Aggregation of therapeutic proteins from dilute towards concentrated conditions

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. Sc. ETH Zurich)

presented by
Lucrèce Hélène NICOUĐ

Dipl. Ing. École Nationale Supérieure des Industries Chimiques
born on 17.09.1988
citizen of France

accepted on the recommendation of
Prof. Dr. Massimo Morbidelli, examiner
Prof. Dr. Peter Fischer, co-examiner
Dr. Edith Norrant, co-examiner

2016
Acknowledgements

Many people were involved in one way or another in this work, and I wish to deeply thank them for their contribution. Prof. Massimo Morbidelli, for offering me the opportunity to perform my PhD thesis in his group, for giving me both support and freedom to conduct this project, and for setting such a great work environment. Prof. Peter Fischer and Dr. Edith Norrant, for having accepted to be my co-examiners. UCB Pharma for financial support and for material supplying. The foundation Claude et Giuliana and the Swiss National Foundation (grant number 200020_147137/1) for financial support. Dr. Paolo Arosio, for his valuable guidance, especially at the beginning of my doctoral studies. Prof. Marco Lattuada, for his sound advice. My students Margaux Sozo, Daniel Balderas Barragán, Nicholas Cohrs and Jakub Jagielski for their precious help with experimental studies. My colleagues and friends in Zurich, who made these four years extremely enjoyable. Rushd, for all the memorable moments we shared since our very first day at ETH, for his keen sense of humour, and his delicious Lebanese dinners. Marta, with whom it was a pleasure to share my office, for being always happy to give a hand. Ste, for his contribution to several chapters, for wonderful short journeys and thrilling long talks. Anna for her good mood and her boundless enthusiasm. Baptiste for his scientific expertise and his amusing French phrases. Peter for his übergail cocktails and parties. My friends outside Zurich, and in particular Rémy et Fab, who make my weekends in Nancy so delightful. My family, and especially my parents and my brother, for their whole-hearted support and their belief in me. David, who stands at my side through thick and thin, for his love and care.
Abstract

The present thesis aims at gaining fundamental knowledge on the kinetics and thermodynamics of therapeutic protein aggregation, which is an essential requirement in the development of strategies seeking to improve the stability of biopharmaceuticals. To do so, experimental characterization (mainly carried out with size exclusion chromatography and light scattering) is combined with theoretical models based on key concepts from colloid science. The first part of the thesis deals with protein aggregation under dilute conditions. It is shown that, under given operating conditions, different monoclonal antibodies can follow different aggregation pathways, and that electrostatic repulsion alone cannot explain the stability of antibody solutions. Then, the impact of operative parameters (cosolutes and solution pH) on the protein conformational and colloidal stabilities is investigated. Particular attention is drawn to the stabilization of antibodies by polyols, which are commonly used as excipients in drug formulation. Importantly, it is found that the stabilization effect does not depend only on the volume fraction occupied by the polyol molecules, but also on the polyol chemical structure. The second part of this thesis is dedicated to protein aggregation under concentrated conditions, that are relevant for the long-term storage of protein-based drugs. After showing that the rise in solution viscosity due to aggregate formation can be rationalized by using the concept of occupied volume fraction, the effective viscosity experienced by the aggregates of different sizes is quantified by means of Brownian dynamics simulations and introduced in the aggregation rate constant. Finally, a correction term accounting for thermodynamic and hydrodynamic non-idealities arising at high protein concentration due to protein-protein interactions is proposed.
Résumé

L’objectif de cette thèse est d’acquérir une meilleure compréhension de la cinétique et de la thermodynamique d’agrégation de protéines thérapeutiques dans le but d’améliorer la stabilité de produits biopharmaceutiques. Pour ce faire, l’agrégation est étudiée à la fois de manière expérimentale, principalement par chromatographie d’exclusion stérique et diffusion de lumière, et de manière théorique, en utilisant des concepts clés de la science des colloïdes. La première partie de cette thèse traite de l’agrégation de protéines en conditions diluées. Il est montré que, dans des conditions opératoires données, différents anticorps peuvent suivre des mécanismes d’agrégation différents, et que la répulsion électrostatique seule ne permet pas d’expliquer la stabilité des anticorps. Ensuite, l’impact de paramètres opérationnels (cosolutés et pH) sur les stabilités conformationnelle et colloïdale des protéines est examiné. Une attention particulière est portée sur la stabilisation d’anticorps par des polyols, couramment utilisés comme excipients. Notamment, il est observé que l’effet de stabilisation ne dépend pas uniquement de la fraction volumique occupée par le polyol mais aussi de sa structure chimique. La seconde partie de cette thèse est dédiée à l’agrégation de protéines en conditions concentrées, qui sont essentielles pour la formulation de médicaments. Après avoir montré que l’augmentation de viscosité lors de l’agrégation peut être rationalisée par le concept de fraction volumique occupée, la viscosité effective ressentie par des agrégats de différentes tailles est quantifiée grâce à des simulations du mouvement Brownien, puis est introduite dans la constante de vitesse d’agrégation. Enfin, un terme de correction lié aux non-idéalités thermodynamiques et hydrodynamiques dues aux interactions entre protéines est proposé.
“What gets us into trouble is not what we don’t know. It’s what we know for sure that just ain’t so.”

Mark Twain
Contents

Acknowledgements iii
Abstract v
Résumé vii
Contents xi

1 Introduction 1
1.1 Aims and scope of the thesis . . . . . . . . . . . . . . . . . . . . . 1
1.2 Literature review . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
1.2.1 Atomistic simulations . . . . . . . . . . . . . . . . . . . . . . . 4
1.2.2 Molecular interaction potentials . . . . . . . . . . . . . . . . . 5
1.2.2.1 Second virial coefficient . . . . . . . . . . . . . . . 5
1.2.2.2 Colloidal stability and conformational stability . . 8
1.2.3 Formation of mesoscopic aggregates . . . . . . . . . . . . . . 9
1.2.3.1 Protein aggregation mechanisms . . . . . . . . . . 9
1.2.3.2 Experimental characterization . . . . . . . . . . . . 11
1.2.3.3 Aggregate fractal geometry . . . . . . . . . . . . . 12
1.2.3.4 Population balance equations modeling . . . . . . 14
1.2.4 Macroscopic properties . . . . . . . . . . . . . . . . . . . . . . 17
1.2.4.1 Viscosity increase . . . . . . . . . . . . . . . . . . . 17
1.2.4.2 Gel formation . . . . . . . . . . . . . . . . . . . . . 19
1.2.4.3 Phase diagrams . . . . . . . . . . . . . . . . . . . . 19
1.3 Outline of the thesis . . . . . . . . . . . . . . . . . . . . . . . . . 20

2 Multistep aggregation mechanisms of monoclonal antibodies 23
2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
2.2 Materials and methods . . . . . . . . . . . . . . . . . . . . . . . . . 25
2.2.1 Materials . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
2.2.2 Zeta potential measurements . . . . . . . . . . . . . . . . . . . 25
2.2.3 Thermal stability by circular dichroism . . . . . . . . . . . . 26
2.2.4 Thermal stability by dynamic light scattering . . . . . . . 26
2.2.5 Isothermal aggregation kinetics . . . . . . . . . . . . . . . . 27
2.2.6 Size exclusion chromatography with inline light scattering . 27
2.2.7 Evaluation of the fractal dimension by static light scattering 28
3 Fragmentation mechanisms of insulin fibrils 53
3.1 Introduction ........................................... 53
3.2 Materials and methods ............................... 56
  3.2.1 Sample preparation and experimental conditions ........ 56
  3.2.2 Size exclusion chromatography ....................... 57
  3.2.3 ThT fluorescence .................................. 57
  3.2.4 Atomic force microscopy ........................... 58
  3.2.5 Kinetic scheme .................................... 58
  3.2.6 Choice of the breakage kernel ....................... 60
  3.2.7 Population balance equations ....................... 61
  3.2.8 Numerical solution ................................ 62
  3.2.9 Optimization procedure ............................ 64
3.3 Results .............................................. 65
3.4 Discussion ........................................... 71
  3.4.1 Impact of agitation ................................ 71
  3.4.2 Impact of temperature ............................. 72
  3.4.3 Comparison with literature data ..................... 73
3.5 Conclusion ........................................... 76

