MBoC | METHODS AND RESOURCES

NAGPKin: Nucleation-and-growth parameters from the kinetics of protein phase separation

Zsuzsa Sárkány[a,b], Francisco Figueiredo[a,b,†], Sandra Macedo-Ribeiro[a,b], and Pedro M. Martins[a,b,*]
[a]Biomolecular Structure and Function Group, IBMC – Instituto de Biologia Molecular e Celular, Porto 4200-135, Portugal; [b]i3S – Instituto de Investigação e Inovação em Saúde da Universidade do Porto, Porto 4200-135, Portugal

ABSTRACT **The assembly of biomolecular condensate in eukaryotic cells and the accumulation of amyloid deposits in neurons are processes involving the nucleation and growth (NAG) of new protein phases. To therapeutically target protein phase separation, drug candidates are tested in in vitro assays that monitor the increase in the mass or size of the new phase. Limited mechanistic insight is, however, provided if empirical or untestable kinetic models are fitted to these progress curves. Here we present the web server NAGPKin that quantifies NAG rates using mass-based or size-based progress curves as the input data. A report is generated containing the fitted NAG parameters and elucidating the phase separation mechanisms at play. The NAG parameters can be used to predict particle size distributions of, for example, protein droplets formed by liquid-liquid phase separation (LLPS) or amyloid fibrils formed by protein aggregation. Because minimal intervention is required from the user, NAGPKin is a good platform for standardized reporting of LLPS and protein self-assembly data. NAGPKin is useful for drug discovery as well as for fundamental studies on protein phase separation. NAGPKin is freely available (no login required) at https://nagpkin.i3s.up.pt.**

SIGNIFICANCE STATEMENT

- Protein phase separation (PPS) is a ubiquitous phenomenon with vital roles in human health and disease. Drug discovery in PPS has been hampered by the lack of quantitative methods characterizing nucleation and growth rates.
- NAGPKin is the first computational tool dedicated to quantifying PPS kinetics from the mass -or size-increase of biomolecular condensates over time. Fundamental insights are gained from the automatic analysis of raw experimental data.
- NAGPKin is useful for the development of new drugs targeting PPS, for standardized data reporting and, on the whole, for understanding the molecular mechanisms of functional and pathological PPS.

Monitoring Editor
Huaiying Zhang
Carnegie Mellon University

Received: Aug 7, 2023
Revised: Nov 27, 2023
Accepted: Dec 13, 2023

This article was published online ahead of print in MBoC in Press (http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E23-07-0289) on December 20, 2023.

Conflict of interest: The authors declare no financial conflict of interest.

Author contributions: Z.S. and F.F. performed the experiments; Z.S., S.M., and P.M. analyzed the Data; S.M. and P.M. drafted the Article; P.M. conceived and designed the experiments.

†Present address: Institute of Science and Technology Austria, ISTA, Klosterneuburg 3400, Austria.

*Address correspondence to: Pedro M. Martins (pmartins@ibmc.up.pt).

Abbreviations used: Atx3, ataxin-3; DA, dopamine; DLS, dynamic light scattering; LCD, low-complexity domain; LLPS, liquid-liquid phase separation; mTTR, monomeric transthyretin; NAG, nucleation and growth; polyQ, polyglutamine; STEs, surface tension effects; TDP-43, TAR DNA-binding protein 43; ThT, Thioflavin-T.

INTRODUCTION

The liquid-liquid phase separation (LLPS) of biomolecular condensates and the aggregation of proteins into amyloid fibrils are examples of protein phase separation in cell physiology and disease. Pathological hallmarks of several neurodegenerative diseases, amyloid fibrils form through the phase separation step of primary nucleation and proliferate through the secondary nucleation of new fibrils on the surface of existing ones (Hadi Alijanvand *et al.*, 2021). While nucleation increases the number of fibrils, autocatalytic growth increases the size of the fibrils (Bentea *et al.*, 2017), which can reach >1000 nm in length (Gade Malmos *et al.*, 2017). Nucleation and growth (NAG) are also present during the formation of aberrant and functional membraneless organelles through the LLPS of proteins and nucleic acids (Strom and Brangwynne, 2019; Darling and Shorter, 2021). Distinctively from the highly ordered structure of amyloid fibrils, biomolecular condensates formed by LLPS present liquid-like properties such as the spherical shape, the ability to coalesce and deform under shear flow, and the short recovery times from fluorescence recovery after photobleaching (FRAP) experiments (Garaizar *et al.*, 2022).

A good understanding of the NAG kinetics is important for knowing how candidate drugs modulate the phase separation rates and the size distributions of protein assemblies (Shimobayashi *et al.*, 2021; Kar *et al.*, 2022). For the kinetic analysis of amyloid aggregation, the mass concentration of protein assemblies ([*M*]) is measured over time (*t*) and subject to a min-max normalization (Figure 1, A and B). The normalized curves are then fitted by a physical model to yield rate constants for the primary nucleation, growth, and secondary nucleation steps (Knowles *et al.*, 2009; Linse, 2019; Sárkány *et al.*, 2023). Mass-based progress curves are readily measured using amyloid dyes such as Thioflavin-T (ThT) or by estimating the mass of aggregates from the depleted concentration of monomer in solution (Gade Malmos *et al.*, 2017; Xue *et al.*, 2017; Bellomo *et al.*, 2018). To test whether the fitted model does provide reliable mechanistic information about the NAG steps, complementary measurements of size-based progress curves and particle size distributions are needed using dynamic light scattering (DLS) and advanced microscopy techniques (Silva *et al.*, 2017, 2018), for example. Size-based progress curves are, however, less commonly studied because the size distributions of amyloid fibrils can be too complex to be deconvoluted by conventional light scattering or image analysis techniques.

