SimpleDSFviewer: a tool to analyse and view differential scanning fluorimetry data for characterising protein thermal stability and interactions

Changye Sun, Yong Li, Edwin A. Yates, David G. Fernig

Department of biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK

Corresponding author: David G Fernig, Department of biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK
donferni@liv.ac.uk

Abstract:

Differential scanning fluorimetry (DSF) is used widely as a thermal shift assay to study protein stability and protein-ligand interactions. The benefit of DSF is that it is simple, cheap and can generate melting curves in 96-well plates providing good throughput. However, data analysis remains a challenge, and requires different methods to optimise and analyse the collected raw data. Here, the program SimpleDSFviewer is introduced to help view and analyse DSF data in an efficient way and with a user-friendly interface. The data analysis, optimisation and view methods provided by the program are described, using sample melting curves of fibroblast growth factors.
Introduction

Differential scanning fluorimetry (DSF) is a thermal shift assay technique used to measure the denaturation of proteins caused by increasing temperature, which breaks the non-covalent bonds that underlie protein folding (Niesen et al. 2007; Pantoliano et al. 2001; Semisotnov et al. 1991). A high temperature is required to denature a stable protein, while an unstable protein will be denatured at lower temperature. The fluorescent dye, which gives high fluorescence in a non-polar environment, is used in the DSF assay to probe the hydrophobic sites exposed on the unfolded proteins. So, as the temperature is increased, the protein unfolds to progressively expose more hydrophobic residues, until the protein is fully unfolded. This in turn produces more fluorescence by the interaction of fluorescent dye and the exposed hydrophobic residues. Since the DSF method is being adopted by an ever wider community, an increasing number of applications have been developed to test protein stability and protein-ligand interactions (Vivoli et al. 2014). The equipment requirement is so modest - a common RT-PCR instrument and 96 multi-well plates - that the experiment can be conducted easily and a large volume of data can be acquired (Niesen et al. 2007) and the experimental procedure is now described in video format, which will further facilitate its adoption (Vivoli et al. 2014). However, analysis of the data is more challenging and time consuming. Consequently, the present program for DSF data analysis has been written to help users view and analyse their data. The program has a user-friendly interface and a number of additional functions have been included, such as normalisation, smoothing and melting temperature extraction.

Materials and Methods

Proteins and ligands – In the given example, the reagents His-FGF10 (hexahistidine tagged Fibroblast Growth Factor 10) and His-FGF7 were produced, as described (Sun et al. 2015; Xu et al. 2013). Heparin sodium powder (from porcine intestinal mucosa, Sigma-Aldrich, Dorset, UK) was dissolved in phosphate-buffered saline (PBS: 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ and 137 mM NaCl, pH 7.4) to prepare a series of concentrations of ligands (supplementary: Example data) for FGF binding.
**Differential Scanning Fluorimetry (DSF) setup** – The purified FGFs, heparin ligand and SYPRO Orange dye (Life technologies, Paisley, UK) were added to the corresponding buffer and mixed gently, as listed in Table 1 (Uniewicz et al. 2010; Xu et al. 2013). The samples were split into three wells, 10 µL in each, and placed in a fast optical 96-well plate (Life technologies). The plate was then covered with Optical Adhesive Film to prevent evaporation and were read by a 7500 Fast Real-time PCR machine (Life technologies). The running method was designed to raise the temperature from 32°C to 81°C in 0.5°C steps, and to measure the fluorescence every 0.5°C for 30 seconds. The protein concentration used for data acquisition was 5 µM. The raw data were exported from the running software (7500 Fast Real-time PCR system) following the reading. The details of the program for data analysis are described in the supplementary.

**Table 1 DSF reaction components for three wells:**

<table>
<thead>
<tr>
<th>Components</th>
<th>Protein</th>
<th>Ligand (10X)</th>
<th>SYPRO orange (100X)</th>
<th>PBS</th>
<th>Total volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume / µL</td>
<td>a¹</td>
<td>3.5</td>
<td>3.5</td>
<td>(28-a)</td>
<td>35</td>
</tr>
</tbody>
</table>

a¹: volume of the protein for specified concentration.

**Results and Discussion** (Data analysis and display with SimpleDSFviewer using Matlab)

**1.1 Basic functions for data analysis and view** – SimpleDSFviewer automatically processes the raw data or the re-organised data (Fig. S2), as is shown in the flow chart (Fig. 1). The melting curve data are normalised and smoothed with the given range and the first derivative of the melting curve is calculated, as described in 1.2. The melting temperature of each melting curve is extracted (1.3). The analysed data may be saved in an Excel file for further analysis.