4 Role of cosolutes in the kinetics of mAb aggregation 77
4.1 Introduction ........................................... 77
4.2 Materials and methods ............................... 78
4.3 Results .............................................. 79
  4.3.1 Impact of cosolutes on aggregation thermal stability, aggregate morphology and electrostatic interactions ..................... 79
  4.3.2 Impact of cosolutes on the mAb aggregation kinetics .......... 81
  4.3.2.1 mAb-1 - Impact of NaCl .......................... 81
4.3.2.2 mAb-1 - Impact of sorbitol 84
4.3.2.3 mAb-2 - Impact of NaCl 86
4.3.2.4 mAb-2 - Impact of sorbitol 89
4.4 Discussion 90
4.5 Conclusion 94

5 Stabilization of monoclonal antibodies by polyols 95
5.1 Introduction 95
5.2 Materials and methods 97
5.3 Results and discussion 97
5.3.1 Models proposed 98
5.3.2 Stabilization by sorbitol 103
5.3.3 Impact of polyol size on mAb stabilization 105
5.4 Conclusion 109

6 Impact of aggregation on the viscosity of mAb solutions 111
6.1 Introduction 111
6.2 Materials and methods 113
6.2.1 Materials 113
6.2.2 Hydrodynamic radius measurement 113
6.2.2.1 Batch mode 114
6.2.2.2 Flowing mode 115
6.2.3 Fractal dimension measurement 115
6.2.4 Viscosity measurement with tracer nanoparticles 116
6.2.5 Rheological measurements 119
6.3 Results 119
6.3.1 Viscosity of monomeric protein solutions 119
6.3.2 Viscosity of aggregated protein solutions 120
6.3.3 Interaction parameter 121
6.3.4 Occupied volume fraction 124
6.3.5 Aggregate fractal morphology 126
6.3.6 Master curve 128
6.4 Discussion 128
6.4.1 Protein-protein interactions 129
6.4.2 Master curve of the macroscopic viscosity 130
6.5 Conclusion 132

7 Viscosity scaling in concentrated dispersions and its impact on colloidal aggregation 133
7.1 Introduction 133
7.2 Methods 135
7.2.1 Brownian dynamics simulations 135
7.2.2 Effective viscosity computation 138
7.3 Results and discussion 139
7.3.1 Viscosity scaling in concentrated dispersion 139
8 Complex phenomena arising at high mAb concentration
8.1 Introduction .................................................. 157
8.2 Materials and methods .................................. 159
  8.2.1 Sample preparation .................................. 159
  8.2.2 Isothermal aggregation kinetics .................. 160
  8.2.3 Size exclusion chromatography with inline light scattering . 160
  8.2.4 Aggregate fractal dimension ......................... 160
  8.2.5 Viscosity measurements ................................ 161
  8.2.6 Protein-protein interactions ....................... 161
    8.2.6.1 Dynamic light scattering ..................... 161
    8.2.6.2 Static light scattering ........................ 162
  8.2.7 Aggregation rate constant .......................... 164
8.3 Results ....................................................... 167
  8.3.1 Aggregation under dilute conditions ............... 168
  8.3.2 Aggregation under concentrated conditions ......... 172
8.4 Discussion .................................................... 181
  8.4.1 Relation between the second virial coefficient and aggregation kinetics under dilute conditions .......... 181
  8.4.2 Aggregation kernel under concentrated conditions ...... 184
    8.4.2.1 Protein solution non-idealities ............... 184
    8.4.2.2 Increase in solution viscosity during aggregation . 185
    8.4.2.3 Increase in aggregate compactness ............. 186
  8.4.3 Stability ranking from low to high concentration ...... 186
8.5 Conclusion .................................................... 188

9 Concluding remarks ........................................ 191
  9.1 Approaches to assess protein stability ............... 191
  9.2 Input from colloid science ................................ 193
Contents

A Details about SEC-MALS experiments 197

B PBE discretization with Gaussian functions (fibril breakage) 201

C Details about viscosity measurements 207

D Boundaries for the approximate volume fraction 211

E PBE discretization with Gaussian functions (effective viscosity) 217

F Collective and self-diffusion coefficients 223

G Supplementary light scattering data 231

H Relation between Fuchs ratio and second virial coefficient 233

Bibliography 237

Abbreviations 263

Physical Constants 265

Symbols 267

Curriculum Vitae 273

Publications 275

Conference presentations 277
Chapter 1

Introduction

1.1 Aims and scope of the thesis

Protein-based drugs have emerged as a major class of pharmaceuticals since the 1980s and are nowadays widely used in the treatment of numerous human diseases including diabetes, rheumatoid arthritis and cancer [1]. During manufacturing, therapeutic proteins are exposed to various types of stress (such as acidic, thermal, mechanical, interaction with surfaces), which often promote aggregate formation [2, 3]. In particular, it is troublesome to guarantee long-term protein stability in the final product formulation, as protein-based drugs generally require administration at high protein concentration, in the order of 80 g/L or more [4]. Since the presence of protein aggregates may reduce drug efficacy or trigger unwanted immunogenicity [5], the protein aggregate content must be strictly controlled to ensure satisfactory product quality [3].

With a view to improving protein stability, considerable effort is put forth to unravel the fundamental mechanisms underlying the aggregation process. The protein stability depends on a variety of factors, both intrinsic (i.e. related to the protein properties) and extrinsic (i.e. related to the environmental conditions). As a consequence, strategies aiming at improving protein stability focus either (i) on designing aggregation resistant proteins (for instance, by engineering the protein...
primary sequence [6, 7] or by performing controlled chemical modifications such as PEGylation [8]); or (ii) on optimizing the operative conditions, and in particular the buffer composition [9–12]. In both approaches, a solid understanding of the aggregation process is needed for the development of rational strategies seeking to enhance protein stability [12, 13].

Gaining such understanding is challenging because protein aggregation is a complex multistep process, which involves characteristic times and length scales spanning many orders of magnitude, from conformational changes occurring in milliseconds to the formation of visible particulates over months. At present, the available set of experimental and computational techniques can hardly cover the whole range of time and length scales relevant for protein aggregation, making predictions of the aggregation rates barely feasible [12, 14–16].

Concepts from polymer [17, 18] and colloid science are emerging as powerful tools in the analysis of protein aggregation. Colloidal dispersions are two-component systems consisting of a discontinuous phase finely dispersed in a continuous medium, in which the elements of the dispersed phase are (i) sufficiently small to be subject to Brownian motion, and (ii) sufficiently large so that the solvent can be regarded as a continuum [19, 20]. Such boundaries correspond to sizes typically comprised between 1 nm and 1 \(\mu\)m. Paints, clay slurries, tobacco smokes, and aerosols are familiar examples of colloidal systems. Although proteins in aqueous media are molecular mixtures, and thus represent true solutions, they can be treated in many respects as colloidal dispersions.