The intense research presently devoted to the LLPS of biomolecular condensates instigates an interest in size-based progress curves in which a characteristic droplet size (*R*) is monitored over time (Figure 1C) either in the test tube or in cells (Berry *et al.*, 2015; Matsarskaia *et al.*, 2019). Typically, these curves are represented in a log-log scale to extract power-law scaling exponents: the scaling exponents take values of ~1/2 in purely diffusional regimes and of ~1 in kinetic regimes determined by surface attachment (Berry *et al.*, 2018); spontaneous LLPS by spinodal decomposition (without nucleation) is characterized by a scaling exponent of ~1/3 (Matsarskaia *et al.*, 2019). More in-depth mechanistic investigations have been hampered by the absence of phase separation models able to describe the *R* versus *t* dependencies and, above all, explain the recurrent observation of sharp NAG of a few droplets with large and uniform sizes (Mohapatra *et al.*, 2017; Sárkány *et al.*, 2023).

These challenges were recently addressed with the proposal of a "general" NAG model that, besides describing the kinetics of protein crystallization and aggregation (Crespo *et al.*, 2012), also predicts the evolution of particle size distributions of amyloid fibrils (Silva *et al.*, 2017) and liquid droplets (Sárkány *et al.*, 2023). The initial motivation behind the general NAG model was to study protein aggregation as a phase separation process and not as a protein polymerization reaction (Crespo *et al.*, 2012). Analogous to crystallization processes, the elementary rate equations of primary nucleation, growth, and secondary nucleation were expressed as a function of supersaturation thus helping to elucidate the important role of protein solubility as a thermodynamic determinant of protein aggregation (Crespo *et al.*, 2012). Moreover, the phase separation view accounted for hyperbolic-shaped aggregation curves measured in the absence of preformed aggregates, which was a fundamental advance relative to the existing physical models (Crespo *et al.*, 2012). More recently, the crystallization-like model was developed to include the effect of surface tension on protein aggregation and LLPS (Sárkány *et al.*, 2023).

Here we present the NAGPKin web server for automated quantification NAG parameters from the kinetics of protein phase

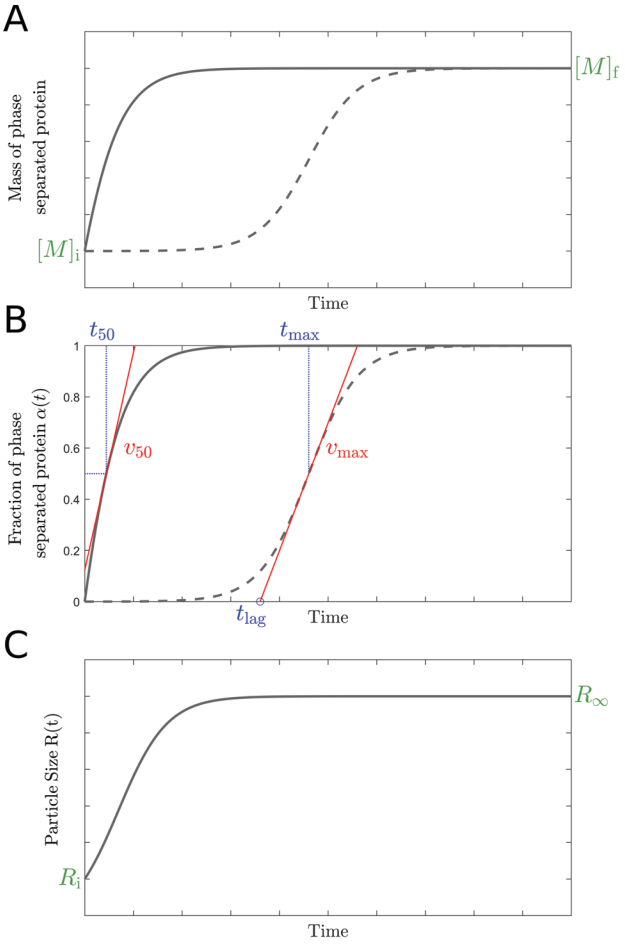

FIGURE 1. Schematic representations of mass-based and size-based progress curves. (A) Mass-based progress curves monitor the mass of protein assemblies either directly or indirectly (e.g., through fluorescence measurements) from an initial point with signal $[M]_i$ to a final point $[M]_f$. These curves can have hyperbolic (solid line) or sigmoidal (dashed line) shapes. (B) The same curves are represented in normalized units of reaction conversion (α). The maximal growth rate (v_{max}, red line), the time required to reach v_{max} (t_{max}), the half-life coordinates t_{50} and v_{50} (red line), and the duration of the lag phase (t_{lag}) are examples of commonly used kinetic measurables. (C) Size-based progress curves monitor the increase in the size of protein assemblies from an initial value R_i to a final value R_∞.