The melting curve patterns of the tested proteins are also automatically screened and presented as an image, as is shown in 1.4 and 1.5, which provides a view of the proteins’ thermal stabilities across the 96 wells. The protein melting curves and the first derivative curves of selected wells can be plotted in the preview box or in a separate figure, which can be automatically or manually...
saved in different file formats. The melting temperature point for each protein is also marked on the melting curve or first derivative curve.

**Flow chart of DSF data analysis and display**

**Figure 1 Flow chart of DSF data analysis and display.** The processing steps for the input data are described in this diagram. The red box and green boxes are the input (experimental data) and one output (melting temperature). The yellow boxes are the automatically analysed outputs. The purple boxes are to display data and to save data. The ‘Screen Data’ function directs the user to select the desired data for viewing.

**1.2 Normalise, smooth melting curve and first derivative curve** – The background is removed by subtracting the minimum value of each curve from all points, and then the melting curve is normalised by dividing by the maximum value (Equation 1). The normalisation of melting curves enables these to be compared to each other (Figs 2 A-B). The first derivative is calculated using Equation 2 and is normalised in the same way as the melting curves. The melting curves
may be slightly noisy, e.g., due to low protein concentration (Figs 2 C-D), so each normalised curve is smoothed in three parts: from the beginning to the minimum value, from the minimum value to the maximum value and from the maximum to the end, to ensure the original shape of the curve is maintained. The smoothing method is based on Matlab’s smooth function (MathWorks 2015) and the smoothing range can be defined by the user. Note that the ‘normalise’ and ‘smooth’ functions are optional.

**Figure 2 Normalise and smooth the melting curves.** Normalising the melting curves changes the display range of measured fluorescence intensity from 0 to 1, but the shape of the melting curve is unaltered. Smoothing the melting curves reduces the noise level, but also does not alter the shape of the curves. (A): Original melting curves of His-FGF10 (red) and His-FGF10 stabilised with 1.25 µM heparin (green). (B): Normalised melting curves of the same samples in (A). (C): Original melting curve (pink) and smoothed melting curve (black) of His-FGF7 stabilised with 50 µM heparin. (D): First derivative curves of melting curves in (C).
Equation 1: \[ MC = \frac{MC - \text{min}(MC)}{\text{max}(MC) - \text{min}(MC)} \]

Equation 2: \[ FD(n) = \frac{MC(n+1) - MC(n-1)}{2 \Delta t} \]

\( MC \): Melting curve values; \( \text{min}(MC) \): minimum value for the melting curve; \( \text{max}(MC) \): maximum value for the melting curve; \( MC(n+1) \) is the \((n+1)\)th value in the melting curve; \( FD(n) \) is the \(n\)th value in the first derivative curve. \( \Delta t \) is the difference in temperature between two neighbouring measurement points.

1.3 Calculate melting temperature – (1) First derivative curve method: The melting temperature is generally defined as the temperature at the maximum of the first derivative curve (Fig. 3 A). (2) Melting curve method: The melting temperature can also be determined by calculating the half maximal denaturation temperature (\( DT_{50} \)), which is the temperature when the fluorescence intensity reaches 50% of maximum value (Fig. 3 B). The \( DT_{50} \) is a slightly higher value, because it calculates the first temperature after the ideal \( DT_{50} \). But, the difference of the ideal \( DT_{50} \) and acquired \( DT_{50} \) is less than the temperature increase for each measuring step.

Figure 3 Extraction of melting temperatures. The stabilisation of His-FGF10 in the presence of a series of heparin ligands (concentrations are listed in Supplementary) was measured. The melting temperatures may be calculated by (A) the temperatures corresponding to half maximal fluorescence intensities and by (B) the temperatures corresponding to the maxima of the first derivative curves.
1.4 Auto-screen a melting curve – The melting curves are screened by the default parameters when the data are input into the program. The \textit{findpeak} function is used to find the peaks of the first derivative curves (Fig. 4 B). Only when the width and height of a peak are over 10% that of the highest peak in the assay is the peak recognised as a second melting temperature. However, if the number of recognised peaks is >=4 or the melting curve decreases from the maximal value to the minimal value as the temperature increases, the protein corresponding to the melting curve is considered to be a denatured or unfolded protein (Fig. 4 A).

\textbf{Figure 4} Auto-screening the melting curves and preview of melting temperatures in 96 wells. Three different types of melting curves (A) can be screened and recognised by the program. The screened results and the melting temperature information are automatically presented in two figures (C and D). (A): Melting curves of His-FGF10 (red line), His-FGF10 stabilised with 0.625 \( \mu \)M heparin (yellow line) and His-FGF10 (pink line) denatured by 8 M
urea. (B): First derivative curves calculated from melting curves in (A). (C): 2D map view of melting temperatures of the proteins in 96 wells. The dark blue wells contain unfolded proteins or unstable proteins (e.g., proteins in an inappropriate buffer). The colours of the melting temperatures of folded proteins are distributed between 50 and 250 (in the colour bar), which is from the lowest melting temperature to the highest temperature. (D): 3D view of melting temperatures. The melting temperature of each well is plotted with a bar in the 3D graph.

