

EDUCATIONAL

Open Access



Processing binding data using an open-source workflow

Errol L. G. Samuel^{1,2*} , Secondra L. Holmes^{1,2} and Damian W. Young^{1,2,3*}

Abstract

The thermal shift assay (TSA)—also known as differential scanning fluorimetry (DSF), thermofluor, and T_m shift—is one of the most popular biophysical screening techniques used in fragment-based ligand discovery (FBLD) to detect protein–ligand interactions. By comparing the thermal stability of a target protein in the presence and absence of a ligand, potential binders can be identified. The technique is easy to set up, has low protein consumption, and can be run on most real-time polymerase chain reaction (PCR) instruments. While data analysis is straightforward in principle, it becomes cumbersome and time-consuming when the screens involve multiple 96- or 384-well plates. There are several approaches that aim to streamline this process, but most involve proprietary software, programming knowledge, or are designed for specific instrument output files. We therefore developed an analysis workflow implemented in the Konstanz Information Miner (KNIME), a free and open-source data analytics platform, which greatly streamlined our data processing timeline for 384-well plates. The implementation is code-free and freely available to the community for improvement and customization to accommodate a wide range of instrument input files and workflows.

Keywords: Thermal shift assay, Differential scanning fluorimetry, Fragment-based drug discovery, Fragment-based ligand discovery, KNIME, Education, Data analysis, Data processing, Workflows, Flow-based programming

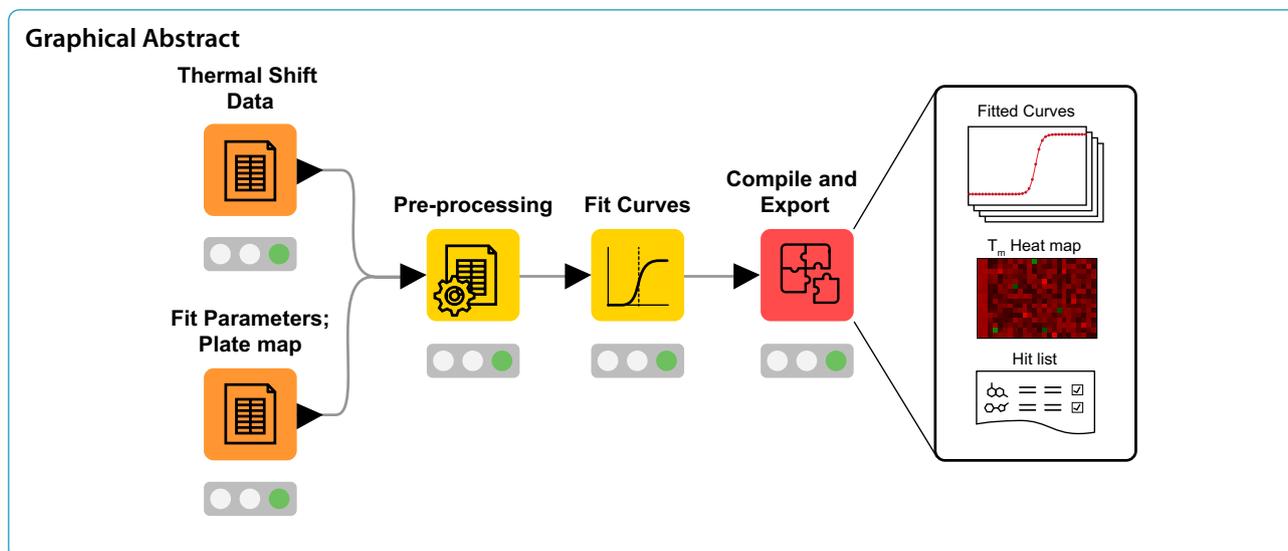
*Correspondence: errol.samuel@bcm.edu; damian.young@bcm.edu

¹ Center for Drug Discovery, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



Background

Modern small molecule protein modulators are developed using an array of drug discovery technologies such as fragment-based ligand discovery (FBLD), high-throughput screening (HTS), and DNA-encoded libraries (DELs). Library sizes and specific assays vary greatly, but irrespective of the preferred discovery technique, the ability to evaluate target engagement of every molecule in a screening library is essential to properly identifying novel binders and unlocking mechanistic details of newly developed drugs. Since screening libraries can be sizable collections of compounds, prodigious data sets are produced which require accurate and efficient methods of analysis to quickly identify hits. Academic labs are typically at a disadvantage when processing this data, as there are few methods that are both user-friendly and freely available. For users interested in such a method, this tutorial summarizes the widely popular thermal shift assay, discusses the existing approaches and workflows for thermal shift data processing, then outlines the development of a new, freely available, efficient cheminformatics workflow.