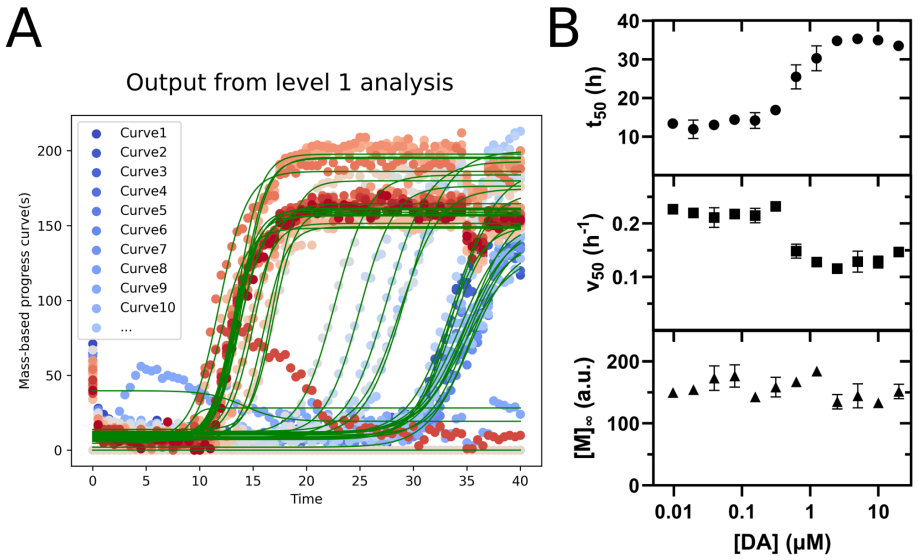

FIGURE 2. Systematic analysis of 36 aggregation curves of Atx3-77Q measured by Figueiredo *et al.* in the presence of different concentrations of DA (Figueiredo *et al.*, 2023). (A) NAGPKin's output graph with the uploaded (symbols) and fitted (line) curves. (B) The values of t_{50} (top), v_{50} (middle), and $[M]_\infty$ (bottom) produced by NAGPKin are represented as a function of DA concentration. Symbols and error bars: mean and SD values. In two of the conditions studied (5.0 and 0.31 μM DA), one replicate is excluded.

separation. NAGPKin requires minimal intervention from the users, who, after uploading their raw data, follow an informative tour through the steps of chemical kinetic analysis. A final report provides quantitative information about the general NAG parameters that best fit the measured mass or size progress curves. To improve the quality of the predictions, global fittings to multiple curves obtained at different protein concentrations are encouraged. The application examples that follow illustrate the usefulness of NAGPKin for drug discovery, standardized kinetic data reporting and, on the whole, for understanding the molecular mechanisms of protein phase separation.

RESULTS AND DISCUSSION

NAGPKin provides different types of information depending if mass-based or size-based curves are uploaded (Supplemental Table S1). In the first level of analysis, "descriptive" kinetic measurables such as the half-life coordinates are determined by a systematic, user-independent method. In the subsequent levels, the phase separation mechanism and the elementary steps of primary nucleation, secondary nucleation, and growth are characterized quantitatively from the kinetic analysis of scaling laws (level 2) or through global numerical fitting (level 3) using the physical NAG model of Sárkány *et al.*, 2023.

As an application example of level 1 analysis, we study how different concentrations of dopamine (DA) affect the aggregation of an Atx3 variant containing a pathological polyglutamine (polyQ) tract size of 77Q (Atx3-77Q). The curves of ThT fluorescence increase were previously measured by us in triplicate experiments for 12 serially diluted concentrations of DA (Figueiredo *et al.*, 2023). The NAGPKin's template was filled with the values of time (first column) and ThT fluorescence (36 subsequent columns) and then saved keeping the .ods extension. The used concentration of Atx3-77Q (3 μM) is not a piece of relevant information if only level 1 analysis is performed. So, for descriptive-only studies, the row destined for protein concentration values can be left with blank cells. Because NAGPKin does not analyze the effect of modulators on protein phase separation, no information about the adopted DA concentrations is requested in the spreadsheet template. After the .ods file is uploaded, the initial data processing produces a graphic with experimental and numerically fitted curves (Figure 2A) and two .csv files summarizing the fitted kinetic measurables and progress curves. The user can download these files and use the data to customize new graphs or to find dose–response relationships (Figure 2B). In general, kinetic modulators of protein aggregation influence the values of t_{50} (via the primary nucleation and autocatalytic steps) and v_{50} (via the autocatalytic steps), whereas thermodynamic modulators influence the values of $[M]_\infty$ (via the effect on protein solubility) (Sárkány *et al.*, 2019, 2023). Because the uploaded data were obtained at a fixed protein concentration (c_0), moving forward to the second level of analysis is not possible – the minimum requirement is four different c_0 values. A global fit to all curves to estimate a single set of NAG parameters (level 3 analysis) is also undue because changeable inhibitor concentrations are adopted.

By default, NAGPKin assumes the submitted data to consist of mass-based progress curves. For the analysis of size-based progress curves, the "size-based" option in the lefthand expander should be selected at any point, before or after file upload. To exemplify the fundamental interpretation of particle size measurements, we used DLS data obtained during the LLPS of the low-complexity domain (LCD) of TAR DNA-binding protein 43 (TDP-43) (Van Lindt *et al.*, 2021), and during the aggregation of nonexpanded Atx3 (Silva *et al.*, 2017). The evolution of the hydrodynamic radii follows dissimilar trends in each of the cases, either increasing with time for liquid droplets of TDP-43-LCD (Figure 3A) or decreasing with time for Atx3 fibrils (Figure 3B). Importantly, the two types of curves are resolved by NAGPKin both in level 1 and level 3 of analysis. Level 2 analysis is again not performed because the required c_0-scaling data is not included in the input file. The final reports (Supplemental Figure S1) disclose that the importance of secondary nucleation relative to the other autocatalytic step (growth) is minimal for TDP-43-LCD (~0%) and maximal for Atx3 (~100%). Because the secondary nucleus is smaller than the primary nucleus, systems dominated by secondary nucleation exhibit decreasing mean sizes over time, even in the absence of significant breakage/fragmentation, as confirmed for Atx3 (Silva *et al.*, 2017). As we move from level 1 to level 3, the quality of the numerical fittings improves for TDP-43-LCD and worsens for Atx3. The explanation for this fact can also be found in the NAGPKin's report (Supplemental Figure S1): in the case of TDP-43-LCD, the underperformance of the numerical fittings in level 1 is due to STEs that play a significant role in the nucleation of these liquid droplets (25% importance) but are not considered in the first stage of analysis. Level 3 is a more constrained analysis because it takes into account the effects of protein concentration and solubility. Therefore, level 3 is more trustworthy than level 1 even if the quality of the numerical result is worse. For reproducible assays, imperfect goodness-of-fit statistics are suggestive of the occurrence of parallel phenomena, which, in the case of Atx3 are indeed present through nonamyloid aggregation pathways (Silva *et al.*, 2017).