1.5 Preview of the protein stabilities in the 96 wells – This function is to show the wells containing unfolded proteins and to present the difference of their melting temperatures as an image map. The wells coloured with a low value colour (at the beginning of the colour bar, Fig. 4 C) are supposed to contain the unfolded proteins (Fig. 4 C). The melting temperatures are presented as different coloured squares from a value 50 colour (protein with lowest melting temperature) to value 256 colour (protein with highest melting temperature). A 3D bar plot containing the melting temperatures of the 96 wells is also prepared, to give a view of difference of proteins’ thermal stabilities (Fig. 4 D).

**Figure 5 Statistical analysis of different samples with ANOVA.** The melting temperatures of different samples can be compared with ANOVA. The melting temperatures of His-FGF7 in PBS, His-FGF7 stabilised with 50 µM heparin in PBS, His-FGF7 in PBS containing 0.5% (v/v) Tween-20 and His-FGF7 stabilised with 50 µM heparin in PBS containing 0.5% (v/v) Tween-20.
1.6 Comparison of melting temperatures of different samples – The melting temperatures of different samples can be compared by analysis of variance (ANOVA), as is described in Supplementary (User instructions). This will provide an appropriate statistical analysis (Fig. 5).

Conclusion

This program provides for the analysis of large amounts of data and can be easily transferred into analogous assays, such as spectral data acquired on 96-well plates. The program makes it more efficient to rapidly view and inspect the analysed data and produce a report graph. Further development can be introduced by the user or by contacting the authors.

References


Supplementary:

1 File Table:

(1) Code files (Folder: SimpleDSFviewer MATLAB code):

SimpleDSFviewer.m (SimpleDSFviewer.fig) is the main script, which displays the user interface and contains the buttons and parameter inputs.

Normalisemc.m is connected to the user interface script to normalise the melting curves by using the parameters loaded into the user interface.

Smomcurve.m is used to smooth the melting curves if it required.

Peaknumber.m is used to screen the melting curves, which automatically classifies the curves. The different curves shown in Fig. 4 A can be recognised automatically, and the parameters for classifying the curves can be changed manually.

mapTm.m is to generate the preview map (Fig. 4 C) and 3 D bar plot graph (Fig. 4 D).

(2) Runtime app files (Folders: SimpleDSFviewer Mac/Win32/Win64 runtime):

SimpleDSFviewer.exe is the software package for running the program. Choose the appropriate version, depending on the computer (Windows 32 or Windows 64 or Mac), and double click this file to run the program. The required Matlab Runtime should be installed before running this software.

readme.txt contains the details which introduce the requirement of installation of MATLAB runtime.

Other files are icon pictures and supporting files.

(3) Example data (‘example raw.xlsx’ and ‘example sorted.xlsx’).

All the above files for the DSF data analysis program can be freely downloaded from website:

https://github.com/hscsun/SimpleDSFviewer.git.
2 User instructions:

1. **Start the program:** Open SimpleDSFviewer.m in MATLAB R2015a (containing ‘Signal Processing Toolbox’, ‘Statistics and Machine Learning Toolbox’ and ‘Curve Fitting Toolbox’) and run the script. Alternatively, install the desired MATLAB Runtime 2015 (download from webpage: [http://uk.mathworks.com/products/compiler/mcr/](http://uk.mathworks.com/products/compiler/mcr/)) and open the compiled app (SimpleDSFviewer), if MATLAB software is not installed. The interface (Fig. S1) will be loaded and ready to use. A Windows computer is preferred, since the analysed data cannot be written into an Excel file for ‘Save Data’ function on a Mac computer.

![User interface of SimpleDSFviewer](image)

**Figure S1 User interface of SimpleDSFviewer.** This figure shows the user interface of SimpleDSFviewer, in which the user can analyse and view the collected DSF data. The top-left three panels are used to load the parameters and collected DSF data. The 96 well radio buttons and the data view buttons are distributed on the bottom left. The right side contains a graphical view box for displaying the selected data and the path of input data in a text box.
2. **Load parameters:** The temperatures, including start temperature, temperature step and end temperature are required, if a raw data file (Fig. S2 A) is to be loaded. In the software, ‘1’ signifies Yes and 0 signifies No for ‘Smooth’, ‘Normalise’ and ‘Legend’ parameters. The *Smooth* and *Normalise* parameters should be kept as ‘1’ as the data are loaded, but the range (*Smooth Range*) can be changed to be increased or decreased. The ‘Normalise’ and ‘Smooth’ functions can be turned off when the original melting curves are desired.