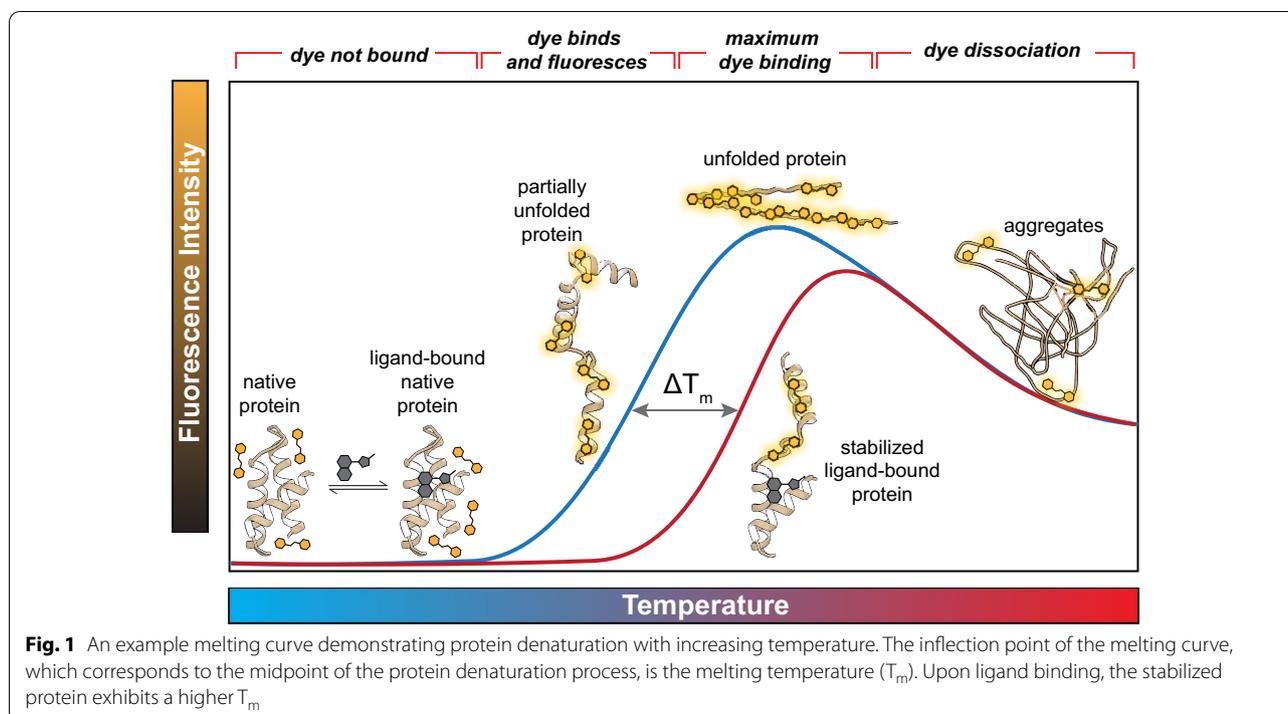
The thermal shift assay

The thermal shift assay (TSA) has become a mainstream technique to evaluate ligand binding owing to its ease of execution and high rate of throughput. Proteins fold in a manner that minimizes free energy. Hydrophobic residues pack into the interior of a structure and hydrophilic residues are exposed on the exterior [1]. The structures are stabilized by intramolecular forces such as electronic van der Waals attractions, dipole interactions, and hydrogen bonds [2]. As thermal energy increases, these intramolecular protein forces are disrupted, leading to protein

structure destabilization, denaturation, and the exposure of interior hydrophobic sites. Thermal denaturation can be monitored by fluorescence detection with real-time PCR instruments using intrinsic tryptophan fluorescence, or environmentally sensitive dyes, i.e., dyes which have increased quantum yield upon binding hydrophobic environments [3]. A graph of fluorescence intensity vs. temperature yields a characteristic melting curve, and the temperature of heat denaturation (T_m) is determined as the midpoint of the melting transition region. The T_m is influenced by sample conditions; bound ligands alter T_m values and therefore, experimentally significant ΔT_m observances are generally indicative of binding (Fig. 1) [4].

Existing approaches and workflows for thermal shift data processing

At the bench, preparing for and performing a TSA protocol is straightforward. However, the subsequent data analysis, especially in a high-throughput format, is less intuitive and requires substantial preparation and manual manipulation. The two most employed methods to obtain melting temperatures from TSA data are the first derivative method and the curve fitting method—if the melting curve is not symmetrical, the T_m values determined by these two methods may differ. The first-derivative method determines this point by searching for the maximum of the first-derivative curve, or the minimum of the negative first-derivative curve. In the curve-fitting method, the melting curve is fit to a sigmoidal function, and the melting temperature can be determined by the inflection point parameter. Curve-fitting methods are known to be more robust and less sensitive to random noise than the first-derivative method [5, 6]. Moreover,



a 5-parameter sigmoid provides a better fit particularly around the asymptotes compared to 4- or 3-parameter functions [7].

The data analysis process varies between laboratories, but often involves exporting the fluorescence data from the instrument, importing it to a separate software package to obtain melting temperatures, then exporting to another software package(s) for further analysis, hit determination, and visualization. The number of data transfers depends on the chosen form of T_m determination, available software, and the use of optional Microsoft® Excel plug-ins. Some qPCR instrument manufacturers provide software for analysis of thermal shift data, to make this increasingly popular assay more user friendly, but these become less user friendly when analyzing data from multiple plates. Moreover, if the software use requires additional licensing, this introduces a potentially costly roadblock.

To meet this need, several groups have developed and shared tools for processing and analyzing thermal shift data, including curve fitting tools written in the MATLAB® [7–9] and Python® [10] programming languages, standalone all-purpose tools written in the Perl® [6] and Java™ programming languages [11], and visualization tools based on Microsoft® Excel [5]. DMAN, Meltdown, SimpleDSFviewer, and ThermoQ employ the first derivative method, while TSA-CRAFT and MTSA employ the Boltzmann fitting method, using

4-parameter and 5-parameter sigmoidal functions, respectively. Though these tools are designed to increase efficiency, in practice their usage becomes increasingly challenging due to the requirements of licensed software and proficiency in specific programming languages. Further barriers include the inability to customize the software if the lab qPCR instrument changes, or if new analysis features are desired. Screening hit identification is also optimal with structure comparisons and to the best of our knowledge, no existing tool integrates the chemical structures into the data analysis or visualization.

Workflow/pipeline software. KNIME as an open-source data analysis tool

Workflow-based environments provide a common platform for multiple tools and have swiftly been adopted by the scientific community [12, 13]. The Konstanz Information Miner (KNIME) [14] is a free and open-source data analytics workflow platform which supports a wide range of functionality and has an active cheminformatics and bioinformatics community [15–25]. Its modular environment allows users to visually assemble and adapt the analysis flow from standardized building blocks called nodes, which are then connected through pipes carrying data or models. Nodes can be executed selectively and intermediate results checked via a graphical user interface. Metanodes and components can be used to encapsulate parts of a workflow and can even be re-used

between workflows, allowing for the creation of complex but well-structured workflows. The capabilities of KNIME can be expanded by incorporating open-source third-party nodes including R integration, and open-source cheminformatics toolkits such as the RDKit, both of which are used in our workflow.

