol.

11. Any curves identified by Meltdown as belonging to controls – either through the 'Control' tag in the contents map file or by one of the standard names for controls: 'Lysozyme'; 'No Dye'; 'No Protein'; 'Protein as supplied' are not displayed on the summary graph, but controls that fail – for example, don't superpose well - are noted on the front page of the report.

12. The number of values along the x-axis of the summary graph is determined from the contents map file. In the simplest case, identical experimental replicate curves are grouped, and each group would be labelled separately on the summary graph, Figure 3. Each member of a group is identified by having the same string in the 'Condition Variable 1' column and the same value in the pH column of the contents map file. A second layer of differentiation may be used, this is defined in the 'Condition Variable 2' column; up to 24 unique secondary identifiers may be used. The order of the values along the X-axis is defined by

the 'pH' column of the contents map file, with the lowest values on the left and the highest values on the right of the graph.

13. The normalized melt curves for each of the groups are shown in separate detail plots. Curves distinguished by the secondary differentiator within a group are drawn in different colors. There is one detail plot for each of the X-axis strings in the summary graph (Figure 1-C,D).

All the calculations are performed using SciPy / Pandas, as implemented in the Anaconda Python distribution (http://continuum.io),[19,20] and the final pdf report is generated using ReportLab (www.reportlab.com). All of the software used in Meltdown is freely available.

2. Results and Discussion:

2.1. Normalization and outliers:

The absolute fluorescence response seen in a single protein concentration DSF experiment is generally not quantifiable[21]; with the exception of step 2 above, the Meltdown analysis assumes the curves in the DSF experiment contain only relative information and thus can be normalized to integrate to unity. The normalization process allows direct comparison of the replicate curves and is an essential pre-requisite for outlier rejection.

The DSF experiments are generally robust, but some curves are outliers, Figure 1-C. The outliers are likely the result of mis-dispensing during experimental setup, or could be an indication that a particular combination of protein/pH/salt/buffer is inappropriate for this analysis – for instance, the protein precipitates under the starting conditions. The threshold for outlier rejection in this study was set on the basis of the 168 individual lysozyme melt curves from 56 different BS9 experiments, Figure 1-B. These curves were normalized and for each pairwise combination of lysozyme curves, the sum of squared differences between every point was calculated. The mean and standard deviation of these distances were determined. The threshold for outlier rejection in the experimental curves has been set to the mean distance between two curves in this lysozyme curve set (5×10^{-7} normalized RFU units). If the Aitchison distance between two curves of a replicate set falls outside this threshold the measurement is considered unreliable and the curve is tagged as an outlier and is not included in the subsequent

analyses.   Both Euclidian and Aitchison distance were tested (over 56 BS9 experiments) for outlier rejection, and the results were similar, but the Aichison distance resulted in a set of outliers that was slightly larger - and a superset - of the set produced using a Euclidian distance. Although we have only included lysozyme data collected on a single RT-PCR machine in our basis set for the rejection threshold, we assume that the normalization of the curves makes the threshold value appropriate for curves collected from other machines as well.

The threshold for outlier selection is necessarily somewhat arbitrary, and by this criterion rejected outliers made up 10.2% of the experimental curves in the 56 exemplar BS9 experiments.

2.2. T_m estimation and curve shape

Melting temperature, although not the only measurable outcome of a DSF experiment, is the most widely reported. Although there are different approaches to obtaining T_m[14], most find the region of the fluorescence curve between the global maximum and minimum, assume that this region displays a sigmoidal transition and either perform a least squares fit to the truncated region (e.g. MTSA) or take the maximum of the first derivative (e.g. DMAN). The two approaches generally give a slightly different T_m, although the variation is probably not important for comparative analyses.

Meltdown uses both of these approaches, but uses both only to gauge the reliability of the T_m estimate. The T_m reported in the Meltdown report is derived from the minimum of the negative first derivative curve. In the case of more complex curves this simple approach will return the global minimum of the derivative curve, but Meltdown tags these curves to let the user know that the assumption of a single sigmoidal transition is not valid (Figure 1-D). A second test for T_m reliability is the comparison of the T_m values calculated in the two different ways. The cutoff for the ΔT_m ($5°C$ or greater) was derived empirically from inspection of some pathological examples.

