Modelling Early Events In Amyloid Aggregation

Alexander John Dear

Trinity College

University of Cambridge

Department of Chemistry

This dissertation is submitted for the degree of *Doctor of Philosophy*

July 2018

Supervisor: Prof Tuomas P. J. Knowles

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Introduction and Acknowledgements and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution, except as declared in the Acknowledgements.

This dissertation does not exceed the word limit of 60,000 words prescribed by the Degree Committee, including summary/abstract, tables, and footnotes, but excluding table of contents, diagrams, figure captions, bibliography, appendices and acknowledgements.

Abstract

The formation of macroscopic biofilaments from monomeric peptides is a process central to both normal and disease biology, and a very active area of research. An example of particular significance is the formation of amyloid plaques from normally soluble proteins in several neurodegenerative conditions including Alzheimer's disease. Traditionally, attention has focused on the proliferation of macroscopic quantities of biofilaments, and experimentalists have relied on bulk measurements to quantify this. By contrast, the initiation of the aggregation reaction has been comparatively understudied, despite the involvement of oligomeric intermediates that are considered to be the key pathogenic agents in many amyloidogenic diseases. This is due in large part to the key quantities measured in bulk experiments being insensitive to the details of the initiation step, and to the difficulties faced until recently in detecting comparatively rare oligomeric species. In this thesis I describe a number of theoretical approaches towards filling this crucial gap in our knowledge. These approaches fall into 3 main themes: modelling the early stages of bulk aggregation kinetics in greater detail; developing general kinetic theories of filament formation via initial oligomeric intermediates; and investigating heterogeneity within oligomer populations using equilibrium statistical mechanical modelling. In several instances I have already been able to apply these theories to experimental results, furthering our practical understanding of these processes in vitro. A final chapter is devoted to a highly general and insightful theoretical method for developing analytical solutions for the kinetics of a wide range of self-assembly phenomena. Although not focused on early events in amyloid aggregation, it is already finding application in this field, as in many others.

Acknowledgements

I have had an unusually collaborative PhD, and so owe a lot of thanks to a lot of people. I mention here only those whose influence has been particularly large. Firstly, my principal experimental collaborators:

- Jie Yang, Si Wu, and Sarah Perrett from the Perrett Lab in Beijing, for their high-quality data on oligomer kinetics, their ingenious experiment design, and their conception of the Ure2 oligomers project;
- Marija Iljina, Suman De, Patrick Flagmeier, David Klenerman and the entire Klenerman group for initiating fruitful collaborations regarding oligomer thermodynamics, and providing both high-quality data and ideas;
- Magnus Kjaergaard and Franziska Kundel, both for their high-quality data on Tau aggregation and for Magnus' conception of the tau oligomers project;
- Matthew Horrocks and Laura Tosatto for kindly inviting me to participate in their projects developing and applying high-throughput single-molecule FRET methods;
- Jonathan Taylor, Maria Andreasen, Steve Matthews and the entire Matthews Lab for kindly inviting me to participate in their project on bacterial amyloid formation.

Secondly, my technical collaborators:

- Georg Meisl, for his data-fitting expertise, his frequent coding assistance, some invaluable discussions around oligomer kinetics and many informal chats;
- Andela Saric, for thought-provoking Monte-Carlo simulations, discussions and feedback;
- Julius Kirkegaard, whose Gillespie simulations helped me discover and rectify a crucial mathematical error both in my work and in the established literature.

Thirdly, my advisory collaborators:

- David Klenerman, for being very generous with his time in drafting papers, for many very useful suggestions on modelling, and for feedback on manuscripts and on presentations;
- Daan Frenkel, for helping un-confuse me on statistical mechanical issues that were subtle for me but probably simple for him;
- Chris Dobson, for some insightful suggestions and feedback on manuscripts;
- and Lakshminarayanan Mahadevan, for generously hosting a productive research visit to Harvard.

I owe an especially big debt of thanks to three people, without whom I would have had a very different and probably much less successful PhD.

Gonzalo Garcia deserves primary credit for recruiting me to the Knowles Group, whose website at the time made almost no mention of theoretical research and had thus initially escaped my notice. I couldn't imagine a better host research group for my PhD, so I am profoundly grateful for his advice to join. Moreover, by developing some of the first quantitative models of amyloid oligomer chemistry during his PhD, Gonzalo inspired a lot of interest from experimentalists in this field, which I was later able to take advantage of by forming many productive experimental collaborations. He himself also made some of the introductions that made these collaborations possible. Finally, having left science at the end of his PhD, he invited me to complete his project on Aβ co-oligomers, which resulted in a joint-first authorship publication for me, and which forms the basis of one of the chapters of my thesis. He deserves primary credit for the theory developed in this project; my contributions were primarily in the form of data analysis & interpretation, coding, minor changes to the theoretical modelling approach, and calculations justifying this new approach. An earlier version of this project, with different data, results and modelling approach, were submitted as a part of his PhD thesis.

At the commencement of my studies I asked the other theorists in my research group whether they would like any help with any of their active research projects, reasoning that this would be a more effective and interesting way to get up to speed with the latest developments in the field than the traditional approach of a literature review. In response, Thomas C. T. Michaels invited me to participate in the project on stochastic filamentous growth described in the second chapter of this thesis; I am very grateful to him for a perfect start to my PhD. Our working styles complemented one another very well, and within 3 months we had a draft paper which was later published in *Phys Rev Lett.* Although the project was led by Thomas, it has been included in both of our PhD theses since we share roughly equal credit for the theoretical results and analysis. This was only the beginning of a highly productive and enjoyable ongoing collaboration, which has resulted in 14 completed manuscripts to date. In addition, Thomas Michaels has been unfailingly generous with his advice on all matters academic and professional, and has been a source of inspiration to me over the past 4 years. I look forwards to many more interesting joint projects in the future. Doubtless my academic career so far would have been significantly less productive had we never met.

Finally and most importantly, I owe a huge debt of thanks to my supervisor Tuomas Knowles, whose research group I feel very lucky to be in. He has provided vital input into several of my research projects, and has provided sage advice on direction, and on drafting, in many more. Moreover, he has allowed me an extraordinary amount of freedom in my research priorities and in my working habits, which I believe has paid off as evidenced by my list of publications! He has finally also been incredibly supportive of my career in

general. It is in large part due to his attitude and abilities as a supervisor that I have so thoroughly enjoyed my PhD.

For their financial support, I thank the Schiff Foundation, Trinity College, the Department of Chemistry, Tuomas P. J. Knowles and the Rouse Ball / Eddington Research Fund.

Contents

Part I

Preliminary Material

Chapter 1

List of Publications

Below are listed in reverse chronological order my 15 completed manuscripts, 11 of which have been published or accepted for publication in scientific journals. The remainder have either been submitted or are expected to be submitted before my viva. For brevity I neglect the 4 remaining manuscripts that are likely to be submitted by this date but on which I am not first author.

"Amyloid oligomer classification using chemical kinetics"
A. J. Dear, T. C. T. Michaels, G. Meisl, A. Saric, M. Kjaergaard, and T. P. J. Knowles
Manuscript under preparation for submission to Nat. Struct. Mol. Biol.
This manuscript is reproduced in Chapter 6; its Supplementary Information is included as Appendix D.

"Statistical mechanics of globular oligomer formation"
A. J. Dear, A. Saric, T. C. T. Michaels, and T. P. J. Knowles
Manuscript under preparation for submission to Biophys. J.
This manuscript is reproduced in Chapter 9; its Supplementary Information is included as Appendix G.

"Renormalization group for protein filament formation"
T. C. T. Michaels*, **A. J. Dear***, and T. P. J. Knowles
Manuscript submitted to Phys. Rev. Lett.
(*these authors contributed equally to this work)
This manuscript is reproduced in Chapter 10.

"Quantifying co-oligomer formation by alpha-synuclein"
M. Iljina*, **A. J. Dear***, G. A. Garcia, S. De, P. Flagmeier, D. R. Whiten, T. C. T. Michaels, D. Frenkel, C. M. Dobson, T. P. J. Knowles, and D. Klenerman
Manuscript under review at ACS Nano.
(*these authors contributed equally to this work)
This manuscript is reproduced in abbreviated form in Chapter 8; some Supplementary

Information sections are included in Appendix F.

"Oligomer diversity during the aggregation of the repeat region of tau"
M. Kjaergaard*, **A. J. Dear***, F. Kundel, S. Qamar, G. Meisl, T. P. J. Knowles, and D. Klenerman
In press, ACS Chem. Neurosci.
(*these authors contributed equally to this work)

"Stochastic calculus of protein filament formation under spatial confinement"
T. C. T. Michaels, **A. J. Dear**, and T. P. J. Knowles
New J. Phys. **20**, 055007

"Direct observation of oligomerization by single molecule fluorescence reveals a multistep aggregation mechanism for the yeast prion protein Ure2"
J. Yang*, **A. J. Dear***, T. C. T. Michaels, C. M. Dobson, T. P. J. Knowles, Si Wu, and S. Perrett
J. Am. Chem. Soc. **140**, 2493-2503 (2018).
(*these authors contributed equally to this work)
The material I was primarily responsible for is reproduced in Chapter 5; Supplementary Information sections for which I was primarily responsible are included in Appendix C.

"Dynamics of Heteromolecular Filament Formation"
A. J. Dear, T. C. T. Michaels, and T. P. J. Knowles
J. Chem. Phys. **145**, 175101 (2016).
This manuscript is reproduced in abbreviated form in Chapter 4 and in Appendix B.

"Scaling and dimensionality in the chemical kinetics of protein filament formation"
T. C. T. Michaels, **A. J. Dear**, and T. P. J. Knowles
International Reviews in Physical Chemistry, **35:4**, 679 (2016).

"Fluctuations in the Kinetics of Linear Protein Self-Assembly"
T. C. T. Michaels, **A. J. Dear**, J. B. Kirkegaard, K. L. Saar, D. A. Weitz, and T. P. J. Knowles
Phys. Rev. Lett. **116**, 258103 (2016).
The majority of this manuscript is reproduced in Chapter 3; Supplementary Information sections for which I was primarily responsible are included in Appendix A.

"Quantitative analysis of co-oligomer formation by amyloid-beta peptide isoforms"
M. Iljina*, G. A. Garcia*, **A. J. Dear***, J. Flint*, P. Narayan, T. C. T. Michaels, C. M. Dobson, D. Frenkel, T. P. J. Knowles and D. Klenerman
Sci. Rep. **6**, 28658 (2016).
(*these authors contributed equally to this work)
This manuscript is reproduced in abbreviated form in Chapter 7; Supplementary Infor-

mation sections for which I was primarily responsible are included in Appendix E.

“Electrostatically-guided inhibition of Curli amyloid nucleation by the CsgC-like family of chaperones”
J. D. Taylor, W. J. Hawthorne, J. Lo, **A. J. Dear**, N. Jain, G. Meisl, M. Andreasen, C. Fletcher, M. Koch, N. Darvill, N. Scull, A. Escalera-Maurer, L. Sefer, R. Wenman, S. Lambert, J. Jean, Y. Xu, B. Turner, S. G. Kazarian, M. R. Chapman, D. Bubeck, A. de Simone, T. P. J. Knowles and S. J. Matthews
Sci. Rep. **6**, 24656 (2016).

“Single-molecule FRET studies on alpha-synuclein oligomerization of Parkinson’s disease genetically related mutants”
Laura Tosatto* , M. H. Horrocks*, **A. J. Dear**, T. P. J. Knowles, M. D. Serra, N. Cremades, C. M. Dobson and D. Klenerman
Sci. Rep. **5**, 16696 (2015).
(*these authors contributed equally to this work)

“Fast Flow Microfluidics and Single-Molecule Fluorescence for the Rapid Characterization of alpha-synuclein Oligomers”
M. H. Horrocks, L. Tosatto, **A. J. Dear**, G. A. Garcia, M. Iljina, N. Cremades, M. D. Serra, T. P. J. Knowles, C. M. Dobson, and D. Klenerman
Anal. Chem. **87**, 8818 (2015).

“Effect of disorder on condensation in the lattice gas model on a random graph”
T. P. Handford, **A. J. Dear**, F. J. Pérez-Reche, and S. N. Taraskin
Phys. Rev. E **90**, 012144 (2014).

Chapter 2

Introduction

2.1 Background

Proteins are diverse biological macromolecules with a common fundamental architecture but widely varying structures and properties, and are responsible for much of the functionality underlying living organisms. They consist of one or more covalently bonded chains of amino acids, that fold into specific three-dimensional structures in aqueous solution. Their conformation or structure is closely connected to their function; failure to fold into the "correct" structure for which they were designed is responsible for a range of pathologies [1–5]. A notable example is the aggregation of misfolded proteins into linear noncovalently bonded polymers known as amyloid fibrils [6, 7]. The amyloid form is in fact often the most thermodynamically stable peptide state even under physiological conditions, and thus many different proteins have a tendency to form these structures [8, 9]. Although the kinetic barrier to the initial formation of amyloid is high, once formed the fibrils proliferate rapidly. This uncontrolled proliferation is closely linked to several important diseases, such as Alzheimer's disease and type II diabetes [10–12].

Much progress has been made in uncovering the molecular mechanisms behind the proliferation of fibrils associated with some of these diseases in recent years, by applying analytical kinetic models to bulk time-dependent measurements of fibril concentrations [13–16]. However, there has been only limited insight into the processes responsible for the initiation of the aggregation reaction. This is in large part because the proliferation of macroscopic quantities of fibrils is fairly insensitive to the precise nature of these processes [16, 17]. The complex reaction pathways by which new fibrils are formed have typically been abstracted into a single "primary nucleation" step for the purposes of bulk kinetic modelling [18–21]. It has become clear however, that these pathways often involve diverse oligomeric intermediates. Moreover, it has become clear in recent years that in fact oligomeric species formed at very low concentrations in the early stages of amyloid fibril formation, and not the fibrils themselves, are toxic to cells, and are now believed to be the key pathogenic agent in neurodegenerative and other amyloidogenic diseases [8, 22–24].

With the advent of single-molecule experimental techniques [25, 26], it has recently become feasible to measure oligomer concentrations quantitatively. The opportunity has therefore arisen to explicitly model these low-concentration oligomeric species directly, to gain insight into their role and nature in a variety of amyloidogenic systems. Some early work has been carried out in this field; however, the analytical models developed to date have been either insufficiently generalizable, being tailored to a very specific system [27, 28], or too coarse grained for simple mechanistic interpretation [29].

The main focus of my PhD has been to explore the initiation of aggregation, by developing and applying theoretical methods that can determine the underlying reaction pathways. I have helped develop techniques to study primary nucleation with far greater accuracy than hitherto possible; and I have developed a range of analytical kinetic and thermodynamic models of oligomer formation, using them to interpret experimental data from diverse biologically relevant amyloid aggregation reactions. This approach has led to a range of new insights into the kinetics and thermodynamics of amyloid oligomers.

2.2 Theoretical methods

Much of the work detailed in this dissertation involves developing and solving kinetic equations for the modelling of self-assembly processes, or the spontaneous formation of complex structures from simple components. Several chapters generalize and extend previous theoretical results obtained for the kinetics of simple linear self assembly; hence, to aid understanding of subsequent chapters, a brief overview of the basic methods used in obtaining these original results is presented here [21, 30].

Kinetic equations based on the principle of mass action, originally developed to model small-molecule chemical reactions, have since been used to model a huge range of growth processes outside the remit of traditional chemistry. Such processes include crystal nucleation & growth [31], polymerisation [32], disease epidemics [33], population dynamics [34], and microbial growth [35], amongst countless others. The first application of such equations to protein polymerisation was performed by Oosawa *et al* [18] over 50 years ago. In the past decade especially, there has been a significant increase in interest in the kinetic modelling of protein self-assembly, driven by advances in our understanding of the fundamental role of protein aggregation in neurodegenerative diseases. The overview in this section focuses on the theoretical methods developed in this decade for solving such rate equations.

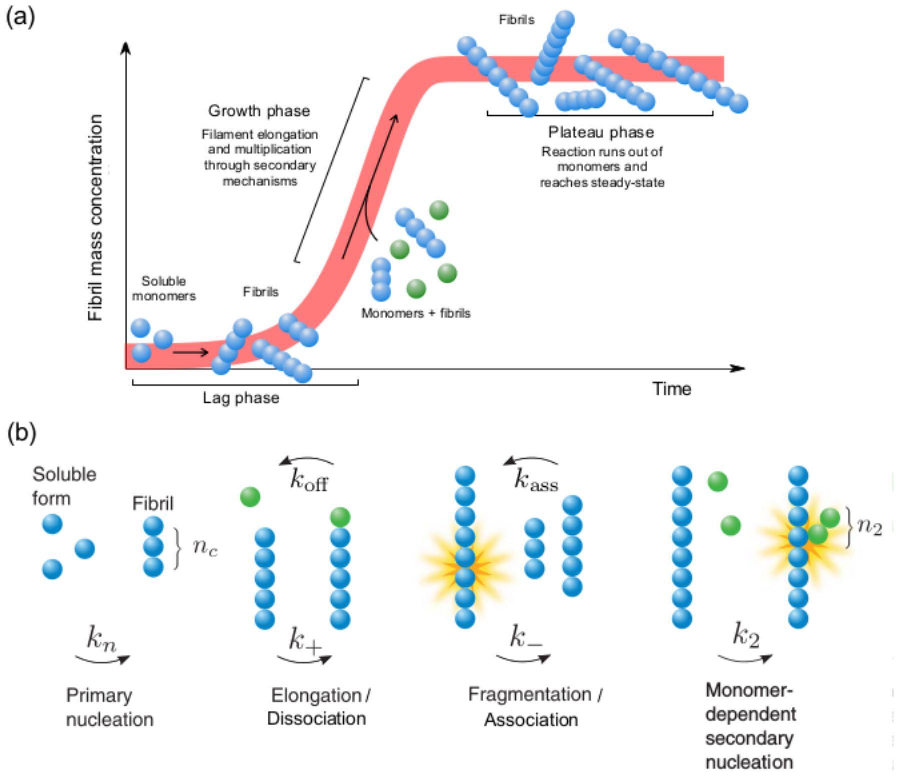

Figure 2.1: (a) A schematic representation of a typical kinetic curve for amyloid growth. Similarly to crystallisation, filamentous growth may begin with the formation and growth of nuclei during the lag phase. There is then a rapid phase of growth ending in a plateau. (b) An overview of the basic microscopic events involved in the growth of linear protein filaments. Figure adapted from [36].

2.2.1 Full master equations

An aggregation reaction involves many fibrils of different lengths; for a complete description of the system one would track the time evolution of the full length distribution of filaments [37–42]. The quantities that are tracked are the concentrations of the species of different lengths $f(t, j)$, which are held to be deterministic. Their time evolution is governed by the so-called mean-field master equation, which accounts explicitly for every process that can occur in the system, under the assumptions of molecular chaos and thermodynamically large system size. Master equation approaches are versatile and have also been used for modelling the growth kinetics of virus capsids and other closed spherical polymers [43–45], as well as the assembly of non-amyloidogenic protein filaments [18]. In the case of filamentous growth, the master equation has the following form: [30]

$$\begin{aligned}\frac{\partial f(t,j)}{\partial t} =& 2k_+ f(t,j-1) - 2k_+ m(t) f(t,j)\\ &+ 2k_{\text{off}} f(t,j+1) - 2k_{\text{off}} f(t,j)\\ &+ 2k_- \sum_{i=j+1}^{\infty} f(t,i) - (j-1)k_- f(t,j)\\ &+ k_n m(t)^{n_c} \delta_{j,n_c} + k_2 m(t)^{n_2} \sum_{i=n_c}^{\infty} i f(t,i) \delta_{j,n_2} \end{aligned} \tag{2.1}$$

$$\frac{dm(t)}{dt} = -\frac{d}{dt}\left[\sum_{j=n_c}^{\infty} j f(t,j)\right], \tag{2.2}$$

where k_+ is the rate constant for filament elongation by monomer addition, k_{off} is its inverse (filament shortening by loss of the terminal monomeric subunit), k_n is the rate constant for new filament formation by free association of solution-phase monomer (primary nucleation), and n_c is the reaction order of primary nucleation with respect to monomer (see Fig. 2.1). The master equation also takes into account optional "secondary processes", by which new filament nuclei are generated from existing filaments, and are often important in amyloid filament assembly. These include filament fragmentation [46] (rate constant k_-), and monomer-dependent secondary nucleation at the surface of existing filaments [19, 20, 47] (rate constant k_2; reaction order with respect to monomer n_2). Note we neglect fibril association, which is the inverse of fragmentation, as this is considered unimportant in many protein polymers, due to spatial constraints [30, 48, 49].

2.2.2 Moment equations

The full master equations are hard to solve, and the length distribution is anyway not the most directly experimentally accessible quantity. For these reasons, previous theoretical work has focused on simpler quantities, namely the 0^{th}- and 1^{st}-order moments of the distribution, or the number and mass concentration of polymers [14, 21, 30, 50]. These are given by the formulae $P(t) = \sum_j f(t,j)$ and $M(t) = \sum_j j f(t,j)$, and, applying these contractions to the master equations, one arrives at the following expression for the 0^{th} moment

$$\begin{aligned}\frac{dP}{dt} =& k_-[M(t) - (2n_c - 1)P(t)]\\ &+ k_2 m(t)^{n_2} M(t) + k_n m(t)^{n_c}, \end{aligned} \tag{2.3}$$

where we have neglected the contribution from $k_{\text{off}}f(t, n_c)$.

The equation for the 1^{st} moment is given by

$$\begin{aligned}\frac{dM}{dt} =&2[m(t)k_+ - k_{\text{off}} - k_- n_c(n_c-1)/2]P(t) \\ &+ n_2 k_2 m(t)^{n_2} M(t) + n_c k_n m(t)^{n_c}\end{aligned} \tag{2.4}$$

The last two terms are typically very small relative to the elongation term.

Note that it is not strictly necessary to start with the full master equations in order to derive these expressions.

2.2.3 Fixed-point scheme

The moment equations themselves are in general analytically intractable, so our aim becomes to find an accurate approximation scheme instead. We first recast equations (2.3) and (2.4) as integral equations. They can be represented as

$$\vec{x}(t) = \mathcal{A}\left[\vec{x}(t)\right], \tag{2.5}$$

where $\vec{x}(t) = [P(t), M(t)]$ and $\mathcal{A}$ is the integral operator. The fixed point $\vec{x}^*(t)$ of this operator is the point in the space of possible values for the moments that is returned unchanged upon application of the operator. The fixed point is therefore precisely the solution to eqs (2.3) and (2.4).

The contraction mapping principle [51] now guarantees that these integral equations can be solved iteratively

$$\vec{x}^*(t) = \lim_{N\to\infty} \mathcal{A}^N\left[\vec{x}_0(t)\right], \tag{2.6}$$

for starting values of the moments $\vec{x}_0(t)$ sufficiently close to the final solution [30,36]. Note this principle is taken advantage of by numerical methods for the solution of equations, such as Newton's method. Each iteration then gives rise to a solution that is more accurate, and that approaches the true solution more closely.

This method requires initial input of a reasonably accurate trial solution; and the speed of convergence of the iterative scheme depends upon the accuracy of the initial guess. An effective choice has been found from investigating the linearised versions of equations (2.3) and (2.4).

2.2.4 Linearised moment equations

If we assume that not too much time has elapsed such that monomers have not been significantly depleted from solution, the monomer concentrations on the RHS of equations

(2.3) and (2.4) can be held constant at their initial values, giving the early-time equations

$$\frac{dP}{dt} = [k_- + k_2 m_{\mathrm{tot}}^{n_2}]M(t) + k_n m_{\mathrm{tot}}^{n_c} \tag{2.7}$$

$$\frac{dM}{dt} = 2k_+ m_{\mathrm{tot}} P(t), \tag{2.8}$$

neglecting terms insignificant at early time.

These are linear 1^{st}-order ODEs and thus can be straightforwardly solved [21,30], yielding a solution valid in the limit t. This proves to be a good trial solution for the fixed point scheme, such that upon application of the fixed point operator we obtain, after just 1 iteration, a good approximation to the full time-course behaviour of the mass concentration.

2.3 Outline of thesis

Stochastic aggregation kinetics

I first present, in Chapter 3, a stochastic theory describing how monomers aggregate into fibrils in sub-thermodynamic volumes. Crucially, it is found that an aggregation reaction becomes stochastic under confinement in small volumes in a way that causes primary nucleation to have a dominant role in the kinetics. By contrast, in bulk aggregation reactions, the primary nucleation rate enters the kinetic equations only logarithmically and coupled to elongation, and is thus challenging to measure accurately. Thus, small volume aggregation reactions, carried out in microdroplets using microfluidic devices, allows us to accurately study the primary nucleation mechanism for the first time.

The early stages of coaggregation

In Chapter 4 I develop general kinetic equations describing coaggregation of multiple monomer types into mixed filaments, and solve them analytically for early times, before significant monomer depletion has occurred. This provides valuable insight into the effect of coaggregation on the early stages of amyloid aggregation. This is a problem of significant real-world interest, as amyloidogenic proteins rarely exist in isolation *in vivo*, instead typically being found in heterogeneous environments that are supportive of coaggregation.

The kinetics of amyloid aggregation via oligomeric intermediates

A major part of my thesis is given over to investigating the kinetics of amyloid fibril formation coupled to amyloid oligomer formation. In Chapter 5 I develop partly analytical models to describe formation of oligomers of the functional yeast prion protein Ure2, discovering that although most Ure2 oligomers dissociate rather than form fibrils, they are still on-pathway intermediates of the Ure2 fibril formation process. Drawing on this, on my work on tau oligomer formation not included here, and on other insights obtained by Thomas C. T. Michaels on Aβ oligomer formation, Chapter 6 then explains the general features of amyloid oligomer formation, and how insight may be derived from them using chemical kinetics. It includes new, accurate and fully analytical, models for the kinetics of oligomers formed as intermediates of both primary and secondary nucleation reaction steps, that are suitable for modelling all amyloid oligomer species heretofore discovered.

Statistical mechanics of oligomer formation

Next, the thermodynamics of oligomer and co-oligomer formation is considered. In Chapter 7 I investigate co-oligomer formation at equilibrium between Aβ40 and Aβ42 peptides using a simple dimer model; in Chapter 8 I then investigate equilibrium co-oligomerization between α-synuclein mutants using a model of linear co-oligomer formation. In both cases, co-oligomers are found to be at least as stable as self-oligomers. It has proved difficult to measure oligomer size distributions and geometry in kinetic experiments, even for self-oligomers. To aid in understanding these features of oligomer diversity, in Chapter 9 I investigate the equilibrium statistical mechanics of peptide oligomers consisting of a single monomer type. I draw on micelle theory to model size distributions of globular oligomers, compare the results to those of an established linear oligomer model, and discuss how they may be distinguished from experiment.

Universal solutions to linear self-assembly using renormalization group (RG) theory

Although the initiation of aggregation has been my main focus, I have carried out a broad range of research and several projects do not fit solely into this category. One such project of particular significance is presented in Chapter 10. Using dynamical RG theory I develop a strategy for obtaining highly accurate analytical solutions to the kinetics of linear self-assembly across a wide range of different systems. We used this to investigate the universal behaviour of filamentous growth kinetics. Due to its unsurpassed generality and accuracy, and to its clear conceptual underpinnings, I believe this will find many applications in my future work. Moreover, it has already been used successfully in this thesis (Chapter 6), to develop particularly accurate solutions to oligomerizing systems featuring secondary nucleation of filaments.

To summarize, in this thesis I present a wide range of works relating to the early stages of amyloid aggregation. These include several studies combining theory and experiment, as well as pure theoretical studies. Together they provide a universal framework for understanding the nature of amyloid oligomers and their role in amyloid filament formation, exerted via the coarse grained primary nucleation reaction step. These theories were derived for homomolecular aggregation; I also begin to investigate the early stages of coaggregation of multiple monomeric peptide types into mixed oligomers and fibrils. I finally present a powerful and highly general technique for studying linear self-assembly that I believe will supplant most existing techniques.

Part II

The Early Stages of Fibril Formation

Chapter 3

Studying Primary Nucleation Using Confinement

Chapter Abstract

Biological systems are characterized by compartmentalization from the subcellular to the tissue level, and thus reactions in small volumes are ubiquitous in living systems. Under such conditions, statistical number fluctuations, which are commonly negligible in bulk reactions, can become dominant and lead to stochastic behaviour. We present here a stochastic model of protein filament formation in small volumes. We show that two principal regimes emerge for the system behaviour, a small fluctuation regime close to bulk behaviour and a large fluctuation regime characterised by single rare events. Our analysis shows that in both regimes the reaction lag-time scales inversely with the system volume, unlike in bulk. Finally we use our stochastic model to connect data from small volume microdroplet experiments of amyloid formation to bulk aggregation rates, and show that digital analysis of an ensemble of protein aggregation reactions taking place under micro-confinement provides an accurate measure of the rate of primary nucleation of protein aggregates, a process which has been challenging to quantify from conventional bulk experiments.

In this project, a stochastic model for amyloid aggregation in the sub-thermodynamic small-volume regime was developed. A key result was the discovery that stochastic kinetic experiments provided a way of studying the early stages of amyloid filament formation directly, and with much greater accuracy than pre-existing methods. The work presented in this chapter was carried out in close collaboration between T. C. T. Michaels and myself; since I provided roughly half of the theoretical content, our contributions cannot be disentangled into coherent separate manuscripts. Instead, our work is presented together, and results sections for which I do *not* wish to claim primary credit are labelled with a dagger (†). Derivations for which I am responsible are included in Appendix A. Our work was published in *Phys Rev Lett* in 2016, under the title “Fluctuations in the Kinetics of Linear Protein Self-Assembly”. An abbreviated version of this paper, focussing somewhat more on the implications for studying the early stages of amyloid aggregation, is presented in this chapter.

A precise list of my contributions are as follows. I discovered and proved the correlation between n and x, allowing me to reduce the insoluble full master equation Eq. (3.1) to the tractable form Eq. (3.2). I discovered that the naive boundary conditions used in [52] were incorrect, and subsequently determined the correct ones; and used these new boundary conditions to correct both T. C. T. Michaels’ solution to the simplified master equations and A. Szabo’s solution for the lag time distribution. Finally I uncovered the limiting behaviour of the lag times in key regimes, and identified that stochastic aggregation experiments would permit much more accurate investigation of the primary nucleation reaction step.

3.1 Introduction

The formation of protein filaments is a process of central importance for both normal [18, 53] and aberrant biology [1, 4], as well as for the development of novel materials for nanotechnology [54–57]. The fundamental kinetic equations describing such processes in bulk are well-established in the literature and have been studied extensively over the past 50 years [13, 18, 19, 21, 30, 47, 53, 58–60]. These descriptions rely on the mean-field assumption [59] and therefore neglect statistical mechanical fluctuations. Yet, protein aggregation processes in typical cellular environments (fL-pL) involve significantly smaller numbers of molecules than conventional bulk experiments and thus stochastic variability is expected to play an important role [19,47,61,62]. Moreover, recent experimental advances in microdroplet techniques [62, 63] allow volumes in the picolitre range, comparable to intracellular volumes, to be probed for synthetic systems, creating the need for a general theoretical framework capable of describing protein filament assembly in small volumes.

Current theoretical descriptions of protein filament formation in small volumes focus on systems characterized by aggregate propagation from a single primary nucleation event [64]. A key question is, however, the nature of the full fluctuation behaviour bridging the gap between the limit of classical nucleation theory and bulk behaviour. In this Letter, we present a general theoretical description of stochastic effects in filamentous growth processes with secondary pathways [13,19,21,39,46,47,58,65–69] and derive closed-form expressions for the distribution of lag times. Our theoretical framework describes currently available microdroplet experimental data which are characterized by aggregate proliferation from multiple nucleation sites. Moreover, our results suggest a powerful method for characterizing the primary nucleation step, which is typically difficult to access from current bulk methods. We demonstrate the power of this approach by obtaining a value for the rate of primary nucleation for bovine insulin aggregation which is significantly better constrained than results obtained from analysis of bulk data.

3.2 Stochastic moment equations†

We consider a system of volume V containing a mixture of fibrillar aggregates and monomeric proteins in solution. Since we are interested in the early-stages of the assembly process, we assume a constant chemical potential for the available soluble precursor proteins [21,30,58]. We describe the state of the system by a vector (n, x), where n is the number of fibrils and x is the number of monomers incorporated into aggregates, parameters which relate directly to experimental observables [70]. The probability distribution function (PDF) $\mathbb{P}(n, x, t)$ of states (n, x) evolves according to a master equation [64]

$$\begin{aligned}\frac{\partial \mathbb{P}(n,x,t)}{\partial t} &= \alpha_1 \mathbb{P}(n-1, x-x_c, t) - \alpha_1 \mathbb{P}(n,x,t) \\ &+ \mu n \mathbb{P}(n, x-1, t) - \mu n \mathbb{P}(n,x,t) \\ &+ \alpha_2 (x - x_2) \mathbb{P}(n-1, x-x_2, t) \\ &- \alpha_2 x \mathbb{P}(n,x,t), \\ \mathbb{P}(n,x,t) &= 0 \quad \forall (n,x) < 0,\end{aligned} \tag{3.1}$$

where α_1, μ and α_2 are the transition rates (units s^{-1}) for primary nucleation, filament elongation and secondary mechanisms, respectively (Fig. 1(a)). This description explicitly considers time variations of the PDF in terms of probability fluxes: the positive expressions represent gain terms that account for system transitions into state (n, x), whereas the negative terms describe losses from transitions from (n, x) into other states. The terms on the first line of Eq. (3.1) describe the initial primary nucleation step as the spontaneous formation of new growth-competent aggregates from x_c monomers. The increase of aggregate mass through elongation is described by the terms on the second line of Eq. (3.1). Secondary processes, in which new filaments of initial size x_2 are produced by means of a reaction involving existing fibrils, are captured by the third and fourth lines of Eq. (3.1). These cover several options, including filament breakage ($x_2 = 0$) [21, 46, 65, 66], lateral branching ($x_2 = 1$) [58,67] and surface-catalyzed secondary nucleation ($x_2 \geq 2$) [19,68,69]. Note that in general monomer dissociation from filament ends and re-joining of fibrils are necessary components to ensure microscopic reversibility [71]. The assumption of vanishing rates of monomer dissociation and polymer re-joining employed here, however, is justified as these processes do not significantly affect the early stages of the reaction [49].

The transition rates entering Eq. (3.1) can be related to the total concentration of proteins, m_{tot}, and the bulk rate parameters k_n, k_+, k_2 for primary nucleation, elongation and secondary pathways, respectively, by requiring the rate equations for the averages $\langle n\rangle$ and $\langle x\rangle$ being in agreement with existing early-time deterministic models [21, 30, 58, 59]. This condition yields $\alpha_1 = k_n m_{\text{tot}}^{n_c} N_A V$, $\mu = 2k_+ m_{\text{tot}}$ and $\alpha_2 = k_2 m_{\text{tot}}^{n_2}$, where N_A is the Avogadro number [64], n_c is the reaction order for primary nucleation, and n_2 the reaction order for secondary nucleation. Importantly, the transition rate for primary nucleation, α_1, explicitly depends on the system size, V, while the parameters μ and α_2 describing autocatalytic growth are determined only by the associated bulk quantities. We expect therefore that reducing system size leads to a transition from a situation when the kinetics are controlled by autocatalytic growth to a situation when the fibrillization reaction is limited by primary nucleation. Thus, primary nucleation events becoming infrequent is at the origin of the stochastic behaviour of filamentous growth processes in small volumes.

3.3 Analytical solution for the PDF

The master equation Eq. (3.1) yields differential equations for the principal moments of the PDF through summation over system compositions. The equations for the j^{th}-order moments can be determined for general j and written in matrix form (see Appendix A.1). The largest eigenvalue and its associated eigenvector may then be calculated. The largest eigenvalue for the j^{th}-order moments is found to be $j\kappa$, with $\kappa = \sqrt{\mu\alpha_2} = \sqrt{2k_+ k_2 m_{\text{tot}}^{n_2+1}}$ being the characteristic time scale for aggregate proliferation [21, 30, 58]. The moments are therefore dominated by the contribution from this eigenvalue-eigenvector pair in the limit $t \gg \kappa^{-1}$. From this it follows that n and x are perfectly correlated for times greater than κ^{-1}, but still sufficiently short for the constant monomer approximation to be valid. This is confirmed by calculation of the Pearson's correlation coefficient for n and x, $\rho_{n,x} = [\langle nx\rangle - \langle n\rangle\langle x\rangle][(\langle n^2\rangle - \langle n\rangle^2)(\langle x^2\rangle - \langle x\rangle^2)]^{-1/2}$, which equals 1 in this limit (see Appendix A.2). These results imply the existence of a linear correlation in this regime between the random variables n and x, whereby the constant of proportionality is $x = \frac{\kappa}{\alpha_2} n$, $t \gg \kappa^{-1}$. We can directly test this prediction from numerical realizations of Eq. (3.1) generated using Gillespie algorithm [72] which reveal that n and x are indeed linearly correlated even before aggregation is detected [73]. The linear correlation between n and x allows recasting the master equation (3.1) into an equivalent one with a single variable (see Appendix A.3 for calculation)

$$\begin{aligned}\frac{\partial \mathbb{P}(n,t)}{\partial t} &= \alpha_1 \mathbb{P}(n-1,t) - \alpha_1 \mathbb{P}(n,t)\\ &\quad + \kappa(n-1)\mathbb{P}(n-1,t) - \kappa n \mathbb{P}(n,t).\end{aligned} \tag{3.2}$$

Interestingly, Eq. (3.2) is analogous to the master equation of bacterial growth [52], whereby bacteria are constantly introduced into the system at rate α_1 and multiply with

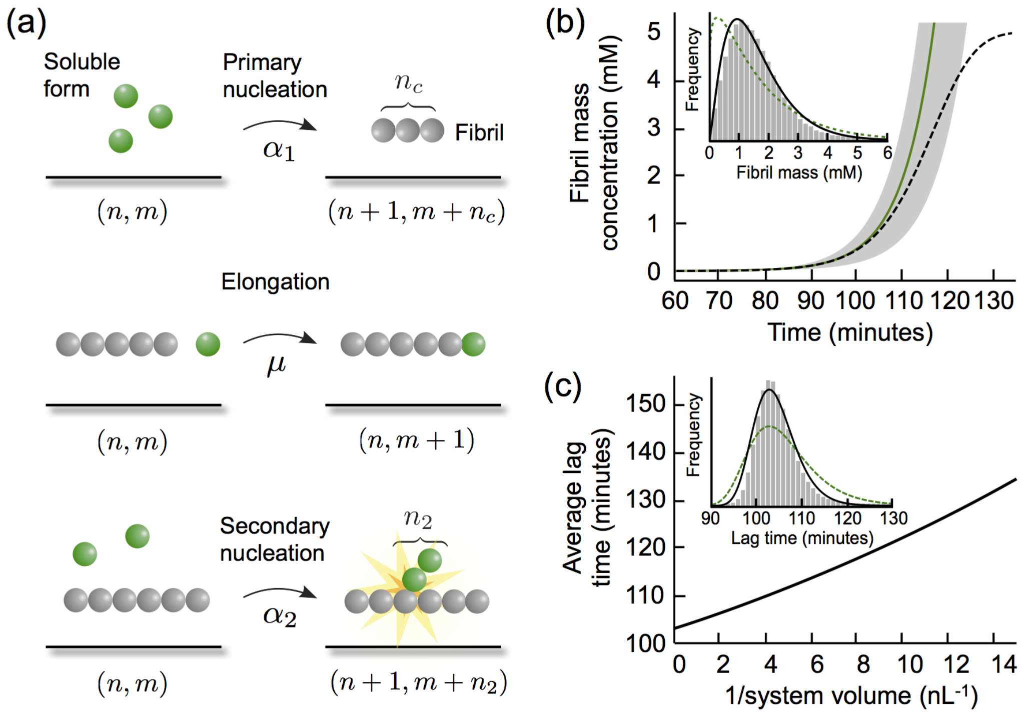

Figure 3.1: (a) Different transitions in stochastic protein aggregation. (b) Time evolution of average mass concentration and 68% confidence bands. Inset: PDF for x at $t = 106$ minutes predicted by Eq. (3.3) (solid line) is compared to numerics. The dashed line is the solution of [52]. (c) The scaling behaviour of the average lag time with system volume predicted by Eq. (3.5). Inset: PDF of lag times for $V = 1$ nL predicted by Eq. (3.4) (solid line) is compared to numerics. The dashed line is from [52]. Calculation parameters: $k_n = 4 \times 10^{-13}$ M^{-1}s^{-1}, $x_c = 2$, $x_2 = 0$, $k_2 = 2.5 \times 10^{-8}$ s^{-1}, $k_+ = 2.5 \times 10^4$ M^{-1}s^{-1}, $m_{\text{tot}} = 5$ mM, $V = 1$ nL. Numerical simulations performed by J. B. Kirkegaard [73].

rate κ. This analogy, first hypothesized by Szabo [52], is a statement of the fact that for times bigger than κ^{-1} the average length of aggregates is constant. To define appropriate initial conditions for Eq. (3.2), we match the first moments of the PDFs of Eqs. (3.1) and (3.2) for times $t \gg \kappa^{-1}$, yielding $\langle n \rangle (t = \log 2/\kappa) = 0$. This condition translates necessarily into $\mathbb{P}(n, t = \log 2/\kappa) = \delta_{n,0}$ therefore ensuring that also all higher moments of the PDFs of Eqs. (3.1) and (3.2) match for $t \gg \kappa^{-1}$ at leading order. The exact solution of Eq. (3.2) subject to the above initial conditions is (see Appendix A.4)

$$\mathbb{P}(n,t) = \frac{2^{\alpha_1/\kappa}\Gamma\left(n + \frac{\alpha_1}{\kappa}\right)}{\Gamma(n+1)\Gamma\left(\frac{\alpha_1}{\kappa}\right)} e^{-(\alpha_1 + \kappa n)t}(e^{\kappa t} - 2)^n, \tag{3.3}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function. The PDF in terms of the variable x is obtained by implementing the correlation between n and x in Eq. (3.3). Figure 1(b) shows that Eq. (3.3) is in agreement with numerical realizations of Eq. (3.1).

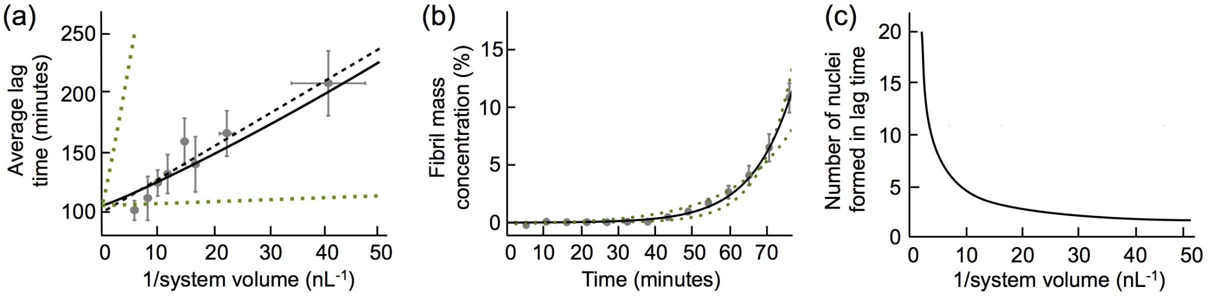

Figure 3.2: (a) Analysis of small-volume experiments of bovine insulin fibrillization kinetics from Ref. [57]. Dashed line: best fit to $\langle\tau\rangle \propto c_n/V$ with $c_n^{-1} = 6 \times 10^6$ s^{-1} l^{-1}; solid line: prediction from Eq. (3.5); dotted lines: k_n is decreased and increased by an order of magnitude. $\langle\tau\rangle$ shows marked volume dependence despite the presence of multiple nucleation sites as demonstrated by the plot of the number of nuclei formed on average during the mean lag time against $1/V$. (b) Kinetic analysis of insulin aggregation in bulk. Solid line: best fit curve to initial exponential growth; dotted lines: k_n is decreased/increased by an order of magnitude relative to best fit. Data provided by K. L. Saar. [73] (c) Average number of nuclei formed in lag time.

3.4 Lag times†

A common qualitative feature of filamentous growth processes is the observation of a lag phase before aggregation can be detected. A commonly used measure of this polymerization delay is the lag time, τ, defined as the time at which the aggregate mass concentration $x(t)/(N_A V)$ reaches an arbitrarily chosen concentration threshold M_{th}. Although it is often the case that a halfway point for the reaction is taken, this may fall outside the realm over which our approximations are valid. Therefore a 10% extent or the experimental limit for aggregate detection [70] are simple choices for M_{th} more in keeping with our solution. Because τ is a random variable, the quantity of interest is the PDF of lag times, i.e. the probability $T(t)$ that τ equals t.

Although previous authors have studied lag times in kinetic growth equations [74], these have typically been focused on the bulk mass-action limit, and have not investigated stochastic effects. Moreover, the systems typically explored do not feature secondary growth processes, and are thus qualitatively different in behaviour. On the other hand, Szabo *et al* [52] investigated lag times in stochastic bacterial proliferation. Since the master equation for bacterial propagation was found to be identical to our simplified master equation Eq. (3.2), we were able to apply the same techniques to determine lag times in stochastic filamentous growth with secondary processes.

According to the theory of first passage times [52], $T(t)$ is computed as $T(t) = -\frac{dQ(t)}{dt}$, where $Q(t) = \sum_{n=0}^{n_{\text{th}}-1} \mathbb{P}(n,t)$ is the probability that at time t the process $x(t)/(N_A V)$ has not yet reached M_{th}, and n_{th} is the number of filaments associated with this threshold.

Using Eq. (3.3) we find (see Appendix A.5)

$$T(t) = \frac{\kappa 2^{\alpha_1/\kappa}\Gamma\left(n_{\text{th}} + \frac{\alpha_1}{\kappa}\right)}{\Gamma\left(n_{\text{th}}\right)\Gamma\left(\frac{\alpha_1}{\kappa}\right)} e^{-(\alpha_1 + (n_{\text{th}}-1)\kappa)t}(e^{\kappa t} - 2)^{n_{\text{th}}-1}, \tag{3.4}$$

where $n_{\text{th}} = \alpha_2 N_A V M_{\text{th}}/\kappa$. The average lag time, $\langle\tau\rangle = \int_0^\infty tT(t)dt$, is obtained from adaptation of the result in [52] to our boundary conditions (see Appendix A.5) as

$$\langle\tau\rangle = \frac{\log(2)}{\kappa} + \sum_{j=0}^{n_{\text{th}}-1} \frac{1}{\alpha_1 + j\kappa} \tag{3.5}$$

and the extent of fluctuations is similarly obtained as

$$\sigma^2 = \langle\tau^2\rangle - \langle\tau\rangle^2 = \sum_{j=0}^{n_{th}-1} \frac{1}{(\alpha_1 + j\kappa)^2}. \tag{3.6}$$

3.5 Limiting behaviour of lag time in key regimes

Inspection of Eqs. (3.5) and (3.6) reveal that the level of stochasticity in the system is controlled by the dimensionless parameter $\gamma = \kappa/\alpha_1$. Based on this parameter, we can distinguish three natural regimes of stochastic behaviour: bulk ($\gamma = 0$), onset of stochasticity ($\gamma \to 0$) and single-event controlled ($\gamma \to \infty$). We now discuss how Eqs. (3.5) and (3.6) can be simplified in these regimes. In bulk ($\gamma = 0$), the sum in Eq. (3.5) is replaced by an integral and Eq. (3.5) is determined solely by the propagation time associated with the secondary nucleation chain reaction, τ_{bulk} (see Appendix A.6.1).

In the opposite limit of very small volumes or slow nucleation ($\gamma \to \infty$), the dominant contribution to Eq. (3.5) is the inverse nucleation rate, $\langle\tau\rangle = 1/\alpha_1 = c_n/V$, where $c_n = 1/(k_n m_{\text{tot}}^{n_c} N_A)$ [64]. In this regime, V is small enough that eventually only a single nucleation event occurs ahead of the threshold being reached, at which point $\langle\tau\rangle$ is dominated by the waiting time for formation of a single nucleus, which scales inversely proportional to system volume as expected from classical nucleation theory [75, 76]. The growth of this nucleus into sufficient amounts of aggregates to pass the mass concentration threshold M_{th} occurs far faster than the formation of a second nucleus in this limit. Furthermore, the extent of fluctuations is now controlled by $\sigma^2 = (c_n/V)^2$ (see Appendix A.6.3).

These simplifications apply in the limit of very small volumes, and cannot be used at intermediate volumes that are small enough for stochastic fluctuations to be visible, but still large enough that several primary nucleation events can occur. Series approximation of Eq. (3.5) around $\gamma = 0$ allows us to explore the onset of stochasticity and reveals that $\langle\tau\rangle$ approaches the bulk value τ_{bulk} as $\langle\tau\rangle = \tau_{\text{bulk}} + c_n/(2V)$, where $c_n = 1/(k_n m_{\text{tot}}^{n_c} N_A)$ is the average time of forming nuclei in volume V (see Appendix A.6.2). This system size expansion shows therefore that, in this regime, $\langle\tau\rangle$ approximately decomposes into a sum of the deterministic lag time and half of the inverse nucleation rate (a term proportional to $1/V$). The extent of fluctuations in this regime is approximately given by $\sigma^2 = c_n/(\kappa V)$ (see Appendix A.6.2).

Equations (3.5) and (3.6) interpolate smoothly between these limiting regimes. In particular, across the entire range of system sizes $\langle\tau\rangle$ approximately writes as the sum of the deterministic lag time and a nucleation term proportional to $1/V$, whereby the constant of proportionality satisfies $c_n/2 \leq d\langle\tau\rangle/d(1/V) \leq c_n$. The transition between these two limiting regimes occurs approximately at the critical volume $V_c = \kappa c_n$ at which $\gamma = 1$. This critical volume corresponds to the radius of convergence of the system size expansion of $\langle\tau\rangle$ around $\gamma = 0$ and hence marks the upper volume limit for the small fluctuation result to be accurate.

3.6 Stochastic analysis provides strong constraints for probing primary nucleation events

Heretofore, rate constants for protein aggregation have been determined by carrying out kinetic experiments in bulk, in which the mass concentration of fibrils over time is measured [70, 77], and by fitting such data to rate laws derived from deterministic master equations [77]. Typically, however, the rate constants characterising the elementary processes of primary and secondary nucleation and growth occur as combinations, and thus it remains challenging to obtain accurate values for the rates of these processes from experimental data. In particular the process of primary nucleation has proven challenging to quantify, in part as the major experimental observables such as the lag-phase display only a weak logarithmic dependence on this parameter. Remarkably, however, our results suggest that a measurement of the volume dependence of the lag time allows the rate of primary nucleation to be determined directly from the slope of a plot of $\langle\tau\rangle$ versus $1/V$. This approach has several advantages over attempting to characterize primary nucleation from bulk polymerization fraction experiments: (i) While fitting of bulk experiments fixes only the combined rate parameter k_+k_n, the analysis of stochastic data allows the rate of primary nucleation to be determined directly, without the necessity of estimating the elongation rate constant or the length of the aggregates, from other experimental techniques, factors which are intrinsically sources of significant error. (ii) Microdroplet experiments can be tightly controlled, and lag time experiments are digital in nature with the exact value of the threshold not entering the gradient and thus not contributing to error. (iii) In the presence of secondary mechanisms, the fitting of bulk data over the full time course is predominantly constrained by the autocatalytic processes rather than the primary nucleation step, contributing to uncertainties of several orders of magnitude for the determined nucleation rate [78]. Even if fitting is limited to the early-times, the exponential form of the fitting equation allows for substantial leeway in the values of the fitting parameters, as numerous combinations of such parameters give rise to fairly similar-looking curves. For example, changing the best fit value for k_n by an order of magnitude would not significantly affect the performance of the bulk analysis fit, but would give rise to a dramatically poorer fit of $\langle \tau \rangle$ vs $1/V$ in a linear relationship (Figs. 2(c,e)).

3.7 Connecting small-volume experiments of bovine insulin fibrillization kinetics to bulk experiments†

We have applied this technique to analyse data from microdroplet experiments on bovine insulin [57] and obtained a value for the rate of primary nucleation of $c_n^{-1} = (6 \pm 1) \cdot 10^6$ $\mathrm{s}^{-1}\mathrm{L}^{-1}$ (Fig. 2(a)). K. L. Saar then carried out bulk experiments of insulin aggregation, and the data were fitted to a standard deterministic model, in conjunction with our calculated value for the rate of primary nucleation in microdroplets (Fig. 2(b)). The resultant calculated rate of elongation $\mu = 2 \cdot 10^5$ s^{-1} agrees with those reported in the literature, to within the levels of error expected from the method of calculation [79], showing that the stochastic analysis presented in this paper allows small volume behaviour to be related to conventional bulk experiments. Moreover, the value of the nucleation rate constant is constrained to within better than an order of magnitude, a result that is very challenging to achieve with analysis of bulk data (Fig. 2(b)).

3.8 Conclusions

We have reported of a theoretical study on stochastic effects in nucleated polymerization phenomena in small volumes. We have derived fully analytical results describing the distribution of lag times that allows linking the bulk parameters characterizing large-volume experiments with the statistical properties of polymerization curves in small volumes across the entire range of fluctuation behaviour. Our results provide a practical route towards an accurate determination of primary nucleation rates which represent a key event in the transition of soluble proteins into their aggregated forms. Of particular interest is the ability to probe the primary nucleation reaction order with respect to monomers in detail, both for homogeneous systems and for copolymerising systems; which may reveal new information about oligomeric species formed as intermediates of primary nucleation. Moreover, microdroplet-based experiments admit the possibility of exploring the role of interfaces in primary nucleation, through modulation of the chemical properties of droplet surfaces.

Chapter 4

The Early Stages of Heteromolecular Filament Formation

Chapter Abstract

The self-assembly of molecular building blocks into linear filaments is a common form of self-organization in nature, and underlies the formation of supramolecular polymers in a variety of contexts, including in both functional and aberrant biology. To date, attention has focused mainly on homomolecular assembly phenomena; however, it has recently become apparent that heteromolecular assemblies can be common, and for instance pathological protein filaments such as amyloid aggregates form *in vivo* in environments supporting copolymerization. Here, we present a general kinetic scheme for heteromolecular filament formation, and derive closed-form analytical expressions that describe the dynamics of such systems under constant-monomer conditions. Our results reveal the existence of a demixing transition time controlled by the relative rates of depletion of the different aggregating species, after which predominantly homomolecular polymers are formed even when the initial solution is heteromolecular. Furthermore, these results may be applied to the analysis of experimental kinetic data on the aggregation of mixtures of proteins, to determine which fundamental reaction steps occur between unlike proteins, and to provide accurate estimates of their rate constants.

This chapter investigates the effect of heterogeneity on the early stages of aggregation. It outlines effectively exact analytical solutions to the kinetics of formation of mixed filaments from coaggregation of multiple monomer types at early reaction times. The work in this chapter was published in *J Chem Phys* under the title “Dynamics of heteromolecular filament formation”. I was responsible for all results and writing, receiving helpful advice and feedback on the drafting from Tuomas P. J. Knowles and Thomas C. T. Michaels. All sections from the paper dealing with the full-time kinetics of closed aggregation reactions have been omitted, and some technical calculations and derivations have been removed to Appendix B.

4.1 Introduction

Protein filament formation is a molecular-level process of central importance for both normal [18,53] and aberrant biology [1,4], as well as for the development of novel materials for nanotechnology [54–57]. In recent years, it has become increasingly clear that copolymerization is an important phenomenon in the context of protein filament formation. Copolymerization has been directly implicated in amyloidogenic diseases *in vivo* [80–82], and has been observed and studied in numerous *in vitro* experiments [83, 84]. Specific examples include cross-nucleation between Aβ40 and Aβ42 [85], formation of heteromolecular fibrils consisting of multiple Aβ42 variants [86], and copolymerization of IAPP and Aβ40 proteins [80]. The latter is of especial interest given the known link between Alzheimer's disease and diabetes [87]. There is therefore a clear motivation for the development of models analogous to existing single-species models, that are capable of quantitatively interpreting data from such experiments, and elucidating the important microscopic mechanisms responsible for copolymerization of different biologically-relevant combinations of proteins.

Hitherto, models of linear aggregation kinetics have focused on self-aggregation of single protein species. Such models include the Oosawa model of polymer nucleation and elongation [18], that has been employed to successfully describe a range of biophysical phenomena, such as cell cytoskeleton assembly [53]; and models incorporating secondary processes such as fragmentation of polymer chains, and surface-catalysed nucleation, that are capable of describing protein aggregation in amyloidogenic diseases such as Alzheimer's and Parkinson's [13, 19, 21, 30, 39, 46–48, 58, 65–69, 88]. Although accurate analytical solutions have been derived for these models, they are single-species only and not applicable to multiple-protein systems. This historical direction of development has been motivated by the assumption that a single protein is responsible for a single amyloidogenic disease, and also by the much greater mathematical complexity involved in any attempt to model a multiple-species system.

In this work, we present a new analytic model for heteromolecular fibril formation that explicitly accounts for the interactions of multiple protein types in terms of the various possible fundamental processes. We illustrate much of the work using a system of two protein types, although the model can be generalized to three or more. We expect that this theory will find practical use in interpreting the significant amount of experimental work that is being undertaken in this field.

The paper is organized as follows. In Sec. 4.2, we introduce the approach used to formulate the theory. In Sec. 4.3-4.4, we consider an Oosawa-like system with no secondary processes, and show how to introduce copolymerization into such a system. In Sec. 4.5-4.6, we then generalize our methodology to model a general copolymerizing system that includes secondary processes. In Sec. 4.7, we investigate the internal structure of coaggregated fibrils that are formed during the self-assembly reaction; and finally, in Sec. 4.9, we discuss with examples how this new theory may be applied to the analysis of experimental data, and what can be learned from such analysis.

4.2 Composition Distribution and Principal Moments

On a macroscopic scale, the kinetics of a system of chemical species can be described by a master equation that expresses the rate of change of concentrations of the various species in terms of elementary reaction steps [48]. Invoking the law of mass action, the rates of these elementary reaction steps can be written as simple powers of the concentrations of the species involved. In order to proceed in the development of a kinetic theory of mixed fibril formation, we must therefore first identify the relevant chemical species that we wish to track, and then establish the equations that govern their kinetics.

4.2.1 Introducing the composition distribution

A single-species aggregating system is fully characterized by knowledge of the time-dependent length distribution $f(t, j)$ [21, 30], giving the concentration of an aggregate composed of j monomers at time t. When multiple species are present and can coaggregate this description is no longer sufficient, as in addition to their length, one must keep track of their internal composition. For example, Fig. 1 shows the variety of possible copolymeric structures that can result from the linear assembly of two species, such as random copolymers, block copolymers, and alternating copolymers. In order to encapsulate this possible variety, a full characterization must thus be given in terms of multi-indices $f(t, \vec{j})$, where $\vec{j}$ is a vector representing an aggregate of length $j = \dim(\vec{j})$. Each component j_i can take values from 1 to M, with each number representing a species and M is the number of different species in the system.

4.2.2 Defining the moments of the composition distribution

The full composition distribution equations for a general copolymerizing system contain a very large number of degrees of freedom; however, knowledge of the full composition distribution is often not necessary for the understanding of the system behaviour at a macroscopic level. A fruitful analysis strategy is therefore to follow the time evolution of the fibril mass concentration and the number concentration of fibril ends only, which are related to the composition distribution by:

$$\begin{aligned} P_x(t) &= \sum_{\vec{j}} (\delta_{j_1 x} + \delta_{j_{|j|} x}) f(t, \vec{j}) \\ M_x(t) &= \sum_{\vec{j}} \sum_{i} \delta_{j_i x} f(t, \vec{j}), \end{aligned} \tag{4.1}$$

and contain only that information which is currently accessible in bulk experiments. Note that these quantities (4.1) are analogous to the principal moments commonly considered in a single-species system [30], defined as:

$$\begin{aligned} P(t) &= \sum_{j} f(t, j) \\ M(t) &= \sum_{j} j \; f(t, j), \end{aligned} \tag{4.2}$$

with the key difference that $P(t)$ tracks the concentration of whole fibrils, whereas $P_x(t)$ tracks the concentration of fibril ends. Note also that, in the single-species case, there is no difference between the number of ends P_x and the number of filaments P, aside from a factor of two. However, in the copolymerizing case, tracking the concentration of ends rather than concentration of fibrils allows us to write P_x and M_x as the simplest possible closed set of dynamic equations for this system. These can be directly solved in the case where we have no secondary processes, a task which is more complicated if instead P is used.

In the following sections, we discuss closed-form expressions for these quantities $M_x(t)$, $P_x(t)$ under constant-monomer conditions for different models of practical importance, including the Oosawa model and models featuring secondary processes. In a general

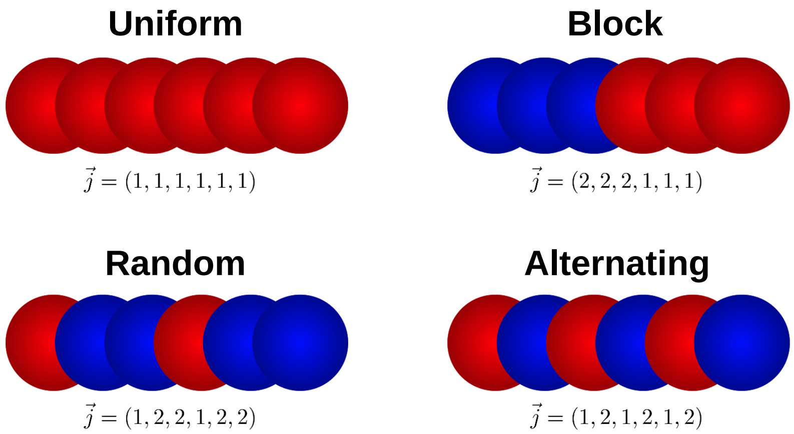

Figure 4.1: Examples of types of copolymers that can be observed, for $j = 6$. Red is labelled monomer type 1, blue is type 2. Subsequently in the paper we refer to red as type a, and blue as type b.

copolymerizing system, we may have any number of monomer species interacting and forming fibrils together; for simplicity, however, we restrict our attention to systems comprising just 2 monomer types, labelled a and b, while noting that our methodology is easily generalizable to systems involving arbitrary numbers of monomer types.

4.3 The Mixed Oosawa Model: From One to Multiple Species

Oosawa and coworkers were the first to study the kinetics of protein polymerization in the absence of secondary processes (e.g. actin polymerization) [18], and in this section we demonstrate how this model can be generalized to account for multiple species.

4.3.1 Introducing the microscopic processes

Differential equations governing the length distribution of a single-species system can be written down, in which each microscopic reaction step is explicitly accounted for [59]. Differential equations governing the principal moments can also be written down, or alternatively derived from the length distribution equations using (4.2) [89]. In the original Oosawa model, we have just 2 active microscopic processes: formation of linear, growth-competent nuclei from free monomers with reaction order n_c and rate constant k_n, known as “primary nucleation”; and addition of free monomers to the ends of these nuclei and of other existing aggregates with reaction order 1 and rate constant k_+, known as “elongation”. In principle there may be inverse elongation and nucleation steps, described by an “off” and a “dissociation” rate; however, for many experimentally-relevant systems these rates are very small compared to the forward ones and here we choose to neglect them for simplicity.

Generalizing nucleation to more than one species

In the multiple-species case, nuclei of a variety of reaction orders with respect to monomers a and b may in principle form. For notational simplicitly, we illustrate our methodology with nuclei of total reaction order 2 only, and assume that reaction orders are directly equivalent to nuclei produced. We thus have only 3 nucleation reactions: production of nuclei of type aa, ab and bb. We note that both of these limitations can be relaxed, and explore this possibility in Appendix B.1. For the case considered, we write the rate constants for these 3 reactions as $2k_n(aa)$, $k_n(ab)$, and $2k_n(bb)$. These microscopic reaction steps, along with their associated rate constants, are illustrated in Fig. 4.2.

Generalizing elongation to more than one species

We consider the most general markovian first-order elongation reactions in which the rate depends both on the identity of the monomer and of the fibril end, but not on the history of the fibril. We write $k_+(x|y)$ as the elongation rate constant for addition of monomer of type x onto an end of type y. Typical elongation reaction steps are illustrated in Fig. 4.2. In some systems, it has been shown that elongation follows a 2-step mechanism [16, 90, 91], with the elongation rate becoming monomer-independent under certain conditions. However, in this first study of copolymerization we ignore this possibility and assume that the rate of elongation is always linearly proportional to the monomer concentration, even though our approach can in principle be generalized to take this effect into account.

4.3.2 Moment equations

We have made the assumption that only nearest-neighbour interactions are important, and therefore that internal composition does not affect the elongation rates. Therefore the multiple-species analogues of the 0th- and 1st-order moment equations alone can encapsulate the key determinants of the dynamics of the multiple-species Oosawa model. Writing $P_a(t)$, $M_a(t)$ and $m_a(t)$ for the number concentration of filament ends of type a, mass concentration of filament-incorporated monomer of type a, and concentration of free monomer of type a respectively, and $P_b(t)$, $M_b(t)$ and $m_b(t)$ for the equivalent quantities

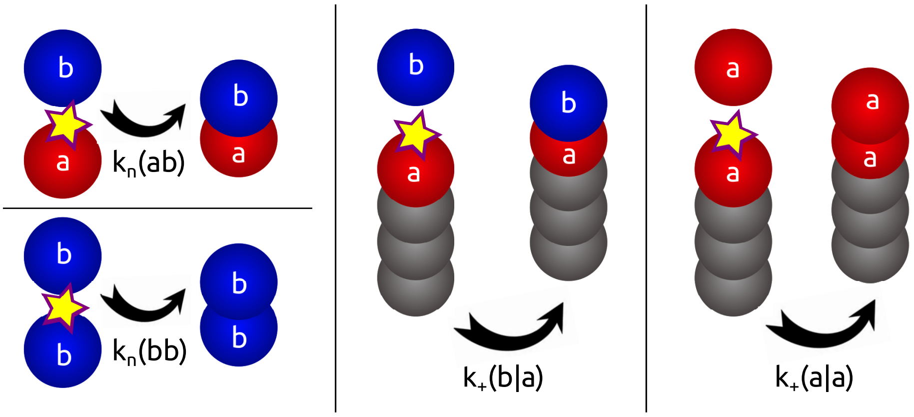

Figure 4.2: A representative set of the types of fundamental primary nucleation and elongation processes involved in heteromolecular filament formation.

of type b, the following moment equations can be written down:

$$\begin{aligned}
\frac{d}{dt}P_a(t) &= 2k_n(aa)m_a(t)^2 + k_n(ab)m_a(t)m_b(t) \\
&\quad + k_+(a|b)m_a(t)P_b(t) - k_+(b|a)m_b(t)P_a(t) \\
\frac{d}{dt}P_b(t) &= 2k_n(bb)m_b(t)^2 + k_n(ab)m_a(t)m_b(t) \\
&\quad - k_+(a|b)m_a(t)P_b(t) + k_+(b|a)m_b(t)P_a(t) \qquad (4.3a) \\
\frac{d}{dt}M_a(t) &= k_+(a|b)m_a(t)P_b(t) + k_+(a|a)m_a(t)P_a(t) \\
\frac{d}{dt}M_b(t) &= k_+(b|b)m_b(t)P_b(t) + k_+(b|a)m_b(t)P_a(t). \qquad (4.3b)
\end{aligned}$$

Note that in arriving at these equations we have assumed the contribution of nucleation to monomer depletion to be negligible [48].

The different terms in (4.3) have straightforward interpretations. The first two terms in the same equations are analogous to single-species terms and represent the production of ends of the relevant type through nucleation. The terms in the final two equations are similarly analogous and represent the contribution of self- and cross-elongation to monomer depletion. However, the final two terms in the first two equations have no direct equivalence to single-species processes and are “drift” terms that represent the inter-conversion of ends of type a and type b due to elongation.

Note that if one of the monomer populations is depleted from solution before the other, the dynamics reduce to that of a single-species system, and may be qualitatively different from the pre-depletion dynamics if the elongation rate constants differ significantly. If this is the case the reaction is governed by two timescales: a copolymer timescale and a single-species timescale, with the dominant timescale switching from the former to the latter as the first monomer species is depleted. This feature is in contrast to the single-species Oosawa model, whose dynamics is controlled by a single timescale. This behaviour is visible in the numerical solutions to the rate equations (4.3) plotted in Fig. 4.3(b) and (d). Once one monomer species is depleted from the solution, the rate of depletion of the other monomer shows a “kink” and slows significantly, as it comes to be controlled by the slower single-species timescale.

4.4 The Mixed Oosawa Model: Early-Time Behaviour

Having formulated the fundamental set of equations governing the dynamics of $P_x(t)$ and $M_x(t)$, we now study the steady-state case in which there is a flux of monomer into the system such that the monomer concentration remains constant at all times. We find that this condition renders the moment equations (4.3) exactly soluble. Moreover, since monomer concentration varies little at early times, this solution is valid in the limit of early time, even when mass conservation is imposed. We consider the most general case in which there is significant cross-elongation, as the early-time mixed Oosawa model without significant cross-elongation is essentially trivial to solve.

4.4.1 Linearized early-time moment equations

At early times $m_a(t)$ and $m_b(t)$ can be approximated by their initial values, $m_a(0)$ and $m_b(0)$, as monomer has not been significantly depleted. We write $M_a^{(0)}(t)$ and $M_b^{(0)}(t)$ as the early-time fibril mass concentrations, and $P_a^{(0)}(t)$ and $P_b^{(0)}(t)$ as the early-time fibril end concentrations. The moment equations can then be written as:

$$\begin{aligned}
\frac{dP_a^{(0)}}{dt} &= \alpha_a + \mu_{ab} P_b^{(0)}(t) - \mu_{ba} P_a^{(0)}(t) \\
\frac{dP_b^{(0)}}{dt} &= \alpha_b - \mu_{ab} P_b^{(0)}(t) + \mu_{ba} P_a^{(0)}(t) \qquad (4.4a) \\
\frac{dM_a^{(0)}}{dt} &= \mu_{ab} P_b^{(0)}(t) + \mu_{aa} P_a^{(0)}(t) \\
\frac{dM_b^{(0)}}{dt} &= \mu_{bb} P_b^{(0)}(t) + \mu_{ba} P_a^{(0)}(t), \qquad (4.4b)
\end{aligned}$$

where α_a and α_b describe the initial rates of primary nucleation; and μ_{ab}, μ_{ba}, μ_{aa} and μ_{bb} describe the initial rates of elongation. The rates are constant and are given by $\alpha_x = \sum_y (1+\delta_{xy}) k_n(xy) m_x(0) m_y(0)$, and $\mu_{xy} = k_+(x|y) m_x(0)$.

4.4.2 Solving the early-time equations

The linearized ordinary differential equations (ODEs) for the polymer ends (4.4a) are solved, when cross-elongation is present, to give:

$$\begin{aligned}
P_a^{(0)}(t) = P_a(0)e^{-\mu_c t} &+ \frac{\alpha_t}{\mu_c}\mu_{ab} t \\
&+ \frac{1}{\mu_c^2}\Big[(y + \mu_{ab} P(0)\mu_c)(1 - e^{-\mu_c t})\Big] \\
P_b^{(0)}(t) = P_b(0)e^{-\mu_c t} &+ \frac{\alpha_t}{\mu_c}\mu_{ba} t \\
&+ \frac{1}{\mu_c^2}\Big[(\mu_{ba} P(0)\mu_c - y)(1 - e^{-\mu_c t})\Big], \qquad (4.5)
\end{aligned}$$

where the total early-time nucleation rate $\alpha_t = \alpha_a + \alpha_b$, the early-time cross-elongation rate $\mu_c = \mu_{ab} + \mu_{ba}$, $P(0) = P_a(0) + P_b(0)$, and $y = \alpha_a \mu_{ba} - \alpha_b \mu_{ab}$.