Just like lyophobic colloids, most of the protein solutions are thermodynamically unstable and evolve spontaneously towards an aggregated state, which has a lower energy compared to the dispersed state, as represented in Figure 1.1(a). Depending on the protein and the operating conditions under investigation, aggregation can lead to various thermodynamically stable states, such as gels or precipitates, as illustrated in the qualitative phase diagram of Figure 1.1(b).

However, if the energy barrier that the system must overcome in order to reach the thermodynamic equilibrium is sufficiently high (Figure 1.1(a)), the aggregation
Chapter 1. Introduction

Figure 1.1: (a) Qualitative scheme of an energy profile, illustrating the thermodynamic and the kinetic aspects of protein aggregation. (b) Qualitative scheme of a phase diagram indicating various possible thermodynamically stable states of protein solutions. The operating parameters on the x and y axes of (b) are typically protein concentration, ionic strength or pH value.

process can be significantly delayed and the colloidal system is defined as *kinetically stable* [19–21].

In this thesis, we borrow fundamental principles of colloid science to gain knowledge on the thermodynamics and kinetics of therapeutic protein aggregation. One of the main objectives of this work is to understand how the protein aggregation behavior is impacted when moving from dilute to concentrated protein solutions. Particular attention will be drawn to monoclonal antibodies (mAbs), which represent a large part of the biopharmaceuticals market [22, 23].

In this first chapter, we review the state-of-the-art literature on therapeutic aggregation. We highlight how key concepts from colloid science can be applied to address the challenging topic of protein stability across different length scales, as illustrated in Figure 1.2. After considering briefly atomistic approaches which provide a high resolution of the protein structure, we discuss the use of coarse-grained molecular potentials to describe protein interactions. We then elaborate on the formation of mesoscopic aggregates, and on their impact on the macroscopic properties of protein solutions.
1.2 Literature review

1.2.1 Atomistic simulations

The protein aggregation propensity is strongly connected to the protein amino-acid sequence. A clear evidence of this relationship arises from the fact that single point mutations impact dramatically protein aggregation rates [24]. In silico approaches represent an invaluable tool to investigate biomolecular systems at the atomistic scale, offering the possibility to reveal details of the protein structure otherwise difficult to access experimentally [25]. In this context, a large effort has been given to the development of simulation techniques for the prediction of aggregation-prone regions [26], as for instance the spatial aggregation propensity (SAP) tool [27]. This technology was successfully applied to engineer antibodies with enhanced stability, and therefore represents a promising strategy to assess the developability of protein drug candidates [7, 27]. In addition, molecular dynamics simulations at the atomistic level allow designing proteins with enhanced solubility [28], and can provide insights into protein-solvent interactions, as well as osmolyte-induced protein refolding or denaturation [29–31].
Chapter 1. Introduction

However, the high resolution of the protein spatial conformation in atomistic simulations is achieved at the expense of high computational efforts. This limitation makes the approach not suitable to describe aggregation kinetics of complex therapeutic proteins on a time scale relevant for accelerated studies and storage conditions. Consequently, coarse-grained models at the amino-acid or at the monomer level may offer a compromise between structure resolution and computational time [32–34]. A more comprehensive summary of the computational tools used to study therapeutic protein aggregation can be found in specific reviews on this topic [26, 35].

1.2.2 Molecular interaction potentials

1.2.2.1 Second virial coefficient

With a view to decreasing the level of complexity of protein structures, intermolecular forces can be described by coarse-grained molecular interaction potentials. One of the most common formalisms consists in quantifying the sum of all protein-protein interactions in a global effective parameter known as the second virial coefficient, commonly denoted as $\bar{B}$, which is formally expressed as the average over space of the total protein-protein potential in the limit of low protein concentration [36, 37]:

$$\bar{B} = -\frac{1}{16\pi^2} \int_{\Omega} \int_{0}^{\infty} \left( \exp \left( -\frac{V_T(r, \Omega)}{k_B T} \right) - 1 \right) r^2 \text{d}r \text{d}\Omega \quad (1.1)$$

where $k_B$ is the Boltzmann constant, $T$ is the temperature, and $V_T$ is the total protein-protein potential that accounts for different types of intermolecular interactions (such as electrostatics, Van der Waals, hydration or hydrophobic forces). This potential is a function of all the configuration variables, i.e. the intermolecular center to center distance $r$, and the angular variables that are collectively represented by $\Omega$ [36, 37].
The sign and the absolute value of $B$ reflect the direction and the magnitude of the two-body protein-protein interactions, respectively. While a negative value of $B$ is indicative of predominantly attractive protein-protein interactions, a positive $B$ value corresponds to net repulsive protein-protein interactions.

The contribution of the electrostatic repulsion to the protein colloidal stability can be quantified through the assessment of the surface charge density, either from theoretical computations based on the protein amino-acid sequence, or through experimental measurements of the effective protein charge, for example from the electrophoretic mobility [38–40].

The simplest approach to evaluate the contribution of electrostatics to the protein colloidal stability is to compute interaction potentials within the DLVO theory, named after Derjaguin, Landau, Verwey and Overbeek, which accounts for electrostatic repulsion and Van der Waals attraction [19, 41]. The sum of these two forces can result in an energy barrier repelling the approaching molecules, as illustrated in Figure 1.3, and further detailed in Chapter 2. This energy barrier is expected to increase with the molecule charge and to decrease with the ionic strength of the surrounding medium, since the presence of salt induces charge screening.

![Image of interaction potentials](a) Born, Electrostatics, Van der Waals, Total Interaction energy vs. Interparticle distance. (b) Energy barrier, Primary minimum, Secondary minimum Interaction energy vs. Interparticle distance.

**Figure 1.3:** Interaction potentials as a function of intermolecular distance as described within the framework of the DLVO theory, highlighting (a) the different forces and (b) the energy barrier that the colliding molecules must overcome to aggregate, i.e. in order to fall in the primary energy minimum.
On top of electrostatic and Van der Waals forces, short-range Born repulsion forces prevent the approaching molecules from overlapping. The sum of these three contributions leads to a very deep attractive well in the total interaction potential, often referred to as primary minimum. Molecules falling in this deep attractive well cannot freely diffuse away from each other, thus triggering the formation of an aggregate. In the case where the energy introduced in the system (e.g. by heating or mechanical agitation) is larger than the depth of the well, aggregate breakage can be observed. This situation often occurs for amyloid fibrils, as will be studied in Chapter 3.

The DLVO theory has long set the basis for the interpretation of the stability of colloidal systems [19, 20] and has proven to be applicable to some extent also to protein solutions. The impact of pH and ionic strength on the aggregation propensity of some small globular proteins exhibiting random charge distribution could indeed be qualitatively rationalized within the DLVO framework [41, 42].

In addition to the aforementioned forces, a variety of non-DLVO forces such as hydration forces, hydrophobic forces, and hydrogen bonding often play a prominent role in protein stability [43–45]. Moreover, the protein surface patchiness tends to increase the colloidal stability of proteins in solution as compared to the case of traditional colloids [46]. The determination of the second virial coefficient allows to quantify in a coarse-grained manner the two-body protein-protein interactions, lumping together all types of interactions and protein surface heterogeneities.