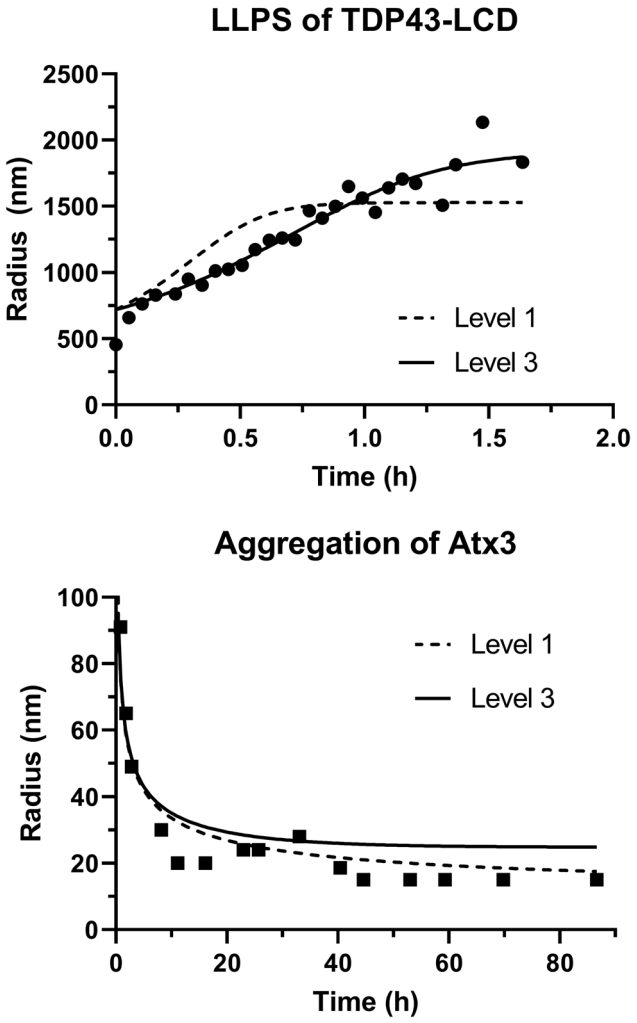

FIGURE 3. Comparison between measured and theoretical size-based curves. (A and B) Evolution of the hydrodynamic radius measured by DLS (symbols) and numerically fitted after level 1 (dashed line) and level 3 (solid line) analyses by NAGPKin. Experimental data obtained by (A) Van Lindt *et al.* during LLPS of TDP-43-LCD in the absence of NaCl (Van Lindt *et al.*, 2021) and (B) Silva *et al.* during the formation of Atx3 fibrils (Silva *et al.*, 2017).

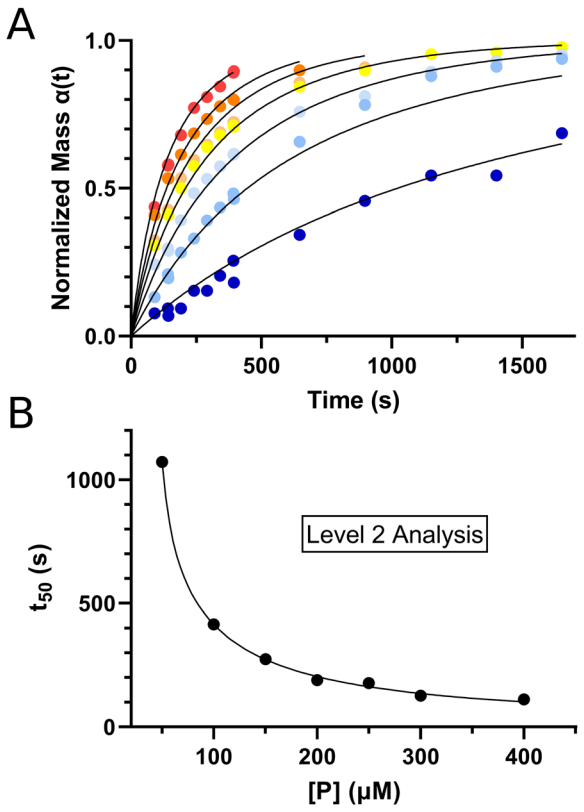

FIGURE 4. NAGPKin analysis of the aggregation kinetics of mTTR. (A) Symbols: normalized values of ThT fluorescence increase measured by Hurshamnn *et al.* for mTTR concentrations of (from top to bottom) 400, 300, 250, 200, 150, 100, and 50 μg/mL (Hurshman *et al.*, 2004). Lines. NAGPKin's global fit (level 3). (B) Representation of the measured (symbols) and numerically fitted (line) dependencies of t_{50} on c_0 (level 2).

As an illustrative example of all levels of NAGPKin analysis, we use mass-based progress curves measured by Hurshman *et al.* (Hurshman *et al.*, 2004) for different concentrations of monomeric transthyretin (mTTR) (Figure 4). The ThT fluorescence data obtained by those authors were previously digitized and normalized by us (Sárkány *et al.*, 2023) and are made available to NAGPKin users as a workable "prefilled example" for which no file upload is required. The kinetic measurables t_{max}, v_{max}, or t_{lag} are not quantified by NAGPKin because only t_{50} and v_{50} are suitable to describe sigmoidal, but also hyperbolic curves (Figure 1), such as those obtained during mTTR aggregation. The selection between the results from levels 2 or 3 is automatic and relies on the values of r-squared (r^2) obtained using t_{50} vs. c_0 scaling properties (level 2) or the the global numerical fitting (level 3). The t_{50} versus c_0 analysis will be preferred over the global fit analysis only if the difference in the values of r^2 is > 0.03 in favor of level 2. Values of $r^2 > 0.98$ are obtained in the example of mTTR for both levels of analysis (Supplemental Figure S2). Even in such cases of high r^2, the occurrence of (undetected) parallel processes cannot be excluded; in fact, significant amounts of TTR dimers were identified in the initial sample of mTTR (Hurshman *et al.*, 2004), an observation that may explain differences between the values of fitted parameters in levels 2 and 3.