For extended fibrils to form, the timescale of elongation must be much more rapid than the timescale of the overall aggregation reaction. Thus, when cross-elongation is significant,

$\mu_c t \gg 1$ for any time of interest, and (4.5) reduces to:

$$\begin{aligned} P_a^{(0)}(t) &= \frac{\mu_{ab}}{\mu_c}\left(\alpha_t t + P(0)\right) \\ P_b^{(0)}(t) &= \frac{\mu_{ba}}{\mu_c}\left(\alpha_t t + P(0)\right). \end{aligned} \tag{4.6}$$

This result can be understood as the fibril end concentrations rapidly equilibrating at early times due to cross-elongation. The same results can therefore be arrived at by calculating the total early-time fibril end concentration, and applying to it the equilibrium ratio of end types, given by setting the cross-elongation rates equal to one another. These expressions (4.6) can be substituted into the equations for the mass concentrations (4.4b), to give early-time solutions:

$$\begin{aligned} M_a^{(0)}(t) &= \frac{\alpha_t}{2\mu_c}\mu_{ab}(\mu_{aa} + \mu_{ba})t^2 \\ &\quad + \frac{P(0)}{\mu_c}\mu_{ab}(\mu_{aa} + \mu_{ba})t + M_a(0) \\ M_b^{(0)}(t) &= \frac{\alpha_t}{2\mu_c}\mu_{ba}(\mu_{ab} + \mu_{bb})t^2 \\ &\quad + \frac{P(0)}{\mu_c}\mu_{ba}(\mu_{ab} + \mu_{bb})t + M_b(0). \end{aligned} \tag{4.7}$$

These expressions (4.7) are plotted against numerical solutions to the full-time rate laws (4.3) in Fig. 4.3, and show excellent agreement at early times. At later times they continue to grow indefinitely as t^2, whereas the numerical solutions, after an initial t^2 growth phase, converge to a finite maximum filament mass concentration. This difference is due to the fact that monomer concentration is kept constant in the early-time solutions; thus, the rates of reaction remain constant and fibril concentration increases unboundedly. On the other hand, the full-time moment equations feature depletion of monomer and thus the rate of filament growth tends to zero as the monomers all become incorporated into filaments.

4.4.3 Lag times and scaling

The availability of closed form expressions for the early time dynamics is already able to reveal important insights into the problem of copolymerization. For instance, the scaling of the time taken to deplete a given proportion of each monomer type in the absence of seed can be obtained, provided the proportion of monomer consumed is relatively low. These times are known as lag times. Rearranging our early-time relations (4.7), setting

$P(0) = M(0) = 0$, we obtain:

$$\begin{aligned}
t_{\mathrm{lag},x;a} &= \left(\frac{2x\mu_c/\alpha_t}{k_+(a|a)\mu_{ab} + k_+(a|b)\mu_{ba}}\right)^{1/2} \\
t_{\mathrm{lag},x;b} &= \left(\frac{2x\mu_c/\alpha_t}{k_+(b|a)\mu_{ab} + k_+(b|b)\mu_{ba}}\right)^{1/2} \\
t_{\mathrm{lag},x} &= \left(\frac{2xm(0)\mu_c/\alpha_t}{(\mu_{aa} + \mu_{ba})\mu_{ab} + (\mu_{ab} + \mu_{bb})\mu_{ba}}\right)^{1/2},
\end{aligned} \tag{4.8}$$

where $t_{\mathrm{lag},x;y}$ is time taken to deplete a proportion x of the monomer of type y (or of all types in the absence of a subscript x).

4.4.4 Average size of aggregates

Another important parameter that can be accessed from Eqs. (4.6) and (4.7) is the average size of fibrils $L(t)$, which is simply $M(t)/P(t)$. In the absence of seed, i.e. setting $P(0) = M(0) = 0$, we obtain:

$$L(t) = \frac{\mu_{aa}\mu_{ab} + 2\mu_{ab}\mu_{ba} + \mu_{ba}\mu_{bb}}{2\mu_c}t. \tag{4.9}$$

As expected, the average size of aggregates increases linearly with time, as in the single species case.

4.5 Secondary Processes: Internal Composition

4.5.1 Motivation for study

Secondary processes, such as surface catalyzed secondary nucleation and fragmentation, are critical for the rapid formation of fibrils in amyloid diseases, and often directly produce those species responsible for the observed cell toxicity [13, 39, 92].

Unlike primary processes, which only involve the ends of polymer chains, secondary processes typically involve the internal parts of the polymer. For instance, in surface-catalysed nucleation, monomers of either type adsorb onto the surface all along the fibril length [20]. In a copolymerizing system, the physical and chemical properties of the polymers are not expected to be homogeneous along their length. Thus, to model a copolymerizing system in the presence of secondary processes, it is necessary to consider the internal compositions of the polymer chains away from the ends, which in principle requires us to solve the full master equation. However, such detailed knowledge is not actually necessary to describe the kinetics when we consider only nearest-neighbour interactions; instead, the local structure around a site of secondary nucleation or of fragmentation is all that is

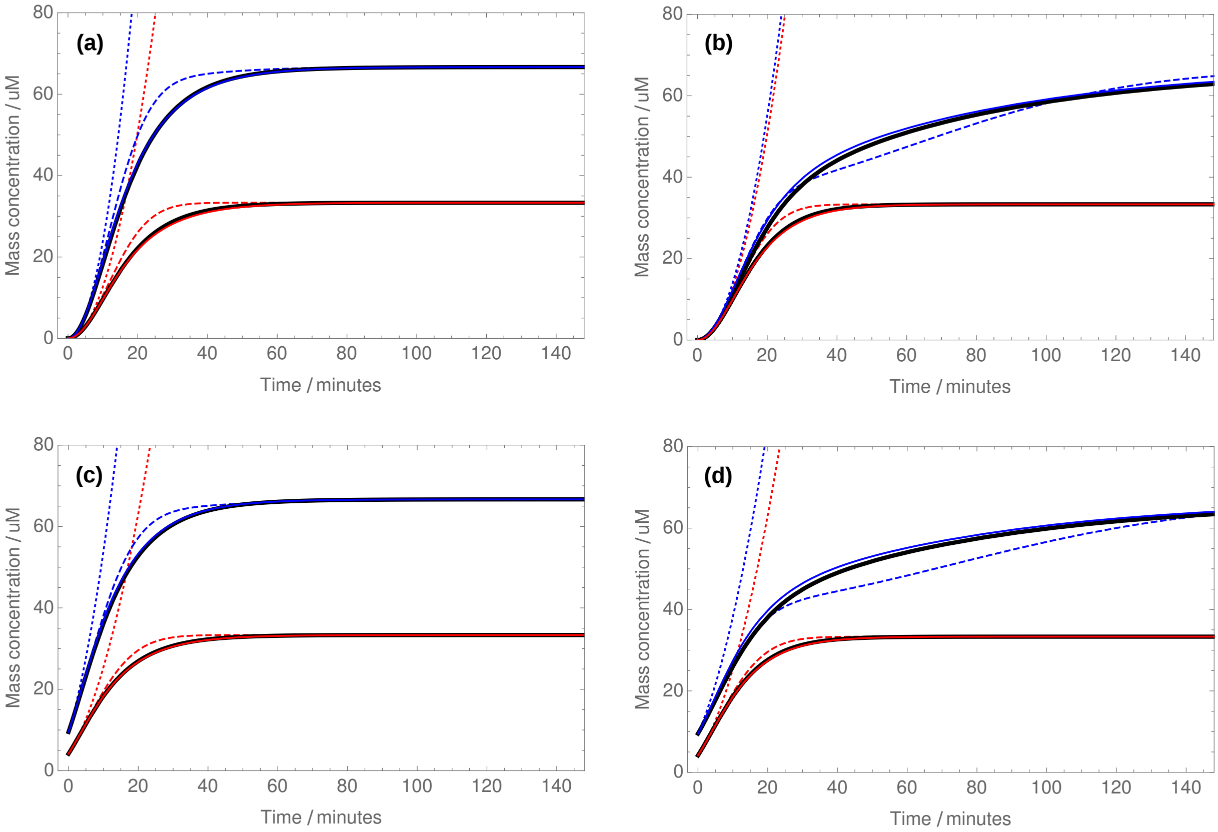

Figure 4.3: A plot of mass concentration vs time for a primary nucleation-dominated system, for parameter values consistent with extended fibril formation (see Appendix B.5). Red is the rapidly depleting species a, blue is species b. Dotted lines are early-time solutions (4.7), solid black lines are numerical solutions to the moment equations (4.3). (a) is unseeded, and with similar monomer depletion rates. (b) is unseeded, and with different monomer depletion rates. (c) is seeded, and with similar monomer depletion rates. (d) is seeded, and with different monomer depletion rates. Dashed and solid coloured lines are approximate full-time analytical solutions not presented in this thesis.

needed to know. This can be most obviously seen in the case of fragmentation, where the rate of breakage depends on the identity of the species being separated. Furthermore, the rate of formation of secondary nuclei of reaction order $n_2 = 2$ will also depend on the concentration of adjacent pairs of subunits of different types with which they may interact. We illustrate these processes, and their rate constants, in Fig. 4.4. We use $n_2 = 2$ throughout, although the methodology can be applied in principle to any reaction order.

4.5.2 Generalizing the moment equations to account for secondary processes

For a copolymerizing system involving subunit-pair-dependent secondary processes, the previously-derived moment equations (4.3) are extended to include these new processes as follows:

$$\begin{aligned}
\frac{dP_a}{dt} &= 2k_n(aa)m_a(t)^2 + k_n(ab)m_a(t)m_b(t) \\
&\quad + k_+(a|b)m_a(t)P_b(t) - k_+(b|a)m_b(t)P_a(t) \\
&\quad + \sum_{pq,xy} (\delta_{pa} + \delta_{qa})k_2(pq|xy)m_p(t)m_q(t)N_{xy}(t) \\
\frac{dP_b}{dt} &= 2k_n(bb)m_b(t)^2 + k_n(ab)m_a(t)m_b(t) \\
&\quad - k_+(a|b)m_a(t)P_b(t) + k_+(b|a)m_b(t)P_a(t) \\
&\quad + \sum_{pq,xy} (\delta_{pb} + \delta_{qb})k_2(pq|xy)m_p(t)m_q(t)N_{xy}(t) \qquad \text{(4.10a)} \\
\frac{dM_a}{dt} &= k_+(a|b)m_a(t)P_b(t) + k_+(a|a)m_a(t)P_a(t) \\
\frac{dM_b}{dt} &= k_+(b|b)m_b(t)P_b(t) + k_+(b|a)m_b(t)P_a(t), \qquad \text{(4.10b)}
\end{aligned}$$

where $N_{xy}(t)$ is the concentration of adjacent pairs of subunits of type xy at time t, $k_2(pq|xy)$ is the rate constant for production of secondary nuclei of type pq on adjacent subunit pairs of type xy, and the sums are over the possible monomer and subunit pairs. This notation is valid assuming we can directly interpret $k_2(pq|xy)$ as the rate constant for formation of a nucleus of type pq, at the site of an xy subunit pair. If we wish to relax this assumption we can use a more general notation $k_2(z; pq|xy)$ as the rate constant for formation of ends of type z through secondary processes with reaction order 1 with respect to each of $m_p(t)$, $m_q(t)$, and $N_{xy}(t)$ (if $p = q$ then the reaction order sums to 2). This allows for instance that ends of type a form through secondary processes with reaction order 0 with respect to $m_a(t)$, and 2 with respect to $m_b(t)$.

If fragmentation is occurring but secondary nucleation is absent, we instead have:

$$\begin{aligned}
\frac{dP_a}{dt} &= 2k_n(aa)m_a(t)^2 + k_n(ab)m_a(t)m_b(t) \\
&\quad + k_+(a|b)m_a(t)P_b(t) - k_+(b|a)m_b(t)P_a(t) \\
&\quad + \sum_{xy}(\delta_{xa} + \delta_{ya})k_-(xy)N_{xy}(t) \\
\frac{dP_b}{dt} &= 2k_n(bb)m_b(t)^2 + k_n(ab)m_a(t)m_b(t) \\
&\quad - k_+(a|b)m_a(t)P_b(t) + k_+(b|a)m_b(t)P_a(t) \\
&\quad + \sum_{xy}(\delta_{xb} + \delta_{yb})k_-(xy)N_{xy}(t), \qquad \text{(4.11)}
\end{aligned}$$

where $k_-(xy)$ is the rate constant for filament fragmentation occurring at a subunit pair of composition xy. Note that at early times, when we can treat monomer concentrations $m(t)$ as constant, all these equations for different possible secondary processes have identical functional form.

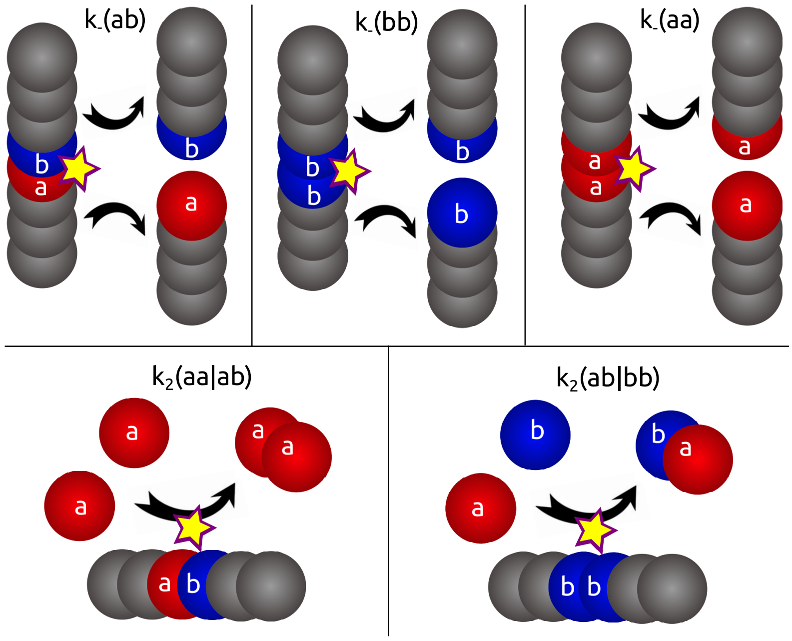

Figure 4.4: Representative possible fragmentation and secondary nucleation events in a 2-species copolymerizing system.

It is important to note that, due to the inclusion of $N_{xy}(t)$ terms, these equations are no longer closed. We can, however, write down ODEs describing the time evolution of the pair concentrations:

$$\begin{aligned}\frac{dN_{aa}}{dt} &= k_+(a|a)m_a(t)P_a(t)\\ \frac{dN_{ab}}{dt} &= k_+(a|b)m_a(t)P_b(t) + k_+(b|a)m_b(t)P_a(t)\\ \frac{dN_{bb}}{dt} &= k_+(b|b)m_b(t)P_b(t).\end{aligned} \tag{4.12}$$

As with the 1st moments, these equations are accurate in the limit that elongation is much faster than nucleation. They depend only on the 0th- and 1st moments of the distribution; thus, the extended moment equations featuring these equations along with those for the 0th and 1st moments is closed when combined with a conservation-of-mass relation, and can in principle be solved uniquely.

4.6 Secondary Processes: Analytical Early-Time Solutions

The early-time behaviour of the system is interesting to study in its own right and provides a framework for interpreting lag time experiments. Here we discuss the hardest possible case, in which cross-elongation plays a significant role. In Appendix B.3 we discuss the simpler case in which no appreciable cross-elongation occurs, and also the case in which no cross-secondary processes occur.

4.6.1 Early-time moment equations

At early time, the rates $(\delta_{px} + \delta_{qx})k_2(pq|yz)m_p(t)m_q(t)$ as well as the rates of primary nucleation become time-independent, as we fix $m_a(t)$ and $m_b(t)$ at their initial values. The system of moment equations are now linear. Furthermore, the 1st moment ODEs are now decoupled from the rest and we need only solve the filament end concentration and subunit pair concentration equations simultaneously to calculate the mass concentrations. We write these in matrix form:

$$\frac{d}{dt}\begin{pmatrix} P_a^{(0)}(t) \\ P_b^{(0)}(t) \\ N_{aa}^{(0)}(t) \\ N_{ab}^{(0)}(t) \\ N_{bb}^{(0)}(t) \end{pmatrix} = \begin{pmatrix} \alpha_a \\ \alpha_b \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -\mu_{ba} & \mu_{ab} & \alpha_{aa}^{2a} & \alpha_{ab}^{2a} & \alpha_{bb}^{2a} \\ \mu_{ba} & -\mu_{ab} & \alpha_{aa}^{2b} & \alpha_{ab}^{2b} & \alpha_{bb}^{2b} \\ \mu_{aa} & 0 & 0 & 0 & 0 \\ \mu_{ba} & \mu_{ab} & 0 & 0 & 0 \\ 0 & \mu_{bb} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} P_a^{(0)}(t) \\ P_b^{(0)}(t) \\ N_{aa}^{(0)}(t) \\ N_{ab}^{(0)}(t) \\ N_{bb}^{(0)}(t) \end{pmatrix}, \tag{4.13}$$

where $\alpha_{yz}^{2x} = \sum_{pq}(\delta_{px} + \delta_{qx})k_2(pq|yz)m_p(0)m_q(0)$, and $\alpha_{yz}^{2x}N_{yz}(t)$ represents the early-time rate of formation of ends of type x through secondary nucleation at sites of composition yz. For a system undergoing fragmentation but not secondary nucleation, the results are identical except $k_2(pq|yz)m_p(0)m_q(0)$ is replaced with $k_-(pq)$.

4.6.2 Solutions to the extended moment equations

These equations cannot be exactly solved in a useful manner, since they give rise to a quintic eigenvalue equation. We instead non-dimensionalise the eigenvalue equation and identify the dimensionless parameter $\nu = (\sum_{i=a,b}[\alpha_{aa}^{2i} + \alpha_{ab}^{2i} + \alpha_{bb}^{2i}])/(\mu_{ab} + \mu_{ba})$, and note that $\nu << 1$, under the assumption that there is significant cross-elongation. We perform dominant balance on the eigenvalue equation using this small parameter to obtain approximate eigenvalues (see Appendix B.2). We can then obtain the corresponding approximate eigenvectors, by considering only the leading-order behaviour in ν of the eigenvector equations. After a very short initial adjustment time, Eq. (4.13) is controlled by the only positive eigenvalue λ_+, and its associated eigenvector. We neglect contributions to the full inhomogeneous solution from the remaining eigenvalue-eigenvector pairs, since these have a negligible effect over most of the reaction time course. Ultimately we obtain the following expressions for the early-time fibril end concentrations:

$$\begin{aligned} P_a^{(0)}(t) &= \frac{\alpha_t\mu_{ab}}{2\lambda_+\mu_c}\left(e^{\lambda_+ t} - 1\right) \\ P_b^{(0)}(t) &= \frac{\alpha_t\mu_{ba}}{2\lambda_+\mu_c}\left(e^{\lambda_+ t} - 1\right), \end{aligned} \tag{4.14}$$

with

$$\lambda_+ = \sqrt{(\alpha^{2t}_{aa}\mu_{aa}\mu_{ab} + 2\alpha^{2t}_{ab}\mu_{ab}\mu_{ba} + \alpha^{2t}_{bb}\mu_{ba}\mu_{bb})/\mu_c}, \tag{4.15}$$

and $\alpha^{2t}_{xy} = \alpha^{2a}_{xy} + \alpha^{2b}_{xy}$; and subunit pair concentrations:

$$\begin{aligned} N^{(0)}_{aa}(t) &= \frac{\alpha_t \mu_{aa}\mu_{ab}}{2\lambda_+^2 \mu_c}\left(e^{\lambda_+ t} - 1\right) \\ N^{(0)}_{ab}(t) &= \frac{\alpha_t \mu_{ab}\mu_{ba}}{\lambda_+^2 \mu_c}\left(e^{\lambda_+ t} - 1\right) \\ N^{(0)}_{bb}(t) &= \frac{\alpha_t \mu_{ba}\mu_{bb}}{2\lambda_+^2 \mu_c}\left(e^{\lambda_+ t} - 1\right). \end{aligned} \tag{4.16}$$

We can then use the early-time mass concentration equations with these solutions to calculate:

$$\begin{aligned} M^{(0)}_{a}(t) &= \frac{\alpha_t (\mu_{aa} + \mu_{ba})\mu_{ab}}{2\lambda_+^2 \mu_c}\left(e^{\lambda_+ t} - 1\right) \\ M^{(0)}_{b}(t) &= \frac{\alpha_t (\mu_{ab} + \mu_{bb})\mu_{ba}}{2\lambda_+^2 \mu_c}\left(e^{\lambda_+ t} - 1\right). \end{aligned} \tag{4.17}$$

Seeded solutions are given in Appendix B.2. These are plotted against numerical solutions to the rate equations (4.10), (4.12) in Fig. 4.5, and show excellent agreement at early times, as expected. As with the Oosawa early-time solutions, they diverge away from the numerical solutions at later times due to the lack of monomer conservation in the early-time rate equations.

4.6.3 Lag Times

For a system without seed, it is interesting to obtain an expression for the time taken to deplete a given proportion of the monomer. We can use early-time expressions (4.17) to do this provided the proportion of depleted monomer is relatively low. Neglecting small terms, we obtain:

$$t_{\text{lag},x} = \frac{1}{\lambda_+}\ln\left[\frac{2\lambda_+^2 x\, m_{\text{tot}}\mu_c}{\alpha_t(\mu_{aa}\mu_{ab} + 2\mu_{ab}\mu_{ba} + \mu_{ba}\mu_{bb})}\right], \tag{4.18}$$

where x is the proportion of monomer that has been depleted. We immediately see that the behaviour of the lag time is controlled predominantly by λ_+, with primary nucleation entering the expression only logarithmically. This is in analogy with the role of κ in a single-species system with secondary processes.

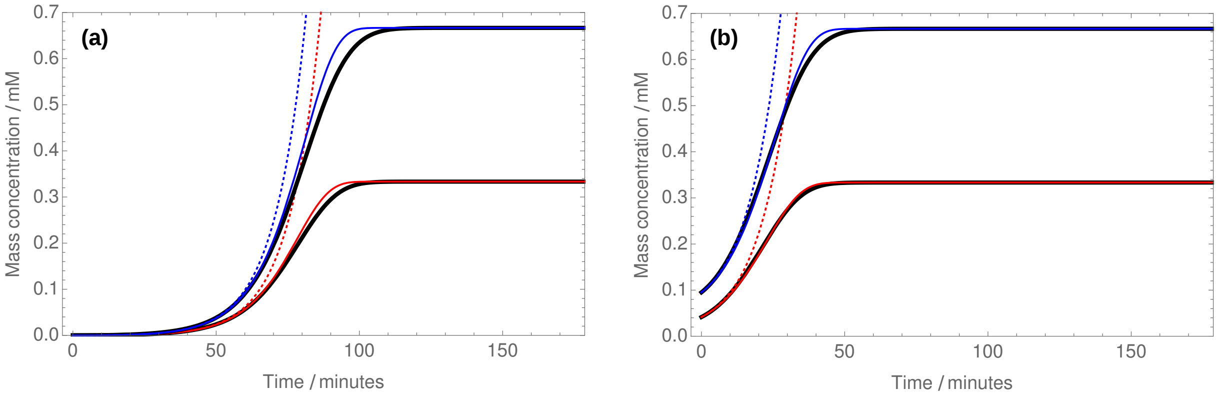

Figure 4.5: A plot of mass concentration vs time for a fragmentation-dominated system, for parameter values consistent with extended fibril formation (see Appendix B.6). Red is the rapidly depleting species a, blue is species b. Dotted lines are early-time solutions (4.17), and black lines are numerical solutions to the moment equations (4.10), (4.12). (a) is unseeded, and (b) seeded. Solid coloured lines are approximate analytical solutions not presented in this thesis.

4.6.4 Average size

The average size of fibrils is simply $M(t)/P(t)$. We use the unseeded early-time expressions (4.14) and (4.17) to obtain:

$$L(t) = \frac{\mu_{aa}\mu_{ab} + 2\mu_{ab}\mu_{ba} + \mu_{ba}\mu_{bb}}{\lambda_+\mu_c}. \tag{4.19}$$

This expression reduces straightforwardly to the single-species expression [30].

4.7 Investigating Internal Composition

So far, we have derived general solutions to the dynamics of mixed fibril number and mass concentration with and without secondary processes. The next question is how the internal structure of these fibrils evolves with time. To answer this question we introduce the pair correlation function, $\mathcal{P}(xy;t)$, which is defined as the probability that a randomly selected fibril subunit pair will have a particular identity xy at time t, and measures the time-dependent spatial correlation of a given pair of subunits in a fibril. We also introduce the related correlation function $\mathcal{P}(x-n-y;t)$, in which the subunits are separated by n unknown residues; and also consider the n-tuple correlation function $\mathcal{P}(xyz...m...pqr;t)$, giving the probability that a randomly selected fibril subunit n-tuple will have a particular identity $xyz...m...pqr$ at time t. We wish to explore analytically the behaviour of these quantities. Note that the obvious approach to accessing the pair correlations, based on normalizing the quantities $N_{xy}(t)$, although conceptually simpler, is less easily generalizable to n-tuple correlations. We demonstrate the equivalence of this approach in Appendix B.4.

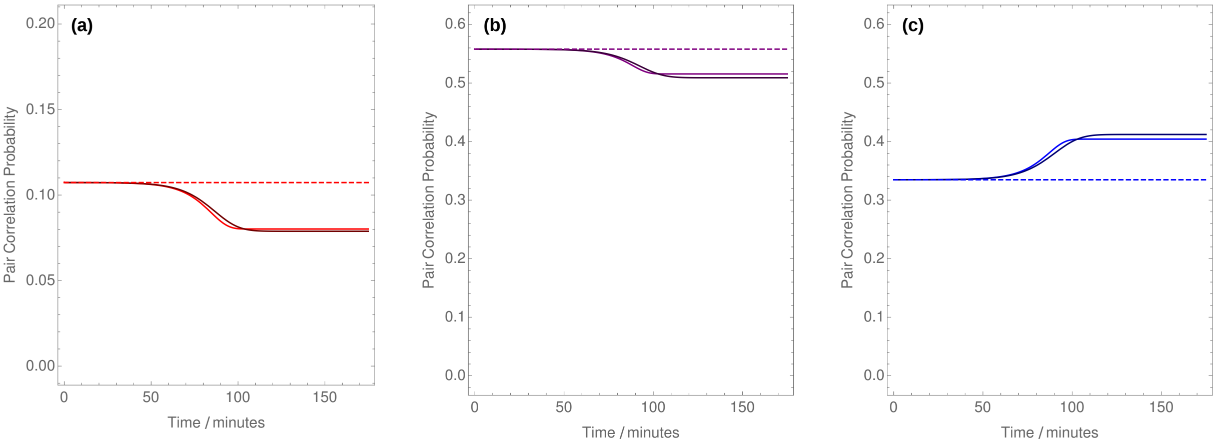

Figure 4.6: A plot of the pair correlation functions vs time for a fragmentation-dominated system, for parameter values used in Fig. 4.5(a) (see Appendix B.6). Dotted lines are early-time solutions (accessible from Eqs. (4.16) and (4.17)), and solid lines are approximate analytical solutions (not presented here). Darker solid lines are numerical solutions to the extended moment equations (4.10), (4.12). (a) examines aa pairs ($\mathcal{P}(aa)(t)$); (b) is for ab pairs, and (c) is for bb pairs.

4.7.1 Formulating the pair correlation

Provided elongation is much faster than primary and secondary nucleation (as it must be for extended fibrils to form, see Sec. 4.6.4) the vast majority of subunits are formed by elongation, and are not ends. Thus, for a given pair of residues in a fibril that formed at early times, to a good approximation, one will have formed by elongation from the other. Given that the first residue in the pair to form is of type x, the probability that the next one to form is of type y is given by $\mathcal{G}(y|x)$, where:

$$\begin{aligned}\mathcal{G}(a|a) &= \frac{\mu_{aa}}{\mu_{aa}+\mu_{ba}}\\ \mathcal{G}(b|b) &= \frac{\mu_{bb}}{\mu_{bb}+\mu_{ab}}\\ \mathcal{G}(a|b) &= \frac{\mu_{ab}}{\mu_{ab}+\mu_{bb}}\\ \mathcal{G}(b|a) &= \frac{\mu_{ba}}{\mu_{ba}+\mu_{aa}}.\end{aligned} \tag{4.20}$$

Since the elongation process has been assumed to be fast, formation-time-specific conditional composition probabilities for non-adjacent sites can be written down as a concatenation of these terms. We define the matrix:

$$\mathbf{G} = \begin{pmatrix} \mathcal{G}(a|a) & \mathcal{G}(a|b) \\ \mathcal{G}(b|a) & \mathcal{G}(b|b) \end{pmatrix}. \tag{4.21}$$

We can then construct the formation probabilities of particular n-tuples in a given direction simply by concatenating the appropriate matrix elements, and obtain the directionless formation probability by adding the expressions for each direction if they differ.

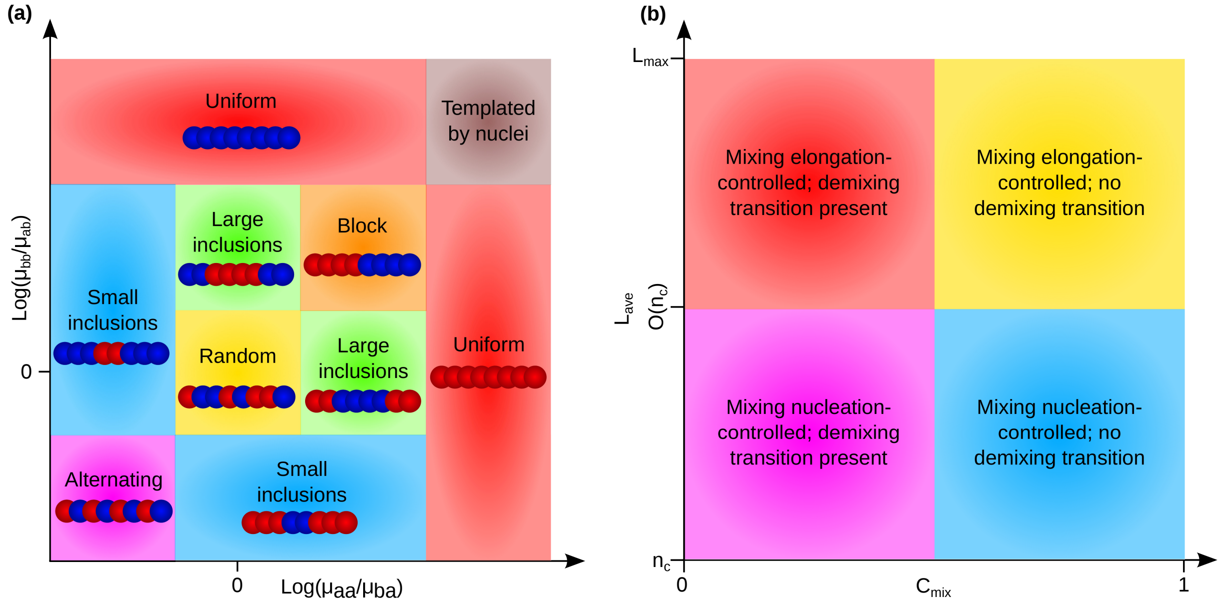

Figure 4.7: (a) A plot of the possible kinds of internal structure in the initial stages of a two-species copolymerizing system, before significant monomer depletion has occurred. Monomers of type a are labelled red, and b are blue. For large enough c_{xy}, no cross-elongation is seen during the early stages of aggregation and perfectly uniform filaments are formed; for small enough c_{xy}, no self-elongation occurs, and perfectly alternating filaments form. (b) Charting the different possible mixing and demixing behaviours of a copolymerizing system, dependent on the average filament length L_{ave}, and the mixing coefficient C_{mix}. For values of C_{mix} approaching zero, one monomer type is depleted before the other, and at later times homo-polymers of the remaining monomer type are formed. For short enough filaments the internal composition of aggregates is controlled by the nucleation processes forming them, not the subsequent elongation reactions. See Sec. 4.7.2 for full details.

Also of interest is the probability that, given a selected fibril subunit of type x, a subunit of type y is found n sites away from it. This is given by:

$$\begin{aligned}
\mathcal{P}(a|n|a) &= [\mathbf{G}^n]_{11}\\
\mathcal{P}(a|n|b) &= [\mathbf{G}^n]_{12}\\
\mathcal{P}(b|n|a) &= [\mathbf{G}^n]_{21}\\
\mathcal{P}(b|n|b) &= [\mathbf{G}^n]_{22}.
\end{aligned} \tag{4.22}$$

Fully general directionless site correlation functions are given by:

$$\mathcal{P}(x;n;y) = \frac{\mathcal{P}(x|n|y) + \mathcal{P}(y|n|x)}{1+\delta_{xy}}. \tag{4.23}$$

The probability of a randomly selected subunit pair having a particular composition is then $\mathcal{P}(x;1;y)$ which we shorten for convenience to $\mathcal{P}(xy)$.

4.7.2 Steady-state composition and de-mixing

From the form of the correlation function calculated (4.20), we see that the correlation behaviour depends only on the dimensionless parameters $c_{xy} = \mu_{yy}/\mu_{xy}$. These can be interpreted as the preference for an end of type y to add a monomer of the same type or a different type. We may use these to construct a convenient diagram (Fig. 4.7(a)) that classifies the possible kinds of internal structure being produced at early times in a two-species system. Where the parameters c_{xy} are smaller, there is a greater preference for cross-bonds to form, and domain lengths are smaller. In the limit that the parameters c_{xy} approach zero, a perfect alternating copolymer is obtained. In the opposite limit $c_{xy} \to \infty$, perfectly uniform homo-polymers are obtained. In practice, however, uniform homo-polymers are seen at finite values of c_{xy}, since for sufficiently large c_{xy}, the timescale of cross-elongation becomes longer than the time taken for the polymerization reaction to complete. Under these conditions no mixed filaments or domain walls are seen during the early stages of the reaction. For the same reason, perfectly alternating polymers are produced at finite values of $\log(c_{xy})$. The axes of Fig. 4.7(a) are $\log(c_{xy})$ in order that these limits can be represented in a single plot.

If the monomer species deplete at the same rate, then the structure produced at all times is accurately given by Fig. 4.7(a); however, typically one monomer depletes before the other (type a by convention depletes first). The condition for this happening, in the absence of seed, can be derived by noting that the relative early-time rates of depletion must be different if one monomer is to deplete first. Therefore the ratio $m_a(0)M_b^{(0)}(t)'/m_b(0)M_a^{(0)}(t)'$ must not be close to 1. We name this ratio the mixing coefficient C_{mix}, and calculate it as $k_+(b|a)(\mu_{ab}+\mu_{bb})/(k_+(a|b)(\mu_{aa}+\mu_{ba}))$. This argument can be equivalently formulated in terms of the tenth times rather than the early-time rates of depletion, leading to the same formula. The type of structure then being formed migrates upward and to the left on Fig. 4.7(a) as the reaction progresses; and once a is totally depleted we have total de-mixing and only uniform type-b polymer forms. De-mixing can also occur in a fragmenting system where the rate of hetero-fragmentation $k_-(ab)$ is large, and much bigger than the other fragmentation rates.

A final parameter that is important in determining both steady-state composition and de-mixing at later times is the average filament length L_{ave}. For most systems of interest, extended filaments form such that the average filament length is far greater than the critical nucleus sizes, and the composition of fibrils at any time are determined principally by the elongation rates. However, for systems undergoing very little elongation, fibril composition becomes controlled by the relative rates of homomolecular and heteromolecular nucleation processes. Moreover, de-mixing is also controlled by concentration-dependent changes in these rates. The influence of C_{mix} and of the average filament length L_{ave} on the mixing behaviour of a copolymerizing system is shown schematically in Fig. 4.7(b).

4.8 Correlation lengths

One of the few prior analytical studies of copolymerisation kinetics was performed by Whitelam *et al* [93]. In their paper, they consider mixed filaments of two subunit types, growing under conditions of constant monomer concentration. Their work differs to ours in several key regards. Firstly, they do not consider nucleation processes, studying instead the growth of pre-existing filaments. Secondly, they investigate the near-equilibrium limit where fibril depolymerisation is appreciable, whereas we consider situations where depolymerisation is negligible (as is typically the case in amyloid fibril formation). Thirdly, they consider stochastic kinetics whereas our model is deterministic. Finally, they consider elongation to occur only at one rate, regardless of the composition of the fibre end or of the type of monomer being added.

They conclude that under these circumstances the fibre composition is dependent on the extent of supersaturation, such that the equilibrium fibre structure cannot be attained when there is any supersaturation. Interestingly, they also conclude that a desired "equilibrium" structure can be achieved for a given level of supersaturation by instead changing the depolymerisation rate constants. To achieve this in practice, they provide an analytical formula linking the mean domain length, the supersaturation and the rate constants together.

This control can also be achieved in our model, by varying the rate constants. It is a worthwhile goal to derive analytical expressions for the mean domain lengths, which we do as follows. We have already derived the probability of adding a type-a monomer onto a type-a end as $\mathcal{G}(a|a) = (1 + \mu_{ba}/\mu_{aa})^{-1}$. The probability of forming a type-a domain of length n_a is $P(n_a) = \mathcal{G}(a|a)^{n_a-1}\mathcal{G}(b|a)$. Writing $\mathcal{G}(a|a)$ as g_a, the mean domain length is then given by

$$\begin{aligned}\langle n_a\rangle &= \sum_{n_a=1}^{\infty} n_a P(n_a) = \mathcal{G}(b|a)\sum_{n_a=1}^{\infty} n_a \mathcal{G}(a|a)^{n_a-1}\\ &= (1-g_a)\frac{d}{dg_a}\sum_{n_a=1}^{\infty} g_a^{n_a} = (1-g_a)\frac{d}{dg_a}\left(\frac{g_a}{1-g_a}\right)\\ &= (1-g_a)\left(\frac{1}{1-g_a} + \frac{g_a}{(1-g_a)^2}\right) = \frac{1}{1-g_a}. \end{aligned} \tag{4.24}$$

The result for domains of type-b is identical but with indices permuted; substituting the expressions for g_a and g_b, we can finally write:

$$\langle n_a\rangle = 1 + \frac{k_+(a|a)}{k_+(b|a)}\frac{m_a}{m_b}; \qquad \langle n_a\rangle = 1 + \frac{k_+(b|b)}{k_+(a|b)}\frac{m_b}{m_a}. \tag{4.25}$$

We see clearly that, for given rate constants, we can obtain any desired type-a or type-b domain length by varying the ratio of monomer concentrations. However, we cannot control both domain lengths simultaneously in this way; changing the ratio increases the mean length of one type of domain but shrinks the other. Alternatively, as suggested by Whitelam *et al*, target type-a and type-b domain lengths can be obtained for any ratio of monomer concentrations by tailoring the cross-elongation rate constants $k_+(b|a)$ and $k_+(a|b)$.

4.9 Implications for Experimental Investigations

This paper has focused on quantities analogous to those typically measurable in experiments on single-species filament formation. An arsenal of formulae has been developed that interprets these quantities in terms of microscopic rate constants. Hence, by applying techniques similar to those currently used in single-species studies to mixed systems [16], the extent to which different protein species coaggregate can now be determined quantitatively. In this section, details of how this may be done are given for two key classes of experimentally measurable quantities.

4.9.1 Scaling of tenth times

Lag times are the times taken for a given proportion x of monomer to be converted into filamentous material. Lag times are often measured experimentally for $x = 0.1$ and $x = 0.5$, whereupon they are named tenth times and half times respectively. The scaling of these quantities with initial monomer concentration can yield a wealth of information on the underlying kinetics. Herein, we focus solely on tenth times and consider only systems with primary processes as an example, but similar arguments can be made for the other possible cases.

Using Eq. (4.8) as our starting point, and given the single-species rate constants, lag time experiments can be used to determine the cross-elongation rate constants. The tenth times for type a and type b monomers are calculated for fixed total monomer concentration, but varying partitioning into type a and type b monomers. The square ratio of the two tenth times, r, is taken at each composition, and should obey the formula:

$$\left(\frac{t_{\mathrm{tenth;a}}}{t_{\mathrm{tenth;b}}}\right)^2 = \frac{k_+(b|a)k_+(a|b)f + k_+(b|b)k_+(b|a)(1-f)}{k_+(a|a)k_+(a|b)f + k_+(a|b)k_+(b|a)(1-f)}, \tag{4.26}$$

where $f = m_a(0)/m(0)$. This ratio is plotted in Fig. 4.8 for varying values of f.
Equation (4.26) provides a valuable tool for extracting important information about the system's kinetics. For example, the cross-elongation rate constants can be determined from a fit of the experimentally-determined r to Eq. (4.26). In the limits of $m_a(0) \to 0$ and $m_b(0) \to 0$, the square tenth time ratio reduces to $k_+(b|b)/k_+(a|b)$ and $k_+(b|a)/k_+(a|a)$ respectively, and full fitting is not needed, if these limits are well-determined by the

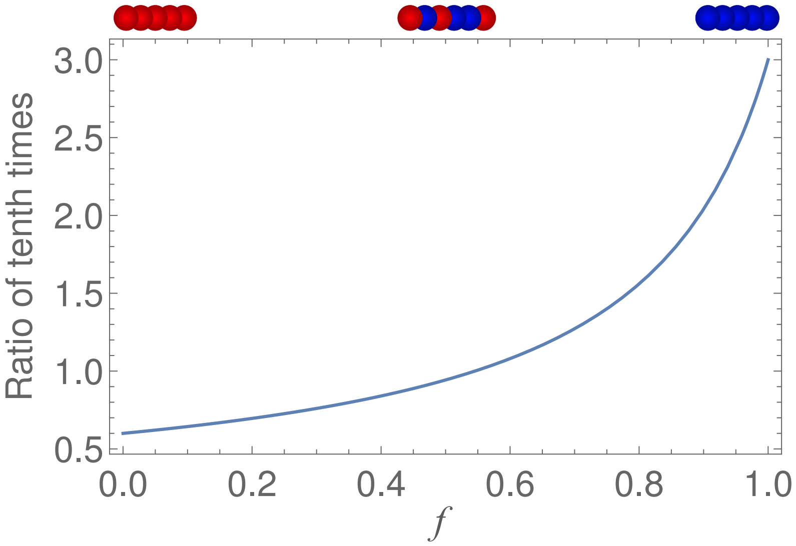

Figure 4.8: Square ratio of tenth times, $r = (t_{\text{tenth;a}}/t_{\text{tenth;b}})^2$, vs ratio of initial monomer concentrations, $f = m_a(0)/m(0)$, as given by (4.26), for the same parameter values as used in Fig. 4.3(a) (see Appendix B.5).

experimental data.

Finally, by taking the logarithm of the total tenth time given by (4.8), and setting $m_a(0) = m_b(0)$ we obtain:

$$\begin{aligned}\ln(t_{\text{tenth}}) = \frac{1}{2} \ln\left(\frac{0.2(k_+(a|b) + k_+(b|a))}{\sum_{i,j} k_+(i|j)k_+(j|j')}\right) \\ - \frac{1}{2}\ln\left(k_n(aa) + k_n(ab) + k_n(bb)\right) - \ln(m_a(0)),\end{aligned} \qquad (4.27)$$

where $j' \neq j$. Therefore, a log-log plot of lag time vs initial monomer concentration will give a straight line at $m_a(0) = m_b(0)$; and the sum of the primary nucleation rate constants can be found from the intercept of the log-log plot, if the elongation rate constants are known.

4.9.2 Scaling of filament mass concentration

Filament mass concentration, or free monomer concentration, can both be measured using a variety of experimental techniques, and analysed in exactly the same way by taking advantage of conservation of mass. The resultant kinetic traces will be qualitatively different for different types of copolymerizing behaviour. An example of this behaviour is the strong “kink” seen in the kinetic trace for a system without secondary processes exhibiting cross-elongation that is significantly faster than the self-elongation of one of the species, but not seen when the elongation rates are all similar. These qualitative differences can be taken advantage of to select the most appropriate formulae of the ones presented in this paper. At this point, seeded experiments can be used to determine the cross-elongation rate constants relative to the single-species rate constants. For instance, under highly seeded conditions the initial gradients of the filament mass concentrations, Eq. (4.4b), when divided through by the initial gradients of single-species aggregation reactions, become:

$$\begin{aligned} \frac{1}{\mu_{aa}P(0)} \left.\frac{dM_a^{(1)}(t)}{dt}\right|_{t=0} &= \frac{\mu_{ab} + \frac{k_+(a|b)}{k_+(a|a)}\mu_{ba}}{\mu_{ab}+\mu_{ba}} \\ \frac{1}{\mu_{bb}P(0)} \left.\frac{dM_b^{(1)}(t)}{dt}\right|_{t=0} &= \frac{\frac{k_+(b|a)}{k_+(b|b)}\mu_{ab} + \mu_{ba}}{\mu_{ab}+\mu_{ba}}. \end{aligned} \tag{4.28}$$

By fitting initial gradients to this expression for a range of initial monomer concentrations we can therefore obtain the cross-elongation rate constants.

4.10 Conclusions

We have examined copolymerizing behaviour in linear filament self-assembly for both primary and secondary process-dominated systems. We have then presented a mathematical approach that yields closed-form analytical expressions for the early-time dynamics of copolymerizing systems without secondary processes. Furthermore we have developed a formalism for modelling the internal composition of a copolymerizing system in a computationally feasible way. We have used this approach to generalize our methodology and obtain analytical closed form expressions for the early-time dynamics of copolymerizing systems with secondary processes.

We anticipate that this work will be of great interest to researchers studying the interaction between multiple protein types in the course of amyloid diseases. Particularly we expect that our work will be used to interpret experimental data on heteromolecular fibril formation, to discover the underlying dynamics for different coaggregating systems. However, given the paucity of quantitative experimental data on coaggregation at the time of publication, subsequently to this research I moved on to theoretical projects in the more experimentally-accessible field of oligomer kinetics. A priority for my research after my PhD will be to actually enter a lab and take a direct hand in helping develop quantitative experimental techniques for coaggregation, in collaboration with experimentalists in the Linse group in Lund, Sweden, to get this hugely exciting field moving.

Part III

The Kinetics of Amyloid Oligomers

Chapter 5

Oligomeric Intermediates Initiate Fibril Formation by the Yeast Prion Protein Ure2

Chapter Abstract

The self-assembly of polypeptides into amyloid structures is associated with a range of increasingly prevalent neurodegenerative diseases as well as with a select set of functional processes in biology. The phenomenon of self-assembly results in species with dramatically different sizes, from small oligomers to large fibrils; however, the kinetic relationship between these species is challenging to characterize. In the case of prion aggregates, these structures can self-replicate and act as infectious agents. Here I analyse quantitative data on the oligomer populations formed during aggregation of the yeast prion protein Ure2 obtained by single molecule spectroscopy. Global analysis of the aggregation kinetics reveals the molecular mechanism underlying oligomer formation and depletion. Quantitative characterization indicates that the majority of Ure2 oligomers are relatively short-lived, and their rate of dissociation is much higher than their rate of conversion into growing fibrils. An initial metastable oligomer is identified, which can subsequently convert into a structurally distinct oligomer, which in turn converts into growing fibrils. I also show that fragmentation is responsible for the autocatalytic self-replication of Ure2 fibrils, but that preformed fibrils do not promote oligomer formation, indicating that secondary nucleation of the type observed for peptides and proteins associated with neurodegenerative disease does not occur at a significant rate for Ure2. These results establish a framework for elucidating the temporal and causal relationship between oligomers and larger fibrillar species in amyloid forming systems, and provide insights into why functional amyloid systems

are not toxic to their host organisms.

The previous two chapters seek to model early events in amyloid aggregation using an explicit primary nucleation reaction step, the same coarse-grained approach to new filament formation as has been employed in almost all bulk studies of amyloid kinetics to date. The current chapter instead explores primary nucleation in more detail by explicitly investigating kinetic intermediates of this reaction step. The work described in this chapter was published in *J Am Chem Soc* with the title "Direct Observation of Oligomerization by Single Molecule Fluorescence Reveals a Multistep Aggregation Mechanism for the Yeast Prion Protein Ure2". I took joint first authorship and was responsible for the great majority of the numerical modelling, theoretical work and analysis, as well as a large part of the drafting. Jie Yang and Si Wu carried out all experiments and the remainder of the drafting. The extensive Supporting Information section is reproduced in Appendix C. It contains further details of experimental techniques, further details of the model including partial analytical solutions, and additional experimental evidence for some of our conclusions (including a fibril size distribution analysis carried out with the assistance of Thomas C. T. Michaels).

5.1 Introduction

The self-assembly of soluble proteins into insoluble and highly structured amyloid fibrils rich in β-sheet structure is associated with a variety of human disorders, including Alzheimer's and Parkinson's diseases, type II diabetes, and the prion diseases. [8, 94] In addition, the formation of amyloid fibrils has been found to be a common or generic property of polypeptide molecules, and also to be associated with a number of diverse biological functions in living organisms. [94, 95] Over the past decade, oligomeric intermediates that form during the early stages of amyloid fibril formation or dissociate from mature fibrils have become of increasing interest [12] because such species, rather than the fibrils, have been shown to be toxic to cells and are now thought to be the major pathogenic agent in neurodegenerative disease. [8, 22–24, 94] Understanding the nature and dynamics of the oligomers is not only of intrinsic interest, but has the potential to provide a catalyst for the development of therapeutic strategies for protein misfolding diseases. [8, 96, 97]

Several conventional biochemical and biophysical methods are able to provide ensemble information about the aggregation kinetics of amyloidogenic proteins as well as the conformation and morphology of amyloid fibrils and the intermediate prefibrillar species populated during their formation. The metastable and heterogeneous nature of protein oligomers, however, has historically limited the detailed characterization of such prefibrillar species. Recently, a variety of novel approaches have been developed and applied for this purpose. Of these, single molecule fluorescence spectroscopy offers a particularly powerful approach for exploring the formation and properties of oligomers, as it has the ability to investigate individual molecular species and to reveal conformational dynamics that may be averaged in ensemble experiments. [98,99] These techniques have been shown to be able to identify and characterize the low-populated, heterogeneous and transient species formed during fibril assembly of several amyloidogenic proteins. [26, 28, 29, 100, 101]

Ure2 from *Saccharomyces cerevisiae* is the protein determinant of the yeast prion state [*URE3*] [102] and provides an important system for probing amyloid formation and prion propagation. Ure2 is a negative regulatory factor of nitrogen metabolism, as in its native state the protein interacts with the transcription factor Gln3 and represses the uptake of poor nitrogen sources. [103] Therefore, when Ure2 converts into the aggregated prion state, Gln3 is released and activates the expression of the genes related to the metabolism of less favorable nitrogen sources.

While the aggregated prion form of Ure2 is tolerated by yeast cells, precursor aggregates of Ure2 are toxic to mammalian cells, [104, 105] as are amyloid aggregates of other proteins. [106] It is therefore interesting to study the oligomers formed and the structural changes occurring during Ure2 fibril formation, and to compare them to disease-related models. In a previous study prefibrillar intermediates of Ure2 with a range of different sizes formed during the aggregation process were identified by AFM. [107] A soluble oligomeric species formed during the early stages of Ure2 aggregation was separated and characterized by biochemical and spectroscopic methods. Taken together, these results suggest a connection between the population of oligomeric species and the course of Ure2 amyloid assembly into mature fibrils.

Theoretical modelling has previously enabled the kinetic parameters that describe the growth and breakage of Ure2 fibrils to be defined, allowing the contribution of individual molecular steps to be correlated with prion propensity; [108] however, data on the oligomeric populations of Ure2, which would allow this type of mechanistic analysis to be carried out, have not previously been available. In the present study, single molecule fluorescence resonance energy transfer (smFRET) was applied to investigate in detail the intermolecular assembly and aggregation process of Ure2. This approach enabled oligomerization during the initial lag phase to be observed, and two types of Ure2 oligomers with different assembly modes were identified. Using theoretical analysis combined with single molecule and ensemble kinetic data, I could describe the formation and depletion pathway of oligomers, and propose a multistep mechanism for Ure2 fibril formation, in which initial oligomerization is followed by conformational conversion to β-sheet-containing oligomers that are then able to grow to form mature amyloid fibrils.

5.2 Methods

5.2.1 Developing a model for kinetic data fitting

We set out here to develop a quantitative model that describes the experimental observables: the total fibril mass concentration $M(t)$ and the total oligomer concentration $O(t)$. In addition to $M(t)$ and $O(t)$, the model explicitly considers the concentration of native state dimeric Ure2 $m(t)$, and of fibrils $P(t)$. In particular, the model describes explicitly the formation of oligomers through dimer association with rate constant k_{oligo}, their conversion to fibrils with rate constant k_c, their destruction (rate constant k_d), and fibril growth and fragmentation (rate constants k_+ and k_-, respectively). Addition of further complexity to this coarse-grained model is in principle readily possible within the master equation formalism; such additional details, including differentiating between multiple structural classes within the oligomer sub-populations, would, however, require further experimental constraints than are currently available, in order to avoid overfitting. [109] We used the Ure2 dimer concentration for $m(t)$ rather than the monomer concentration, as evidence suggests Ure2 remains in its dimeric form throughout the aggregation reaction (see Results).

The rate equations for the model can be written as a master equation:

$$\frac{dO}{dt} = k_{\text{oligo}} m(t)^2 - k_c O(t) - k_d O(t) \tag{5.1}$$

$$\frac{dP}{dt} = k_c O(t) + k_- M(t) \tag{5.2}$$

$$\frac{dM}{dt} = 2k_+ m(t) P(t) \tag{5.3}$$

$$m_{\text{tot}} = M(t) + m(t) + x_{\text{oligo}} O(t), \tag{5.4}$$

where x_{oligo} is the number of Ure2 dimers in an average oligomer, and where we have left out processes with negligible contributions to the overall kinetics, such as filament annealing [48,49]. m_{tot} is the total concentration of Ure2 in dimeric, oligomeric or fibrillar state at any time, and is equal to $m(0)$ in an unseeded reaction. Any larger oligomers are expected to form from growth of smaller oligomers; all oligomers ultimately grow from the initial interaction of a pair of dimeric Ure2 molecules. The physically reasonable choice of overall reaction order for oligomer formation is therefore 2.0. (For further explanation, see Appendix C.1.) Equations (5.1)-(5.3) were solved for early times in the aggregation process.

Since observed oligomer concentrations are very low compared to that of the other species, the contribution from $x_{\text{oligo}} O(t)$ to Eq. (5.4) is negligible (and determining x_{oligo} is not needed to successfully model the kinetics). Dropping this term, we could combine Eq. (5.4) with Eq. (5.3) and with the early-time solutions to derive a first-order self-consistent solution for $M(t)$ valid for all times (see also Appendix C.1):

$$M^{(1)}(t) = m(0)\left(1 - \exp\left(-A\left(e^{\kappa t} - 1\right)\right)\right) \tag{5.5}$$

with $A = \alpha k_+ k_c/\kappa^2(kl + \kappa)$, $k_l = k_c + k_d$, $\kappa = \sqrt{2k_+k_-m(0)}$ and $\alpha = k_{\text{oligo}}m(0)^2$.

The time-dependent evolution of the oligomer population depends only on k_{oligo}, k_l and $m(t)$. In turn, $m(t)$ depends only on κ and A, or κ, k_l, and $\alpha k_+ k_c$. Overall therefore, the dynamics of the dimer and oligomer populations depend on the following four combinations of rate parameters: k_+k_-, k_+k_c, k_l, and k_{oligo}. Moreover the fitted values of k_{oligo} and k_l are approximately independent of the values chosen for k_+k_- and k_+k_c, provided that these two parameter combinations give a reasonable fit to the fibril mass concentration.

5.2.2 Relating model rate constants to fundamental reaction steps

Experiments indicate that we can resolve the observed oligomers into two structurally distinct populations, with a low-FRET oligomer formed initially and subsequently converting into a high-FRET oligomer, which in turn converts to fibrils. The data are not, however, sufficiently detailed to allow a full kinetic analysis to be carried out on both populations individually. The avoidance of overfitting necessitates, therefore, that we consider together the different structural classes of oligomers, and examine the overall fluxes that lead to their generation or depletion. We can, however, incorporate elements of our knowledge of the oligomer sub-populations in the overall interpretation of the results. We did so by determining how each species contributes to the total oligomer formation, depletion and conversion rate constants in our coarse-grained model. The concentration of the later high-FRET oligomer species changes very little over the time course of the aggregation reaction compared to that of the earlier low-FRET species, and is present at significantly lower concentrations than the earlier low-FRET species over the times most relevant to the fitting procedure. Moreover, we show below that high-FRET oligomers are likely to be formed from conversion of low-FRET oligomers. Therefore, the rate constants for total oligomer formation and dissociation obtained from the fitting process can be interpreted as approximately the rate constants for low-FRET oligomer formation and dissociation. The rate constant k_c gives the approximate proportionality between the overall oligomer concentration and the rate of formation of fibrils from oligomers, and so contains information on both the conversion of low-FRET to high-FRET oligomers, as well as the conversion of high-FRET oligomers to fibrils. The concentration data on high-FRET oligomers indicated that the steady-state approximation is likely to be valid here, in which case we can explicitly write k_c in terms of the rate constants of a more detailed kinetic model featuring two separate oligomeric species (see Appendix C.1). The "conversion" rate constant would then be proportional to the rate constant for transformation of low-FRET oligomers to high-FRET oligomers, as well as to the rate constant for conversion of high-FRET oligomers to fibrillar species.

5.2.3 Fitting the combined smFRET/ThT data to the model

The ThT component of the data was fitted globally to the analytical expression for the full time-course fibril concentration to obtain values for A and κ using the online fitting platform Amylofit. [16] Then, the smFRET component of the data was fitted to our early time expression for $O(t)$ using Mathematica to give approximate values for k_{oligo} and k_l. Having established that $k_d \gg k_c$, we can set $k_d = k_l$. Combining these conclusions with A and κ yields approximate values for k_+k_- and k_+k_c. The availability of these approximate rate constants as trial parameters enabled a numerical fit of our combined smFRET/ThT data to Eqs (5.1)-(5.4) to be carried out, yielding robust rate constants and verifying the consistency of our model with the experimental data. For the fitting of the combined smFRET/ThT data, a ratio of 1.5:1 was chosen for k_+k_-(S68C):k_+k_-(V9C) (see Results). For a full description of the kinetic model and fitting methods, see Appendix C.1.

5.3 Results

5.3.1 SmFRET measurements reveal the absence of significant oligomer formation via secondary nucleation during Ure2 fibril formation

The two generic mechanisms that lead to the formation of fibrils are the primary nucleation pathway, during which new oligomers are generated by the direct association of soluble protein or peptide molecules, and secondary pathways, where existing fibrils have the propensity to generate the formation of new fibrils, either through fragmentation or through surface catalyzed secondary nucleation. In the latter case, nucleation of new fibrils takes place on, and is catalyzed by, the surface of existing fibrils, [30] the rate of which therefore depends on the mass concentration of existing fibrils. In the case of Aβ42 aggregation, [13] whose kinetics are dominated by the surface-catalyzed secondary nucleation pathway, it has been shown explicitly that the majority of small oligomers present during the reaction are produced during secondary nucleation. In earlier studies of Ure2, it has been shown that fragmentation events are important determinants of the rate of fibril formation under both quiescent and shaking conditions [108, 110, 111] although the production of oligomers through surface-catalyzed secondary nucleation in the case of Ure2 has not previously been discounted.

The generation of Ure2 oligomers was measured directly using smFRET in the presence of 1% preformed mature fibrils [112] to provide a surface for oligomer formation if secondary nucleation were to occur at a significant rate for Ure2. The time course of oligomerization observed by smFRET in the presence of the added fibrils shows a similar initial rate to that of the unseeded system, and the quantity of oligomers detected at each time point is not increased, indicating that the rate of oligomer production during secondary nucleation is insignificant compared to the rate of direct association of dimers to form oligomers during primary nucleation. This was confirmed by explicit fitting to kinetic models featuring oligomer formation during primary nucleation (Figure 5.1A) and during secondary nucleation (Figure 5.1B). The former yielded a good fit; the latter a poor fit, a result that is very different from the findings for Aβ42 [13] (Figure 5.1C). The effect of seeding can be fully captured by supplying appropriate non-zero initial conditions to the kinetic models. The initial mass concentration is known since the mass concentration of the seeds is simply 1% of the initial monomer concentration, or 0.15µM. The initial fibril number concentration was estimated from the initial mass concentration divided by the average fibril length as calculated from TEM measurements. The quantity of Ure2 oligomers under seeded conditions appears to be lower than in the absence of preformed fibrils, which is likely to be a result of the rapid depletion of native Ure2 by association with, and elongation of, the pre-existing fibril ends. Together with the kinetic analysis of the ensemble fibril formation of Ure2 (Appendix C.1 and Figure C.1), this observation confirms that the proliferation of Ure2 fibrils results from fragmentation and not from secondary nucleation.

5.3.2 Analysis of oligomer populations reveals the existence of an oligomer conformational conversion step

The results so far have established that Ure2 oligomers are formed predominantly from the free association of dimers during primary nucleation. We can demonstrate using the following simple argument that only a minority of these oligomers ultimately become fibrils, and that both an oligomer dissociation pathway and a conformational conversion step are needed. Taking a conservative estimate of the initial oligomer formation rate of 20 nM/h, and noting that minimal native Ure2 depletion occurs over the first 4 h of aggregation, a concentration of oligomers of at least 80 nM will be formed in the first 4 h. The rate of oligomer formation declines subsequently, but does not cease until all Ure2 is depleted from solution at ~$10-12$ h; we therefore estimate the lower bound on the total concentration of oligomers that form during primary nucleation to be 100 nM. Mature Ure2 fibrils are observed to have lengths typically greater than 100 nm, and mostly on the micrometer scale. Given that the interchain distance within an amyloid fibril is ca. 0.5 nm, fibrils must typically contain at least 100 Ure2 dimers. In the smFRET

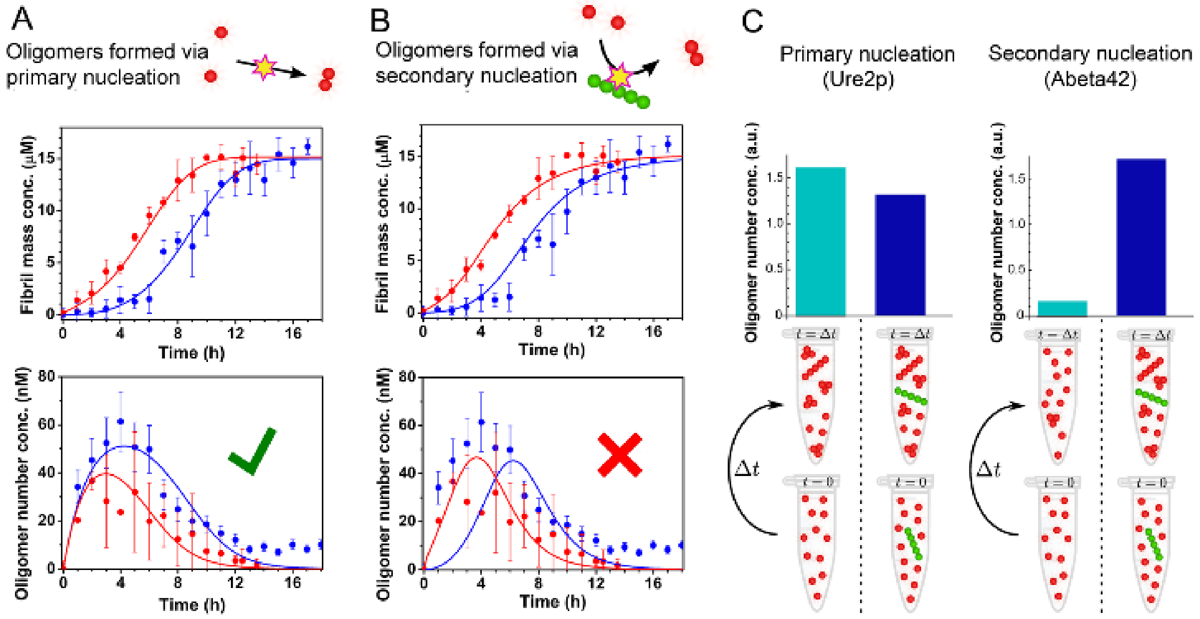

Figure 5.1: **Absence of a fibril-catalyzed secondary nucleation process for Ure2.** (**A**,**B**) Ensemble aggregation kinetics of 15 µM unlabeled Ure2-S68C monitored by ThT fluorescence under unseeded (blue) or seeded (red) conditions (*upper panels*). Ure2 oligomers were then detected under unseeded (blue) or seeded (red) conditions by confocal single molecule FRET (*lower panels*). (**A**) The data fit well to a model that generates oligomers during primary nucleation. (**B**) A model that generates oligomers during secondary nucleation cannot fit the data. (**C**) The presence of seeds (*right hand columns*) drastically increases the concentration of Aβ42 oligomers measured at a single time point in the lag phase of an Aβ42 aggregation experiment [13] (*right panel*), but do not increase the production of Ure2 oligomers (*left panel*), indicating fundamentally different mechanisms of oligomer formation for these two systems. Data provided by J. Yang and S. Wu. [112]

experiments the concentration of Ure2 dimers incorporated into fibrils is approximately 15 µM, so the final concentration of fibrils is at most 150 nM. Ure2 has also been shown to follow fragmentation-dominant kinetics, so fibril formation through primary nucleation is insignificant compared to the total formation of fibrils through fragmentation. The total concentration of fibrils formed through primary nucleation must therefore be at least an order of magnitude less than the total concentration of fibrils formed, and so is significantly less than 15 nM. This value is far lower than the concentration of oligomers formed during primary nucleation observed in our smFRET experiments, and leads us to conclude that most oligomers must dissociate rather than elongate. Given that oligomers undergo faster dissociation than the fibrils, the oligomers must be structurally distinct from fibrils. In the following two sections we demonstrate that new fibrils are likely to originate from structural conversion of these oligomers, although this must occur much more slowly than dissociation. This result was confirmed by comparing fitted values for the oligomer conversion and dissociation rate constants (see kinetic analysis below).

5.3.3 Two types of on-pathway oligomeric species with different structures can be observed

Two oligomer populations could be resolved in experiments, distinguished by their FRET efficiency, and with distinct structural properties [112]. High-FRET oligomers appear after low-FRET oligomers but before fibrils, suggesting that they are formed by conversion of the low-FRET oligomers.

Next, the structures of the Ure2 species dissociated from fibrils were probed by smFRET. Two major populations centered at around 0.64 and around 0.8 were observed, the lower of which was similar to the high-FRET distribution observed during the aggregation reaction, and can be attributed to disaggregated oligomers. The species showing higher FRET values, with a maximum at around 0.8, can be attributed to small fibrils, indicating the more compact structure within amyloid fibrils. This observation is remarkably similar to that in previous studies of α-synuclein. [100] The results of TIRF experiments, therefore, demonstrate that the oligomers that disaggregate from Ure2 fibrils have the same structural properties as the oligomeric species formed in the later stages of the aggregation reaction. This finding strongly suggests that the high-FRET oligomers contain β-sheet structure similar to that found in mature amyloid fibrils, and that this type of oligomeric species is able to convert to elongation-competent fibril-type species.

5.3.4 Kinetic analysis of combined smFRET and ThT data yields a quantitative understanding of oligomer formation, dissociation and conversion

Analysis of the kinetics of the aggregation reaction is a crucial step in understanding the microscopic mechanism of amyloid formation. Previous theoretical work has provided an analytical solution to the kinetics of fibril formation involving fragmentation, [21] and by globally fitting the ThT curves over a range of concentrations [77] to this expression, two combined kinetic parameters, $k_n k_+$ and $k_+ k_-$ can be obtained, where k_n, k_+ and k_- represent the amyloid nucleation, elongation and fragmentation rates, respectively. To study how the mutations affect these rates, and thus to gain further structural and mechanistic insight into the nucleation process, we performed smFRET experiments to compare the oligomer formation of the two Ure2 mutants, V9C and S68C.

Previous kinetic modeling of amyloid aggregation used a single coarse-grained reaction step to represent the "primary nucleation" pathway by which new fibrils are generated via an initial association step. Here, the availability of accurate kinetic data on the total concentration of oligomeric intermediates allows us to devise a less coarse-grained kinetic model that explicitly includes intermediates in the nucleation step. The model remains partly coarse-grained, however, as it makes no distinction between different oligomer types; nevertheless, it provides additional insights into the nature of the nucleation process. In this model, oligomers are formed through an initial assembly process, occurring with rate constant k_{oligo}, and subsequently convert into growth-competent fibril-type species with a rate constant k_c. These species can then elongate by dimer addition with rate constant k_+, and fragment with rate constant k_-. The oligomers can also dissociate with rate constant k_d (that we have shown above is much larger than k_c). Note that we can approximately interpret oligomerization and dissociation as fundamental reaction steps; however, the conversion step is in fact a coarse-grained step that contains information on the transformation of low-FRET to high-FRET oligomers, as well as on the subsequent conversion of the latter species to fibrils (see Methods and Appendix C.1 for full details). The accurate determination of reaction orders with respect to dimers requires kinetic data for a range of initial dimer concentrations. Given just one initial dimer concentration, however, we can make the reasonable assumption of a reaction order of 0 for conversion and 2 for oligomer formation. Any inaccuracy in these reaction orders is effectively incorporated into our definitions of k_c and k_{oligo}, and does not significantly affect the quality of the fitting (see Appendix C.1). An accurate analytical solution for the time dependence of the fibril mass concentration in our model can be derived by extension of previous approaches. [21, 27] The solution is identical to the analytical solution for the kinetics of a fragmenting system, upon which the kinetic analysis of the bulk ThT experiments in Appendix C.1 is based, except that the fibril nucleation rate $k_n m(0)^{n_c}$ is resolved in terms

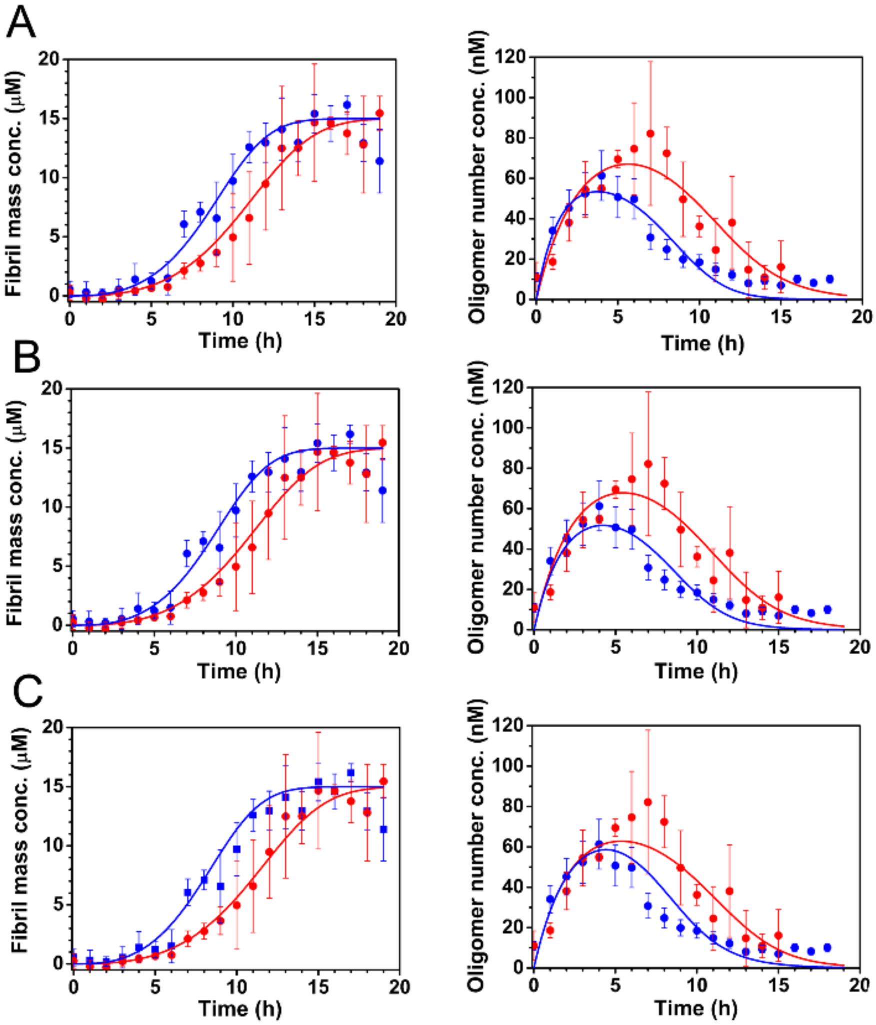

Figure 5.2: **Fitting of combined smFRET/ThT data to models indicates a probable effect of mutations on the dissociation of oligomers, but not on their formation.** (**A-C**) The bulk aggregation kinetics of 15 µM unlabeled Ure2-V9C (red) and Ure2-S68C (blue) monitored by ThT fluorescence (*left panels*) and the concentration of AF555/AF647 labeled Ure2-V9C (red) and Ure2-S68C (blue) oligomers throughout the aggregation reaction monitored by confocal smFRET (*right panels*) were globally fitted to a theoretical model (see Methods) including the formation, dissociation, and conversion of oligomers, and the elongation and fragmentation of fibrils. (**A**) Allowing both k_{oligo} and k_d to differ for each mutant gives good fits, with a mean squared error of 1.54. (**B**) If k_{oligo} is constrained to be the same for both mutants, the model fits the data equally well, with a mean squared error of 1.58. (**C**) If neither k_{oligo} nor k_d is allowed to differ, the fit is less good, especially around the time when the oligomer concentration is at a maximum, with a mean squared error of 1.87. This result therefore implies that k_{oligo} is the same for the two variants, while the values of k_d may differ slightly.

of the microscopic processes introduced in our oligomer model. Specifically, we obtain

$$k_n m(0)^{n_c} = \alpha \frac{k_c}{k_c + k_d + \kappa} \tag{5.6}$$

where $\kappa = \sqrt{2k_+ k_- m(0)}$ and $\alpha = k_{\text{oligo}} m(0)^{n_c}$. This result, combined with an analysis of the equation governing oligomer kinetics, reveals that the kinetics of this system are controlled by the parameter combinations $k_{\text{oligo}}, k_d, k_+ k_c$, and $k_+ k_-$. The combined rate parameters $k_+ k_c$ and $k_+ k_-$ can be determined with order-of-magnitude accuracy.