Static light scattering (SLS) is one of the most conventional techniques used to measure the second virial coefficient of protein solutions [38, 39, 45, 47, 48]. This method is relatively material- and labor-intensive and alternative methods have been recently developed to simplify measurements. For example, a dual-detector cell which allows the simultaneous measurement of the protein concentration and scattered intensity after the protein elutes from a size exclusion chromatography (SEC) column can be employed [49]. Another approach relies on the estimation of $\overline{B}$ through the orthogonal determination of the diffusion and sedimentation
coefficients from dynamic light scattering (DLS) and sedimentation velocity measurements, respectively [50]. Finally, self-interaction chromatography represents a time efficient approach for the assessment of protein-protein interactions, and thus has potential applications in the screening of the buffer composition during drug formulation [51]. The experimental measurement of the second virial coefficient remains a rather delicate task and the absolute values of $B$ measured with different techniques must be taken with some cautions [47]. Nonetheless, comparative studies provide robust information about the change of intermolecular interactions in different solution compositions [38, 39, 45, 48, 52].

It is worth underlining that the second-virial coefficient is by definition a dilute solution property. Nevertheless, a similar approach can be employed at high protein concentration by using a Kirkwood-Buff analysis of light scattering experiments [53–55], which will be introduced in Chapter 8.

1.2.2.2 Colloidal stability and conformational stability

Although valuable information on protein-protein interactions can be extracted from the measurement of the second virial coefficient, the correlation between $B$ values and the aggregation rate is not straightforward, and is still a matter of debate in the literature [39, 45, 47, 50, 56, 57]. This question will be discussed in Chapter 8.

Here, we only highlight that the protein stability does not result exclusively from the protein colloidal stability, i.e. the energy barrier that two protein molecules need to overcome upon collision in order to aggregate, but also from the protein conformational stability, i.e. the energy barrier preventing protein molecules from unfolding [10, 13].

While the protein colloidal stability depends mainly on intermolecular protein-protein interactions, the protein conformational stability is dictated by the delicate balance between solvation and intramolecular protein interactions. The protein colloidal stability and the protein conformational stability are highly sensitive
to the solution composition and are strongly interconnected [43]. For instance, denaturing conditions promote the exposure of buried hydrophobic patches upon protein unfolding, which then leads to additional strong attractive hydrophobic intermolecular interactions favoring aggregation.

Therefore, it is paramount to characterize both the colloidal stability and the conformational stability when assessing the protein aggregation propensity, bearing in mind the prominent role of protein misfolding in the formation of aggregation-prone species. The protein conformational stability is commonly evaluated by applying differential scanning calorimetry (DSC) [58], circular dichroism (CD) spectroscopy [59], intrinsic tryptophan [60] and extrinsic dyes [61] fluorescence spectroscopy.

1.2.3 Formation of mesoscopic aggregates

1.2.3.1 Protein aggregation mechanisms

Proteins can aggregate following various elementary reaction paths. Nevertheless, most of the reaction schemes can be included in a general model, defined as the Lumry-Eyring Nucleated Polymerization (LENP) model [62–65], which reflects the contributions of both the protein conformational stability and the protein colloidal stability to the aggregation process, as shown in Figure 1.4.

The model involves first the formation of a non-native aggregation-prone conformational state of the protein, which results from changes in the protein structure (Stage I in Figure 1.4). This intermediate then promotes nucleation of oligomers (Stages II and III), which subsequently grow to larger aggregates, either by monomer addition (Stage IV) or by aggregate-aggregate assembly (Stage V). For the sake of generality, aggregate breakage (Stage VI), which is known to play a prominent role in the kinetics of amyloid fibril formation [66], has been included in the reaction scheme, and will be further investigated in Chapter 3. It is worth
Specific to proteins (compared to usual colloids)

I. Monomeric conformational changes

II. Reversible oligomerization

III. Irreversible rearrangement

IV. Aggregate growth by monomer addition

V. Aggregate growth by aggregate-aggregate assembly

VI. Aggregate breakage

Fibrillar / amorphous aggregates Precipitates / gels ...

Figure 1.4: Illustrative scheme of the events involved in the mechanism of protein aggregation, reflecting both the conformational stability (related to protein unfolding) and the colloidal stability (related to aggregation events).

mentioning that in the case where aggregation is mediated by interactions at interfaces [67] or by protein hydrolysis [68], additional steps may become relevant in the mechanism.

This multistep aggregation scheme is sufficiently versatile to describe a large variety of aggregation mechanisms occurring in bulk solutions, including the formation of both amorphous aggregates [62, 63] and amyloid fibrils [66]. Nevertheless, it must be emphasized that the relative importance of the aforementioned elementary steps depends strongly on the system under investigation, i.e. both on the protein considered (see Chapter 2) and on the environmental conditions (see Chapters 4 and 8), leading to different aggregation pathways [57, 69].

The reaction scheme shown in Figure 1.4 can then be translated into mass balance equations, where each elementary reaction is characterized by a proper reaction rate constant. This type of mean-field kinetic analysis is emerging as a powerful tool in the mechanistic description of the aggregation process, as further discussed in the next paragraphs.
1.2.3.2 Experimental characterization

The rate of the various elementary events involved in the global aggregation process can be determined from the comparison between simulations based on mass action laws and experimental data [62–64].

No single experimental technique can analyze the complete size range of protein aggregates, from small oligomers to large particulates. Thus, the proper experimental characterization of the kinetics of protein aggregates formation requires the combination of multiple orthogonal techniques [70, 71]. SEC analysis is the most common analytical method to quantify the residual monomer content, as well as the oligomer distribution. However, off-line SEC analysis suffers from several limitations, including possible protein adsorption on the column matrix, aggregate dissociation during the analysis run, and removal of large aggregates in the frit [70]. Therefore, orthogonal methods, such as sedimentation velocity analytical ultracentrifugation or flow field-flow fractionation technique may be implemented to confirm the accuracy of SEC results [71].

Monitoring the decrease in the monomer content at several protein concentrations provides pertinent information on the elementary step(s) limiting the rate of monomer consumption [62, 64, 65]. On this purpose, it is convenient to describe the kinetics of monomer depletion in terms of an apparent reaction order $n_{app}$, defined as:

$$\frac{dM}{dt} = -k_{app}M^{n_{app}} \tag{1.2}$$

where $M$ is the monomer concentration followed over a few half-lives, and $k_{app}$ is the apparent reaction rate for monomer loss.

Values of $n_{app}$ lying between one and two are typically observed. In particular, $n_{app} = 1$ and $n_{app} = 2$ indicate, respectively, that unimolecular conformational changes and bimolecular aggregation events are limiting the rate of monomer depletion [65]. Monomer depletion kinetics will be analyzed in terms of apparent reaction order in Chapter 2.
A valuable strategy involves the combination of SEC with an inline multi-angle light scattering (MALS) detector, in order to monitor the increase in average aggregate molecular weight with time [71]. The shape of the aggregate growth profile as a function of monomer conversion indeed allows to gain knowledge on the mechanism of aggregate growth: a linear increase of the aggregate molecular weight with conversion is characteristic of a monomer addition mechanism, whereas an upwards profile indicates that aggregate-aggregate assembly is appreciable [63, 72].

In addition, the average hydrodynamic radius of the non-fractionated population can be measured by batch-mode DLS, which offers the possibility to analyze aggregate sizes covering the whole colloidal range, from a few nanometers to roughly one micrometer, provided that the aggregate concentration is high enough. Larger particles are instead typically detected by light obscuration, microscopy or coulter counter. However, the accurate quantification of subvisible particles still represents an open challenge in protein drug formulation [69, 71].

1.2.3.3 Aggregate fractal geometry

Besides the measurements of the aggregate size and of the aggregate content, it is relevant to characterize the aggregate morphology, which can be investigated with microscopy techniques, such as atomic force microscopy (AFM) or transmission electron microscopy, as well as with other approaches such as micro flow imaging and static light scattering (SLS). Depending on the solution conditions and on the environmental stress, a given protein can form aggregates of various morphologies. For instance, it was recently reported that freeze-thaw cycles lead to mAb aggregates which are relatively fibrillar in nature, while stirring rather induces the formation of large spherical particles [73].