DISCUSSION

NAGPKin is the first computational tool dedicated to the kinetic analysis of phase separation processes. To the best of our knowledge, none of the practical examples here investigated could have been processed by alternative software for automated kinetic analysis. In the first of the given examples, the inhibition of protein aggregation was characterized through the quantification of both kinetic (t_{50} and v_{50}) and thermodynamic measurables ($[M]_\infty$). Detailed mechanistic information could be extracted from size-based progress curves obtained during the LLPS of TDP-43-LCD and the aggregation of Atx3. Similar mechanistic insights were then obtained from hyperbolic curves of ThT fluorescence increase and chemical kinetic properties of mTTR aggregation. We expect that NAGPKin will be useful for processing data of condensate size increase measured in cells (Zwicker *et al.*, 2014; Snead *et al.*, 2022), especially during the initial phase of coalescence-free self-assembly (Berry *et al.*, 2018; Matsarskaia *et al.*, 2019). Additional applications such as the study of NAG during protein and industrial crystallization processes (Arruda *et al.*, 2023) will be investigated in the future.

When mass-based progress curves showing a sigmoidal shape are obtained, fundamental studies can alternatively be performed using the web tools AmyloFit (Meisl *et al.*, 2016) and Kfits

(Rimon and Reichmann, 2018). These tools are based on polymerization principles that are distinct from phase separation principles of protein self-assembly. In particular, the used protein polymerization model was originally developed by Oosawa and Kasai to describe actin-like polymerization (Oosawa and Kasai, 1962) and then expanded to include secondary nucleation steps and Michaelis–Menten-like kinetics for elongation and secondary nucleation (Knowles *et al.*, 2009; Meisl *et al.*, 2016). Although the analogies with "a gas-liquid transition" or "condensation phenomenon" are used since the original work by Oosawa and Kasai (Oosawa and Kasai, 1962), the "nucleation" and elongation steps during actin-like polymerization do not follow phase equilibrium principles. For this reason, the AmyloFit and Kfits tools cannot be used to process 1) mass-based, hyperbolic curves, 2) LLPS data, or 3) the modulation of phase equilibrium parameters such as protein solubility. Moreover, NAGPKin is the only web server analyzing size-based progress curves, whereas the general NAG model (on which NAGPKin is based) is the first to rationalize the sharp formation of large assemblies with Gaussian size distributions (Sárkány *et al.*, 2023).

The NAGPKin's report may reveal unsatisfactory fitting results, even when high-quality reproducible data are subject to analysis. Those situations likely result from parallel phenomena such as off-pathway aggregation, coalescence, and mixed liquid-liquid and liquid-solid nucleation pathways, whose occurrence can be checked by complementary methods for monitoring phase separation (Crespo *et al.*, 2016, 2017; Ferreira *et al.*, 2017; Lee *et al.*, 2018; Babinchak and Surewicz, 2020; Martins *et al.*, 2020; Garaizar *et al.*, 2022). To avoid the problem of overparameterization, our algorithm does not include additional parameters describing parallel microscopic events whose number and nature cannot be known a priori. Instead, the report issued by NAGPKin identifies possible reasons for the poor fittings and suggests ways to address this problem. It may happen, for example, that the scaling of t_{50} with c_0 follows the expected trend but the global fitting results are poor. In cases such as this, the NAGPKin report would alert for the possible occurrence of off-pathway aggregation, which is known to severely affect the shape of the kinetic curves but has little effect on the values of t_{50} (Crespo *et al.*, 2016).

The predictions from the general NAG model are testable. For example, the NAG rate constants fitted to size-based progress curves should also describe mass-based progress curves (Silva *et al.*, 2017, 2018). Similarly, the fitted NAG rate constants can be used to predict the evolution of size distributions of protein droplets (Sárkány *et al.*, 2023). It is therefore advised that, when possible, results obtained using mass- and size-based methods are submitted to NAGPKin for consistency analysis.

MATERIALS AND METHODS

Overview

NAGPKin performs automated kinetic analysis with different levels of detail depending on what type of experimental data is uploaded (Figure 5). In level 1a, the half-life coordinates t_{50} and v_{50} and the total mass increase $[M]_\infty = [M]_f - [M]_i$ are estimated from mass-based progress curves. If, instead, size-based curves are uploaded, the fitted values of the primary nucleus size (R_1) and final size (R_∞) are provided together with an indicative estimate of the half-life time t_{50} expected for mass-based progress curves (level 1b). In levels 2 and 3, NAG rate constants are obtained following different methods independently of what type of progress curves are used as input. The NAGPKin algorithm automatically selects what method is more reliable and what rate constants should be included in the final report. Level 2 is based on the scaling properties of measurable t_{50} with protein concentration c_0. As such, level 2 analysis is only possible if four or more progress curves are measured at different c_0 values. Level 3 analysis includes a global fit of the model equations to the experimental data. Although this can be done using a single progress curve as the input, the simultaneous analysis of multiple curves decreases the risk of numerical convergence to local minima rather than to the global optimum. From the comparison of the level 2 and level 3 results, the NAGPKin algorithm can identify the possible occurrence of parallel self-assembly pathways or suggest ways to improve the quality of the kinetic analysis.