A numerical procedure was used to fit the combined smFRET and bulk ThT data to this model (see Methods for further details). The numerical fits are reasonable given the accuracy of the data, and show that our coarse-grained model provides a good description of the system (Figure 5.2). The minor divergence between the fitted curve and the data for S68C oligomers at the latest times is consistent with the fact that disaggregation from fibrils yields a small population of oligomers at equilibrium, yet to avoid overfitting there is no explicit fibril disaggregation step in the model. Fitting gives a value for k_{oligo} of 1.6×10^{-3} μM^{-1} h^{-1} for both variants (Figure 5.2A,B), and a value of k_d of 0.60 h^{-1} for S68C, and 0.45 h^{-1} for V9C. This result indicates that the kinetics of oligomer formation by S68C and V9C are broadly similar, but that the S68C oligomers dissociate more readily by a factor of approximately 1.3. Given that the rates of oligomer formation are very similar, it is likely that the reaction order of oligomerization is approximately the same in the two mutants at this concentration range.

The differences in $k_+ k_c$ and $k_+ k_-$ between the V9C and S68C variants are less than an order of magnitude, indicating that these parameters are similar for both variants. An order-of-magnitude estimate for k_+ of 40 µM^{-1} h^{-1} was obtained from analysis of the seeded aggregation experiment monitored by ThT fluorescence (see Appendix C.1 and Figure C.2 for details). This value allows us to calculate order-of-magnitude estimates for k_- and k_c, of 1×10^{-4} h^{-1} and 2×10^{-3} h^{-1} respectively. We note that k_c is 2 orders of magnitude smaller than k_d, as expected from our analysis of oligomer dissociation. To estimate the differences in these parameters between each variant, we used a ratio k_+(S68C):k_+(V9C) of 1.5:1, as indicated by seeded bulk experiments (see Appendix C.1 and Figure C.2), and a ratio k_-(S68C):k_-(V9C) of 1:1 as indicated by fragmentation rate measurements (see Appendix C.1 and Figure C.3). This series of steps then allows us to calculate a ratio k_c(S68C):k_c(V9C) of 1.4:1. Although this value is rather sensitive both to experimental error and to errors in the parameter ratios that are used, it can be interpreted as demonstrating that the conversion rates are similar, although possibly somewhat larger in S68C. Taken together, these results indicate that the V9C mutation decreases the fibril elongation rate and oligomer dissociation rate, leaving the fibril fragmentation rate and oligomer formation rate unaffected. Furthermore, they suggest that the oligomer conversion rate may also be slightly decreased by the V9C mutation.

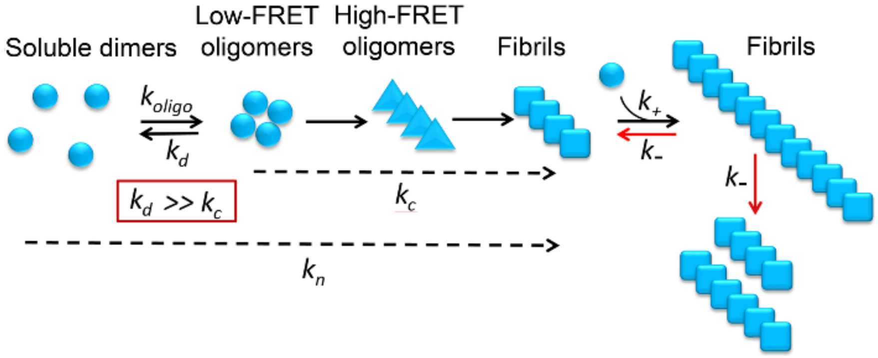

Figure 5.3: **Proposed model for the aggregation pathway of Ure2.** Native dimeric Ure2 forms loosely structured oligomers driven by hydrophobic interactions and either dissociates back to the native state or undergoes conformational conversion to form more compact oligomers containing β-sheet structure, which can in turn convert into growth-competent fibrillar species. Fragmentation of fibrils then contributes to their proliferation.

5.4 Discussion

Single molecule FRET measurements were carried out to investigate the aggregation behaviour of the yeast prion protein Ure2 and to observe the low populations of transient oligomers formed during the aggregation reaction that are challenging to detect by ensemble methods. Analytical modelling of the single molecule FRET observations indicate that the majority of the oligomers formed during the initial step of the nucleation process dissociate back to the native dimeric state, but a small population of oligomers undergoes a conformational conversion step leading to formation of elongation-competent species. Quantitative analysis of a combination of bulk and single molecule data has provided detailed information about the rates of the microscopic kinetic steps in the formation of amyloid fibrils, and how these rates are altered by point mutations in the prion domain. Based on our observations, a mechanism has been proposed for the formation of Ure2 amyloid fibrils in which native Ure2 forms relatively disordered oligomers, probably driven by hydrophobic intermolecular interactions, only a small proportion of which then rearrange to form structurally more compact β-sheet containing oligomers that are able to convert further to elongation-competent fibrillar species and grow by addition of native dimers to form mature amyloid fibrils (Figure 5.3).

Primary nucleation of amyloidogenic proteins, in which a native protein converts into an elongation-competent species, is a crucial step in fibril formation. Molecular simulations of the aggregation of Aβ42 indicate that primary nucleation occurs via intermediate disordered non-β oligomers, which not only facilitates encounters between the monomeric proteins but also provides an environment that facilitates their conversion to fibrillar β-structure. [113, 114] Nonspecific intermolecular interactions, such as hydrophobic interactions, play a crucial role in the formation of the initial disordered oligomers, [113] while the intra- and intermolecular hydrogen bonding within β-sheets is considered to be the driving force for the subsequent conformational conversion. [115] This formation of hydrogen bonds compensates for the disruption of the hydrophobic interactions in the initial disordered oligomers, thus favoring conformational reorganization to β-sheet structure in order to reach the lowest energy state. The conformational reorganization between early relatively disordered oligomers and β-rich elongation-competent species during amyloid formation has also been suggested by experimental studies of other amyloidogenic proteins such as α-synuclein, [100] Aβ40, [116] and the yeast prion protein Sup35. [117] In agreement with theoretical and experimental results for other amyloid proteins, both theoretical analysis of oligomer concentrations and direct observation of two types of Ure2 oligomers possessing different assembly and emergence times provide additional evidence for the oligomerization/conversion model as a generic feature of amyloid nucleation.

In order to obtain a quantitative understanding of the fundamental reaction steps that contribute to the amyloid formation process, analytical methods have been used to describe fibril growth and to obtain the microscopic kinetic parameters. [21, 30] Since the advent of single molecule techniques, it has become feasible to probe the nucleation process directly in real time throughout the aggregation reaction. [28,29,100,101] However, the kinetic analysis in previous studies has either focused on the early stage of oligomer formation [28,100] or made use of a highly coarse-grained nucleated polymerization model in which secondary processes were not considered and in which the fitted rate constants were difficult to relate to specific reaction steps. [29] In this study, we provide a kinetic analysis of the combined single molecule FRET and ensemble ThT data (Figure 5.2 and Figure 5.1) that considers not only the oligomer formation process but also the elongation and fragmentation processes, providing quantitative information about the complete aggregation pathway of an amyloid protein. Another advance in the present study is the determination of an expression for the bulk primary nucleation rate k_n, which is found to depend not only on the rate constants of oligomer formation, dissociation and conversion, but also on the rate constants for fibril elongation and fragmentation. Global analysis of the oligomer formation kinetics, measured by smFRET, and the fibril formation kinetics, monitored by the ensemble ThT assay, results in two independent parameters k_oligo and k_d and two combined parameters k_ck_+ and k_+k_-, where the latter two can be decomposed by direct measurement of k_+ or of k_-. Thus, the theoretical analysis of the smFRET data provides detailed quantitative information about the fundamental steps in the process of amyloid nucleation.

Our study also provides new structural insights into the aggregation mechanism of Ure2. We observed that the V9C mutation causes an increase in the length of the lag phase for fibril formation compared with that of wild-type Ure2. Kinetic analysis of the bulk ThT variable-concentration data under both unseeded and cross-seeded conditions reveals that this effect is predominantly due to a decrease in the elongation rate caused by the V9C mutation. We then investigated the effect of the mutation on the nucleation process in detail by kinetic analysis of the combined smFRET and bulk ThT data. The rates of both oligomer dissociation and oligomer conversion are slightly decreased in the V9C mutant relative to WT Ure2, while the rate of oligomer formation is unaffected, indicating that V9C forms a more stable oligomer. However, the V9C mutation has a negligible effect on the overall rate of formation of elongation-competent oligomers (k_n). Taken together, these results show that the V9 residue plays an important role in the elongation of Ure2 fibrils.

The toxicity of amyloid oligomers has been demonstrated to be correlated with the degree of exposure of hydrophobic surface. [118,119] Since oligomers assembled by functional amyloidogenic proteins, such as Ure2 in this study, and Sup35 in a previous study, [117] have similar conformational features (for example, reactivity with A11 and OC antibodies) and a similar formation and conversion pathway to that of toxic amyloidogenic proteins such as α-synuclein [100] and Aβ, [113,116] the difference between functional and disease-related amyloid remains to be elucidated. In this study, we observed a gradual decrease in the concentration of oligomers accompanying the onset of ThT fluorescence. Calculations and theoretical analysis of the smFRET data has led to the conclusion that the majority of Ure2 oligomers formed do not ultimately become fibrils but are depleted by dissociation. In contrast, the populations of Aβ40 and α-synuclein oligomers have been observed to follow a single exponential process and to remain at a significant equilibrium concentration even after all the monomers are depleted. [28, 100, 101] A possible explanation for the accumulation of Aβ40 and α-synuclein oligomers is that they convert very slowly relative to their formation, and become kinetically trapped once monomers are depleted, as the residual population of oligomers can no longer grow through monomer addition into fibrils. Another possible explanation for the difference is that the fibrils of Aβ40 and α-synuclein may not be as stable as the Ure2 fibrils, such that the equilibrium between oligomers and fibrils favors oligomers to a much greater extent than is the case with Ure2. In either of the above cases, Aβ40 and α-synuclein oligomers should have higher stability relative to the native state than Ure2 oligomers, and therefore would not be expected to dissociate significantly. The accumulation of relatively high concentrations of oligomers formed by disease-related amyloidogenic proteins could therefore be the cause of their toxicity.

A further potential reason for the higher toxicity of disease-related amyloidogenic proteins is that oligomers are generated not only during primary nucleation but also during secondary nucleation. For the case of Ure2, we have been able to demonstrate directly the lack of formation of surface-catalyzed secondary nuclei using smFRET measurements (Figure 5.1). The absence of surface-catalyzed secondary nucleation will greatly reduce the generation of oligomers, [13] which is another possible reason for the lower toxicity of functional amyloids. Taken together, the low stability of Ure2 oligomers and the absence of secondary nucleation suggests that the functional yeast prions may replicate and propagate by fragmentation rather than secondary nucleation and hence avoid significant populations of potentially toxic oligomers, as occurs in the aggregation of neurodegenerative disease-related proteins. This study paves the way for the detailed study and comparison of further examples of both disease-related and functional amyloid systems, particularly the variety of amyloidogenic proteins that play structural and functional roles in bacteria. [120, 121] The understanding gained from such studies will not only shed light on the mechanisms by which amyloid structures are harnessed for functional roles, but may also provide clues as to possible new therapeutic strategies to combat amyloid disease.

5.5 Conclusions

In summary we used kinetic theory to uncover basic details of the mechanism of Ure2 fibril formation via oligomeric intermediates. We found that oligomers were formed through primary nucleation only, and that unlike with previously-studied α-synuclein, most oligomers dissociated rather than convert into fibrillar species. As a result of the latter feature, and of the filament fragmentation present in this system, a substantially new kinetic model was developed. I carried this out at the same time as another, similar, project on tau oligomerization (currently under review at *ACS Chem Neurosci*), that I have not included in this thesis due to constraints on the word count. These two projects highlighted to me the need for a more general analytical theory of oligomer formation kinetics, that I outline in the following chapter.

Chapter 6

Classifying Amyloid Oligomers Using Chemical Kinetics

Chapter Abstract

The misfolding and aberrant aggregation of proteins into fibrillar structures is a key factor in some of the most prevalent human diseases, including diabetes and dementia. Oligomers present during the early stages of amyloid fibril formation are thought to be the major pathogenic agent in these diseases. Understanding the nature of these oligomers and the reactions they undergo is thus not only of intrinsic interest, but is likely to aid in the successful development of treatments for protein misfolding diseases. Here, we present a general approach, based on chemical kinetics, for understanding the key processes of oligomer formation and depletion, and their role in fibril formation. We reveal the kinetic equations that govern filament self-assembly proceeding via oligomeric intermediates; and develop analytical solutions to them. We apply this methodology to experimental data on 3 unrelated amyloid systems to discover the commonalities and differences between oligomeric intermediates of the primary and secondary fibril nucleation processes. In all cases we find that oligomers are mechanistically distinct from small fibrils, and overwhelmingly dissociate back to monomeric form rather than convert to fibrillar species.

To date, oligomer kinetics have been dealt with in an ad-hoc way; the previous chapter provides an example of this. A natural and important next step is to create a complete, unified and mathematically rigorous theory of amyloid oligomer kinetics. The manuscript presented in this chapter sets out to do precisely this. I am primarily responsible for all work described herein with the exception of the simulations, carried out by Anđela Šarić, and the early-time solutions to secondary oligomerization and the Aβ42 data fitting,

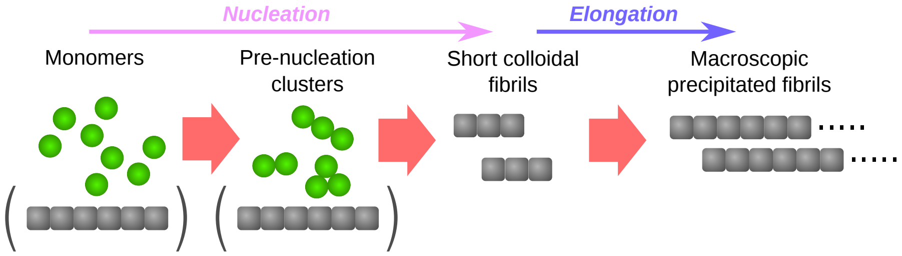

Figure 6.1: Schematic of the role of oligomers in amyloid filament formation. Pre-nucleation clusters and short colloidal fibrils are both oligomeric and both detectable by experiment. Pre-existing fibrils may in principle catalyze both pre-nucleation cluster formation and their conversion to fibrillar species.

both carried out by T. C. T. Michaels. Appendix D contains derivations, supplemental theoretical results and details of the Monte-Carlo simulations.

6.1 Introduction

Oligomers are of paramount importance in the field of protein aggregation. They are believed to be the primary species responsible for the pathology of human disorders associated with amyloid filament formation, such as Alzheimer's disease and Parkinson's disease [22–24, 122]. There has often been confusion surrounding the term and what aggregate species it includes. However, a precise definition for oligomers has in fact long existed [123], originating in supramolecular chemistry: molecules whose structure comprises a small plurality of units derived from molecules of lower relative molecular mass, and whose properties vary significantly with the removal of one or a few of the units. The protein oligomer species meeting this definition are highly heterogeneous, varying significantly both in structure and in size; and in propensity to grow into filaments. The protein filament formation process has previously been studied using bulk Thioflavin T (ThT) assays among other techniques, providing kinetic data on the mass concentration of fibrils formed. The fundamental kinetic equations describing the formation and growth of amyloid fibrils in bulk are now well-established in the literature [13, 18, 19, 21, 30, 47, 48, 53, 58], and have permitted modelling of kinetic data to discover the underlying reaction mechanisms. These models do not explicitly consider oligomer populations, with formation of new filaments from monomers occurring directly via a slow "primary nucleation" step (rate constant k_n), that treats any oligomeric intermediates implicitly; and filaments subsequently growing through monomer addition, via a rapid "elongation" mechanism (rate constant k_+). They also account for filament-dependent "secondary processes", such as fragmentation of existing filaments (rate constant k_-) and "secondary nucleation" of new filaments on the surface of existing filaments (rate constant k_2). The treatment of a multiple microscopic reaction steps involving intermediate species as a single simplified reaction process is known as coarse-graining, and is a crucial component of kinetic modelling [109].

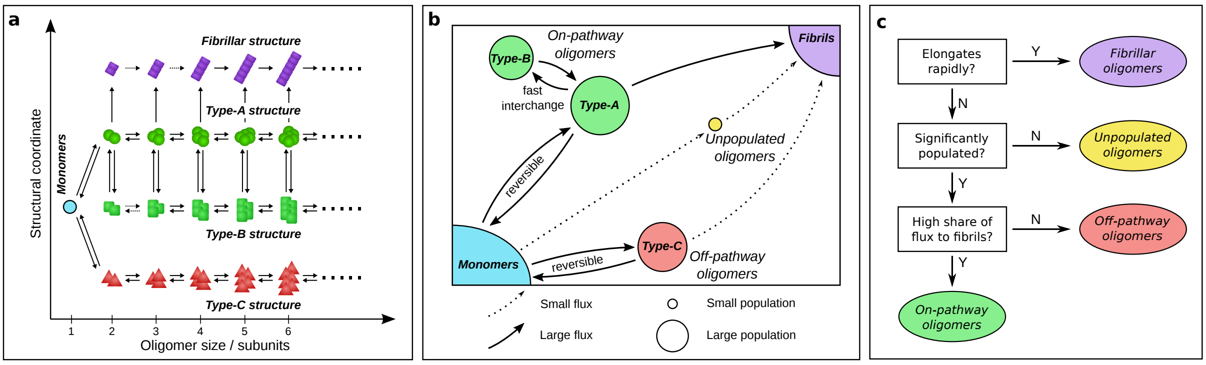

Figure 6.2: Classifying oligomers according to their role in filament formation. (a) The full amyloid oligomer reaction network, organized according to conformation and number of constituent monomers. Experiments can typically only distinguish conformation (or structure type) and not size. (b) The coarse-grained reaction network with no size distinctions. Oligomer types A and B are in rapid exchange on the experimental timescale and are thus part of the same ensemble, and both on-pathway. (c) Decision tree for categorizing different types of oligomers, once both experimental and timescale coarse graining procedures have been carried out.

Recent advances in single molecule experimental techniques have allowed researchers to record the time-dependence of the concentration of oligomeric species present during amyloid fibril formation [100, 124]. These advances, coupled with the recent recognition of the fundamental importance of oligomers in amyloid diseases, have given rise to a need for new kinetic theory, capable of describing the formation and loss of oligomers during protein filament formation. To date, however, the only analytical theory developed has either neglected non-fibrillar species [59, 125, 126]; or has neglected oligomer dissociation, and been limited to early-time, constant-monomer conditions in the absence of secondary processes [27]. The latter has had notable success in the study of alpha-synuclein oligomers (a protein associated with Parkinson's disease) [28, 100, 124, 127]; but has proven insufficient to model more recently-studied systems [112, 128], including the data on Aβ42 presented here.

In this paper, we therefore develop a complete framework for understanding the dynamics of amyloid oligomer populations. We illustrate throughout with coarse-grained Monte-Carlo simulations of oligomer-driven fibril formation, and with experimental data on the oligomerisation and aggregation of the yeast prion protein Ure2, and the amyloid proteins Aβ42 and tau associated with Alzheimer's disease.

6.2 Most amyloid oligomers are nonfibrillar

The single-molecule experimental techniques currently used to track oligomer concentrations cannot directly report on oligomer size. They instead yield quantities that are only approximately correlated with size, such as time taken for a species to cross the confocal volume. The delineation between oligomers and fibrils has so far been made using arbitrary values of these quantities, that are high enough to ensure that few if any nonfibrillar oligomers are mislabelled as fibrils [124]. We therefore expect measured oligomer populations to contain both nonfibrillar prenucleation clusters, and very small colloidally-suspended ThT-active linear species with high β-sheet content, capable of rapid elongation (Fig. 6.1). These cannot properly be viewed as intermediates of the nucleation process, as their propensity to undergo rapid elongation makes them kinetically equivalent to longer fibrils; instead, these "fibrillar oligomers" [129] should be considered as products of the coarse-grained nucleation steps employed in earlier kinetic models of filament formation. In all amyloid systems hitherto studied, fibrillar oligomers can be shown to be present at negligible concentrations compared to non-fibrillar oligomers. The maximum possible concentration of a fibrillar oligomer of length j produced via primary nucleation is that attained under constant monomer conditions in the late-time limit. This is given by the late-time limit of Eq. (26) in ref. [59]: $k_n m(0)^{n_c-1}/2k_+$. The maximum possible concentration of a fibrillar oligomer of length j produced via secondary nucleation, on the other hand, is given by Eq. (30) of ref. [126] as $k_2 m(0)^{n_2}/2k_+$. Note these are both independent of the length. Choosing an oligomer-fibril cut-off of x monomers, we can therefore calculate a generous upper bound on the total concentration of fibrillar oligomers present at any time, S_F:

$$S_F < \frac{x k_{\text{nuc}} m(0)^{\beta}}{2k_+}, \tag{6.1}$$

where k_{nuc} is the rate of the dominant coarse-grained fibril nucleation process (primary or secondary), and β is related to the reaction order of the dominant nucleation mechanism as $n_c - 1$ (primary), or n_2 (secondary), and can also be determined by dimensionality arguments [130]. For Aβ42, Ure2 and tau aggregation, we find that S_F is far lower than experimentally-observed concentrations of oligomers; thus, observed oligomers are predominantly non-fibrillar, and not able to undergo the rapid elongation characteristic of fibrils. This negates the need for explicitly considering the contribution of fibrillar oligomers when modelling experimental data. We illustrate this conclusion with the example of Aβ42. Here, for an $m(0)$ of 5 μM, $k_+ \approx 3 \times 10^6$ M^{-1}s^{-1}, $k_2 \approx 10^4$ M^{-2}s^{-1}, $\beta = 2$ [13]. Then, choosing a generous $x = 100$, S_F is at most 4 pM, compared to an observed oligomer concentration S of around 75 nM. Full results for other systems are detailed in Appendix D.

6.3 Characterizing the oligomer reaction network

Since non-fibrillar oligomers greatly outnumber short filaments, it is expected that the majority of the former do not ultimately convert into filamentous species. This does not imply, however, that they are off-pathway species. To understand this, it is useful to consider the underlying reaction network connecting monomers to fibrillar species. The network contains a node for monomers (reactants), for fibrils (products), and for each distinct oligomeric species. Species are considered distinct if the free energy barrier separating them is $> O(kT)$, the energy of thermal fluctuations (k is the Boltzmann constant; T is the temperature in Kelvin). Amyloid oligomers can be uniquely identified by their conformation, or structure, and by the number of monomers they contain (their size). The oligomer reaction network can then be conveniently organized into a 2-dimensional grid; with one axis corresponding to oligomer size, and the other serving to distinguish different conformations (see Fig. 6.2(a)).

Each reaction process directly connecting distinct species is represented by an arrow. We assume that all reaction processes are either unimolecular or involve addition or loss of monomers. By treating monomer addition or dissociation implicitly we may then represent all reaction processes by arrows with only 1 origin and 1 destination.

This picture represents the maximum desirable level of information we could have about the reacting system. However, we almost never have sufficient data to fully constrain a kinetic model of this complexity. Model selection theory [109] dictates that our chosen kinetic model should contain no more detail than is necessary to describe the data using mass-action kinetics; we must therefore coarse-grain our network according to experimental considerations. Firstly, although available experimental techniques can often obtain separate reaction profiles for different types of oligomer conformations, they cannot usually collect kinetic data separately on oligomers of the same generic conformation but different numbers of constituent monomers. For instance, in [112] we identified 2 separate oligomer structures distinguished by their Förster Resonance Energy Transfer (FRET) efficiency, and in [128] we identified 2 generic structure types by their differing stabilities in different buffer solutions. In both cases, however, we could not further distinguish oligomers with the same structure type but different size. We must therefore typically coarse-grain reaction networks such that oligomer populations are summed over size (see Fig. 6.2(a)-(b)) and sometimes also structure, and consider only total fluxes between these combined populations. Secondly, species can in practice be considered distinct only if the timescale of their interconversion is not much shorter than either the timescale over which experiments can be accurately carried out, or the timescale of the overall aggregation reaction (see Fig. 6.2(b)). For instance, in one of the tau oligomer species exchanges so rapidly with monomer above 22°C (data not shown) that they must properly be considered part of the reactant ensemble, and not an intermediate at all at the available experimental resolution.

6.4 Identification of on- and off-pathway oligomers

Having identified the level of coarse-graining appropriate for the available data, we may now identify “reaction pathways” as the chains of processes connecting the reactant node

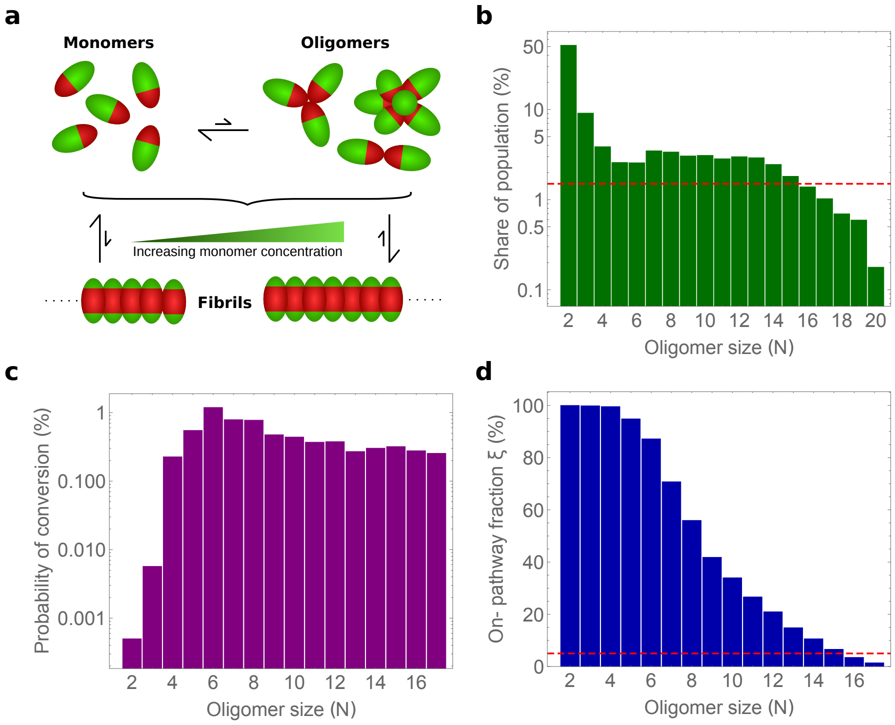

Figure 6.3: Simulations of the kinetics of filament formation via oligomers of a single bonding type, but a range of sizes. (a) Schematic of the coarse-grained Monte-Carlo simulations. (b) Size distribution of oligomers at steady state; the lowest-concentration species (below the red line) contribute to less than 5% of the total oligomer concentration and are labelled unpopulated. (c) Probability of conversion to fibrils vs oligomer size N. This steeply increases with N initially, peaks at $N = 6$ and then plateaus. (d) Contribution of oligomers of different size to the overall flux to fibrils. Larger oligomers grow via smaller oligomers; therefore, the smaller the oligomer, the more reaction pathways it sits on and the larger its ξ value. The species below the red line contribute less than 5% of the flux to fibrils, and would be considered off-pathway if significantly populated. However, the population falls off with increasing size faster than the propensity to convert; coarse-graining by size is therefore not expected to lead to mixed on/off-pathway oligomer populations.

in our coarse-grained network to the product node without loops. Each pathway X carries a portion f_X of the total flux f of reactants to products. Defining the on-pathway coefficient ξ_i of an oligomeric species S_i as $\xi_i = \sum_{i\in X} f_X/f$, a species is on-pathway if it lies on a set of reaction pathways with a high total flux, or has a high value of ξ_i. By contrast, it is off-pathway if it is significantly populated during the reaction but has a low ξ_i. See Fig. 6.2(b) for an illustration of these distinctions. If it has a low ξ_i and a low share of total oligomer concentration, we label it an unpopulated oligomer. See Fig. 6.2(c) for a summary of this decision-making process. An on-pathway species of oligomer may then be greatly in excess of fibrils if it becomes kinetically trapped, or if it forms reversibly from monomers.

To illustrate the properties of a network in which we do not need to coarse-grain by size, we simulate a simple aggregation reaction containing oligomers of a range of sizes but just 2 possible conformations: non-fibrillar and fibrillar (Fig. 6.3(a)). Our simulations sug-

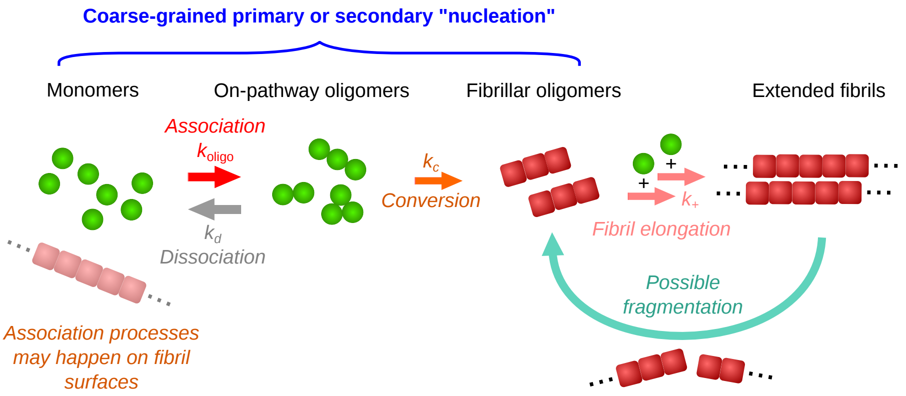

Figure 6.4: Schematic of the coarse-grained reaction network employed in the remainder of this paper. Its underlying kinetic equations are solved analytically and found to successfully describe all kinetic data collected on amyloid oligomers to date. Note fibrillar oligomers are typically present at far lower concentrations than non-fibrillar species, and are thus ignored.

gest that significantly-populated oligomers of different size but the same structure, when distinguishable, are likely to be either all on-pathway or all off-pathway. The propensity to convert to fibrils (Fig. 6.3(c)) increases with size initially, as it only takes one subunit to convert to a fibrillar state for the whole oligomer to rapidly convert. It then appears to peak at an intermediate oligomer size, possibly due to increased crowding inhibiting the required conformational rearrangement; however, oligomer concentrations fall off faster than the conversion propensity (Fig. 6.3(b)). It is therefore unlikely that small on-pathway oligomers can coexist with a significant population of larger, off-pathway oligomers of the same structural type (Fig. 6.3(d)). Note that Fig. 6.3(d) is computed as the reverse-cumulative flux directly to fibrils, since all larger oligomers are formed by growth of smaller oligomers. A given oligomer in these simulations therefore sits on all the reaction pathways in which larger oligomers convert directly to fibrils.

It is fortunately often possible to experimentally distinguish off-pathway oligomer populations from those that are on-pathway. For example, in [128] the more stable grouping of oligomer species formed significantly more slowly than fibrils and are thus likely off-pathway. In [112] and [28], separate low- and high-FRET oligomer species were discovered; and the latter was found to form from the former. Additionally, the high-FRET species was found to form by disaggregation from mature fibrils. Microscopic reversibility thus implied them both to be on-pathway. Inhibitors or activators may also be employed to probe reaction pathways: if the rate of formation of a measurable coarse-grained oligomeric species is decreased or increased but the rate of fibril formation is not, then this species must be off-pathway.

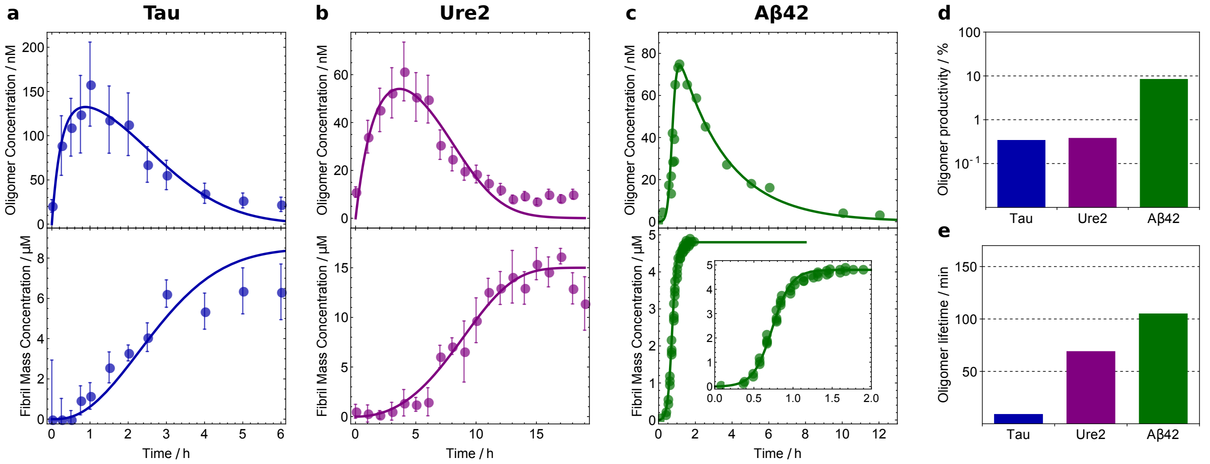

Figure 6.5: Fitting of available experimental kinetic data on a range of amyloid systems using the analytical models derived in this study. Experimental data on oligomer formation (top) and fibril aggregation (bottom) are fitted simultaneously (solid lines), with fitting parameters summarized in table S1. (a) Tau data, with solution to primary nucleation model (Eqs. (6.2)-(6.4)). (b) Ure2 data, with breakable filament model containing additional fragmentation step. (c) Aβ42 data, with model containing additional secondary nucleation step for oligomer formation. (d) Fraction of oligomers that convert to fibrillar species. (e) A comparison of oligomer lifetimes. Note Ure2 and tau form predominantly primary oligomers, whereas Aβ42 forms predominantly secondary oligomers.

6.5 Analytical solutions to the kinetics of filament assembly via an oligomeric intermediate

The coarse-grained reaction networks for Aβ42, tau and Ure2 all have the generic form shown in Fig. 6.4. Having determined this form, we may now obtain general rate equations for filament assembly via an oligomeric intermediate that explicitly include optional secondary processes and oligomer dissociation. Such equations can be obtained in closed form by considering only the species $m(t)$, $S(t)$, $P(t)$, and $M(t)$, representing the concentrations of monomers, oligomers, fibrils and fibril mass respectively. For a system in which oligomers are generated through primary nucleation ("primary oligomers"), we can then write:

$$\frac{dS}{dt} = k_{\text{oligo}} m(t)^{n_o} - (k_c + k_d)S(t) \tag{6.2}$$

$$\frac{dP}{dt} = k_c S(t) + k_2 m(t)^{n_2} M(t) \tag{6.3}$$

$$\frac{dM}{dt} = 2k_+ m(t)P(t) \qquad m(t) + M(t) = m_{\text{tot}}, \tag{6.4}$$

where we have neglected filament annealing and depolymerisation of fibrils, and the contributions to monomer depletion from non-elongation processes, which are typically insignificant. Fibrils can elongate with rate constant k_+. Oligomers are formed through monomer association (occurring with rate constant k_{oligo} and reaction order n_o); and subsequently convert to growth-competent fibril-type species with a rate constant k_c, or dissociate back to monomers with rate constant k_d. Further fibril formation via secondary processes occurs with rate constant k_2 (=0 in the case of no secondary processes), and reaction order n_2 (=0 in the case of filament fragmentation). These equations may be straightforwardly generalized to allow an explicit monomer dependence for the conversion step (see Appendix D), and also to include multiple on- and off-pathway oligomer species; for simplicity, we neglect these possibilities here.

In the early-time limit these equations become linear and admit exact solutions. By applying similar fixed point techniques to those employed in refs [21,30,50], we can extend the range of validity of these solutions to the full time course. These solutions are given in closed form in Appendix D. After a short initial adjustment period, the solutions for $M(t)$ both with and without secondary processes reduce to the same functional forms as their counterparts for filament formation without oligomeric intermediates (see refs [21,30,36]), validating the use of a coarse-grained nucleation step to model filament assembly kinetics. The timescale of this adjustment period is $k_e = k_c + k_d$; it arises from the vanishing of a term in the solutions, proportional to $e^{-k_e t}$, that is the only component not also present in the solutions for filament formation via direct nucleation. For a full discussion of this phenomenon, see Appendix D.1.1. Note this timescale and this functional simplification arises due to the onset of steady-state conditions for oligomers; these results can thus equivalently be obtained by application of the steady-state condition $dS/dt = 0$.

The kinetics of filament assembly *without* oligomeric intermediates feature two timescales, λ and κ, that describe aggregate proliferation through primary and secondary processes respectively. The latter timescale is unchanged when primary oligomeric intermediates are introduced, and is still given by $\kappa = \sqrt{2k_+k_2m(0)^{n_2+1}}$, since we have not altered the elongation or secondary nucleation steps. The other timescale, however, changes from $\lambda = \sqrt{2k_+k_nm(0)^{n_c}}$ to $\lambda_p = \sqrt{2k_+k_{\text{oligo}}m(0)^{n_o}k_c/(k_e+\kappa)}$. Thus, the key effect of introducing an intermediate into the primary nucleation step is to recast the primary nucleation rate as $k_nm(0)^{n_c} = k_{\text{oligo}}m(0)^{n_o}k_c/(k_e+\kappa)$, or k_c multiplied by the steady-state oligomer concentration.

In addition to these timescales, the solutions for $S(t)$ are also explicitly controlled by k_e and by the initial rate of oligomer formation, $\alpha = k_{\text{oligo}}m(0)^{n_o}$. Overall, then, our systems are controlled by the 4 timescales α, λ, κ, and k_e. From these come 5 fitting parameters: k_{oligo}, n_o, k_e, k_+k_c, and κ.

6.6 Contrasting primary and secondary oligomers

We have so far considered a system featuring primary oligomeric intermediates; closely related equations can be written down to account for formation of filaments or oligomers through secondary processes (“secondary oligomers”):

$$\frac{dS}{dt} = k_{\mathrm{o1}}m(t)^{n_{\mathrm{o1}}} + k_{\mathrm{o2}}m(t)^{n_{\mathrm{o2}}}M(t) - (k_c + k_d)S(t) \tag{6.5}$$

$$\frac{dP}{dt} = k_c S(t) \tag{6.6}$$

$$\frac{dM}{dt} = 2k_+ m(t)P(t) \qquad m(t) + M(t) = m_{\mathrm{tot}}, \tag{6.7}$$

Similarly to before, these equations may be solved using self-consistent techniques; the solutions are presented in Appendix D. As with primary oligomers, introducing oligomeric intermediates to the secondary nucleation step of filament formation does not change the functional form of the solution for $M(t)$ first given in Ref [30]. Unlike before, both timescales are different to the no-oligomer case, and are given by $\lambda_s = (2k_+k_ck_{\mathrm{o1}}m(0)^{n_{\mathrm{o1}}})^{1/3}$, and $\kappa_s = (2k_+k_ck_{\mathrm{o2}}m(0)^{n_{\mathrm{o2}}+1})^{1/3}$. Equating our timescales we resolve k_2 in terms of both secondary oligomerization and conversion as $(k_ck_{\mathrm{o2}})^{2/3}$; and k_n intriguingly in terms of conversion and both primary and secondary oligomerization as $2k_{\mathrm{o1}}k_{\mathrm{conv}}^{2/3}k_{\mathrm{o2}}^{-1/3}/3$.

Secondary oligomers may be tested for using single-molecule aggregation experiments with and without the presence of pre-formed fibril seeds. If the presence of fibril seeds does not increase the speed at which oligomers are formed, or the concentration of oligomers formed, then they are produced predominantly by direct association of soluble monomers, as a primary nucleation intermediate if on-pathway. By this method, we confirm that Ure2 oligomers are intermediates only of primary nucleation (“primary oligomers”), and that Aβ42 oligomers are predominantly intermediates of secondary nucleation.

Having solved the dynamical equations governing oligomer formation, we now proceed to analyzing experimental kinetic data on the oligomer formation process for 3 systems of interest. The fitting procedure requires simultaneous consideration of oligomer data and aggregate mass concentration data. The data for each protein were fitted to the relevant expressions, yielding good fits and demonstrating that these models indeed explain the observed behaviour (Fig. 6.5).

6.7 Oligomers are depleted predominantly by dissociation

In an aggregating system, some on-pathway oligomers must be consumed by conversion to fibrillar species; however, oligomer dissociation may also be important. From fitting unseeded kinetic data alone, it is not possible to determine the rate constants for these processes: our fitting parameters are the combined rate constants $k_e = k_c + k_d$, and k_+k_c. For Ure2, tau and Aβ42, however, almost all oligomers are ultimately depleted;

having established earlier that the vast majority do not become fibrils, non-conversion mechanisms must dominate. It is important to again emphasize here that a dominant non-conversion oligomer depletion mechanism does not imply that the oligomers are off-pathway.

For a more precise calculation, analytical expressions can be developed for the average filament size; these may be used in conjunction with microscopy data to approximately disentangle these rate constants. Alternatively k_+ may be extracted directly from seeded kinetic experiments. A convenient measure is then the oligomer productivity, defined as k_c/k_e; a low value indicates that most oligomers dissociate rather than convert. The results for Ure2, tau and Aβ42 are displayed in Fig. 6.5(d), and confirm that the majority of reactive flux from monomers to fibrils goes through intermediates that simply have much quicker rates for dissociation back to monomers than for conversion to fibrils over the reaction timecourse. It is important to note, however, that if the conversion reaction step has a monomer dependence, the conversion rate may initially be larger than the dissociation rate constant without violating the rule that most oligomers ultimately do not become fibrils. Examining Fig. 6.6, it seems that this is an almost universal feature of amyloid oligomers, the only known exception being αS under certain conditions.

We have hitherto assumed that non-conversion depletion (represented by k_d) is dominated by the most obvious candidate mechanism: unimolecular dissociation back to monomers. Other possible contributing mechanisms, however, include sticking to fibrils. To eliminate these possibilities, one can measure the rate of oligomer dissociation directly by taking an aliquot from an aggregation reaction, and diluting heavily to effectively suppress all non-unimolecular processes. k_d(dissociation) is then determined by fitting the observed oligomer concentration to an exponential decay function. If this is similar to the k_d value determined from fitting data to our kinetic models above, then oligomer depletion must be dominated by dissociation. For Ure2, this experiment reveals that this is the case, and oligomers predominantly deplete through dissociation [112]. These conclusions are supported by our simulations, in which oligomers have a much stronger propensity to dissociate than to rearrange to form fibrils.

6.8 Kinetic stability varies widely between systems

Determining $k_e = k_c + k_d$ also allows us to calculate the average half-lives, $t_h = \ln(2)/k_e$, of the measured total oligomer populations. For types A and B tau oligomers, these are approximately 9 minutes and 208 minutes respectively; for Ure2, 69 minutes; and for Aβ42, 105 minutes. As the key toxic species in amyloid aggregation diseases, the comparatively long lifetimes of Aβ42 oligomers and type B tau oligomers carries direct clinical significance: it is possible that their comparative kinetic stability hinders their clearance by cellular mechanisms and that their resulting greater persistence increases

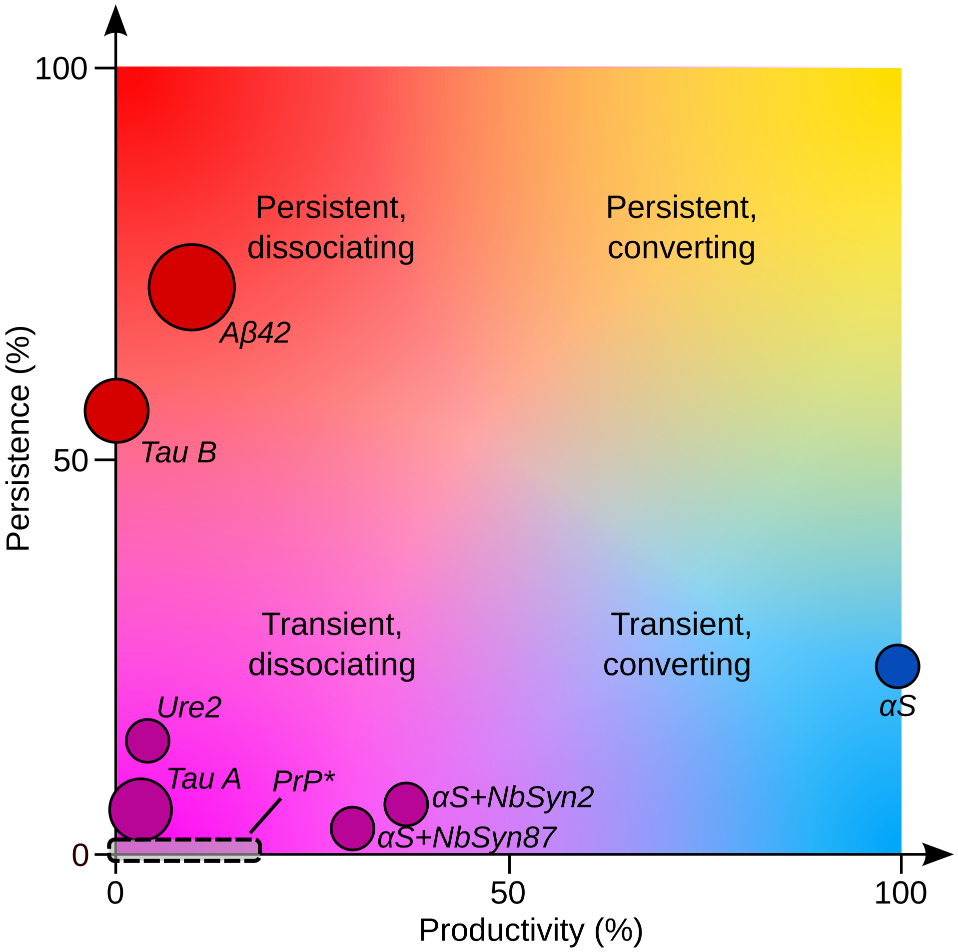

Figure 6.6: Categorizing nonfibrillar oligomers by their key properties. Bubble width: maximum abundance relative to monomers (approximate logarithmic scale). Persistence: $t_h/(t_h + \tau_h)$, where $t_h = \ln(2)/k_e$ is oligomer half-life and τ_h is observed monomer depletion half-time; this is a measure of the relative kinetic stability of oligomers. Productivity: k_c/k_e, or contribution of conversion to overall oligomer depletion. Of systems hitherto studied, only αS oligomers convert more rapidly than they dissociate; and only Aβ42 oligomers and type-B tau oligomers persist longer than their monomers. Tau data taken from [128]; "αS" represents the low-FRET αS WT oligomer population from [28]; "αS+NbSyn2/NbSyn87" represents the low-FRET αS WT oligomer population from [131], aggregated in the presence of NbSyn2/NbSyn87 camelid nanobodies; "Ure2" represents the overall Ure2 S68C mutant oligomer population from [112], believed to have identical kinetics to WT-Ure2; Aβ42 data provided by Sara Linse. *: Prion protein PrP oligomer persistence and productivity not known precisely, but can be constrained to fall within the highlighted region. Relative abundance unknown. Data provided by Jason Sang.

their toxic effect. Interestingly, while in Ure2 aggregation oligomer depletion is much faster than fibril proliferation (half time 510 minutes), in the case of disease-related Aβ42 and tau oligomer depletion is slower than fibril proliferation (half times 45 and 162 minutes). A convenient measure of relative oligomer stability is given by $t_h/(t_h + \tau_h)$, where t_h is the oligomer depletion half-life as defined above, and τ_h is the observed monomer depletion half-time. This measure is termed persistence, and reported as a percentage. Examining Fig. 6.6, it appears that most amyloid oligomers hitherto investigated are less kinetically stable than their monomers; therefore, oligomers and monomers have time to approach equilibrium, and observed oligomer decay is instead controlled by monomer depletion. By contrast, observed decay of secondary Aβ42 oligomers and type-B tau oligomers occur over longer timescales than monomer depletion, and hence must be controlled by oligomer dissociation.

It is interesting to note an apparent gap in the distribution of oligomer properties illustrated in Fig. 6.6, with no oligomers being both persistent relative to monomers and predominantly converting in nature. We cannot formally include fibrillar oligomers in this matrix, but can nonetheless assign them a location based on their analogous behaviour. Their incidence is very low, leading to a very small bubble; and they do not undergo significant dissociation, putting them on the far right. Moreover, in the absence of monomers their persistence is high, being unable to elongate into fibrils, and they may then be placed in the top-right corner of the matrix, filling the gap identified. It is finally also interesting to note that the disease-associated tau and Aβoligomers are formed significantly more readily than oligomers of the functional amyloid Ure2, suggesting a possible link between oligomer abundance and toxicity.

6.9 Conclusions

In summary, we have presented a general chemical kinetic approach to understanding the nature, formation and disappearance of peptide oligomers generated during amyloid filament formation.

We first described a rigorous conceptual framework for identifying amyloid oligomers and for classifying them according to their role in fibril formation. A key therapeutic goal in neurodegenerative diseases is to interrupt or halt the fibril formation process; charting the reaction network connecting native protein to fibrillar aggregates is a vital component of this endeavour. On-pathway oligomers, properly identified, present attractive targets for rational drug design; not only due to their inherent and well-documented toxicity, but also due to the critical dependence of the overall fibrillation process on their formation.

We next developed general coarse-grained kinetic models of filament formation via on-pathway oligomeric intermediates of both primary and secondary nucleation. We derived analytical solutions to the oligomer and fibril concentrations, demonstrating their validity by comparison to numerical solutions. We confirmed that these models are capable of explaining in-vitro aggregation behaviour by generating good fits to our experimental data. Examining these models revealed the emergence of key additional timescales upon introduction of oligomeric intermediates into the nucleation process, notably the adjustment timescale $k_e^{-1} = (k_c + k_d)^{-1}$, where k_c and k_d are the rate constants for oligomer conversion to fibrils and oligomer dissociation to monomers, respectively. At times greater than this timescale, these models reduce to the same functional form of the simpler no-intermediate models used previously to describe a wide range of filamentous growth processes. This may explain the success of these direct-nucleation models despite the fact that most filamentous growth processes to which they have previously been applied are known or suspected to proceed via oligomeric intermediates.

We next defined two useful metrics based on the rate constants in these models: oligomer persistence, and oligomer productivity. Persistence, a convenient measure of the relative kinetic stability of oligomers compared to monomers, was defined to be $t_h/(t_h + \tau_h)$, (where $t_h = \ln(2)/k_e$ is the oligomer half-life, and τ_h is the observed monomer depletion half-time); productivity was defined as k_c/k_e, and indicates the proportion of oligomers that convert rather than dissociate. Combined with oligomer thermodynamic stability compared to monomers, we were able to use these metrics to categorize all oligomers hitherto modelled according to their kinetic properties. This analysis revealed that, despite the fact that the mechanisms producing both oligomers and fibrils differ widely in different systems, in almost every case the majority of oligomers formed dissociate back to monomers rather than go on to form fibrils. This revelation as to the nature of the oligomers formed in such systems may have far-reaching implications for the understanding of amyloid diseases, and for rational design of drugs to target toxic oligomers. Our methodology will find ready application to other systems when new data become available.

Part IV

Oligomer Thermodynamics

Chapter 7

Thermodynamics of Aβ co-oligomer formation

Chapter Abstract

Multiple isoforms of aggregation-prone proteins are present under physiological conditions and have the propensity to assemble into co-oligomers with potentially very different properties from self-oligomers, but this process has not been quantitatively studied to date. We have investigated the amyloid-β (Aβ) peptide, associated with Alzheimer's disease, and the aggregation of its two major isoforms, Aβ40 and Aβ42, using a statistical mechanical modelling approach in combination with *in vitro* single-molecule fluorescence measurements. We find that at low concentrations of Aβ corresponding to its physiological abundance, there is little free energy penalty in forming co-oligomers, suggesting that the formation of both self-oligomers and co-oligomers is possible under these conditions and that mixing entropy can promote the formation of mixed species. Our model can be used to predict the oligomer concentration and size at physiological concentrations of Aβ and suggests mechanisms by which the ratio of Aβ42 to Aβ40 can affect cell toxicity. An increased ratio of Aβ42 to Aβ40 raises the fraction of oligomers containing Aβ42, which can potentially increase the hydrophobicity of the oligomeric species and thus promote deleterious binding to the cell membrane and increase neuronal damage. Our results suggest that co-oligomers are a common form of aggregate when Aβ isoforms are present in solution and may therefore potentially play a significant role in Alzheimer's disease.

Having considered only total oligomer populations in prior chapters, in this chapter Aβ oligomer heterogeneity in both size and composition are investigated. The work presented here was published in *Sci Rep* in 2016, under the title "Quantitative analysis of co-oligomer formation by amyloid-beta peptide isoforms". My contributions were to combine theory and experiment to infer that most oligomers must be dimeric, rendering geometry irrelevant; to reduce an unpublished linear co-oligomer model developed by G. A. Garcia into a simpler mixed dimer one; and to use this simpler model to analyse a dataset on Aβ co-oligomer formation at equilibrium. The parts of the Supporting Information section for which I was responsible are reproduced in Appendix E. It contains a derivation of the dimer model used, and supporting theoretical results.

7.1 Introduction

Neurodegenerative diseases, such as Alzheimer's disease (AD), are devastating and incurable conditions associated with the misfolding and aggregation of native monomeric proteins [4]. The deposition of aggregated amyloid-β peptide (Aβ) in the brain is a pathological hallmark of AD [132]. Aβ is formed from the cleavage of a transmembrane receptor, the amyloid precursor protein (APP), in various locations to generate peptides of varying lengths, most commonly 40 and 42 residues (Aβ40 and Aβ42) [133]. The Aβ42 isoform has an additional Ile-Ala dipeptide at its C terminus making it more hydrophobic and more aggregation-prone than Aβ40 [14, 134]. Hence, while the relative ratio of the Aβ40 to Aβ42 in cerebrospinal fluid (CSF) is approximately 9:1, the amount of Aβ42 is enriched relative to Aβ40 in deposits such as amyloid plaques [135, 136]. Moreover, some early-onset versions of AD have been related to the overproduction of Aβ42 relative to Aβ40 [137], and an increase in the ratio of Aβ42 to Aβ40 cleaved from APP has been correlated to increases in toxicity both *in vitro* and *in vivo* [138–143].

Although solid fibrillar deposits of Aβ accumulate in AD brains, the major cytotoxic effects causing the earliest pathological events are associated with smaller aggregates, Aβ oligomers [144]. Such species are formed via the association of monomeric Aβ and ultimately polymerize into amyloid fibrils when the total protein concentration exceeds the critical aggregation concentration (CAC) [27]. Due to their transient presence and low abundance, the oligomers have been difficult to characterise using conventional experimental techniques [145], particularly in the systems containing multiple isoforms of Aβ. There have been numerous studies of the mixtures of Aβ isoforms, demonstrating that Aβ40 and Aβ42 co-interact during the aggregation reaction [146–150]. Furthermore, there is evidence that Aβ40 and Aβ42 can form co-oligomers *in vitro* [139], [146], and on the surface of neurons [151]. A detailed study revealed that while Aβ40 and Aβ42 form separate fibrils in solution, the peptides co-interact in the early stages of Aβ aggregation, during primary nucleation [85].

In these previous studies, it has not been possible to determine the concentration and composition of the formed self- or co-oligomeric species of Aβ40 and Aβ42. Moreover, since

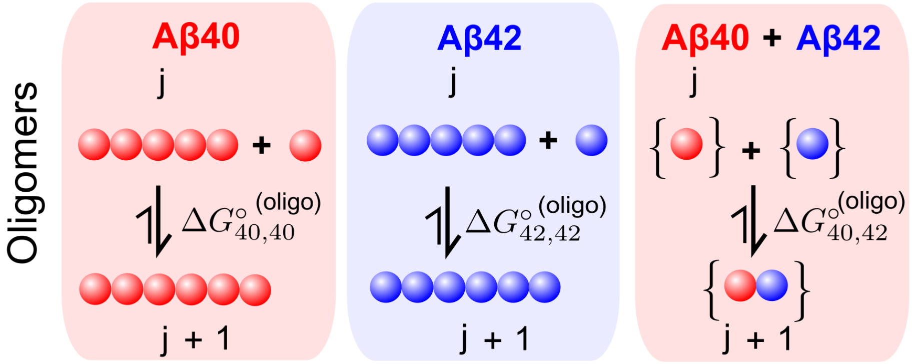

Figure 7.1: Schematic of the statistical mechanical model used to estimate Aβ oligomer numbers and relative composition. For the single-species datasets, the model considers oligomers of any length, whereas for the co-oligomerising datasets it considers monomers and dimers for mathematical simplicity. This simplification is justified by the single-species analysis predicting a very low number of oligomers larger than dimers.

most biophysical studies are typically performed at non-physiological high-micromolar concentrations of Aβ, it has not been possible to extrapolate the observations to very low total concentrations of Aβ peptide observed *in vivo* [152]. Because of the demonstrated strong and non-linear concentration dependence of Aβ aggregation [153–156], a meaningful extrapolation would require direct measurements of Aβ oligomer populations at sub-micromolar peptide concentrations. In order to address this, we combine here direct single-molecule measurements of oligomer populations at low Aβ concentrations with a statistical mechanical model to estimate the number and composition of the oligomers present under equilibrium conditions, and subsequently investigate how changing the ratios of the two Aβ isoforms affects the resulting oligomer populations.

7.2 Results and Discussion

7.2.1 Modelling approach

In this study, the relevant thermodynamic parameter characterising oligomerization is the free energy of monomer addition, ΔG°, independent of oligomer size, and this single parameter forms the basis for our model, as described in detail in Appendix E.

In the model we consider the major contribution to the energetics of the oligomeric aggregates to emerge from nearest neighbour interactions. We thus treat self-oligomers as simple non-interacting one-dimensional chain structures with nearest-neighbour interactions independent of the chain length (Fig. 7.1), and thus ΔG° is an elongation free energy. We note that the assumption of one-dimensional chain structures is not restrictive under our experimental conditions, where both self- and mixed oligomers can be inferred by our self-oligomer model to be predominantly dimeric (see Appendix E.3), and therefore larger non-linear structures where geometric effects can play a major role are not expected to perturb the analysis. This result permits us to formulate and employ a simple model for co-oligomers, containing monomers of both Aβ40 and Aβ42, that considers only dimers. The system behaviour is thus effectively governed by the Gibbs free energy released upon adding two monomers together to form a new intermolecular interaction. Note that while the assumption of the size-independent binding free energy is valid for the studied Aβ system, it is not applicable to non-filamentous growth assemblies.

In general for a linear aggregation process involving size-independent aggregation and disaggregation rates, the equilibrium constant K for filament elongation is given by:

$$K = \frac{f(j+1)c_0}{f(j)c}, \tag{7.1}$$

where $f(j)$ is the equilibrium concentration of an aggregate of length j, c is the equilibrium monomer concentration and c_0 the standard concentration (taken to be 1 M). We may therefore write:

$$\frac{c}{c_0} e^{-\Delta G^\circ/(RT)} = \frac{f(j+1)}{f(j)}. \tag{7.2}$$

We can then identify $e^{\Delta G^\circ/(RT)} = c/c_0$ with a CAC c [53]: this is the maximal equilibrium monomer concentration that can be attained by the system, since the filament size distribution must always be monotonic decreasing for a finite population of filaments. The nature of the species present at equilibrium depends strongly on the initial concentration of the monomeric peptides. When the initial monomer concentration is below the CAC, the majority of the peptides in the system are in their monomeric states and only a few aggregates are formed consisting of a small number of monomers. By contrast, above the CAC, most molecules are present as aggregates. These aggregates are either oligomers or fibrils. Previous single-molecule [100] and bulk data [157–159] indicate that these two species differ in their structures, and thus we allow for separate ΔG° for the oligomeric and fibrillar states, $\Delta G^{\circ(oligo)}$ and $\Delta G^{\circ(fib)}$. Moreover, oligomers are populated only for small aggregation numbers, while mature fibrils are observed for sizes that exceed 1000 monomers [101]. At low concentrations, therefore, below the CAC, the formation of large aggregates is suppressed, and the majority of aggregates are oligomeric. When the total concentration reaches the CAC, the majority of monomers are sequestered into fibrillar forms, and the concentration of oligomers does not increase even when the total peptide concentration is increased. Thus we expect the initial increase in aggregate concentration to be controlled by the free energy of oligomer formation $\Delta G^{\circ(oligo)}$. Once the total peptide concentration reaches the CAC, $c_0 e^{\Delta G^{\circ(fib)}/(kT)}$, the theory predicts a plateau in the concentration of oligomers, controlled by the free energy of fibril formation $\Delta G^{\circ(fib)}$. A recent study shows that the formation of self-fibrils of the Aβ isoforms is favoured *in vitro* [85], which implies that there is a significant difference between the $\Delta G^{\circ(fib)}$ when adding a monomer to a self-fibril or a fibril of different composition for Aβ40 and Aβ42.

However, the same study suggests that the difference is smaller for $\Delta G^{\circ(oligo)}$ when adding a monomer to a self- or mixed oligomer.

7.2.2 Single-molecule measurements

Having established the described theoretical approach, we then used single-molecule two-colour coincidence detection (TCCD) [160] in order to measure directly the concentration of Aβ oligomers present in solutions below and around the CAC, for Aβ40, Aβ42 and a 1:1 mixture of Aβ40 and Aβ42 [161].

Initially, we measured the fibril CAC through two independent methods: firstly by determining the concentration of soluble species in equilibrium with fibrils, which coincides with the fibril CAC at high concentration, above the fibril CAC [53] [161]. Secondly, we determined the concentration at which the oligomer concentration ceases to increase with total peptide concentration and reaches a plateau phase; the theory predicts that this transition should take place at the fibril CAC.

The total concentration of the released species in the former approach, which corresponds to the CAC, was measured to be 94 ± 37 nM for Aβ40, and 28 ± 4 nM for Aβ42 at pH 7.4. The value for Aβ40 is in good agreement with the previous result of 100 nM at pH 7.440, and the value for Aβ42 is lower than a previously reported value of 0.2 µM at pH 8 [70], consistent with a reported decrease in CAC with lowering the pH [162]. This gives values of the free energy for fibril formation as $\Delta G^{\circ(fib)}_{42,42} = -44.8 \pm 0.4$ kJ mol^{-1} for Aβ42 and $\Delta G^{\circ(fib)}_{40,40} = -41.7 \pm 1.1$ kJ mol^{-1} for Aβ40. The result for Aβ40 is within the range of previously reported values for the unlabelled peptide [163, 164], which were -37.7 kJ mol^{-1} and -46.7 kJ mol^{-1}, indicating that the presence of the fluorophore labels at the N-terminus does not substantially alter the free energy of fibril formation. The observation that Aβ42 fibrils disaggregate to a lesser extent than fibrils of Aβ40 suggests that the Aβ42 fibrils are more stable than their Aβ40 counterparts, correlating well with previous reports of Aβ disassembly and stability [165, 166].

Next, we combined equal quantities of monomeric peptide singly labelled with a blue-fluorophore with monomeric peptides singly labelled with a red-fluorophore, using low concentrations of total Aβ, 1-250 nM [161]. The solutions were left until equilibrium was attained. As the monomeric peptides self-associate to generate oligomers, we can distinguish them from monomers by the criteria of coincidence and quantify the oligomeric populations by TCCD [161]. The results are shown in Fig. 7.2, and the oligomer concentrations are in the range of 0–20 nM for Aβ40, 0–4 nM for Aβ42 and, strikingly, around 0–3 nM for mixed Aβ40-Aβ42 species. The error bars are relatively high in these experiments due to the low oligomer concentrations and inherent sample to sample variations. However, the results appear to follow the prediction from the theory and allow an estimate of the $\Delta G^{\circ(oligo)}$ values to be obtained in each case, as is described below.

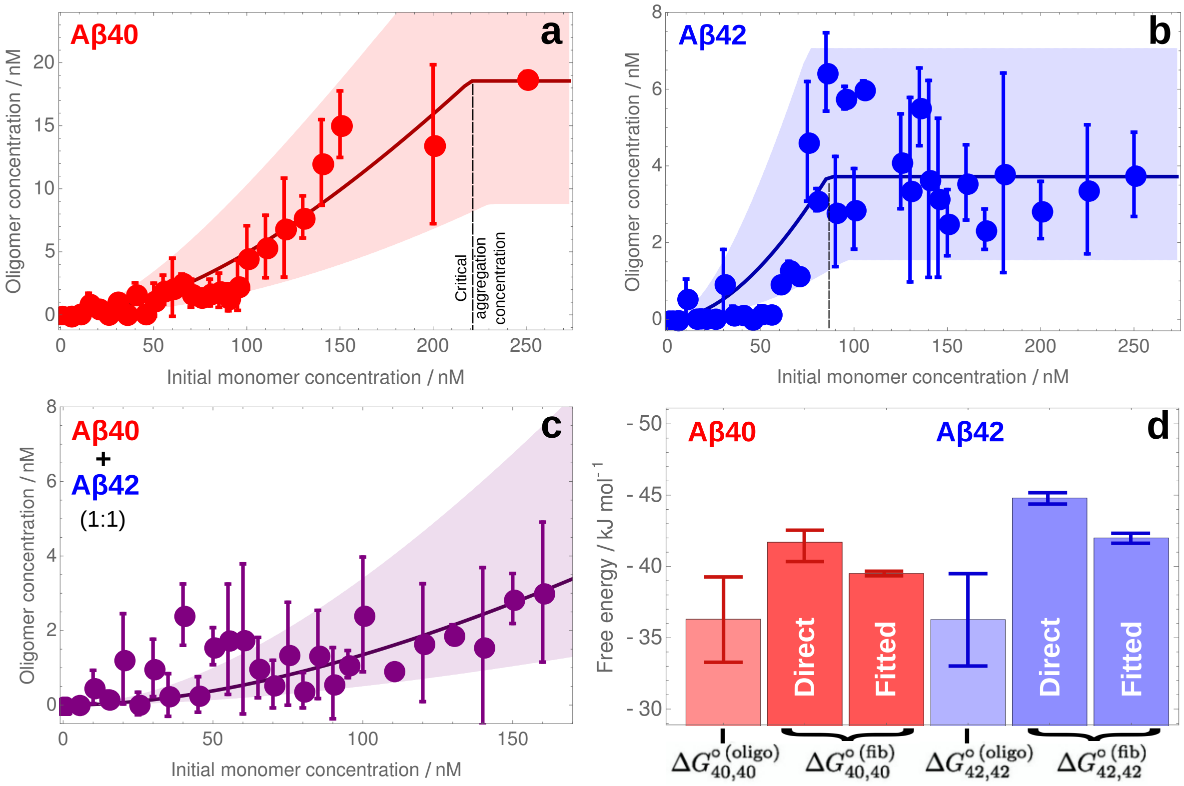

Figure 7.2: Equilibrium oligomer concentrations as a function of the total initial monomer concentration in the aggregation reaction. (Error bars SD, N(samples) = 3). The oligomer concentration was modelled and fitted separately for both Aβ40 (**a**), Aβ42 (**b**), and the 1:1 mixture (**c**); allowing extraction of the free energies of oligomerization and estimation of the CAC for Aβ40 and Aβ42 (fitted curves shown overlaid). The shaded bounds on these charts are curves plotted using the maximum and minimum free energies of oligomerization, and of fibril formation (given by the CAC) that still lie within the majority of the error bars. (**d**) The fitted free energies of oligomerization are also shown in comparison to the free energies of fibril formation obtained by direct measurement of the CAC ("Direct"), and also the free energies of fibril formation obtained from the fitted estimation of the CAC ("Fitted").

7.2.3 Estimations of the free energies of oligomer and fibril formation

From the results in Fig. 7.2a and b, the similarity in the slopes of the growth regions below the CAC of the Aβ40 and Aβ42 self-oligomerizing systems suggests that there is no large difference in the mean free energy of oligomerization in both cases. By fitting our model to the self-oligomerizing systems (Appendix E), we estimate the free energy of oligomerization for Aβ40, $\Delta G^{\circ(oligo)}_{40,40}$, as -36.3 ± 3.0 kJ mol^{-1}, and similarly $\Delta G^{\circ(oligo)}_{42,42}$ for Aβ42 as -36.3 ± 3.2 kJ mol^{-1} (Fig. 7.2). These values are different from those for the fibrils, which is consistent with the expected differences in the structure of oligomers and fibrils. The CAC for Aβ40 is estimated as 222 ± 10 nM by the same fitting procedure, and the CAC for Aβ42 is estimated as 86 ± 10 nM; these values allow independent estimation of $\Delta G^{\circ(fib)}_{40,40}$ as -39.5 ± 0.1 kJ mol^{-1} and $\Delta G^{\circ(fib)}_{42,42}$ as -42.0 ± 0.3 kJ mol^{-1}, demonstrating broad consistency with the direct measurements. The value of $\Delta G^{\circ(oligo)}_{40,42}$ is estimated to be -32.6 ± 2.6 kJ mol^{-1}, and the absence of apparent plateau in the co-oligomer plot (Fig. 7.2c) is consistent with both isoforms being present below their CAC values. According to these results, summarized in Fig. 7.2d, in all cases the free energy of oligomerization is large and negative. The seemingly small difference in the free energy of oligomerization for the formation of co-oligomers in comparison to the self-oligomers, however, leads to lower abundance of these species, as will be described later. To point out, while there have been previous reports of the free energy for fibril formation of Aβ [163, 164] and other amyloidogenic proteins [167], the directly measured free energies of oligomerization for Aβ40, Aβ42 and Aβ40-Aβ42, to our knowledge, are reported for the first time. The formation of the spectator co-oligomers means that, in the presence of both Aβ40 and Aβ42, fewer self-oligomers of Aβ40 or Aβ42 will be formed, so growth into Aβ40 or Aβ42 fibrils may be suppressed. This may provide an explanation of why the aggregation kinetics of both isoforms were observed to be mutually affected in the previous related studies [139], [146].

7.2.4 Predictions of oligomer populations at 1 nM concentration of Aβ

The obtained experimental values for the free energies of oligomerization can be used to predict the total oligomer concentration and the fraction of mixed and self-oligomers at pre-defined Aβ concentrations and ratios of Aβ40 and Aβ42. The measurements in this study have been carried out at 0–250 nM starting concentrations of Aβ, the range which is substantially lower than what can be accessed using more conventional experimental methods [145]. However, it is known that the physiologically related total concentration of this peptide is in the range of 1–10 nM [152]. To infer the information about oligomer types and sizes at these extremely low concentrations of Aβ, we can use the derived free energy values and set the starting total Aβ concentration to a chosen value within the physiological range. Figure 3a shows how the distributions of oligomer sub-populations are predicted to change in Aβ40 and Aβ42 mixture as a function of the Aβ42 proportion, when the total Aβ concentration is chosen to be 1 nM. Similar predictions with the total concentrations set to 5 nM and 10 nM are shown in Appendix Fig. E.1. Due to less negative free energy of co-oligomerization, the resulting predicted co-oligomer populations are lower than the self-oligomer populations at all mixing ratios of Aβ40 and Aβ42. The predominant oligomers at a physiologically-relevant ratio of 9:1 of Aβ40 to Aβ42 will be the oligomers of Aβ40, then a small fraction of co-oligomers with only a tiny fraction of Aβ42 oligomers. Moreover, the size distributions can be also inferred, as shown in Fig. 7.3b. At 1 nM of the total protein concentration, the main oligomers present are dimers, and the number of oligomers is predicted to decrease exponentially with oligomer size.