Previous studies have indicated that real DSF curves do not necessarily adopt the canonical melt curve shape, and we note that some of the less orthodox curves – for example, those where the highest fluorescence response is seen at the start of the melt curve and the lowest response is at the end – are currently not robust in the Meltdown analyses. Earlier studies classified curves into three or more classes[5,22], but performed the classification by eye. Although this approach is certainly useful, a programmatic or statistical method would allow automated and reproducible classification of ambiguous

curves in high throughput experiments. The interpretation of different curve classifications is not completely clear. Although a common interpretation for a monotonically decreasing curve with no obvious unfolding transition is that the protein is unfolded from the outset of the experiment, other work has suggested that curves with no clear unfolding transition may be the result of a protein having a limited hydrophobic core which limits dye binding on thermal challenge[22].

2.3. pH and temperature

Although it is recognized that the pK_a of a buffering chemical can change with temperature[23], this does not seem to be taken into consideration in many of the studies done on the formulation of the protein solution for protein stability studies. As a way of bringing this to the attention of the user, we include field for temperature correction in the contents map. Although the absolute pH at which a transition occurs is probably of little consequence when using this method to select an appropriate formulation, it may be very important when DSF is used to measure ligand binding, as the charge of the binding site might change according to the pH of formulation used in the assay. The measurement of the variation in pH with temperature is complex, as the response of pH probes varies with both temperature and pH and these effects have to be teased out from the fundamental changes in buffer pKa with temperature[24]. A guide to the values that might be used for the temperature dependence of some different buffering chemicals is given in the supplementary information.

2.4. Meltdown Report

The Meltdown analyses are presented in a pdf (portable document format) report, which is arranged so that the "high information content" summaries come first. In the ideal DSF experiment – that is, one that includes all four types of control, and contains replication - the front page of the Meltdown report includes what system gives the 'best' (highest) T_m, a summary graph of the experimental replicates, and estimates of the reliability of the whole experiment (Figure 3). The summary graph presents the average T_m values from the experimental curves, and includes a reference line which is the T_m determined for the protein baseline control "protein as supplied". Robust T_m estimations are plotted as solid circles and less reliable T_m estimations are plotted with diamond

shapes. If any of the control replicates (lysozyme only, dye only, protein only) gave aberrant results this is presented in the report on the front page. This allows a user to rapidly see if any of the tested formulations give a stability increase relative to the current formulation, and gives an immediate indication of how much faith one should have in the results of the experiment.

The remaining pages of the report show individual detail plots, each showing a group of normalized curves. The groups are assigned by string matching in the 'contents map' input file; the number of plots will depend on how many groups of replicates are defined in the 'contents map' file. In the case of BS9, there are 14 groups defined by the 'Condition Variable 1' column in the 'contents map' file, and thus there are 14 individual detail plots. In the BS9 experiment, the 'Condition Variable 2' column has either low (50 mM) or high (200 mM) salt, so there are two subsets of curves (distinguished by color) in each of the 14 details plots. Meltdown groups melt curves into three sets: those for which no attempt is made to obtain T_m values; those for which T_m values may be unreliable and those which give a robust T_m estimation. Invalid, monotonic and outlier curves are found in the first set, and are drawn in the detail plots as dashed lines. Complex and shallow curves may give unreliable T_m values, these are plotted as dotted lines. All remaining curves are considered robust, and are drawn in solid lines, Figure 1-C 1-D. A number of different sample reports are provided (along with the raw data and matching 'contents map' files) in the supplementary information.