Input file

A single spreadsheet file containing mass-based or size-based progress curves is used as the input for the NAGPKin web server. An example of this file is available for download at the beginning of the analysis and can be modified to remove the exemplifying data and add the user's data. The OpenDocument Spreadsheet (.ods) format of this file is compatible with applications such as Microsoft Excel or LibreOffice and should be kept upon saving the file. The first rows of the template contain instructional information that can be left unmodified. The experimental data should be added as indicated by the instructions, with the values of time over the blue column, the values of protein concentration over the yellow row, and the values of protein mass or size over the green cells (Figure 5, left). It is assumed that all curves are obtained simultaneously and, therefore, the values in the first column (time) are common to all the individual contours. No distinction between mass or size progress curves is required at this point. Data obtained during mass-based experiments (such as turbidimetry and ThT fluorescence measurements) or size-based experiments (such as average size measurements) can be copied and pasted directly from the raw data files. Each curve may contain blank cells corresponding to, for example, experimental outliers or noisy data. After saving the modified .ods template, the user can upload it into NAGPKin by dragging and dropping the file or by file browsing.

Algorithm

The NAGPKin algorithm is based on the general NAG model of phase separation (Crespo *et al.*, 2012; Silva *et al.*, 2017; Sárkány *et al.*, 2023). In this model, the principal moments of the size distribution (Eq. 1 in Supplemental Figure S3) can be numerically solved to obtain the evolution with time of the mass concentration $[M](t)$, number concentration $[P](t)$, and mean size (in number of protein units) $N(t)$ of new assemblies. Through mathematical programming, it is also possible to fit the master equation to size-based progress curves and estimate the rate constants for primary nucleation (k_n), growth (k_+) and secondary nucleation (k_2). While the growth and secondary nucleation steps affect particle size in drastically different ways, their effect on the total mass of assemblies is difficult to discriminate as k_+ and k_2 are both elementary rate constants describing a supersaturation-dependent, autocatalytic step (Crespo *et al.*, 2012; Silva *et al.*, 2017; Sárkány *et al.*, 2023). This means that numerical fittings to mass-based progress curves yield an autocatalytic rate constant $k_\alpha = k_+ + k_2$ and not the individual estimates of k_+ and k_2. The importance of the primary nucleation rate relative to the autocatalytic rates is given by the parameter $k_b = k_n/k_\alpha$. In the NAGPKin algorithm, the exact solution of model equations is only used in the last stage (level 3) of the analysis (Figure 5, right). Before, simplified model equations (level 1) and, if possible, NAG scaling laws (level 2) are fitted to the experimental data to obtain complementary mechanistic information and good initial guesses for the global fitting. The simplified

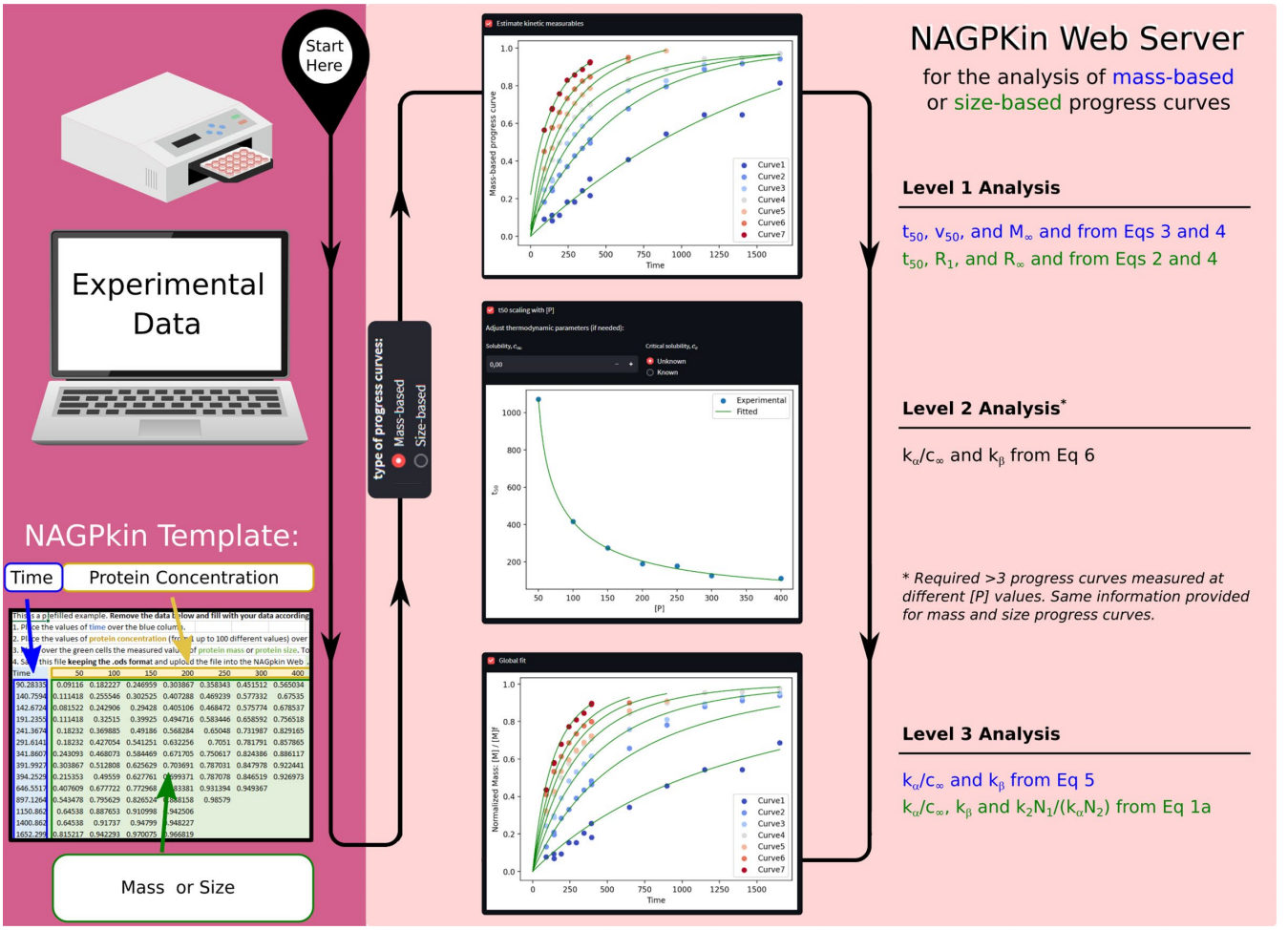

FIGURE 5. Overview of the NAGPKin web server. Left: the raw data measured during protein phase separation experiments can be copied and pasted into the NAGPKin's spreadsheet file. Right: NAGPKin's workflow. The user selects whether the uploaded data correspond to mass or size progress curves; then, automatic curve fitting analyses are performed to provide level 1, 2, and 3 information. The equations used for parameter estimation are numbered as in Supplemental Figure S3.