Description of the KNIME workflow

While the workflow we describe here processes TSA data, the concepts and processes can be adapted to accommodate data from virtually any screening technique. Our workflow takes as inputs thermal shift assay data, a plate map containing chemical structures, and fitting parameters, and outputs melting temperature information for each compound. It has been designed as a sequence of 10 steps which guide the user through data analysis process and can be sequentially repeated for multiple plates (Fig. 2). Each step is color-coded in keeping with the KNIME node color scheme: orange is reserved for data input, yellow for data transformation, blue for exploration/visualization, and red for data output. This work was developed using KNIME version 4.4.0 and uses nodes from the KNIME Analytics Platform, open-source KNIME Extensions, and Community Extensions by RDKit, R, and the High-Throughput Technology Development Studio (TDS) of the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG). The RDKit extension is a collection of nodes that provides basic chemistry and cheminformatics functionality. The Interactive R Statistics Integration comprises nodes that interact with an external R™ installation [26] to write and execute R scripts. The high-content screening nodes provided by the MPI-CBG provide a powerful solution for mining screening data that is easy to use, but flexible enough to cope with various types of assays. These extensions should be downloaded once prior to first use. The KNIME workflow is freely available for download at <https://github.com/loicsamuel/knime-tsa-analysis>. Instructions for users explaining how to install and use the workflow are also provided at the same link and in Additional file 1.

Step 0: R server initialization

The first step of the workflow is to start the R server by running R and executing the command:

```
library(Rserve); Rserve(args = "--vanilla")
```

After command execution, the KNIME platform can communicate with the R server for executing nodes of the Interactive R Statistics Integration. Non-linear curve fitting of the melting curves is carried out using the Dose Response (R) node which uses the 'drc' R package [27] as a back end and 'ggplot2' for plotting the fit output. In the event that this workflow is used successively, step 0 is not required on subsequent runs.

Step 1: input data

The components in Step 1 reformat the raw output data from qPCR instruments into a table containing columns for well position, temperature, and fluorescence. These fields are the guiding requirement for workflow compatibility with a text file from any qPCR instrument, so users are asked to define these columns for the second component of this step. The data is then split into 2 outputs. The first is pivoted into a single temperature column and columns containing fluorescence values for each well at each temperature point, then transferred to step 2 for plotting, while the second output is transferred unchanged to step 3 for further processing (Fig. 3). We have successfully imported and processed the data file output from a Roche LightCycler® 480 II instrument, but the step 1 component can be modified to accept and reformat any common file output. For example, we have also successfully imported a pre-processed Microsoft® Excel output from the Applied Biosystems QuantStudio 7 Flex Real-Time PCR System.

Step 2: examine melting curves

The user is encouraged to examine the melting curves to confirm that the majority are well-behaved, ideally with a flat initial pretransition region followed by a steep sigmoidal unfolding region, and finally an aggregation region, as shown in Fig. 1. The user should decide on a threshold for excluding wells which do not have sufficient fluorescence signal for reliable fitting and will therefore not yield sensible results. This step is also critical in the selection of an appropriate initial fitting window, as restricting the fitting algorithm to the transition region improves workflow performance.

(See figure on next page.)

Fig. 2 Thermal Shift Assay data analysis workflow in KNIME

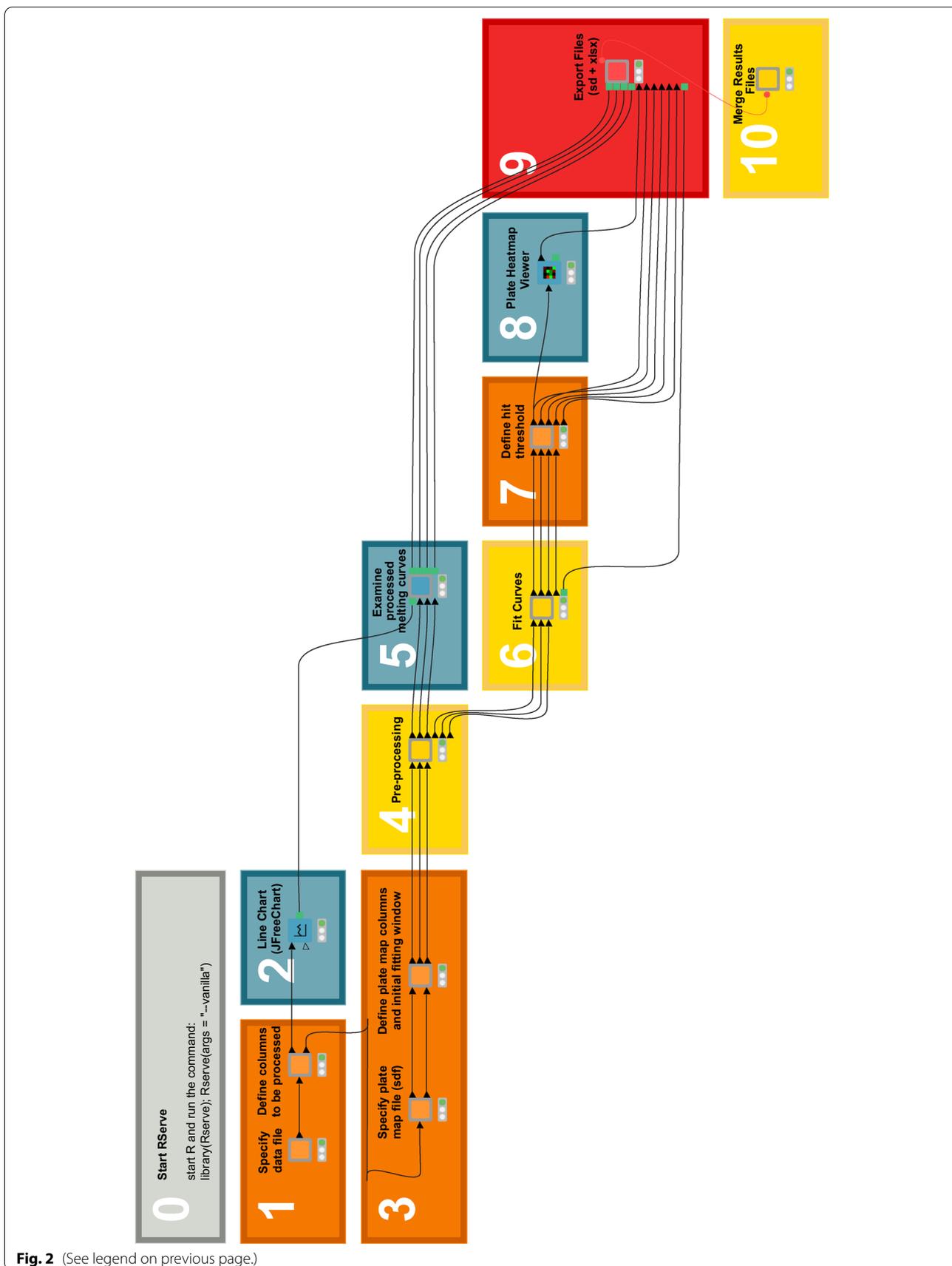
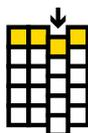