There were two major considerations in designing the report format – the report must show a useful summary of the experiments, yet present the data in a manner that discourages facile over-interpretation of the data. To this end, the graph presenting an overview of the results is shown after a summary that shows two things: the replicate melt curves of the protein as supplied, and a box which gives overall statistics. The statistics box starts with the caution that "Full interpretation requires you to look at the individual melt curves". This is followed by a brief 'reality check' – how many of the experimental melt curves were used to estimate a T_m value; the average estimation of error for the plate; and a summary of the "protein as supplied" curves. If any of these three checks fail to meet minimum standards, then a further cautionary statement is printed: "The summary graph below appears to be unreliable". By implementing these cues it is hoped that the investigator will then take the time to investigate the individual graphs that follow the summary, and make a cautious decision as to whether any information can be inferred from the data.

3. Conclusion:

We have written a Python program to help summarize and interpret DSF experiments, particularly those which are run with replication and controls. This program, Meltdown, requires a text file containing raw fluorescence data, along with a file describing of the contents of each well. The program locates the control curves and replicate experiments as well as finding and rejecting curves which are inappropriate for use in T_m estimation. A simple quadratic fit to the global minimum of the inverse first derivative curve is used to estimate T_m, and the results – the best experimental system (by T_m), and a summary overview are presented as a pdf report. The code, instructions for installation and use and some sample data are freely available *via* GitHub (https://github.com/C3-CSIRO/Meltdown).

Acknowledgements

References

(1)   Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; et al. High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screen.* **2001**, *6*, 429–440.
(2)   Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.; Fedorov, O. Y.; Finerty, P. J.; Wasney, G. A.; Yeung, R.; Arrowsmith, C.; Ball, L. J.; Berglund, H.; et al. Chemical Screening Methods to Identify Ligands That Promote Protein Stability, Protein Crystallization, and Structure Determination. *Proc. Natl. Acad. Sci.* **2006**, *103*, 15835.
(3)   McMahon, R. M.; Scanlon, M. J.; Martin, J. L. Interrogating Fragments Using a Protein Thermal Shift Assay. *Aust. J. Chem.* **2013**, *66*, 1502.

(4)   Lavinder, J. J.; Hari, S. B.; Sullivan, B. J.; Magliery, T. J. High-Throughput Thermal Scanning: A General, Rapid Dye-Binding Thermal Shift Screen for Protein Engineering. *J. Am. Chem. Soc.* **2009**, *131*, 3794–3795.

(5)   Crowther, G. J.; He, P.; Rodenbough, P. P.; Thomas, A. P.; Kovzun, K. V.; Leibly, D. J.; Bhandari, J.; Castaneda, L. J.; Hol, W. G. J.; Gelb, M. H.; et al. Use of Thermal Melt Curves to Assess the Quality of Enzyme Preparations. *Anal. Biochem.* **2010**, *399*, 268–275.

(6)   Layton, C. J.; Hellinga, H. W. Thermodynamic Analysis of Ligand-Induced Changes in Protein Thermal Unfolding Applied to High-Throughput Determination of Ligand Affinities with Extrinsic Fluorescent Dyes. *Biochemistry (Mosc.)* **2010**, *49*, 10831–10841.

(7)   Lea, W. A.; Simeonov, A. Differential Scanning Fluorometry Signatures as Indicators of Enzyme Inhibitor Mode of Action: Case Study of Glutathione S-Transferase. *PLoS ONE* **2012**, *7*, e36219.

(8)   Senisterra, G.; Chau, I.; Vedadi, M. Thermal Denaturation Assays in Chemical Biology. *ASSAY Drug Dev. Technol.* **2012**, *10*, 128–136.

(9)   He, F.; Hogan, S.; Latypov, R. F.; Narhi, L. O.; Razinkov, V. I. High Throughput Thermostability Screening of Monoclonal Antibody Formulations. *J. Pharm. Sci.* **2009**, n/a–n/a.

(10)   Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2*, 2212–2221.

(11)   Ericsson, U. B.; Hallberg, B. M.; DeTitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-Based High-Throughput Stability Optimization of Proteins for Structural Studies. *Anal. Biochem.* **2006**, *357*, 289–298.