The default values are loaded automatically and are generally ideal for analysis.

![Figure S2 Templates of input data for SimpleDSFviewer.](image1.png)

**Figure S2 Templates of input data for SimpleDSFviewer.** Two formats of input data can be loaded into the program. (A): Raw data contains all the fluorescence reading from wells 1 to 96 and from the lowest temperature to the highest temperature in one column. The number of desired column needs to be specified. (B): Sorted data contains the melting temperature in the first column and the 96 wells’ fluorescence in 96 columns.
3. **Load experimental result:** Two formats of data can be loaded, as shown in Fig. S2. The first format is the raw data, which contains all the data in a single column and the column that contains the desired data should be specified (Text box next to ‘Load RawData’ pushing button, Fig. S1). Column 3 was selected in the example (Figs. S1 and S2 A). The second format is sorted data (Fig. S2 B), which contains 96 columns corresponding to 96 wells and fluorescence intensities of each temperature in each row. The loading data should be stored in the first sheet of the Excel file.

4. **Functions:** Multiple wells can be selected by clicking the corresponding radio buttons for data view. The melting curves and first derivative curves can be previewed in the graph box (Fig. S1) by pushing the buttons (containing MC or FD) in the Select Wells panel or be viewed and saved by pushing the buttons in the Save Result panel. ‘+Tm’ means the melting temperature is marked on the melting curve. MC (Half-Value) shows the melting curves with half denaturation point (main paper 3.3). The legend can be added by changing the legend parameter to ‘1’.

Different samples can be compared by ANOVA: Repeating samples are selected by clicking the radio buttons and a name should be input into the text box in the ‘Load Explanatory Values (X)’ panel. Then, the sample is added by pushing the ‘Group (String)’ button. The selected wells should be deselected and more samples can be added in the same way. The ANOVA (string) button will give the statistical analysis and a box-plot graph for the samples under comparison.

5. **Save function:** The sorted (Sheet1) and normalised melting curves (Sheet: nMC and nFD) and the first derivative curves and the melting temperatures (Sheet: Tm) calculated
by the three different methods (Fig. 1) may be saved as an Excel file by pushing ‘Save Data’ button.

6. **Errors:** When the data are input into the program, a running status bar will be displayed on the screen and it will be turned off after the auto-analysis. However, if the running cannot be completed (bar status), the input data should be checked. For example, if the gradient numbers of temperature are longer than the fluorescence measuring numbers for each well, an error will be generated and running will not complete. Missing values in the loaded data also cause the same problem.

7. **Example data:** Protein and the heparin (Hep) ligand concentrations are listed in the following table. Every three consecutive wells contain three same samples to ensure the experimental results.

<table>
<thead>
<tr>
<th></th>
<th>FGF10 PBS</th>
<th>FGF10 Hep 0.2 µM</th>
<th>FGF10 Hep 0.3 µM</th>
<th>FGF10 Hep 0.4 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGF10</td>
<td>Hep 0.5 µM</td>
<td>FGF10 Hep 0.625 µM</td>
<td>FGF10 Hep 0.75 µM</td>
<td>FGF10 Hep 0.875 µM</td>
</tr>
<tr>
<td>Hep 1 µM</td>
<td>FGF10 Hep 1.25 µM</td>
<td>FGF10 Hep 2.5 µM</td>
<td>FGF10 Hep 5 µM</td>
<td></td>
</tr>
<tr>
<td>Hep 50 µM</td>
<td>FGF10 Hep 500 µM</td>
<td>FGF10 PBS</td>
<td>FGF10 Hep 0.2 µM</td>
<td></td>
</tr>
<tr>
<td>FGF10</td>
<td>Hep 0.4 µM</td>
<td>FGF10 Hep 0.625 µM</td>
<td>FGF10 Hep 0.75 µM</td>
<td></td>
</tr>
<tr>
<td>FGF10</td>
<td>Hep 0.875 µM</td>
<td>FGF10 Hep 1.25 µM</td>
<td>FGF10 Hep 2.5 µM</td>
<td></td>
</tr>
<tr>
<td>FGF10</td>
<td>Hep 5 µM</td>
<td>FGF10 Hep 50 µM</td>
<td>FGF10 Hep 500 µM</td>
<td>FGF10 8 M urea</td>
</tr>
<tr>
<td>PBS</td>
<td>FGF10 Hep 50 µM</td>
<td>FGF10 0.5% Tween-20 (v/v)</td>
<td>FGF10 0.5% Tween-20 (v/v)</td>
<td>FGF10 Hep 50 µM</td>
</tr>
</tbody>
</table>