Since Aβ42 peptide is more hydrophobic than Aβ40, it is plausible that this difference would be conserved in the derived oligomers, which could influence their properties. Our previous study suggested that Aβ40 and Aβ42 oligomers are both cytotoxic, once formed [168]. Furthermore, our previous experimental data on the binding of Aβ40 and Aβ42 oligomers to neuronal cells suggested that, at the lowest concentration measured, the relative affinity of Aβ42 oligomers for the cell membrane was 4 times that of the Aβ40 oligomers [169]. If we assume that the affinity of the co-oligomers is 2 times that of the Aβ40 oligomers, a value intermediate between Aβ40 and Aβ42 oligomers, and that the majority of oligomers are dimers, according to Fig. 7.3b, we can then predict how the relative concentration of membrane-bound oligomers varies as a function of Aβ42 proportion, as is presented in Fig. 7.3c. This analysis predicts a clear increase in the relative number of oligomers bound to the cell surface with the increase in the proportion of Aβ42. Interestingly, the minimum number of cell-bound oligomers in this simulation occurs at a ratio of 9:1 of Aβ40 to Aβ42. Note that the oligomer size distribution (Fig. 7.3b) is not significantly altered by the ratios of Aβ40 and Aβ42 since the free energies of oligomerization are all comparable and in all cases are dominated by dimers. However, more of these dimers will contain Aβ42 as the proportion of Aβ42 increases. We note that while our analysis in Fig. 7.3c considers dimers, as they are the most abundant oligomers in our system, the prediction of absolute concentrations of large surface-bound oligomers is beyond the scope of this analysis due to the absence of additional oligomer to membrane interactions.

Clearly, this model may not be fully applicable to the Aβ oligomers in AD, since their formation under more complex *in vivo* environment is potentially affected by numerous extrinsic factors such as, for instance, the presence of small molecules and proteins, lipid surfaces, altered pH or ionic strength and the underlying assumption of thermodynamic equilibrium may not be correct. Nevertheless, it is interesting to compare the predictions of our model to what is actually observed in humans. From the results of a previous quantitative study where stable synthetic Aβ dimers were used as standards, the concentrations of Aβ oligomers in CSF of AD patients and controls were identified to be in the sub-picomolar range, in agreement with our predictions of the oligomer concentration at a total Aβ concentration of 1 nM, although the low concentration prevented the determination of the oligomer sizes in that work [170]. It is also interesting that the oligomer concentration measured *in vivo* appears to be determined by the Aβ monomer concentration in the CSF. AD patients will also have amyloid plaques containing Aβ40 and predominantly Aβ42 fibrils. In our experiments, the oligomer concentrations above fibrils are those shown in the plateau regions in Fig. 7.2. Overall the total oligomer concentration is about 20 nM, which is two orders of magnitude larger than around 0.1 pM observed *in vivo* [170]. This suggests that either the exchange between oligomers and fibrillar plaques does not occur to any significant extent *in vivo*, or that there are additional contributing factors which are not present in our analysis, for example, active degradation mechanisms that remove oligomers [171]. To note, even though the amount of Aβ42 in the CSF is generally observed to decrease in AD, our model would predict that this has little effect on the total oligomer concentration, because their population is largely dominated by Aβ40 oligomers. This may provide a simple explanation for why most diagnostic tests for AD to date based on detecting the Aβ oligomer concentration in CSF observe little significant difference between controls and AD patients [172].

Our model can be applied to predict how the number of membrane-bound oligomers changes upon increasing the ratio of Aβ42 to Aβ40 using pre-defined concentrations of Aβ which correlate with the onset of AD. While this analysis does not take account of any additional factors that may contribute to the disease in man [132], it serves to illustrate how significantly the starting concentrations of the two isoforms influence the resulting populations of potentially pathogenic oligomers. For example, in the case of the Beyreuther/Iberian mutation [173,174] where the ratio of Aβ42 to Aβ40 is as high as 22:1 [175], early onset of AD occurs before 40 years of age. If we use a starting peptide ratio of 22:1 in our simulations, the number of oligomers on the cell surface is predicted to increase by a factor of 4 relative to Aβ40 self-oligomers. Not only can a raised proportion of Aβ42 be pathogenic *in vivo*, but also the overall overproduction of Aβ. For example, in Down's syndrome there is an extra copy of the gene for APP, meaning that the total Aβ concentration is elevated by a factor of 1.5, leading to an early-onset AD at around 40 years [176]. If we increase the total peptide concentration by a factor of 1.5 in our model, the total Aβ oligomer concentration increases by 125%, and the predicted number of cell-bound oligomers increases by a factor of 2.1 relative to the number of oligomers bound for 100% Aβ40 at the initial total Aβ concentration. While a change in Aβ40 to Aβ42 ratio from 9:1 to 7:3 results in no overall increase in the total number of oligomers, there

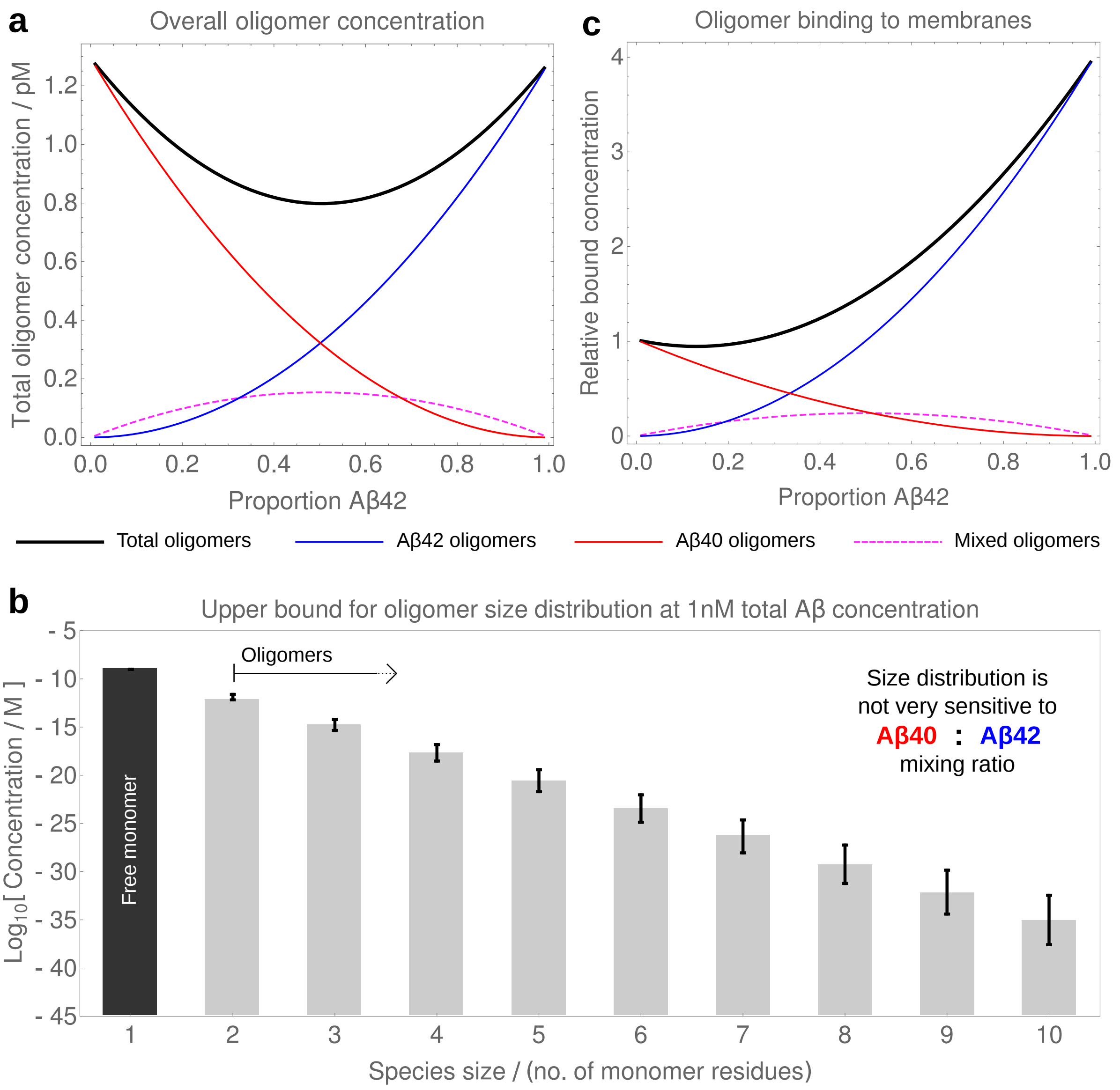

Figure 7.3: Simulation of Aβ40-Aβ42 co-oligomerization equilibrium behaviour at a total Aβ concentration of 1 nM for a range of Aβ42 proportions, using $\Delta G^{\circ(oligo)}_{40,40}$, $\Delta G^{\circ(oligo)}_{40,42}$, $\Delta G^{\circ(oligo)}_{42,42}$. Simulations at 5 nM and 10 nM of total Aβ are shown in Appendix Fig. E.1. (**a**) Total oligomer concentration and composition as a function of Aβ42 proportion. (**b**) Estimated concentrations of oligomers of different sizes at 1 nM total protein concentration, calculated by assuming that $\Delta G^{\circ(oligo)}_{40,42}$ is unchanged from the single-species value (in which case the ratio of Aβ40:Aβ42 is irrelevant). The true distribution will decline with oligomer size even more rapidly, as visual inspection of the data shows $\Delta G^{\circ(oligo)}_{40,42}$ to be less favourable than the single-species values. The error bars correspond to averaged uncertainty in the ΔG measurements. (**c**) The relative concentration of oligomers estimated to be bound to the surface of a neuronal cell, expressed relative to the concentration of oligomers bound to the surface at 1 nM of Aβ40. This result assumes that the relative affinity of co-oligomers for the cell membrane is 2 times higher than the affinity of Aβ40 oligomers, and that the relative affinity of Aβ42 oligomers is 4 times higher than that of Aβ40 oligomers.

is a significant difference in their predicted composition, with more co-oligomers being bound. In addition, the co-oligomers may be more persistent than self-oligomers, since they cannot grow into less toxic fibrils [85], so it is possible that the increased persistency of co-oligomers additionally contributes to the increased toxicity.

7.3 Summary and Conclusions

Our results show that co-oligomers of Aβ40 and Aβ42 can be formed at sub-micromolar concentrations of Aβ with little free energy penalty. This finding suggests that the environment of the Ile-Ala dipeptide on Aβ42 does not change significantly between the monomeric and oligomeric state, and that the effect of these additional peptides on the free energy of oligomerization is small. There are multiple other isoforms of Aβ present because of truncations, mutations, ubiquitination or post-translational modifications. If there is no high penalty in the free energy of co-oligomerization, then these species may potentially be formed by various isoforms of the peptide since mixing entropy under such conditions favours the formation of mixed rather than purely segregated aggregates. It is likely therefore that under *in vivo* conditions where multiple isoforms are present, such mixed aggregates are prevalent. Thus, any comprehensive therapeutic strategy based on antibodies that bind Aβ may need to take account of the presence of co-oligomers in addition to self-oligomers of Aβ. At present it is still unclear which forms of Aβ are the true pathogens in AD [177], and the contribution of Aβ co-oligomers to AD may not have been recognized to date.

The techniques developed in this paper soon found use in the investigation of α-synuclein oligomers in the next chapter. The uncertainty around oligomer geometry that prompted my initial involvement in this project later inspired me to directly investigate non-linear oligomer geometries, the results of which form the chapter-after-next.

Chapter 8

Thermodynamics of αS co-oligomer formation

Chapter Abstract

Small oligomers of the protein α-synuclein (αS) are highly cytotoxic species associated with Parkinson's disease (PD). In addition, αS can form co-aggregates with its mutational variants and with other proteins such as amyloid-beta (Aβ) and tau, implicated in Alzheimer's disease. The processes of self-oligomerisation and co-oligomerisation of αS are however, challenging to study quantitatively. Here, we have utilised single-molecule techniques to measure the equilibrium populations of oligomers formed *in vitro* by mixtures of wild-type αS with its mutational variants and with Aβ40, Aβ42 and a fragment of tau. Using a statistical mechanical model, we find that co-oligomer formation is generally more favourable than self-oligomer formation at equilibrium. Furthermore, self-oligomers more potently disrupt lipid membranes than co-oligomers. However, this difference is sometimes outweighed by the greater formation propensity of co-oligomers when multiple proteins co-exist. Our results suggest that co-oligomer formation may be important in PD and related neurodegenerative diseases.

Here, we extend the methodology of the previous chapter and employ a statistical mechanical model to study both the propensity of co-oligomers to form between wild-type α-synuclein and other peptides, and their relative ability to disrupt membranes. The work presented here forms part of a manuscript under review at *ACS Nano*, on which I am joint first author. I was responsible for all modelling and analysis presented in this chapter. Experimental details have been largely neglected; the experiments themselves were carried out by M. Iljina, P. Flagmeier, and S. De. The statistical mechanical model employed, derived by G. A. Garcia, is detailed in Appendix F.

8.1 Introduction

The aggregation of the protein alpha-synuclein (αS) from its soluble monomeric form into amyloid fibrils is associated with a range of devastating neurodegenerative disorders such as Parkinson's disease (PD) and a series of related synucleinopathies. [178] In these conditions, amyloid fibrils of αS are segregated into cytoplasmic brain inclusions, Lewy Bodies or Lewy Neurites, [179] although molecular complexes consisting of a small number of αS monomers, termed oligomers, are increasingly recognised as the most cytotoxic forms giving rise to the disease etiology. [1, 100, 122, 180, 181] It has also been found that single residue mutational variants of the protein, including A30P, E46K and A53T, are associated with familial forms of PD. [182–184] Moreover, aggregates of αS are observed in more than half of the patients suffering from Alzheimer's disease (AD) along with the more characteristic deposits of the amyloid-beta (Aβ) peptide and the protein tau. [185] The co-occurrence of aggregates of αS, Aβ and tau has been reported in a variety of neurodegenerative conditions, and it has been suggested that this enhancement could be due to co-interactions between the proteins. [186]

αS has been shown *in vitro* to assemble into a large variety of oligomers with distinct morphologies, structures and functional properties. Such oligomers can be formed via different mechanisms and include nonfibrillar off-pathway species and prefibrillar intermediates of the amyloid fibril formation process. [100, 187–190] The mutational variants of αS have been found to affect the rates of αS aggregation [191–193] and oligomer formation, [127] and to alter its interactions with lipid membranes. [15] The effects of the amino acid substitutions on the rate of oligomer formation and the number of oligomers remain unclear. Previous studies have reported either increased [122, 192, 194], or unaltered [127] levels of oligomers formed by the mutational variants relative to wild-type αS. Additionally, the formation of co-oligomers between different variants of αS has been observed to occur in aqueous solution. [195]

In addition to its ability to self-assemble into a variety of oligomeric species, αS has been reported to interact with other proteins, including Aβ and tau. For example, the effects of αS and Aβ co-interaction on the aggregation of Aβ42 *in vitro* have recently been delineated, [196] and the direct binding between αS and Aβ40 and Aβ42 in solution has been characterised, [197–199] as well as the direct interaction of αS and tau. [200–202] The formation of co-oligomers containing both αS and Aβ has been predicted by several molecular dynamics simulations. [203–205] Furthermore, the formation of a dimer comprising αS and Aβ domains upon their coordination to Cu^{2+} has been reported, [206] and the formation of co-oligomers of αS and tau has been observed *in vitro*. [207] The accumulation of co-oligomers of αS and tau in the human brain has been shown using novel conformational-specific antibodies. [208] In addition, the formation of co-oligomers of αS with Aβ and tau in human red blood cells has recently been reported. [209]

Despite the accumulating evidence that αS can form self-oligomers as well as co-oligomers with other proteins, the extent of formation and the properties of these co-oligomeric species remain to be determined and quantified. Most *in vitro* investigations, particularly at low physiologically-relevant protein concentrations, have focussed on the self-oligomerisation of αS, since its co-oligomerisation is difficult to detect and quantify under these conditions. A comprehensive characterisation of the self-oligomerisation process of αS and its co-oligomerisation with the mutational variants or Aβ and with tau can help to compare their relative abundances, and functional assays can reveal the relative cytotoxicities of the self- and co-oligomeric species. To address these issues, we have applied statistical mechanical modelling to experimental measurements of self- and co-oligomers. Using these methodologies, we have analysed the self-oligomerisation of wild-type αS, its co-oligomerisation with the mutational variants A30P, A53T and E46K, and its co-oligomerisation with Aβ40 and Aβ42 and tau construct k18, that contains four repeats of the aggregation-prone region of full-length tau, [210] at physiologically-relevant protein concentrations in aqueous solution. We have derived the equilibrium free energies of oligomer formation for the studied protein combinations, enabling a quantitative comparison of the self- and co-oligomerisation processes. Furthermore, we have investigated the ability of the characterised self- and co-oligomers to permeabilize lipid vesicles, allowing us to quantify and compare their potential for lipid membrane disruption.

8.2 Results and Discussion

8.2.1 TCCD measurements of the formation of self-oligomers

Measurements of equilibrium self-oligomer concentrations formed by wild-type αS (subsequently abbreviated to “WT”), A30P, A53T and E46K αS over a range of initial monomer concentrations and under non-fibril-forming conditions were performed by M. Iljina using a previously-reported method [161]. In addition to αS, concentrations of self-oligomerisation under the same conditions of tau protein k18, which comprises the central region of this protein that is included in the microtubule binding region, and is frequently used to study tau aggregation, were provided. [211]

In this method, half of the monomers are labelled with one fluorescent dye and the other half are labelled with another dye. Only oligomers containing monomers of both types are detected, and we therefore term the resulting oligomer concentrations as “apparent oligomer concentrations”. The resulting plots of the apparent concentrations of self-oligomers of αS A30P, A53T, E46K, WT and of tau k18 are shown in Fig. 8.1. The data show that the oligomer populations formed under these conditions are below 200 nM, highlighting the challenge of studying these species by less sensitive bulk experimental techniques. The oligomer populations of the αS variants A30P, A53T and E46K reached

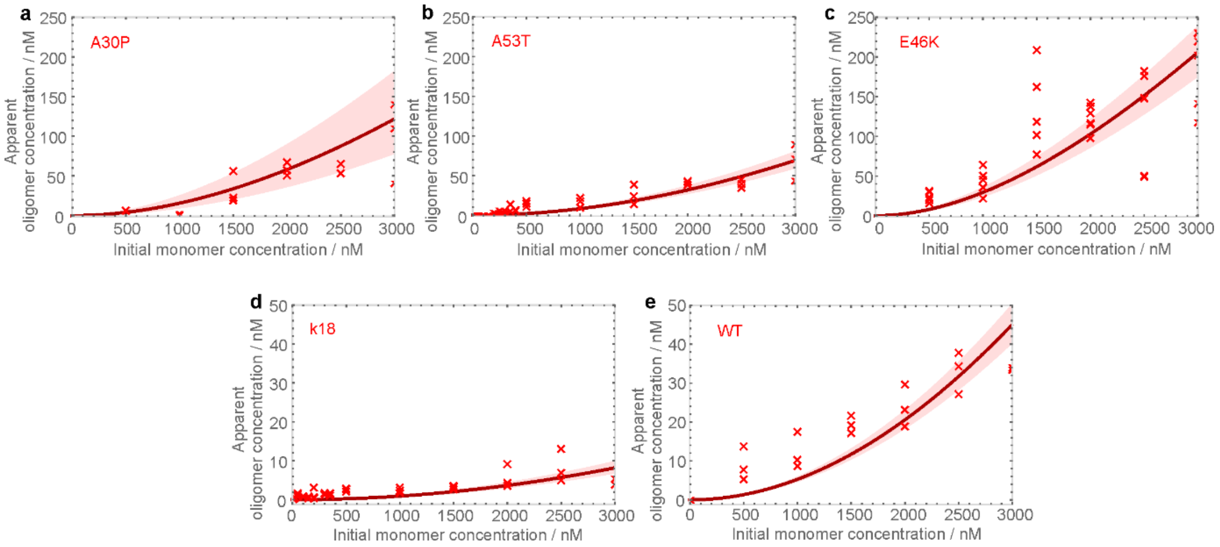

Figure 8.1: Equilibrium populations of self-oligomers plotted against the total initial protein concentrations. Three to five separate samples were analysed for each initial concentration of monomeric protein. The resulting values are represented by red crosses, and the fit to a single-peptide oligomerisation model is shown as a solid line. The shaded bounds represent the fitting error derived using nonparametric bootstrapping, as detailed in the Methods section. Data provided by M. Iljina.

higher levels than the WT protein, with E46K giving the highest apparent oligomer concentrations. The highest apparent equilibrium concentrations of oligomers were observed at the highest total starting αS concentrations, as expected by mass action, in agreement with our previous study of oligomer formation by αS. [28] The apparent concentrations of oligomers formed by k18 were mostly below 10 nM, which is consistent with the expected low aggregation propensity of this protein in aqueous buffer solution in the absence of aggregation inducers. [211] They were nevertheless sufficiently above the background (evidence not shown), confirming the formation of a low population of oligomeric species under the incubation conditions used here.

8.2.2 TCCD measurements of the formation of co-oligomers

We next set out to determine whether or not αS WT could co-oligomerise with its mutational variants, as well as with Aβ40, Aβ42 and k18, by combining equimolar ratios of the WT with the other proteins. The 1:1 mixtures of αS WT with its mutational variants are likely to be physiologically-relevant as the variants of αS are known to co-exist with αS WT *in vivo*. [182–184] The interactions of αS with Aβ40 and with Aβ42, both of which are largely extracellular, is of interest in the light of the reported role extracellular αS may play in neurodegeneration. [212] The potential co-interaction of αS WT with the tau construct k18 is of interest since both αS and tau proteins are predominantly intracellular, and therefore are likely to co-exist under the same conditions.

To quantify the extent of co-oligomerisation of the various protein combinations, we combined and incubated 1:1 stoichiometric ratios of αS WT, singly-labelled with one dye, with each of these other proteins, singly-labelled with another dye, and selectively monitored the formation of co-oligomeric species that contain both fluorophore labels by sm-TCCD following the same protocols as described above for the detection of self-oligomers. The apparent oligomer concentrations measured in this way are shown in Fig. 8.2, indicating that co-oligomeric species are formed with all of the protein combinations. The apparent equilibrium concentrations of co-oligomers of αS WT and A30P (Fig. 8.2(a)) can be seen to be similar to the apparent concentrations of self-oligomers of αS WT (Fig. 8.1(e)). Interestingly, the apparent concentrations of αS WT co-oligomers with A53T and E46K (Fig. 8.2(b) and (c)) are similar to the apparent concentrations of oligomers generated by the mutational variants alone (Fig. 8.1(b) and (c)), and are higher than the apparent concentrations of the self-oligomers of αS WT (Fig. 8.1(e)). Our observation that αS WT can readily co-oligomerise with all of the chosen mutational variants differs from the conclusion of a relatively recent single-molecule study that identified selectivity in the co-interactions with these different isoforms. [195] The experimental conditions in the two studies are, however, different and the incubations were carried out over much longer timescales in the present experiments, thus making the results not directly comparable. The observed populations of co-oligomers between WT and k18 can be seen to be below 20 nM over a similar monomer concentration range (Fig. 8.2(d)). Despite the low levels of such species, the direct detection of co-oligomerisation of αS with tau is interesting, especially considering that there are at least six major isoforms of tau in the human brain. [213] The concentrations of co-oligomers between αS and Aβ detected appear even lower, but in fact can be predicted to be higher than for all other co-oligomers when comparison is made over the same initial monomer concentration range. This is also an interesting result, especially given that multiple isoforms of Aβ can be present *in vivo*. [214] Indeed, it is likely that various isoforms of tau or of Aβ may also co-assemble since the extent of co-aggregation between different proteins is known to be determined by the identity of their primary sequences. [215]

8.2.3 Theoretical modelling of self- and co-oligomer datasets to determine the free energies of oligomer formation

We then analysed the datasets observed for self- and co-oligomer formation by sm-TCCD (Figs. 8.1,8.2) using a statistical mechanical model similar to a previously reported model. [161] In this theoretical model, the key parameter describing the oligomerisation process is the Gibbs free energy of monomer addition to an oligomer or another monomer, ΔG°, independent of oligomer size, as illustrated schematically in Fig. 8.3(a). This parameter characterises the ease of oligomer formation by monomeric species, and the more negative the value the more favourable the oligomerisation.

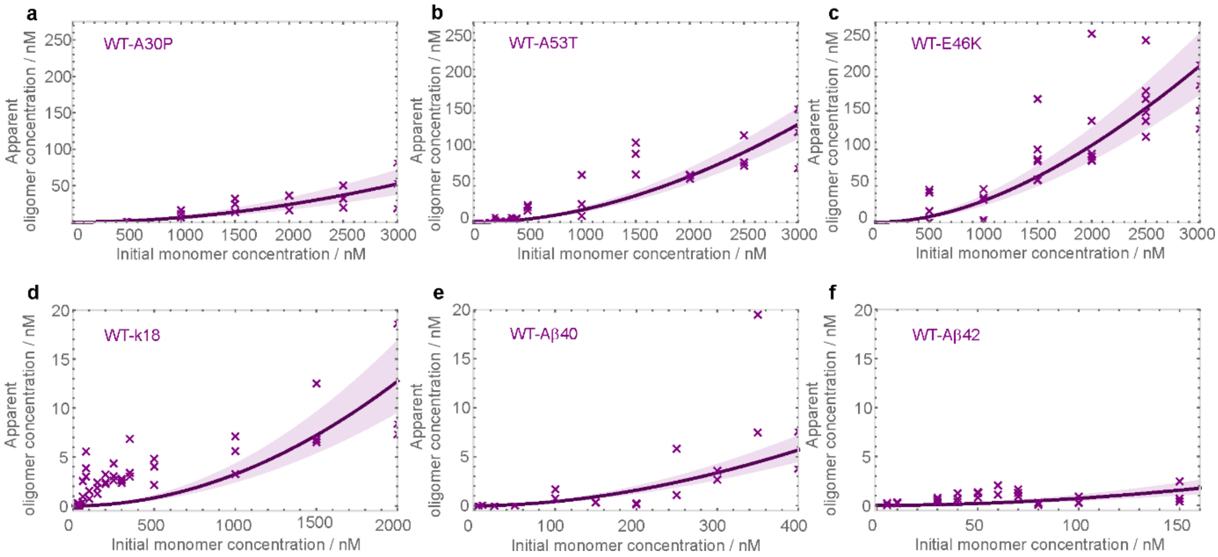

Figure 8.2: Equilibrium populations of co-oligomers formed in 1:1 mixtures of the various protein combinations as indicated in the panels. Two to five separate samples were studied at each concentration. Individual values are represented as purple crosses, and the fit to a 2-peptide model of oligomer formation is denoted by a solid line. The shaded bounds represent the fitting error, as in Fig. 8.1. Note the apparent low oligomer concentrations in **(e)** and **(f)** are due to the lower monomer concentration ranges investigated (see x axes). In fact these concentrations are predicted to be higher than for other co-oligomers over the same monomer concentration range. Data provided by M. Iljina.

In the present model, described in the Methods section, the oligomers were treated as simple non-interacting one-dimensional chain structures with nearest-neighbour interactions independent of chain length. The approximation of non-interacting oligomers is reasonable given the very low apparent concentrations of oligomers in the present experiments. For self-oligomers our model is identical to that previously reported. [161] Note again that the modelling explicitly takes into account the fact that only oligomers containing both dyes will be detected whereas the singly-labelled oligomers are not detectable. The modelling of linear co-oligomers requires substantial new theory to be developed. Under the present conditions, most oligomers are inferred to be dimeric (see Appendix F). Therefore, although we cannot rule out the existence of larger nonlinear oligomeric species (e.g. tetrahedral clusters) not explicitly included in our linear oligomer model, they cannot in any case be present at high enough concentrations under the present experimental conditions to significantly affect our analysis. We emphasize that oligomers have been shown to have a different structure, stability and toxicity compared to fibrils; [28, 100] thus, our ΔG° is not the same as for fibril elongation.

Our model is not applicable to the analysis of the data derived for the samples containing Aβ at the highest concentrations investigated, since we expect fibrillar aggregates to exist at equilibrium above the critical aggregation concentration for fibril formation, which is the lowest total monomer concentration that is required for the formation of fibrils. [161] Using the same fitting procedure, we previously estimated these critical ag-

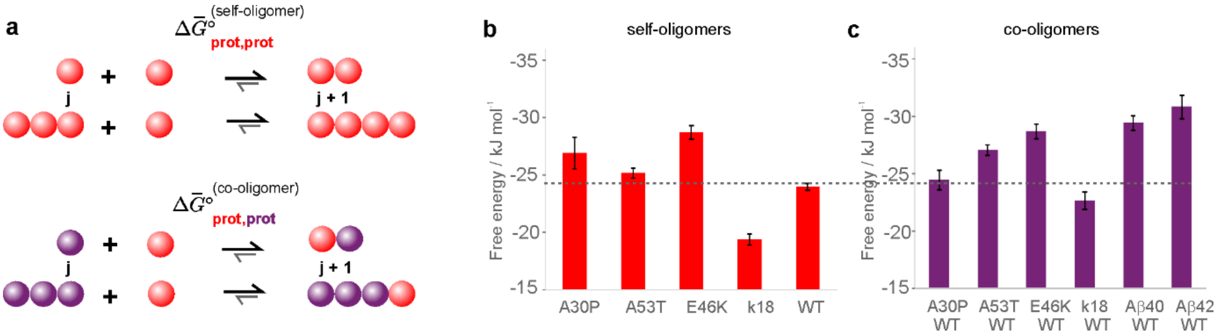

Figure 8.3: Modelling of self- and co-oligomer formation. **(a)** Schematic representation of the two processes. **(b)** The fitted values of free energies of oligomerisation, ΔG°s, derived from the analysis of sm-TCCD data for the formation of self-oligomers and **(c)** of co-oligomers. The standard errors in the ΔG° values were determined using a nonparametric bootstrap approach as defined in Methods. The dotted line denotes the ΔG° for the formation of oligomers from WT.

gregation concentration values to be 222±10 nM and 86±10 nM for Aβ40 and Aβ42 isoforms, respectively. [161] Since Aβ comprises only a half of the protein molecules in the αS-Aβ solutions, we restricted our analysis of these datasets to total protein concentrations that were up to twice those previously derived critical aggregation concentration values.

Following the fitting of self- and co-oligomer datasets, as detailed in the Methods section, we derived the value of ΔG° of oligomerisation of all the protein combinations investigated in this study. The resulting values, summarised in Figs. 8.3(b)-(c) and listed in Appendix Table F.1, are all large and negative, similar in magnitude to the values previously reported for the formation of Aβ40/42 oligomerising systems. [161] The least negative ΔG° value of -19.4±0.5 kJ mol^{-1} was obtained for k18, which is consistent with its lowest propensity to self-assemble under our experimental conditions in the absence of aggregation inducers. Interestingly, the resulting ΔG° value for αS WT of -24.0±0.3 kJ mol^{-1} is less negative than the resulting values for all mutational variants of αS, except WT-A30P. In addition, the ΔG° values for WT-Aβ co-oligomers (-29.4±0.6 kJ mol^{-1} for WT-Aβ40 and -30.8±1.0 kJ mol^{-1} for WT-Aβ42) are more negative than the value for αS WT, and less negative than our previously obtained values for Aβ40 and Aβ42 self-oligomers (-36.3 kJ mol^{-1} for both isoforms). [161] Overall, the more negative ΔG° values for the mixtures containing the WT protein compared to αS WT alone suggest that the co-oligomer formation by αS WT is more favourable than the self-oligomer formation for most of the mixtures under our experimental conditions.

8.2.4 Predicted equilibrium concentrations of oligomer populations based on the derived free energies of oligomer formation

The derivation of the free energies of oligomer formation enables the prediction of the total rather than apparent equilibrium concentrations of self- and co-oligomers at any chosen protein concentration. Given that oligomer concentrations depend exponentially on ΔG°s (as detailed in Appendix E; in particular Eqs (E.7) and (E.11)), even slight differences in the derived ΔG° values correspond to large differences in the resulting equilibrium oligomer populations. Indeed, under the present conditions, a stabilization of less than 2 kJ mol^{-1} is sufficient to result in a doubling of the observed oligomer concentration. To illustrate this statement, we set out to explore the relative concentrations of different types of oligomers formed by the mixtures of αS WT with its mutational variants. To this end, we predicted the overall oligomer concentrations and the concentrations of the separate oligomer sub-populations for the 1:1 mixtures of WT with its variant A30P, A53T and E46K generated over the total range of protein concentrations from 10 to 1,000 nM, and compared them to the predicted oligomer populations for αS WT at the same starting protein concentrations. These predictions reveal that the co-oligomers of αS WT-A53T and WT-E46K are the dominant sub-populations across the entire ranges of the protein concentrations (Fig. 8.4). In addition to the concentrations of the self- and co-oligomer sub-populations in the αS WT-mutational variant mixtures, we also predicted the concentrations of oligomers for αS WT at a concentration of αS of 1.5×, which is relevant to the scenario of αS overproduction [216, 217] and corresponds to the conditions of its gene duplication. [218] According to these simulations, the total oligomer concentration in the αS WT-mutational variant mixtures can reach similar levels to the 1.5WT system, and even higher concentrations of oligomers are generated in the case of WT-E46K combination.

Importantly, this analysis shows that the generation of high concentrations of oligomers can be caused not only by the increased concentration of soluble αS WT, but can also occur as a result of the presence of the protein variants that have a favourable interaction with the αS WT owing to sequence similarity. [215] The propensity identified in this study of the mutational variants of αS to generate elevated equilibrium concentrations of oligomers and co-oligomers may explain their pathogenicity, along with other previously established disease-relevant properties of these mutational variants such as perturbed aggregation rates, [191–193] altered binding to lipid membranes [15, 219] and the altered structural organisation of the oligomers. [127]

Subsequently, we used the derived ΔG°s to predict the populations of the defined oligomeric species at the protein concentrations that correspond to their physiological abundance, focussing on the mixture of αS WT and Aβ isoforms. Using the protein concentrations that might be present inside a cell, 0.9 nM and 0.1 nM for Aβ40 and Aβ42,

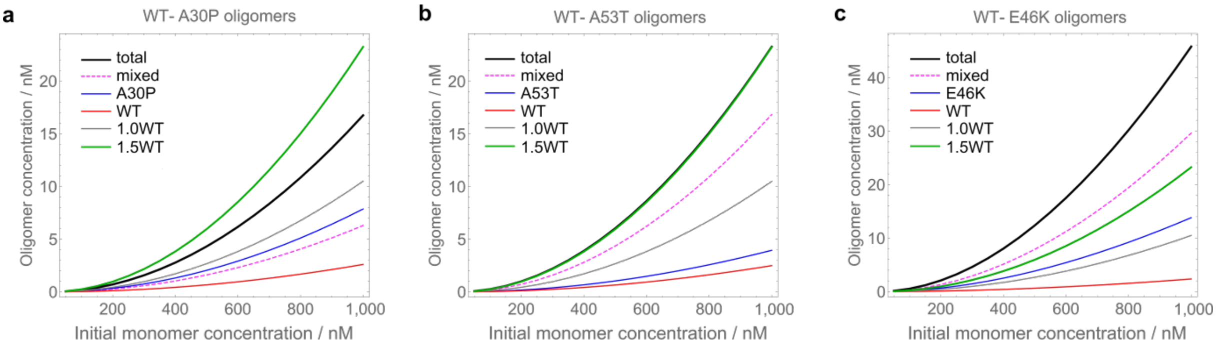

Figure 8.4: Predicted equilibrium concentrations of protein oligomer populations over the concentration range of 10-1,000 nM. In **(a)-(c)**, the total concentrations of oligomers generated in 1:1 mixtures of αS WT-mutational variant are shown ("total"), and the separate oligomer sub-populations that are present in these mixtures (co-oligomers are denoted as "mixed", and the self-oligomers of either component as "WT" and the corresponding mutational variant). In addition, αS WT self-oligomers generated at the same total protein concentrations are shown ("1.0WT"), and the oligomers generated at 1.5-times higher total starting total concentration of αS WT ("1.5WT"). Note that in **(b)**, the curves "1.5WT" and "total" overlap.

assuming the intacellular concentration is similar their extracellular concentrations, [152] vs 300 nM for αS WT, based on the measured K_d for membrane binding, [220] the major oligomeric form is predicted to be self-oligomers of αS WT (0.96 nM), followed by co-oligomers of αS WT and Aβ (0.05 nM and 0.01 nM for WT-Aβ40 and WT-Aβ42), with Aβ self-oligomers being at concentrations that are orders of magnitude lower. Thus, under these conditions, most of the Aβ peptides within the oligomeric species are incorporated into the co-oligomers with αS WT. At the protein concentrations that correspond to their extracellular abundance in cerebrospinal fluid (CSF), (0.9 nM and 0.1 nM for Aβ40 and Aβ42 vs 1 nM for αS WT), [152, 221] the predicted dominant type of oligomer are self-oligomers of Aβ40 (83%) at a concentration of 1.1 pM. Aβ42-containing oligomers are present at 0.04 pM: 70% of these latter oligomers are predicted to be co-oligomers of αS WT and Aβ42, with 30% being self-oligomers of Aβ42. Remarkably, most of Aβ42 is thus predicted to be a constituent of co-oligomers under these conditions.

8.2.5 Co-oligomer formation leads to increased calcium influx into lipid vesicles due to membrane disruption

Next, we set out to examine if the self- and co-oligomers that αS forms with its mutational variants and with the Aβ peptide are able to permeabilise lipid membranes, a process that has been associated with neurotoxicity. [222] Measurements of the oligomer-induced permeabilization of lipid membranes were provided by P. Flagmeier and S. De, using a recently developed ultrasensitive single vesicle assay, that quantifies Ca^{2+} influx into lipid vesicles upon membrane disruption based on fluorescence intensity changes of a calcium sensitive dye. [223] Using this assay and identical experimental conditions, it was previously reported that neither monomeric Aβ peptide nor monomeric αS WT cause significant calcium influx; [223] therefore, any influx observed can be attributed to the action of peptide aggregates on the vesicle membranes. The detection of a higher fluorescence intensity corresponds to a larger calcium influx due to a higher level of membrane permeabilization caused by the aggregates. [223]

Firstly, the action of αS WT with and without its mutational variant E46K was examined, since this combination of αS proteins generated the highest steady-state concentrations of mixed oligomers, was expected to have comparable lipid-binding properties and not to form lipid-induced aggregates under our experimental conditions. Separate solutions of 150 nM monomeric αS WT and of 150 nM monomeric αS E46K variant were prepared and incubated under the same conditions as were used for the oligomer concentration measurements above (Fig. 8.1(e),(c)), and probed with the assay (Fig. 8.5(a)). Significant Ca^{2+} influx was observed in both cases, confirming that the self-oligomers formed by these αS variants can disrupt membranes. We then used our theoretical model to calculate the equilibrium concentration of oligomers formed in each of these solutions (Table 8.1). Comparing these figures to the permeabilization data allows us to quantitatively determine the ability of the respective self-oligomers to induce Ca^{2+} influx (the permeabilization propensity, measured in units of percentage point (ppt) Ca^{2+} influx / pM).

αS solution (initial monomer concentrations)	WT self-oligomer concentration (pM)	E46K self-oligomer concentration (pM)	WT-E46K co-oligomer concentration (pM)
WT (150 nM)	240	-	-
E46K (150 nM)	-	1500	-
WT + E46K (75 + 75 nM)	60	370	740

Table 8.1: Predicted oligomer concentrations for αS WT-E46K solutions.

αS or Aβ solution (initial monomer concentrations)	WT self-oligomer concentration (pM)	Aβ42 self-oligomer concentration (pM)	WT-Aβ42 co-oligomer concentration (pM)
αS WT (40 nM)	17	-	-
Aβ42 (4 nM)	-	21	-
WT + Aβ42 (40 + 4 nM)	17	20	49

Table 8.2: Predicted oligomer concentrations for αS WT-Aβ42 solutions.

Next, solutions containing both monomeric αS WT and the E46K variant at equimolar concentrations of 75 nM (total concentration of 150 nM) were prepared, incubated under the conditions described above to allow the formation of oligomer populations, and tested (Fig. 8.5(a)). We again used our theoretical model to determine the concentrations of WT self-oligomers (60 pM), E46K self-oligomers (370 pM), and WT-E46K co-oligomers (740 pM) present in these solutions (Fig. 8.5(b)). Based on the concentrations of the self-oligomers and their ability to induce Ca^{2+} influx that we have calculated above, we can determine how much of the Ca^{2+} influx of the mixture can be attributed to αS WT self-oligomers (3.7 ppt) and αS E46K self-oligomers (5.6 ppt). The co-oligomers are therefore responsible for 37.1 ppt of the Ca^{2+} influx. Dividing this by the predicted co-oligomer concentration gives us the permeabilization propensity of the WT-E46K co-oligomers.

These results show that although E46K-containing oligomers are formed much more readily than WT oligomers (Table 8.1), the ability of E46K self-oligomers to cause Ca^{2+} influx is lower compared to the αS WT-containing oligomers, indicated by their lower permeabilization propensity (Fig. 8.5(b)). The co-oligomers both readily form and significantly disrupt membranes, and are thus responsible for the great majority, namely 80%, of the observed Ca^{2+} influx caused by the mixed αS WT-E46K solution (Fig. 8.5(b)).

Since the toxicity of Aβ42 self-oligomers is also well-established, [101, 223] it is of interest to test the potential toxicity of αS WT-Aβ42 co-oligomers. Choosing monomeric concentrations at which co-oligomer formation was predicted to be particularly high, we performed membrane permeabilization experiments (Fig. 8.5(c)) for solutions containing only self-oligomers (generated from incubation of 4 nM Aβ42 or 40 nM αS WT, respectively) and solutions containing self- and co-oligomers (generated from incubation of 4 nM Aβ42 + 40 nM αS WT). We employed our theoretical model to calculate the concentrations of αS WT self-oligomers (17 pM), Aβ42 self-oligomers (20 pM), and WT-Aβ42 co-oligomers (49 pM) formed in the mixed solution. Combined with theoretical predictions of the oligomer concentrations formed in the single-species solutions (Table 8.2), the data enabled us to perform the same analysis as above to determine the relative ability of these different oligomeric species to permeabilise membranes. Our results show that although more than half of the oligomers formed in the mixed solution are co-oligomers, their contribution to the ability of the solution to permeabilise membranes is small, and therefore co-oligomers are significantly less disruptive than the self-oligomers (Fig. 8.5(d)). Interestingly, we also find that Aβ oligomers are over an order of magnitude more disruptive to membranes than αS oligomers; and that they are present at similar concentrations (Table 8.2). The greater number of co-oligomers in this case is insufficient to outweigh the greater membrane permeabilization propensity of Aβ42 self-oligomers, which remain responsible for the majority (80%) of the Ca^{2+} influx caused by the oligomers present in the WT-Aβ42 mixed solutions (Fig. 8.5(d)).

As the αS WT protein co-exists in the human brain with its E46K variant in subjects displaying a single E46K mutation, which means the person carries one WT αS allele and one allele with the E46K mutation, species formed during co-aggregation may be involved in membrane permeabilization and the loss of protein homeostasis. Our results show that the presence of E46K mutation leads to the increased formation of E46K self- and co-oligomers and not to these structures being inherently more toxic. Our results also indicate substantial co-oligomerisation between the PD-related αS protein and the AD-associated Aβ42 peptide. However, although this hints at a possible cross-seeding

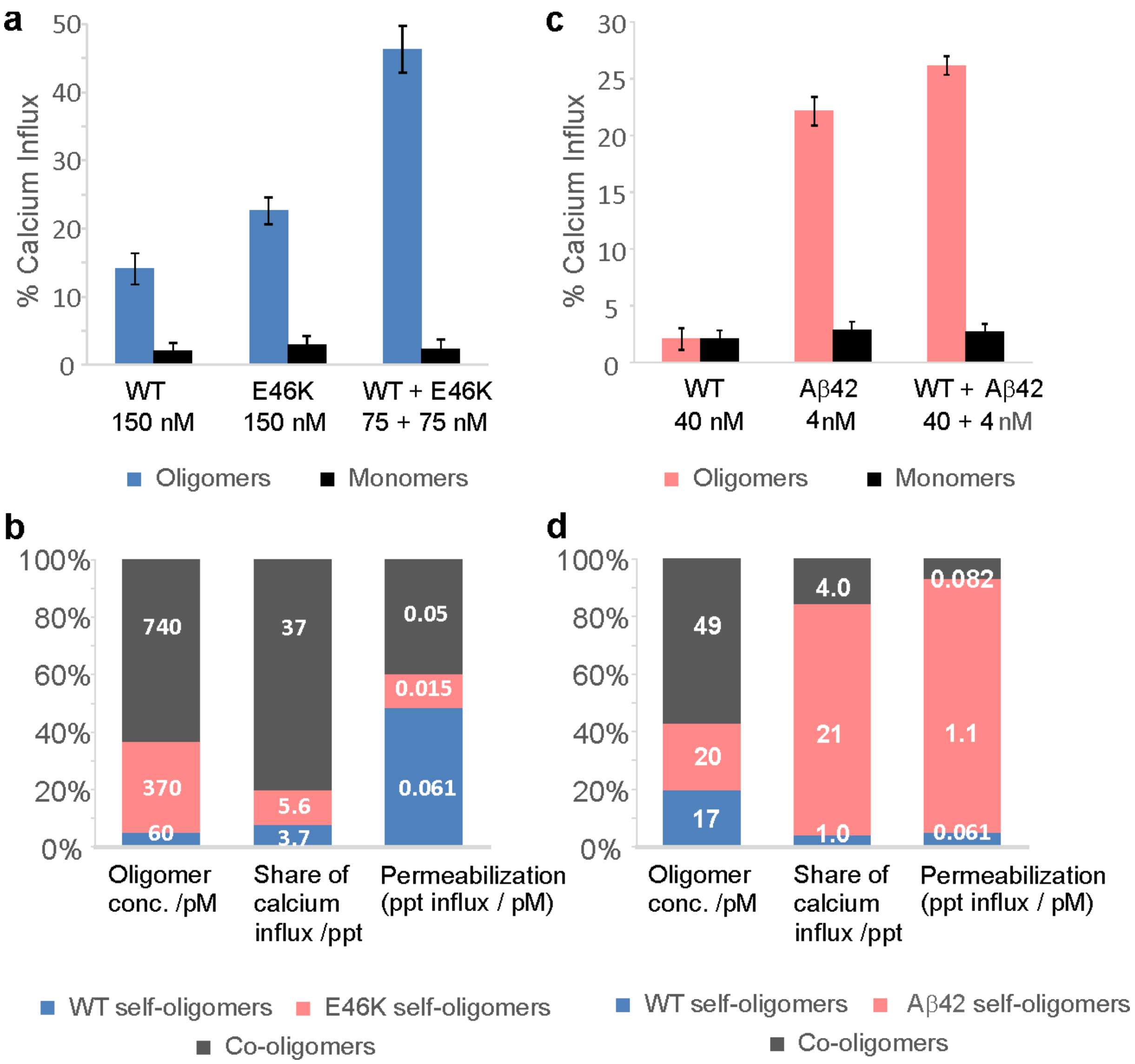

Figure 8.5: Quantification of the Ca^{2+} influx induced by oligomers using the single vesicle assay. **(a)** Experimentally measured average values of Ca^{2+} influx, induced by the oligomers formed at equilibrium in solutions containing αS WT and E46K (concentrations in monomer equivalents). Error bars correspond to the standard deviations from three separate experiments for each experimental condition ($p = 0.032$). **(b)** Comparing self- and co-oligomer populations to their contributions to total Ca^{2+} influx, and their relative permeabilization propensity, in the αS solution containing both 75 nM WT and 75 nM E46K. **(c)** Average values of Ca^{2+} influx, induced by the oligomers formed at equilibrium in solutions containing αS WT and Aβ42 (concentrations in monomer equivalents). Error bars correspond to the standard deviations from four separate experiments for each experimental condition ($p = 0.017$). **(d)** Comparing self- and co-oligomer populations to their contributions to total Ca^{2+} influx, and their relative permeabilization propensity, in the solution containing both αS (40 nM WT) and Aβ42 (4 nM). Data provided by P. Flagmeier and S. De.

effect that may have relevance for amyloid fibril formation, the co-oligomers themselves appear to be relatively inert in the single vesicle assay compared to the self-oligomers of Aβ42 and thus may have comparatively little direct effect on cell membrane disruption.

Taken together, our combined experimental and theoretical methodology permits us not only to determine the concentrations of different types of oligomers formed from mixtures of amyloidogenic monomeric peptides and proteins at biologically-relevant concentrations, but also to gain insight into their likely relative toxicities due to cell membrane disruption. In addition, the close agreement between the permeabilisation propensity values calculated for αS self-oligomers formed from 40 nM and 150 nM monomeric αS solutions (Fig. 8.5(b),(d)) clearly demonstrates the self-consistency of our modelling approach. The number of possible cytotoxic αS-containing co-oligomer types can, in principle, be very large *in vivo*, considering that new αS mutational variants are still being discovered [224, 225] and numerous post-translationally modified and truncated proteoforms of αS are physiologically abundant, [226] as well as multiple isoforms of Aβ, [227] and other amyloidogenic proteins and their complexes. [228] More generally, most proteins associated with neurodegenerative diseases are found to be post-translationally modified in the aggregated state so that the formation of co-oligomers, either between different post-translationally modified forms of the same protein or different proteins present in the same cellular compartment, is likely to play a much more important role in neurodegenerative diseases than previously thought.

8.3 Conclusions

In this study, we have performed a detailed quantitative analysis of self- and co-oligomer formation by αS and its mutational variants *in vitro*, and investigated its co-oligomerisation with two major AD-related proteins, Aβ and tau k18 construct. We have compared the potential for lipid membrane disruption by the different oligomer types using a single-vesicle assay. Our results show that at low physiologically-relevant protein concentrations, co-oligomer formation in most cases is more favourable than self-oligomer formation by αS, and thus αS-containing co-oligomers may be highly abundant under conditions where multiple proteins co-exist. We also find that although the co-oligomers examined are less potent membrane disruptors in comparison to self-oligomers, their adverse effects can become dominant when the co-oligomers are present at high steady-state concentrations. The identified favourable formation of co-oligomers between αS and other proteins and the high potential of the resulting species to disrupt lipid membranes are important in the context of synucleinopathies, and in the development of therapeutics against the onset and progression of these diseases.

8.4 Methods

8.4.1 Modelling of self- and co-oligomer datasets

Having ascertained that the experiments detailed here have reached equilibrium with respect to oligomer formation, we analysed our results using two equilibrium statistical mechanical models: a previously described single-peptide oligomerization model outlined in Appendix E.2, [161] and a new 2-peptide model of co-oligomer formation outlined in Appendix F.2. The first model considers a solution of a single species of monomeric peptide that can reversibly self-associate to form linear chains of any length. The second model considers a solution containing 2 species of monomeric peptide that can reversibly associate to form both mixed and single-species chains. The equilibrium constants of these reactions can be related to the free energies of monomer addition to like and unlike peptides, or to self- and co-oligomerization. We work in the grand canonical ensemble, and we set the chemical potentials by insisting on conservation of peptide. Further, we consider concentrations not numbers, thus the only free thermodynamic parameter is the temperature. Thus, the systems described by both models are completely specified by the standard free energies, the peptide concentrations, and the temperature. We can therefore determine the standard free energies by least-squares fitting the apparent oligomer concentrations as a function of initial monomer concentrations.

We first fit the self-oligomer data to the single-species model (Appendix Eqs (E.15) and (E.16)) to obtain the self-oligomerization standard free energies. Note that the model contains a correction factor (calculated exactly analytically) to account for the fact that only the oligomers containing both donor and acceptor dye are detected by the experimental measurements, with donor-only and acceptor-only oligomers being invisible. We then enter these free energies into the 2-species model, which we fit to the 2-species oligomerization data to determine the co-oligomerization standard free energy. These data only measure the co-oligomer concentration so we fit only to the mixed oligomer component of the model (Appendix Eqs (F.4)-(F.12)).

Due to the complicated analytical fitting procedure, a formal parametric approach to error estimation is not suitable. Therefore, we prefer to use a nonparametric bootstrap approach. [229] Our data are effectively stratified by initial monomer concentration; therefore, we also stratify our sampling to better reproduce the structure of the data in our resamples. [230] Furthermore, we prefer a subsampling technique rather than full bootstrap, as subsampling is valid under much weaker conditions than full resampling. [231] Our full method is then to randomly sample one data point at each monomer concentration and fit the resulting sub-dataset to our model, recording the resultant ΔG°. This process is then repeated many times (we choose to do 500 repeats), generating a distribution of ΔG°s. An estimate of the standard error for ΔG° is then given by the standard deviation of this distribution.

Chapter 9

Statistical Mechanics of Oligomer Formation

Chapter Abstract

The aggregation of proteins into oligomers and linear fibrils finds widespread relevance in human biology, particularly in connection with the pathology of neurodegenerative disorders such as Alzheimer's and Parkinson's diseases. Even though oligomeric peptide aggregates are intimately connected with amyloid disease toxicity, their size distributions and morphologies remain poorly understood. We discuss here formation of globular oligomers via hydrophobic interactions. Such oligomers are expected to be micelle-like, although much smaller than typical micelles. We therefore generalize a long-wavelength equilibrium thermodynamical micelle model to small cluster sizes, finding that this explains Monte-Carlo simulations of oligomer formation well. We identify the controlling parameters of the model, which are closely related to simple quantities that may be fitted directly from experiment. We predict that globular oligomers are unlikely to form at equilibrium in many amyloid systems, but instead may form transiently in the early stages of aggregation. We finally contrast the globular model to a well-known model for linear oligomer formation, highlighting how the differing ensemble properties of linear and globular oligomers offers a potential strategy for identifying oligomer morphology from experimental measurements.

A natural question that arises following the studies on Aβ and αS oligomers presented in the previous chapters is to what extent linear oligomer models are appropriate for analysing oligomer populations. The research presented in this chapter seeks to answer this question, and investigates globular oligomers, which are expected to be important components of some oligomer populations. The work in this chapter is being prepared for submission to *Biophys J* under the title “Statistical mechanics of globular oligomer formation”. I am primarily responsible for all theory and writing. Anđela Šarić is responsible for the Monte-Carlo simulations, and Thomas C. T. Michaels assisted in the drafting and discussion. Appendix G contains supplemental theoretical results and details of the Monte-Carlo simulations.

9.1 Introduction

The process of linear self-assembly, whereby monomeric protein units aggregate spontaneously to form larger filamentous structures, is central to both normal and aberrant human biology and has been studied extensively over the past five decades [18, 20, 21, 30, 46, 50, 53, 89, 232]. Such a process is fundamental to the pathology of a number of widespread and deadly afflictions, most prominently protein misfolding diseases such as Alzheimer’s Disease (AD) and Parkinson’s Disease (PD) [1–4, 11, 233–236], Sickle-Cell Anaemia [20, 61, 237] and prion disorders [5, 238–241].

In many of these conditions, it has been shown that cell damage is caused mainly by small oligomeric species, rather than the large fibrillar aggregates that ultimately are formed. [8] Whilst most studies of filamentous protein aggregation have focussed on the formation of these fibrils, relatively little work targets the oligomers. This discrepancy arises from difficulties in detecting oligomers under physiological conditions, where they are present at picomolar concentrations. Recent advances in single molecule detection techniques have greatly facilitated the observation of peptide oligomer populations at low concentrations, and even allow approximate inference of oligomer size distributions at equilibrium via confocal two-color coincidence detection. [25, 26] Through these experimental advances, many peptide oligomer geometries have been reported in the literature, including discs, [157] rings, [194, 242] chains, [242] and spheres [242, 243]. Additionally several studies have investigated oligomer size distributions without determining their geometry [29, 244, 245]. Some of these geometries were only reported under unusual conditions [157]; and it is also possible that some are artefacts of the methods of preparation and imaging. Furthermore, measurements of apparent size distributions are currently highly approximate. Mathematical modelling of oligomer formation therefore has the potential to provide much-needed clarity on oligomer properties and morphology that experiments alone cannot currently provide. Developing such an understanding will have important implications for both the kinetics and the thermodynamics of amyloid protein aggregation; and likely also for understanding the mechanisms behind oligomer toxicity.

In this paper, we draw on micelle theory to present a statistical mechanical model for equilibrium globular oligomer formation. We investigate the key properties of globular oligomers using the model and how its parameters may be determined experimentally.

We finally compare the results to those for linear oligomers using a well-known model, and discuss how these geometries may be distinguished in practice.

9.2 Statistical Mechanics of Globular Amyloid Oligomers

Using the theoretical framework offered by the grand canonical ensemble (see Appendix G.1), a general expression for the concentration $f(j)$ of noninteracting oligomers of size j in a system at temperature T can be found, in terms of the internal oligomer partition function $q_{\text{int}}(j) = \exp(-\beta\Delta G^{\circ}_{\text{int}}(j))$ and the fundamental volume $v_0(j)$ associated with the translational partition function:

$$f(j) = \frac{1}{N_A\, v_0(j)}\, q_{\text{int}}(j)\, e^{\beta j \mu}, \tag{9.1}$$

where $\beta = 1/(k_B\, T)$ is the inverse temperature, N_A is Avogadro's number, and μ is the chemical potential of the monomers, which is set implicitly by the conservation-of-mass condition, $p = \sum_{j=1}^{\infty} j f(j)$, where p is the total peptide concentration. For notational consistency with previous kinetic and thermodynamic studies of oligomerization [112,161], we hereafter write $f(1) = m$. We further assume that $v_0(j) \equiv v_0$ is independent of oligomer size. From Eq. (9.1), it follows that to model equilibrium oligomer size distributions, we must identify the internal free energy change associated with oligomer formation, $\Delta G^{\circ}_{\text{int}}(j)$.

9.2.1 Internal free energy

Globular morphology can be expected on physical grounds to be common for oligomers assembled via hydrophobic interactions: only part of the surface of an amyloidogenic peptide is typically expected to be hydrophobic. If this part is contiguous and not too large [246], the peptide has amphiphilic character, and oligomers are likely to be finite and spherical, and micellar in nature. An analytical expression for spherical micelle free energy relative to monomers has been derived by Maibaum, Dinner and Chandler [247]:

$$\Delta G^{\circ}_{\text{int}}(j) = \alpha\, j^{2/3} + \Delta G_{\text{bulk}}\, j + h\, j^{5/3}/\beta, \tag{9.2}$$

where ΔG_{bulk} is the favourable free energy of transferring the hydrophobic "tail" of an amphiphile into a pure-tail phase; and $\alpha j^{2/3}$ is an energy penalty arising from surface tension between the bulk phase and the new "tail" phase. $hj^{5/3}$ is a connectivity-enforcing term that penalises the separation of head and tail groups that occurs if the micelle becomes too large whilst retaining its spherical geometry [248], originally computed by utilizing an analogy between micelle formation and charge repulsion [249]. The large-j penalty from steric clash between head groups can be shown to be insignificant in front of this term (see Appendix G.2). For simplicity we consider only uncharged micelles in the present work.

This model was developed for typical surfactant micelles with aggregation numbers $50 < j < 100$ and whose constituent amphiphiles have long hydrocarbon tail groups, for which the concept of a micelle surface tension is well-defined. It was intended that the surface tension be measured directly from experiments featuring a water phase and a phase consisting of oil molecules similar to the tail groups. For the small aggregation numbers ($5 < j < 15$) [113] typically seen in oligomers, however, the micelle core can hardly be considered a distinct phase separated from the bulk phase by a well-defined surface, whose nature would instead vary significantly with j. Moreover, measurements of the kind outlined above are not readily possible with amphiphilic peptides instead of hydrocarbon surfactants. We instead write the free energy of forming a micelle core as the sum of pairwise interactions between amphiphiles, yielding an overall free energy:

$$\Delta G^\circ_{\text{int}}(j) = j\,\frac{z_j}{2} G_b + h\,j^{5/3}/\beta, \tag{9.3}$$

where z_j is a size-dependent mean-field coordination number. For nonspecific interactions we expect this description to be reasonable [250]; even where internal structures lack fluidity bonding isomers are likely to only be relevant for $j = 3 - 4$.

To utilize this model we require a functional form for z_j that can reproduce the coordination numbers seen in oligomer ensembles. We know that $\lim_{j\to\infty} z_j$ exhibits perfect spherical scaling; at small j, however, deviations from this mean-field relation are possible. We nonetheless try the exact spherical form:

$$z_j = z_\infty(1 - \gamma\,j^{-1/3}). \tag{9.4}$$

Combined with the choice $\gamma = 1.21$ and $z_\infty = 12$ this reproduces known small Van der Waals cluster coordination numbers [250], e.g. $z_6 = 4$. Furthermore, this matches average coordination numbers measured from Monte-Carlo simulations of globular oligomers remarkably well (Fig. 9.2). Equating Eq. (9.2) and Eq. (9.3) yields $\gamma = -2\alpha/z_\infty G_b$; we have thus eliminated a model degree of freedom without resorting to the impractical experimental strategy previously recommended. We note that determining α experimentally is often unsuitable even for conventional micelles, being unable to reproduce Monte Carlo simulations of size distributions [251]. Our bond-counting approach instead makes clear the link between the surface tension and the bulk bonding free energy, which seems to be a more effective constraint than surface tension measurements.

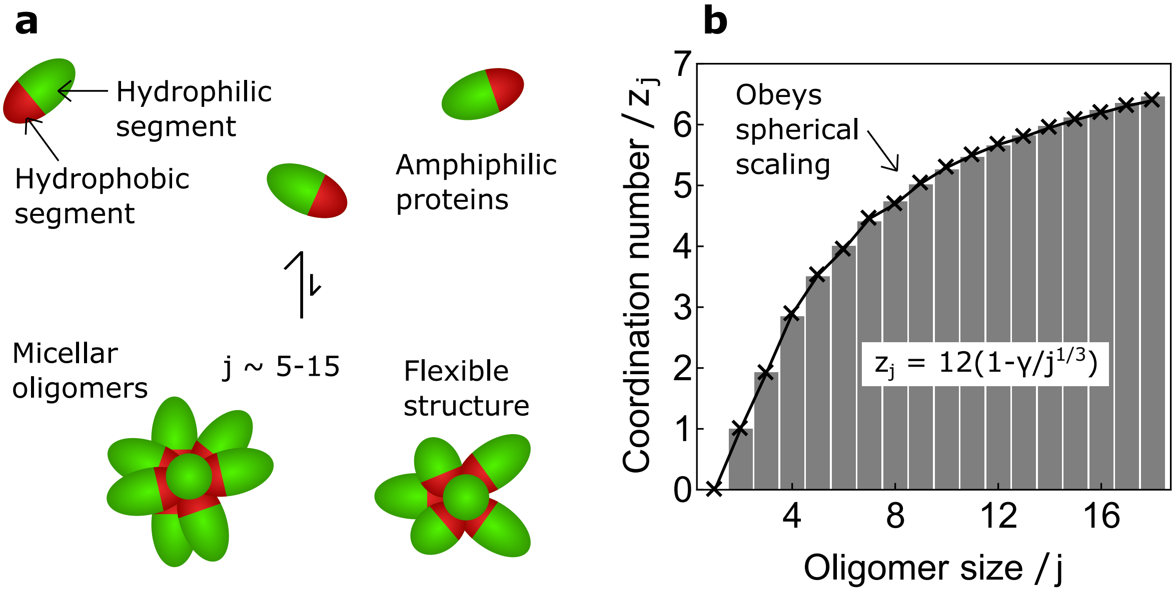

Figure 9.1: **a**: Schematic of cluster oligomer model. Peptides in the model have a non-bonding hydrophilic region and a hydrophobic region capable of non-directional bonding. Bonding interactions are counted pairwise. Typical protein oligomer aggregation numbers are expected to range from 5-15 monomers. **b**: We replace bulk and surface energy terms with a bonding energy term depending on an average spherical coordination number function (bars), and eliminate a parameter by matching known small cluster geometries [250] using the value $\gamma = 1.21$ and $z_\infty = 12$. This matches the results from Monte Carlo simulations (data points), reflecting the flexibility of oligomer structure. Minor deviations below the function values at larger sizes suggests a tendency to form spherocylindrical micelles; this effect has been observed previously [252].

9.2.2 Size distribution

Having determined an expression for the free energy of globular oligomerization, we are now in a position to determine the size distribution for globular oligomers. Using (9.3) in (9.1), we find:

$$f(j) = \frac{1}{N_A v_0} \exp\Big(-\beta j \, \frac{z_j}{2} G_b - h \, j^{5/3} + \beta j \mu\Big) \,. \tag{9.5}$$

The origin of this size distribution in micelle theory ensures a total peptide transition concentration p^*, above which the size distribution is peaked, and below which it is monotonically decreasing. The equilibrium monomer concentration m^* associated with the transition is approximately the critical micelle concentration (CMC).

A further key advantage of Eq. (9.5) over the Maibaum-Dinner-Chandler formulation is that we can more naturally interpret the properties of the size distribution in terms of its parameters. We expect the modal oligomer size above p^* to be controlled by the ratio $\beta G_b/h$: the trade-off between the bonding free energy, that increases the favourability of larger oligomers, and the connectivity term, that decreases it. Given a known average oligomer size, we then expect the polydispersity of the size distribution to be set by the magnitude of βG_b or h; the smaller their magnitude the flatter the internal free energy, and the smaller the difference in stability between oligomers of different sizes. The concentration units are then set by v_0.

We were able to globally fit our model successfully to our Monte-Carlo simulations of oligomer formation at two different total peptide concentrations (Fig. 9.2), yielding $\beta G_b/h = -3.62$, $h = 0.58$, and $v_0 = (11.8 \text{ nm})^3$. The h value is comparable to the range predicted by *ab initio* calculations [248]. Details of the simulations are given in Appendix G.3.

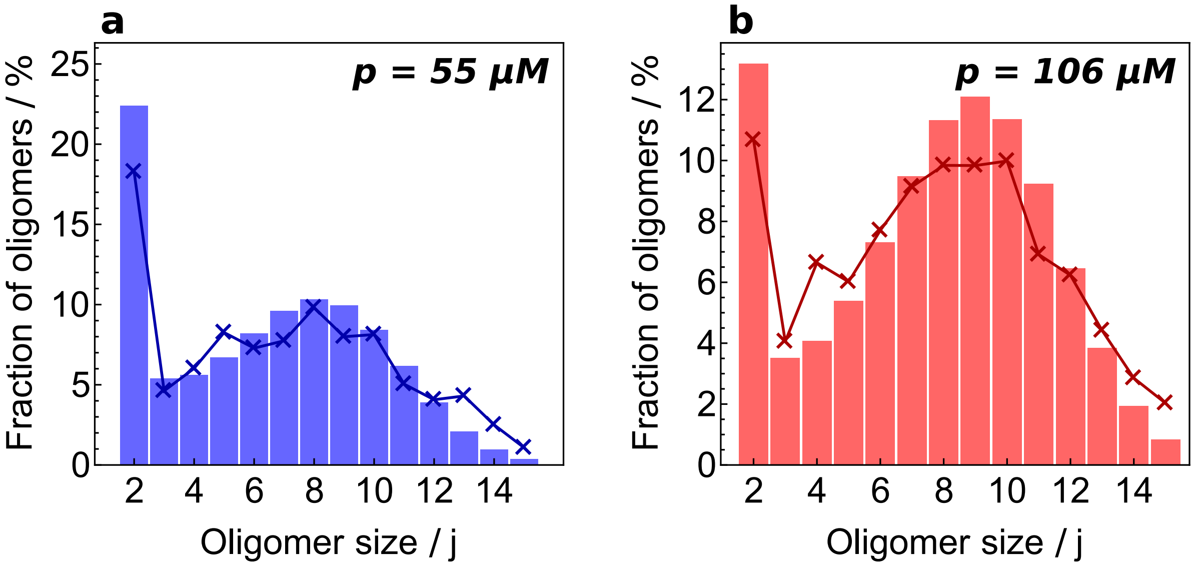

Figure 9.2: The analytical size distribution model (bars) can be globally fitted to coarse-grained numerical Monte Carlo results (data points) with reasonable accuracy (fitted parameters in Sec. 9.2.2).

9.3 Key Characteristics of Globular Amyloid Oligomers at Equilibrium

Having predicted how the model parameters will relate to the key properties of the system, we now seek explicit expressions for these properties. A key advantage of doing so is that these properties are more easily measurable from experiment than the full size distribution itself; by measuring them we can thus constrain the model and simulate the expected size distribution. Furthermore, we can rewrite Eq. (9.5) in a particularly convenient form in terms of these properties. We finally discuss how the size distribution changes in the presence of amyloid fibrils.

9.3.1 Average oligomer size

The most probable oligomer size above m^* is a key property often directly measurable from experiment, and is given by solving $df(j)/dj = 0$. A lower bound on this is given by the location of the point of inflection j^* of the size distribution at $m = m^*$ (Appendix G.4; first derived for micelles in Ref. 247):

$$j^* = -\frac{6\gamma\beta G_b}{5h}, \tag{9.6}$$

where we have set $z_\infty = 12$. Maibaum *et al.* [247] also derived an approximate expression for the most likely micellar size, by approximating the size distribution above the CMC as monodisperse at size $j = j'$. Combining Eq's (16) and (17) from Ref. 247 using our notation, the value of j' is given by solving:

$$\left[\frac{\partial \beta \Delta G/j}{\partial j}\right]_{j=j'} = \frac{z_\infty \gamma \beta G_b}{6} j'^{-4/3} + \frac{2h}{3} j'^{-1/3} = \frac{1}{j'^2} \ln\left(f_j N_A v_0\right), \tag{9.7}$$

where x is the fraction of peptide incorporated into micelles ($x = j' f_j / f_1$). The RHS was further approximated as zero to arrive at an analytical expression for j'. We note, however, that the RHS actually approaches zero from below as f_j increases towards $1/N_A v_0$ M. We further note that this limit is approximately the highest concentration possible for the system, corresponding to a volume fraction of unity, since v_0 is of the order of the monomer dimensions; therefore their expression is in fact an upper bound on the peak position. In particular, from the inequality:

$$\frac{z_\infty \gamma \beta G_b}{6} j_p^{-4/3} + \frac{2h}{3} j_p^{-1/3} < 0, \tag{9.8}$$

one obtains the following expression for the peak j_p when $z_\infty = 12$:

$$j_p < j' = -\frac{6\gamma\beta G_b}{2h}. \tag{9.9}$$

We expect the peak j_p to increase rapidly from j^*, as the lower bound j^* is an inflection point; and then to approach the upper bound j' asymptotically. Therefore, j_p will typically be closer to j' than to j^*. Indeed, in the current case the choice $j_p = -5\gamma\beta G_b/2h$ is in good agreement with our system (see Fig. 9.3). Crucially, both bounds show the expected scaling behaviour G_b/h.

Once a peak in the globular oligomer size distribution forms, we expect the mass of the system to shift rapidly into oligomers with increasing p, with a higher peak size, or $|G_b|/h$, corresponding to a faster shift. This is inferred from the fact that above $p = p^*$ the oligomer concentration scales initially as approximately p^{j_p}. Indeed, explicit calculations using parameters fitted from the Monte-Carlo simulations demonstrate that the total peptide concentration for which 50% of peptide, p_{50}, is in oligomers is only slightly higher than p^* (Fig. 9.3).

9.3.2 m^* and rescaling

Size distributions have units of concentration. Rescaling by an experimentally-measurable concentration should therefore dedimensionalize the model, and effectively replace one degree of freedom with another more accessible one. A natural concentration for globular oligomers is m^*, at which a new maximum first appears in the size distribution; this can be shown to be (see Appendix G.4):

$$m^* = \frac{1}{N_A v_0} \exp\left(6\beta\, G_b + (5h)^{1/3}\left(-6\gamma\beta\, G_b\right)^{2/3}\right) \tag{9.10a}$$

$$= \frac{1}{N_A v_0} \exp\left(\frac{z_{j^*}}{2}\beta\, G_b\right), \tag{9.10b}$$

where z_{j^*} is the coordination number at the point of inflection of the size distribution at $f(1) = m^*$, and where we have set $z_\infty = 12$. We can then identify the convenient dimensionless form for the size distribution:

$$\hat{f}(j) = \hat{f}(1)^j \exp\left(h\left[-\frac{\beta G_b}{h}\left(j\frac{z_j}{2} - (j-1)\frac{z_{j^*}}{2}\right) - j^{5/3}\right]\right), \tag{9.11}$$

where $\hat{f}(j) = f(j)/m^*$, and the exponential term is the size distribution at $\hat{f}(1) = 1$, which is monotonically decreasing apart from a point of inflection at j^*. We clearly see that at concentrations below $\hat{f}(1) = 1$, the size distribution is monotonic decreasing, and above m^* it has a peak.