(12)   Crowther, G. J.; Napuli, A. J.; Thomas, A. P.; Chung, D. J.; Kovzun, K. V.; Leibly, D. J.; Castaneda, L. J.; Bhandari, J.; Damman, C. J.; Hui, R.; et al. Buffer Optimization of Thermal Melt Assays of Plasmodium Proteins for Detection of Small-Molecule Ligands. *J. Biomol. Screen.* **2009**, *14*, 700.

(13)   Wang, C. K.; Weeratunga, S. K.; Pacheco, C. M.; Hofmann, A. DMAN: A Java Tool for Analysis of Multi-Well Differential Scanning Fluorimetry Experiments. *Bioinformatics* **2012**, *28*, 439–440.

(14)   Schulz, M. N.; Landström, J.; Hubbard, R. E. MTSA—A Matlab Program to Fit Thermal Shift Data. *Anal. Biochem.* **2013**, *433*, 43–47.

(15)   Phillips, K.; de la Peña, A. H. 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. In *Current Protocols in Molecular Biology*; Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011.

(16)   Seabrook, S. A.; Newman, J. High-Throughput Thermal Scanning for Protein Stability: Making a Good Technique More Robust. *ACS Comb. Sci.* **2013**, *15*, 387–392.

(17)   Boivin, S.; Kozak, S.; Meijers, R. Optimization of Protein Purification and Characterization Using Thermofluor Screens. *Protein Expr. Purif.* **2013**, *91*, 192–206.

(18)   Lovell, D.; Müller, W.; Taylor, J.; Zwart, A.; Helliwell, C. Proportions, Percentages, PPM: Do the Molecular Biosciences Treat Compositional Data Right? In *Compositional Data Analysis: Theory and Applications*; Pawlowsky-Glahn, V.; Buccianti, A., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2011; pp. 191–207.

(19)   Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20.

(20)   McKinney, W. PROC. OF THE 9th PYTHON IN SCIENCE CONF. (SCIPY 2010) http://conference.scipy.org/proceedings/scipy2010/mckinney.html (accessed Sep 26, 2014).

(21)   Seo, D.-H.; Jung, J.-H.; Kim, H.-Y.; Park, C.-S. Direct and Simple Detection of Recombinant Proteins from Cell Lysates Using Differential Scanning Fluorimetry. *Anal. Biochem.* **2014**, *444*, 75–80.

(22)   Dupeux, F.; Röwer, M.; Seroul, G.; Blot, D.; Márquez, J. A. A Thermal Stability Assay Can Help to Estimate the Crystallization Likelihood of Biological Samples. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 915–919.

(23)   Beynon, R. J. *Buffer Solutions*; The basics; IRL Press at Oxford University Press: Oxford ; New York, 1996.

(24)   Goodng, C. C.; Swenson, C. A.; Bull, H. B. A Mathematical Method for Temperature Compensation of pH Measurements. *Anal. Biochem.* **1975**, *67*, 220–225.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1.

Structure and contents of the contents map file

Column Header	Description
Well	Positional identifier, usually between 'A1' and 'H12' (for a 96 well plate). These must match the strings found in the header row of the text file containing the raw fluorescence values. Thus if the positional identifier in the fluorescence file is "A01" then the positional identifier in the contents map file must also be "A01" (rather than "A1", for example).
Condition Variable 1	This is the condition variable that is used to group replicate wells. There is no limit to the number of unique entries, up to the total number of wells in the experiment. Examples might be "50 mM sodium acetate" or "ligand one". Grouping is done on string matching, so case and white space must be identical for Meltdown to recognize these as being the same. The grouping defined by this column dictates how many values are shown in the summary graph, and how many 'detail' plots are drawn. If the strings 'Lysozyme', 'No Dye', 'No Protein', 'Protein as supplied' are included in this column, Meltdown recognizes these as control curves, and does not draw them on the summary graph.
Condition Variable 2	This allows wells that have the same Condition Variable 1 to be distinguished. A maximum of 24 unique values can be entered. Each set of curves defined as having the same 'Condition Variable 1' and 'Condition Variable 2' strings are considered replicates.
pH	This is the pH of the well, and is used in conjuction with 'Condition Variable 1' to uniquely identify the primary replicate sets. However, one can leave this blank and use pH as either 'Condition Variable 1' or 'Condition Variable 2'
d(pH)/dT	Entering a value here will direct Meltdown to calculate and display an adjusted pH value on the 'detail' graphs. The adjusted pH value is the calculated pH at the melt temperature, given the initial buffer pH (as given in the pH column) and assumes a linear pH / temperature dependence.
Control	This is used to distinguish which wells are controls and thus should not be used in the Meltdown analysis. Control wells are tagged either by having '1' in this column or by having the appropriate strings in the 'Condition Variable 1' column.