Fig. 2 (See legend on previous page.)

```

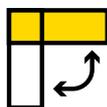
Raw Data Experiment - Thermal Shift Assay - Plate 10001 (Run on LCS480 1.5.0.39)
SamplePos SampleName Prog# Seg# Cycle# Time Temp 465-580
A1 Sample 1 1 2 1 24250 20.04 7.70
A1 Sample 1 1 2 1 25333 20.05 7.69
A1 Sample 1 1 2 1 26450 20.36 7.73
A1 Sample 1 1 2 1 27550 20.28 7.71
A1 Sample 1 1 2 1 28633 20.36 7.67
A1 Sample 1 1 2 1 29750 20.51 7.69
A1 Sample 1 1 2 1 30850 20.59 7.66
A1 Sample 1 1 2 1 31950 20.66 7.67

```



1. Read starting from line 2
2. Rename columns
3. Keep only well position, temperature, fluorescence

	Well Position	Temperature	Fluorescence
Row0	A1	20.04	7.70
Row1	A1	20.05	7.69
Row2	A1	20.36	7.73
Row3	A1	20.28	7.71
Row4	A1	20.36	7.67
Row5	A1	20.51	7.69
Row6	A1	20.59	7.66
Row7	A1	20.66	7.67



- Pivot
- group by temperature
 - pivot on well position
 - aggregate by fluorescence

	Temperature	A1	A2	A3	A4	A5
Row0	20.04	7.7	7.64	8.2	8.19	8.34
Row1	20.05	7.69	7.65	8.25	8.16	8.31
Row2	20.28	7.71	7.61	8.19	8.18	8.34
Row3	20.36	7.73	7.65	8.22	8.18	8.37
Row4	20.51	7.69	7.62	8.23	8.21	8.39
Row5	20.59	7.66	7.62	8.2	8.19	8.34
Row6	20.66	7.67	7.61	8.2	8.19	8.36

Fig. 3 Reformatting a tab-separated text file containing thermal shift data exported from the Roche LightCycler® 480 II for further processing in the workflow

Step 3: experiment details

In step 3, the user is prompted to provide a plate map, in the form of an SD file, containing all library compounds

and their locations (chemical structure, compound ID, plate ID, row, column, etc.), the initial fitting window, and a threshold for excluding wells which lack a suitable

fluorescence signal. As an alternative to the SD format, a workflow is provided which accommodates a plate map of compound SMILES in CSV format. Example data, along with plate maps and KNIME workflows are provided as Additional files 2, 3, 4, 5, 6 and 7. The plate number, row, and column data contained in the plate map are used to associate the chemical structures and compound IDs with the temperature and fluorescence data. The joined data is then passed on to subsequent nodes which remove data outside of the fitting window and plot the resulting curves.

Step 4: flag problematic wells

First, the range of fluorescence values in each well is calculated and compared to the threshold specified in step four. Wells that do not meet the fluorescence signal threshold are flagged and separated for plotting and review in step 5. The user should examine these melting curves to determine whether they were correctly flagged. The fluorescence range threshold should be adjusted until the filter captures, as much as possible, only the low-signal curves. Melting curves that meet or exceed the threshold are then analyzed for the presence of a downward slope or high initial fluorescence, i.e., whether the global maximum fluorescence occurs at a lower temperature than the global minimum fluorescence. After this, the curves are analyzed for the presence of gaps, jumps, and spikes. By default, these are defined as regions where Δf (the change in fluorescence between two successive data points) is more than 25 times the mean Δf for that well. This criterion can be modified to suit the data being processed—noisy data would require a higher threshold. We suggest starting with a lower value then increasing it if necessary.

Step 5: review pre-processing results

Here, the user can review the processed melting curves to confirm that the shapes are sigmoidal with well-behaved

minima and maxima, and a sigmoidal transition region that is completely inside the fitting window. Importantly, the user is also advised to confirm whether the flagged spectra are indeed problematic, and to adjust the thresholds and the fitting window of step 3 to avoid anomalies while capturing as much real data as possible. Figure 4 illustrates several types of melting curves which are flagged by this component. The well selection feature in Step 1 allows for the isolation of problematic individual wells if desired, and a T_m can be calculated on these by adjusting the initial window for a more custom fit.