Rescaling concentrations by m^* thus eliminates v_0 and collapses all possible size distributions onto a set of curves indexed only by G_b and h. To map to SI units we now require m^* instead of v_0, which may often be conveniently determined experimentally.

9.3.3 Polydispersity of the size distribution

An indicator of the polydispersity of the globular size distribution above m^* is given by its second derivative with respect to j at $j = j_p$ normalized by the peak height, which we label $\mathcal{C}$:

$$\mathcal{C} = \frac{1}{f(j)}\frac{d^2 f(j)}{dj^2}\bigg|_{j=j_p}. \tag{9.12}$$

Using Eq. (9.3) in Eq. (9.12), we find:

$$\mathcal{C} = \frac{1}{9\, j_p^{1/3}}\left[\frac{\gamma z_\infty}{j_p}(-\beta G_b) - 10\, h\right]. \tag{9.13}$$

Interestingly there is no explicit concentration dependence; thus j^* is always a point of inflection with zero curvature, and curvature is always negative at higher j since the positive first term is reduced more than the negative second term.

We see that the polydispersity of the distribution is thus controlled by the magnitude of G_b or h, as expected. Upon closer inspection, doubling $(-\beta G_b)$ reduces the curvature by $2^{1/3}$, whereas halving h reduces $\mathcal{C}$ by $2^{4/3}$; thus, the polydispersity is substantially more sensitive to h than to G_b. Halving both $|G_b|$ and h gives no change in j_p and thus halves $\mathcal{C}$.

An expected consequence of a more polydisperse size distribution is a greater range of peptide concentrations over which the transition from a dimer-dominated distribution to a peaked distribution occurs. Thus a greater $|G_b|$ or h should lead to a more rapid transition from one regime to the other (Fig. 9.3). The form of Eq. (9.11) ensures that, given a moderate average oligomer size, the distribution below m^* is highly dimer-dominated for all but the smallest h.

Our model considers only spherical oligomers; however, at sizes above j_p we expect spherocylindrical oligomers to become important [252], since it becomes favourable to reduce average coordination number to avoid head and tail group separation. We therefore expect our model to underestimate the polydispersity. Fitting our model to full size distributions will then only lead to lower bounds on the magnitudes of $|G_b|$ and h, and consequently an upper bound on v_0, rather than exact values. Indeed, fitting to the Monte-Carlo simulations yielded $v_0 = (11.8\text{ nm})^3$, which is slightly larger than expected given monomer dimensions of $(2\text{ nm})^3$. However, given only m^* and j^*, the model still has good predictive power apart from the concentration range immediately above $p = p^*$ (Fig. 9.3). Should greater accuracy for parameter magnitudes be required, h may in some cases be computed relatively simply analytically [247, 248].

9.3.4 Oligomer formation in the presence of fibrils

We have characterized oligomer formation in isolation; in practice, we are frequently interested in peptides that can form both oligomeric pre-nucleation clusters and macroscopic fibrillar aggregates, with different structures and therefore different chemistries. The coupling of these processes has important implications for oligomer formation.

Filament formation by monomeric peptides involves a phase transition, below which fibrils do not form; the total monomer concentration at which this occurs is known as the critical aggregation concentration (CAC). To show this, we first define $P(t)$ as the number concentration of filaments of any size, and k_+ and k_{off} as the size-independent filament elongation and depolymerization rate constants respectively. Following from earlier notation, the monomer concentration at time t is $m(t)$. The equilibrium concentration of monomers in filament forming conditions is then independent of initial monomer concentration:

$$2k_+\, m(\infty)\, P(\infty) = k_{\text{off}} P(\infty) \implies m(\infty) = \frac{k_{\text{off}}}{2k_+}. \tag{9.14}$$

We may then identify $m(\infty) \equiv m_{\text{CAC}}$: if the initial monomer concentration is lower than this, no macroscopic filament formation occurs; if it is higher, m_{CAC} becomes the equilibrium monomer concentration.

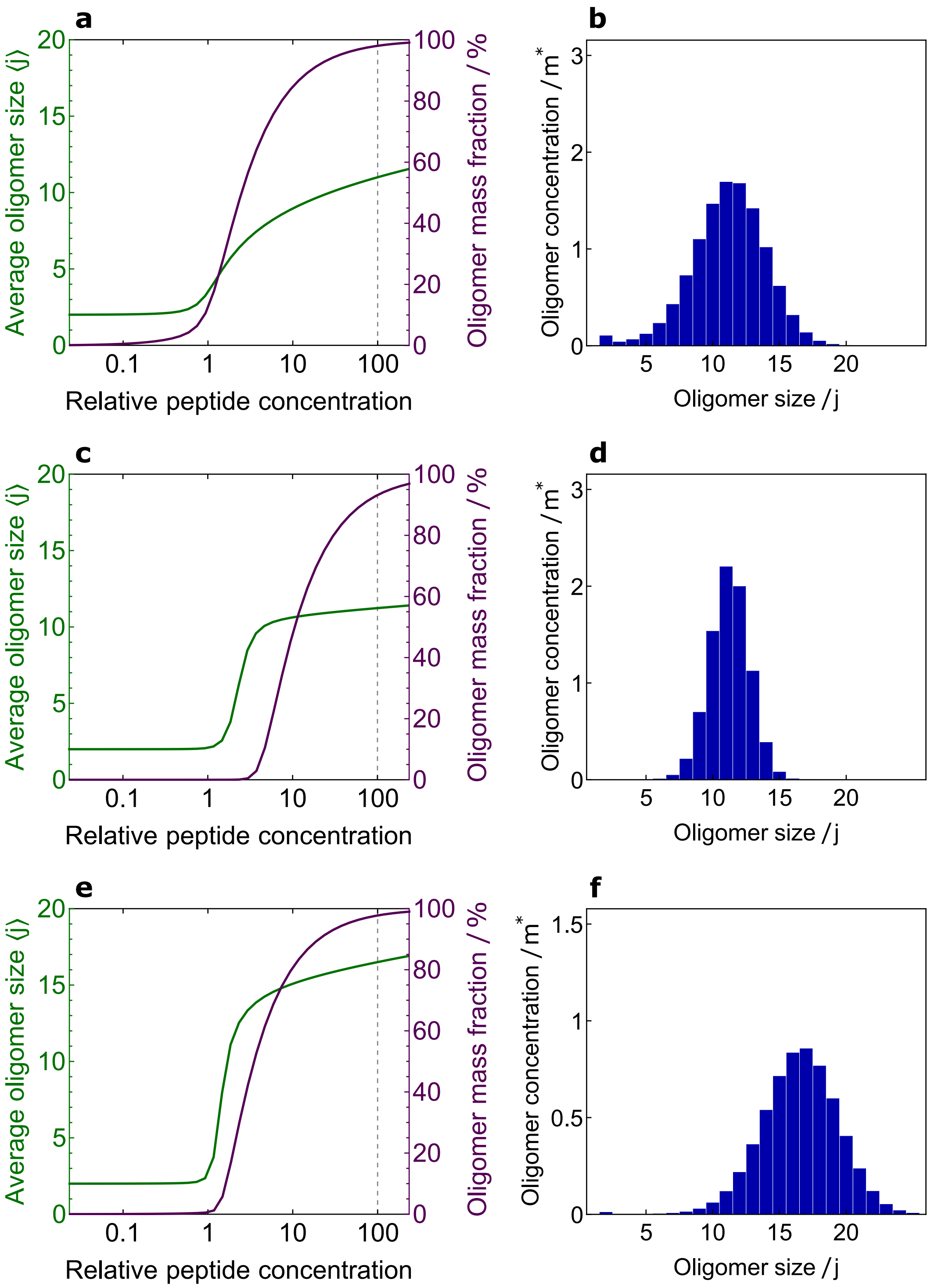

Figure 9.3: The key parameters controlling the globular model are the transition concentration m^* and the transition size j^*. **a**: Average size and mass fraction as a function of peptide concentration scaled by m^*; parameters fitted from Monte Carlo data (Sec. 9.2.2). Systems identical apart from v_0 collapse onto single curves of this form upon rescaling concentrations by m^*. **b**: Size distribution at scaled $p = 100$ (gridline on **a**). Tripling G_b and h does not affect j^*, but increases the sharpness of the transition from small to large size (**c**). By delaying the appearance of large oligomers, it also delays the onset but not the speed of the mass fraction transition relative to the size transition. These effects are achieved by the reduction in the polydispersity of the size distribution (**d**). Increasing G_b by 1.5× increases the average size by the same factor (**e**), without affecting the polydispersity significantly (**f**). The larger average size increases the rapidity of the oligomerization transition.

Equilibrium oligomer formation

Given that amyloid fibrils are expected to be more thermodynamically stable than prenucleation clusters, the analogue of the CMC for any globular oligomers will typically be higher than the CAC ($m^* > m_{\mathrm{CAC}}$). Globular oligomers will therefore never dominate over monomers at chemical equilibrium, no matter what the initial monomer concentration is. Furthermore, globular prenucleation clusters are unimportant compared to dimeric prenucleation clusters at equilibrium. This is borne out by studies of α-synuclein oligomers at equilibrium [ref ACSNano], and of equilibrium Aβ oligomer concentrations [161]; both of which demonstrate oligomers to be predominantly dimeric.

Transient oligomer formation

Although we have demonstrated that globular prenucleation clusters are unlikely to feature at equilibrium with fibrils, they are expected to form transiently at the much higher monomer concentrations visited during kinetic experiments. It is often the case that the timescale of oligomerization is faster than that of fibril formation, such that a pre-equilibrium between monomers and oligomers may be a reasonable approximation. In such experiments, our equilibrium model of globular oligomer formation is therefore expected to be useful for understanding observed size distributions of prenucleation clusters, as well as filament formation when the oligomers are on-pathway nucleation intermediates.

9.4 Contrasting Linear and Globular Oligomers

Here we derive or present the key characteristics of linear oligomer systems based on the well-known Oosawa model [253]. We explain how they differ from globular oligomers (Fig. 9.4) and how this may be taken advantage of to infer oligomer geometry from experiments.

9.4.1 Size distribution

In Sec. 9.2 we reduced the problem of finding a size distribution for an oligomerizing system to one of finding a suitable internal free energy form. In the case of linear oligomers, we have to a good approximation $\Delta G^{\circ}_{\mathrm{int}}(j) = (j-1)\varepsilon$, where ε is the size-independent nearest-neighbour binding energy. Thus, using (9.1) we recover Oosawa's result for the size distribution [253] (see [161] for further details):

$$f(j) = \frac{1}{N_A\, v_0}\, e^{-\beta\,\varepsilon\,(j-1)}\, e^{\beta\, j\, \mu}. \tag{9.15}$$

If we instead absorb v_0 into ε, identifying $\varepsilon = \Delta G^\circ_+$, the size-independent free energy of monomer addition, this may be further simplified to a single-parameter model:

$$f(j) = e^{-\beta\,\varepsilon\,(j-1)}\,e^{\beta\,j\,\mu}. \tag{9.16}$$

We therefore expect a monotonically-decreasing equilibrium oligomer size distribution for all initial monomer concentrations, with increasing initial monomer concentration has no effect beyond increasing the slope of the size distribution in log-space (Fig. 9.4(e)); this is confirmed by coarse-grained Monte-Carlo simulations of linear oligomer formation (Fig. G.1). This is qualitatively different to globular oligomers (Fig. 9.4(b)), which have a peaked size distribution above a critical peptide concentration. Although current experimental measurements of size distributions are highly approximate [28], this distinction is large enough that it may in certain circumstances be detected.

9.4.2 Critical concentration and rescaling

A critical concentration analogous to m^*, but for linear oligomers, is $m^\ddagger = e^{\beta\varepsilon}$; above this total peptide concentration, most peptide added becomes aggregated into more linear oligomers. Since size distributions have dimensions of concentration, rescaling by this concentration reduces the size distribution to the dimensionless form $\hat{f}(j) = \hat{f}(1)^j$, with $\hat{f}(1) = f(1)/m^\ddagger$ computable from the reduced form of the total peptide concentration $\hat{p}$ [253]:

$$\begin{aligned}\hat{p} = p/m^\ddagger &= \sum_{j=1}^{\infty} j\hat{f}(j) = \hat{f}(1)\frac{d}{d\hat{f}(1)}\sum_{j=1}^{\infty}\hat{f}(1)^j \\ &= \hat{f}(1)\frac{d}{d\hat{f}(1)}\left(\frac{\hat{f}(1)}{1-\hat{f}(1)}\right) = \frac{\hat{f}(1)}{1-\hat{f}(1)} + \frac{\hat{f}(1)^2}{(1-\hat{f}(1))^2} \\ &= \frac{\hat{f}(1)}{(1-\hat{f}(1))^2}.\end{aligned} \tag{9.17}$$

Additionally, the reduced total oligomer concentration is given as:

$$\hat{F} = \sum_{j=2}^{\infty}\hat{f}(j) = \sum_{j=2}^{\infty}\hat{f}(1)^j = \frac{\hat{f}(1)^2}{1-\hat{f}(1)}. \tag{9.18}$$

Thus, scaling concentrations by $m^\ddagger$ causes all key properties of linear oligomers to collapse onto single curves parameterized by $\hat{p}$; changing $\varepsilon \to \varepsilon + \delta$ is equivalent to changing the concentrations on the unscaled curves $p \to p\,e^{-\beta\delta}$ (Fig. 9.4(g)-(h)). Since $m^\ddagger$ does not correspond to a sharp transition, we expect that longer linear oligomeric species may be able to exist in equilibrium with fibrils at the end of an aggregation reaction, unlike in the globular case. This crucial difference may be taken advantage of in experiments to determine what oligomer morphologies may be present in an aggregating system.

9.4.3 Average size and polydispersity

The average linear oligomer size may be computed with greater ease using rescaled units. We find the mean oligomer size as:

$$\begin{aligned}\langle j \rangle &= \frac{1}{\hat{F}} \sum_{j=2}^{\infty} j\hat{f}(j) = \frac{\hat{p} - \hat{f}(1)}{\hat{F}} \\ &= \frac{1 - [1 - \hat{f}(1)]^2}{\hat{f}(1)[1 - \hat{f}(1)]} \\ &= \frac{2 - \hat{f}(1)}{1 - \hat{f}(1)},\end{aligned} \tag{9.19}$$

which, for large $\hat{p}$, approaches $\hat{p}^{1/2}$. Note this is unbounded, unlike globular oligomers which reach a well-defined maximum peak size. We expect as a consequence that the polydispersity of the linear size distribution is much greater at high concentrations than the globular size distribution, since there is no free energy penalty for large linear oligomers. The resultant qualitative differences in average size and polydispersity may be detected comparatively easily from appropriate approximate experimental measurements of the size distribution.

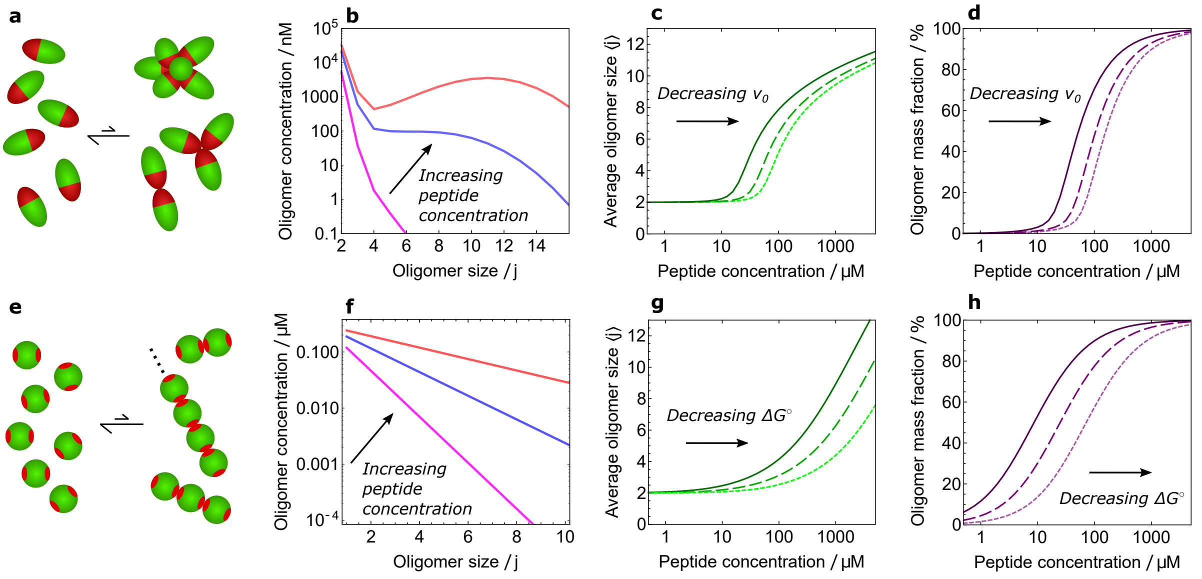

Figure 9.4: The differing properties of globular (**a-d**) and linear (**e-h**) oligomers. Whereas globular oligomers are expected to form from amphiphilic monomers (**a**), linear oligomers may be formed from monomers with 2 nonspecific bonding regions (**e**). Whereas globular oligomer size distributions become peaked at high enough total peptide concentration p (**b**), linear size distributions always decrease exponentially (**f**), simply falling off less rapidly at higher p. Average globular oligomer size (**c**) transitions rapidly from 2 to a well-defined intermediate value at higher concentrations; linear oligomer average size increases more gradually but without bound (**g**). As a consequence, linear oligomer mass fraction (**h**) also increases more gradually than globular (**d**). Decreasing v_0 has the effect of rescaling globular concentrations to higher values, without changing the phase-like properties (**c-d**; dashed lines); the equivalent parameter in the linear case is the sole model parameter $\beta\Delta G^\circ$ (**g-h**; dashed lines). Scaling linear peptide and oligomer concentrations by $m^\ddagger = e^{\beta\Delta G^\circ}$ therefore simplifies the model by dedimensionalizing it just as scaling globular system concentrations by m^* does.

9.5 Conclusions

In summary, we have adapted a model of micelle formation to be suitable for micelle-like globular oligomers with low aggregation number, and demonstrated that this can reproduce coarse-grained Monte-Carlo simulations of oligomer formation. We found that when total peptide concentration p is below a critical value p^*, the size distribution is monotonic decreasing. Increasing p above p^*, however, the distribution is soon dominated by a peak at higher aggregation number, and soon after most peptide rapidly becomes incorporated into oligomers.

We found that a particularly convenient non-dimensional form of the model could be obtained by scaling all concentrations by the fundamental monomer concentration m^* associated with p^*. This could be expressed in terms of the fundamental observables of the model, revealing the crucial scaling laws and simplifying the analysis of experimental data. We also developed bounds for the average globular oligomer size above the transition, finding that the globular size tends asymptotically towards a limiting value.

We have also shown that the presence of fibrils suppresses globular oligomer formation at equilibrium, such that they cannot be present at the end of a fibril aggregation reaction. Finally we contrasted the properties of globular oligomers with those of linear oligomers, highlighting the key distinctions that may be looked for in experiments designed to reveal oligomer morphology. The most notable are the lack of a peak in the linear size distribution, which is instead monotonic decreasing with aggregation number j, and the possibility for larger linear oligomers to coexist with fibrils at equilibrium.

Going forward, I would like to explicitly include the nonlinear spherocylindrical geometries that arise at larger sizes, to better match the experimentally-observed polydispersity. I would also like to extend the globular model to consider globular co-oligomers formed by multiple monomer types.

Part V

A Unified Theory of Linear Self Assembly

Chapter 10

Renormalization Group for Protein Filament Formation

Chapter Abstract

We describe a perturbative renormalization group theory approach to study the kinetics of protein filament self-assembly processes in a unified manner across multiple timescales. Using this approach, we find that, irrespective of the specific molecular details or experimental conditions, filamentous assembly systems display universal behavior in time. Moreover, we show that the renormalization group fixed points for filamentous self-assembly correspond to simple autocatalytic processes and that the diversity of behavior in these systems is determined solely by the reaction order for secondary nucleation, which labels all possible universality classes. We validate these predictions on experimental data of different filamentous proteins. These results establish the power of perturbative renormalization group in distilling the ultimately simple temporal behavior of complex protein aggregation systems, opening the door to the study the kinetics of general self-assembly phenomena in a unified fashion.

This thesis concludes with a powerful method for accurately solving the kinetics of biofilament formation that, although applicable to problems in oligomer formation kinetics (see Appendix D), is most notable for its generality and its application across the whole field of self-assembly. It was conceived and carried out in close collaboration with T. C. T. Michaels; for clarity of reading the chapter therefore retains his contributions. Results sections for which I do *not* wish to claim primary credit are thus labelled with a dagger (†). Whereas I am primarily responsible for development of the method, T. C. T. Michaels is primarily responsible for its use to interpret filament assembly in terms of universality classes and scaling behaviour. The chapter is substantially the same as a manuscript we recently submitted to *Phys Rev Lett*, on which we share joint first authorship.

10.1 Introduction

The formation of protein filaments is an ubiquitous example of a self-assembly phenomenon that has fundamental implications for biology [53, 254], medicine [8, 94, 255], and materials science [56, 256, 257]. For instance, biofilaments of actin and tubulin are key components of the cellular cytoskeleton, which is implicated in cell shape regulation and cell division [53, 254], while a particular class of protein filaments, known as amyloids, are associated with over 50 medical disorders, including Alzheimer's and Parkinson's diseases [8, 94, 255]. One of the most intriguing yet least understood aspects of protein filament formation is its generality, i.e. the fact that many different proteins, with unrelated sequence or fold, are able to self-assemble into such filamentous structures [3, 167]. A central question in this area is thus to establish whether certain aspects of the aggregation process are universal, i.e. are independent of the molecular specifics and are thus conserved across different protein systems.

Historically, renormalization group (RG) theory approaches [258] have emerged as the fundamental tool for explaining how microscopically different systems can display universal features at the macroscopic level in areas ranging from condensed matter physics to particle physics. Asymptotic analysis seeks to obtain the simplified intermediate asymptotic behaviour af a system of equations, for instance through the identification and solution of an amplitude equation, or the discovery of a slow manifold. Since asymptotic analysis involves identifying a reduced, lower-dimensional description of the behaviour of a physical system, we might expect disparate systems to display universal features asymptotically, just as in classical RG problems. In this context, Goldenfeld and coworkers [259–261] demonstrated a deep connection between RG theory and asymptotic analysis of singular perturbation problems, and developed a highly general perturbative RG approach to asymptotic analysis that was shown to be superior in accuracy to many standard methods. In this paper, we bring the power of perturbative RG theory to protein filament self-assembly kinetics. We find that the macroscopic behavior of protein filament formation kinetics is universal in time and that this universal behavior is controlled solely by the dependence of fibril self-replication on the supply of monomeric protein. These results, which we validate using experimental data on protein aggregation, establish perturbative RG theory as a key for understanding the kinetics of diverse filamentous protein self-assembly systems across multiple timescales and in a unified fashion.

10.2 Perturbative RG approach to protein filament self-assembly

10.2.1 Kinetic equations for protein filament formation†

Protein filament self-assembly is the result of an interplay between multiple molecular events (Fig. 1). Generally, protein filament assembly is initiated by a primary nucleation step, whereby monomers in solution come together spontaneously to form the smallest stable aggregate, followed by filament growth through elongation at one or both ends of the aggregates [53]. For many protein systems, including the aggregation of disease-related species such as Alzheimer's Amyloid-β peptide, aggregation is accelerated by secondary processes, that, unlike primary nucleation, depend on the actual fibril concentration and, hence, lead to an autocatalytic multiplication of the fibrillar structures; key examples of such secondary processes are filament fragmentation [21, 39, 46, 262], lateral branching [67, 88] and surface-catalyzed secondary nucleation [13, 20, 68, 263, 264]. The interplay between these different microscopic steps of aggregation can be captured by means of a master equation approach [13, 20, 21, 53, 60, 265], which tracks the time evolution of the population of aggregates of different sizes. In the context of protein aggregation, however, the experimentally accessible information about the aggregation process is encoded in the time evolution of three coarse-grained fields: the number and mass concentrations of fibrils, denoted with $P(t)$ and $M(t)$, and the free monomer concentration $m(t)$. The time evolution of these experimentally observable quantities, which correspond to the principal moments of the filament distribution, can be derived explicitly in terms of the underlying mechanisms of aggregation from the non-linear master equation describing the time evolution of the entire distribution of aggregate sizes [265]; for an aggregating system evolving through primary and secondary nucleation pathways, this procedure results in the following general set of kinetic equations [13, 20, 21, 53, 60, 265]:

$$\frac{dP(t)}{dt} = k_n m(t)^{n_c} + k_2 m(t)^{n_2}[m_{\text{tot}} - m(t)] \tag{10.1a}$$

$$\frac{dm(t)}{dt} = -2k_+ m(t)P(t) = -\frac{dM(t)}{dt}, \tag{10.1b}$$

where $m_{\text{tot}} = M(t) + m(t)$ is the conserved total monomer concentration. Equation (10.1a) describes the rate of formation of new fibrils through primary and secondary nucleation, with the rate constants for these processes being k_n and k_2; n_c, n_2 are the reaction orders of primary and secondary nucleation with respect to the free monomer and embrace several options: filament fragmentation, which is independent of monomer concentration, corresponds to $n_2 = 0$, lateral branching corresponds to a linear dependence on the monomer concentration ($n_2 = 1$), while surface-catalyzed secondary nucleation is usually described by higher reaction orders ($n_2 \geq 2$). It is important to note that

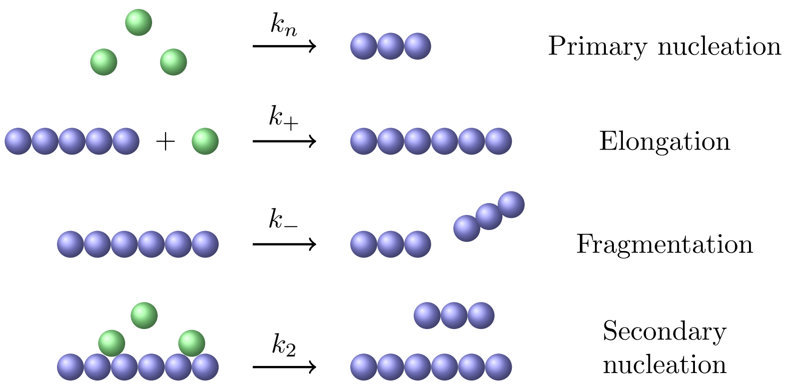

Figure 10.1: Schematic representation of the elementary steps of protein filament formation and definition of the associated rate constants.

both primary and secondary nucleation of new filaments are believed to be non-classical, multi-step processes, such that the reaction orders n_c and n_2 are not equal to the critical nucleus sizes of classical nucleation theory [113, 114, 263]. Equation (10.1b) describes the consumption of monomers and consequent build-up of aggregate mass through the elongation of existing filaments with rate constant k_+.

10.2.2 Perturbation expansion

Equations (10.1a) and (10.1b) do not admit exact solutions in general [20, 60], and we instead seek asymptotic solutions. The perturbative RG approach is intimately connected with asymptotic analysis [259–261]; as a first step, it is thus useful to recast (10.1) into a singular perturbation problem. To do so, we introduce $\mu(t) = m(t)/m_{\text{tot}}$, $t' = \kappa t$, and $\varepsilon = \lambda^2/2\kappa^2$, where $\lambda = \sqrt{2k_+k_nm_{\text{tot}}^{n_c}}$, and $\kappa = \sqrt{2k_+k_2m_{\text{tot}}^{n_2+1}}$ are the characteristic rates that describe aggregate proliferation via primary [53] and secondary nucleation [20,21,30], respectively. Equations (10.1a) and (10.1b) can then be combined and re-formulated as:

$$-\frac{d^2 \log \mu(t')}{dt'^2} = 2\varepsilon\, \mu(t')^{n_c} + \mu(t')^{n_2}\Big[1 - \mu(t')\Big]. \tag{10.2}$$

In many cases, secondary processes dominate over primary nucleation for the formation of new filaments, i.e. $\varepsilon \ll 1$. Typical values for ε are in fact $\varepsilon \sim 10^{-2}$ for the yeast prion Ure2p [266], $\varepsilon \sim 10^{-3}$ for sickle-cell hemoglobin (Hsb) [20], $\varepsilon \sim 10^{-5}$ for the Amyloid-β (Aβ) peptide [13], and $\varepsilon \sim 10^{-7}$ for the Islet Amyloid Polypeptide (IAPP) [68]. Consequently, a second-order perturbation series solution $\mu(t') = \mu^{(0)}(t') + \varepsilon\, \mu^{(1)}(t') + \varepsilon^2\mu^{(2)}(t') + \mathcal{O}(\varepsilon^3)$ can be found for Eq. (10.2) by applying the initial conditions $\mu(0) = 1$ and $d\mu(0)/dt' = 0$, to yield:

$$\mu(t') = 1 - \varepsilon e^{t'} + \frac{\varepsilon^2}{2c}e^{2t'} + \mathcal{R}, \tag{10.3}$$

where $c = 3/(2n_2 + 4)$ and $\mathcal{R}$ denotes other terms that are either of order ε^3 or vanish in comparison to the dominant terms at the respective order in ε for large t'. The perturba-

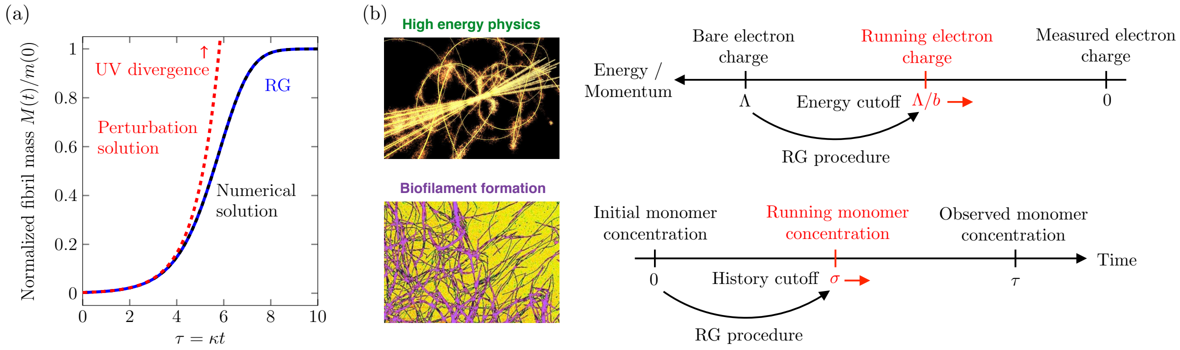

Figure 10.2: (a) Comparison of perturbative solution, Eq. (10.3) (dashed red), and RG solution, Eq. (10.12) (solid blue), with the numerical solution to Eqs. (10.1) (dashed black) for $n_2 = 0$ and $\varepsilon = \lambda^2/(2\kappa^2) = 3 \times 10^{-3}$. (b) Schematics of the RG approach to filamentous aggregation kinetics and analogy to RG procedure for high energy physics. A running time variable σ connects the observed monomer concentration at time $\tau = e^{\kappa t}$ with the known initial monomer concentration. The RG procedure, Eq. (10.9), describes how the initial monomer concentration, ρ, is renormalized as the renormalization scale σ runs between the initial time and τ.

tive solution Eq. (10.3) approximates the exact solution of (10.1) for early times, but it fails to capture the behavior at later times (Fig. 2(a)). This is due to exponentially divergent terms (UV divergence), a phenomenon originally pointed out by Ferrone *et al* [20] and that has challenged the derivation of accurate expressions for the aggregation kinetics; Eq. (10.3) is identical to the solution to the linearized form of Eqs. (10.1) obtained when $m(t) = m_{\text{tot}}$ [20].

10.2.3 Perturbative RG

To tame this divergence we impose timescale invariance on our perturbative solution. This amounts to re-summing the perturbation series for μ at the observation time t' such that its dependence on the time that has elapsed since μ was at its initial concentration (here $t' = 0$) is removed. We start by defining a new variable $\tau = e^{\kappa t}$ to simplify the mathematics. Following the conventional work-flow of perturbative RG [259–261], we then introduce an arbitrary past-time cut-off σ which we will vary between the initial time and the observation point τ, and write $\tau = (\tau - \sigma) + \sigma$ in Eq. (10.3) (Fig. 2(b)). This past-time cut-off is equivalent to the UV cut-off in conventional momentum-space RG; timescale coarse-graining by rewriting $\mu(\tau)$ in terms of $\mu(\tau - \sigma)$ and increasing σ is directly equivalent to coarse-graining by reducing the UV cut-off and integrating out the high-frequency degrees of freedom in momentum-space. Doing so, we obtain from (10.3):

$$\mu(\tau - \sigma, \sigma) = 1 - \varepsilon(\tau - \sigma) + \frac{\varepsilon^2}{2c}(\tau - \sigma)^2 + \varepsilon\sigma + \frac{\varepsilon^2}{2c}\left[\sigma^2 + (\tau - \sigma)\sigma\right] + \mathcal{R}. \tag{10.4}$$

The next step is to renormalize μ by multiplying Eq. (10.4) by a renormalization constant $\rho(\sigma)$, whose second order perturbation expansion in ε is:

$$\rho(\sigma) = \rho_0(\sigma) + \varepsilon\delta\rho_1(\sigma) + \varepsilon^2\delta\rho_2(\sigma). \tag{10.5}$$

$\delta\rho_1(\sigma)$ and $\delta\rho_2(\sigma)$ are counter terms, chosen to absorb the UV divergent terms in σ at the respective orders in ε [259–261]. The constant initial monomer concentration is thus effectively replaced by a running coupling which evolves with the RG scale. This yields:

$$\delta\rho_1(\sigma) = \sigma\rho_0(\sigma), \quad \delta\rho_2(\sigma) = \sigma^2\rho_0(\sigma)\left(1 - \frac{1}{2c}\right) \tag{10.6}$$

and we arrive at the following renormalized second-order expansion:

$$\mu'(\tau-\sigma,\sigma) = \rho_0(\sigma)\Big\{1 - \varepsilon\,(\tau-\sigma) + \varepsilon^2\left[\frac{\tau^2-\sigma^2}{2c} - \sigma(\tau-\sigma)\right] + \mathcal{R}\Big\}. \tag{10.7}$$

The renormalized solution, however, cannot depend on σ once the observation scale τ is reached. We thus require [259–261]:

$$\left.\frac{\partial\mu(\tau,\sigma)}{\partial\sigma}\right|_{\sigma=\tau} = 0. \tag{10.8}$$

This condition gives the perturbative RG equation, which, to second order in ε, reads:

$$\frac{\partial\rho_0(\tau)}{\partial\tau} = -\varepsilon\left(1 - \frac{\varepsilon\tau}{\theta}\right)\rho_0(\tau) + \mathcal{O}(\varepsilon^3) \tag{10.9}$$

where we have introduced the parameter:

$$\theta = \frac{c}{1-c} = \frac{3}{2n_2+1}. \tag{10.10}$$

This procedure is entirely analogous to performing a perturbative momentum-space RG calculation in statistical physics or quantum field theory; in this analogy, the electron charge or mass is replaced by the initial monomer concentration ρ, and the RG procedure yields renormalized values for this quantity at different time scales [260]. The RG equation describes a fixed point; here, it is a similarity solution that no longer diverges in time. The critical points of conventional RG are equivalent to bifurcation points in parameter space in this dynamical RG approach, and can be interpreted similarly: minute fluctuations in the system parameters at these points give rise to qualitative changes in the system properties. The flow of the renormalization constant, or “initial condition”, with respect to the initial time is precisely the long-wavelength motion we seek.

In (10.9), we recognize the expansion of the function $1/(1+\varepsilon\tau/\theta) = 1 - \varepsilon\tau/\theta + \mathcal{O}(\varepsilon^2)$, such that the solution to the second order perturbative RG equation (10.9) can be found to be:

$$\rho_0(\tau) = \left[1 + \frac{\varepsilon\tau}{\theta}\right]^{-\theta}. \tag{10.11}$$

Finally, substituting $\rho_0(\sigma)$ into Eq. (10.7), increasing σ to τ, and rewriting the result in terms of the original dimensional parameters, we find the following renormalized solution for the total aggregate mass:

$$\frac{M(t)}{m(0)} = 1 - \left[1 + \frac{\lambda^2}{2\kappa^2\theta}e^{\kappa t}\right]^{-\theta}. \tag{10.12}$$

A closed-form solution for the aggregate number concentration, $P(t)$, can be obtained from Eq. (10.1b) by simple differentiation, $P(t) = -1/[2k_+ m(t)]dm(t)/dt$, as:

$$\frac{P(t)}{P(\infty)} = \left[1 + \frac{2\kappa^2\theta}{\lambda^2}e^{-\kappa t}\right]^{-1}, \tag{10.13}$$

where $P(\infty) = \kappa\theta/(2k_+)$. Figure 2(a) shows a comparison between the RG solution, Eq. (10.12), and the numerical solution of Eqs. (10.1). By "removing" the UV divergence, the second-order perturbative RG solution succeeds in providing a highly accurate approximate description of aggregation kinetics when, as is usually the case, $\varepsilon \ll 1$ (since then $\varepsilon e^{-t'} \ll 1$, and the terms dropped from the perturbation series in Eq. (10.3) can justifiably be neglected). This method fails only when secondary growth processes are not present, such as in the growth of actin networks without branching; in this case, however, the kinetic equations admit an exact solution [53].

10.3 Universality of filamentous protein self-assembling systems

10.3.1 Self-similarity of protein aggregation

We now discuss some key consequences of our RG approach to protein aggregation. The first key prediction is that the time course of fibril mass formation has self-similar form:

$$\frac{M(t)}{m(0)} = \Phi_\theta\left(\frac{\lambda^2 e^{\kappa t}}{2\kappa^2}\right), \tag{10.14}$$

where

$$\Phi_\theta(x) = 1 - \left(1 + \frac{x}{\theta}\right)^{-\theta}, \quad \theta = \frac{3}{2n_2 + 1}. \tag{10.15}$$

This means that all instances of filamentous growth kinetics dominated by secondary processes are controlled by a single effective variable $x = \lambda^2 e^{\kappa t}/(2\kappa^2)$ containing all dimensional parameters. Therefore, given the same value for the reaction order of secondary

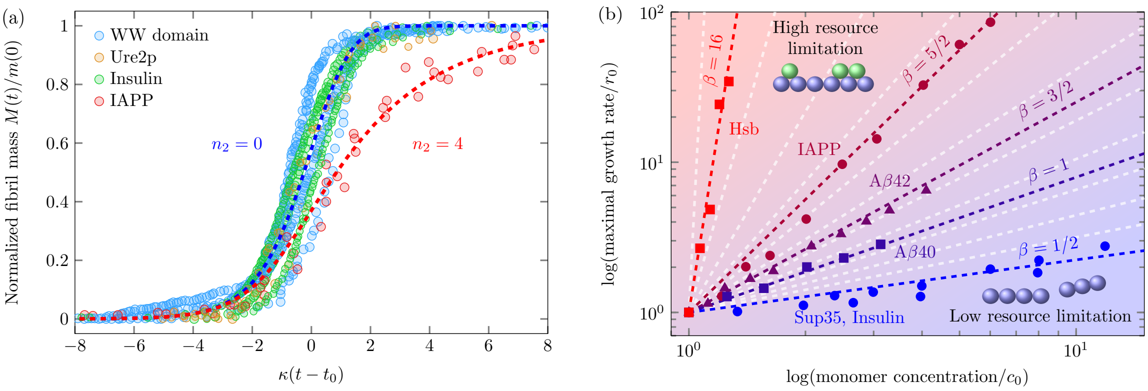

Figure 10.3: **Universality of protein filament formation kinetics.** (a) Experimental data of filament mass formation at different initial monomer concentrations for the WW domain (50, 100, 200, 500μM) [267], the yeast prion Ure2p (20, 25, 38μM) [266] and insulin (2, 4, 6μM) [21] collapse onto a single universal curve upon appropriate rescaling of the time coordinate, where $t_0 = -1/\kappa \log(\varepsilon)$. The universal curve (blue dashed line) is Eq. (10.14) with $n_2 = 0$. Upon rescaling, the aggregation kinetics data for IAPP [68] recorded at 3 different initial monomer concentrations (700, 800, 1000μM) collapse onto a different universal curve, Eq. (10.14) with $n_2 = 4$ (red dashed line). (b) Classification of different filamentous protein systems into universality classes on the basis of the scaling exponent β of maximal growth rate. In this plot, universality classes corresponding to a different reaction order for secondary nucleation, n_2, are described by straight lines with slope $\beta = (n_2 + 1)/2$ (shaded dashed lines). Data are shown for Sup35 and insulin ($n_2 = 0$) [21], Aβ40 ($n_2 = 1$) [14], Aβ42 ($n_2 = 2$), IAPP ($n_2 = 4$) [68], and sickle-cell hemoglobin ($n_2 = 31$) [20]. Panel (b) provided by T. C. T. Michaels.

nucleation, n_2, all realizations of protein aggregation become identical upon an appropriate rescaling of the time variable, irrespective of the specific experimental conditions or the specific molecular details of the protein system under consideration. This arises since all such realizations are in the same basin of attraction of the fixed point indexed by this particular value of n_2, and thus fall into the same universality class, regardless of their individual values for n_c, k_2, etc. In Fig. 3(a), we have tested this prediction under different conditions on a range of unrelated protein aggregation systems, ranging from infectious prion systems to amyloidogenic proteins: the WW domain [267], the yeast prion Ure2p [266], bovine insulin [21] and IAPP [68]. We find that, upon appropriate rescaling, experimental aggregation data recorded at different initial monomer concentrations for WW, Ure2p and insulin collapse onto a single universal curve described by Eq. (10.14) with $n_2 = 0$ and corresponding to systems dominated by filament fragmentation. The aggregation of IAPP is driven by surface catalyzed secondary nucleation with $n_2 = 4$ and thus, as expected, experimental data of IAPP fibril formation at varying initial monomer concentrations collapse but onto a different universal curve.

10.3.2 Scaling behaviour†

RG also provides new insights into the scaling behaviour of protein aggregation [21,30,68]. Indeed, a direct implication of RG is not only that filamentous assembly displays scaling behaviour, but, most importantly, that the characteristic scaling exponents must depend solely on the universality class. This prediction is verified directly for the scaling of the reaction half-time $t_{1/2}$, which, from (10.14), is found to scale as $t_{1/2} \propto m_{\text{tot}}^{\gamma}$, where exponent is $\gamma = -(n_2 + 1)/2$. Another important example is the scaling of the maximal growth rate, $r_{\max} = \max[dM(t)/dt]$, with initial monomer concentration. In this case, $r_{\max} \propto m_{\text{tot}}^{\beta}$ where the scaling exponent $\beta = (n_2 + 1)/2$ is, once again, solely dependent on n_2.

10.3.3 Classification of protein aggregation systems into universality classes†

Since elements of the same universality class share the same scaling behavior, scaling exponents are the most direct way for classifying different protein aggregation systems into universality classes using experimental data, as shown in Fig. 3(b). Notably, the different universality classes that we find for protein aggregation can be related to simple autocatalytic processes. Indeed, the universal curves (10.14) are solutions to the RG equation which is found to be equivalent to the generalized logistic equation $dM(t)/dt = \theta\kappa\Big(1 - [M(t)/K]^{1/\theta}\Big)M(t)$ with $K = m_{\text{tot}}$. Generalized logistic equations represent the simplest description of autocatalytic growth with finite resources and, thus, emerge most commonly in population dynamics, where K is known as the carrying capacity, i.e. the maximum population size that the environment can sustain indefinitely, given the available energy resources (e.g. in form of nutrients).

Since, for protein aggregation, K is directly related to the chemical potential of monomers $k_B T \log(m_{\text{tot}})$, i.e. the thermodynamic driving force for self-assembly, we may interpret monomer concentration as the available energy resources, and we can interpret the reaction order n_2, which labels our different universality classes, as describing the dependence of fibril proliferation on these resources. Protein systems with $n_2 < 1$, such as insulin or the prion protein Sup35, have small $1/\theta$ and are characterized by a weak dependence of proliferation on the available resources, such that their growth curves respond only gently to changes in the monomer chemical potential. In the limiting case of fragmentation ($n_2 = 0$), there is no resource-dependence for the secondary process responsible for autocatalysis, and resource dependence is almost solely due to fibril elongation. By contrast, the assembly of protein systems with $n_2 > 1$ and large $1/\theta$, such as Aβ42 or sickle-cell hemoglobin, is strongly dependent on the available resources, such that the approach of $M(t)$ to the carrying capacity is much more gradual as the remaining resources dwindle. The growth of systems with $n_2 = 1$, such as Aβ40 in the concentration range 5-10 μM, is equally as limited by resource depletion as it is accelerated by the accumulation of substrate the resources are converted into, and is described by a logistic universal function.

10.4 Summary and outlook[†]

We have discussed the use of perturbative RG approaches for protein filament formation. Using this approach, we have found that the kinetics of these phenomena display universal behavior and that the resulting universality classes correspond to simple autocatalytic processes. Looking forward, our work raises the natural question of whether, using RG, the kinetics of general self-assembly systems could also be classified into universality classes similar to those found for filamentous systems. In this sense, the RG approach presented in this work could represent an important starting point for constructing a unified dynamical theory of self-assembly.

I believe this project also motivates and applies dynamical RG with greater clarity than was achieved in the seminal works in the field. I have since prepared a follow-up paper that applies these methods to arrive at a universal theory of saturated biological growth kinetics.

Part VI

Closing Remarks

Chapter 11

Summary & Conclusions

11.1 Impact of work

This PhD has been highly productive, with seventeen papers relating to my doctoral studies published or under review at the time of writing. Several more manuscripts prepared during the final year of my studies are likely to be published in 2019. Although most of my first- and second-authorship works have been published in the past year, some of these have already begun to have an influence on their respective fields.

My and Thomas C. T. Michaels' work on stochastic aggregation kinetics, presented in Chapter 3, is already inspiring experimental protocol development. Scientists in our research group and in others are developing microfluidic platforms to perform massively parallel protein aggregation experiments encapsulated in microdroplets, in part to better understand primary nucleation. Our model provides an easy method for extracting accurate nucleation rates from data analysis of such experiments. In combination with these new protocols it will be possible to study the primary nucleation process in isolation and in much greater detail than hitherto possible. Even where microfluidic experiments are employed for purposes other than studying primary nucleation, it will still be important to account for stochastic small-volume effects. Previous modelling approaches have only been able to do so in the limit where volumes are so small that only single nucleation events occur. Our model, on the other hand, can describe stochastic effects at any volume, including in intermediate volumes where smaller stochastic fluctuations, caused by the interplay of a small number of nucleation events, are observed.

A large portion of my doctoral work has focused on the kinetics of amyloid formation via oligomeric intermediates; this work has resulted in some of my most highly-cited papers, including that presented in Chapter 5. It has helped spur interest in kinetic modelling of this phenomenon in hitherto unstudied protein systems and under hitherto uninvestigated conditions. Oligomers are formed by a wide range of amyloidogenic proteins; collating the results from prior kinetic studies has afforded insight into the commonalities and differences between these species (see Chapter 6). This brings us closer to an understanding of the reasons behind the universal occurrence of oligomers, and of their role in filament formation. This work is likely to be important for developing more successful strategies for the design of drugs targeting neurodegenerative diseases. The introduction of aliquots from oligomerizing systems into cell toxicity assays will, in combination with modelling, allow us to identify harmful sub-populations of oligomers and harmful reaction conditions; the same modelling can then in principle be used to guide drug design to suppress toxic species and toxic reactions.

The early studies into co-oligomer formation presented in Chapters 7 and 8 demonstrate how the relative stability of co-oligomers compared to self-oligomers may be calculated from theoretical modelling, and how the likely concentrations of different co-oligomers under biological conditions can be predicted. Furthermore, they demonstrate the existence and importance of co-oligomers consisting of key amyloidogenic proteins present in the brain and cerebro-spinal fluid. Given the highly heterogeneous nature of biological environments, the observation that co-oligomers can form in significant quantities is an important one. It also provides a starting point for investigations into possible causative links between the aggregation of different amyloidogenic proteins in both the same and different diseases. In Chapter 8, it is also demonstrated that combining this analysis with experiments into oligomer toxicity permits the relative toxicity of different possible self- and co-oligomers to be calculated. This admits the exciting possibility that development of future treatments for neurodegenerative diseases can focus solely on suppressing specific oligomers that are most responsible for disease symptoms.

Finally, the theory of linear coaggregation presented in Chapter 4 is one of the first successful forays into a largely unexplored field. It is a fundamental advance in its own right, and has the potential to additionally provide significant advances in our understanding of amyloid diseases going forward. The impact it ultimately has may largely depend on the extent to which it is used in future collaborative studies into mixed aggregation with experimentalists.

11.2 Outlook for the field

The outlook for the field of oligomer kinetics is bright, with a physico-chemical understanding of amyloid oligomers only just beginning to fall into place. The kinetic theory of amyloid oligomers presented in this thesis will probably continue to be heavily employed as new systems are studied. Relevant experimental techniques continue to be invented and improved at a rapid rate, which will enable a much more detailed investigation of the multiple oligomeric species that co-exist during aggregation reactions. This more detailed view of the oligomer reaction network will necessitate significant further development of theoretical models in order that the physical and chemical properties of these species can be fully understood. To this end, the models in this thesis are likely to provide a starting point for future generalization.

The statistical mechanical theory of globular oligomer thermodynamics presented in this thesis provides a firm theoretical basis for future experimental investigations into oligomer geometry, size distributions and heterogeneity, an area which has been insufficiently studied to date. It also provides a starting point for future development of more general theory. In particular, generalizing this theory to account for globular co-oligomer formation could be highly useful for elucidating interactions between different amyloidogenic peptides, complementing existing theories of linear co-oligomer formation.

The method for solving kinetic models presented in the final chapter of this thesis is likely to be especially influential on the field of linear self-assembly, being capable of yielding solutions of greater accuracy than any methods previously employed. More important than the accuracy of this method, however, is its generality, with a huge range of problems being amenable to this approach. Moreover, all solutions follow the same simple architecture, which is especially advantageous for gaining physical insight into these systems and their differences. This kind of insight is the primary motivation for seeking analytical solutions to such problems. This method therefore has great potential to supersede existing theoretical methods in the field, and will likely be used to replace some of the less accurate existing models as well as to model new systems going forward.

One major aim of my future research will be to make significant inroads into the theory of mixed self-assembly, and its real-world application. Much of the work in this thesis will find ready application in this emerging field, and many of the homomolecular models will likely be generalized to account for heteromolecular self-assembly. As an example of the latter, although it seems likely that my work on the role of oligomers in amyloid fibril formation will continue, given the current high interest in the field, it will take a less prominent role in my research going forward as the models developed in my doctoral studies are already versatile and highly capable. The focus will likely shift onto heterogeneity, both in terms of co-oligomer kinetics and thermodynamics, and in terms of structural and size heterogeneity in oligomer populations.

Part VII

Appendices

Appendix A

Studying Primary Nucleation Using Confinement

This appendix contains key proofs and derivations supporting the material in Chapter 3. It is largely based on the Supplementary Information for the paper "Fluctuations in the Kinetics of Linear Protein Self-Assembly" Those derivations for which I am *not* personally responsible are marked with a dagger (†).

A.1 Linear correlation for number and mass of aggregates in the presence of secondary pathways

In this appendix, we provide a mathematical justification for the existence of a linear correlation for the random variables n and x at times $t \gg \kappa^{-1}$. We employ two equivalent methods to prove that n and x are linearly correlated:

- *Moment approach*: We show that for $t \gg \kappa^{-1}$ and for any $k \geq 1$ the following relationship for the moments of the PDF holds

  $$\frac{\langle x^k \rangle}{\langle n^k \rangle} = \xi^k, \tag{A.1}$$

  where ξ is some constant that depends on environmental conditions. Therefore, a linear relationship exists between the moments $\langle x^k \rangle$ and $\langle n^k \rangle$ of any order, indicating that the random variables x and n are not independent but obey a linear relationship of the form $x = \xi n$.

- *Pearson's correlation approach*: We show that the Pearson's correlation coefficient for n and x, defined by

  $$\rho_{n,x} = \frac{\langle nx \rangle - \langle n \rangle \langle x \rangle}{\sqrt{(\langle n^2 \rangle - \langle n \rangle^2)(\langle x^2 \rangle - \langle x \rangle^2)}}, \tag{A.2}$$

is equal to 1 for $t \gg \kappa^{-1}$. Because the Pearson's correlation coefficient is an indicator of the degree of linear dependence between random variables, this result verifies the existence of an exact linear functional relationship between n and x.

In the following, we describe the first approach; a proof of the linear correlation between n and x using Pearson's correlation coefficient is given in Sec. A.2.

A.1.1 Examining the 1st moments

The full stochastic master equation (Eq. (3.1) of Chapter 3) can be manipulated to give differential equations for the 1st moments $\langle x\rangle$ and $\langle n\rangle$ of the PDF. We multiply the master equation by x and n respectively and sum over all system configurations (n, x) yielding the following set of coupled ordinary differential equations (ODEs) for the 1st moments

$$\begin{aligned}\frac{d\langle n\rangle}{dt} &= \alpha_1 + \alpha_2\langle x\rangle \\ \frac{d\langle x\rangle}{dt} &= \alpha_1 x_c + \mu\langle n\rangle + \alpha_2 x_2\langle x\rangle.\end{aligned} \tag{A.3}$$

Note that with the definitions $P(t) = \frac{\langle n\rangle}{N_A V}$ and $M(t) = \frac{\langle x\rangle}{N_A V}$ Eqs. (A.3) agree with the deterministic equations for filament formation [21, 30, 48, 58]. Assuming the initial conditions $\langle n\rangle = \langle x\rangle = 0$ (no aggregates present initially), the solution to Eq. (A.3) is

$$\langle n\rangle = \frac{\alpha_2 A}{\kappa_+}e^{\kappa_+ t} + \frac{\alpha_1 - \alpha_2 A}{\kappa_-}e^{\kappa_- t} - \frac{\alpha_2 A}{\kappa_+} - \frac{\alpha_1 - \alpha_2 A}{\kappa_-} \tag{A.4}$$

$$\langle x\rangle = Ae^{\kappa_+ t} + (\alpha_1/\alpha_2 - A)e^{\kappa_- t} - \frac{\alpha_1}{\alpha_2}, \tag{A.5}$$

where

$$\begin{aligned}A &= \frac{\alpha_1\kappa_+\kappa_-(x_c - x_2 - \mu/\kappa_-)}{\mu\alpha_2(\kappa_- - \kappa_+)} \\ \kappa_+ &= \frac{\alpha_2 x_2}{2} + \sqrt{\mu\alpha_2 + \frac{\alpha_2^2 x_2^2}{4}} \\ \kappa_- &= \frac{\alpha_2 x_2}{2} - \sqrt{\mu\alpha_2 + \frac{\alpha_2^2 x_2^2}{4}}.\end{aligned} \tag{A.6}$$

We note that for $t \gg \kappa^{-1}$ the exponential growing terms in Eq. (A.4) dominate over the exponential decaying and constant terms, such that the following relationship holds

$$\frac{\langle x\rangle}{\langle n\rangle} = \frac{\kappa_+}{\alpha_2}, \quad t \gg \kappa^{-1}. \tag{A.7}$$

Equation (A.7) shows that for $t \gg \kappa^{-1}$ the average number and mass of aggregates are directly proportional to each other. This result that has been reported previously in the literature [30]. We note that Eq. (A.7) is a necessary but not sufficient condition for the existence of a linear correlation between the random variables n and x. In general, a linear correlation between the random variables n and x requires the relationship

$$\frac{\langle x^k \rangle}{\langle n^k \rangle} = \left(\frac{\kappa_+}{\alpha_2} \right)^k, \quad t \gg \kappa^{-1}. \tag{A.8}$$

being satisfied by all moments $\langle n^k \rangle$, $\langle x^k \rangle$ of order $k \geq 1$.

A.1.2 Examining the 2nd moments

By a similar method, equations for the 2nd moments of the PDF can be derived from the master equation as

$$\begin{aligned}
\frac{d\langle n^2 \rangle}{dt} &= 2\alpha_2 \langle nx \rangle + 2\alpha_1 \langle n \rangle + \alpha_2 \langle x \rangle + \alpha_1 \\
\frac{d\langle x^2 \rangle}{dt} &= 2\mu \langle nx \rangle + 2\alpha_2 x_2 \langle x^2 \rangle + \mu \langle n \rangle + (2\alpha_1 x_c + \alpha_2 x_2^2) \langle x \rangle + \alpha_1 x_c^2 \\
\frac{d\langle nx \rangle}{dt} &= \mu \langle n^2 \rangle + \alpha_2 x_2 \langle nx \rangle + \alpha_2 \langle x^2 \rangle + \alpha_1 x_c \langle n \rangle + (\alpha_2 x_2 + \alpha_1) \langle x \rangle + \alpha_1 x_c.
\end{aligned} \tag{A.9}$$

These are coupled inhomogeneous 1st-order ODEs, which can be most conveniently written in matrix form as

$$\frac{d}{dt} \mathbf{x} = \mathbf{A} \mathbf{x} + \mathbf{g} \tag{A.10}$$

or, explicitly,

$$\frac{d}{dt} \begin{pmatrix} \langle n^2 \rangle \\ \langle nx \rangle \\ \langle x^2 \rangle \end{pmatrix} = \begin{pmatrix} 0 & 2\alpha_2 & 0 \\ \mu & \alpha_2 x_2 & \alpha_2 \\ 0 & 2\mu & 2\alpha_2 x_2 \end{pmatrix} \begin{pmatrix} \langle n^2 \rangle \\ \langle nx \rangle \\ \langle x^2 \rangle \end{pmatrix} + \begin{pmatrix} 2\alpha_1 \langle n \rangle + \alpha_2 \langle x \rangle + \alpha_1 \\ \alpha_1 x_c \langle n \rangle + (\alpha_2 x_2 + \alpha_1) \langle x \rangle + \alpha_1 x_c \\ \mu \langle n \rangle + (2\alpha_1 x_c + \alpha_2 x_2^2) \langle x \rangle + \alpha_1 x_c^2 \end{pmatrix}. \tag{A.11}$$

The matrix **A** for the homogeneous part of Eq. (A.11) has the eigenvalues

$$\begin{aligned}
\lambda_+ &= 2\kappa_+ \\
\lambda_0 &= \kappa_+ + \kappa_- \\
\lambda_- &= 2\kappa_-
\end{aligned} \tag{A.12}$$

and associated eigenvectors

$$\mathbf{v}_+ = \begin{pmatrix} \alpha_2 \\ \kappa_+ \\ -\frac{\mu\kappa_+}{\kappa_-} \end{pmatrix}, \quad \mathbf{v_0} = \begin{pmatrix} \alpha_2 \\ \alpha_2 x_2/2 \\ -\mu \end{pmatrix}, \quad \mathbf{v}_- = \begin{pmatrix} \alpha_2 \\ -\kappa_- \\ \frac{\mu\kappa_-}{\kappa_+} \end{pmatrix}, \tag{A.13}$$

where $\kappa_\pm$ are defined in Eq. (A.6). Transforming Eq. (A.10) to the eigenvector basis by premultiplying the coupled ODEs by the inverse $\mathbf{C^{-1}}$ of the eigenvector-column matrix $\mathbf{C}$ yields the uncoupled set of inhomogeneous ODEs

$$\frac{dy_i}{dt} - \lambda_i y_i = h_i \tag{A.14}$$

where y_i are the components of the solution vector in the eigenvector basis

$$\mathbf{y}(t) = \mathbf{C^{-1}}\mathbf{x}(t)$$

and h_i are the components of the vector

$$\mathbf{h}(t) = \mathbf{C^{-1}}\mathbf{g}(t).$$

The solution of Eqs. (A.14) can be written as the sum of a homogeneous solution and a particular integral

$$y_i(t) = y_{i,h}(t) + y_{i,p}(t).$$

The homogeneous solution is given by

$$y_{i,h}(t) = B_i e^{\lambda_i t}$$

where B_i is a constant of integration. The particular integral $y_{i,p}(t)$ is given by the formula

$$y_{i,p}(t) = e^{\lambda_i t} \int h_i(t) e^{-\lambda_i t} dt.$$

Because the components $g_i(t)$ are linear combinations of the first moments, the particular integrals $y_{i,p}(t)$ are linear combinations of terms in $e^{\kappa_+ t}$, $e^{\kappa_- t}$ and constant terms. In the long-time limit, the homogeneous solution term corresponding to the largest eigenvalue dominates therefore over all other terms, yielding the solution

$$\begin{pmatrix} \langle n^2 \rangle \\ \langle nx \rangle \\ \langle x^2 \rangle \end{pmatrix} = B_+ \begin{pmatrix} \alpha_2 \\ \kappa_+ \\ -\frac{\mu\kappa_+}{\kappa_-} \end{pmatrix} e^{2\kappa_+ t}, \quad t \gg \kappa^{-1}, \tag{A.15}$$

where the constant B_+ is fixed by the initial conditions. From Eq. (A.15), we find that the second moments of the PDF obey the relationship

$$\frac{\langle x^2 \rangle}{\langle n^2 \rangle} = -\frac{\mu\kappa_+}{\alpha_2 \kappa_-} = \left(\frac{\kappa_+}{\alpha_2}\right)^2, \quad t \gg \kappa^{-1}, \tag{A.16}$$

where we have used $\kappa_+\kappa_- = -\mu\alpha_2$.

A.1.3 Examining higher-order moments

The methodology employed in the previous section can be extended to the general case of j-th order moments to show that for any j, the following relationship holds

$$\frac{\langle x^j \rangle}{\langle n^j \rangle} = \left(\frac{\kappa_+}{\alpha_2} \right)^j, \quad t \gg \kappa^{-1}. \tag{A.17}$$

To this purpose, we examine the form of the master equation (3.1) and note that for any $p + q = j$ the following holds

$$\frac{d}{dt} \langle n^p x^q \rangle = q\mu \langle n^{p+1} x^{q-1} \rangle + p\alpha_2 \langle n^{p-1} x^{q+1} \rangle + q x_2 \alpha_2 \langle n^p x^q \rangle + O(\langle n^{j-1} \rangle). \tag{A.18}$$

Equation (A.18) allows the matrix $\mathbf{A}_j$ describing the homogeneous part of the time evolution for the j-th order moments to be written down explicitly

$$\mathbf{A}_j = \begin{pmatrix}
0 & j\alpha_2 & & & & \\
\mu & \alpha_2 x_2 & (j-1)\alpha_2 & & & \\
 & 2\mu & 2\alpha_2 x_2 & (j-2)\alpha_2 & & \\
 & & 3\mu & 3\alpha_2 x_2 & \ddots & \\
 & & & \ddots & \ddots & \alpha_2 \\
 & & & & j\mu & j\alpha_2 x_2
\end{pmatrix}. \tag{A.19}$$

Writing the column vector for the j-th order moments in descending order of increasing powers of x

$$\mathbf{x} = \begin{pmatrix} \langle x^j \rangle \\ \langle x^{j-1} n \rangle \\ \vdots \\ \langle n^j \rangle \end{pmatrix}, \tag{A.20}$$

the matrix $\mathbf{A}_j$ has tridiagonal form; with the top-right diagonal descending in decrements of α_2, from $j\alpha_2$ to α_2, the middle diagonal descending in increments of $x_2\alpha_2$ from 0 to $jx_2\alpha_2$, and the bottom-left diagonal descending in increments of μ from μ to $j\mu$.

For $t \gg \kappa^{-1}$, the time evolution of $\mathbf{x}$ is determined by the largest eigenvalue of $\mathbf{A}_j$ and the associated eigenvector. We now show that the largest eigenvalue of $\mathbf{A}_j$ is $\lambda = j\kappa_+$, where κ_+ is given by Eq. (A.6). To this purpose, we recall that the eigenvalues λ of the matrix $\mathbf{A}_j$ are obtained as solutions of the equation

$$\det(\mathbf{A}_j - \lambda \mathbf{I}) = 0, \tag{A.21}$$

where $\mathbf{I}$ denotes the identity matrix. Now, defining $B_{j,q} = \det(\mathbf{A}_j - \lambda\mathbf{I})_q$, where $(\mathbf{A}_j - \lambda\mathbf{I})_q = (\mathbf{A}_j - \lambda\mathbf{I})$ with the bottom $j - q$ rows and right-most $j - q$ columns deleted, yields to a recursion relation

$$B_{j,q} = (qx_2\alpha_2 - \lambda)B_{j,q-1} - (j-(q-1))q\alpha_2\mu B_{j,q-2} \tag{A.22}$$
$$B_{j,j} = \det(\mathbf{A}_j - \lambda\mathbf{I}). \tag{A.23}$$

The initial terms of this recursion relation are given explicitly by

$$B_{j,0} = -\lambda \tag{A.24}$$
$$B_{j,1} = -\lambda(x_2\alpha_2 - \lambda) - \mu j\alpha_2 \tag{A.25}$$
$$B_{j,2} = (2x_2\alpha_2 - \lambda)B_{j,1} + 2(j-1)\alpha_2\mu\lambda. \tag{A.26}$$

Inserting $\lambda = j\kappa_+$ in the above recursion relation yields

$$B_{j,0} = -j\kappa_+ \tag{A.27}$$
$$B_{j,1} = j(j-1)\kappa_+^2 \tag{A.28}$$
$$B_{j,2} = -j(j-1)(j-2)\kappa_+^3. \tag{A.29}$$

From the form of the first few terms, we find that

$$B_{j,q} = (-\kappa_+)^{q+1}\prod_{i=0}^{q}(j-i) \tag{A.30}$$

is the solution of the recurrence relation. Further, $B_{j,j} = 0$, implying that $\lambda = j\kappa_+$ is indeed an eigenvalue of the jth-order homogeneous matrix $\mathbf{A}_j$.

We now examine the eigenvector equation

$$(\mathbf{A}_j - \lambda\mathbf{I})\mathbf{v} = 0, \tag{A.31}$$

where $\mathbf{v}$ is the eigenvector associated with the eigenvalue $\lambda = j\kappa_+$. In order to solve Eq. (A.31), it is convenient to consider the ratios $r_i = v_i/v_0$. From Eq. (A.31), the r_i satisfy the following recurrence relation

$$r_q = \frac{\lambda - x_2\alpha_2(q-1)}{\alpha_2(j-(q-1))}r_{q-1} - \frac{\alpha_2\mu(q-1)}{\alpha_2(j-(q-1))}r_{q-2} \tag{A.32}$$

subject to the boundary conditions

$$r_0 = 1, \quad r_1 = \frac{\lambda}{j\alpha_2}. \tag{A.33}$$

The solution of Eq. (A.32) subject to the boundary conditions (A.33) is

$$r_i = \left(\frac{\kappa_+}{\alpha_2}\right)^i . \tag{A.34}$$

Therefore, the required eigenvector $\mathbf{v}$ to the largest eigenvalue $\lambda = j\kappa_+$ is

$$\mathbf{v} = v_0 \begin{pmatrix} 1 \\ \frac{\kappa_+}{\alpha_2} \\ \vdots \\ \left(\frac{\kappa_+}{\alpha_2}\right)^j \end{pmatrix} . \tag{A.35}$$

Summarizing the above results, for $t \gg \kappa^{-1}$ the moments of order j are given by

$$\begin{pmatrix} \langle x^j \rangle \\ \langle x^{j-1} n \rangle \\ \vdots \\ \langle n^j \rangle \end{pmatrix} = C \begin{pmatrix} 1 \\ \frac{\kappa_+}{\alpha_2} \\ \vdots \\ \left(\frac{\kappa_+}{\alpha_2}\right)^j \end{pmatrix} e^{j\kappa_+ t}, \tag{A.36}$$

where C is a constant of integration. From Eq. (A.36), we find therefore

$$\frac{\langle x^j \rangle}{\langle n^j \rangle} = \left(\frac{\kappa_+}{\alpha_2}\right)^j, \quad t \gg \kappa^{-1}. \tag{A.37}$$

A.2 Calculation of Pearson's correlation coefficient for n and x

For a mathematically simpler proof of the linear correlation between n and x, we consider Pearson's correlation coefficient for n and x, defined as

$$\rho_{n,x} = \frac{\langle \tilde{n}\tilde{x} \rangle}{\sqrt{\langle \tilde{n}^2 \rangle \langle \tilde{x}^2 \rangle}}, \tag{A.38}$$

where

$$\begin{aligned} \langle \tilde{n}^2 \rangle &= \langle n^2 \rangle - \langle n \rangle^2 \\ \langle \tilde{n}\tilde{x} \rangle &= \langle nx \rangle - \langle n \rangle \langle x \rangle \\ \langle \tilde{x}^2 \rangle &= \langle x^2 \rangle - \langle x \rangle^2 . \end{aligned} \tag{A.39}$$

Pearson's correlation coefficient is a measure of the linear relationship between two random variables: the Pearson's correlation coefficient $\rho_{n,x}$ equals 1 (-1) in the case of a perfect increasing (decreasing) linear relationship between the random variables n, x. In the following, we show the random variables n and x have a Pearson's correlation coefficient of 1 for $t \gg \kappa^{-1}$, indicating that n and x are perfectly linearly correlated for $t \gg \kappa^{-1}$. To this purpose, we note that the quantities $\langle\tilde{n}^2\rangle, \langle\tilde{n}\tilde{x}\rangle, \langle\tilde{x}^2\rangle$ obey the following closed set of ODEs

$$\frac{d}{dt}\begin{pmatrix}\langle\tilde{n}^2\rangle\\ \langle\tilde{n}\tilde{x}\rangle\\ \langle\tilde{x}^2\rangle\end{pmatrix} = \begin{pmatrix}0 & 2\alpha_2 & 0\\ \mu & \alpha_2 x_2 & \alpha_2\\ 0 & 2\mu & 2\alpha_2 x_2\end{pmatrix}\begin{pmatrix}\langle\tilde{n}^2\rangle\\ \langle\tilde{n}\tilde{x}\rangle\\ \langle\tilde{x}^2\rangle\end{pmatrix} + \begin{pmatrix}\alpha_2\langle x\rangle + \alpha_1\\ \alpha_2 x_2\langle x\rangle + \alpha_1 x_c\\ \mu\langle n\rangle + \alpha_2 x_2^2\langle x\rangle + \alpha_1 x_c^2\end{pmatrix} \tag{A.40}$$

obtained by combining Eqs. (A.3) and (A.9). Following the same arguments as in Sec. A.1.2, for $t \gg \kappa^{-1}$ the solution of Eq. (A.40) is obtained by considering the eigenvector to the largest eigenvalue of the matrix

$$\mathbf{A} = \begin{pmatrix}0 & 2\alpha_2 & 0\\ \mu & \alpha_2 x_2 & \alpha_2\\ 0 & 2\mu & 2\alpha_2 x_2\end{pmatrix} \tag{A.41}$$

yielding

$$\begin{pmatrix}\langle\tilde{n}^2\rangle\\ \langle\tilde{n}\tilde{x}\rangle\\ \langle\tilde{x}^2\rangle\end{pmatrix} = C\begin{pmatrix}\alpha_2\\ \kappa_+\\ -\frac{\mu\kappa_+}{\kappa_-}\end{pmatrix} e^{2\kappa_+ t}, \quad t \gg \kappa^{-1}, \tag{A.42}$$

where C is a constant of integration and the parameters $\kappa_\pm$ are defined as in Eq. (A.6). Combining Eq. (A.42) with Eq. (A.38) yields the following result valid for $t \gg \kappa^{-1}$

$$\rho_{n,x} = \frac{\langle\tilde{n}\tilde{x}\rangle}{\sqrt{\langle\tilde{n}^2\rangle\langle\tilde{x}^2\rangle}} = \frac{C\kappa_+ e^{2\kappa_+ t}}{\sqrt{C^2\left(-\frac{\alpha_2\mu\kappa_+}{\kappa_-}\right)e^{4\kappa_+ t}}} = \left(-\frac{\kappa_+\kappa_-}{\alpha_2\mu}\right)^{1/2} = 1, \quad t \gg \kappa^{-1}, \tag{A.43}$$

where we used $\kappa_+\kappa_- = -\alpha_2\mu$. Equation (A.43) implies that for $t \gg \kappa^{-1}$ the random variables n and x are linearly correlated, i.e. n and x obey a relationship of the form

$$x = \beta n + \gamma \tag{A.44}$$

for some β and γ. The constant β is determined from Eq. (A.7) as $\beta = \frac{\kappa_+}{\alpha_2}$, while for γ we find $\gamma = 0$.