Figure 1

A)   A cartoon overview of the (ideal) DSF experiment – initially protein and dye are combined at room temperature in an aqueous environment. Under these initial conditions, the protein is folded, and dye is partitioned into the aqueous medium, where its fluorescence is quenched. On heating, the protein molecules begin to unfold, exposing a hydrophobic region to which the dye binds preferentially, allowing fluorescence to be measured. As the protein continues to unfold, more dye binds and more signal is seen. When the protein is completely unfolded, it is believed to aggregate, masking the hydrophobic regions, and excluding the dye, which is again is quenched. The red dotted curve shows the negative first derivative of the melt curve, the minimum of this derivative curve is often used as the estimation of the melt temperature (T_m).

B)   Shows 168 lysozyme DSF fluorescence curves which have been normalized to integrate to unity, then overlaid. The curves come from 56 discrete BS9 experiments that were run in 2013 in C3. T_m= 70.9 ± 0.7 °C for the 168 replicates. The standard deviation for the least square fit of these normalized curves is 5×10^{-7}, and this is the basis for the outlier rejection in the Meltdown program. There is some variation as to the curve shape of the lysozyme curves in the low temperature range. A plausible explanation for this might be aging of the lysozyme standard solution – the lysozyme control solution is 0.1 mg/mL lysozyme in 50 mM tris chloride pH 8.0, 50 mM NaCl. This is made up in a 1 mL volume, and stored in a refrigerator until it is used up (60 uL are used in each run of BS9). As the protein ages, some of the protein may unfold, leading to a melt curve with a high initial fluorescence.

C)   Shows a detail plot that follows the overall summary in the Meltdown report. In this example, there were 6 replicates grouped together (from the 'Condition Variable 1' column in the contents map file), and two sets within this grouping (either high and low salt) defined in the 'Condition Variable 2' column. The two salt concentrations within the replicate set are colored blue (low salt) or orange (high salt). There is one outlier (which comes from the low salt set) this is colored blue, but drawn with a dashed line.

D)   Shows a detail plot that follows the overall summary in the Meltdown report.   In this example, there were 6 replicates grouped together (from the first 'Condition Variable 1' column in the contents map file), and two sets within this grouping (high and low salt) defined in the 'Condition Variable 2' column.   Here the low salt (blue) curves are flagged as problematic: the dashed blue lines show that these curves were monotonic, and thus have no determinable T_m; the dotted blue curve indicates that this curve had more than one minimum in the first derivative curve, and is thus considered a 'complex' curve.

Figure 2

A)   A "reality checkbox" for a poorly behaved sample.   A 'reality checkbox' is presented first to the user in the Meltdown report. The first line "**Full interpretation requires you to look at the individual melt curves**" always appears. The second line "**The summary graph below appears to be unreliable**" does not appear only if the following three conditions are met: 50% or more of all the experimental curves were used to generate a T_m; the average estimation of error is 1.5°C or less and the 'protein as supplied' is well behaved.   A well behaved 'Protein as supplied' has all curves overlaid (no outliers or monotonic curves), with an estimation of error of 1.5 °C or less. The value of 1.5 °C was chosen as it is twice the spread found in the very well behaved protein lysozyme (Figure 1-B).

B) If available, the 'Protein as supplied' curves are displayed along with the 'reality checkbox'. Here eight (well behaved) 'Protein as supplied' control curves are overlaid.