The concept of fractal scaling relies on the assumption of a scale invariant geometry, also called self-similarity. Namely, it unifies various irregular aggregate morphologies that would be otherwise difficult to rationalize, by relating the
aggregate mass with aggregate size according to the following power law:

\[ i = k_f \times \left( \frac{R_i}{R_p} \right)^{d_f} \]  

(1.3)

where \( d_f \) is the fractal dimension, \( k_f \) is the scaling prefactor usually in the order of unity [74], \( R_p \) is the radius of a primary particle (i.e. the monomer), and \( R_i \) is the radius of an aggregate containing \( i \) primary particles. The gyration radius is commonly used in Equation 1.3. Nevertheless, the concept of fractal scaling also holds in the case of the hydrodynamic or collision radii.

It is worth mentioning that the notion of fractal scaling theoretically refers to scale invariance over infinite dimensions, while in the context of aggregation it is used in an effective manner over the cluster discrete sizes, ranging from small oligomers to the largest aggregate formed in the system.

Aggregate fractal geometry has proven to be a realistic assumption for a wide variety of colloidal systems by means of computer simulations as well as experimental characterization. Indeed, both Monte Carlo algorithms [75, 76], which rely on the generation of random numbers to simulate Brownian motion, and Brownian dynamics algorithms [77], which are based on the numerical integration of the Langevin equation of motion, predict the formation of fractal aggregates. Moreover, SLS experiments revealed that many proteins, from small globular proteins to more complex multi-domain proteins such as antibodies, form aggregates exhibiting fractal morphology [42, 78–82].

Such simulation and experimental techniques allow the determination of the fractal dimension, which quantifies the aggregate compactness: high \( d_f \) values characterize compact aggregates, whereas low \( d_f \) values indicate open structures. The extreme cases of \( d_f = 1 \) and \( d_f = 3 \) correspond to fibrillar and spherical aggregates, respectively. Values of the fractal dimension of mAb aggregates will be reported in Chapter 2. The impact of cosolutes, protein concentration, and solution pH on \( d_f \) will be analyzed in Chapters 4, 6 and 8, respectively.
This simple mathematical description of aggregate geometry, i.e. correlating the aggregate mass with its size through the fractal dimension, reveals particularly useful in the definition of the aggregation kernels implemented in population balance equations modeling [78], as discussed in the next section.

### 1.2.3.4 Population balance equations modeling

The detailed experimental data can be coupled to simulations based on mass action laws in order to extract the values of the kinetic rate constants necessary to describe the kinetics of protein aggregation [45, 62, 72]. This approach was for example successfully applied to study the oligomerization of an antibody under low pH conditions, which may occur during the purification steps with protein A chromatography [83].

Depending on the protein, either small oligomers or large precipitates can be formed in acidic environments encountered during downstream processing [84]. When the aggregate size distribution is broad, the collection of mass action laws describing the formation of each aggregate size is commonly referred to as a set of population balance equations (PBE). PBE simulations can be applied to describe the time evolution of the aggregate mass distribution of a wide diversity of colloidal systems [85, 86], including proteins [78, 79]. The general PBE for an aggregating colloidal system is given below:

\[
\frac{dM_i}{dt} = \sum_{j=1}^{i-1} k_{j,i-j} M_j M_{i-j} - M_i \sum_{j=1}^{\infty} k_{i,j} M_j
\]  

(1.4)

where \( M_i \) is the number concentration of aggregates containing \( i \) monomeric units.

This general PBE can be modified to include additional elementary events that contribute to the mechanism of aggregate formation, such as protein unfolding and reversible oligomerization (see Chapters 2, 4, 5, and 8), as well as aggregate breakage [68, 87, 88] (see Chapter 3).
Solving the PBE shown in Equation 1.4 requires to define the aggregation kernel, i.e. the collection of aggregation rate constants $k_{i,j}$ between two aggregates of any size $i$ and $j$. Several aggregation kernels have been proposed in the literature in the context of polymer colloids, and we mention here only the two most common ones: (i) the diffusion-limited cluster aggregation (DLCA) kernel, which corresponds to the case where the particles are not repelled by the presence of an energy barrier; (ii) the reaction-limited cluster aggregation (RLCA) kernel, which corresponds to the case where the interparticle potential exhibits a maximum, resulting in a reduced sticking efficiency upon collision.

Under DLCA conditions, the aggregation rate constant results from two contributions: the diffusivity of the aggregates, and their collision cross section, as derived by von Smoluchowski [89–92]:

$$k_{i,j} = 4\pi(D_{0,i} + D_{0,j})(R_{c,i} + R_{c,j})$$

(1.5)

where $D_{0,i}$ and $R_{c,i}$ represent the diffusion coefficient and the collision radius of the clusters containing $i$ monomers, respectively.

Under dilute conditions, the protein diffusion coefficient can be computed from the Stokes-Einstein equation according to:

$$D_{0,i} = \frac{k_B T}{6\pi \eta_0 R_{h,i}}$$

(1.6)

where $R_{h,i}$ is the hydrodynamic radius of the cluster of mass $i$.

By using the fractal scaling, the collision and hydrodynamic radii can be estimated from: $R_{c,i} \approx R_{h,i} \approx i^{1/d_f}/k_f$.

Under RLCA conditions, two additional terms must be added and the aggregation rate constant becomes:

$$k_{i,j} = \frac{k_s}{4} \left(i^{1/d_f} + j^{1/d_f}\right)\left(i^{-1/d_f} + j^{-1/d_f}\right)\left(ij\right)^{\lambda}$$

(1.7)
where $k_s$ is the Smoluchowski rate constant, defined as the aggregation rate constant between two primary particles under diffusion-limited conditions:

$$k_s = \frac{8k_B T}{3\eta_0} \quad (1.8)$$

The term $(ij)^\lambda$ in Equation 1.7, with $\lambda \approx 1 - 1/d_f$, accounts for the number of primary particles located on the external surface of the cluster.

Moreover, the factor $W$ is the so-called Fuchs stability ratio which quantifies the energy barrier that clusters must overcome in order to aggregate. It describes the reduced sticking efficiency with respect to diffusion-limited conditions, and can be computed from the expression of the total interaction potential between two primary particles, $V_T$, according to:

$$W = 2R_p \int_{2R_p}^{\infty} \exp \left( \frac{V_T(r)}{k_B T} \right) \frac{dr}{r^2} \quad (1.9)$$

The value of the Fuchs stability ratio tends to be extremely sensitive to small uncertainties in the protein surface charge density, making almost unfeasible the computations of $W$ from the theoretical intermolecular interaction potential. Therefore, the Fuchs parameter is more accurately estimated by fitting model simulations to experimental data.

The analysis of $W$ values in various environmental conditions provides a quantification of the impact of operative parameters (such as temperature [78]) on effective protein-protein interactions. This approach will be applied to study the impact of cosolutes and pH on the aggregation kinetics of monoclonal antibodies in Chapters 4 and 8, respectively.

The expression shown in Equation 1.7 is very valuable as it provides a theoretical relationship between aggregate size and reactivity. However, it must be emphasized that this kernel was derived for uniformly charged spheres, while protein reactivity strongly depends on protein conformation and on the accessibility of
aggregation-prone patches. Modifications of the classical RLCA kernel to account for peculiar features of protein systems will be discussed in Chapter 2.