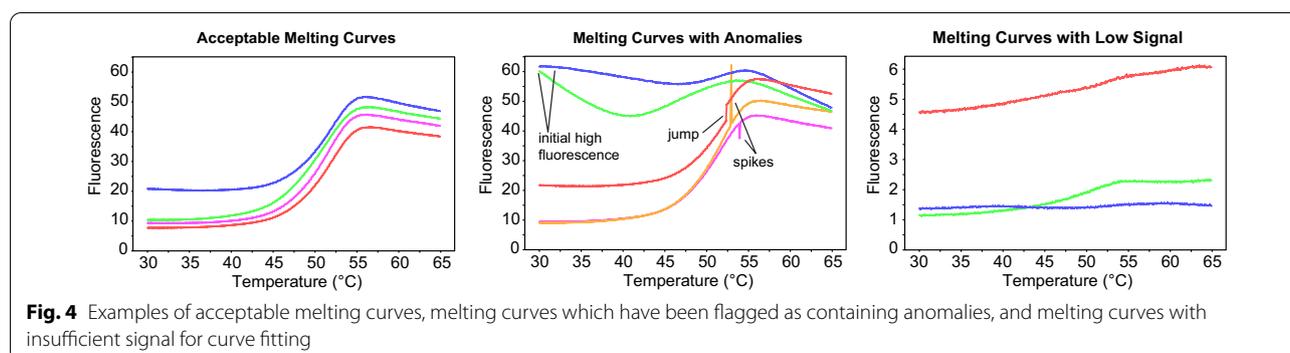
Step 6: non-linear regression

Inside this component is a loop, which for each well,

1. Applies a 21-point center simple moving average for smoothing of the fluorescence values
2. Normalizes fluorescence values between 0 and 1
3. Determines the temperature value at which the minimum fluorescence value occurs
4. Discards fluorescence data which occur at lower temperatures than the value determined in (3)
5. Determines the temperature value at which the maximum fluorescence value occurs
6. Discards fluorescence data which occur at higher temperatures than the value determined in (5)
7. Fits the fluorescence data from each well to a 5-parameter log-logistic model

$$f(x) = c + \frac{d - c}{[1 + \exp(b(\ln(x) - \ln(e)))]^f}$$

where b = slope, c = minimum fluorescence value, d = maximum fluorescence value, e = inflection point, and f = asymmetry factor, i.e. the difference in the rates of approach from the inflection point to the lower and the upper asymptotes (Fig. 5). If the melting curve is symmetrical then $f=1$ and the equation becomes a 4-parameter log-logistic.



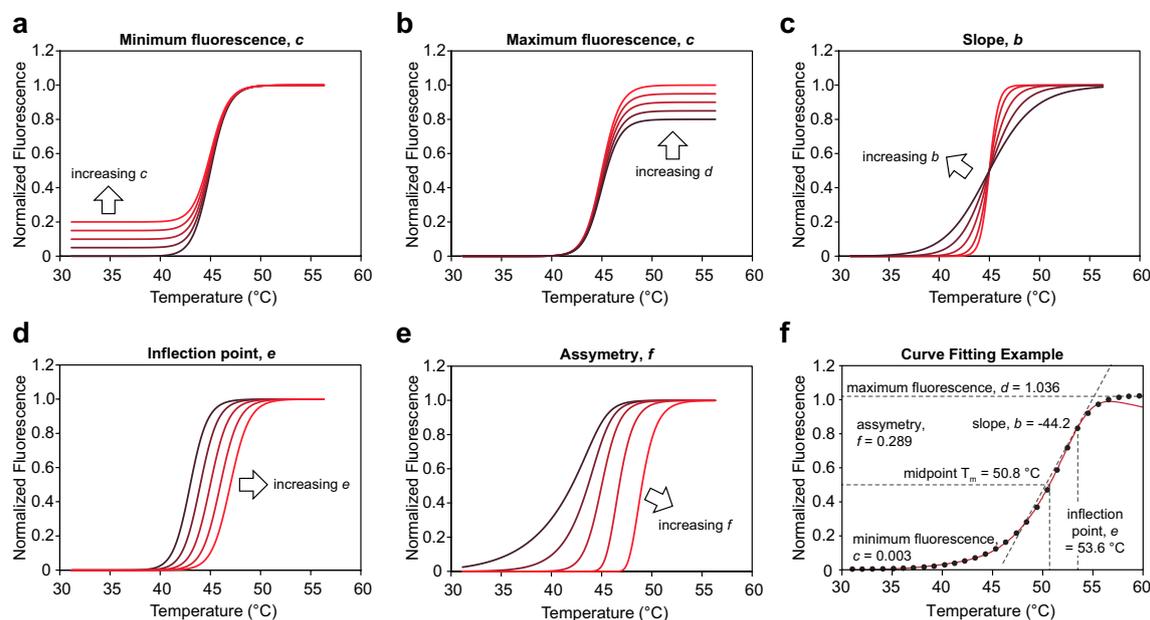


Fig. 5 Illustration of the parameters used in 5-parameter log-logistic fitting. **a–e** Effect of increasing minimum, maximum, slope, inflection point, and asymmetry values, respectively. **f** melting curve data (red line) fit to a 5-parameter log-logistic function (black dots), and the resulting fit parameters

8. The inflection point of an asymmetric sigmoidal curve does not coincide with its midpoint. The midpoint T_m is calculated using the fit parameters according to the equation

$$\text{midpoint } T_m = \frac{e}{\exp\left[-\frac{1}{b} \ln\left(2^{\frac{1}{f}} - 1\right)\right]}$$

The component returns the melting temperature, fit parameters with error estimates, and the fitted curves.

Step 7: define hit threshold and compile results

Once fitting is complete, the user can define the hit threshold for hit selection by double clicking on the component in Step 7. By default, this is set to 3 times the standard deviation of the melting temperatures of the DMSO-treated wells. The component splits the data into two branches for further manipulation. The first branch groups the logistic fit results by molecule ID number and calculates the average melting temperature and standard deviation, thereby providing a summary of replicates if present. The melting temperatures are then compared to the previously defined hit threshold and hits are tagged as such. The second branch creates a heatmap of the plate, colored by melting temperature.