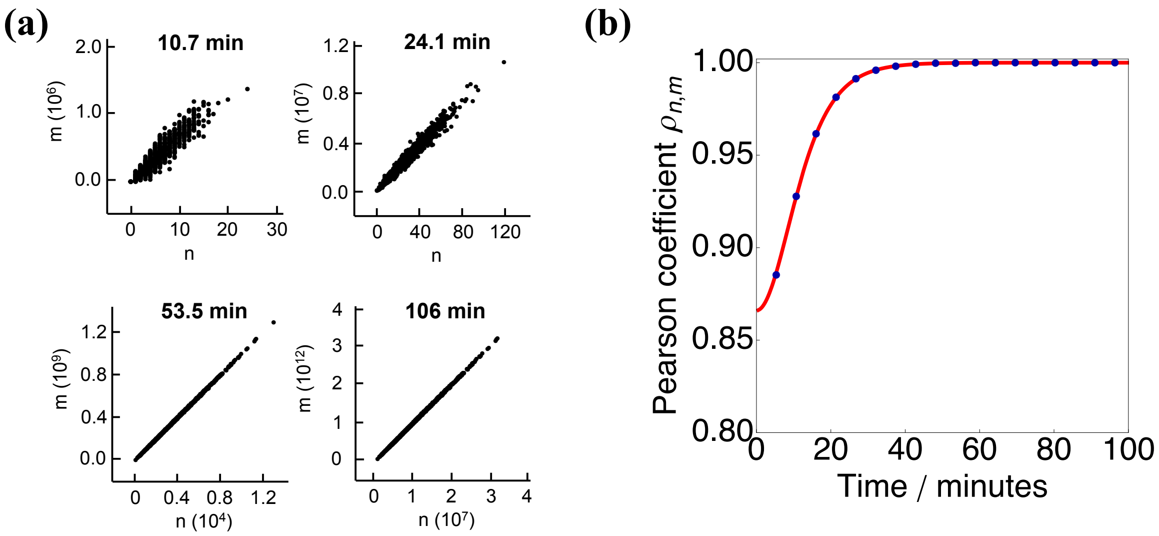

Figure A.1: Time evolution of Pearson's correlation coefficient (Solid line: analytical prediction). Calculation parameters: $k_n = 4 \times 10^{-13}$ M^{-1}s^{-1}, $x_c = 2$, $x_2 = 0$, $k_2 = 2.5 \times 10^{-8}$ s^{-1}, $k_+ = 2.5 \times 10^4$ M^{-1}s^{-1}, $m_{\text{tot}} = 5$ mM, $V = 1$ nL. Data points: numerical simulations of the master equation provided by J. B. Kirkegaard [73].

A.3 Simplification of Master Equation

Using the above relation $x = \beta n$, we can simplify the full master equation Eq. (3.1). We start by summing over all x, imposing the relation using Kronecker deltas:

$$\begin{aligned}\sum_{x=0}^{\infty} \frac{\partial \mathbb{P}(n,x,t)}{\partial t}\delta_{x\,(\beta n)} = {} & \alpha_1 \left[\sum_{x=0}^{\infty} \mathbb{P}(n-1,x-x_c,t)\delta_{x\,(\beta n)} - \sum_{x=0}^{\infty} \mathbb{P}(n,x,t)\delta_{x\,(\beta n)}\right] \\ & + \mu n \left[\sum_{x=0}^{\infty} \delta_{x\,(\beta n)}\mathbb{P}(n,x-1,t) - \sum_{x=0}^{\infty} \delta_{x\,(\beta n)}\mathbb{P}(n,x,t)\right] \\ & + \sum_{x=0}^{\infty} \delta_{x\,(\beta n)}\alpha_2(x-x_2)\mathbb{P}(n-1,x-x_2,t) \\ & - \sum_{x=0}^{\infty} \delta_{x\,(\beta n)}\alpha_2 x\mathbb{P}(n,x,t).\end{aligned}$$

Noting that the Kronecker deltas simply pick individual terms out of these sums, we have:

$$\begin{aligned}\frac{\partial \mathbb{P}(n,t)}{\partial t} &= \alpha_1\left[\mathbb{P}(n-1,t) - \mathbb{P}(n,t)\right] + \mu n\left[\mathbb{P}(n,t) - \mathbb{P}(n,t)\right] \\ &\quad + \alpha_2(\beta(n-1)-x_2)\mathbb{P}(n-1,t) - \alpha_2\beta n\mathbb{P}(n,t) \\ &\simeq \alpha_1\mathbb{P}(n-1,t) - \alpha_1\mathbb{P}(n,t) \\ &\quad + \kappa(n-1)\mathbb{P}(n-1,t) - \kappa n\mathbb{P}(n,t),\end{aligned}$$

where the last line follows from $\beta \gg x_2$.

A.4 Derivation of PDF[†]

In this section, we provide the mathematical details pertaining to the derivation of Eq. (3.3) of Chapter 3. As a strategy for solving Eq. (3.2) we exploit the use of generating functions

$$\mathcal{C}(z,t) = \sum_{n=0}^{\infty} \mathbb{P}(n,t) z^n. \tag{A.45}$$

Generating functions encode complete information about the probability distribution. For example, one can compute the principal moments of the probability distribution as

$$\langle n \rangle = \left. \frac{\partial \mathcal{C}(z,t)}{\partial z} \right|_{z=1}, \quad \langle n(n-1) \rangle = \left. \frac{\partial^2 \mathcal{C}(z,t)}{\partial z^2} \right|_{z=1}, \quad \text{etc.} \tag{A.46}$$

By multiplying Eq. (3.2) with z^n and summing over n from 0 to ∞ on both sides, a differential equation for the generating function $\mathcal{C}(z,t)$ is obtained

$$\frac{\partial \mathcal{C}(z,t)}{\partial t} = \alpha_1 (z-1) \mathcal{C}(z,t) + \kappa (z-1) z \frac{\partial \mathcal{C}(z,t)}{\partial z}. \tag{A.47}$$

Solving Eq. (A.47) subject to the initial condition $\mathcal{C}(z, t = \log(2)/\kappa) = 1$ (corresponding to $\mathbb{P}(n, t = \log(2)/\kappa) = \delta_{n,0}$) yields

$$\mathcal{C}(z,t) = \Big[z - e^{\kappa t}(z-1)/2 \Big]^{-\alpha_1/\kappa}. \tag{A.48}$$

According to Eq. (A.45), $\mathbb{P}(n,t)$ is found by expanding Eq. (A.48) in power series around $z = 0$, yielding

$$\mathbb{P}(n,t) = \frac{2^{\alpha_1/\kappa} \Gamma\left(n + \frac{\alpha_1}{\kappa}\right)}{\Gamma(n+1) \Gamma\left(\frac{\alpha_1}{\kappa}\right)} e^{-(\alpha_1 + \kappa n)t} (e^{\kappa t} - 2)^n, \tag{A.49}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function. In terms of the variable x, we combine Eqs (A.44), (A.49) to give the PDF for the aggregate mass

$$\mathbb{P}(x,t) = \frac{2^{\alpha_1/\kappa} \Gamma\left(\frac{\alpha_2}{\kappa} x + \frac{\alpha_1}{\kappa}\right)}{\Gamma\left(\frac{\alpha_2}{\kappa} x + 1\right) \Gamma\left(\frac{\alpha_1}{\kappa}\right)} e^{-(\alpha_1 + \alpha_2 x)t} (e^{\kappa t} - 2)^{\frac{\alpha_2}{\kappa} x}. \tag{A.50}$$

Equations (A.49) and (A.50) describe the time evolution of the full PDF.

We can use the generating function Eq. (A.48) to recover the time evolution of the principal moments of the PDF. For example, for the average value for n we obtain

$$\langle n \rangle = \left. \frac{\partial \mathcal{C}(z,t)}{\partial z} \right|_{z=1} = \frac{\alpha_1}{2\kappa} \left(e^{\kappa t} - 2 \right). \tag{A.51}$$

Similarly, from the second derivative of the generating function we obtain

$$\langle n(n-1)\rangle = \left.\frac{\partial^2 \mathcal{C}(z,t)}{\partial z^2}\right|_{z=1} = \frac{\alpha_1(\alpha_1+\kappa)}{4\kappa^2}\left(e^{\kappa t}-2\right)^2 \tag{A.52}$$

which can be used to compute the standard deviation of n.

A.5 Derivation of lag time formulae†

In small volumes, progress curves are non reproducible and the lag time τ is a random variable. Under these circumstances, finding τ translates into a first hitting time problem: we define τ as the time at which a sample path of the stochastic process x first hits the threshold $x_{\text{th}} = N_A V M_{\text{th}}$

$$\tau = \inf_{t\geq 0}\{x(t) \geq x_{\text{th}}\}. \tag{A.53}$$

Because τ is random, we are interest in the PDF of lag times, i.e. the probability $T(t)$ that τ equals t. According to the theory of first passage times, [52] the PDF of lag times, $T(t)$, is given by

$$T(t) = -\frac{dQ(t)}{dt}, \tag{A.54}$$

where $Q(t)$ denotes the probability that at time t the process x has not yet reached x_{th}

$$Q(t) = \sum_{n=0}^{n_{\text{th}}-1} \mathbb{P}(n,t), \tag{A.55}$$

where $n_{\text{th}} = \alpha_2 x_{\text{th}}/\kappa$. We can either substitute Eq. (A.50) in Eq. (A.55) in Eq. (A.54), or simply use the result in ref. [52] but replacing t with $t - \log(2)/\kappa$ to reflect our boundary condition at $t = \log(2)/\kappa$ (as our boundary condition is equivalent to a delayed starting time of $\log(2)/\kappa$). Either approach yields

$$T(t) = \frac{\kappa 2^{\alpha_1/\kappa}\Gamma\left(n_{\text{th}}+\frac{\alpha_1}{\kappa}\right)}{\Gamma\left(n_{\text{th}}\right)\Gamma\left(\frac{\alpha_1}{\kappa}\right)} e^{-(\alpha_1+(n_{\text{th}}-1)\kappa)t}(e^{\kappa t}-2)^{n_{\text{th}}-1}. \tag{A.56}$$

Eq. (A.56) gives our central result: an analytical expression for PDF of lag times. In the long time limit, the PDF of lag times decays exponentially

$$T(t) \overset{t\to\infty}{\to} \frac{\kappa 2^{\alpha_1/\kappa}\Gamma\left(n_{\text{th}}+\frac{\alpha_1}{\kappa}\right)}{\Gamma(n_{\text{th}})\Gamma(\alpha_1/\kappa)} e^{-\alpha_1 t}, \tag{A.57}$$

with characteristic decay time α_1^{-1}. In the opposite limit of early times, $T(t)$ and all its derivative decay to zero; hence, $T(t)$ has an essential singularity at $t = 0$. The maximum of the PDF $T(t)$ is attained at

$$\tau^* = \frac{1}{\kappa}\log\left(\frac{2(n_{\mathrm{th}}-1)\kappa}{\alpha_1}+2\right) \approx \frac{1}{\kappa}\log\left(\frac{2\alpha_2 x_{\mathrm{th}}}{\alpha_1}+2\right), \tag{A.58}$$

which can be interpreted as the most probable value for the lag time. Interestingly, the mode of the PDF of lag times equals the deterministic lag time, Eq. (A.66). The expected value for the lag time τ can again be obtained from addition of $\log(2)/\kappa$ to the result in ref. [52], yielding

$$\langle\tau\rangle = \frac{\log(2)}{\kappa} + \sum_{j=0}^{n_{\mathrm{th}}-1} \frac{1}{(\alpha_1 + j\kappa)}. \tag{A.59}$$

Note the expected lag time is different from the most probable lag time. In the limit of large volumes and high nucleation rate, corresponding to $\alpha_1 \gg \kappa$, however, both values agree

$$\langle\tau\rangle \approx \frac{\log(2)}{\kappa} + \int_0^{n_{\mathrm{th}}-1} \frac{dy}{\alpha_1 + \kappa y} = \tau^*. \tag{A.60}$$

The variance of the lag-time distribution is computed similarly and reads

$$\sigma_\tau^2 = \sum_{j=0}^{n_{\mathrm{th}}-1} \frac{1}{(\alpha_1 + j\kappa)^2}. \tag{A.61}$$

A.6 Limiting behaviours for the formula for average lag time and extent of fluctuations

In Chapter 3, we have derived a closed-form expression for the average lag time as a function of volume (Eq. (3.5) of Chapter 3)

$$\langle\tau\rangle = \frac{\log(2)}{\kappa} + \sum_{j=0}^{n_{th}-1} \frac{1}{\alpha_1 + j\kappa}, \quad t \gg \kappa^{-1}, \tag{A.62}$$

where $n_{th} = \frac{\alpha_2 N_A V M_{th}}{\kappa}$. Similarly, we obtain a closed-form expression for the variance of the lag time

$$\sigma^2 = \langle\tau^2\rangle - \langle\tau\rangle^2 = \sum_{j=0}^{n_{th}-1} \frac{1}{(\alpha_1 + j\kappa)^2}, \quad t \gg \kappa^{-1}. \tag{A.63}$$

In this appendix, we discuss several limits of practical importance, for which the functional forms of Eqs. (A.62) and (A.63) can be greatly simplified.

A.6.1 Bulk regime

The bulk regime emerges when the system volume V is large enough that fluctuations can be ignored. Defining the dimensionless ratio $\gamma = \kappa/\alpha_1 = \kappa/(\alpha_0 N_A V)$, where $\alpha_0 = k_n m_{\mathrm{tot}}^{n_c}$,

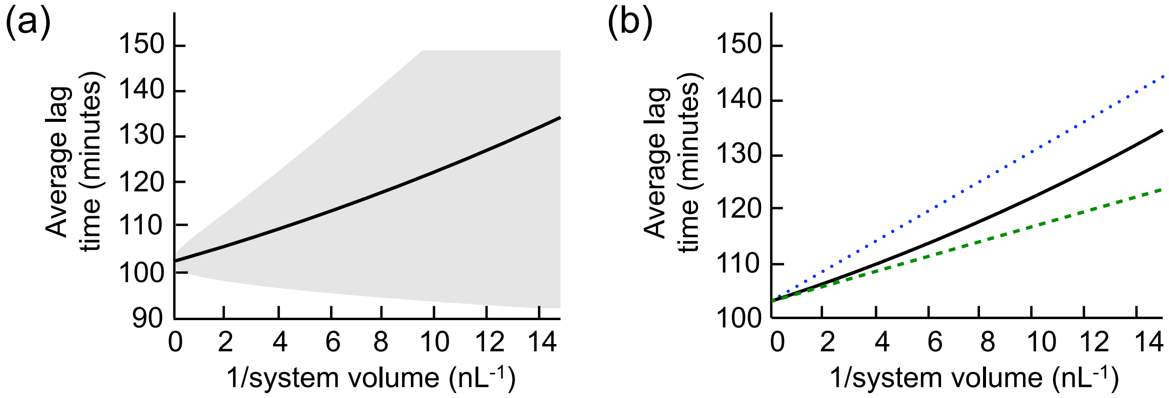

Figure A.2: (a) Volume dependence of average lag time predicted using Eq. (A.62) and 68% confidence bands computed using Eq. (A.63). The parameters are the same as for Fig. 2(a) of Chapter 3. (b) Average lag time (solid line) is approximated by $\langle\tau\rangle = \tau_g + c_n/(2V)$ for $\gamma \to 0$ (dashed line) turning to the prediction of classical nucleation theory for $\gamma \to \infty$ (dotted line).

the bulk regime corresponds to $\gamma \to 0$. Under these circumstances, the sum $\sum_{j=0}^{n_{\mathrm{th}}-1} \frac{\gamma}{1+\gamma j}$ entering the expression for the adimensionalised average lag time $\kappa\langle\tau\rangle$, Eq. (A.62), is replaced by the integral $\int_0^{n_{th}} \frac{\gamma}{1+\gamma x} dx$, yielding

$$\kappa\langle\tau\rangle = \log(2) + \sum_{j=0}^{n_{th}-1} \frac{\gamma}{1+\gamma j} \tag{A.64}$$

$$\to \log(2) + \int_0^{n_{th}} \frac{\gamma}{1+\gamma x} dx = \log\left(\frac{2\alpha_2 M_{th}}{\alpha_0} + 2\right), \tag{A.65}$$

or, equivalently,

$$\langle\tau\rangle \to \frac{1}{\kappa} \log\left(\frac{2\alpha_2 M_{th}}{\alpha_0} + 2\right) = \tau_g, \tag{A.66}$$

where τ_g is the deterministic lag time, which is obtained as solution of $M(\tau_g) = M_{th}$ with $M(t)$ being the aggregate mass concentration in bulk. Therefore, our expression for the volume dependence of the average lag time, Eq. (A.62), converges to the deterministic lag time τ_g in the limit of a very large system.

In the limit $\gamma \to 0$, the expression for the standard deviation of the lag time becomes

$$\kappa^2\sigma^2 = \sum_{j=0}^{n_{th}-1} \frac{\gamma^2}{(1+\gamma j)^2} \tag{A.67}$$

$$\to \int_0^{n_{th}} \frac{\gamma^2}{(1+\gamma x)^2} dx = \gamma\left(1 - \frac{1}{1+\frac{\alpha_2 M_{th}}{\alpha_0}}\right), \tag{A.68}$$

or, equivalently,

$$\sigma \to \frac{1}{\sqrt{\alpha_1 \kappa}} \left(1 - \frac{1}{1 + \frac{\alpha_2 M_{th}}{\alpha_0}} \right)^{1/2} = \sigma_g. \tag{A.69}$$

Note that in this limit, the extent of fluctuations scale as $\sigma_g \sim \kappa^{-1/2} \alpha_0^{-1/2} V^{-1/2}$.

A.6.2 Small fluctuation regime

In Sec. A.6.1, we have shown that in the limit of very large volumes the expression for the average lag time, Eq. (A.62), recovers the deterministic lag time. Formally, this result was obtained by replacing sums with integrals in Eq. (A.62), which is justified only in the limit $\gamma \to 0$. When γ is small but non-zero , however, correction terms appear. In this section, we consider how Eq. (A.62) approaches the bulk limit with increasing volume and look for an approximation of Eq. (A.62) which is valid for small but non-zero γ.

The starting point for this calculation is the Euler-Maclaurin summation formula[1]

$$\sum_{n=0}^{n_{th}-1} f(n) = \int_0^{n_{th}} f(t) dt - \frac{1}{2} \left[f(n_{th}) - f(0) \right] + \sum_{i=1}^{\infty} \frac{B_{2i}}{(2i)!} \left[f^{(2i-1)}(n_{th}) - f^{(2i-1)}(0) \right], \tag{A.70}$$

where f is any smooth real function and B_{2i} are the Bernoulli numbers. Application of the Euler-Maclaurin formula, Eq. (A.70), to Eq. (A.62) with $f(j) = \frac{\gamma}{1+\gamma j}$ gives for the following expression for the average lag time,

$$\kappa \langle \tau \rangle = \kappa \tau_g + \frac{\gamma}{2} \left(1 - \frac{1}{1 + \gamma n_{th}} \right) + \sum_{i=1}^{\infty} \frac{B_{2i}}{2i} \gamma^{2i} \left(1 - \frac{1}{(1 + \gamma n_{th})^{2i}} \right). \tag{A.71}$$

As expected, for $\gamma \to 0$ the bulk limit $\langle \tau \rangle \to \tau_g$ is recovered, in accordance with the results of Sec. A.6.1. Equation (A.71), however, provides higher-order correction terms in γ (or, equivalently, V^{-1}) that describe the departure of $\langle \tau \rangle$ from the deterministic value τ_g when γ is small but non-zero. To leading order in γ, Eq. (A.71) becomes

$$\kappa \langle \tau \rangle = \kappa \tau_g + \frac{\gamma}{2} + \cdots . \tag{A.72}$$

Hence, in a plot of $\langle \tau \rangle$ against V^{-1} the average lag time expression Eq. (A.62) approaches the bulk limit τ_g approximatively as a straight line of gradient (see Fig. 2(b) of Chapter 3)

$$\frac{d\langle \tau \rangle}{d(1/V)} \simeq \frac{1}{2\alpha_0 N_A}, \qquad V^{-1} \ll \frac{\alpha_0 N_A}{\kappa}. \tag{A.73}$$

---

[1] Thomas C. T. Michaels suggested use of the Euler-Maclaurin summation formula for this purpose

A.6.3 Large fluctuation regime

In the previous section, we have considered the behaviour of Eq. (A.62) when the system is close to the bulk regime ($\gamma \to 0$). Results were expressed in terms of a power series in γ, which requires γ to be small. We now consider the opposite limit of very small volumes ($\gamma \to \infty$), where the results of Sec. A.6.2 no longer apply.

In the limit of very small volumes (large γ), Eq. (A.62) diverges. In order to extract the leading order divergent term, it proves easier to consider the first derivative of Eq. (A.62) with respect to γ, yielding

$$\frac{d(\kappa\langle\tau\rangle)}{d\gamma} = \sum_{i=1}^{n_{th}-1}\left[\frac{1}{1+\gamma i} - \frac{\gamma i}{(1+\gamma i)^2}\right] + 1. \tag{A.74}$$

In the limit $\gamma \to \infty$, the leading order term in Eq. (A.74) is

$$\frac{d(\kappa\langle\tau\rangle)}{d\gamma} = 1 + \cdots . \tag{A.75}$$

This equation can be reduced to the simple expression for the gradient of the lag time with the inverse volume, valid in the limit of small volume:

$$\frac{d\langle\tau\rangle}{d(1/V)} \simeq \frac{1}{\alpha_0 N_A}, \quad V^{-1} \gg \frac{\alpha_0 N_A}{\kappa}. \tag{A.76}$$

In the limit $\gamma \to \infty$, the expression for the variance of the lag time, Eq. (A.63), can be treated similarly to yield

$$\frac{d(\kappa^2\sigma^2)}{d\gamma} = \sum_{i=1}^{n_{th}-1}\left[\frac{2\gamma}{(1+\gamma i)^2} - \frac{2\gamma^2 i}{(1+\gamma i)^3}\right] + 2\gamma = 2\gamma + \cdots . \tag{A.77}$$

Therefore, integrating Eq. (A.77) yields the compact expression

$$\sigma \simeq \sqrt{\sigma_g^2 + \frac{1}{(\alpha_0 N_A V)^2}}, \quad V^{-1} \gg \frac{\alpha_0 N_A}{\kappa}. \tag{A.78}$$

A.6.4 Summary

Introducing the quantity $c_n = \frac{1}{k_n m_{\text{tot}}^{n_c} N_A}$, the above results can be summarized by the following simple statement for the gradient of the lag time with the inverse volume

$$\frac{c_n}{2} \leq \frac{d\langle\tau\rangle}{d(1/V)} \leq c_n, \tag{A.79}$$

with the lower bound describing the approximatively linear approach of $\langle\tau\rangle$ to the bulk limit τ_g for small V^{-1} and the upper bound c_n describing system behaviour for very large V^{-1}.

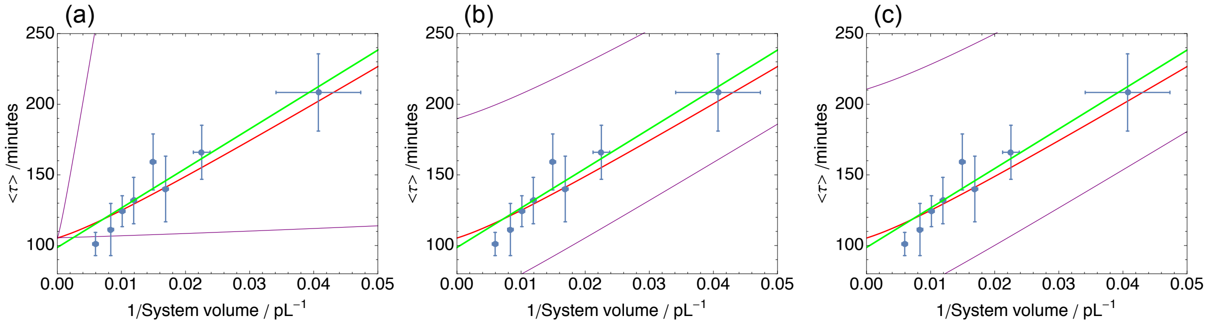

Figure A.3: (Analysis of small-volume experiments of bovine insulin fibrillization as discussed in Chapter 3 (Fig. 2(a)). Data provided by K. L. Saar. [73] Thin purple lines: k_n is decreased/increased by an order of magnitude (a); k_2 is decreased/increased by a factor 4 (b); k_+ is decreased/increased by a factor 4 (c). The plots show that the slope of the average lag time in a $1/V$ plot is controlled by the rate of nucleation; changing k_+ or k_2 shifts the intercept (deterministic lag time), but does not affect the slope. These results demonstrate that an analysis of the volume dependence of the average lag time is able to separate the contributions from primary nucleation (k_n) and auto-catalytic growth (k_+, k_2) and hence provides an accurate route for determining k_n with increased accuracy relative to conventional bulk methods.

Appendix B

The Early Stages of Heteromolecular Filament Formation

This appendix contains the key derivation associated with Chapter 4, as well as a few generalizations and extensions of the theory. It finally contains the rate constants used in the kinetic plots included as figures in Chapter 4. All of these appendices originate from the paper "Dynamics of heteromolecular filament formation" published in *J Chem Phys* in 2016; some appendices from this publication that were not directly relevant to the thesis theme were not reproduced here. All work is my own.

B.1 Considering primary nucleation with arbitrary reaction order

The moment equations can be straightforwardly generalized to arbitrary nucleation reaction order, with nucleus size not necessarily corresponding to reaction order. We write $k_n(x; a^i b^j)$ as the rate constant for the primary nucleation step producing ends of type x of reaction order i with respect to a and j with respect to b. We illustrate with a system without secondary processes. The moment equations become then:

$$\begin{aligned}
\frac{d}{dt}P_a &= \sum_{i,j} k_n(a; a^i b^j) m_a(t)^i m_b(t)^j \\
&\quad + k_+(a|b) m_a(t) P_b(t) - k_+(b|a) m_b(t) P_a(t) \\
\frac{d}{dt}P_b &= \sum_{i,j} k_n(b; a^i b^j) m_a(t)^i m_b(t)^j \\
&\quad - k_+(a|b) m_a(t) P_b(t) + k_+(b|a) m_b(t) P_a(t) \qquad \text{(B.1a)}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}M_a &= k_+(a|b) m_a(t) P_b(t) + k_+(a|a) m_a(t) P_a(t) \\
\frac{d}{dt}M_b &= k_+(b|b) m_b(t) P_b(t) + k_+(b|a) m_b(t) P_a(t). \qquad \text{(B.1b)}
\end{aligned}$$

In practice we might expect to have an overall reaction order n_c for all possible nucleation steps; or at least a maximum overall reaction order. Regardless, at early time, with or without secondary processes, our extended moment equations are the same but with modified α_a and α_b:

$$\begin{aligned}\alpha_a &= \sum_{i,j} k_n(a; a^i b^j) m_a(0)^i m_b(0)^j \\ \alpha_b &= \sum_{i,j} k_n(b; a^i b^j) m_a(0)^i m_b(0)^j.\end{aligned} \tag{B.2}$$

Our methodology is thus unaffected by generalizing to arbitrary primary nucleation reaction order.

B.2 Solving the early-time extended moment equations for a system with secondary processes

B.2.1 Early-time extended moment equations

At early time, the rates $(\delta_{px} + \delta_{qx})k_2(pq|yz)m_p(t)m_q(t)$ as well as the rates of primary nucleation become time-independent, as we fix m_a and m_b at their initial values. The system of extended moment equations are now linear. Furthermore, the 1st moment ODEs are now decoupled from the rest and we need only solve the fibril end and subunit pair concentration equations simultaneously to calculate the fibril mass concentrations. We write the rate equations for the fibril end concentrations and subunit pair concetrations in matrix form:

$$\frac{d}{dt}\begin{pmatrix} P_a^{(0)}(t) \\ P_b^{(0)}(t) \\ N_{aa}^{(0)}(t) \\ N_{ab}^{(0)}(t) \\ N_{bb}^{(0)}(t) \end{pmatrix} = \begin{pmatrix} \alpha_a \\ \alpha_b \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -\mu_{ba} & \mu_{ab} & \alpha_{aa}^{2a} & \alpha_{ab}^{2a} & \alpha_{bb}^{2a} \\ \mu_{ba} & -\mu_{ab} & \alpha_{aa}^{2b} & \alpha_{ab}^{2b} & \alpha_{bb}^{2b} \\ \mu_{aa} & 0 & 0 & 0 & 0 \\ \mu_{ba} & \mu_{ab} & 0 & 0 & 0 \\ 0 & \mu_{bb} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} P_a^{(0)}(t) \\ P_b^{(0)}(t) \\ N_{aa}^{(0)}(t) \\ N_{ab}^{(0)}(t) \\ N_{bb}^{(0)}(t) \end{pmatrix}, \tag{B.3}$$

where $\alpha_{yz}^{2x} = \sum_{pq=a,b}(\delta_{px} + \delta_{qx})k_2(pq|yz)m_p(0)m_q(0)$, and $\alpha_{yz}^{2x}N_{yz}(t)$ represents the early-time rate of formation of ends of type x through secondary nucleation at sites of composition yz.

B.2.2 Solving the equations

To solve the early-time extended moment equations (B.3), its homogeneous matrix must be diagonalized. This matrix has 5 eigenvalues in total. Because this matrix has vanishing determinant, one of its eigenvalues is trivial, $\lambda = 0$. The remaining 4 eigenvalues are determined by the following quartic equation:

$$a\lambda^4 + b\lambda^3 + c\lambda^2 + d\lambda + e = 0 \ , \tag{B.4}$$

where the coefficients are explicitly given by the following equations:

$$\begin{aligned}
a &= 1 && \text{(B.5)}\\
b &= \mu_{ab} + \mu_{ba} = \mu_c && \text{(B.6)}\\
c &= -(\alpha^{2a}_{aa}\mu_{aa} + \alpha^{2b}_{ab}\mu_{ab} + \alpha^{2a}_{ab}\mu_{ba} + \alpha^{2b}_{bb}\mu_{bb}) && \text{(B.7)}\\
d &= -((\alpha^{2a}_{aa} + \alpha^{2b}_{aa})\mu_{aa}\mu_{ab}\\
&\quad + 2(\alpha^{2a}_{ab} + \alpha^{2b}_{ab})\mu_{ab}\mu_{ba}\\
&\quad + (\alpha^{2a}_{bb} + \alpha^{2b}_{bb})\mu_{ba}\mu_{bb}) && \text{(B.8)}\\
e &= (\alpha^{2a}_{aa}\alpha^{2b}_{ab} - \alpha^{2a}_{ab}\alpha^{2b}_{aa})\mu_{aa}\mu_{ab}\\
&\quad + (\alpha^{2a}_{aa}\alpha^{2b}_{bb} - \alpha^{2a}_{bb}\alpha^{2b}_{aa})\mu_{aa}\mu_{bb}\\
&\quad + (\alpha^{2a}_{ab}\alpha^{2b}_{bb} - \alpha^{2a}_{bb}\alpha^{2b}_{ab})\mu_{ba}\mu_{bb}. && \text{(B.9)}
\end{aligned}$$

While quartic equations admit, at least in principle, an exact solution, the resulting expressions tend to be very complicated and, thus, rather unpractical to use. Thus, in order to make progress, it is preferable to employ asymptotic methods (dominant balance argument) to derive simple, accurate approximative expressions for the eigenvalues.

Dominant balance

It is most convenient to employ a dominant balance argument to obtain approximate solutions to the eigenvalue equation (B.4). To get to this point, we have assumed that secondary nucleation is much slower than cross-elongation, which in turn is not significantly slower than self-elongation. We define the dimensionless parameter $\nu = (\sum_{i=a,b}[\alpha^{2i}_{aa} + \alpha^{2i}_{ab} + \alpha^{2i}_{bb}])/(\mu_{ab} + \mu_{ba})$, and note that under our assumptions $\nu << 1$. We then dedimensionalize the eigenvalue equation by dividing through by $b^4 = \mu_c^4$. Our equation now typically has the form:

$$(\lambda/\mu_c)^4 + (\lambda/\mu_c)^3 + O(\nu)(\lambda/\mu_c)^2 + O(\nu)(\lambda/\mu_c) + O(\nu^2) = 0. \tag{B.10}$$

We first guess that $\lambda/\mu_c = O(\nu^0)$. This gives the first two terms as the only non-vanishing terms and thus $\lambda^4 + \mu_c\lambda^3 \simeq 0$. This gives two approximate solutions, $\lambda \simeq 0$ and $\lambda \simeq -\mu_c$, but only the latter is permissible as only it is consistent with the order-of-magnitude assumption used to obtain it. We next guess that $\lambda/\mu_c = O(\nu^1)$. This gives the final two terms as the only non-vanishing terms and thus $d\lambda + e \simeq 0$. This gives the approximate solution $\lambda \simeq -e/d$, which has the correct order of magnitude and is thus permissible. Finally we guess that $\lambda/\mu_c = O(\nu^{\frac{1}{2}})$. This gives the second and fourth terms as the only non-vanishing terms and thus $\mu_c\lambda^3 + d\lambda \simeq 0$. This gives the approximate solutions $\lambda \simeq \pm\sqrt{-d/\mu_c}$, which have the correct order of magnitudes and are thus permissible (again we throw away the zero solution for being inconsistent).

In summary, the approximate eigenvalues are:

$$\lambda_0 = \lambda_+ = \sqrt{-\frac{d}{\mu_c}} \tag{B.11}$$

$$\lambda_1 = -\frac{e}{d} \tag{B.12}$$

$$\lambda_2 = -\sqrt{-\frac{d}{\mu_c}} \tag{B.13}$$

$$\lambda_3 = -\mu_c \ , \tag{B.14}$$

with the long-time behaviour dominated by the largest eigenvector, λ_+.

Obtaining the leading-order eigenvector contribution

We write the ODEs as:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{g}. \tag{B.15}$$

Using the matrix $\mathbf{C}$, whose columns are the eigenvectors (denoted $\mathbf{w_i}$), we transform the ODEs to the eigenvector basis:

$$\frac{d\mathbf{y}}{dt} = \mathbf{D}\mathbf{y}(t) + \mathbf{h}, \tag{B.16}$$

where $\mathbf{y}(t) = \mathbf{C}^{-1}\mathbf{x}(t)$, $\mathbf{h} = \mathbf{C}^{-1}\mathbf{g}$, and the diagonal matrix $\mathbf{D} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$. Due to the lack of time dependence of $\mathbf{g}$ the solutions to the ODEs, when decoupled in the eigenvector basis, have the form [73]:

$$y_i(t) = A_i e^{\lambda_i t} - \frac{[\mathbf{C}^{-1}\mathbf{g}]_i}{\lambda_i}, \tag{B.17}$$

where λ_i are the eigenvalues. A_i is a constant of integration whose value is given by matching the boundary conditions.
At $t = 0$, $y_i(0)$ must equal $[\mathbf{C}^{-1}\mathbf{x}(0)]_i$. Thus A_i can be calculated as:

$$\begin{aligned} A_i &= \frac{\mathbf{h}_i}{\lambda_i} + y_i(0) \\ &= \frac{[\mathbf{C}^{-1}\mathbf{g}]_i}{\lambda_i} + [\mathbf{C}^{-1}\mathbf{x}(0)]_i. \end{aligned} \tag{B.18}$$

For $t >> \lambda_+^{-1}$ the time evolution of $\mathbf{x}(t)$ is dominated by the contribution from the exponential term in y_+, and thus the full solution can be approximated as:

$$\mathbf{x}(t) \simeq A_+ \left(e^{\lambda_+ t} - 1\right) \mathbf{w}_+ + \mathbf{z}, \tag{B.19}$$

where $\mathbf{z}$ is some suitable constant vector. The most obvious choice for this vector would be $\mathbf{x}(0)$, in order to match the initial conditions. However, it is actually more appropriate to follow the treatment in Sec. 4.4.2, and for initial conditions allow the filament ends to equilibrate through cross-elongation, which happens extremely rapidly for any system of interest. We therefore take $\mathbf{z} = (P(0)\mu_{ab}/\mu_c, P(0)\mu_{ba}/\mu_c, N_{aa}, N_{ab}, N_{bb})$.

In practice, significant filament formation can only be detected when $t >> \lambda_+^{-1}$, so this approximation scheme is highly accurate.

Using the approximate eigenvalues to obtain the eigenvectors, we can calculate y_+. Throwing away terms of $O((\alpha_2/\mu_{tot}))$ and smaller, we find:

$$\frac{[\mathbf{C}^{-1}\mathbf{g}]_+}{\lambda_+} = \frac{(\alpha_a + \alpha_b)}{2\lambda_+^2(\mu_{ab} + \mu_{ba})} \tag{B.20}$$

$$\mathbf{w}_+ = \left(\lambda_+\mu_{ab}, \quad \lambda_+\mu_{ba}, \quad \mu_{aa}\mu_{ab}, \quad 2\mu_{ab}\mu_{ba}, \quad \mu_{ba}\mu_{bb}\right). \tag{B.21}$$

In the absence of seed $\mathbf{x}(0) = \mathbf{0}$ and these results give the early-time fibril end and subunit pair concentrations:

$$P_a^{(0)}(t) = \frac{(\alpha_a + \alpha_b)\mu_{ab}}{2\lambda_+(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right) \tag{B.22}$$

$$P_b^{(0)}(t) = \frac{(\alpha_a + \alpha_b)\mu_{ba}}{2\lambda_+(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right) \tag{B.23}$$

$$N_{aa}^{(0)}(t) = \frac{(\alpha_a + \alpha_b)\mu_{aa}\mu_{ab}}{2\lambda_+^2(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right) \tag{B.24}$$

$$N_{ab}^{(0)}(t) = \frac{(\alpha_a + \alpha_b)\mu_{ab}\mu_{ba}}{\lambda_+^2(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right) \tag{B.25}$$

$$N_{bb}^{(0)}(t) = \frac{(\alpha_a + \alpha_b)\mu_{ba}\mu_{bb}}{2\lambda_+^2(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right). \tag{B.26}$$

We can then use the 1st-order early-time extended moment equations to calculate:

$$M_a^{(0)}(t) = \frac{(\alpha_a + \alpha_b)(\mu_{aa} + \mu_{ba})\mu_{ab}}{2\lambda_+^2(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right) \tag{B.27}$$

$$M_b^{(0)}(t) = \frac{(\alpha_a + \alpha_b)(\mu_{ab} + \mu_{bb})\mu_{ba}}{2\lambda_+^2(\mu_{ab} + \mu_{ba})} \left(e^{\lambda_+ t} - 1\right). \tag{B.28}$$

When seeds are present, calculating $[\mathbf{C}^{-1}\mathbf{x}(0)]_+$ is not trivial and leads to:

$$[\mathbf{C}^{-1}\mathbf{x}(0)]_+ = \frac{P(0)}{2\lambda_+\mu_c} + F_{aa}N_{aa}(0) + F_{ab}N_{ab}(0) + F_{bb}N_{bb}(0), \tag{B.29}$$

where F_{ij} are unwieldy functions of the initial rates μ_{ij} and α_{ij}^{2k}. We note that $P(0)/(2\lambda_+\mu_c) \sim L_{\text{sys}}P(0)/\mu_c^2$ and $\sum F_{ij}N_{ij}(0) \sim L_{\text{seed}}P(0)/\mu_c^2$, where L_{seed} is the average seed filament length, and L_{sys} is the average filament length formed in an unseeded reaction. Therefore, if the seeds are sonicated to ensure that $L_{\text{seed}} \ll L_{\text{sys}}$, the F_{ij} can be ignored. This is typically the case since in seeded experiments the aim is usually to have $L_{\text{seed}} \ll L_{\text{sys}}$ in order that we can probe only the elongation rates. Their full forms are:

$$\begin{aligned}
F_{aa} &= \frac{2(\alpha_{ab}^{2a}\alpha_{aa}^{2b} - \alpha_{aa}^{2a}\alpha_{ab}^{2b})\mu_{ab}\mu_{ba}}{2(\alpha_{aa}^{2b}\mu_{aa}\mu_{ab} + 2\alpha_{ab}^{2b}\mu_{ab}\mu_{ba} + \alpha_{bb}^{2b}\mu_{ba}\mu_{bb})d} \\
&\quad + \frac{(\alpha_{bb}^{2a}\alpha_{aa}^{2b} - \alpha_{aa}^{2a}\alpha_{bb}^{2b})\mu_{ba}\mu_{bb} + \alpha_{aa}^{2b}d}{2(\alpha_{aa}^{2b}\mu_{aa}\mu_{ab} + 2\alpha_{ab}^{2b}\mu_{ab}\mu_{ba} + \alpha_{bb}^{2b}\mu_{ba}\mu_{bb})d} \\
F_{ab} &= \frac{(\alpha_{ab}^{2b}\alpha_{aa}^{2a} - \alpha_{ab}^{2a}\alpha_{aa}^{2b})\mu_{aa}\mu_{ab}}{2(\alpha_{aa}^{2b}\mu_{aa}\mu_{ab} + 2\alpha_{ab}^{2b}\mu_{ab}\mu_{ba} + \alpha_{bb}^{2b}\mu_{ba}\mu_{bb})d} \\
&\quad + \frac{(\alpha_{ab}^{2b}\alpha_{bb}^{2a} - \alpha_{ab}^{2a}\alpha_{bb}^{2b})\mu_{ba}\mu_{bb} + 2\alpha_{ab}^{2b}d}{2(\alpha_{aa}^{2b}\mu_{aa}\mu_{ab} + 2\alpha_{ab}^{2b}\mu_{ab}\mu_{ba} + \alpha_{bb}^{2b}\mu_{ba}\mu_{bb})d} \\
F_{bb} &= -\frac{(\alpha_{bb}^{2a}\alpha_{aa}^{2b} - \alpha_{aa}^{2a}\alpha_{bb}^{2b})\mu_{aa}\mu_{bb}}{2(\alpha_{aa}^{2b}\mu_{aa}\mu_{ab} + 2\alpha_{ab}^{2b}\mu_{ab}\mu_{ba} + \alpha_{bb}^{2b}\mu_{ba}\mu_{bb})d} \\
&\quad - \frac{2(\alpha_{bb}^{2a}\alpha_{ab}^{2b} - \alpha_{ab}^{2a}\alpha_{bb}^{2b})\mu_{ba}\mu_{bb} + \alpha_{bb}^{2b}d}{2(\alpha_{aa}^{2b}\mu_{aa}\mu_{ab} + 2\alpha_{ab}^{2b}\mu_{ab}\mu_{ba} + \alpha_{bb}^{2b}\mu_{ba}\mu_{bb})d},
\end{aligned} \tag{B.30}$$

where d is the linear coefficient in the eigenvalue equation.
Using these, A_+ can be modified accordingly to give full seeded early-time solutions.

B.3 Secondary systems with no cross-elongation

B.3.1 Cross-terms in primary nucleation only

For a system with no cross-elongation, the hetero-correlation functions are approximately 0 provided homo-elongation is faster than fragmentation. Then the system is uncoupled for early time, and approximate solutions can be derived for the concentration of each species using the same methodology as the single-species case [21, 30], yielding the same solutions but with appropriately modified early-time nucleation rates:

$$M_x^{(0)}(t) = \frac{\mu_{xx}}{\kappa_x}(A_x e^{\kappa_x t} - B_x e^{-\kappa_x t} + B_x - A_x) + M_x(0), \tag{B.31}$$

where $B_x = P_x(0) - A_x$, $A_x = \alpha_x/2\kappa_x + \alpha_{2x}M_x(0)/2\kappa + P_x(0)/2$, $\kappa_x = \sqrt{\alpha_{2x}\mu_{xx}}$, and α_{2x} is the early-time rate of new end formation of type x through secondary processes per unit of M_x.
Note that if $k_n(aa)m_a^2 >> k_n(ab)m_am_b >> k_n(bb)m_b^2$, then the α for the slower component must be modified to account for depletion of monomer. It may be best to work out how many b-type ends have formed by the a-system lag time, and then use this as the initial

condition for a seeded solution starting at the a-type lag time.

B.3.2 Including secondary process cross-seeding

In this case we still have coupled extended moment equations, but the hetero-correlation functions are again 0 so the pair concentrations are no longer needed to obtain the dynamics. We instead have 4 coupled equations for the number and mass concentrations:

$$\frac{d}{dt}\begin{pmatrix} P_a^{(0)}(t) \\ P_b^{(0)}(t) \\ M_a^{(0)}(t) \\ M_b^{(0)}(t) \end{pmatrix} = \begin{pmatrix} \alpha_a \\ \alpha_b \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \alpha_a^{2a} & \alpha_b^{2a} \\ 0 & 0 & \alpha_a^{2b} & \alpha_b^{2b} \\ \mu_{aa} & 0 & 0 & 0 \\ 0 & \mu_{bb} & 0 & 0 \end{pmatrix} \begin{pmatrix} P_a^{(0)}(t) \\ P_b^{(0)}(t) \\ M_a^{(0)}(t) \\ M_b^{(0)}(t) \end{pmatrix}. \tag{B.32}$$

This can be solved by the same method employed in Appendix B.2, yielding:

$$M_x^{(0)}(t) = \frac{\mu_{xx} G_x}{\lambda_2}(e^{\lambda_2 t} - \lambda_2 t - 1) + M_x(0), \tag{B.33}$$

where

$$G_x = \frac{\alpha_x(\lambda_2^2 - \alpha_{x'}^{2x'}\mu_{x'x'}) + \alpha_{x'}\alpha_{x'}^{2x}\mu_{x'x'}}{2\lambda_2(2\lambda_2^2 - \alpha_a^{2a}\mu_{aa} - \alpha_b^{2b}\mu_{bb})} \tag{B.34}$$

$$\lambda_2 = \sqrt{(\kappa_a^2 + \kappa_b^2)/2 + \sqrt{(\kappa_a^2 - \kappa_b^2)^2 + 4\alpha_a^{2b}\alpha_b^{2a}\mu_{aa}\mu_{bb}}/2}, \tag{B.35}$$

and $\kappa_x^2 = \alpha_x^{2x}\mu_{xx}$ (with $x' \neq x$).

B.4 Equivalence of pair correlation formulations

As discussed in Sec. 4.7, the simplest way of accessing the pair correlations is simply via $N_{xy}(t)$; for instance, $\mathcal{P}_{aa} = N_{aa}(t)/M(t)$. We can demonstrate the equivalence of these formulations as follows. Starting from the $N_{xy}(t)$ formulation given in this appendix:

$$\frac{dN_{aa}}{dt} = \mu_{aa} P_a(t) \frac{\mu_{aa} P_a(t) + \mu_{ba} P_a(t)}{\mu_{aa} P_a(t) + \mu_{ba} P_a(t)}. \tag{B.36}$$

As with a primary nucleation-dominated system, for all but very early times, the drift terms cancel and $\mu_{ba} P_a(t)/\mu_{ab} P_b(t)$ is very nearly 1. Combined with the expression for

dM_a/dt, we can therefore write:

$$\begin{aligned}\frac{dN_{aa}}{dt} &= \mu_{aa}P_a(t)\frac{\mu_{aa}P_a(t)+\mu_{ab}P_b(t)}{\mu_{aa}P_a(t)+\mu_{ba}P_a(t)}\\ &= \frac{\mu_{aa}P_a(t)M'_a(t)}{\mu_{aa}P_a(t)+\mu_{ba}P_a(t)}\\ &= \frac{\mu_{aa}M'_a(t)}{\mu_{aa}+\mu_{ba}}. \end{aligned} \tag{B.37}$$

Integrating and dividing through by $M(t)$, we recover the expression (4.20) for the pair correlation given in Chapter 4.

B.5 Parameters used for mixed aggregation with primary processes only

The following parameters were used to generate all plots of mixed fibril aggregation without the presence of secondary processes:

Parameters	**units**	**Fig. 4.3(a)**	**Fig. 4.3(b)**	**Fig. 4.3(c)**	**Fig. 4.3(d)**
$M_a(0)$	µM	0	0	4.17	4.17
$M_b(0)$	µM	0	0	9.52	9.52
$P_a(0)$	p M	0	0	41.7	41.7
$P_b(0)$	p M	0	0	95.2	95.2
$m_a(0)$	µM	33.3	33.3	29.2	29.2
$m_b(0)$	µM	66.7	66.7	57.1	57.1
$k_n(aa)$	$\times 10^{-5}$ $\mathrm{M^{-1}s^{-1}}$	1	1	1	1
$k_n(ab)$	$\times 10^{-5}$ $\mathrm{M^{-1}s^{-1}}$	6	6	6	6
$k_n(bb)$	$\times 10^{-5}$ $\mathrm{M^{-1}s^{-1}}$	1	1	1	1
$k_+(a\|a)$	$\times 10^{6}$ $\mathrm{M^{-1}s^{-1}}$	4	4	4	4
$k_+(a\|b)$	$\times 10^{6}$ $\mathrm{M^{-1}s^{-1}}$	6	6	6	6
$k_+(b\|a)$	$\times 10^{6}$ $\mathrm{M^{-1}s^{-1}}$	12	12	12	12
$k_+(b\|b)$	$\times 10^{6}$ $\mathrm{M^{-1}s^{-1}}$	3.6	0.8	3.6	0.8

Table B.1: Rate constants and initial conditions used to generate plots in Fig.4.3

B.6 Parameters used to investigate mixed aggregation with secondary processes

The following parameters were used to generate all plots of mixed fibril aggregation in the presence of secondary processes:

Parameters	**units**	**Fig. 4.5(a)**	**Fig. 4.5(b)**
$M_a(0)$	µM	0	41.7
$M_b(0)$	µM	0	95.2
$P_a(0)$	p M	0	417
$P_b(0)$	p M	0	952
$m_a(0)$	µM	333	292
$m_b(0)$	µM	667	571
$k_n(aa)$	$\times 10^{-8}$ M^{-1}s^{-1}	1	1
$k_n(ab)$	$\times 10^{-8}$ M^{-1}s^{-1}	4	4
$k_n(bb)$	$\times 10^{-8}$ M^{-1}s^{-1}	1	1
$k_+(a\|a)$	$\times 10^{5}$ M^{-1}s^{-1}	1	1
$k_+(a\|b)$	$\times 10^{5}$ M^{-1}s^{-1}	1.5	1.5
$k_+(b\|a)$	$\times 10^{5}$ M^{-1}s^{-1}	1.3	1.3
$k_+(b\|b)$	$\times 10^{5}$ M^{-1}s^{-1}	0.9	0.9
$k_-(aa)$	$\times 10^{-8}$ s^{-1}	1	1
$k_-(ab)$	$\times 10^{-8}$ s^{-1}	1	1
$k_-(bb)$	$\times 10^{-8}$ s^{-1}	1	1

Table B.2: Rate constants and initial conditions used to generate plots in Figs. 4.5 and 4.6

Appendix C

Oligomeric Intermediates Initiate Fibril Formation by the Yeast Prion Protein Ure2

This appendix contains additional details of the analytical methodology I employed in Chapter 5.

C.1 Additional Modelling

C.1.1 Detailed description of the kinetic model

The experimentally-measurable quantities that we seek to model are the total oligomer concentration $O(t)$, and the concentration of dimers that have been incorporated into fibrils, called the fibril "mass concentration" $M(t)$. We have established the main reaction processes acting on these quantities in Chapter 5: oligomer formation from dimers; oligomer dissociation back to dimers; oligomer conversion into fibrillar aggregates; elongation of fibrillar aggregates through dimer addition; and fragmentation of fibrils into smaller fibrils. We model the total oligomer concentration despite the existence of two oligomer sub-populations because, as explained in the Methods section of Chapter 5, our data on sub-populations of oligomers is of insufficient completeness and accuracy for kinetic modelling. Our model is thus partially coarse-grained. This modelling approach is rigorously justified in the SI section "Interpreting the kinetic model", in which we also demonstrate that the effects of the coarse-graining are largely limited to the "conversion" step, which contains information on the inter-conversion of distinct oligomer species as well as their ultimate conversion to growing fibrils. This is analogous to the coarse-grained "primary nucleation" step in traditional bulk models of fibrillar growth that actually contains information on all of the oligomeric reactions that lead to new fibril formation.

We note for full rigor here that we have ignored depolymerization, as at equilibrium the dimer concentration is very low (less than 2% as calculated from the single molecule burst density of donor-labeled Ure2 after the aggregation reaction reached plateau) and thus this process does not significantly affect the kinetics. We have also ignored fibril annealing, as it has a negligible effect on the time profile of the measured quantities $M(t)$ and $O(t)$. Furthermore, we may ignore the changes in dimer concentration caused directly by the oligomer formation and depletion processes, as these are insignificant relative to the changes due to fibril elongation, due to the low incidence of oligomers relative to native Ure2 dimers. According to the law of mass action, the rate of a reaction is proportional to the product of the concentrations of the species involved. Neglecting certain other small terms [36], we can therefore write the rate equations for our aggregating system, outlined in Chapter 5 (Eq.1-4):

$$\frac{dO}{dt} = k_{\text{oligo}} m(t)^2 - k_c O(t) - k_d O(t) \tag{C.1}$$

$$\frac{dP}{dt} = k_c O(t) + k_- M(t) \tag{C.2}$$

$$\frac{dM}{dt} = 2k_+ m(t) P(t) \tag{C.3}$$

$$\frac{dm}{dt} = -2k_+ m(t) P(t) \tag{C.4}$$

In order to describe the kinetics of $M(t)$ and $O(t)$ with a closed set of rate equations, it was necessary to also include an explicit rate equation for the fibril concentration $P(t)$, and for the dimer concentration $m(t)$. This is because oligomer formation from dimers depends in general on dimer concentration, and fibril elongation by dimer addition will depend explicitly on fibril concentration and dimer concentration.

Each process of importance is now represented by a rate term in the above closed set of rate equations: fibril elongation by $2k_+ m(t)P(t)$; fibril fragmentation (proportional to the total concentration of potential break sites, and therefore to the fibril mass concentration) by $k_- M(t)$; oligomer dissociation by kd O(t); oligomer conversion by $k_c O(t)$; and oligomer formation by $k_{\text{oligo}} m(t)^2$.

C.1.2 Choosing reaction orders

We do not know the reaction orders of the oligomer formation and conversion processes with respect to dimer concentration for certain *a priori*. We can see from the shape of the oligomer concentration curves that the reaction order of oligomer formation, n_o, with respect to dimers must be at least 1, and unlikely to exceed 5; however, since our dataset involves only one initial dimer concentration, we cannot more accurately determine it from fitting. Including the reaction order as a free parameter in the model would therefore be overfitting, and we must choose a value before proceeding with the data fitting. We believe 2 is a physically reasonable value given that we expect that all oligomers must either arise from the initial interaction of a pair of dimeric Ure2 molecules (i.e. formation of a "dimer-of-dimers"), or have grown from dimers-of-dimers through dimer addition, so the only reaction that forms new oligomers is the dimer-of-dimer formation reaction. In fact, it is hard to physically justify a reaction order substantially differing from this.

Any error in our choice of reaction order relative to the "true" value is expected to have no qualitative effect on the modelling conclusions, for the same reason as why we were forced to choose an a priori value in the first place – for this limited dataset, all physical choices of reaction order will result in similar fit quality and similar parameters. This is provided we realize that with a varying reaction order we must compare values of $k_{\text{oligo}}m(0)^{n_o}$, not k_{oligo}.

By fitting to a single initial dimer concentration it is not possible to determine the reaction order accurately for the oligomer conversion step either. In fact, the conversion reaction order, n_p, is even less visible from analysis of datasets with a single initial dimer concentration, since conversion is of principal importance early in the reaction before significant dimer depletion has occurred. We must therefore again choose a reaction order with respect to dimers; we have chosen zero. As before, the value chosen has no effect on the modelling, beyond a redefinition of the rate constant k_c, provided it is understood that the relevant fitting parameter to compare is really $k_c m(0)^{n_p}$ and not simply k_c.

We verified the insensitivity of the modelling of our available data to these reaction orders by carrying out the fitting procedure with alternative values for the reaction orders, and examining the variability of the fitted rate parameters. For oligomer formation we expect that the only physically reasonable alternatives are between 1 and 2 (as would be seen in the case of saturation effects). We think it is highly unlikely to be greater than 3; therefore, we trialled 1, 2, and 3. The initial slopes of the oligomer concentrations for each Ure2 mutant are very similar, making it extremely unlikely that their formation reaction orders are different; thus, we limit our analysis to considering identical reaction orders for each mutant.

In the case of conversion, we trialled reaction orders of 0, 1 and 3. This time we have no reason to suppose they are the same for each mutant. We find that, to within error, our key fitting results are totally unaffected by changes in conversion reaction order, even when the mutants have different reaction orders (recall that k_- and k_c can only be determined to within an order of magnitude). We further find that changes in oligomerization reaction order have only small quantitative effects on k_{oligo} and k_d for each mutant but do not affect their ratio to within error, or any of our other key conclusions. We are therefore satisfied that the precise values of these reaction orders are unimportant for our analysis. See Tables C.1-C.4 for full results of the sensitivity analysis.

C.1.3 Analytical solution for the aggregation kinetics

At early times, the dimer concentration can be considered to be approximately constant at its initial value $m(0)$ and our kinetic rate equations (C.1)-(C.3) reduce to a simpler form:

$$\frac{d}{dt}O^{(0)}(t) = \alpha - k_l O^{(0)}(t) \tag{C.5}$$

$$\frac{d}{dt}P^{(0)}(t) = k_c O^{(0)}(t) + k_- M^{(0)}(t) \tag{C.6}$$

$$\frac{d}{dt}M^{(0)}(t) = 2k_+ m(0) P^{(0)}(t), \tag{C.7}$$

where $k_l = k_c + k_d$, and $\alpha = k_{\text{oligo}} m(0)^2$. These can be solved for unseeded initial conditions to give early-time analytic expressions. We know from experiment that oligomers are present only at low concentration, and therefore that dissociation is relatively fast. This allows us to simplify our early-time analytical expressions to give:

$$O^{(0)}(t) = \frac{\alpha}{k_l}\left(1 - e^{-k_l t}\right) \tag{C.8}$$

$$P^{(0)}(t) = \frac{\alpha k_c}{2\kappa(k_l + \kappa)}\left(e^{\kappa t} - 1\right), \tag{C.9}$$

where $\kappa = \sqrt{2k_+ k_- m(0)}$. Now, using conservation of mass, $M(t) = m(0) - m(t)$, we rewrite our equation for fibril mass concentration as an integral equation for the dimer concentration:

$$m(t) = m(0)\exp\left(-2k_+ \int_0^t P(t')dt'\right). \tag{C.10}$$

Substituting our early-time expression $P^{(0)}(t)$ into the right-hand side, and again applying conservation of mass, yields a 1st order self-consistent expression for $M(t)$ of remarkable accuracy, in line with Ref. [36] and [27]:

$$m(t) = m(0)\left(1 - \exp\left(-A\left(e^{\kappa t} - 1\right)\right)\right), \tag{C.11}$$

with $A = \frac{\alpha k_+ k_c}{\kappa^2(k_l + \kappa)}$. This expression bears a striking resemblance to the expressions derived in Ref. [21] for the kinetics of fragmenting systems. This is to be expected given that oligomers do not comprise a large proportion of the system mass. Note that the parameters k_+, k_- and k_c cannot therefore be uniquely determined from fitting bulk aggregate mass concentration to this expression.

C.1.4 Interpreting the results of data fitting to the kinetic model

The model in Chapter 5 was selected because it is the simplest possible physically reasonable model capable of fitting the available combined smFRET/ThT experimental data with reasonable accuracy. This is in line with the principle of parsimony that underlies the theory of model selection [109]. The success of the fits demonstrates that, at this level of experimental detail, the full range of low- and high-FRET oligomers of all sizes, and the reactions that connect them, can be well-approximated by a single oligomer species undergoing the formation, dissociation and conversion processes currently incorporated in the model. This is inferred, not assumed. The inclusion of any finer detail in a kinetic model would lead to overfitting, degrading the quality of subsequent inferences made on the basis of the model. It would only be reasonable to test such models given a much larger dataset than is currently available.

Nonetheless, it is of interest to investigate how exactly we might interpret our coarse-grained model reaction processes in terms of more fundamental reaction steps expected to be present, and to what extent they might hold up as true constants in a more detailed study. We know from smFRET data [112] that there are in fact two distinct structures of oligomer, distinguished by their differing FRET efficiencies, and that high-FRET oligomers likely form from low-FRET oligomers and not directly from dimers. Therefore, the total oligomer formation step in the model is likely indeed a fundamental reaction step: that of low-FRET oligomer formation. We also know that at early and intermediate times, low-FRET oligomers dominate the population; thus, even if some high-FRET oligomers were to form directly from dimers, the fitted total oligomer formation rate constant would still be a very good approximation to the low-FRET oligomer formation rate constant.

To investigate what range of sizes oligomers might come in, a size distribution analysis was performed, that indicates that both low-FRET and high-FRET oligomer populations have average apparent sizes that change little during the majority of the time course of the aggregation reaction [112]; implying that the normalized size distribution of each population remains constant. Therefore, the time evolution of the full network of oligomeric species should be faithfully represented by a kinetic model which treats the low-FRET oligomers and the high-FRET oligomers each as a single kinetic species, since the coarse-grained rate constants should remain approximately constant throughout the reaction. The experiments on low/high-FRET oligomers indicate that the majority of oligomer depletion is low-FRET oligomer dissociation; we may therefore identify the total oligomer dissociation rate constant with the low-FRET oligomer dissociation rate constant.

We expect the coarse-grained total oligomer conversion step to contain information both on inter-conversion of low- and high-FRET oligomers, and the ultimate conversion of high-FRET oligomers to growing fibrils. To interpret it more quantitatively, we start by explicitly writing down the more detailed kinetic model that we believe governs the system:

$$\frac{dS_1}{dt} = k_{\text{oligo}} m(t)^2 - (k_{c1} + k_{d1}) S_1(t) \tag{C.12}$$

$$\frac{dS_2}{dt} = k_{c1} S_1(t) - (k_{c2} + k_{d2}) S_2(t) + k_L P(t) \tag{C.13}$$

$$\frac{dP}{dt} = k_{c2} S_2(t) + k_- M(t) \tag{C.14}$$

$$\frac{dM}{dt} = 2k_+ m(t) P(t); \qquad M(t) + m(t) = m(0) \tag{C.15}$$

where $k_{c1}S_1(t)$ is the rate of conversion of low-FRET to high-FRET oligomers (whose concentrations are given by $S_1(t)$ and $S_2(t)$ respectively); and $k_{c2}S_2(t)$ is the rate of conversion of high-FRET oligomers to fibrils; we have shown in Chapter 5 that these processes likely occur; and $k_{d1}S_1(t)$ and $k_{d2}S_2(t)$ are the rates of dissociation of low- and high-FRET oligomers, respectively. Fibril depolymerization experiments outlined in Chapter 5 imply that reverse conversion of high- to low-FRET oligomers is not a major process. The rate "constants" k_{c1} and k_{c2} may have dimer dependence (i.e. $k_{c1}m(t)^{n1}$). We show in Chapter 5 that the basal oligomer concentrations seen at the end of the aggregation reactions are likely due to disaggregation of high-FRET oligomers from fibrils; we represent this here with $k_L P(t)$.

The concentration of $S_2(t)$ changes very little through most of the reaction compared to the other species featuring in these equations, as well as being much lower than $S_1(t)$; thus, applying the steady state approximation and setting its rate of change to zero is reasonable. This gives us:

$$S_2(t) = \frac{k_{c1}}{k_{c2} + k_{d2}} S_1(t) + \frac{k_L}{k_{c2} + k_{d2}} P(t) \tag{C.16}$$

$$O(t) = S_1(t) + S_2(t) = \frac{k_{c1} + k_{c2} + k_{d2}}{k_{c1}} S_2(t) - \frac{k_L}{k_{c1}} P(t) \tag{C.17}$$

$$\frac{dP}{dt} = \frac{k_{c1}k_{c2}}{k_{c1} + k_{c2} + k_{d2}} O(t) + k_- M(t) + \frac{k_L k_{c2}}{k_{c1} + k_{c2} + k_{d2}} P(t). \tag{C.18}$$

If we compare to the coarse-grained equations in Chapter 5, we can identify the coarse-grained "k_c" in the steady-state limit to be approximately equal to $\frac{k_{c1}k_{c2}}{k_{c1}+k_{c2}+k_{d2}}$. This can be simplified given that $S_2(t) \ll S_1(t)$, and so $k_{c2} + k_{d2} \gg k_{c1}$. In fact, as we know most oligomers do not become fibrils, we likely have $k_{d2} \gg k_{c2}$, and so our interpretation of conversion becomes $k_c \simeq k_{c1}k_{c2}/k_{d2}$, i.e. the coarse-grained "conversion" rate constant is proportional to those for both fine-grained conversion steps. We can furthermore identify late-time underestimate of the total oligomer concentration data by the coarse-grained model as being due to the neglected term proportional to $P(t)$, as expected.

These results are essentially unchanged for non-zero conversion reaction orders - the rate constants k_{c1} and k_{c2} are simply replaced by $k_{c1}m(t)^{n1}$ and $k_{c2}m(t)^{n2}$. Given that conversion is of principal importance early in the reaction before significant dimer depletion has occurred, the reaction orders of conversion would be practically invisible from analysis of datasets involving only a single initial dimer concentration, even if we had separate datasets on low- and high-FRET oligomers of sufficient detail to model.

C.1.5 Bulk concentration-variable kinetic assay and analysis of Ure2 fibril formation

A global kinetic analysis (Fig. C.1) of Ure2 fibril formation at different concentrations (2.5 µM – 22.5 µM) was performed. At least three replicates were performed to check the reproducibility and 3–5 repetitions were used for global fitting analysis.

Kinetic analysis was performed with the assistance of an online fitting platform, AmyloFit [16]. In brief, the ThT curves of amyloid formation for Ure2 were normalized, and the time to half-completion ($t_{1/2}$) of each curve was plotted against initial concentration of Ure2 according to the power-law, the slope of which gave the scaling exponent γ. Here, we used the dimer concentration of Ure2 because Ure2 seldom dissociates into monomers either in its native state or during fibril formation, as indicated both by previous studies [268, 269] and our intra-dimer FRET experiment [112]. The slopes of the half-time plots are close to 0.5 (0.64 for WT Ure2, 0.57 for S68C and 0.60 for V9C as shown in Fig. S4A), suggesting a fragmentation-dominant mechanism [77]. The concentration-variable ThT curves were therefore fitted globally to a fragmentation-dominant model [21] to obtain the kinetic parameters for fibril formation of WT, Ure2-S68C and V9C, allowing comparison of the differences between them (Fig. C.1B-D). The normalized ThT data were fitted globally to an analytical solution of the kinetics of breakable filament assembly [21, 30]. In this framework, the evolution of the fibril mass concentration is given in terms of the rate constants as a double exponential form as shown:

$$\frac{M(t)}{m_{\text{tot}}} = 1 - \exp\left(-C_+ e^{\kappa t} + C_- e^{-\kappa t} + \frac{\lambda^2}{\kappa^2}\right) \tag{C.19}$$

where $M(t)$ is the concentrations of protein in fibrillar form at time t, m_{tot} is the total concentration of Ure2 and the constants, C_+ and C_-, are fixed by the initial conditions.

$$C_\pm = \pm\frac{\lambda^2}{2\kappa^2} \tag{C.20}$$

The λ and κ are two combined kinetic parameters related to the primary nucleation rate k_n, fibril elongation rate k_+ and fibril fragmentation rate k_-.

$$\lambda = \sqrt{2k_+k_n m(0)^{n_c}} \tag{C.21}$$

$$\kappa = \sqrt{2k_+k_- m(0)} \tag{C.22}$$

By globally fitting the normalized ensemble kinetics data, the combined parameters k_+k_n and k_+k_- were obtained for each mutant and shown in the figures. k_n/k_- could be obtained by the ratio of the above two parameters. As the fragmentation rate of the two mutants under the same shaking conditions was the same (see Fig. C.2 and below), the ratios of the nucleation rates k_n(S68C)/k_n(V9C) and the elongation rates k_+(S68C)/k_+(V9C) were calculated and compared. The results show that the elongation rate of Ure2-S68C is 2.3-fold greater than that of V9C, and the amyloid nucleation rate of Ure2-S68C is about 2-fold greater than that for V9C under these experimental conditions.

C.1.6 Analysis of cross-seeded bulk experiments

Bulk cross-seeding experiments (Fig. C.2) were carried out at 18 ℃, in order to investigate whether the difference observed in k_+k_- between the two mutants is due to a difference in k_+. First a batch of S68C seeds were produced and used to seed a solution of native S68C and V9C protein separately, and then the same was done with V9C seeds (Fig. C.2). The seed concentration was high enough that the initial slope was controlled by elongation. The normalized ThT data were fitted globally to the same model as the unseeded data, but with a non-zero initial fibril concentration, to yield the values of $k_+P(0)$ for each dimer-seed combination, where $P(0)$ is the initial seed fibril concentration. We cannot calculate $P(0)$ exactly, but we can assume its value is the same for reactions using the same seed batch. Therefore the ratios of $k_+P(0)$ for each seed type give us the ratios of the elongation rate constants onto each seed. These ratios were approximately 1.3–1.8:1 for S68C:V9C, added to S68C seeds; and 1:1 for adding to V9C seeds. Had these ratios been the same, this would have provided strong evidence that the elongation rates were seed-independent and that this ratio was therefore also the ratio of self-elongation rates. However, we instead have evidence that elongation is seed-dependent. Furthermore, these results suggest that the elongation rate for S68C is larger than for V9C (note that this is not a conclusive result, as it is technically possible that the rate of addition of V9C dimer onto S68C seed fibril is slower than that of S68C dimer onto V9C seed fibril).

An estimate of the average length of the seed fibrils (100 nm) further allowed us to obtain an order-of-magnitude estimate for $P(0)$, and therefore the absolute elongation rates (approximately $k_+ = 40\ \mu\mathrm{M}^{-1}\mathrm{h}^{-1}$), and therefore also of the fragmentation and nucleation rates (see Chapter 5). Theoretical analysis shows that these rates are consistent with the formation of extended fibrils.

C.1.7 Determining fragmentation rates from analysis of fibril length distributions

Fibril length distributions were measured at different time points after completion of the aggregation reaction, as detailed in the above section. From these distributions, the average fibril length was calculated at each time point. Knowing the dimensions of the dimers (a = 1 nm) allows us to convert this to the average number of dimers per fibril, $L(t) = M(\text{final})/P(t)$. Since we know all dimers are converted to fibrils, $M(\text{final}) = m_{\text{tot}} = 15$ µM, therefore, we can estimate $P(t)$ from the measured average lengths as $P(t) = m_{\text{tot}}/L(t)$. Examining the rate equation for $P(t)$ (Eq. (4) in Chapter 5), we find that after aggregation is completed $dP/dt = k_{-}M(\text{final})$. Therefore, the quantity $k_{-}M(\text{final})$ is simply the gradient obtained from fitting $P(t)$ vs t to a straight line. We have carried out this procedure for both mutants (see Fig. C.3A and B), determining that k_{-}, averaged over the two sets of measurements is $k_{-} = (5.1 \pm 2.3) \times 10^{-5}$ h^{-1} for V9C and $k_{-} = (5.4 \pm 0.5) \times 10^{-5}$ h^{-1} for S68C. The difference between these two values is within experimental error, thus justifying our assumption in the main modelling that k_{-} is identical for the two mutants.

C.2 Supplementary Tables

	Fitted Parameter Values			Precision
	$n_o = 2$ (default)	$n_o = 1$	$n_o = 3$	
k_{-} /h^{-1}	1.2×10^{-4}	1.4×10^{-4}	1.2×10^{-4}	OM[a]
$k_{\text{oligo}}m(0)^{n_o}$ /µM h^{-1}	3.6×10^{-2}	4.6×10^{-2}	2.9×10^{-2}	$\pm 0.3 \times 10^{-2}$
k_c (S68C) /h^{-1}	2.3×10^{-3}	2.0×10^{-3}	2.3×10^{-3}	OM[a]
k_c (V9C) /h^{-1}	1.7×10^{-3}	1.5×10^{-3}	1.6×10^{-3}	OM[a]
k_c ratio	1.4	1.3	1.4	±0.3
k_d (S68C) /h^{-1}	0.60	0.89	0.47	±0.08
k_d (S68C) /h^{-1}	0.45	0.67	0.36	±0.05
k_d ratio	1.3	1.3	1.3	±0.3
Fitting error /MRE[b]	1.6	2.1	1.4	

Table C.1: Sensitivity analysis for varying oligomer formation reaction orders. [a] OM, order-of-magnitude precision. [b] MRE, mean residual error.