Figure 3

A) Shows an overall summary graph prepared by Meltdown for a 96 well experiment where there was 8-fold replication of 10 experiments, and where there were eight 'Protein as supplied' control wells.   The red dashed horizontal line shows the melt temperature of the 'Protein as supplied'.   The values on the X-axis derive from the 'contents map' file

B) An overall summary graph prepared by Meltdown where up to 6 different salt concentrations were tested (with duplication) for 18 different primary buffer conditions.

Supporting information
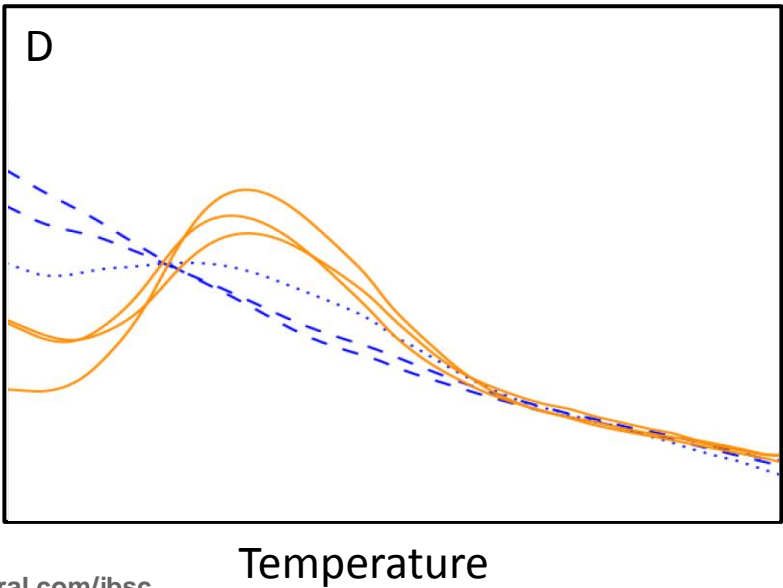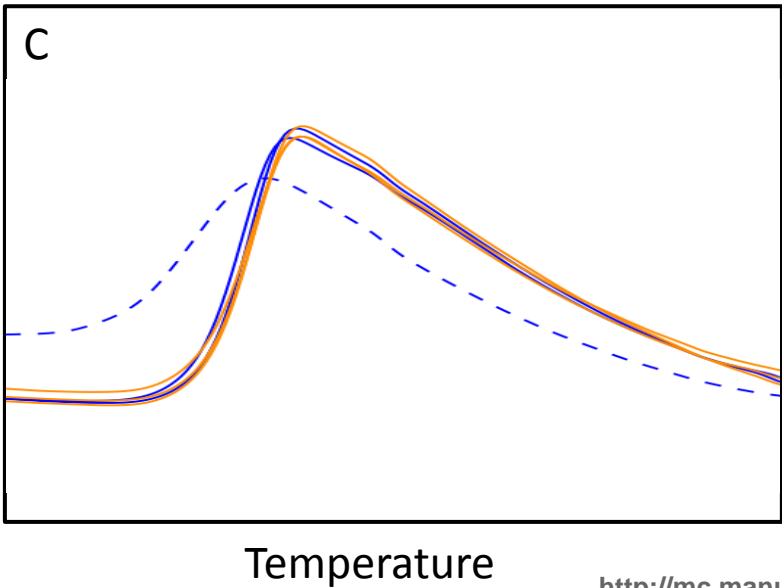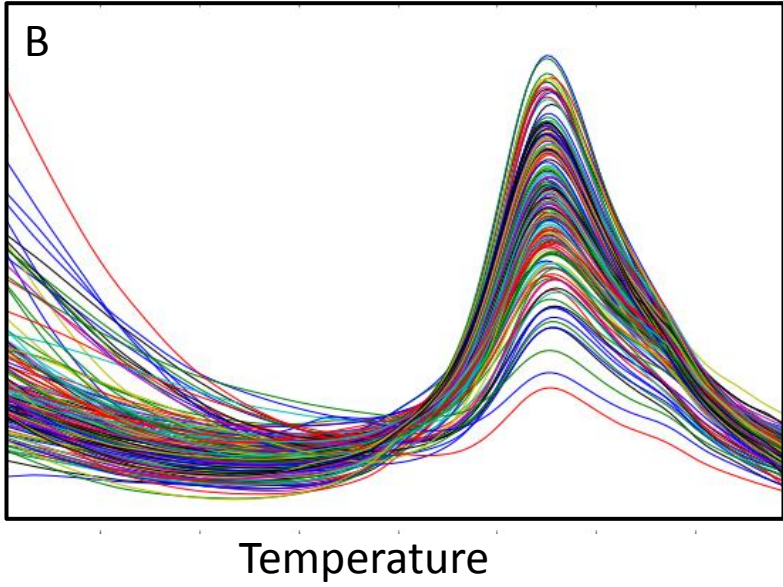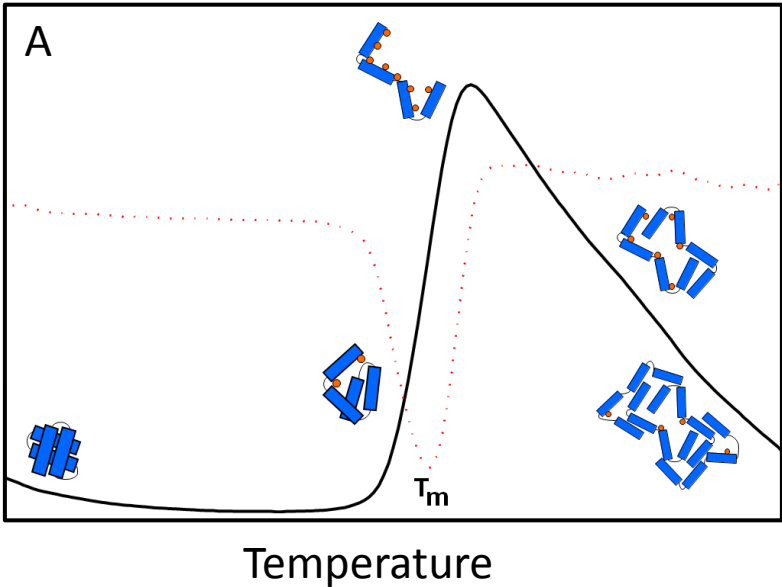
Meltdown reports, raw fluorescence data and content maps are provided for three different systems, along with the installation guide for Meltdown.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