Moreover, it must be pointed out that Equation 1.7 is valid only under sufficiently dilute conditions. Corrections of this expression to include complex effects arising at high protein concentrations will be introduced in Chapters 7 and 8.

1.2.4 Macroscopic properties

The aggregation phenomena described in the previous paragraphs may eventually lead to macroscopic changes in the protein solution properties, both in the bulk and at interfaces [82, 93, 94].

1.2.4.1 Viscosity increase

In particular, the irreversible aggregation and the reversible self-association of protein molecules are often accompanied by an increase in the solution viscosity [69, 95]. This effect, which complicates the handling and the delivery of the drug product, is even more pronounced at the high protein concentrations that are typically encountered in therapeutic protein formulations [4, 96, 97].

Following the approach used in colloid science [98, 99], the rise in viscosity of protein solutions during aggregation can be correlated to the increase in the volume fraction occupied by the aggregates, which depends on aggregate size, concentration and morphology, and can be evaluated from:

\[
\phi = \frac{4}{3} \pi \frac{M_0}{k_f} R_p^{d_f} \langle R \rangle^{3-d_f}
\]  

(1.10)

where \( \langle R \rangle \) is the average radius of the aggregate population and \( M_0 \) is the initial protein concentration. Note that the computed value of the occupied volume fraction depends on the type of radius considered (e.g.: collision, gyration, hydrodynamic) [100].
The notion of occupied volume fraction will be used in Chapter 6 to rationalize the viscosity increase during the heat-induced aggregation of a monoclonal antibody.

It is worth highlighting that the increase in solution viscosity will reduce the thermal diffusion of the clusters, and in turn slow down the aggregation process. However, predicting the impact of the viscosity increase on the inhibition of aggregation is quite challenging since the microscopic viscosity which is experienced by the aggregates varies with aggregate size: while sufficiently large aggregates feel the macroscopic viscosity, small aggregates rather feel the solvent viscosity [101]. This concept of scale dependent viscosity will be deepened in Chapter 7.

If the initial protein concentration is sufficiently high, the occupied volume fraction might exceed unity. The growing clusters then interact with each other, which can eventually lead to the formation of a solid-like gel. In the arrested state, the clusters form a percolating structure, i.e. an interconnected network spanning the entire available space, as represented in Figure 1.5. This type of sol-gel transition has been observed for several proteins of pharmaceutical interest, such as glucagon [102], collagen [103], monoclonal antibodies [104] as well as the amphiphilic peptide RADA 16-I [105].

![Figure 1.5](image_url)

**Figure 1.5:** (a) Schematic representation of the gelation process within the fractal gel model, which can be conceptually divided into two steps, namely the formation of fractal clusters and the interconnection of the clusters in a percolating structure. (b) The scheme of gel formation presented in (a) for amorphous aggregates is extended to fibrillar aggregates. (c) Increase of the occupied volume fraction with time due to the formation of clusters, leading to an increase in solution viscosity, and potentially to the formation of a gel.
1.2.4.2 Gel formation

From a colloidal perspective, it is interesting to determine the morphology of the aggregates constituting the gel network. As a matter of fact, the fractal gel model [105–108], (Figure 1.5), which is commonly applied to describe the gelation of colloidal systems, has been proven successful in correlating macroscopic rheological properties with the gel microscopic structure for both amorphous [109] and fibrillar [105] aggregates.

Briefly, assuming the colloidal gel as a series of interconnected flocks, the storage modulus is related to the protein concentration in a power law manner with a scaling exponent containing information about the fractal dimension of the aggregates [80, 82, 110]. Depending on the type and flexibility of the network junction, the storage modulus is correlated either to the entropic or to the enthalpic elasticity [82]. In the case of protein gels, the entropic contribution is limited and the storage modulus $G'$ is then expressed by the following formula [82]:

$$G' \sim M_0^{\frac{3+d_f}{3+d_b}}$$

(1.11)

where $M_0$ is the protein concentration, and $d_b$ and $d_f$ are the fractal dimensions of the backbone and of the aggregate, respectively.

1.2.4.3 Phase diagrams

The formation of a gel phase is not the only macroscopic change observed in protein solutions. Attractive intermolecular interactions, which lead to reversible self-association or irreversible aggregation events, can indeed result in the formation of various phases, including liquid-liquid separation, precipitates, crystals and nematic phase [82, 111]. In analogy with colloidal systems, the phase behavior of proteins is often represented in phase diagrams showing the boundaries between different phases at thermodynamic equilibrium (Figure 1.1(b)) as a function of
relevant parameters such as pH, temperature, protein and cosolute concentrations [104, 105, 112–115].

1.3 Outline of the thesis

The thesis is divided into two main parts. The first part (Chapters 2 to 5) is dedicated to the study of therapeutic protein aggregation under dilute conditions, which, although not strictly representative of manufacturing conditions, is a necessary starting point to gain fundamental knowledge on the kinetics and thermodynamics of protein aggregation. In the second part (Chapters 6 to 8), the protein concentration is progressively increased, with the aim of identifying and quantifying the complex phenomena that impact protein aggregation under concentrated conditions.

The approach used throughout the thesis relies on the comparison between experimental data (mainly acquired with size exclusion chromatography and light scattering techniques) and simulations from kinetic models (mainly based on population balance equations and key concepts from colloid science). Three therapeutic proteins are used for the experimental studies: two model monoclonal antibodies of industrial origin, denoted as mAb-1 and mAb-2, and insulin.

In Chapter 2, as a first step, we investigate the aggregation mechanisms of mAb-1 and mAb-2 under thermal stress and dilute conditions. The aim of this study is to identify the key elementary events that contribute to the global aggregation process, and to assess whether the DLVO theory can provide reasonable estimates of the aggregation rate.

In Chapter 3, we examine the aggregation mechanism of another therapeutic protein, namely insulin. Interestingly, as compared to mAb-1 and mAb-2, insulin forms breakable fibrillar aggregates under the investigated conditions. In particular, we study how fibril fragmentation is influenced by temperature and mechanical agitation.
In Chapter 4, the aggregation model built in Chapter 2 is used to investigate the impact of cosolutes on the aggregation kinetics of mAb-1 and mAb-2. NaCl and sorbitol, which are commonly used in the formulation of protein-based drugs, are selected as representative examples of excipients. The main goal of this chapter is to discriminate the impact of cosolutes on the protein conformational and colloidal stabilities.

In Chapter 5, we deepen the stabilization mechanism of mAb-1 by polyols, focusing on the role of the cosolute size in the mAb stabilization effect.

Chapter 6 marks the transition from the dilute to the concentrated regime, and is devoted to only one peculiar effect arising at high protein concentration: the viscosity increase during the aggregation process. This chapter aims at developing an experimental technique to monitor the rise in solution viscosity, as well as rationalizing the impact of aggregate formation on the solution viscosity by using the key concept of occupied volume fraction.

Chapter 7 complements Chapter 6 by quantifying the impact of an increase in solution viscosity on the kinetics of aggregation from a theoretical point of view. First, Brownian dynamics simulations are performed to quantify the effective viscosity experienced by a tracer particle immersed in a crowded environment. Then, the impact of viscosity effects on the kinetics of aggregation of colloidal particles is analyzed by means of population balance equations.

In Chapter 8, we go beyond viscosity effects and propose to include additional hydrodynamic and thermodynamic non-idealities due to protein-protein interactions in the aggregation rate. Model simulations are compared to experimental data acquired at several solution conditions, with varying pH and salt content.