	Fitted Parameter Values			Precision
	n_p (V9C) = 0	n_p (V9C) = 1	n_p (V9C) = 3	
k_{-} /h^{-1}	1.2×10^{-4}	1.2×10^{-4}	1.3×10^{-4}	OM[a]
k_{oligo} /µM^{-1}h^{-1}	1.6×10^{-4}	1.6×10^{-4}	1.6×10^{-4}	$\pm 0.1 \times 10^{-4}$
$k_c m(0)^{n_p}$ (S68C) /µM h^{-1}	2.3×10^{-3}	2.2×10^{-3}	2.1×10^{-3}	OM[a]
$k_c m(0)^{n_p}$ (V9C) /µM h^{-1}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	OM[a]
$k_c m(0)^{n_p}$ ratio	1.4	1.3	1.2	±0.3
k_d (S68C) /h^{-1}	0.60	0.60	0.61	±0.08
k_d (S68C) /h^{-1}	0.45	0.45	0.45	±0.05
k_d ratio	1.3	1.3	1.3	±0.3
Fitting error /MRE[b]	1.6	1.6	1.6	

Table C.2: Sensitivity analysis for varying V9C oligomer conversion reaction orders n_p (V9C), with S68C conversion reaction order n_p (S68C) = 0. [a] OM, order-of-magnitude precision. [b] MRE, mean residual error.

	Fitted Parameter Values			Precision
	n_p (V9C) = 0	n_p (V9C) = 1	n_p (V9C) = 3	
k_- /h^{-1}	1.2×10^{-4}	1.2×10^{-4}	1.3×10^{-4}	OM[a]
k_{oligo} /µM^{-1}h^{-1}	1.6×10^{-4}	1.6×10^{-4}	1.6×10^{-4}	$\pm 0.1 \times 10^{-4}$
$k_c m(0)^{n_p}$ (S68C) /µM h^{-1}	2.4×10^{-3}	2.3×10^{-3}	2.2×10^{-3}	OM[a]
$k_c m(0)^{n_p}$ (V9C) /µM h^{-1}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	OM[a]
$k_c m(0)^{n_p}$ ratio	1.5	1.4	1.3	±0.3
k_d (S68C) /h^{-1}	0.59	0.60	0.60	±0.08
k_d (S68C) /h^{-1}	0.45	0.45	0.45	±0.05
k_d ratio	1.3	1.3	1.3	±0.3
Fitting error /MRE[b]	1.6	1.6	1.6	

Table C.3: Sensitivity analysis for varying V9C oligomer conversion reaction orders n_p (V9C), with S68C conversion reaction order n_p (S68C) = 1. [a] OM, order-of-magnitude precision. [b] MRE, mean residual error.

	Fitted Parameter Values			Precision
	n_p (V9C) = 0	n_p (V9C) = 1	n_p (V9C) = 3	
k_- /h^{-1}	1.2×10^{-4}	1.2×10^{-4}	1.3×10^{-4}	OM[a]
k_{oligo} /µM^{-1}h^{-1}	1.6×10^{-4}	1.6×10^{-4}	1.6×10^{-4}	$\pm 0.1 \times 10^{-4}$
$k_c m(0)^{n_p}$ (S68C) /µM h^{-1}	2.7×10^{-3}	2.5×10^{-3}	2.4×10^{-3}	OM[a]
$k_c m(0)^{n_p}$ (V9C) /µM h^{-1}	1.7×10^{-3}	1.7×10^{-3}	1.7×10^{-3}	OM[a]
$k_c m(0)^{n_p}$ ratio	1.6	1.5	1.4	±0.3
k_d (S68C) /h^{-1}	0.59	0.59	0.60	±0.08
k_d (S68C) /h^{-1}	0.45	0.45	0.45	±0.05
k_d ratio	1.3	1.3	1.3	±0.3
Fitting error /MRE[b]	1.6	1.6	1.6	

Table C.4: Sensitivity analysis for varying V9C oligomer conversion reaction orders n_p (V9C), with S68C conversion reaction order n_p (S68C) = 3. [a] OM, order-of-magnitude precision. [b] MRE, mean residual error.

C.3 Supplementary Figures

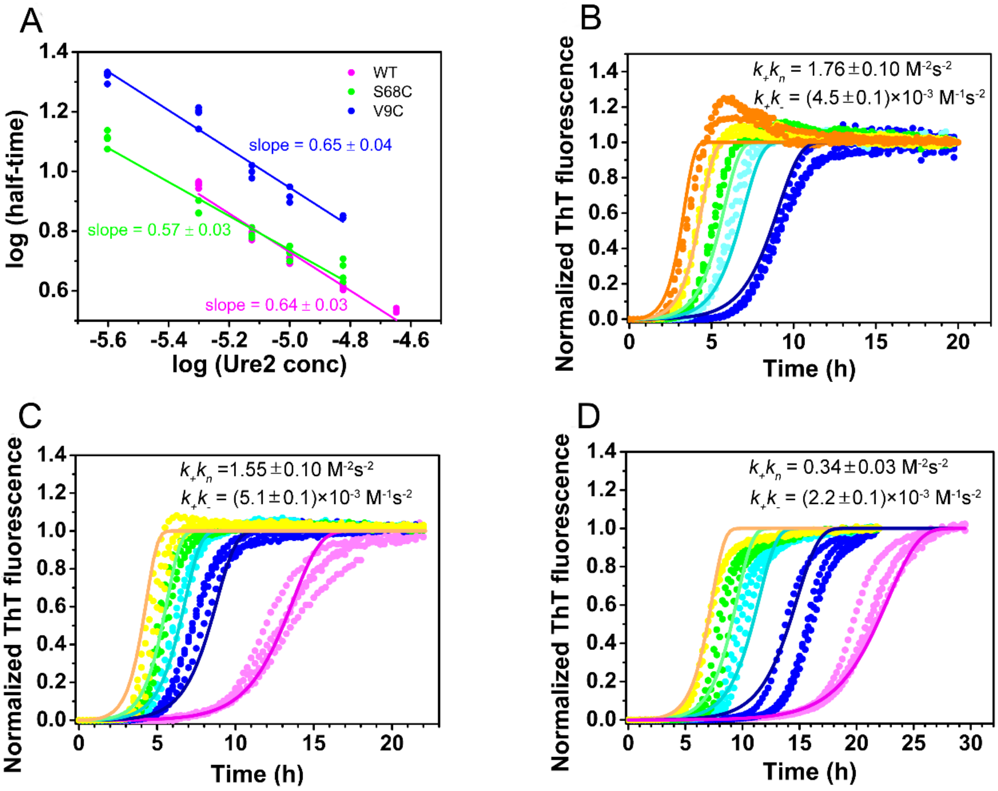

Figure C.1: **Global fitting of the kinetics of fibril formation of WT Ure2 and its variants.** (**A**) Power-law scaling plot of half-time versus initial dimer concentration of WT (red), Ure2-S68C (green), Ure2-V9C (blue). The slopes give the scaling exponent of each protein. (**B**-**D**) The normalized ThT fluorescence curves of five concentrations of (**B**) WT Ure2, (**C**) Ure2-S68C and (**D**) Ure2-V9C were globally fitted using the analytical solution with a fragmentation-dominant mechanism (for details see SI Text). The concentrations shown in (**B**-**D**) are 2.5 µM (pink), 5 µM (blue), 7.5 µM (cyan), 10 µM (green), 15 µM (yellow), and 22.5 µM (orange). The ThT assay was carried out in a Fluostar Omega plate reader (BMGLabtech) at 30 °C with 200 rpm orbital shaking. Three to five repetitions were analyzed for each concentration.

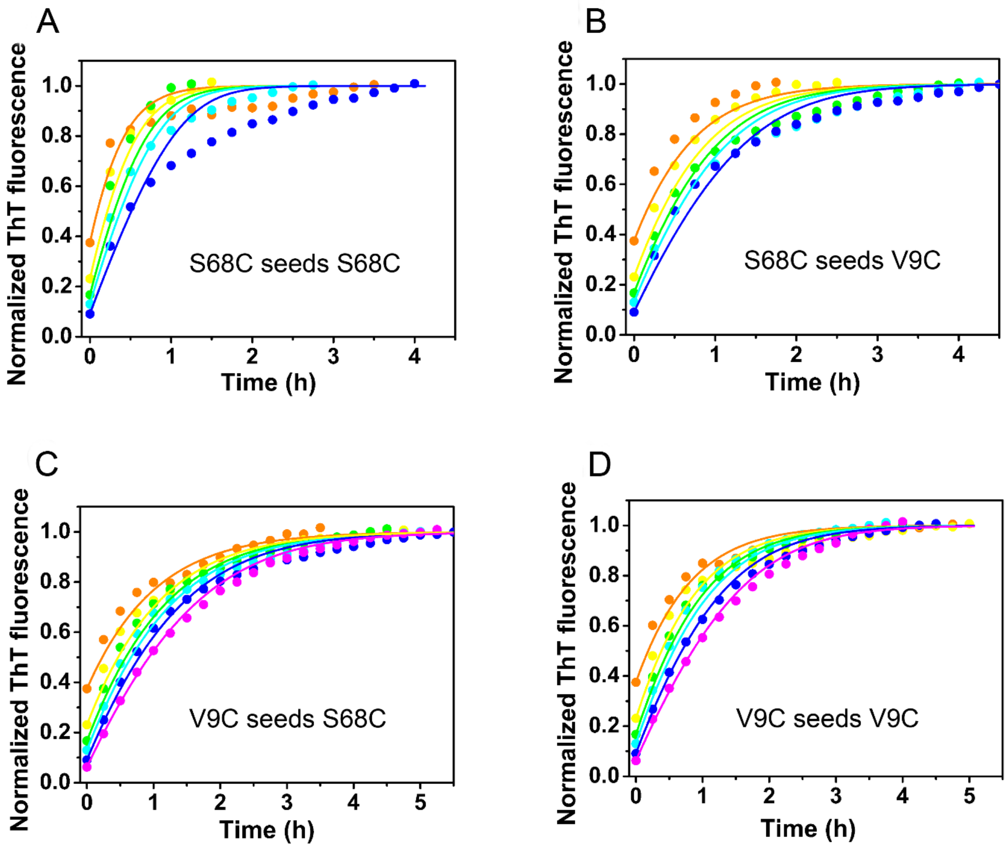

Figure C.2: **Fibril formation kinetics of self-seeding and cross-seeding of Ure2 variants monitored by ThT fluorescence.** Solution of different concentrations of Ure2-S68C or Ure2-V9C, 2.5 µM (orange), 5 µM (yellow), 7.5 µM (green), 10 µM (cyan), 15 µM (blue), 22.5 µM (pink), were incubated with 1.5 µM fibril seeds of Ure2-S68C or Ure2-V9C as indicated in the figure. The seeding experiments were carried out in a Fluostar Omega plate reader (BMGLabtech) at 18 °C with 200 rpm orbital shaking. The data were globally fitted to the saturation-elongation model to obtain the ratio of the initial elongation rate constants for the two protein variants and an estimation of the individual elongation rate constants.

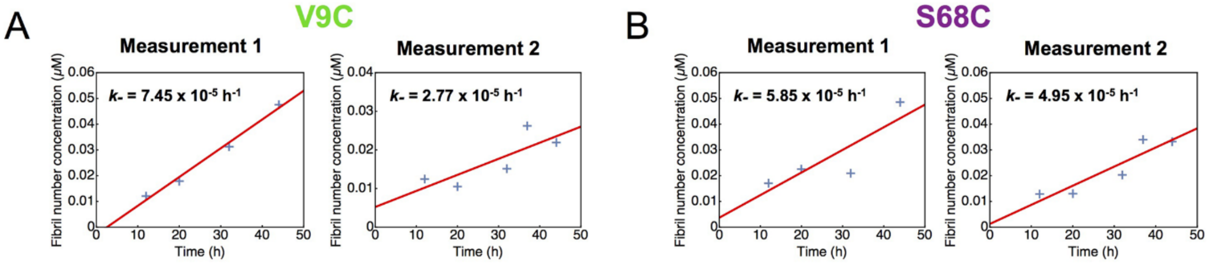

Figure C.3: **Kinetic analysis of fibril length distributions to determine fragmentation rates.** (**A**,**B**) Time evolution of fibril concentration P for each protein variant, as determined from the average length, is fitted to a straight line, whose slope yields the fragmentation rate constant (see ***Determining fragmentation rates from analysis of fibril length distributions*** above).

Appendix D

Classifying Amyloid Oligomers Using Chemical Kinetics

The material in this appendix supports Chapter 6. It contains derivations, supporting theoretical results, fitted rate constants and other quantities. All work is my own apart from the early-time solutions for secondary oligomers, which were carried out by T. C. T. Michaels; and the simulations, which were performed by Anđela Šarić.

D.1 Rate equations and early-time solutions for oligomer-mediated filament assembly

Here, we first obtain general rate equations for filament assembly via oligomeric intermediate species that explicitly include oligomer dissociation and optionally include secondary growth processes such as fragmentation and secondary nucleation. We then solve them for early times, before significant monomer has been depleted. We consider only one oligomeric species; in principle there may be sub-populations, but in many cases such systems can be usefully modelled assuming the oligomer population to be uniform, and valuable insights obtained. We note that the methodology generalizes should we wish to explicitly consider additional oligomeric intermediate species.

In our models, free monomers have concentration $m(t)$, and monomers incorporated into fibrils have concentration $M(t)$. Fibrils can elongate with rate constant k_+. Oligomers (concentration $S(t)$), once formed, may subsequently convert to growth-competent fibril-type species (concentration $P(t)$) with a rate constant k_{conv}, and an optional reaction order with respect to monomers of n_{conv}, or dissociate back to monomers with rate constant k_d. We ignore depolymerisation throughout, as at equilibrium monomer concentration is very low in the systems studied and thus this process does not significantly affect the kinetics. We further ignore filament annealing, as it has a negligible effect on the time profile of the measured quantities $M(t)$ and $S(t)$.

D.1.1 Primary oligomers

We first consider systems in which oligomers are formed only as intermediates of primary nucleation. Secondary processes may be present, but do not contribute directly to oligomer formation.

Rate equations

Oligomers are formed only through a nucleation process (occurring with rate constant k_{o1} and reaction order n_{o1}). Neglecting certain small terms, we can then write rate equations describing the system:

$$\frac{dS}{dt} = k_{\mathrm{o1}} m(t)^{n_{\mathrm{o1}}} - (k_{\mathrm{conv}} m(t)^{n_{\mathrm{conv}}} + k_d) S(t) \tag{D.1a}$$

$$\frac{dP}{dt} = k_{\mathrm{conv}} m(t)^{n_{\mathrm{conv}}} S(t) + k_2 m(t)^{n_2} M(t) \tag{D.1b}$$

$$\frac{dM}{dt} = 2k_+ m(t) P(t) \tag{D.1c}$$

$$\frac{dm}{dt} = -2k_+ m(t) P(t), \tag{D.1d}$$

When the secondary rate constant $k_2 = 0$, we have no secondary processes. Where both k_2 and n_2 are non-zero, we have secondary nucleation occurring on fibril surfaces (but not contributing significantly to measured oligomer populations). We can also model fragmentation with these equations by setting $k_2 = k_-$ and $n_2 = 0$. We finally define $\rho_c(t) = k_{\mathrm{conv}} m(t)^{n_{\mathrm{conv}}}$, and $\rho_e(t) = \rho_c(t) + k_d$.

Solutions for early times

At early times, the monomer concentration can be considered to be approximately constant at its initial value $m(0)$ and we have a reduced set of equations:

$$\frac{d}{dt} S^{(0)}(t) = \alpha_1 - k_e S^{(0)}(t) \tag{D.2a}$$

$$\frac{d}{dt} P^{(0)}(t) = k_c S^{(0)}(t) + k_2 m(0)^{n_2} M^{(0)}(t) \tag{D.2b}$$

$$\frac{d}{dt} M^{(0)}(t) = 2k_+ m(0) P^{(0)}(t), \tag{D.2c}$$

where $\alpha_1 = k_{\mathrm{o1}} m(0)^{n_{\mathrm{o1}}}$, $k_e = \rho_e(0)$, and $k_c = \rho_c(0)$. These can be solved for unseeded initial conditions to give early-time analytic expressions. These are given by:

$$S^{(0)}(t) = \frac{\alpha_1}{k_e}(1 - e^{-k_e t}) \tag{D.3a}$$

$$P^{(0)}(t) = \frac{\alpha_1 k_c}{2\kappa(k_e + \kappa)} e^{\kappa t} + \frac{\alpha_1 k_c}{2\kappa(\kappa - k_e)} e^{-\kappa t} + \frac{\alpha_1 k_c}{(k_e^2 - \kappa^2)} e^{-k_e t}, \tag{D.3b}$$

where $\kappa = \sqrt{2k_+ k_2 m(0)^{n_2+1}}$.
This is identical in form to the early-time solution for $P(t)$ in the equivalent model without oligomeric intermediates [30], apart from the final component proportional to k_e. When κ is non-vanishing, this component vanishes in comparison to the leading-order term when

$$\begin{aligned} e^{(\kappa+k_e)t} &\gg \left|\frac{2\kappa}{k_e-\kappa}\right| \\ (\kappa+k_e)t &> \ln(2) \text{ (taking the largest possible value of the RHS)} \\ \therefore t &> (\kappa+k_e)^{-1}. \end{aligned}$$

In this (usually-attained) limit, ignoring rapidly-decaying components gives:

$$S^{(0)}(t) = \frac{\alpha_1}{k_e}(1-e^{-k_e t}) \tag{D.4a}$$

$$P^{(0)}(t) = \frac{\alpha_1 k_c}{2\kappa(k_e+\kappa)}(e^{\kappa t}-1). \tag{D.4b}$$

On the other hand, when $k_2 = 0$ such that minimal secondary processes are present, Eqs (D.3a) and (D.3b) reduce to:

$$S^{(0)}(t) = \frac{\alpha_1}{k_e}(1-e^{-k_e t}) \tag{D.5a}$$

$$P^{(0)}(t) = \frac{\alpha_1 k_c}{k_e^2}(e^{-k_e t}+k_e t-1), \tag{D.5b}$$

and we see that the solution for $P(t)$ is functionally identical to that for the direct-nucleation model [36] in the limit $t \gg k_e^{-1}$. From Eq. (D.5a) we see that this is merely the steady-state condition for oligomers. We also note that if this steady-state condition is satisfied, then trivially the condition for simplified secondary processes kinetics $t > (\kappa+k_e)^{-1}$ is also satisfied.

We later derive full-time solutions for $M(t)$ from a self-consistent relation Eq. (D.23b) that depends only on k_+, $m(0)$ and $P^{(0)}(t)$. In the limits discussed above, the solutions for the fibril mass concentration are therefore identical in form to those obtained in prior bulk studies that consider filament formation via single-step nucleation [21]. These limits can then be interpreted as adjustment timescales, indicating the time that must elapse before the kinetics of oligomer-mediated fibril formation become identical to those of fibril formation without intermediates.

D.1.2 Secondary oligomers

We now consider the case in which secondary nucleation occurs via significant levels of oligomeric intermediates ("secondary oligomers"). The early-time equations have a substantially different form, necessitating their separate treatment.

Rate equations

With the rate of primary oligomer formation given by $k_{\mathrm{o}1}m(t)^{n_{\mathrm{o}1}}$, and the rate of secondary oligomer formation given by $k_{\mathrm{o}2}m(t)^{n_{\mathrm{o}2}}M(t)$, we may write down the moment equations:

$$\frac{dS(t)}{dt} = k_{\mathrm{o}1}m(t)^{n_{\mathrm{o}1}} + k_{\mathrm{o}2}m(t)^{n_{\mathrm{o}2}}M(t) - [k_{\mathrm{conv}}m(t)^{n_{\mathrm{conv}}} + k_d]S(t) \tag{D.6}$$

$$\frac{dP(t)}{dt} = k_{\mathrm{conv}}m(t)^{n_{\mathrm{conv}}}S(t) \tag{D.7}$$

$$\frac{dM(t)}{dt} = 2k_+m(t)P(t). \tag{D.8}$$

Solutions for early times

At early times, the monomer concentration can again be considered to be approximately constant at its initial value $m(0)$ and we have a reduced set of equations:

$$\frac{dS^{(0)}(t)}{dt} = \alpha_1 + \alpha_2 M^{(0)}(t) - k_e S^{(0)}(t) \tag{D.9}$$

$$\frac{dP^{(0)}(t)}{dt} = k_c S^{(0)}(t) \tag{D.10}$$

$$\frac{dM^{(0)}(t)}{dt} = 2k_+m(0)P^{(0)}(t), \tag{D.11}$$

where $\alpha_2 = k_{\mathrm{o}2}m(0)^{n_{\mathrm{o}2}}$. These may be solved (see TCTM); the dominant terms of the solution are:

$$S^{(0)}(t) = \frac{\alpha_1}{3\bar{\kappa}}\left(e^{\bar{\kappa}t} - 1\right) \tag{D.12}$$

$$P^{(0)}(t) = \frac{\alpha_1 k_c}{3\bar{\kappa}^2}\left(e^{\bar{\kappa}t} - 1\right) \tag{D.13}$$

$$M^{(0)}(t) = \frac{\bar{\lambda}^3}{3\bar{\kappa}^3}\left(e^{\bar{\kappa}t} - 1\right), \tag{D.14}$$

where

$$\bar{\lambda} = \left(2k_{\mathrm{conv}}k_+k_{\mathrm{o}1}m(0)^{n_{\mathrm{o}1}+n_{\mathrm{conv}}}\right)^{1/3}, \text{ and} \tag{D.15}$$

$$\bar{\kappa} = \left(2k_{\mathrm{conv}}k_+k_{\mathrm{o}2}m(0)^{n_{\mathrm{o}2}+n_{\mathrm{conv}}+1}\right)^{1/3}. \tag{D.16}$$

Comparison to bulk models

In bulk models, the early-time solutions have the same functional forms, but instead of $\bar{\kappa}$ we have $\kappa = (2k_+k_2m(0)^{n_2+1})^{1/2}$. These parameters must be equivalent; equating them we can resolve the bulk secondary nucleation rate parameters as:

$$k_2 = \frac{(k_{\mathrm{conv}} k_{\mathrm{o2}})^{2/3}}{(2k_+)^{1/3}}, \text{ and} \tag{D.17}$$

$$n_2 = (2n_{\mathrm{o2}} + 2n_{\mathrm{conv}} - 1)/3. \tag{D.18}$$

Furthermore, in traditional bulk models we have the exponential prefactor $\lambda^2/(2\kappa^2)$, where $\lambda = \sqrt{2k_+ k_n m(0)^{n_c}}$. Simplifying the prefactors gives:

$$\frac{\lambda^2}{2\kappa^2} = \frac{2k_+ k_n m(0)^{n_c}}{4k_+ k_2 m(0)^{n_2+1}} = \frac{k_n}{2k_2} m(0)^{n_c - n_2 - 1} = \frac{k_n}{2k_2} m(0)^{n_c - (2n_{\mathrm{o2}} + 2n_{\mathrm{conv}} - 1)/3 - 1} \tag{D.19}$$

$$\frac{\bar{\lambda}^3}{3\bar{\kappa}^3} = \frac{2k_{\mathrm{conv}} k_+ k_{\mathrm{o1}} m(0)^{n_{\mathrm{o1}} + n_{\mathrm{conv}}}}{6k_{\mathrm{conv}} k_+ k_{\mathrm{o2}} m(0)^{n_{\mathrm{o2}} + n_{\mathrm{conv}} + 1}} = \frac{k_{\mathrm{o1}}}{3k_{\mathrm{o2}}} m(0)^{n_{\mathrm{o1}} - n_{\mathrm{o2}} - 1} \tag{D.20}$$

Thus, equating the two we may express the bulk primary nucleation parameters as:

$$k_n = \frac{2k_{\mathrm{o1}} k_{\mathrm{conv}}^{2/3}}{3(2k_+ k_{\mathrm{o2}})^{1/3}} \tag{D.21}$$

$$n_c = n_{\mathrm{o1}} + (2n_{\mathrm{conv}} - n_{\mathrm{o2}} - 1)/3. \tag{D.22}$$

D.2 Full-time solutions without secondary nucleation

Our approach for calculating solutions valid for all times is different for systems without secondary nucleation to those with secondary nucleation. Here, we follow the approach used in [21] to derive full-time expressions for $M(t)$ and $S(t)$, for systems featuring primary nucleation via oligomeric intermediates, elongation, and optionally fragmentation.

D.2.1 Solutions for fibril concentrations for all times

We first reformulate Eqs.(D.1c) and (D.1d) as an integral equation for $m(t)$:

$$\frac{d \ln m(t)}{dt} = -2k_+ P(t) \tag{D.23a}$$

$$M(t) = m(0)\left(1 - \exp\left[-2k_+ \int_0^t P(\tau) d\tau\right]\right) \tag{D.23b}$$

By substituting a suitable approximate expression for $P(t)$ in the right-hand side, we can derive a 1st-order self-consistent approximate expression for $M(t)$ [21]. For fragmentation, using Eq.(D.4b) gives an accurate solution:

$$M^{(1)}(t) = m(0)\left(1 - \exp\left[-\frac{\lambda^2}{2\kappa^2}(e^{\kappa t} - \kappa t - 1)\right]\right), \tag{D.24}$$

with

$$\lambda = \left(2k_+ k_{o1} \frac{k_{\rm conv}}{k_e + \kappa} m(0)^{n_{o1}+n_{\rm conv}}\right)^{1/2}. \tag{D.25}$$

D.2.2 Connecting to bulk models

This expression has the same functional form the expressions derived in ref. [21] for the kinetics of fragmenting systems. This is to be expected given the assumption that we have made that oligomers do not comprise a large proportion of the system mass. Comparing, we can firstly see that both expressions have the same dependence on the parameter κ. Secondly we can identify the bulk nucleation rate $k_n m(0)^{n_c}$ as:

$$k_n m(0)^{n_c} = k_{o1} \frac{k_{\rm conv}}{(k_{\rm conv} m(0)^{n_{\rm conv}} + k_d + \kappa)} m(0)^{n_{o1}+n_{\rm conv}}. \tag{D.26}$$

In the limit of fast conversion or at high enough concentrations, oligomerisation becomes the rate-limiting step, and we can identify $k_n = k_{o1}$, as expected (and also $n_c = n_{o1}$). At low enough concentration, we identify instead $n_c = n_{o1} + n_{\rm conv}$. We see immediately, therefore, that introducing an oligomeric intermediate permits the bulk nucleation reaction order to vary over a sufficiently large concentration range, a phenomenon that has been observed but not explained in bulk studies of aggregation. Finally, in the limit of fast dissociation, we identify $k_n = \frac{k_{o1}}{k_d} k_{\rm conv}$, in other words the conversion rate multiplied by the equilibrium constant for oligomerization; and $n_c = n_{o1} + n_{\rm conv}$. This is simply the pre-equilibrium oligomer concentration multiplied by the conversion rate.

The time-dependent evolution of $m(t)$ can be seen to approximately depend only on κ and λ, or $k_+ k_2$, n_2, $k_+ k_{o1} k_{\rm conv}$, $n_{o1} + n_{\rm conv}$ and k_e. In the right concentration regime, such that $k_{\rm conv} m(0)^{n_{\rm conv}} \simeq k_d + \kappa$, we may resolve k_e into its components; more typically, we expect instead $k_e \simeq k_d$.

For no fragmentation, the same approach leads to:

$$M^{(1)}(t) = m(0)\left(1 - \exp\left[-\frac{\lambda^2}{k_e^2}\left(1 - k_e t + \frac{1}{2}(k_e t)^2 - e^{-k_e t}\right)\right]\right). \tag{D.27}$$

D.2.3 Solution for oligomer concentrations for all times

We can now use our first-order self-consistent solution for fibril mass to calculate a second-order self-consistent solution for oligomer concentration. We re-formulate Eq. (D.1) as an integral equation:

$$S^{(2)}(t) = e^{-\int^t \rho_e(\tau) d\tau} \int_0^t e^{\int^t \rho_e(\tau) d\tau} k_{o1} m^{(1)}(t)^{n_{o1}} d\tau, \tag{D.28}$$

where $\rho_e(t) = k_{\rm conv} m^{(1)}(t)^{n_{\rm conv}} + k_d$. For a non-zero $n_{\rm conv}$ this is intractable. However, by assuming $\rho_e(t)$ to be slowly varying (as, indeed, it usually is) we may assume it to be constant for the purposes of integration, permitting the integral to be solved analytically. Firstly we approximate $\int^t \rho_e(\tau)d\tau = t\rho_e(t)$ in line with the above assumption. Using conservation of mass, substituting Eq. (D.24), and the substitution x $= e^{\rho_e \tau}$, we arrive at the solution:

$$S^{(2)}(t) = \frac{\alpha_1}{\kappa} e^z \left(e^{-\rho_e(t)t} \, \mathrm{Ei}\left[1 - z - \frac{\rho(t)}{\kappa}, z\right] - e^{z\kappa t} \, \mathrm{Ei}\left[1 - z - \frac{\rho(t)}{\kappa}, z e^{\kappa t}\right]\right), \tag{D.29}$$

where $z = n_{o1}\lambda^2/\kappa^2$, and $\mathrm{Ei}[\nu, z]$ is the generalized exponential integral function.

In the case of no secondary processes, we obtain:

$$S^{(2)}(t) = \alpha_1 \, e^{-\rho_e t} \int_0^t \exp\left[\rho_e \tau \ + B\left(e^{-k_e \tau} - \frac{1}{2}(k_e \tau)^2 + k_e \tau - 1\right)\right] d\tau, \tag{D.30}$$

where $B = n_{o1}\lambda^2/k_e^2$. Typically $B \ll 1$ so, for $k_e \tau \lesssim 1$, the integrand exponent is dominated by the first term, $\rho_e \tau$. Where $k_e \tau \gg 1$, the exponent is dominated by the non-exponential terms. Thus at all times it is typically acceptable to neglect the exponential term in the integrand exponent. Doing so allows us to evaluate the integral, giving:

$$S^{(2)}(t) = \frac{\alpha}{2}\sqrt{\frac{\pi}{g}} \exp\left[\frac{k^2}{4g} - B\right] e^{-k_e t} \left(\mathrm{erf}\left[\frac{k}{2\sqrt{g}}\right] + \mathrm{erf}\left[\frac{2\,g\,t - k}{2\sqrt{g}}\right]\right) \tag{D.31}$$

with $g = k_e^2 B/2$ and $k = k_e(1 + B)$. This simplifies to:

$$S^{(2)}(t) = \frac{\alpha}{k_e}\sqrt{\frac{\pi}{2B}} \exp\left[\frac{(1+B)^2}{2B} - B\right] e^{-k_e t} \left(\mathrm{erf}\left[\frac{1+B}{\sqrt{2B}}\right] + \mathrm{erf}\left[\frac{B k_e t - 1 - B}{\sqrt{2B}}\right]\right). \tag{D.32}$$

D.2.4 Controlling parameters

We see here that oligomer populations are to a good approximation controlled by α_1, κ, z and $\rho(t)$. In the usual case where $k_d \gg k_c$, $\rho(t)$ reduces to k_d. Thus essentially the same dimensionless combinations of rate parameters control $S(t)$ as control $m(t)$. However, $S(t)$ can be used to more easily resolve k_d from k_e.

D.3 Full-time solutions for systems featuring secondary nucleation

Systems featuring secondary nucleation are better treated with an alternative approach, as the self-consistent approach described in [21] is less accurate in this case. [30, 50].

D.3.1 Full-time expressions for $M(t)$

There is no single way to derive accurate approximate expressions for $M(t)$. We prefer the approach taken in Chapter 10, and write:

$$\frac{M(t)}{m(0)} = 1 - \frac{1}{(1 + M^{(0)}(t)/c)^c}, \tag{D.33}$$

where $M^{(0)}(t) = \frac{\lambda^2}{2\kappa^2}(e^{\kappa t} - 1)$ for secondary nucleation without secondary oligomers, and $M^{(0)}(t) = \frac{\bar{\lambda}^3}{3\bar{\kappa}^3}(e^{\bar{\kappa} t} - 1)$ for secondary nucleation via secondary oligomers. Furthermore, $c = 3/(2n_2 + 1)$. Where secondary oligomerization occurs, n_2 is given by Eq. (D.18); otherwise it is a fundamental parameter in the rate equations.

D.3.2 Full-time expressions for $S(t)$

Using Eq. (D.33) as an input, we may once more take a self-consistent approach to calculating expressions for $S(t)$.

No secondary oligomers

In the case of secondary nucleation without secondary oligomers, we may again use the self-consistent expression Eq. (D.28). Taking the same approach to $\rho_e(t)$ as before, we yield:

$$\begin{aligned} S^{(1)}(t) = \frac{\alpha_1 e^{-\rho_e(t)t}}{\rho_e(t)} \left(\frac{c}{c - A}\right)^{c\,n_{o1}} \Bigg(& e^{\rho_e(t)t}{}_2F_1\left[c\,n_{o1}, \frac{\rho_e(t)}{\kappa}, 1 + \frac{\rho_e(t)}{\kappa}, \frac{A}{A - c}e^{\kappa t}\right] \\ & - {}_2F_1\left[c\,n_{o1}, \frac{\rho_e(t)}{\kappa}, 1 + \frac{\rho_e(t)}{\kappa}, \frac{A}{A - c}\right]\Bigg), \end{aligned} \tag{D.34}$$

where ${}_2F_1(a, b, c, z)$ denotes the hypergeometric function, and $A = \lambda^2/2\kappa^2$.

Including secondary oligomers

In the case of secondary nucleation via oligomeric intermediates, we instead turn Eq. (D.6) into an integral equation:

$$\begin{aligned} S^{(1)}(t) = {} & e^{-\int^t \rho_e(\tau)d\tau} \int_0^t e^{\int^t \rho_e(\tau)d\tau} k_{o1} m^{(1)}(t)^{n_{o1}} d\tau \\ & + e^{-\int^t \rho_e(\tau)d\tau} \int_0^t e^{\int^t \rho_e(\tau)d\tau} k_{o2} m^{(1)}(t)^{n_{o2}} (m(0) - m^{(1)}(t)) d\tau. \end{aligned} \tag{D.35}$$

We note that the first integral is identical to that used just previously. We may then write:

$$S^{(1)}(t) = S^{(1)}_{\text{prim}}(t) + S^{(1)}_{\text{sec}}(t) \tag{D.36}$$

with:

$$
\begin{aligned}
S^{(1)}_{\mathrm{sec}}(t) = {} & \frac{\alpha_2 m(0) e^{-\rho_e(t)t}}{\rho_e(t)} \left(\frac{c}{c-\bar{A}}\right)^{c\,n_{\mathrm{o2}}} \left(e^{\rho_e(t)t} {}_2F_1\left[c\,n_{\mathrm{o2}}, \frac{\rho_e(t)}{\bar{\kappa}}, 1+\frac{\rho_e(t)}{\bar{\kappa}}, \frac{\bar{A}}{\bar{A}-c} e^{\bar{\kappa}t}\right] \right. \\
& \left. - {}_2F_1\left[c\,n_{\mathrm{o2}}, \frac{\rho_e(t)}{\bar{\kappa}}, 1+\frac{\rho_e(t)}{\bar{\kappa}}, \frac{\bar{A}}{\bar{A}-c}\right] \right) \\
& - \frac{\alpha_2 m(0) e^{-\rho_e(t)t}}{\rho_e(t)} \left(\frac{c}{c-\bar{A}}\right)^{c\,(n_{\mathrm{o2}}+1)} \left(e^{\rho_e(t)t} {}_2F_1\left[c\,(n_{\mathrm{o2}}+1), \frac{\rho_e(t)}{\bar{\kappa}}, 1+\frac{\rho_e(t)}{\bar{\kappa}}, \frac{\bar{A}}{\bar{A}-c} e^{\bar{\kappa}t}\right] \right. \\
& \left. - {}_2F_1\left[c\,(n_{\mathrm{o2}}+1), \frac{\rho_e(t)}{\bar{\kappa}}, 1+\frac{\rho_e(t)}{\bar{\kappa}}, \frac{\bar{A}}{\bar{A}-c}\right] \right),
\end{aligned}
\tag{D.37}
$$

where $\bar{A} = \bar{\lambda}^3/(3\bar{\kappa}^3)$, and:

$$
\begin{aligned}
S^{(1)}_{\mathrm{prim}}(t) = {} & \frac{\alpha_1 e^{-\rho_e(t)t}}{\rho_e(t)} \left(\frac{c}{c-\bar{A}}\right)^{c\,n_{\mathrm{o1}}} \left(e^{\rho_e(t)t} {}_2F_1\left[c\,n_{\mathrm{o1}}, \frac{\rho_e(t)}{\bar{\kappa}}, 1+\frac{\rho_e(t)}{\bar{\kappa}}, \frac{\bar{A}}{\bar{A}-c} e^{\bar{\kappa}t}\right] \right. \\
& \left. - {}_2F_1\left[c\,n_{\mathrm{o1}}, \frac{\rho_e(t)}{\bar{\kappa}}, 1+\frac{\rho_e(t)}{\bar{\kappa}}, \frac{\bar{A}}{\bar{A}-c}\right] \right).
\end{aligned}
\tag{D.38}
$$

D.4 Key oligomer properties

We are interested in using kinetics to uncover commonalities and differences between different amyloid oligomer populations. To do so, we focus on 3 key properties of oligomers that can be revealed through kinetic analysis: kinetic stability relative to monomer, thermodynamic stability relative to monomer, and fertility: the degree to which oligomers convert to fibrils rather than dissociate.

Kinetic stability of oligomers relative to monomers can be simply indicated from a ratio of half-times. The half-time for monomer depletion, τ_h, can be computed from fitted parameters, but can also simply be read off from kinetic plots of $M(t)$ vs t. The half-life for oligomers under constant-monomer conditions, t_h, is given by $\ln(2)/k_e$. For primary oligomers, relative thermodynamic stability is given by the steady-state concentration of oligomers attained when $m(t)$ is held at its initial value of $m(0)$: $S_{\mathrm{ss}}/m_{\mathrm{tot}} = k_{\mathrm{oligo}} m(0)^{n_o}/(k_e m(0))$. Finally, initial oligomer fertility is given by k_c/k_e. These parameters, computed for several different oligomer types, are given in Table D.1.

	Ure2 (S68C)	Tau A	Tau B	Aβ42	αS	αS + Nb2	αS + Nb87
k_{oligo}	1.6×10^{-4}	7.6×10^{-2}	2.8×10^{-3}	2.9×10^{-2}	4.0×10^{-4}	1.0×10^{-3}	1.0×10^{-3}
m_{tot}	15	8.5	8.5	5	35	70	70
n_o	2	1	1	0.9	0.9	1	1
k_d	0.60	4.6	0.20	0.36	0	0.20	0.25
k_c	2.3×10^{-3}	1.6×10^{-2}	0	3.3×10^{-2}	9.5×10^{-2}	1.2×10^{-1}	1.2×10^{-1}
t_h	1.2	0.15	3.5	1.8	7.3	2.2	1.9
k_c/k_e	0.38%	0.34%	0%	8.4%	100%	38%	32%
τ_h	8.5	2.7	2.7	0.75	34.3	40	120
$t_h/(t_h + \tau_h)$	12%	5.3%	56%	70%	18%	5.1%	1.5%
$S_{\mathrm{ss}}/m_{\mathrm{tot}}$	0.4%	1.7%	1.4%	8.3%	0.28%	0.31%	0.27%

Table D.1: Best-fit rate constants, and time scales, for the oligomers considered in Fig. 6.6 in the main text. Concentration units: μM. Time units: h. The oligomer lifetime $t_h = \ln 2/k_e$, and the monomer half-life τ_h can be read directly from plots of fibril mass concentration vs time at the desired concentration.

D.5 Secondary oligomerization of Aβ42

Fitting the above analytical model to data on Aβ42 oligomerization and fibril formation by T. C. T. Michaels in an unpublished manuscript led to the following rate constants (units h, μM): $k_+ = 1.1 \times 10^4\ \mu\mathrm{M}^{-1}\mathrm{h}^{-1}$, $k_{\mathrm{o}1} = 3.8 \times 10^{-3}\ \mu\mathrm{M}^{0.7}\mathrm{h}^{-1}$, $k_{\mathrm{o}2} = 2.9 \times 10^{-2}\ \mu\mathrm{M}^{-0.9}\mathrm{h}^{-1}$, $n_{\mathrm{o}1} = 0.3$, $n_{\mathrm{o}2} = 0.9$, $k_e = 4.0 \times 10^{-1}\ \mathrm{h}^{-1}$, $k_{\mathrm{conv}} = 4.3 \times 10^{-4}\ \mu\mathrm{M}^{-2.7}\mathrm{h}^{-1}$, $n_{\mathrm{conv}} = 2.7$.

The dominant mechanism of oligomer formation in Aβ42 aggregation is secondary oligomerization. For the purposes of computing thermodynamic stability of oligomers relative to monomers (S/m) we therefore neglect primary oligomers.

The maximum rate of secondary oligomer formation can be easily shown to occur at an intermediate reaction time t_i where $m(t_i) = \frac{n_{\mathrm{o}2}}{n_{\mathrm{o}2}+1} m_{\mathrm{tot}}$. The maximum possible oligomer concentration that can be formed in a generic secondary oligomerization reaction occurs if oligomers equilibrate with monomers far faster than monomers are depleted at this intermediate stage of the aggregation reaction. In practice this concentration is not usually reached, as monomer depletion is usually rapid at these reaction times. However, it is a useful analogue of the theoretical maximum primary oligomer concentration, being the maximal concentration obtainable if $m(t)$ and $M(t)$ are held at steady-state.

We thus calculate $S_{\mathrm{ss}}/m_{\mathrm{tot}}$ by equilibrating oligomer formation and depletion at this time:

$$\frac{S_{\mathrm{ss}}}{m_{\mathrm{tot}}} = \frac{k_{\mathrm{o}2} m(t_i)^{n_{\mathrm{o}2}} (m(0) - m(t_i))}{k_e m(0)}. \tag{D.39}$$

D.6 Fibrillar oligomer bounds

	Ure2 (S68C)	Tau A	Tau B	Aβ42	αS
m_{tot}	15	8.5	8.5	5	35
k_+	50	59^+	59^+	1.1×10^4	6.0^*
k_n	4.0×10^{-7}	3.0×10^{-4}	3.0×10^{-4}	1.1×10^{-6}	$4.3 \times 10^{-5**}$
n_c	2	1	1	2	0.26^{***}
k_2	1.2×10^{-4}			3.6×10^{-5}	
n_2	0			2	
$S_{\max}$	6.0×10^{-2}	0.16	4.8×10^{-2}	7.5×10^{-2}	$0.3^\dagger$
S_F	6.0×10^{-6}	2.2×10^{-4}	2.2×10^{-4}	4.2×10^{-6}	$6.0 \times 10^{-5^\ddagger}$
$S_F/S_{\max}$	1.0×10^{-4}	1.4×10^{-3}	4.6×10^{-3}	5.6×10^{-5}	2×10^{-4}

Table D.2: Calculating theoretical upper bound for fibrillar oligomer concentration, S_F, using $x = 100$; and comparing it to observed peak oligomer concentration, $S_{\max}$. Concentration units: μM. Time units: h. *Taken from ref. [270]; conditions similar to those in oligomer-forming experiments. **Highest reported value (ref. [15]). ***Corresponding to highest reported value for k_n (ref. [15]). $^+$Lower bound estimated from average fibril length. †Using either low- or high-FRET populations. ‡Using twice the highest reported k_n.

D.7 Oligomer simulations

We used a near-minimal coarse-grained model model developed in Ref. 113. In this model a protein is described as a hard rod decorated with an attractive patch that represents the generic interaction between the proteins, such as charged, hydrophobic, polar interactions, and H-bonding. For soluble proteins, one weakly-attractive patch is placed on the tip of the rod, driving the formation of oligomers. Proteins in the fibril-forming state interact via a side-positioned patch, which are strong and drive the formation of stable fibrils. We run Monte Carlo simulations, with small translational and rotational moves, with random swaps between the soluble and fibril-forming state. The swap is penalised with an excess in chemical potential to capture the fact that amyloidogenic proteins are rarely found in a β-sheet-prone state on their own, without binding partners. All the simulations are done by placing 600 proteins in a box of constant volume, which gives rise to the targeted protein concentration. Oligomer concentration is analysed using an in-house cluster algorithm, while the probability of oligomer conversion is computed as in Ref. 114.

Appendix E

Thermodynamics of Aβ co-oligomer formation

The material in this appendix contains derivations and additional theoretical results in support of Chapter 7. The first section contains results derived by G. A. Garcia that are helpful for understanding the main text. The remaining sections are my own work.

E.1 Self-Oligomer Modelling

A streamlined statistical mechanical model was developed by G. A. Garcia [161] to describe equilibrium oligomer size distributions and to extract the free energy of oligomer growth by single-species monomer addition. For the simple case of a single aggregating protein type, the equivalence of the proposed model to a well-established thermodynamic approach was demonstrated. Filamentous growth was assumed, such that oligomers are treated as one-dimensional chains. Nearest-neighbour interactions were taken to dominate the partition function of each oligomer and were assumed independent of oligomer size. A single equilibrium constant K then describes oligomer formation and growth at temperature T in a reaction volume V. We summarize here the key equations describing this model.

E.1.1 Statistical Mechanical Linear Oligomer Model

The grand canonical partition function for a system containing linear oligomers up to size M, assuming no interaction between oligomers, is found to be

$$\Xi(T, V, \mu) = \exp\left(\sum_{j=1}^{M} q_j(T, V)\, e^{\beta j \mu}\right). \tag{E.1}$$

where $q_j(T,V)$ is the canonical partition function of a j-mer. The number of clusters of size j is then given by

$$N_j(T,V) = q_j(T,V)\, e^{\beta j \mu}. \tag{E.2}$$

The partition function $q_j(T,V)$, or $q(j)$, can be factorized into translational and internal components: $q(j) = q_{\text{trans}}(j)\, q_{\text{int}}(j)$. The translational partition function $q_{\text{trans}}(j)$ is proportional to the system volume, and so can be written as $q_{\text{trans}} = V/v_0(j)$, where $v_0(j)$ is a fundamental volume (in the gaseous phase, it is given by the cube of the thermal wavelength). The concentration of j-mers is given by $f(j) \equiv N_j(T,V)/(N_A\, V)$, where N_A is Avogadro's number:

$$f(j) = \frac{1}{N_A\, v_0(j)}\, q_{\text{int}}(j)\, e^{\beta j \mu}. \tag{E.3}$$

The chemical potential μ is set implicitly by conservation of the initial total monomeric protein concentration m_{tot}:

$$\sum_{j=1}^{M} j f(j) = -\frac{1}{N_A\, V}\, \frac{k_B\, T}{\Xi}\, \frac{\partial \Xi}{\partial \mu} = m_{\text{tot}}. \tag{E.4}$$

E.1.2 Oligomer partition function

We assume size-independent oligomer growth and shrinkage rates, such that $q_{\text{int}}(j) \approx e^{-\beta\, \varepsilon\, (j-1)}$ and $v_0(j) \equiv v_0$. The resulting expression for the oligomer size distribution is:

$$f(j) = \frac{1}{N_A\, v_0}\, e^{-\beta\, \varepsilon\, (j-1)}\, e^{\beta\, j\, \mu}. \tag{E.5}$$

The now size-independent standard free energy change ΔG° upon addition of a monomer to an oligomer is:

$$\Delta G^\circ = -RT \ln K = -RT \ln \frac{\frac{f(j+1)}{c_0}}{\frac{f(1)}{c_0} \cdot \frac{f(j)}{c_0}} = \varepsilon - R\,T \ln\left(N_A\, c_0\, v_0\right). \tag{E.6}$$

This permits us to rewrite Eq. (E.5) in terms of ΔG° as:

$$f(j) = e^{-\beta\,(j-1)\Delta G^\circ}\, e^{\beta\, j\,(\mu - R\,T \ln(N_A\, c_0\, v_0))} = e^{-\beta\,(j-1)\Delta G^\circ}\, e^{\beta\, j\, \mu'}, \tag{E.7}$$

with μ' set implicitly as before. This also shows that ϵ is the internal free energy measured at standard concentration $c_0 = 1/(N_A v_0)$.

E.2 Co-oligomerization model

Fitting the above single-species model to the data shown in Chapter 7 allowed the dependence of oligomer concentration on oligomer length to be determined, and was found that, at the concentrations studied, almost all oligomers are dimeric. The following section therefore shows how, considering only monomers and dimers, the modelling approach can be extended to describe the co-oligomerization of more than one monomer species. Such a technique provides a usable model that was fitted to Aβ40-Aβ42 co-oligomerization data to give the free energy of cross-elongation between the two species.

The grand canonical partition function for a co-oligomerizing ensemble of 2 different monomer types, labelled a and b, up to a maximum oligomer size of 2, can be written as:

$$\Xi(T,V,\vec{\mu}) = \sum_{N_a=0}^{\infty}\sum_{N_b=0}^{\infty}\sum_{N_{aa}=0}^{\infty}\sum_{N_{bb}=0}^{\infty}\sum_{N_{ab}=0}^{\infty} \frac{\left(q_a(T,V)\,e^{\beta\mu_a}\right)^{N_a}}{N_a!}\frac{\left(q_b(T,V)\,e^{\beta\mu_b}\right)^{N_b}}{N_b!} \\ \frac{\left(q_{aa}(T,V)\,e^{\beta 2\mu_a}\right)^{N_{aa}}}{N_{aa}!}\frac{\left(q_{bb}(T,V)\,e^{\beta 2\mu_b}\right)^{N_{bb}}}{N_{bb}!}\frac{\left(q_{ab}(T,V)\,e^{\beta(\mu_a+\mu_b)}\right)^{N_{ab}}}{N_{ab}!} \tag{E.8}$$

This becomes:

$$\Xi(T,V,\vec{\mu}) = \exp\left(q_a\,e^{\beta\mu_a} + q_b\,e^{\beta\mu_b} + q_{aa}\,e^{2\beta\mu_a} + q_{bb}\,e^{2\beta\mu_b} + q_{ab}\,e^{\beta(\mu_a+\mu_b)}\right), \tag{E.9}$$

where e.g. q_{ab} gives the canonical partition function for a mixed dimer. The concentration $f(ij)$ of dimers of composition ij is given by:

$$f(ij) = \frac{1}{N_A V} q_{ij}\,e^{\beta(\mu_i+\mu_j)}. \tag{E.10}$$

As with self-oligomers, if nearest-neighbour interactions dominate we may rewrite this in terms of the interaction free energy ΔG°_{ij} between monomers of types i and j as:

$$f(ij) = e^{\beta\Delta G^{\circ}_{ij}}\,e^{\beta(\mu_i+\mu_j)}. \tag{E.11}$$

The chemical potential μ_i of monomer type i is defined for corresponding initial type i total monomeric protein concentration m^i_{tot} by

$$\begin{aligned} f(a) + f(ab) + 2f(aa) &= -\frac{1}{N_A V}\frac{k_B T}{\Xi}\frac{\partial\,\Xi}{\partial\,\mu_a} = m^a_{\mathrm{tot}}, \\ f(b) + f(ab) + 2f(bb) &= -\frac{1}{N_A V}\frac{k_B T}{\Xi}\frac{\partial\,\Xi}{\partial\,\mu_b} = m^b_{\mathrm{tot}}. \end{aligned} \tag{E.12}$$

E.3 Application to present work

The single-species model, in conjunction with the fitted single-species values for oligomerization free energy, can be used to calculate the proportion of oligomers in the single-species experiments that are dimers. At the range of concentrations explored in this study, this proportion ranges from 90-100%. Total observed oligomer concentrations are lower in the 2-species experiments, implying cross-oligomerization to be less favorable than single-species oligomerization, and therefore that the proportion of oligomers larger than dimers is even lower than in the single species experiments. Under such conditions it is therefore appropriate to fit the 2-species oligomer data to our dimer-only model derived to estimate ϵ_{40-42}. Error inherent in the model due to neglection of larger oligomers is expected to be smaller than experimental error.

E.4 Correcting for experimental observations and fitting

E.4.1 Self-oligomers

In each experiment, half of the monomer species are labeled "red", and the other half "blue". Only oligomers which contain at least one red monomer and at least one blue monomer are accounted for; all single-color species detected are branded free monomer molecules by the experimental procedure, and the two labels are assumed to have no differing effects besides introducing a degree of distinguishability between molecules.

Eq. (E.3) is thus modified to take into account the two colors:

$$f(j_1, j - j_1) = \frac{1}{N_A\, v_0}\, e^{-\beta\, \varepsilon (j-1)} e^{\beta\, j\, \mu} \binom{j}{j_1} \tag{E.13}$$

where $f(j_1, j_2)$ denotes the concentration of oligomers of size $j = j_1 + j_2$ monomers, of which j_1 are red and j_2 are blue. The apparent size distribution $f_{\text{exp}}(j)$ is then:

$$f_{\text{exp}}(j) = \begin{cases} \sum_{j_1=1}^{M} f(j_1, 0) + \sum_{j_2=1}^{M} f(0, j_2) & j = 1 \\ \sum_{j_1=1}^{j-1} f(j_1, j - j_1) & j > 1 \end{cases}. \tag{E.14}$$

Combining Eqs. (E.13) and (E.14), the total apparent oligomer concentration F_{exp} is given by:

$$F_{\text{exp}} = \sum_{j=2}^{M} \frac{2^{j-1}}{N_A\, v_0}\, e^{-\beta\, \varepsilon (j-1)} e^{\beta\, j\, \mu} \tag{E.15}$$

The conservation-of-mass condition is also modified, and becomes:

$$\sum_{j=1}^{M} \sum_{j_1=0}^{j} j_1 \cdot f(j_1, j - j_1) = -\frac{1}{2} \frac{1}{N_A\, V} \frac{k_B\, T}{\Xi} \frac{\partial \Xi}{\partial \mu} = \frac{1}{2} m_{\text{tot}}. \tag{E.16}$$

Eq. (E.15), with M chosen to be large enough to encompass virtually all oligomers, can be fitted directly via least-squares methods to experimental measurements of equilibrium oligomer concentrations, with ε as the sole fitting parameter. At each step, μ is evaluated numerically from the implicit analytical conservation-of-mass condition Eq. (E.16).

E.4.2 Co-oligomers

For two species, one labeled red (type 1) and the other blue (type 2), we can fit the expression for $f(ab)$ directly to experimental measurements, where $f(ab)$ is explicitly given by

$$f(ab) = \frac{V}{v_0}\, e^{-\beta\epsilon_{ab}}\, e^{\beta(\mu_a+\mu_b)} \tag{E.17}$$

The fitting procedure now involves evaluating both chemical potentials μ_a and μ_b at each stage numerically from implicit algebraic expressions, obtained via Eq. (E.12). The sole fitting parameter is $\varepsilon_{1,2}$, once $\varepsilon_{1,1}$ and $\varepsilon_{2,2}$ are established from fits to self-oligomerization data sets [161]. An appropriate choice of v_0 allows us to identify $\varepsilon_{1,2}$ with the dimerization free energy $\Delta G^{\circ}_{1,2}$ [161]. Errors are estimated by manually varying the parameter until the resulting curve no longer falls within the majority of the error bars on the datapoints. All experiments were carried out at $T = 37$ °C, and this is used to calculate an appropriate value of β for the fitting.

E.5 Simulating different concentrations and ratios

Once all elongation free energies have been found from fitting, the expressions above can be used to simulate co-oligomerization equilibrium concentrations of different species for a range of concentrations of each monomer type (provided we remain within the range of concentrations at which dimers dominate), with the chemical potentials set in all cases by Eq. (E.12). The composition distribution $f(ij)$ can be calculated explicitly, and the total oligomer concentrations can be predicted from their sum. The results can be seen in Fig. E.1, and in Fig. 3 of Chapter 7. Finally, it is of interest to consider the maximum predicted total equilibrium oligomer concentration. This is capped because the equilibrium monomer concentration is limited by the CAC. Using the CAC as an accurate estimate for the total concentration of abeta in monomers and in oligomers, we can calculate the maximum possible total equilibrium oligomer concentration to be approximately 44 nM. The approximation that oligomers are predominantly dimers is at its least accurate at this upper limit of the monomer concentration; however, we estimate that the error introduced to the total oligomer concentration by this approximation is no more than 5%. To arrive at this error estimate, we simulated a single-species system at the same total monomer concentration, using the single-species oligomerisation free energy of Aβ40, and discovered that when the permitted oligomer species were restricted to dimers the calculated total oligomer concentration is 95% of the total oligomer concentration obtained when all oligomer lengths are permitted.

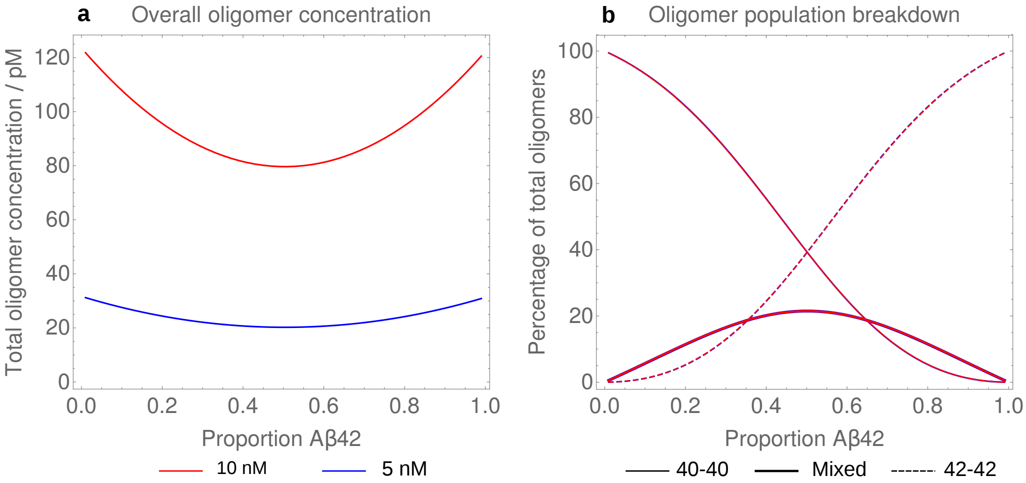

Figure E.1: Simulation of Aβ40-Aβ42 co-oligomerization equilibrium behavior at 5 nM (blue) and 10 nM (red) total Aβ concentration for a range of Aβ40:Aβ42 ratios, using elongation free energies obtained from fitting to experimental data (see Figure 2 in Chapter 7). A clear difference is seen in total oligomer concentrations between the two monomer concentrations **(a)**, although the oligomer percentage breakdowns at both concentrations are virtually identical, with red and blue lines overlapping **(b)**.

Appendix F

Thermodynamics of αS co-oligomer formation

This appendix is mostly given over to an unpublished linear co-oligomer model developed by G. A. Garcia, employed throughout Chapter 8. It is prefaced by a brief analysis of oligomer sizes carried out by myself.

Protein combination	ΔG° **/kJmol^{-1}**	**Error in** ΔG°
A30P-A30P	-26.9	1.37
A53T-A53T	-25.2	0.44
E46K-E46K	-28.7	0.59
k18-k18	-19.4	0.49
WT-WT	-24.0	0.31
WT-A30P	-24.4	0.85
WT-A53T	-27.1	0.46
WT-E46K	-28.7	0.64
WT-k18	-22.7	0.75
WT-Aβ40	-29.4	0.62
WT-Aβ42	-30.8	1.03

Table F.1: ΔG° values of oligomerisation, derived from the fitting of sm-TCCD data as described in the main text. The summary of these values is in Fig. 8.3 (main text).

F.1 Oligomer sizes

Having used the linear model outlined in [161] to determine self-oligomer ΔG°s, it is possible to simulate oligomer size distributions formed by different αS mutants at different initial monomer concentrations. It may be shown that, at the concentrations visited by our study, oligomers formed by the mutants under investigation are predominantly dimeric. For example, at 1 μM, 99% of oligomers formed by WT αS are dimeric. The steep decline in the size distributions implies that possible deviations from linear geometry at larger sizes will not significantly perturb our analysis, justifying our choice of a linear model.

F.2 Co-oligomer modelling

We now present a model for co-oligomerization of two species to form linear oligomers of any length developed by G. A. Garcia, building on the single-species linear model presented in [161].

F.2.1 Exact solution

The grand canonical partition function for a co-oligomerizing ensemble of 2 different monomer types is given as a direct generalization of the single-species partition function in [161]:

$$\Xi(T, V, \vec{\mu}) = \exp\left(\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} q_{\vec{j}}(T, V)\, e^{\beta\, \vec{j}\cdot\vec{\mu}}\right), \tag{F.1}$$

where $\vec{j} = (j_1, j_2)$ is a composition vector whose components j_i describe the number of monomeric residues of type i. Similarly, $\vec{\mu} = (\mu_1, \mu_2)$ with μ_i being the chemical potential associated with species i, $q_{\vec{j}}$ gives the canonical partition function for an oligomer with composition described by $\vec{j}$. If nearest-neighbor interactions dominate, $q_{\vec{j}}(T, V)$ is given by:

$$q_{\vec{j}}(T, V) \approx \frac{V}{v_0} \sum_{o(\vec{j})} e^{-\beta\, U(\vec{j}, o(\vec{j}))}, \tag{F.2}$$

where $U(\vec{j}, o(\vec{j}))$ denotes the internal energy of an oligomer with composition $\vec{j}$ and with monomer units arranged in an order specified by $o(\vec{j})$; a sum must be carried out over all possible arrangements. The concentration of oligomers with composition $\vec{j}$ is then given by:

$$f(\vec{j}) = \frac{1}{N_A v_0} \sum_{o(\vec{j})} e^{-\beta\, U(\vec{j}, o(\vec{j}))}\, e^{\beta\, \vec{j}\cdot\vec{\mu}}. \tag{F.3}$$

The chemical potential μ_i of monomer type i is defined for corresponding initial type i total monomeric protein concentration m^i_{tot} by

$$\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} j_i\, f(\vec{j}) = -\frac{1}{N_A\, V}\frac{k_B\, T}{\Xi}\frac{\partial\, \Xi}{\partial\, \mu_i} = m^i_{\text{tot}}. \tag{F.4}$$

Carrying out the sum over configurations

The sum over arrangements $o(\vec{j})$ can be carried out by noting that if the composition $\vec{j} = (j_1, j_2)$ and the number of cross-bonds x present in a J-mer is known, then $U(\vec{j}, o(\vec{j}))$

is almost uniquely defined. There are three cases to consider.
Case 1: x is even and the chain terminates with type 1 monomers on both ends. We then have $x/2$ distinct sequences of type-2 monomers within the oligomer, and the number of 2-2 bonds is $j_2 - x/2$. The remaining $j_1 - 1 - x/2$ bonds are 1-1, and the maximum value of x is the smaller of $2(j_1 - 1)$ and $2j_2$.
Case 2: x is even and the chain terminates with type 2 monomers on both ends, and the number of 1-1 and 2-2 bonds are $j_1 - x/2$ and $j_2 - 1 - x/2$ respectively, with the maximum value of x being the smaller of $2(j_2 - 1)$ and $2j_1$.
Case 3: x is odd. There are $(x - 1)/2$ internal domains of each type, and 1 external domain. Thus the number of self-bonds of type-i is $j_i - (x + 1)/2$, and the maximum allowed value of x is the smaller of $2j_1 - 1$ and $2j_2 - 1$.

These three cases all give different energies for a given value of x, so indexing the case number with c, we may write the energy $U(\vec{j}, o(\vec{j}))$ as $E(j_1, j_2, x, c)$:

$$E(j_1, j_2, x, 1) = \left(j_1 - 1 - \frac{x}{2}\right) \varepsilon_{1,1} + \left(j_2 - \frac{x}{2}\right) \varepsilon_{2,2} + x\,\varepsilon_{1,2},\ 1 \leq \frac{x}{2} \leq \ \mathrm{Min}(j_1 - 1, j_2) \tag{F.5}$$

$$E(j_1, j_2, x, 2) = \left(j_1 - \frac{x}{2}\right) \varepsilon_{1,1} + \left(j_2 - 1 - \frac{x}{2}\right) \varepsilon_{2,2} + x\,\varepsilon_{1,2},\ 1 \leq \frac{x}{2} \leq \ \mathrm{Min}(j_1, j_2 - 1) \tag{F.6}$$

$$E(j_1, j_2, x, 3) = \left(j_1 - \frac{x+1}{2}\right) \varepsilon_{1,1} + \left(j_2 - \frac{x+1}{2}\right) \varepsilon_{2,2} + x\,\varepsilon_{1,2},\ 1 \leq \frac{x+1}{2} \leq \ \mathrm{Min}(j_1, j_2), \tag{F.7}$$

where the smaller of the two numbers j_1 and j_2 is given by $\mathrm{Min}(j_1, j_2)$. The Boltzmann factors for each of these energies must then be multiplied by the relevant degeneracy and added together to give the contribution to $q(\vec{j})$ for each value of x. For case 1, we must arrange $x/2$ type-2 internal domains with $j_1 - 1 - x/2$ type-1 self bonds, and then $x/2 - 1$ type-1 internal domains with $j_2 - x/2$ type-2 self bonds, yielding $\binom{j_1-1}{x/2}\binom{j_2-1}{x/2-1}$. The degeneracies for the other cases can be obtained similarly. A sum can then be carried out over all possible values of x, multiplying the contribution of case 3 to the sum by a factor of 2 as such chains have directionality, and setting $x = 2r$ or $x = 2r - 1$ in the relevant sums to ensure only odd or even values are included as required. We ultimately arrive at the expression:

$$\frac{q(\vec{j})}{V/v_0} = \begin{cases} e^{-\beta\varepsilon_{2,2}(J-1)}, & j_1 = 0 \\ e^{-\beta\varepsilon_{1,1}(J-1)}, & j_2 = 0 \\ \begin{aligned} &\qquad 2\sum_{r=1}^{\mathrm{Min}(j_1,j_2)} \binom{j_1-1}{r-1}\binom{j_2-1}{r-1} e^{-\beta E(j_1,j_2,r,3)} \\ &+\mathcal{H}[j_1-2]\sum_{r=1}^{\mathrm{Min}(j_1-1,j_2)}\binom{j_1-1}{r}\binom{j_2-1}{r-1}e^{-\beta E(j_1,j_2,r,1)} \\ &+\mathcal{H}[j_2-2]\sum_{r=1}^{\mathrm{Min}(j_1,j_2-1)}\binom{j_2-1}{r}\binom{j_1-1}{r-1}e^{-\beta E(j_1,j_2,r,2)}, \end{aligned} & j_1>0, j_2>0, \end{cases} \quad, \tag{F.8}$$

where $J = j_1 + j_2$, the discrete Heaviside step function is denoted $\mathcal{H}[z]$, defined such that $H[0] = 1$, and the remaining functions are defined as follows:

$$E(j_1, j_2, r, 1) = (j_1 - r - 1)\,\varepsilon_{1,1} + (j_2 - r)\,\varepsilon_{2,2} + 2r\,\varepsilon_{1,2}, \tag{F.9}$$

$$E(j_1, j_2, r, 2) = (j_1 - r)\,\varepsilon_{1,1} + (j_2 - r - 1)\,\varepsilon_{2,2} + 2r\,\varepsilon_{1,2}, \tag{F.10}$$

$$E(j_1, j_2, r, 3) = (j_1 - r)\,\varepsilon_{1,1} + (j_2 - r)\,\varepsilon_{2,2} + (2r - 1)\,\varepsilon_{1,2}. \tag{F.11}$$

F.2.2 Correcting for experimental observations and fitting

Where one monomer type is labelled with donor dye and one with acceptor dye, then only mixed oligomers will be recorded. The total mixed oligomer concentration $F(\vec{m_{\mathrm{tot}}})$ is given from combining Eqs. (F.3) and (F.8):

$$\begin{aligned} F_{\mathrm{exp}}(\vec{m_{\mathrm{tot}}}) = \sum_{j=2}^{M}\sum_{j_1=1}^{j-1} &\qquad 2\sum_{r=1}^{\mathrm{Min}(j_1,j-j_1)}\binom{j_1-1}{r-1}\binom{j-j_1-1}{r-1}e^{-\beta E(j_1,j-j_1,r,3)} \\ &+\ \mathcal{H}[j_1-2]\sum_{r=1}^{\mathrm{Min}(j_1-1,j-j_1)}\binom{j_1-1}{r}\binom{j-j_1-1}{r-1}e^{-\beta E(j_1,j-j_1,r,1)} \\ &+\ \mathcal{H}[j-j_1-2]\sum_{r=1}^{\mathrm{Min}(j_1,j-j_1-1)}\binom{j-j_1-1}{r}\binom{j_1-1}{r-1}e^{-\beta E(j_1,j-j_1,r,2)} \end{aligned} \tag{F.12}$$

This expression can then be fitted directly to experimental measurements.

The fitting procedure now involves evaluating both chemical potentials at each stage numerically from implicit algebraic expressions, obtained via Eq. (F.4) and Eq. (F.8). The sole fitting parameter is $\varepsilon_{1,2}$, once $\varepsilon_{1,1}$ and $\varepsilon_{2,2}$ are established from fits to self-oligomerization data sets. The same standard value for v_0 is used for all fits.

Appendix G

Statistical Mechanics of Oligomer Formation

The material in this appendix supports Chapter 9. It contains derivations, supporting theoretical results, and details of simulations. All work is my own apart from the simulations, which were performed by Anđela Šarić; and the first section, which is a reworking of material from ref. [161] and is included for clarity.

G.1 Derivation of general statistical mechanical formulation for oligomers

The grand canonical partition function is given by:

$$\Xi(T, V, \mu) = \sum_{N}^{\infty} Q(T, V, N) e^{\beta N \mu} \tag{G.1}$$

where $Q(T, V, N)$ denotes the canonical partition function of a system that contains N molecules, $\beta = 1/(k_B T)$ is the inverse temperature (k_B is the Boltzmann constant) and μ is the chemical potential of the monomers. We consider the formation of oligomeric aggregates up to an arbitrary maximum size of M monomers. By assuming that the oligomers do not interact, the grand canonical partition function can be expressed in terms of the individual oligomer partition functions:

$$\Xi(T, V, \mu) = \sum_{N_1=0}^{\infty} \sum_{N_2=0}^{\infty} \cdots \sum_{N_M=0}^{\infty} \frac{\left(q_1(T, V)\, e^{\beta \mu}\right)^{N_1}}{N_1!} \frac{\left(q_2(T, V)\, e^{\beta 2 \mu}\right)^{N_2}}{N_2!} \cdots \frac{\left(q_M(T, V)\, e^{\beta M \mu}\right)^{N_M}}{N_M!}, \tag{G.2}$$

where $q_j(T,V)$ is the canonical partition function of a j-mer. Eq. (G.2) can be re-summed to yield

$$\Xi(T,V,\mu) = \exp\left(\sum_{j=1}^{M} q_j(T,V)\, e^{\beta j \mu}\right). \tag{G.3}$$

From Eq. (G.3), the concentration of oligomers of size j, $f(j) \equiv N_j(T,V)/(N_A\, V)$, is given by

$$f(j) = \frac{1}{N_A\, V}\, q(j)\, e^{\beta j \mu}, \tag{G.4}$$

where N_A is Avogadro's number and we have used the short hand notation $q(j) \equiv q_j(T,V)$. Note that the chemical potential μ is set implicitly by letting $M \to \infty$ in Eq. (G.3) and imposing the conservation-of-mass condition, $p = \sum_{j=1}^{\infty} j f(j)$, which can be written more conveniently as

$$p = -\frac{1}{N_A\, V} \frac{k_B\, T}{\Xi} \frac{\partial \Xi}{\partial \mu} \tag{G.5}$$

We must now address the partition function $q(j)$. This is given in general by a product of independent contributions from translational and internal degrees of freedom: $q(j) = q_{\text{trans}}(j)\, q_{\text{int}}(j)$. In particular, the translational partition function $q_{\text{trans}}(j)$ is proportional to the system volume, and so can be written as $q_{\text{trans}} = V/v_0(j)$, where $v_0(j)$ is a fundamental volume (in the gaseous phase, it is given by the cube of the thermal wavelength). Making this substitution, we arrive at Eq. (9.1).

G.2 Head group sterics

We expect that head groups have some translational, rotational and conformational degrees of freedom, and that these are restricted through steric interactions upon addition of amphiphiles to a micelle. Indeed, it is known that head group size can influence micelle geometry as a result of these interactions [271]. This free energy penalty for formation of larger micelles was neglected by Chandler *et al*, presumably because it is small relative to the connectivity-enforcing energy penalty.