Step 8: review melting temperature heatmap

At this stage, the user is encouraged to inspect the array of melting temperatures using the Plate Heatmap Viewer node. This node provides access to three levels of data with corresponding levels of detail: screen \rightarrow plate \rightarrow well. The screen view (Fig. 6a) is the default view, a trellis containing each plate in the experiment. This view is invaluable for visualizing row and column artifacts and for comparing plate performance across a screening campaign. For this workflow there is only one plate. The *View* menu provides the ability to change the colormap, while the trellis toolbar allows the user to change the readout as well as overlaid information, such as a hit tag. The second data level is the plate view (Fig. 6b) which is accessed by right-clicking on a plate in the heatmap trellis. There are identical controls as in the screen view, and individual wells can be inspected by hovering over them with the mouse. Right-clicking on a well provides access to the third and final data level, the well view (Fig. 6c). This view provides a detailed view of the wells along with information such as well position, compound ID, and chemical structure. The user is advised to note any patterns that may indicate the presence of problematic wells, e.g., lack of consistency in DMSO control wells, or systematic errors such as the presence of edge effects, outlier rows/columns, or unexpected patterns for ostensibly randomly plated compounds.

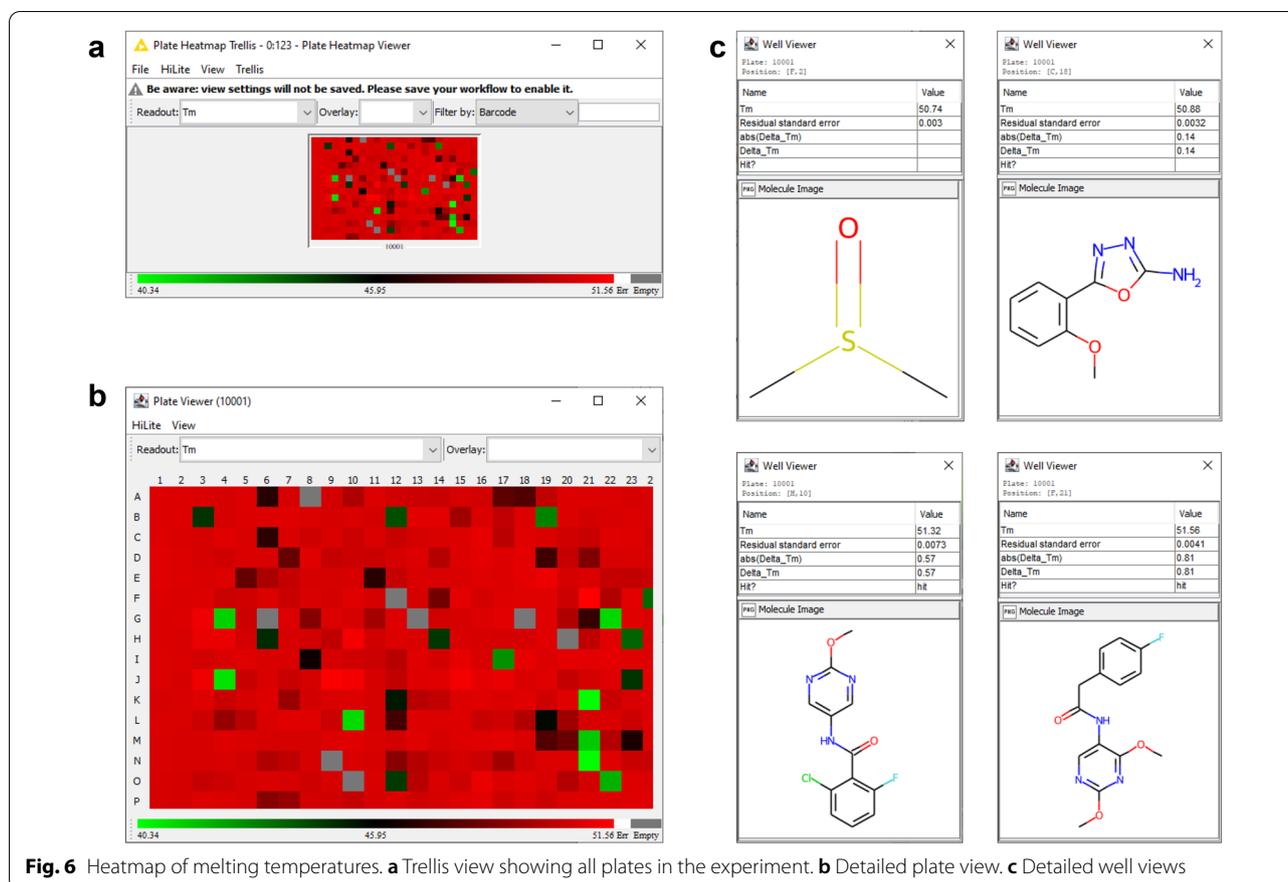


Fig. 6 Heatmap of melting temperatures. **a** Trellis view showing all plates in the experiment. **b** Detailed plate view. **c** Detailed well views