equations used in level 1 are only valid in the absence of surface tension effects (STEs) (Sárkány *et al.*, 2023). Since the simplified equations produce robust numerical fittings both in the presence and absence of STEs (Sárkány *et al.*, 2023), they are used to obtain descriptive measurables but not in-depth mechanistic information. STEs change the protein solubility of small, curved assemblies from the saturation value c_∞ to a critical value c_c, with $c_c \geq c_\infty$. Both c_∞ and c_c can be obtained experimentally, for example, from the linear plot of the total mass of assemblies versus protein concentration (Sárkány *et al.*, 2023). By default, NAGPKin assumes very low solubility ($c_\infty << c_0$), while the ratio c_c/c_∞ is a fitted parameter. However, the user may choose to set c_∞ and c_c to fixed values known experimentally. Level 2 analysis is only performed if four or more progress curves are available for different c_0 values. Global fit analysis (level 3) of mass-based progress curves produces estimates of k_α/c_∞ and $k_\beta = k_b\,(c_\infty/c_c)^2$; in addition, the combined parameter $k_2N_1/(k_\alpha N_2)$ can be fitted during the analysis of size progress curves. Here, N_1 and N_2 are the number of protein units constituting primary and secondary nuclei, respectively. A ratio of $N_1/N_2 = (R_1/R_2)^3 = 223$ was previously estimated for amyloid fibrils of Atx3 (Silva *et al.*, 2018). Whenever possible, a final report is generated summarizing the values and physical meaning of the NAG rate constants, quantifying the importance of STEs, and evaluating the possible occurrence of parallel processes such as off-pathway aggregation and coalescence. The selection between NAG rate constants obtained during level 2 and 3 analysis is based on goodness-of-fit statistics; in the case of equally good statistics, the global fit analysis (level 3) is preferred.

Implementation

NAGPKin's code was written with the programming language Python (version 3.9.16) and the web tool is made using Streamlit (www.streamlit.io). The up-to-date codes and new releases are available on GitHub (https://github.com/pmartins2106/NAGpkin/). The GNU general public license is also added to the source package on GitHub.

CONCLUSIONS

NAGPKin analyzes mass-based progress curves (exhibiting both hyperbolic and sigmoidal shapes) as well as size-based progress curves. Descriptive information but also detailed mechanistic insights are provided into phase separation dynamics. The uniqueness of the NAGPKin toolkit is demonstrated in a series of application examples, from drug discovery to fundamental chemical kinetic analysis. The user-friendly interface of NAGPKin makes it an

accessible tool to both experimentalists and bioinformaticians with or without previous experience in kinetic data modeling.

ACKNOWLEDGMENTS

We thank Professor José Paulo Leal, Department of Computer Science – Faculdade de Ciências da Universidade do Porto, for his invaluable help during the Implementation of NAGPKin. This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 952334 (PhasAGE). This research was funded by the Portuguese Foundation for Science and Technology (FCT) in the framework of project PTDC/QUI-COL/2444/2021.

REFERENCES

Arruda RJ, Cally PAJ, Wylie A, Shah N, Joel I, Leff ZA, Clark A, Fountain G, Neves L, Kratz J, *et al*. (2023). Automated and material-sparing workflow for the measurement of crystal nucleation and growth kinetics. Cryst Growth Des 23, 3845–3861.

Babinchak WM, Surewicz WK (2020). Liquid–liquid phase separation and its mechanistic role in pathological protein aggregation. J Mol Biol 432, 1910–1925.

Bellomo G, Bologna S, Gonnelli L, Ravera E, Fragai M, Lelli M, Luchinat C (2018). Aggregation kinetics of the Aβ1–40 peptide monitored by NMR. Chem Commun 54, 7601–7604.

Bentea L, Watzky MA, Finke RG (2017). Sigmoidal nucleation and growth curves across nature fit by the Finke–Watzky model of slow continuous nucleation and autocatalytic growth: explicit formulas for the lag and growth times plus other key insights. J Phys Chem C 121, 5302–5312.

Berry J, Brangwynne CP, Haataja M (2018). Physical principles of intracellular organization via active and passive phase transitions. Rep Prog Phys 81, 046601.

Berry J, Weber SC, Vaidya N, Haataja M, Brangwynne CP (2015). RNA transcription modulates phase transition-driven nuclear body assembly. P Natl Acad Sci USA 112, E5237–E5245.

Crespo R, Rocha FA, Damas AM, Martins PM (2012). A generic crystallization-like model that describes the kinetics of amyloid fibril formation. J Biol Chem 287, 30585–30594.

Crespo R, Villar-Alvarez E, Taboada P, Rocha FA, Damas AM, Martins PM (2016). What can the kinetics of amyloid fibril formation tell about off-pathway aggregation? J Biol Chem 291, 2018–2032.

Crespo R, Villar-Alvarez E, Taboada P, Rocha FA, Damas AM, Martins PM (2017). Insoluble off-pathway aggregates as crowding agents during amyloid fibril formation. J Phys Chem B 121, 2288–2298.