Chapter 2

Multistep aggregation mechanisms of monoclonal antibodies

2.1 Introduction

In this chapter, we investigate the aggregation mechanisms of two model monoclonal antibodies at low protein concentration under thermal stress. Forced degradation studies at elevated temperatures indeed play a central role in the screening of potential drug candidates as they allow to assess rapidly product stability [2]. The rationalization of the protein aggregation mechanism is a key requisite to evaluate the product shelf-life from these accelerated studies [15].

As described in Chapter 1, most of the aggregation behaviors of therapeutic proteins can be described in the frame of a generalized Lumry-Eyring model, originally proposed by Roberts and co-workers [62–65]. This multistep aggregation scheme involves first the formation of an aggregation-prone intermediate in a non-native conformational state. This reactive intermediate promotes the nucleation of oligomers, which is then followed by growth to larger aggregates.
Despite the main steps involved in the aggregation process have been qualitatively identified [62–65], the quantification of the relative importance of the individual steps in a specific system remains challenging. One of the main problems originates from the fact that only few quantities, such as the monomer conversion and the aggregate average molecular weight, are accessible experimentally. Kinetic analysis has been proven to be a powerful tool to complement experimental characterization and to quantify from macroscopic measurements the contribution of single microscopic events in the global aggregation process [116].

In the first part of this chapter, we apply this kinetic approach to gain insight into the aggregation mechanisms of two model monoclonal antibodies. The aggregate growth steps are described by population balance equations (PBE), which are conservation laws applicable to a variety of colloidal systems undergoing aggregation, including proteins [78, 79, 85, 86]. In this study, the aggregates remain soluble for a sufficient time to allow the detection of both small oligomers and larger aggregates. Accordingly, we modify the classical Smoluchowski’s PBE in order to include nucleation steps, i.e. protein unfolding and oligomer formation, in addition to the usual aggregate growth events. From the comparison between model simulations and experimental data, we estimate a coarse-grained intermolecular expressed in terms of the Fuchs stability ratio, commonly denoted as $W$.

In the second part of this chapter, we compare the fitted $W$ values with predictions from the DLVO theory, which accounts for electrostatic repulsion and Van der Waals attraction. As discussed in Chapter 1, the DLVO theory which has long set the basis for the interpretation of colloidal stability, can be applied to some extent to protein systems. Typically, the qualitative effect of salt and pH on the aggregation of model globular proteins exhibiting a random charge distribution can be rationalized within the DLVO framework [41]. However, the situation is more challenging for large and complex biomacromolecules such as antibodies. Indeed, the protein surface is generally highly heterogeneous and aggregation-prone domains can be confined to a few hydrophobic patches. Moreover, the protein stability is often affected by non DLVO forces and peculiar effects such as ion binding [43–45].
2.2 Materials and methods

2.2.1 Materials

The two monoclonal antibodies used for this study are a glycosylated IgG1 and a non-glycosylated IgG2, which will be denoted in the following as mAb-1 and mAb-2, respectively. The theoretical isoelectric point (pI) of mAb-1 is between 8 and 9.2, while the theoretical pI of mAb-2 lies between 7.35 and 8.15.

The antibody solutions were dialyzed at the protein concentration of 20 g/L against a 20 mM histidine buffer at pH 6.5 using Slide A Lyzers cassettes from Thermo Fisher Scientific, with a cut-off molecular weight of 7 kDa. The volume of the dialysis buffer was 500-fold larger than the volume of the sample to be dialyzed. The buffer was renewed a first time after 2 h, and a second time after 4 h of dialysis. The dialysis was performed at 4 °C under gentle stirring for at least 18 h. The protein concentration of the stock solution after dialysis was checked by UV absorption at 280 nm.

All the samples for this study were prepared by diluting the stock solution to the targeted concentration with a 20 mM histidine buffer at pH 6.5.

The chemicals were purchased from Sigma with the highest purity available, and the buffers were filtered through a 0.1 µm cut-off membrane filter (Millipore).

2.2.2 Zeta potential measurements

Zeta potential values (ζ) of proteins have been evaluated by a Zetasizer Nano (Malvern, Worcestershire, UK) measuring the electrophoretic mobility \( \mu_{el} \) via laser Doppler effect. The zeta potential is calculated from the electrophoretic mobility according to the Henry equation:

\[
\mu_{el} = \frac{2\epsilon_0\epsilon_r\zeta f_H(\kappa R_p)}{3\eta_0}
\]  

(2.1)
where $\epsilon_0$ is the vacuum permittivity, $\epsilon_r$ is the relative dielectric constant of the solvent, and $\eta_0$ is the viscosity of the solvent, respectively. The Henry function $f_H(\kappa R_p)$, where $R_p$ is the protein radius and $\kappa$ is the inverse Debye length, was set to 1 according to the Smoluchowski approximation.

Measurements were performed at 25 °C and at the protein concentration of 1 g/L. Five repetitions of two independent samples were recorded for each condition and average values have been considered.

The protein net charge were then estimated from the measure of the zeta potential according to the following equation [117]:

$$z = \frac{4\pi\epsilon_0\epsilon_r R_p e}{\eta_0} (1 + \kappa R_p) \zeta$$

where $e$ is the elementary charge.

### 2.2.3 Thermal stability by circular dichroism

CD spectra were acquired with a Jasco-815 CD spectrophotometer (Mary’s Court, Easton, MD). Temperature was increased by a step of 1 °C with a rate of 1 °C/min and the samples were equilibrated 90 s at each temperature step before acquiring the spectra. CD measurements were performed with samples at protein concentration of 0.25 g/L in a 10 mM phosphate buffer at pH 6.5. It is worth precising that CD measurements could not be carried out in a histidine buffer due to the strong absorption of the histidine molecules.

### 2.2.4 Thermal stability by dynamic light scattering

Aggregation thermal stability was determined by DLS using a Zetasizer Nano (Malvern, UK). The temperature was automatically increased from 40 °C to 84 °C, with a step of 10 °C between 40 °C and 60 °C and a step of 2 °C between 60 °C and 84 °C. The samples were equilibrated 2 min at each temperature step before
recording the scattered intensity during 3 min. Each data point reported in the plots is an average of three measured intensity values and the error bars correspond to the standard deviation of three measurements.

2.2.5 Isothermal aggregation kinetics

Isothermal aggregation kinetics were performed by incubating antibody samples at 70 °C at protein concentrations in the range from 1 to 5 g/L in hermetically sealed HPLC vials containing 250 µL inserts (Agilent Technologies, part number 5182-0716, 5181-1270 and 5182-0721 for vials, inserts and caps, respectively). The vials were placed in a block-heater containing oil (Rotilabo H 250, Roth, Karlsruhe) for predetermined times. To improve heat transfer, 1 mL of aggregation buffer was added in the space delimited by the vial and the insert. The temperature of the block-heater was set to 71 °C in order to reach 70 °C inside the vials, as verified with a thermocouple.

Aggregated samples were quenched in an ice-water bath for at least 3 min and analyzed immediately after by SEC with inline MALS, or by batch SLS. Moreover, it was verified that the aggregated samples were not significantly impacted by the change in temperature during off-line analysis, as shown in Figure A.3 (see Appendix A).