Step 9: output

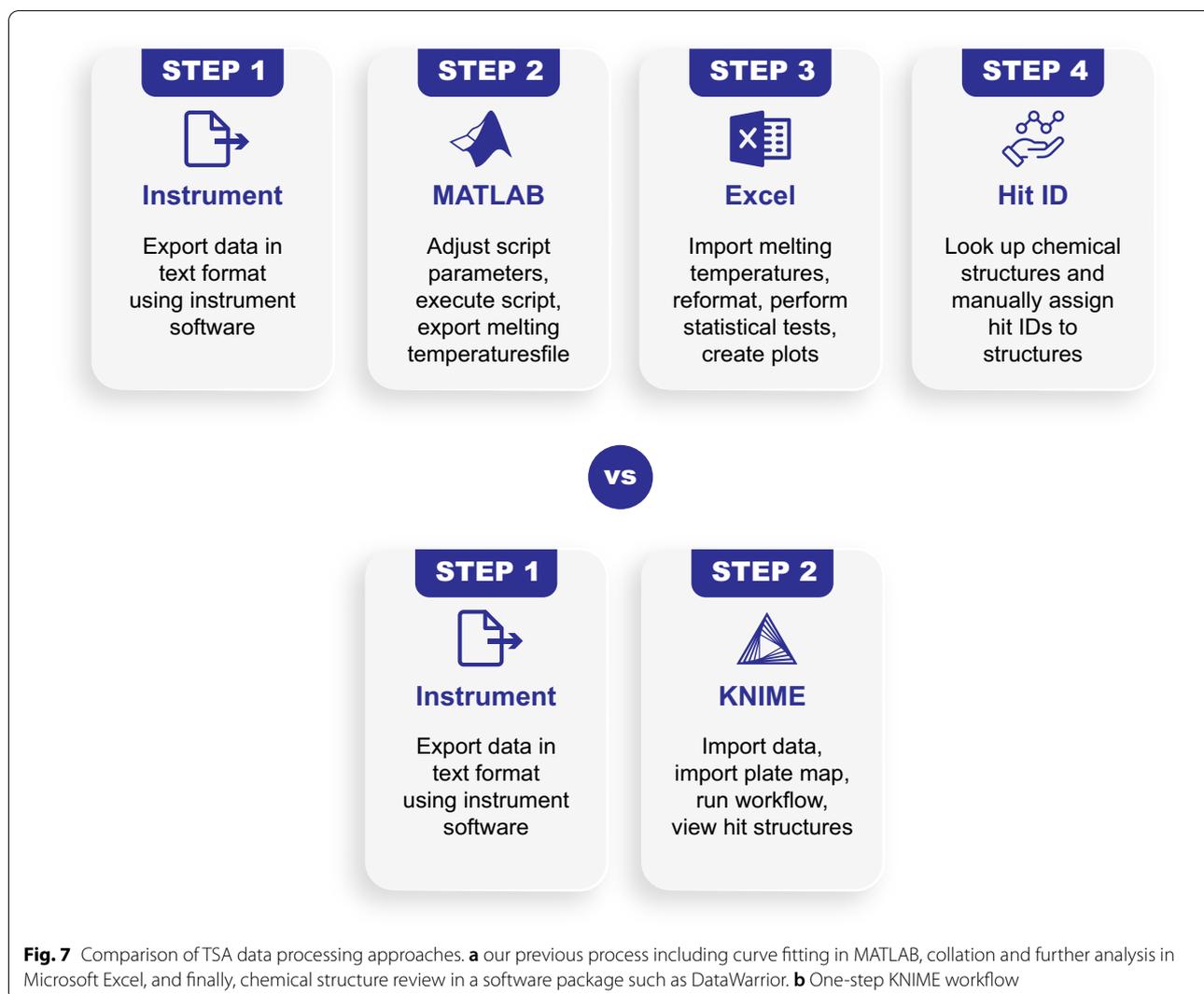
The workflow produces two files, located in the same folder as the initial thermal shift data file and sharing the same name. The first is an SD file containing all the columns of the initial plate map file along with melting temperature, residual standard error of the fit, ΔT_m , $|\Delta T_m|$ and hit status. Alternately, a second workflow is provided which outputs a CSV file with compound SMILES instead of an SD file. The second output is an XLSX document containing various sheets:

- Plate Heatmap*: this includes a heatmap of the melting temperatures across the plate
- Fit Results with hits annotated*: a list of all fitted compounds including (i) an image of the molecule, (ii) molecule ID number, (iii) well position, (iv) plate row, (v) plate column, (vi) plate ID number, (vii) melting temperature, (ix) residual standard error, (x) ΔT_m , (xi) $|\Delta T_m|$, (xii) hit status
- DMSO mean and SD*: this includes (i) molecule ID, (ii) control T_m , (iii) standard deviation of control T_m , (iv) hit threshold as $3 \times$ standard deviation of control T_m , (v) hit threshold as control $T_m + 3 \times$ standard deviation of control T_m

- Fitted Melting Curves*: this includes the well position and a graph of the sigmoidal fit for each fitted well
- Melting Curves*: this includes the graphs showing the overlaid raw melting curves, cropped melting curves, accepted melting curves, melting curves with anomalies, melting curves with low signal, and processed melting curves used for fitting.

Step 10: data consolidation

Data files loaded into the KNIME workflow during step 1 will typically consist of TSA raw fluorescence data from a single plate run of the qPCR instrument, but screening libraries are typically large enough to span multiple plates. The next steps in the screening project are to compile and compare all available data and it is inconvenient to have that data spread across the multiple files that will be produced from multiple runs of the workflow. This



last (optional) workflow component merges the results files produced during step 9. When executed, the component in step 10 reads all SD (or CSV) files saved in the “Results” folder and combines the data into a single table which is then saved as both XLSX and SD (or CSV) files.

Conclusion

The KNIME workflow developed and described in this paper provides an efficient and user-friendly method for processing both small and large TSA datasets. It provides significant savings both in time and financial resources (Fig. 7), and we hope it will serve as a valuable tool for researchers, particularly in academic institutions with

smaller budgets and lower data analytics and programming expertise.

This type of workflow is accessible and can be used as a model for analysis of large datasets from virtually any other screening technique. We encourage the community to apply this tool in their scientific work and to improve it based on their knowledge, experience, and specific data processing needs.

Abbreviations

CSV: Comma separated variable; DEL: DNA-encoded library; DNA: Deoxyribonucleic acid; DSF: Differential scanning fluorimetry; FBLD: Fragment-based ligand discovery; KNIME: Konstanz information miner; PCR: Polymerase chain

reaction; PROTAC: Proteolysis targeting chimera; SDF: Structure data file; TSA: Thermal shift assay; T_m : Melting temperature.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13321-021-00577-1>.

Additional file 1. Workflow Instructions.

Additional file 2. Thermal Shift Data (Plate 10001).

Additional file 3. Thermal Shift Data (Plate 10002).

Additional file 4. Plate Map (SD format).

Additional file 5. Plate Map (CSV format).

Additional file 6. KNIME Workflow (SD format).

Additional file 7. KNIME Workflow (CSV format).

Acknowledgements

The authors thank Mr. Shelton Boyd for his critical evaluation of the manuscript and workflow instructions for clarity.

Authors' contributions

ELGS and SLH developed the workflows presented in the paper. DWY supervised the project. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R01GM139295 and R01GM138781).