Darling AL, Shorter J (2021). Combating deleterious phase transitions in neurodegenerative disease. Biochim Biophys Acta Mol Cell Res 1868, 118984.

Ferreira C, Barbosa S, Taboada P, Rocha FA, Damas AM, Martins PM (2017). The nucleation of protein crystals as a race against time with on- and off-pathways. J Appl Crystallogr 50, 1056–1065.

Figueiredo F, Sárkány Z, Silva A, Vilasboas-Campos D, Maciel P, Teixeira-Castro A, Martins PM, Macedo-Ribeiro S (2023). Drug repurposing of dopaminergic drugs to inhibit ataxin-3 aggregation. Biomed Pharmacother 165, 115258.

Gade Malmos K, Blancas-Mejia LM, Weber B, Buchner J, Ramirez-Alvarado M, Naiki H, Otzen D (2017). ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. Amyloid 24, 1–16.

Garaizar A, Espinosa JR, Joseph JA, Collepardo-Guevara R (2022). Kinetic interplay between droplet maturation and coalescence modulates shape of aged protein condensates. Sci Rep 12, 4390.

Hadi Alijanvand S, Peduzzo A, Buell AK (2021). Secondary nucleation and the conservation of structural characteristics of amyloid fibril strains. Front Mol Biosci 8, 669994.

Hurshman AR, White JT, Powers ET, Kelly JW (2004). Transthyretin aggregation under partially denaturing conditions is a downhill polymerization. Biochemistry 43, 7365–7381.

Kar M, Dar F, Welsh TJ, Vogel LT, Kühnemuth R, Majumdar A, Krainer G, Franzmann TM, Alberti S, Seidel CAM, *et al*. (2022). Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. Proc Natl Acad Sci 119, e2202222119.

Knowles TPJ, Waudby CA, Devlin GL, Cohen SIA, Aguzzi A, Vendruscolo M, Terentjev EM, Welland ME, Dobson CM (2009). An analytical solution to the kinetics of breakable filament assembly. Science 326, 1533–1537.

Lee M-C, Yu W-C, Shih Y-H, Chen C-Y, Guo Z-H, Huang S-J, Chan JCC, Chen Y-R (2018). Zinc ion rapidly induces toxic, off-pathway amyloid-β oligomers distinct from amyloid-β derived diffusible ligands in Alzheimer's disease. Sci Rep 8, 4772.

Linse S (2019). Mechanism of amyloid protein aggregation and the role of inhibitors. Pure Appl Chem 91, 211–229.

Martins PM, Navarro S, Silva A, Pinto MF, Sárkány Z, Figueiredo F, Pereira PJB, Pinheiro F, Bednarikova Z, Burdukiewicz M, *et al*. (2020). MIRRAGGE – Minimum information required for reproducible AGGregation experiments. Front Mol Neurosci 13, 582488.

Matsarskaia O, Da Vela S, Mariani A, Fu Z, Zhang F, Schreiber F (2019). Phase-separation kinetics in protein–salt mixtures with compositionally tuned interactions. J Phys Chem B 123, 1913–1919.

Meisl G, Kirkegaard JB, Arosio P, Michaels TCT, Vendruscolo M, Dobson CM, Linse S, Knowles TPJ (2016). Molecular mechanisms of protein aggregation from global fitting of kinetic models. Nat Protoc 11, 252–272.

Mohapatra L, Lagny TJ, Harbage D, Jelenkovic PR, Kondev J (2017). The limiting-pool mechanism fails to control the size of multiple organelles. Cell Syst 4, 559–567.

Oosawa F, Kasai M (1962). A theory of linear and helical aggregations of macromolecules. J Mol Biol 4, 10–21.

Rimon O, Reichmann D (2018). *Kfits* : a software framework for fitting and cleaning outliers in kinetic measurements. Bioinformatics 34, 129–130.

Sárkány Z, Rocha F, Bratek-Skicki A, Tompa P, Macedo-Ribeiro S, Martins PM (2023). Quantification of surface tension effects and nucleation-and-growth rates during self-assembly of biological condensates. Adv Sci 10, e2301501.

Sárkány Z, Rocha F, Damas AM, Macedo-Ribeiro S, Martins PM (2019). Chemical kinetic strategies for high-throughput screening of protein aggregation modulators. Chem Asian J 14, 500–508.

Shimobayashi SF, Ronceray P, Sanders DW, Haataja MP, Brangwynne CP (2021). Nucleation landscape of biomolecular condensates. Nature 599, 503–506.

Silva A, Almeida B, Fraga JS, Taboada P, Martins PM, Macedo-Ribeiro S (2017). Distribution of amyloid-like and oligomeric species from protein aggregation kinetics. Angew Chem Int Ed 56, 14042–14045.

Silva A, Sárkány Z, Fraga J, Taboada P, Macedo-Ribeiro S, Martins P (2018). Probing the occurrence of soluble oligomers through Amyloid Aggregation Scaling Laws. Biomolecules 8, 108.

Snead WT, Jalihal AP, Gerbich TM, Seim I, Hu Z, Gladfelter AS (2022). Membrane surfaces regulate assembly of ribonucleoprotein condensates. Nat Cell Biol 24, 461–470.

Strom AR, Brangwynne CP (2019). The liquid nucleome – phase transitions in the nucleus at a glance. J Cell Sci 132, jcs235093.

Van Lindt J, Bratek-Skicki A, Nguyen PN, Pakravan D, Durán-Armenta LF, Tantos A, Pancsa R, Van Den Bosch L, Maes D, Tompa P (2021). A generic approach to study the kinetics of liquid–liquid phase separation under near-native conditions. Commun Biol 4, 77.

Xue C, Lin TY, Chang D, Guo Z (2017). Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. R Soc Open Sci 4, 160696.

Zwicker D, Decker M, Jaensch S, Hyman AA, Jülicher F (2014). Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. Proc Natl Acad Sci USA 111, E2636–E2645.