2A

Full interpretation requires you to look at the individual melt curves.
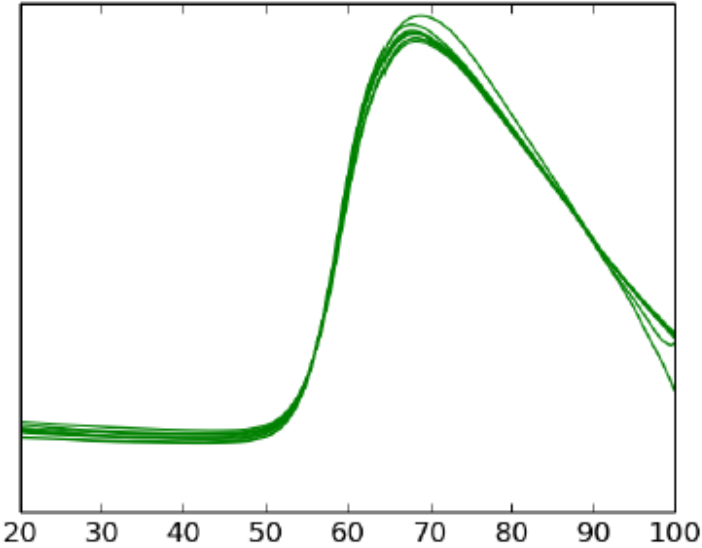
The summary graph below appears to be unreliable