We here verify this hypothesis by calculating the steric penalty explicitly. In previous MT approaches, this has been dealt with using a test particle approach [272], yielding the expression:

$$G_{\text{st}} = -jkT \ln\left[1 - \frac{a_h}{a}\right], \tag{G.6}$$

where a_h is the average cross-sectional area of the head group (in our Monte Carlo simulations $\pi\sigma^2/4$), and a is the core surface area per monomer. We first calculate a for spherical micelles:

$$V = jv = \frac{4}{3}\pi r^3 \implies r = \left(\frac{3jv}{4}\right)^{1/3} \tag{G.7a}$$

$$a = \frac{S}{j} = \frac{Sv}{V} = \frac{3v}{r} \tag{G.7b}$$

$$\therefore a = 3v\left(\frac{4}{3jv}\right)^{1/3} = \left(\frac{36v^2}{j}\right)^{1/3}. \tag{G.7c}$$

Now since v scales with σ^3, we have the scaling behaviour for G_{st}:

$$G_{\mathrm{st}} \simeq -jkT \ln\left[1 - Cj^{\frac{1}{3}}\right], \tag{G.8}$$

with constant C. Thus, G_{st} has lower scaling with j than the connectivity-enforcing free energy penalty ($j^{5/3}$). Furthermore, C is given approximately by:

$$C < \frac{\pi\sigma^2/4}{(36\sigma^6)^{1/3}} \simeq 0.24, \tag{G.9}$$

which is smaller than the connectivity coefficient h (0.58 for our system; expected to be in the range 0.3-1 for many peptides). Thus it is reasonable to neglect the steric contribution for modelling the Monte Carlo simulations, and oligomers in general.

G.3 Monte Carlo simulations

To study the formation of linear and globular oligomers we used Monte Carlo (MC) simulations, where $N = 600$ (in case of the linear model) or $N = 1000$ (in case of the globular model) monomers were placed inside a cubic box to achieve the target concentration. Simulations were run for at least $3 \cdot 10^8$ MC steps. The mapping between the simulation and physical units was done by using the $\sigma = 2$nm mapping, where σ is the length-scale in our simulations. This mapping was chosen such that the simulation model reproduces the sizes of typical small peptides and proteins, such as Aβ. The size distribution of the resulting oligomers was analysed using an in-house clustering algorithm.

G.3.1 Linear oligomers

A monomer is described as a hard sphere of a diameter σ. The interaction between two monomers is created by placing two ideal spheres inside the monomer at a distance of 0.4σ from its centre, diametrically opposite from each other. Two ideal spheres belonging to different monomers interact with the interaction enthalpy of $-\epsilon$ if their centre-to-centre distance is less than 0.3σ. This insures that a single ideal sphere cannot interact with more than one other monomer at a time, giving rise to linear oligomers. A value for ϵ of

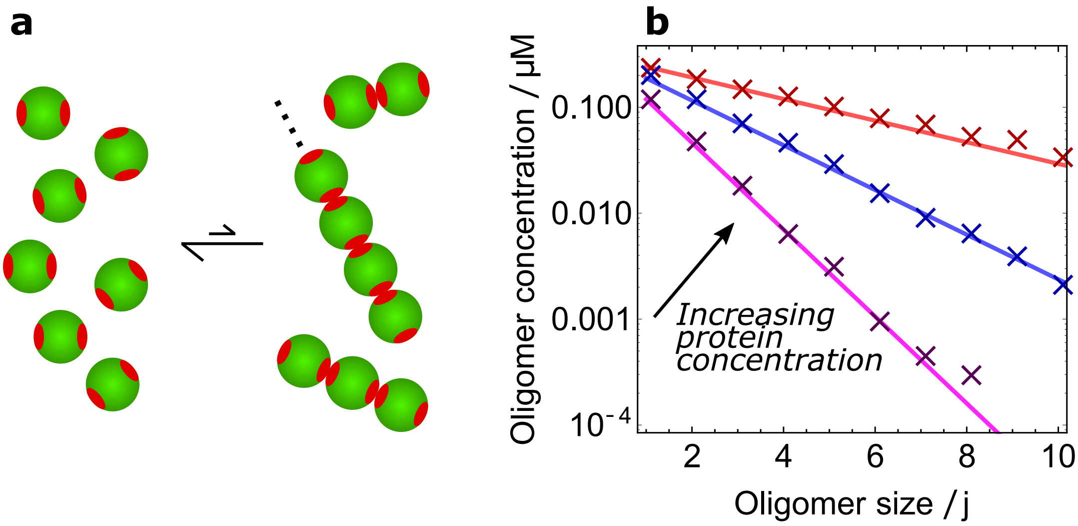

Figure G.1: The analytical size distribution model for linear oligomers can be globally fitted to coarse-grained numerical Monte Carlo results (data points) with high accuracy. $\Delta G^\circ = -15kT$; $p = 0.3\mu\text{M}$ (purple); $p = 1.3\mu\text{M}$ (blue); $p = 5.3\mu\text{M}$ (red).

$-15kT$ was used.

G.3.2 Globular oligomers

A monomer is described as a hard spherocylinder of a diameter of σ and a length of 4σ. An interaction centre is placed at one pole of the spherocylinder, at a distance 1.5σ from its centre. Two spherocylinders interact with the interaction enthalpy of $-\epsilon$ if their interaction centres are less than 1.3σ apart. This interaction drives the formation of micellar-like oligomers, where tips of participating monomers are in contact in the micelle centre. A value for ϵ of $-5kT$ was used; this does not include entropic effects.

G.4 Calculating the critical monomer concentration m^*

The maximum of the length distribution is found by solving $df(j)/dj = 0$:

$$\frac{df(j)}{dj} = \frac{1}{N_A v_0}\frac{d}{dj}e^{-\beta(\Delta G(j)-j\mu)} = f(j)\cdot -\beta\left(\frac{d\Delta G(j)}{dj} - \mu\right) = 0 \tag{G.10a}$$

$$\implies \frac{d\Delta G(j)}{dj} - \mu = \frac{5}{3}\frac{h}{\beta}j^{2/3} + \left(\frac{z_\infty}{2}G_b - \mu\right) - \frac{2}{3}j^{-1/3}\frac{z_\infty}{2}\alpha' G_b = 0. \tag{G.10b}$$

By making the substitution $x = j^{1/3}$, this can be transformed into the simpler problem $\mathcal{F}(x) = 0$, where:

$$\begin{aligned}\mathcal{F}(x) &= ax^3 + cx + d &\text{(G.11a)}\\ &= \frac{5}{3}\frac{h}{\beta}x^3 + \left(\frac{z_\infty}{2}G_b - \mu\right)x - \frac{z_\infty}{3}\alpha' G_b. &\text{(G.11b)}\end{aligned}$$

We in fact need both a maximum and a minimum, so we need at least 2 real solutions to this equation. The cubic function must then cross the x axis three times. For this to happen it must possess two real stationary points $x_\pm$:

$$\mathcal{F}'(x) = 3ax^2 + c = 0 \quad \implies x_\pm = \pm\sqrt{\frac{-c}{3a}}. \tag{G.12}$$

We thus need $c = \frac{z_\infty}{2}G_b - \mu < 0$ as a necessary but not sufficient condition. This translates to a fairly tiny minimum monomer concentration so we look for a further constraint. We need that the first of these stationary points gives a value of $\mathcal{F} > 0$, and the second $\mathcal{F} < 0$:

$$\begin{aligned}\mathcal{F}(x_\pm) &= \pm a\left(\frac{-c}{3a}\right)^{3/2} \pm c\left(\frac{-c}{3a}\right)^{1/2} + d &\text{(G.13a)}\\ &= \pm\left(\frac{(-c)^3}{27a}\right)^{1/2} \mp (-c)^{2/2}\left(\frac{-c}{3a}\right)^{1/2} + d &\text{(G.13b)}\\ &= \pm\frac{1}{3}\left(\frac{(-c)^3}{3a}\right)^{1/2} \mp \left(\frac{(-c)^3}{3a}\right)^{1/2} + d &\text{(G.13c)}\\ \therefore \mathcal{F}(x_-) &= \frac{2}{3}\left(\frac{(-c)^3}{3a}\right)^{1/2} + d, \quad \mathcal{F}(x_+) = -\frac{2}{3}\left(\frac{(-c)^3}{3a}\right)^{1/2} + d. &\text{(G.13d)}\end{aligned}$$

Given that $d > 0$, the first condition is always satisfied. To satisfy the second, we require (using $z_\infty = 12$):

$$\begin{aligned}\frac{2}{3}\left(\frac{(-c)^3}{3a}\right)^{1/2} &> d &\text{(G.14a)}\\ (-c)^3 &> 3a\frac{9d^2}{4} = 5\frac{h}{\beta}\frac{(-z_\infty\alpha' G_b)^2}{4} &\text{(G.14b)}\\ \mu - 6G_b &> \left(\frac{5h}{\beta}\right)^{1/3}(-6\alpha' G_b)^{2/3}. &\text{(G.14c)}\end{aligned}$$

This gives the critical monomer concentration:

$$m^* = \frac{1}{N_A v_0}e^{\beta\mu} = \frac{1}{N_A v_0}\exp\left(6\beta\, G_b + (5h)^{1/3}\left(-6\alpha'\beta\, G_b)^{2/3}\right)\right). \tag{G.15}$$

The value x^* for which the roots merge, or for which $\Delta = 0$, is given by $9d/2\Delta_0 = -3d/2ac$, which gives the location of the new maximum:

$$j^* = \frac{d}{2} = -\frac{6\alpha'\beta G_b}{5h}. \tag{G.16}$$

Bibliography

[1] F. Chiti and C. M. Dobson, Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333–366 (2006).

[2] D. J. Selkoe, Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).

[3] C. M. Dobson, Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**, 329–332 (1999).

[4] C. M. Dobson, Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

[5] A. Aguzzi and T. O'Connor, Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov* **9**, 237–248 (2010).

[6] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Madsen, C. Riekel, and D. Eisenberg, Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).

[7] M. Sunde and C. Blake, The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem* **50**, 123–159 (1997).

[8] T. P. J. Knowles, M. Vendruscolo, and C. M. Dobson, The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15**, 384–396 (2014).

[9] P. Ciryam, G. G. Tartaglia, R. I. Morimoto, C. Dobson, and M. Vendruscolo, Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep* **5**, 781 – 790 (2013).

[10] F. Boller, T. Mizutani, U. Roessmann, and P. Gambetti, Parkinson disease, dementia, and Alzheimer disease: Clinicopathological correlations. *Ann Neurol* **7**, 329–335 (1980).

[11] P. T. Lansbury and H. A. Lashuel, A century-old debate on protein aggregation and neurodegeneration enters the clinic. *Nature* **443**, 774–779 (2006).

[12] F. Bemporad and F. Chiti, Protein misfolded oligomers: Experimental approaches, mechanism of formation, and structure-toxicity relationships. *Chem Biol* **19**, 315 – 327 (2012).

[13] S. I. A. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc Natl Acad Sci USA* **110**, 9758–9763 (2013).

[14] G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J. B. Kirkegaard, S. I. A. Cohen, C. M. Dobson, S. Linse, and T. P. J. Knowles, Differences in nucleation behavior underlie the contrasting aggregation kinetics of the Aβ40 and Aβ42 peptides. *Proc Natl Acad Sci USA* **111**, 9384–9389 (2014).

[15] P. Flagmeier, G. Meisl, M. Vendruscolo, T. P. J. Knowles, C. M. Dobson, A. K. Buell, and C. Galvagnion, Mutations associated with familial Parkinson's disease alter the initiation and amplification steps of α-synuclein aggregation. *Proc Natl Acad Sci USA* **113**, 10328–10333 (2016).

[16] G. Meisl, J. B. Kirkegaard, P. Arosio, T. T. C. Michaels, M. Vendruscolo, C. M. Dobson, S. Linse, and T. P. J. Knowles, Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat Protoc* **11**, 252–272 (2016).

[17] P. Arosio, T. P. J. Knowles, and S. Linse, On the lag phase in amyloid fibril formation. *Phys Chem Chem Phys* **17**, 7606–7618 (2015).

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

[19] F. A. Ferrone, J. Hofrichter, and W. A. Eaton, Kinetics of sickle hemoglobin polymerization. i. studies using temperature-jump and laser photolysis techniques. *J Mol Biol* **183**, 591–610 (1985).

[20] F. A. Ferrone, J. Hofrichter, and W. A. Eaton, Kinetics of sickle hemoglobin polymerization. ii. a double nucleation mechanism. *J Mol Biol* **183**, 611–631 (1985).

[21] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, An analytical solution to the kinetics of breakable filament assembly. *Science* **326**, 1533–1537 (2009).

[22] J. P. Cleary, D. M. Walsh, J. J. Hofmeister, G. M. Shankar, M. A. Kuskowski, D. J. Selkoe, and K. H. Ashe, Natural oligomers of the amyloid-β protein specifically disrupt cognitive function. *Nat Neurosci* **8**, 79–84 (2005).

[23] B. Winner, R. Jappelli, S. K. Maji, P. A. Desplats, L. Boyer, S. Aigner, C. Hetzer, T. Loher, M. Vilar, S. Campioni, C. Tzitzilonis, A. Soragni, S. Jessberger, H. Mira, A. Consiglio, E. Pham, E. Masliah, F. H. Gage, and R. Riek, In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc Natl Acad Sci USA* **108**, 4194–4199 (2011).

[24] M. J. Guerrero-Muñoz, D. L. Castillo-Carranza, S. Krishnamurthy, A. A. Paulucci-Holthauzen, U. Sengupta, C. A. Lasagna-Reeves, Y. Ahmad, G. R. Jackson, and R. Kayed, Amyloid-β oligomers as a template for secondary amyloidosis in Alzheimer's disease. *Neurobiol Dis* **71**, 14 – 23 (2014).

[25] H. Li, L. Ying, J. J. Green, S. Balasubramanian, and D. Klenerman, Ultrasensitive coincidence fluorescence detection of single DNA molecules. *Anal Chem* **75**, 1664–1670 (2003).

[26] A. Orte, N. R. Birkett, R. W. Clarke, G. L. Devlin, C. M. Dobson, and D. Klenerman, Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci USA* **105**, 14424–14429 (2008).

[27] G. A. Garcia, S. I. A. Cohen, C. M. Dobson, and T. P. J. Knowles, Nucleation-conversion-polymerization reactions of biological macromolecules with prenucleation clusters. *Phys Rev E* **89**, 032712 (2014).

[28] M. Iljina, G. A. Garcia, M. H. Horrocks, L. Tosatto, M. L. Choi, K. A. Ganzinger, A. Y. Abramov, S. Gandhi, N. W. Wood, N. Cremades, C. M. Dobson, T. P. J. Knowles, and D. Klenerman, Kinetic model of the aggregation of alpha-synuclein provides insights into prion-like spreading. *Proc Natl Acad Sci USA* **113**, E1206–E1215 (2016).

[29] S. L. Shammas, G. A. Garcia, S. Kumar, M. Kjaergaard, M. H. Horrocks, N. Shivji, E. Mandelkow, T. P. J. Knowles, E. Mandelkow, and D. Klenerman, A mechanistic model of tau amyloid aggregation based on direct observation of oligomers. *Nat Commun* **6**, 7025 (2015).

[30] S. I. A. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles, Nucleated polymerization with secondary pathways. i. time evolution of the principal moments. *J Chem Phys* **135**, 065105 (2011).

[31] T. Vetter, M. Iggland, D. R. Ochsenbein, F. S. HÃ¤nseler, and M. Mazzotti, Modeling nucleation, growth, and ostwald ripening in crystallization processes: A comparison between population balance and kinetic rate equation. *Cryst Growth Des* **13**, 4890–4905 (2013).

[32] A. Goto and T. Fukuda, Kinetics of living radical polymerization. *Prog Polym Sci* **29**, 329 – 385 (2004).

[33] K. L. Cooke and J. A. Yorke, Some equations modelling growth processes and gonorrhea epidemics. *Math Biosci* **16**, 75 – 101 (1973).

[34] M. A. Nowak and C. R. M. Bangham, Population dynamics of immune responses to persistent viruses. *Science* **272**, 74–79 (1996).

[35] D. A. Mitchell, O. F. von Meien, N. Krieger, and F. D. H. Dalsenter, A review of recent developments in modeling of microbial growth kinetics and intraparticle phenomena in solid-state fermentation. *Biochem Eng J* **17**, 15 – 26 (2004).

[36] T. C. T. Michaels and T. P. J. Knowles, Kinetic theory of protein filament growth: Self-consistent methods and perturbative techniques. *Int J Mod Phys B* **29**, 1530002 (2015).

[37] M. A. Nowak, D. C. Krakauer, A. Klug, and R. M. May, Prion infection dynamics. *Integr Biol* **1**, 3 (1998).

[38] D. Hall and H. Edskes, Silent prions lying in wait: a two-hit model of prion/amyloid formation and infection. *J Mol Biol* **336**, 775–786 (2004).

[39] W.-F. Xue, S. W. Homans, and S. E. Radford, Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc Natl Acad Sci USA* **105**, 8926–8931 (2008).

[40] N. Carulla, G. L. Caddy, D. R. Hall, J. Zurdo, M. Gairí, M. Feliz, E. Giralt, C. V. Robinson, and C. M. Dobson, Molecular recycling within amyloid fibrils. *Nature* **436**, 554–558 (2005).

[41] M. M. Pallitto and R. M. Murphy, A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state. *Biophys J* **81**, 1805–1822 (2001).

[42] T. Poeschel, N. V. Brilliantov, and C. Froemmel, Kinetics of prion growth. *Biophys J* **85**, 3460–3474 (2003).

[43] D. Endres and A. Zlotnick, Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J* **83**, 1217–1230 (2002).

[44] T. Keef, C. Micheletti, and R. Twarock, Master equation approach to the assembly of viral capsids. *J Theor Biol* **242**, 713 – 721 (2006).

[45] T. C. T. Michaels, M. M. J. Bellaiche, M. F. Hagan, and T. P. J. Knowles, Kinetic constraints on self-assembly into closed supramolecular structures. *Sci Rep* **7**, 12295 (2017).

[46] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol* **2**, e321 (2004).

[47] F. A. Ferrone, J. Hofrichter, H. R. Sunshine, and W. A. Eaton, Kinetic studies on photolysis-induced gelation of sickle cell hemoglobin suggest a new mechanism. *Biophys J* **32**, 361–380 (1980).

[48] T. C. T. Michaels and T. P. J. Knowles, Mean-field master equation formalism for biofilament growth. *Am J Phys* **82**, 476–483 (2014).

[49] T. C. T. Michaels and T. P. J. Knowles, Role of filament annealing in the kinetics and thermodynamics of nucleated polymerization. *J Chem Phys* **140**, 214904 (2014).

[50] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Nucleated polymerization with secondary pathways. ii. determination of self-consistent solutions to growth processes described by non-linear master equations. *J Chem Phys* **135**, 065106 (2011).

[51] A. Granas and J. Dugundji, *Fixed Point Theory.* Springer-Verlag (2003).

[52] A. Szabo, Fluctuations in the polymerization of sickle hemoglobin. a simple analytic model. *J Mol Biol* **199**, 539–542 (1988).

[53] F. Oosawa and S. Asakura, *Thermodynamics of the Polymerization of Protein.* Academic Press (1975).

[54] C. Li, S. Bolisetty, and R. Mezzenga, Hybrid nanocomposites of gold single-crystal platelets and amyloid fibrils with tunable fluorescence, conductivity, and sensing properties. *Adv Mater* **25**, 3694–3700 (2013).

[55] T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze, and M. E. Welland, Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nat Nanotechnol* **5**, 204–207 (2010).

[56] E. Gazit, Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem Soc Rev* **36**, 1263–1269 (2007).

[57] T. P. J. Knowles and M. J. Buehler, Nanomechanics of functional and pathological amyloid materials. *Nat Nanotechnol* **6**, 469–479 (2011).

[58] M. F. Bishop and F. A. Ferrone, Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. *Biophys J* **46**, 631–644 (1984).

[59] T. C. T. Michaels, G. A. Garcia, and T. P. J. Knowles, Asymptotic solutions of the Oosawa model for the length distribution of biofilaments. *J Chem Phys* **140**, 194906 (2014).

[60] T. C. T. Michaels, S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Hamiltonian dynamics of protein filament formation. *Phys Rev Lett* **116**, 038101 (2016).

[61] J. Hofrichter, Kinetics of sickle hemoglobin polymerization. iii. nucleation rates determined from stochastic fluctuations in polymerization progress curves. *J Mol Biol* **189**, 553–571 (1986).

[62] T. P. J. Knowles, D. A. White, A. R. Abate, J. J. Agresti, S. I. A. Cohen, R. A. Sperling, E. J. D. Genst, C. M. Dobson, and D. A. Weitz, Observation of spatial propagation of amyloid assembly from single nuclei. *Proc Natl Acad Sci USA* **108**, 14746–14751 (2011).

[63] C. A. Stan, G. F. Schneider, S. S. Shevkoplyas, M. Hashimoto, M. Ibanescu, B. J. Wiley, and G. M. Whitesides, A microfluidic apparatus for the study of ice nucleation in supercooled water drops. *Lab Chip* **9**, 2293–2305 (2009).

[64] J. Szavits-Nossan, K. Eden, R. J. Morris, C. E. MacPhee, M. R. Evans, and R. J. Allen, Inherent variability in the kinetics of autocatalytic protein self-assembly. *Phys Rev Lett* **113**, 098101 (2014).

[65] M. Tanaka, S. R. Collins, B. H. Toyama, and J. S. Weissman, The physical basis of how prion conformations determine strain phenotypes. *Nature* **442**, 585–589 (2006).

[66] A. Wegner and P. Savko, Fragmentation of actin filaments. *Biochemistry (Mosc)* **21**, 1909–1913 (1982).

[67] C. B. Andersen, H. Yagi, M. Manno, V. Martorana, T. Ban, G. Christiansen, D. E. Otzen, Y. Goto, and C. Rischel, Branching in amyloid fibril growth. *Biophys J* **96**, 1529–1536 (2009).

[68] A. M. Ruschak and A. D. Miranker, Fiber-dependent amyloid formation as catalysis of an existing reaction pathway. *Proc Natl Acad Sci USA* **104**, 12341–12346 (2007).

[69] A. Cacciuto, S. Auer, and D. Frenkel, Onset of heterogeneous crystal nucleation in colloidal suspensions. *Nature* **428**, 404–406 (2004).

[70] E. Hellstrand, B. Boland, D. M. Walsh, and S. Linse, Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem Neurosci* **1**, 13–18 (2010).

[71] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics*. Cambridge University Press (2010).

[72] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 2340–2361 (1977).

[73] T. C. T. Michaels, A. J. Dear, J. B. Kirkegaard, K. L. Saar, D. A. Weitz, and T. P. J. Knowles, Fluctuations in the kinetics of linear protein self-assembly. *Phys Rev Lett* **116**, 258103 (2016).

[74] M. F. Hagan and O. M. Elrad, Understanding the concentration dependence of viral capsid assembly kinetics–the origin of the lag time and identifying the critical nucleus size. *Biophys J* **98**, 1065–1074 (2010).

[75] J.-L. Barrat and J.-P. Hansen, *Basic Concepts for Simple and Complex Liquids*. Cambridge University Press (2003).

[76] R. P. Sear, Quantitative studies of crystal nucleation at constant supersaturation: experimental data and models. *CrystEngComm* **16**, 6506–6522 (2014).

[77] S. I. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. Knowles, From macroscopic measurements to microscopic mechanisms of protein aggregation. *J Mol Biol* **421**, 160 – 171 (2012).

[78] K. Eden, R. Morris, J. Gillam, C. MacPhee, and R. Allen, Competition between primary nucleation and autocatalysis in amyloid fibril self-assembly. *Biophys J* **108**, 632 – 643 (2015).

[79] F. A. Ferrone and M. A. Rotter, Crowding and the polymerization of sickle hemoglobin. *J Mol Recognit* **17**, 497–504 (2004).

[80] L. M. Young, R. A. Mahood, J. C. Saunders, L.-H. Tu, D. P. Raleigh, S. E. Radford, and A. E. Ashcroft, Insights into the consequences of co-polymerisation in the early stages of IAPP and Aβ peptide assembly from mass spectrometry. *Analyst* **140**, 6990–6999 (2015).

[81] R. J. Moss, A. R. Mastri, and L. J. Schut, The coexistence and differentiation of late onset Huntington's disease and Alzheimer's disease: A case report and review of the literature. *J Am Geriatr Soc* **36**, 237–242 (1988).

[82] M. Tada, E. A. Coon, A. P. Osmand, P. A. Kirby, W. Martin, M. Wieler, A. Shiga, H. Shirasaki, M. Tada, T. Makifuchi, M. Yamada, A. Kakita, M. Nishizawa, H. Takahashi, and H. L. Paulson, Coexistence of Huntington's disease and amyotrophic lateral sclerosis: a clinicopathologic study. *Acta Neuropathol (Berl)* **124**, 749–760 (2012).

[83] C. J. Sarell, P. G. Stockley, and S. E. Radford, Assessing the causes and consequences of co-polymerization in amyloid formation. *Prion* **7**, 359–368 (2013).

[84] K. Dubey, B. G. Anand, M. K. Temgire, and K. Kar, Evidence of rapid coaggregation of globular proteins during amyloid formation. *Biochemistry (Mosc)* **53**, 8001–8004 (2014).

[85] R. Cukalevski, X. Yang, G. Meisl, U. Weininger, K. Bernfur, B. Frohm, T. P. J. Knowles, and S. Linse, The Aβ40 and Aβ42 peptides self-assemble into separate homomolecular fibrils in binary mixtures but cross-react during primary nucleation. *Chem Sci* **6**, 4215–4233 (2015).

[86] O. Szczepankiewicz, B. Linse, G. Meisl, E. Thulin, B. Frohm, C. S. Frigerio, M. T. Colvin, A. C. Jacavone, R. G. Griffin, T. Knowles, D. M. Walsh, and S. Linse, N-terminal extensions retard Aβ42 fibril formation but allow cross-seeding and coaggregation with Aβ42. *J Am Chem Soc* **137**, 14673–14685 (2015).

[87] M. A. Schilling, Unraveling Alzheimer's: Making sense of the relationship between diabetes and Alzheimer's disease. *J Alzheimers Dis* **51**, 961–977 (2016).

[88] K. J. Amann and T. D. Pollard, The Arp2/3 complex nucleates actin filament branches from the sides of pre-existing filaments. *Nat Cell Biol* **3**, 306 (2001).

[89] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Nucleated polymerisation in the presence of pre-formed seed filaments. *Int J Mol Sci* **12**, 5844–5852 (2011).

[90] T. Scheibel, J. Bloom, and S. L. Lindquist, The elongation of yeast prion fibers involves separable steps of association and conversion. *Proc Natl Acad Sci USA* **101**, 2287–2292 (2004).

[91] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh, and J. E. Maggio, Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry (Mosc)* **39**, 6288–6295 (2000).

[92] W.-F. Xue, A. L. Hellewell, W. S. Gosal, S. W. Homans, E. W. Hewitt, and S. E. Radford, Fibril fragmentation enhances amyloid cytotoxicity. *J Biol Chem* **284**, 34272–34282 (2009).

[93] S. Whitelam, R. Schulman, and L. Hedges, Self-assembly of multicomponent structures in and out of equilibrium. *Phys Rev Lett* **109**, 265506 (2012).

[94] F. Chiti and C. M. Dobson, Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu Rev Biochem* **86**, 27–68 (2017).

[95] D. M. Fowler, A. V. Koulov, W. E. Balch, and J. W. Kelly, Functional amyloid–from bacteria to humans. *Trends Biochem Sci* **32**, 217–224 (2007).

[96] B. Cheng, H. Gong, H. Xiao, R. B. Petersen, L. Zheng, and K. Huang, Inhibiting toxic aggregation of amyloidogenic proteins: A therapeutic strategy for protein misfolding diseases. *Biochim Biophys Acta, Gen Subj* **1830**, 4860 – 4871 (2013).

[97] M. J. Guerrero-Muñoz, D. L. Castillo-Carranza, and R. Kayed, Therapeutic approaches against common structural features of toxic oligomers shared by multiple amyloidogenic proteins. *Biochem Pharmacol* **88**, 468 – 478 (2014).

[98] P. R. Banerjee and A. A. Deniz, Shedding light on protein folding landscapes by single-molecule fluorescence. *Chem Soc Rev* **43**, 1172–1188 (2014).

[99] B. Schuler and H. Hofmann, Single-molecule spectroscopy of protein folding dynamics - expanding scope and timescales. *Curr Opin Struct Biol* **23**, 36 – 47 (2013).

[100] N. Cremades, S. I. A. Cohen, E. Deas, A. Y. Abramov, A. Y. Chen, A. Orte, M. Sandal, R. W. Clarke, P. Dunne, F. A. Aprile, C. W. Bertoncini, N. W. Wood, T. P. J. Knowles, C. M. Dobson, and D. Klenerman, Direct observation of the interconversion of normal and toxic forms of alpha-synuclein. *Cell* **149**, 1048–1059 (2012).

[101] P. Narayan, A. Orte, R. W. Clarke, B. Bolognesi, S. Hook, K. A. Ganzinger, S. Meehan, M. R. Wilson, C. M. Dobson, and D. Klenerman, The extracellular chaperone clusterin sequesters oligomeric forms of the amyloid-β(1-40) peptide. *Nat Struct Mol Biol* **19**, 79 (2011).

[102] R. B. Wickner, [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. *Science* **264**, 566–569 (1994).

[103] D. Blinder, P. W. Coschigano, and B. Magasanik, Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in Saccharomyces cerevisiae. *J Bacteriol* **178**, 4734–6 (1996).

[104] L. Pieri, M. Bucciantini, D. Nosi, L. Formigli, J. Savistchenko, R. Melki, and M. Stefani, The yeast prion Ure2p native-like assemblies are toxic to mammalian cells regardless of their aggregation state. *J Biol Chem* **281**, 15337–15344 (2006).

[105] C. Zhang, A. P. Jackson, Z.-R. Zhang, Y. Han, S. Yu, R.-Q. He, and S. Perrett, Amyloid-like aggregates of the yeast prion protein Ure2 enter vertebrate cells by specific endocytotic pathways and induce apoptosis. *PLoS One* **5**, 1–12 (2010).

[106] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani, Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–11 (2002).

[107] Y. Jiang, H. Li, L. Zhu, J.-M. Zhou, and S. Perrett, Amyloid nucleation and hierarchical assembly of Ure2p fibrils: Role of asparagine/glutamine repeat and nonrepeat regions of the prion domain. *J Biol Chem* **279**, 3361–3369 (2004).

[108] Y.-Q. Wang, A. K. Buell, X.-Y. Wang, M. E. Welland, C. M. Dobson, T. P. J. Knowles, and S. Perrett, Relationship between prion propensity and the rates of individual molecular steps of fibril assembly. *J Biol Chem* **286**, 12101–12107 (2011).

[109] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference.* Springer-Verlag New York (2002).

[110] L. Chen, L.-J. Chen, H.-Y. Wang, Y.-Q. Wang, and S. Perrett, Deletion of a Ure2 C-terminal prion-inhibiting region promotes the rate of fibril seed formation and alters interaction with Hsp40. *Protein Eng Des Sel* **24**, 69–78 (2011).

[111] L.-Q. Xu, S. Wu, A. K. Buell, S. I. A. Cohen, L.-J. Chen, W.-H. Hu, S. A. Cusack, L. S. Itzhaki, H. Zhang, T. P. J. Knowles, C. M. Dobson, M. E. Welland, G. W. Jones, and S. Perrett, Influence of specific hsp70 domains on fibril formation of the yeast prion protein Ure2. *Philos Trans R Soc London, Ser B* **368** (2013).

[112] J. Yang, A. J. Dear, T. C. T. Michaels, C. M. Dobson, T. P. J. Knowles, S. Wu, and S. Perrett, Direct observation of oligomerization by single molecule fluorescence reveals a multistep aggregation mechanism for the yeast prion protein Ure2. *J Am Chem Soc* **140**, 2493–2503 (2018).

[113] A. Šarić, Y. C. Chebaro, T. P. J. Knowles, and D. Frenkel, Crucial role of nonspecific interactions in amyloid nucleation. *Proc Natl Acad Sci USA* **111**, 17869–17874 (2014).

[114] A. Šarić, T. C. T. Michaels, A. Zaccone, T. P. J. Knowles, and D. Frenkel, Kinetics of spontaneous filament nucleation via oligomers: Insights from theory and simulation. *J Chem Phys* **145**, 211926 (2016).

[115] S. Auer, F. Meersman, C. M. Dobson, and M. Vendruscolo, A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comput Biol* **4**, e1000222 (2008).

[116] J. Lee, E. K. Culyba, E. T. Powers, and J. W. Kelly, Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. *Nat Chem Biol* **7**, 602–609 (2011).

[117] R. Krishnan, J. L. Goodman, S. Mukhopadhyay, C. D. Pacheco, E. A. Lemke, A. A. Deniz, and S. Lindquist, Conserved features of intermediates in amyloid assembly determine their benign or toxic states. *Proc Natl Acad Sci USA* **109**, 11172–11177 (2012).

[118] B. Bolognesi, J. R. Kumita, T. P. Barros, E. K. Esbjorner, L. M. Luheshi, D. C. Crowther, M. R. Wilson, C. M. Dobson, G. Favrin, and J. J. Yerbury, ANS binding reveals common features of cytotoxic amyloid species. *ACS Chem Biol* **5**, 735–740 (2010).

[119] S. Campioni, B. Mannini, M. Zampagni, A. Pensalfini, C. Parrini, E. Evangelisti, A. Relini, M. Stefani, C. M. Dobson, C. Cecchi, and F. Chiti, A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat Chem Biol* **6**, 140–147 (2010).

[120] E. B. Sawyer, D. Claessen, S. L. Gras, and S. Perrett, Exploiting amyloid: how and why bacteria use cross-β fibrils. *Biochem Soc Trans* **40**, 728–734 (2012).

[121] M. L. Evans and M. R. Chapman, Curli biogenesis: Order out of disorder. *Biochim Biophys Acta, Mol Cell Res* **1843**, 1551 – 1558 (2014).

[122] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, and P. T. Lansbury, Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc Natl Acad Sci USA* **97**, 571–576 (2000).

[123] A. D. Mcnaught and A. Wilkinson, *IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book").* Blackwell Scientific Publications; 2nd edition (1997), ISBN 0-9678550-9-8.

[124] M. H. Horrocks, L. Tosatto, A. J. Dear, G. A. Garcia, M. Iljina, N. Cremades, M. Dalla Serra, T. P. J. Knowles, C. M. Dobson, and D. Klenerman, Fast flow microfluidics and single-molecule fluorescence for the rapid characterization of α-synuclein oligomers. *Anal Chem* **87**, 8818–8826 (2015).

[125] T. C. T. Michaels, P. Yde, J. C. W. Willis, M. H. Jensen, D. Otzen, C. M. Dobson, A. K. Buell, and T. P. J. Knowles, The length distribution of frangible biofilaments. *J Chem Phys* **143**, 164901 (2015).

[126] T. C. T. Michaels, H. W. Lazell, P. Arosio, and T. P. J. Knowles, Dynamics of protein aggregation and oligomer formation governed by secondary nucleation. *J Chem Phys* **143**, 054901 (2015).

[127] L. Tosatto, M. H. Horrocks, A. J. Dear, T. P. J. Knowles, M. Dalla Serra, N. Cremades, C. M. Dobson, and D. Klenerman, Single-molecule FRET studies on alpha-synuclein oligomerization of Parkinson's disease genetically related mutants. *Sci Rep* **5**, 16696 (2015).

[128] M. Kjaergaard, A. J. Dear, F. Kundel, S. Qamar, G. Meisl, T. P. J. Knowles, and D. Klenerman, Oligomer diversity during the aggregation of the repeat region of tau. *ACS Chem Neurosci* **9**, 3060–3071 (2018).

[129] C. G. Glabe, Structural classification of toxic amyloid oligomers. *J Biol Chem* **283**, 29639–29643 (2008).

[130] T. C. Michaels, A. J. Dear, and T. P. Knowles, Scaling and dimensionality in the chemical kinetics of protein filament formation. *Int Rev Phys Chem* **35**, 679–703 (2016).

[131] M. Iljina, L. Hong, M. H. Horrocks, M. H. Ludtmann, M. L. Choi, C. D. Hughes, F. S. Ruggeri, T. Guilliams, A. K. Buell, J.-E. Lee, S. Gandhi, S. F. Lee, C. E. Bryant, M. Vendruscolo, T. P. J. Knowles, C. M. Dobson, E. De Genst, and D. Klenerman, Nanobodies raised against monomeric alpha-synuclein inhibit fibril formation and destabilize toxic oligomeric species. *BMC Biol* **15**, 57 (2017).

[132] D. J. Selkoe, Alzheimer's disease: Genes, proteins, and therapy. *Physiol Rev* **81**, 741–766 (2001).

[133] C. Haass, Take five—BACE and the γ-secretase quartet conduct Alzheimer's amyloid β-peptide generation. *EMBO J* **23**, 483–488 (2004).

[134] J. T. Jarrett, E. P. Berger, and P. T. Lansbury, The carboxy terminus of the β amyloid protein is critical for the seeding of amyloid formation: Implications for the pathogenesis of Alzheimer's disease. *Biochemistry (Mosc)* **32**, 4693–4697 (1993).

[135] S. A. Gravina, L. Ho, C. B. Eckman, K. E. Long, L. Otvos, L. H. Younkin, N. Suzuki, and S. G. Younkin, Amyloid beta protein (Aβ) in Alzheimer's disease brain. biochemical and immunocytochemical analysis with antibodies specific for forms ending at Aβ40 or Aβ42(43). *J Biol Chem* **270**, 7013–7016 (1995).

[136] T. Iwatsubo, A. Odaka, N. Suzuki, H. Mizusawa, N. Nukina, and Y. Ihara, Visualization of Aβ42(43) and Aβ40 in senile plaques with end-specific Aβ monoclonals: Evidence that an initially deposited species is Aβ42(43). *Neuron* **13**, 45 – 53 (1994).

[137] D. Scheuner, C. Eckman, M. Jensen, X. Song, M. Citron, N. Suzuki, T. D. Bird, J. Hardy, M. Hutton, W. Kukull, E. Larson, L. Levy-Lahad, M. Viitanen, E. Peskind, P. Poorkaj, G. Schellenberg, R. Tanzi, W. Wasco, L. Lannfelt, D. Selkoe, and S. Younkin, Secreted amyloid β-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and app mutations linked to familial Alzheimer's disease. *Nat Med* **2**, 864 (1996).

[138] K. N. Dahlgren, A. M. Manelli, W. B. Stine, L. K. Baker, G. A. Krafft, and M. J. LaDu, Oligomeric and fibrillar species of amyloid-β peptides differentially affect neuronal viability. *J Biol Chem* **277**, 32046–32053 (2002).

[139] K. Pauwels, T. L. Williams, K. L. Morris, W. Jonckheere, A. Vandersteen, G. Kelly, J. Schymkowitz, F. Rousseau, A. Pastore, L. C. Serpell, and K. Broersen, Structural basis for increased toxicity of pathological Aβ42:Aβ40 ratios in Alzheimer disease. *J Biol Chem* **287**, 5650–5660 (2012).

[140] I. Kuperstein, K. Broersen, I. Benilova, J. Rozenski, W. Jonckheere, M. Debulpaep, A. Vandersteen, I. Segers-Nolten, K. Van Der Werf, V. Subramaniam, D. Braeken, G. Callewaert, C. Bartic, R. D'Hooge, I. C. Martins, F. Rousseau, J. Schymkowitz, and B. De Strooper, Neurotoxicity of Alzheimer's disease Aβ peptides is induced by small changes in the Aβ42 to Aβ40 ratio. *EMBO J* **29**, 3408–3420 (2010).

[141] K. Duff, C. Eckman, C. Zehr, X. Yu, C.-M. Prada, J. Perez-tur, M. Hutton, L. Buee, Y. Harigaya, D. Yager, D. Morgan, M. N. Gordon, L. Holcomb, L. Refolo, B. Zenk, J. Hardy, and S. Younkin, Increased amyloid-β42(43) in brains of mice expressing mutant presenilin 1. *Nature* **383**, 710 (1996).

[142] M. Citron, D. Westaway, W. Xia, G. Carlson, T. Diehl, G. Levesque, K. Johnson-wood, M. Lee, P. Seubert, A. Davis, D. Kholodenko, R. Motter, R. Sherrington, B. Perry, H. Yao, R. Strome, I. Lieberburg, J. Rommens, S. Kim, D. Schenk, P. Fraser, P. St George Hyslop, and D. J. Selkoe, Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid β-protein in both transfected cells and transgenic mice. *Nat Med* **3**, 67 (1997).

[143] E. Hellstroem-Lindahl, M. Viitanen, and A. Marutle, Comparison of Aβ levels in the brain of familial and sporadic Alzheimer's disease. *Neurochem Int* **55**, 243–252 (2009).

[144] C. Haass and D. J. Selkoe, Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat Rev Mol Cell Biol* **8**, 101–112 (2007).

[145] G. Bitan, E. A. Fradinger, S. M. Spring, and D. B. Teplow, Neurotoxic protein oligomers - what you see is not always what you get. *Amyloid* **12**, 88–95 (2005).

[146] D. Frost, P. M. Gorman, C. M. Yip, and A. Chakrabartty, Co-incorporation of Aβ40 and Aβ42 to form mixed pre-fibrillar aggregates. *Eur J Biochem* **270**, 654–663 (2003).

[147] K. Hasegawa, I. Yamaguchi, S. Omata, F. Gejyo, and H. Naiki, Interaction between Aβ(1-42) and Aβ(1-40) in Alzheimer's β-amyloid fibril formation in vitro. *Biochemistry (Mosc)* **38**, 15514–15521 (1999).

[148] H. J. Kim, E. Chatani, Y. Goto, and S. R. Paik, Seed-dependent accelerated fibrillation of alpha-synuclein induced by periodic ultrasonication treatment. *J Microbiol Biotechnol* **17**, 2027–2032 (2007).

[149] M. M. Murray, S. L. Bernstein, V. Nyugen, M. M. Condron, D. B. Teplow, and M. T. Bowers, Amyloid β protein: Aβ40 inhibits Aβ42 oligomerization. *J Am Chem Soc* **131**, 6316–6317 (2009).

[150] Y. Yoshiike, D.-H. Chui, T. Akagi, N. Tanaka, and A. Takashima, Specific compositions of amyloid-β peptides as the determinant of toxic β-aggregation. *J Biol Chem* **278**, 23648–23655 (2003).

[151] R. Johnson, J. Schauerte, C.-C. Chang, K. Wisser, J. Althaus, C. Carruthers, M. Sutton, D. Steel, and A. Gafni, Single-molecule imaging reveals Aβ42:Aβ40 ratio-dependent oligomer growth on neuronal processes. *Biophys J* **104**, 894 – 903 (2013).

[152] P. D. Mehta, T. Pirttilä, S. P. Mehta, E. A. Sersen, P. S. Aisen, and H. M. Wisniewski, Plasma and cerebrospinal fluid levels of amyloid β proteins 1-40 and 1-42 in Alzheimer disease. *Arch Neurol* **57**, 100–105 (2000).

[153] J. R. Brender, J. Krishnamoorthy, M. F. M. Sciacca, S. Vivekanandan, L. D'Urso, J. Chen, C. La Rosa, and A. Ramamoorthy, Probing the sources of the apparent irreproducibility of amyloid formation: Drastic changes in kinetics and a switch in mechanism due to micellelike oligomer formation at critical concentrations of IAPP. *J Phys Chem B* **119**, 2886–2896 (2015).

[154] A. R. A. Ladiwala, J. Litt, R. S. Kane, D. S. Aucoin, S. O. Smith, S. Ranjan, J. Davis, W. E. Van Nostrand, and P. M. Tessier, Conformational differences between two amyloid β oligomers of similar size and dissimilar toxicity. *J Biol Chem* **287**, 24765–24773 (2012).

[155] R. Sabaté and J. Estelrich, Evidence of the existence of micelles in the fibrillogenesis of β-amyloid peptide. *J Phys Chem B* **109**, 11027–11032 (2005).

[156] B. Soreghan, J. Kosmoski, and C. Glabe, Surfactant properties of Alzheimer's Aβ peptides and the mechanism of amyloid aggregation. *J Biol Chem* **269**, 28551–28554 (1994).

[157] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand, and S. O. Smith, Structural conversion of neurotoxic amyloid-β(1-42) oligomers to fibrils. *Nat Struct Mol Biol* **17**, 561–567 (2010).

[158] K. Ono, M. M. Condron, and D. B. Teplow, Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proc Natl Acad Sci USA* **106**, 14745–14750 (2009).

[159] M. Fändrich, Oligomeric intermediates in amyloid formation: Structure determination and mechanisms of toxicity. *J Mol Biol* **421**, 427 – 440 (2012).

[160] A. Orte, R. Clarke, and D. Klenerman, Single-molecule two-colour coincidence detection to probe biomolecular associations. *Biochem Soc Trans* **38**, 914–918 (2010).

[161] M. Iljina, G. A. Garcia, A. J. Dear, J. Flint, P. Narayan, T. C. T. Michaels, C. M. Dobson, D. Frenkel, T. P. J. Knowles, and D. Klenerman, Quantitative analysis of co-oligomer formation by amyloid-β peptide isoforms. *Sci Rep* **6**, 28658 (2016).

[162] K. Brännström, A. Ã–hman, L. Nilsson, M. Pihl, L. Sandblad, and A. Olofsson, The N-terminal region of amyloid β controls the aggregation rate and fibril stability at low pH through a gain of function mechanism. *J Am Chem Soc* **136**, 10956–10964 (2014).

[163] B. O'Nuallain, S. Shivaprasad, I. Kheterpal, and R. Wetzel, Thermodynamics of Aβ(1-40) amyloid fibril elongation. *Biochemistry (Mosc)* **44**, 12709–12718 (2005).

[164] A. D. Williams, S. Shivaprasad, and R. Wetzel, Alanine scanning mutagenesis of Aβ(1-40) amyloid fibril stability. *J Mol Biol* **357**, 1283 – 1294 (2006).

[165] L. Sánchez, S. Madurga, T. Pukala, M. Vilaseca, C. López-Iglesias, C. V. Robinson, E. Giralt, and N. Carulla, Aβ40 and Aβ42 amyloid fibrils exhibit distinct molecular recycling properties. *J Am Chem Soc* **133**, 6505–6508 (2011).

[166] A.-C. Brorsson, B. Bolognesi, G. G. Tartaglia, S. L. Shammas, G. Favrin, I. Watson, D. A. Lomas, F. Chiti, M. Vendruscolo, C. M. Dobson, D. C. Crowther, and L. M. Luheshi, Intrinsic determinants of neurotoxic aggregate formation by the amyloid β peptide. *Biophys J* **98**, 1677 – 1684 (2010).

[167] A. J. Baldwin, T. P. J. Knowles, G. G. Tartaglia, A. W. Fitzpatrick, G. L. Devlin, S. L. Shammas, C. A. Waudby, M. F. Mossuto, S. Meehan, S. L. Gras, J. Christodoulou, S. J. Anthony-Cahill, P. D. Barker, M. Vendruscolo, and C. M. Dobson, Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc* **133**, 14160–14163 (2011).

[168] P. Narayan, K. M. Holmström, D.-H. Kim, D. J. Whitcomb, M. R. Wilson, P. St. George-Hyslop, N. W. Wood, C. M. Dobson, K. Cho, A. Y. Abramov, and D. Klenerman, Rare individual amyloid-β oligomers act on astrocytes to initiate neuronal damage. *Biochemistry (Mosc)* **53**, 2442–2453 (2014).

[169] P. Narayan, K. A. Ganzinger, J. McColl, L. Weimann, S. Meehan, S. Qamar, J. A. Carver, M. R. Wilson, P. St. George-Hyslop, C. M. Dobson, and D. Klenerman, Single molecule characterization of the interactions between amyloid-β peptides and the membranes of hippocampal cells. *J Am Chem Soc* **135**, 1491–1498 (2013).

[170] M. Hölttä, O. Hansson, U. Andreasson, J. Hertze, L. Minthon, K. Nägga, N. Andreasen, H. Zetterberg, and K. Blennow, Evaluating amyloid-β oligomers in cerebrospinal fluid as a biomarker for Alzheimer's disease. *PLoS One* **8**, 1–8 (2013).

[171] R. Hao, P. Nanduri, Y. Rao, R. S. Panichelli, A. Ito, M. Yoshida, and T.-P. Yao, Proteasomes activate aggresome disassembly and clearance by producing unanchored ubiquitin chains. *Mol Cell* **51**, 819–828 (2013).

[172] T. Yang, T. T. O'Malley, D. Kanmert, J. Jerecic, L. R. Zieske, H. Zetterberg, B. T. Hyman, D. M. Walsh, and D. J. Selkoe, A highly sensitive novel immunoassay specifically detects low levels of soluble Aβ oligomers in human cerebrospinal fluid. *Alzheimer's Research & Therapy* **7**, 14 (2015).

[173] S. F. Lichtenthaler, R. Wang, H. Grimm, S. N. Uljon, C. L. Masters, and K. Beyreuther, Mechanism of the cleavage specificity of Alzheimer's disease γ-secretase identified by phenylalanine-scanning mutagenesis of the transmembrane domain of the amyloid precursor protein. *Proc Natl Acad Sci USA* **96**, 3053–3058 (1999).

[174] C. Guardia-Laguarta, M. Pera, J. Clarimón, J. L. Molinuevo, R. Sànchez-Valle, A. Lladó, M. Coma, T. Gómez-Isla, R. Blesa, I. Ferrer, and A. Lleó, Clinical, neuropathologic, and biochemical profile of the amyloid precursor protein I716F mutation. *J Neuropathol Exp Neurol* **69**, 53–59 (2010).

[175] M. Suárez-Calvet, O. Belbin, M. Pera, N. Badiola, J. Magrané, C. Guardia-Laguarta, L. Muñoz, M. Colom-Cadena, J. Clarimón, and A. Lleó, Autosomal-

dominant Alzheimer's disease mutations at the same codon of amyloid precursor protein differentially alter Aβ production. *J Neurochem* **128**, 330–339 (2014).

[176] M. I. Olson and C.-M. Shaw, Presenile dementia and Alzheimer's disease in mongolism. *Brain* **92**, 147–156 (1969).

[177] K. H. Ashe and A. Aguzzi, Prions, prionoids and pathogenic proteins in Alzheimer disease. *Prion* **7**, 55–59 (2013).

[178] M. G. Spillantini, Parkinson's disease, dementia with Lewy bodies and multiple system atrophy are α-synucleinopathies. *Parkinsonism & Related Disorders* **5**, 157 – 162 (1999).

[179] M. G. Spillantini, M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes, and M. Goedert, Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).

[180] L. V. Kalia, S. K. Kalia, P. J. McLean, A. M. Lozano, and A. E. Lang, α-synuclein oligomers and clinical implications for Parkinson disease. *Ann Neurol* **73**, 155–169 (2013).

[181] M. J. Volles, S.-J. Lee, J.-C. Rochet, M. D. Shtilerman, T. T. Ding, J. C. Kessler, and P. T. Lansbury, Vesicle permeabilization by protofibrillar α-synuclein: Implications for the pathogenesis and treatment of Parkinson's disease. *Biochemistry (Mosc)* **40**, 7812–7819 (2001).

[182] M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe, and R. L. Nussbaum, Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047 (1997).

[183] Z. J. J., A. Javier, G.-E. J. C., L. Elena, R. Raquel, A. Israel, V. Lídice, H. Janet, R. Olga, A. Begoña, L. Verónica, T. E. Gomez, del Ser Teodoro, M. D. G., and de Yebenes Justo G., The new mutation, E46K, of α-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* **55**, 164–173 (2004).

[184] R. Krüger, W. Kuhn, T. Müller, D. Woitalla, M. Graeber, S. Kösel, H. Przuntek, J. T. Epplen, L. Schols, and O. Riess, AlaSoPro mutation in the gene encoding α-synuclein in Parkinson's disease. *Nat Genet* **18**, 106 (1998).

[185] G. W. R. and L. A. E., Interface between tauopathies and synucleinopathies: A tale of two proteins. *Ann Neurol* **59**, 449–458 (2006).

[186] T. L. Spires-Jones, J. Attems, and D. R. Thal, Interactions of pathological proteins in neurodegenerative diseases. *Acta Neuropathol (Berl)* **134**, 187–205 (2017).

[187] L. Pieri, K. Madiona, and R. Melki, Structural and functional properties of prefibrillar α-synuclein oligomers. *Sci Rep* **6**, 24526 (2016).

[188] K. M. Danzer, D. Haasen, A. R. Karow, S. Moussaud, M. Habeck, A. Giese, H. Kretzschmar, B. Hengerer, and M. Kostka, Different species of α-synuclein oligomers induce calcium influx and seeding. *J Neurosci* **27**, 9220–9232 (2007).

[189] S. W. Chen, S. Drakulic, E. Deas, M. Ouberai, F. A. Aprile, R. Arranz, S. Ness, C. Roodveldt, T. Guilliams, E. J. De-Genst, D. Klenerman, N. W. Wood, T. P. Knowles, C. Alfonso, G. Rivas, A. Y. Abramov, J. M. Valpuesta, C. M. Dobson, and N. Cremades, Structural characterization of toxic oligomers that are kinetically trapped during α-synuclein fibril formation. *Proc Natl Acad Sci USA* **112**, E1994–E2003 (2015).

[190] N. Lorenzen, S. B. Nielsen, A. K. Buell, J. D. Kaspersen, P. Arosio, B. S. Vad, W. Paslawski, G. Christiansen, Z. Valnickova-Hansen, M. Andreasen, J. J. Enghild, J. S. Pedersen, C. M. Dobson, T. P. J. Knowles, and D. E. Otzen, The role of stable α-synuclein oligomers in the molecular events underlying amyloid formation. *J Am Chem Soc* **136**, 3859–3868 (2014).

[191] L. Narhi, S. J. Wood, S. Steavenson, Y. Jiang, G. M. Wu, D. Anafi, S. A. Kaufman, F. Martin, K. Sitney, P. Denis, J.-C. Louis, J. Wypych, A. L. Biere, and M. Citron, Both familial Parkinson's disease mutations accelerate α-synuclein aggregation. *J Biol Chem* **274**, 9843–9846 (1999).

[192] J. Li, V. N. Uversky, and A. L. Fink, Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human α-synuclein. *Biochemistry (Mosc)* **40**, 11604–11613 (2001).

[193] R. A. Fredenburg, C. Rospigliosi, R. K. Meray, J. C. Kessler, H. A. Lashuel, D. Eliezer, and P. T. Lansbury, The impact of the E46K mutation on the properties of α-synuclein in its monomeric and oligomeric states. *Biochemistry (Mosc)* **46**, 7107–7118 (2007).

[194] H. A. Lashuel, D. Hartley, B. M. Petre, T. Walz, and P. T. Lansbury, Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature* **418**, 291 (2002).

[195] E. Sierecki, N. Giles, Q. Bowden, M. E. Polinkovsky, J. Steinbeck, N. Arrioti, D. Rahman, A. Bhumkar, P. R. Nicovich, I. Ross, R. G. Parton, T. Böcking, and Y. Gambin, Nanomolar oligomerization and selective co-aggregation of α-synuclein pathogenic mutants revealed by single-molecule fluorescence. *Sci Rep* **6**, 37630 (2016).

[196] S. Chia, P. Flagmeier, J. Habchi, V. Lattanzi, S. Linse, C. M. Dobson, T. P. J. Knowles, and M. Vendruscolo, Monomeric and fibrillar α-synuclein exert opposite effects on the catalytic cycle that promotes the proliferation of Aβ42 aggregates. *Proc Natl Acad Sci USA* **114**, 8005–8010 (2017).

[197] M. Yoshimoto, A. Iwai, D. Kang, D. A. Otero, Y. Xia, and T. Saitoh, NACP, the precursor protein of the non-amyloid beta/A4 protein (Aβ) component of Alzheimer disease amyloid, binds Aβ and stimulates Aβ aggregation. *Proc Natl Acad Sci USA* **92**, 9141–9145 (1995).

[198] P. H. Jensen, P. Hojrup, H. Hager, M. S. Nielsen, L. Jacobsen, O. F. Olesen, J. Gliemann, and R. Jakes, Binding of Aβ to alpha- and beta-synucleins: identification of segments in alpha-synuclein/NAC precursor that bind Aβ and NAC. *Biochem J* **323**, 539–546 (1997).

[199] P. K. Mandal, J. W. Pettegrew, E. Masliah, R. L. Hamilton, and R. Mandal, Interaction between Aβ peptide and α synuclein: Molecular mechanisms in overlapping pathology of Alzheimer's and Parkinson's in dementia with Lewy body disease. *Neurochem Res* **31**, 1153–1162 (2006).

[200] P. H. Jensen, H. Hager, M. S. Nielsen, P. Højrup, J. Gliemann, and R. Jakes, α-synuclein binds to tau and stimulates the protein kinase a-catalyzed tau phosphorylation of serine residues 262 and 356. *J Biol Chem* **274**, 25481–25489 (1999).

[201] A. Esposito, C. P. Dohm, P. Kermer, M. Bähr, and F. S. Wouters, α-synuclein and its disease-related mutants interact differentially with the microtubule protein tau and associate with the actin cytoskeleton. *Neurobiol Dis* **26**, 521 – 531 (2007).

[202] H. Y. Qureshi and H. K. Paudel, Parkinsonian neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) and α-synuclein mutations promote tau protein phosphorylation at ser262 and destabilize microtubule cytoskeleton in vitro. *J Biol Chem* **286**, 5055–5068 (2011).

[203] I. F. Tsigelny, L. Crews, P. Desplats, G. M. Shaked, Y. Sharikov, H. Mizuno, B. Spencer, E. Rockenstein, M. Trejo, O. Platoshyn, J. X.-J. Yuan, and E. Masliah, Mechanisms of hybrid oligomer formation in the pathogenesis of combined Alzheimer's and Parkinson's diseases. *PLoS One* **3**, 1–15 (2008).

[204] J. C. Jose, P. Chatterjee, and N. Sengupta, Cross dimerization of amyloid-β and α-synuclein proteins in aqueous environment: A molecular dynamics simulations study. *PLoS One* **9**, 1–13 (2014).

[205] Y. Atsmon-Raz and Y. Miller, Non-amyloid-β component of human α-synuclein oligomers induces formation of new Aβ oligomers: Insight into the mechanisms that link Parkinson's and Alzheimer's diseases. *ACS Chem Neurosci* **7**, 46–55 (2016).

[206] R. De Ricco, V. Daniela, D. Simone, C. Luigi, H. Christelle, and F. Peter, Copper(i/ii), α/β-synuclein and amyloid-β: Menage à trois? *Chembiochem* **16**, 2319–2328 (2015).

[207] G. Nübling, B. Bader, J. Levin, J. Hildebrandt, H. Kretzschmar, and A. Giese, Synergistic influence of phosphorylation and metal ions on tau oligomer formation and coaggregation with α-synuclein at the single molecule level. *Mol Neurodegener* **7**, 35 (2012).

[208] U. Sengupta, M. J. Guerrero-Muñoz, D. L. Castillo-Carranza, C. A. Lasagna-Reeves, J. E. Gerson, A. A. Paulucci-Holthauzen, S. Krishnamurthy, M. Farhed, G. R. Jackson, and R. Kayed, Pathological interface between oligomeric alpha-synuclein and tau in synucleinopathies. *Biol Psychiatry* **78**, 672 – 683 (2015).

[209] S. Daniele, D. Pietrobono, J. Fusi, C. Iofrida, L. Chico, L. Petrozzi, A. L. Gerfo, F. Baldacci, F. Galetta, G. Siciliano, U. Bonuccelli, G. Santoro, M. L. Trincavelli, F. Franzoni, and C. Martini, α-synuclein aggregates with β-amyloid or tau in human red blood cells: Correlation with antioxidant capability and physical exercise in human healthy subjects. *Mol Neurobiol* **55**, 2653–2675 (2018).

[210] H. Wille, G. Drewes, J. Biernat, E. M. Mandelkow, and E. Mandelkow, Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau in vitro. *J Cell Biol* **118**, 573–584 (1992).

[211] S. Jeganathan, M. von Bergen, E.-M. Mandelkow, and E. Mandelkow, The natively unfolded character of tau and its aggregation to Alzheimer-like paired helical filaments. *Biochemistry (Mosc)* **47**, 10526–10539 (2008).

[212] H.-J. Lee, E.-J. Bae, and S.-J. Lee, Extracellular α-synuclein–a novel and crucial factor in Lewy body diseases. *Nat Rev Neurol* **10**, 92 (2014).

[213] A. Himmler, D. Drechsel, M. W. Kirschner, and D. W. Martin, Tau consists of a set of proteins with repeated C-terminal microtubule-binding domains and variable N-terminal domains. *Mol Cell Biol* **9**, 1381–1388 (1989).

[214] C. S. Atwood, R. N. Martins, M. A. Smith, and G. Perry, Senile plaque composition and posttranslational modification of amyloid-β peptide and associated proteins. *Peptides* **23**, 1343 – 1350 (2002).

[215] C. F. Wright, S. A. Teichmann, J. Clarke, and C. M. Dobson, The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**, 878–881 (2005).

[216] A. B. Singleton, M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum, S. Lincoln, A. Crawley, M. Hanson, D. Maraganore, C. Adler, M. R. Cookson, M. Muenter, M. Baptista, D. Miller, J. Blancato, J. Hardy, and K. Gwinn-Hardy, α-synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841–841 (2003).

[217] T. B. Ahn, S. Y. Kim, J. Y. Kim, S. S. Park, D. S. Lee, H. J. Min, Y. K. Kim, S. E. Kim, J. M. Kim, H. J. Kim, J. Cho, and B. S. Jeon, α-synuclein gene duplication is present in sporadic Parkinson disease. *Neurology* **70**, 43–49 (2008).

[218] P. Ibáñez, A.-M. Bonnet, B. Débarges, E. Lohmann, F. Tison, Y. Agid, A. Dürr, A. Brice, and P. Pollak, Causal relation between α-synuclein locus duplication as a cause of familial Parkinson's disease. *The Lancet* **364**, 1169 – 1171 (2004).

[219] C. Galvagnion, J. W. P. Brown, M. M. Ouberai, P. Flagmeier, M. Vendruscolo, A. K. Buell, E. Sparr, and C. M. Dobson, Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of α-synuclein. *Proc Natl Acad Sci USA* **113**, 7065–7070 (2016).

[220] C. Galvagnion, A. K. Buell, G. Meisl, T. C. T. Michaels, M. Vendruscolo, T. P. J. Knowles, and C. M. Dobson, Lipid vesicles trigger α-synuclein aggregation by stimulating primary nucleation. *Nat Chem Biol* **11**, 229 – 234 (2015).

[221] E. Emmanouilidou, D. Elenis, T. Papasilekas, G. Stranjalis, K. Gerozissis, P. C. Ioannou, and K. Vekrellis, Assessment of α-synuclein secretion in mouse and human brain parenchyma. *PLoS One* **6**, 1–9 (2011).

[222] A. Demuro, E. Mina, R. Kayed, S. C. Milton, I. Parker, and C. G. Glabe, Calcium dysregulation and membrane disruption as a ubiquitous neurotoxic mechanism of soluble amyloid oligomers. *J Biol Chem* **280**, 17294–17300 (2005).

[223] P. Flagmeier, S. De, D. C. Wirthensohn, S. F. Lee, C. Vincke, S. Muyldermans, T. P. J. Knowles, S. Gandhi, C. M. Dobson, and D. Klenerman, Ultrasensitive measurement of Ca2+ influx into lipid vesicles induced by protein aggregates. *Angew Chem Int Ed* **56**, 7750–7754 (2017).

[224] A. P. Kiely, Y. T. Asi, E. Kara, P. Limousin, H. Ling, P. Lewis, C. Proukakis, N. Quinn, A. J. Lees, J. Hardy, T. Revesz, H. Houlden, and J. L. Holton, α-synucleinopathy associated with G51D SNCA mutation: a link between Parkinson's disease and multiple system atrophy? *Acta Neuropathol (Berl)* **125**, 753–769 (2013).

[225] P. Pasanen, L. Myllykangas, M. Siitonen, A. Raunio, S. Kaakkola, J. Lyytinen, P. J. Tienari, M. Pöyhönen, and A. Paetau, A novel α-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol Aging* **35**, 2180.e1 – 2180.e5 (2014).

[226] J. F. Kellie, R. E. Higgs, J. W. Ryder, A. Major, T. G. Beach, C. H. Adler, K. Merchant, and M. D. Knierman, Quantitative measurement of intact alpha-synuclein proteoforms from post-mortem control and Parkinson's disease brain tissue by intact protein mass spectrometry. *Sci Rep* **4**, 5797 (2014).

[227] M. Takami, Y. Nagashima, Y. Sano, S. Ishihara, M. Morishima-Kawashima, S. Funamoto, and Y. Ihara, γ-secretase: Successive tripeptide and tetrapeptide release from the transmembrane domain of β-carboxyl terminal fragment. *J Neurosci* **29**, 13042–13052 (2009).

[228] C. Guerrero, C. Tagwerker, P. Kaiser, and L. Huang, An integrated mass spectrometry-based proteomic approach: Quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (qtax) to decipher the 26 s proteasome-interacting network. *Mol Cell Proteomics* **5**, 366–378 (2006).

[229] B. Efron, Bootstrap methods: Another look at the jackknife. *Ann Statist* **7**, 1–26 (1979).

[230] J. N. K. Rao and C. F. J. Wu, Resampling inference with complex survey data. *J Am Stat Assoc* **83**, 231–241 (1988).

[231] P. J. Bickel, F. Götze, and W. R. van Zwet, Resampling fewer than n observations: Gains, losses, and remedies for losses. *Stat Sinica* **7**, 1–31 (1997).

[232] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Nucleated polymerization with secondary pathways. iii. equilibrium behavior and oligomer populations. *J Chem Phys* **135**, 065107 (2011).

[233] F. E. Cohen and J. W. Kelly, Therapeutic approaches to protein-misfolding diseases. *Nature* **426**, 905–909 (2003).

[234] M. B. Pepys, Pathogenesis, diagnosis and treatment of systemic amyloidosis. *Phil Trans R Soc Lond B* **356**, 203–210; discussion 210–211 (2001).

[235] J. Hardy and D. J. Selkoe, The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).

[236] J. C. Sacchettini and J. W. Kelly, Therapeutic strategies for human amyloid diseases. *Nat Rev Drug Discov* **1**, 267–275 (2002).

[237] J. Hofrichter, P. D. Ross, and W. A. Eaton, Kinetics and mechanism of deoxyhemoglobin s gelation: a new approach to understanding sickle cell disease. *Proc Natl Acad Sci USA* **71**, 4864–4868 (1974).

[238] S. B. Prusiner, Molecular biology of prion diseases. *Science* **252**, 1515–1522 (1991).

[239] J. H. Come, P. E. Fraser, and P. T. Lansbury, A kinetic model for amyloid formation in the prion diseases: importance of seeding. *Proc Natl Acad Sci USA* **90**, 5959–5963 (1993).

[240] A. Aguzzi and C. Haass, Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science* **302**, 814–818 (2003).

[241] J. Falsig, K. P. Nilsson, T. P. J. Knowles, and A. Aguzzi, Chemical and biophysical insights into the propagation of prion strains. *HFSP J* **2**, 332–341 (2008).

[242] H. A. Lashuel, C. R. Overk, A. Oueslati, and E. Masliah, The many faces of α-synuclein: from structure and toxicity to therapeutic target. *Nat Rev Neurosci* **14**, 38–48 (2013).

[243] Y. Ren and N. Sahara, Characteristics of tau oligomers. *Front Neurol* **4**, 102 (2013).

[244] K. R. Patterson, C. Remmers, Y. Fu, S. Brooker, N. M. Kanaan, L. Vana, S. Ward, J. F. Reyes, K. Philibert, M. J. Glucksman, and L. I. Binder, Characterization of prefibrillar tau oligomers in vitro and in Alzheimer disease. *J Biol Chem* **286**, 23063–23076 (2011).

[245] D. M. Walsh, B. P. Tseng, R. E. Rydel, M. B. Podlisny, and D. J. Selkoe, The oligomerization of amyloid β-protein begins intracellularly in cells derived from human brain. *Biochemistry (Mosc)* **39**, 10831–10839 (2000).

[246] J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham, Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *J Chem Soc, Faraday Trans 2* **72**, 1525–1568 (1976).

[247] L. Maibaum, A. R. Dinner, and D. Chandler, Micelle formation and the hydrophobic effect. *J Phys Chem B* **108**, 6778–6781 (2004).

[248] H.-J. Woo, C. Carraro, and D. Chandler, Assembly of extended interfaces and micelles: charge frustrated models of amphiphilic mixtures. *Faraday Discuss* **104**, 183–191 (1996).

[249] F. H. Stillinger, Variational model for micelle structure. *J Chem Phys* **78**, 4654–4661 (1983).

[250] G. Meng, N. Arkus, M. P. Brenner, and V. N. Manoharan, The free-energy landscape of clusters of attractive hard spheres. *Science* **327**, 560–563 (2010).

[251] R. Pool and P. G. Bolhuis, Accurate free energies of micelle formation. *J Phys Chem B* **109**, 6650–6657 (2005).

[252] P. H. Nelson, G. C. Rutledge, and T. A. Hatton, On the size and shape of self-assembled micelles. *J Chem Phys* **107**, 10777–10781 (1997).

[253] F. Oosawa, Size distribution of protein polymers. *J Theor Biol* **27**, 69–86 (1970).

[254] Y. Mimori-Kiyosue, Shaping microtubules into diverse patterns: Molecular connections for setting up both ends. *Cytoskeleton* **68**, 603–618 (2011).

[255] C. M. Dobson, The amyloid phenomenon and its links with human disease. *Cold Spring Harbor Perspect Biol* **9** (2017).

[256] G. Wei, Z. Su, N. P. Reynolds, P. Arosio, I. W. Hamley, E. Gazit, and R. Mezzenga, Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chem Soc Rev* **46**, 4661–4708 (2017).

[257] T. P. J. Knowles and R. Mezzenga, Amyloid fibrils as building blocks for natural and artificial functional materials. *Adv Mater* **28**, 6546–6561 (2016).

[258] K. G. Wilson, The renormalization group: Critical phenomena and the kondo problem. *Rev Mod Phys* **47**, 773–840 (1975).

[259] N. Goldenfeld, *Lectures on phase transitions and the renormalization group*. Frontiers in physics, Addison-Wesley, Advanced Book Program (1992), ISBN 9780201554083.

[260] L.-Y. Chen, N. Goldenfeld, and Y. Oono, Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory. *Phys Rev E* **54**, 376–394 (1996).

[261] L. Y. Chen, N. Goldenfeld, and Y. Oono, Renormalization group theory for global asymptotic analysis. *Phys Rev Lett* **73**, 1311–1315 (1994).

[262] A. Wegner, Kinetic analysis of actin assembly suggests that tropomyosin inhibits spontaneous fragmentation of actin filaments. *J Mol Biol* **161**, 217–227 (1982).

[263] A. Šarić, A. Buell, G. Meisl, T. C. T. Michaels, C. M. Dobson, S. Linse, T. P. J. Knowles, and D. Frenkel, Physical determinants of the self-replication of protein fibrils. *Nat Phys* **12**, 874–880 (2016).

[264] A. Cacciuto, S. Auer, and D. Frenkel, Breakdown of classical nucleation theory near isostructural phase transitions. *Phys Rev Lett* **93**, 166105 (2004).

[265] T. C. T. Michaels, A. Šarić, J. Habchi, S. Chia, G. Meisl, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation. *Annu Rev Phys Chem* **69**, 273–298 (2018).

[266] M. Zhu, J. Li, and A. L. Fink, The association of alpha-synuclein with membranes affects bilayer structure, stability, and fibril formation. *J Biol Chem* **278**, 40186–40197 (2003).

[267] N. Ferguson, J. Berriman, M. Petrovich, T. D. Sharpe, J. T. Finch, and A. R. Fersht, Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc Natl Acad Sci USA* **100**, 9814–9819 (2003).

[268] Z.-R. Zhang and S. Perrett, Novel glutaredoxin activity of the yeast prion protein Ure2 reveals a native-like dimer within fibrils. *J Biol Chem* **284**, 14058–14067 (2009).

[269] D. Galani, A. R. Fersht, and S. Perrett, Folding of the yeast prion protein Ure2: kinetic evidence for folding and unfolding intermediates. *J Mol Biol* **315**, 213 – 227 (2002).

[270] A. K. Buell, C. Galvagnion, R. Gaspar, E. Sparr, M. Vendruscolo, T. P. J. Knowles, S. Linse, and C. M. Dobson, Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. *Proc Natl Acad Sci USA* **111**, 7671–7676 (2014).

[271] C. Tanford, Theory of micelle formation in aqueous solutions. *J Phys Chem* **78**, 2469–2479 (1974).

[272] S. Puvvada and D. Blankschtein, Molecular-thermodynamic approach to predict micellization, phase behavior and phase separation of micellar solutions. i. application to nonionic surfactants. *J Chem Phys* **92**, 3710–3724 (1990).