Availability of data and materials

The KNIME workflow and accompanying data are included within the article (and its additional files) and on GitHub (<https://github.com/loicsamuel/knime-tsa-analysis>).

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Center for Drug Discovery, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. ²Department of Pharmacology and Chemical Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. ³Department of Pathology and Immunology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

Received: 8 October 2021 Accepted: 24 November 2021

Published online: 11 December 2021

References

- Šali A, Shakhnovich E, Karplus M (1994) How does a protein fold? *Nature* 369:248–251. <https://doi.org/10.1038/369248a0>
- Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A relationship between protein stability and protein function. *Proc Natl Acad Sci* 92:452–456. <https://doi.org/10.1073/pnas.92.2.452>
- Huynh K, Partch CL (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* 79:2891–28914. <https://doi.org/10.1002/0471140864.ps2809s79>
- Waldron TT, Murphy KP (2003) Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry* 42:5058–5064. <https://doi.org/10.1021/bi034212v>
- Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2:2212–2221. <https://doi.org/10.1038/nprot.2007.321>
- Lee P-H, Huang XX, Teh BT, Ng L-M (2019) TSA-CRAFT: a free software for automatic and robust thermal shift assay data analysis. *SLAS Discov Adv Sci Drug Discov* 24:606–612. <https://doi.org/10.1177/2472555218823547>
- Schulz MN, Landström J, Hubbard RE (2013) MTSA—a Matlab program to fit thermal shift data. *Anal Biochem* 433:43–47. <https://doi.org/10.1016/j.ab.2012.10.020>
- Phillips K, de la Peña AH (2011) 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 Protoc Mol Biol* 94:1–15. <https://doi.org/10.1002/0471142727.mb1028s94>
- Sun C, Li Y, Yates EA, Fernig DG (2020) SimpleDSFviewer: a tool to analyze and view differential scanning fluorimetry data for characterizing protein thermal stability and interactions. *Protein Sci* 29:19–27. <https://doi.org/10.1002/pro.3703>
- Rosa N, Ristic M, Seabrook SA et al (2015) Meltdown: a tool to help in the interpretation of thermal melt curves acquired by differential scanning fluorimetry. *J Biomol Screen* 20:898–905. <https://doi.org/10.1177/1087057115584059>
- Wang CK, Weeratunga SK, Pacheco CM, Hofmann A (2012) DMAN: a java tool for analysis of multi-well differential scanning fluorimetry experiments. *Bioinformatics* 28:439–440. <https://doi.org/10.1093/bioinformatics/btr664>
- Shon J, Ohkawa H, Hammer J (2008) Scientific workflows as productivity tools for drug discovery. *Curr Opin Drug Discov Devel* 11:381–388
- Warr WA (2012) Scientific workflow systems: Pipeline Pilot and KNIME. *J Comput Aided Mol Des* 26:801–804. <https://doi.org/10.1007/s10822-012-9577-7>
- Berthold MR, Cebron N, Dilll F, et al (2008) KNIME: The Konstanz Information Miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds) *Data analysis, machine learning and applications. studies in classification, data analysis, and knowledge organization*. Springer Berlin Heidelberg, pp 319–326.
- Beiskens S, Meint T, Wiswedel B et al (2013) KNIME-CDK: workflow-driven cheminformatics. *BMC Bioinformatics* 14:257. <https://doi.org/10.1186/1471-2105-14-257>
- Leonis G, Melagraki G, Afantitis A (2017) Open source cheminformatics software including KNIME analytics. *Handbook of computational chemistry*. Springer International Publishing, Cham, pp 2201–2230
- Falcón-Cano G, Molina C, Cabrera-Pérez MÁ (2020) ADME prediction with KNIME: development and validation of a publicly available workflow for the prediction of human oral bioavailability. *J Chem Inf Model* 60:2660–2667. <https://doi.org/10.1021/acs.jcim.0c00019>
- Mazanetz MP, Marmon RJ, Reisser CBT, Morao I (2012) Drug discovery applications for KNIME: an open source data mining platform. *Curr Top Med Chem* 12:1965–1979. <https://doi.org/10.2174/156802612804910331>
- Fillbrunn A, Dietz C, Pfeuffer J et al (2017) KNIME for reproducible cross-domain analysis of life science data. *J Biotechnol* 261:149–156. <https://doi.org/10.1016/j.jbiotec.2017.07.028>
- Gadaleta D, Lombardo A, Toma C, Benfenati E (2018) A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *J Cheminform* 10:60. <https://doi.org/10.1186/s13321-018-0315-6>
- Sydow D, Wichmann M, Rodríguez-Guerra J et al (2019) TeachOpenCADD-KNIME: a teaching platform for computer-aided drug design using KNIME workflows. *J Chem Inf Model* 59:4083–4086. <https://doi.org/10.1021/acs.jcim.9b00662>
- Mazanetz MP, Goode CHF, Chudyk EI (2020) Ligand- and structure-based drug design and optimization using KNIME. *Curr Med Chem* 27:6458–6479. <https://doi.org/10.2174/0929867326666190409141016>
- Saldívar-González FI, Huerta-García CS, Medina-Franco JL (2020) Cheminformatics-based enumeration of chemical libraries: a tutorial. *J Cheminform* 12:64. <https://doi.org/10.1186/s13321-020-00466-z>
- Tuerkova A, Zdrzil B (2020) A ligand-based computational drug repurposing pipeline using KNIME and programmatic data access: case studies for rare diseases and COVID-19. *J Cheminform* 12:71. <https://doi.org/10.1186/s13321-020-00474-z>
- Afantitis A, Melagraki G (2020) Cheminformatics toolboxes and workflows within KNIME analytics. *Curr Med Chem* 27:6442–6443. <https://doi.org/10.2174/092986732738201014102814>
- R Core Team (2021) R: a language and environment for statistical computing
- Ritz C, Baty F, Streibig JC, Gerhard D (2015) Dose-response analysis using R. *PLoS ONE* 10:e0146021. <https://doi.org/10.1371/journal.pone.0146021>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.