22% of curves were used in Tm estimations

Average estimation of error is **3.4° C**

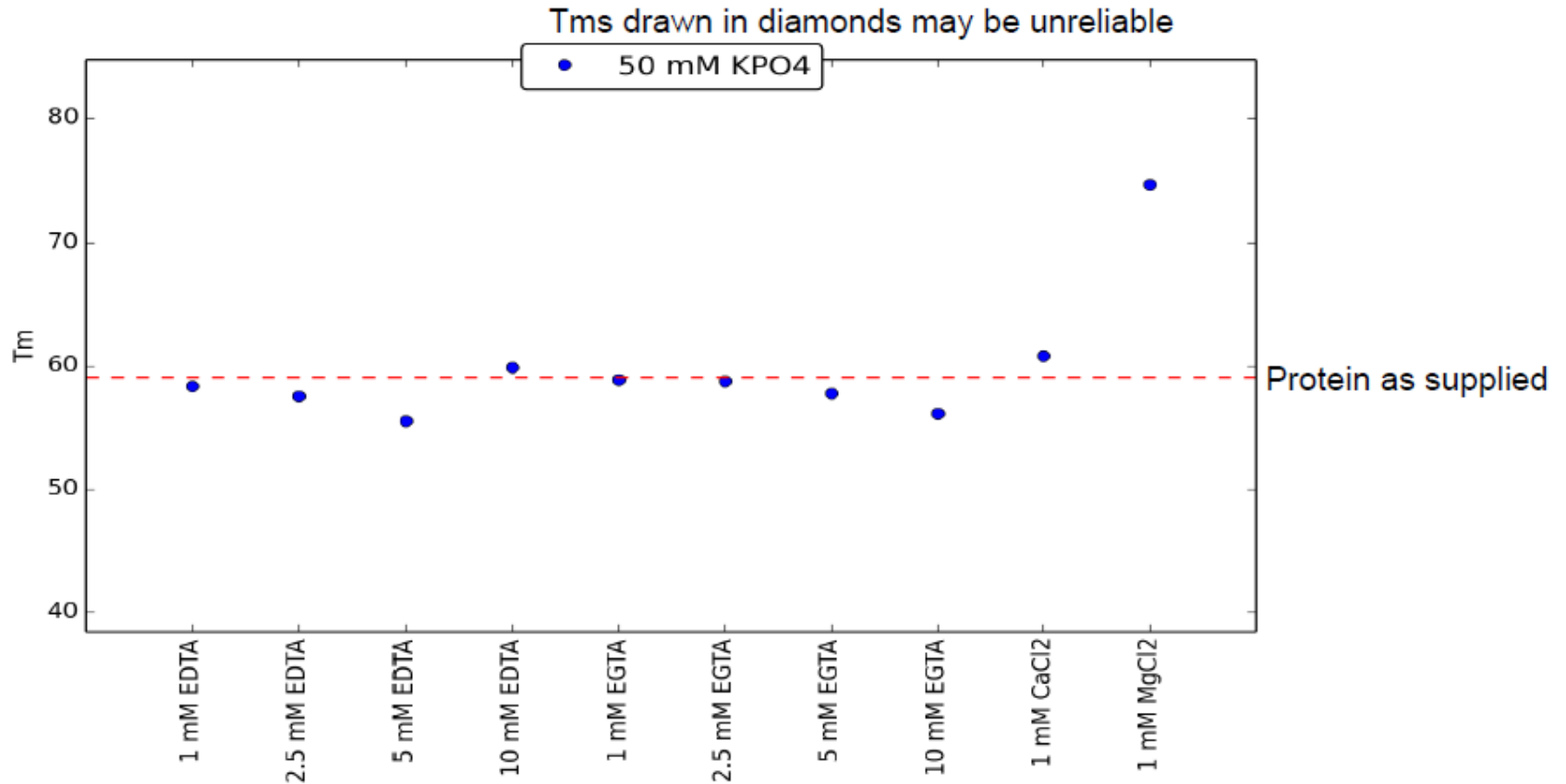Protein as supplied is **not well behaved**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

2B

Protein as supplied: Tm = 59.07(+/-0.1)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

3A

3B