2.2.6 Size exclusion chromatography with inline light scattering

Monomer conversion and oligomer formation was monitored by SEC with a Superdex 200 10/300 GL, 10 mm 300 mm size-exclusion column (GE Healthcare, Uppsala, Sweden) assembled on an Agilent series HPLC unit (Santa Clara, CA, USA). The samples were eluted for 45 min at a constant flow rate of 0.5 mL/min using as mobile phase a 100 mM phosphate buffer containing 200 mM arginine at pH 7.0, which has been shown to improve sample recovery [118].
The eluting species were detected by UV absorbance at 280 nm and by MALS using a Dawn-Heleos II device (Wyatt, Santa Barbara, CA, USA). While the monomer peak was well resolved, the peaks corresponding to the dimer and trimer were partially overlapped with larger aggregates. The chromatograms were deconvoluted using OriginPro 8.5 (Academic) in order to determine the concentrations of monomer, dimer and trimer. Light scattering results were processed with the Astra software (Wyatt, Santa Barbara, CA, USA) to obtain the weight-average molecular weight of the aggregate population. In Appendix A, Figure A.1 shows a representative example of a SEC chromatogram, with the applied peak deconvolution as well as the molecular weight determination from inline MALS. Moreover, Figure A.2(a) shows an example of raw static light scattering results.

The data presented in the Results section correspond to the average and standard deviation (error bars) of at least two independent measurements.

2.2.7 Evaluation of the fractal dimension by static light scattering

Static light scattering measurements were performed using a goniometer BI-200SM (Brookhaven Instruments, Holtsville, NY, USA) covering angles $\theta$ from $10^\circ$ to $145^\circ$. A solid-state laser, Ventus LP532 (Laser Quantum, Manchester, UK) with a wavelength $\lambda_0 = 532$ nm was used as a light source. For clusters significantly larger than primary particles, the fractal dimension $d_f$ can be estimated from the power-law regime of the structure factor, $S(q)$ [119]:

$$S(q) \sim q^{-d_f} \quad \text{for} \quad \frac{1}{\langle R_g \rangle} \ll q \ll \frac{1}{R_p}$$

(2.3)

where $R_p$ is the radius of the primary particles inside the cluster, $\langle R_g \rangle$ is the average radius of gyration of the cluster distribution and $q$ is the scattering vector defined as:

$$q = \frac{4\pi n}{\lambda_0} \sin \left( \frac{\theta}{2} \right)$$

(2.4)
where $\theta$ is the scattering angle, $n$ is the refractive index of the solvent and $\lambda_0$ the wavelength of the laser beam.

For each sample, three measurements were recorded and average values were considered.

### 2.2.8 In situ dynamic light scattering

The time evolution of the average hydrodynamic radius was followed by DLS in situ at fixed angle $\theta = 173^\circ$ using a Zetasizer Nano (Malvern). Briefly, the fitting of the autocorrelation function with the method of cumulants provides the average particle diffusion coefficient $D_0$, which is connected to the average hydrodynamic radius $\langle R_h \rangle$ by the Stokes-Einstein equation:

$$D_0 = \frac{k_B T}{6 \pi \eta_0 \langle R_h \rangle} \quad (2.5)$$

where $k_B$ is the Boltzmann constant, $T$ the temperature and $\eta_0$ the solvent viscosity.

The sample was incubated in a low volume quartz batch cuvette (ZEN 2112, Malvern). In order to prevent evaporation, a custom made plastic cap was added in the cuvette to reduce the air volume on top of the sample.

The data presented in the graphs correspond to the average and standard deviation (error bars) of at least two independent measurements.

### 2.2.9 Kinetic model

The implementation of the PBE requires the definition of the aggregation rate constant between two colliding aggregates containing $i$ and $j$ monomeric units. In this chapter, we used the traditional reaction-limited cluster aggregation (RLCA) kernel, which has been shown to successfully describe the aggregation kinetics of a large variety of colloidal systems, including proteins [78, 79, 85]. The aggregation
kernel is presented in Equation 2.6, while the description of the various terms is
given in Chapter 1.

\[ k_{i,j} = k_s B_{i,j} P_{i,j} W_{i,j}^{-1} \quad \text{with:} \]
\[
\begin{align*}
  k_s &= (8/3) k_B T / \eta_0 \\
  B_{i,j} &= \frac{1}{4} \left( (i^{-1/d_f} + j^{-1/d_f}) \left( i^{1/d_f} + j^{1/d_f} \right) \right) \\
  P_{i,j} &= (ij)^\lambda
\end{align*}
\]

(2.6)

It is worth mentioning that several expressions have been proposed in the
literature for the factor \( P_{i,j} \). In the absence of precise information about the
correlation between aggregate size and reactivity, we selected the product kernel
\( P_{i,j} = (ij)^\lambda \), which has been proven to describe well experimental data in a broad
range of conditions [85]. Based on scaling arguments, Schmitt et al.[120] showed
that the number of primary particles located on the external surface of a fractal
aggregate scales as \( i^{1-1/d_f} \). This suggests that \( \lambda \) can be roughly estimated as
\( \lambda \approx 1 - 1/d_f \).

In the RLCA regime, the presence of a repulsive energy barrier between the
particles delays the aggregation process as compared to diffusion-limited condi-
tions. In this case, only a fraction of collisions is successful in forming larger
aggregates and the reduced sticking efficiency is generally expressed by the so-
called Fuchs stability ratio, which can be computed from the expression of the
total interaction potential between two primary particles, \( V_T \), as follows:

\[ W_{11} = 2 R_p \int_{2R_p}^{\infty} \exp \left( \frac{V_T(r)}{k_B T} \right) \frac{dr}{r^2} \]

(2.7)

This approach can be extended to the aggregation of two clusters of size \( i \)
and \( j \) provided that the aggregate-aggregate interactions can be approximated to
the interactions between the two colliding primary particles, thus implying that
\( W_{11} \) can capture the interactions between two clusters, irrespective of their size.
However, it is unlikely that this assumption holds true in the case of protein
aggregation, where the reactivity strongly depends on protein conformation and on the accessibility of aggregation-prone patches, and thus changes with aggregate size. Therefore, in the present study we introduced different values of the Fuchs ratio in order to characterize the stability of different sub-populations of species characterized by a similar reactivity. These sub-populations include monomer, oligomers, nucleus and larger aggregates.

Simulations were compared to experimental data by defining a normalized residual sum of squares function for each measurable quantity $y$:

$$RSS_n = \sum_{i}^{N_{exp}} \left( \frac{(y_{i}^{exp} - y_{i}^{sim})}{y_{i}^{exp}} \right)^2$$  \hspace{1cm} (2.8)

where $y_{i}^{exp}$ is the $i$-th experimental data point, $y_{i}^{sim}$ is the corresponding value simulated with the model, and $N_{exp}$ is the total number of experimental points considered. The quantities $y$ correspond to the monomer concentration, the dimer concentration, the trimer concentration, the aggregate average molecular weight and the overall average hydrodynamic radius. A total objective function was defined by adding up the contributions of all the measurable quantities.

The weight-average molecular weight of the simulated aggregate distribution was evaluated from the moments of the aggregate distribution and the average hydrodynamic radius of the overall population was computed as described by Lattuada et al. [121, 122].

### 2.2.10 Computation of the Fuchs stability ratio within the DLVO theory

As mentioned previously, the Fuchs stability ratio between two primary colloidal particles, $W_{11}$, can theoretically be computed from the expression of the interaction potential using Equation 2.7. The DLVO theory (named after Derjaguin, Landau, Verwey and Overbeek [123, 124]) has long been a cornerstone for the modeling
of colloidal interactions. It accounts for the competing effects between Van der Waals attraction and electrostatic repulsion.

Assuming that proteins in solution can be modeled as spheres of radius \( R_p \), the Van der Waals and electrostatic interaction potentials as a function of intermolecular distance \( r \) can be computed from the Hamaker expression (Equation 2.9) and from the modified Hogg-Healy-Fursteneau expression (Equation 2.10), respectively [19, 125, 126]:

\[
V_{\text{VdW}}(r) = -\frac{A_H}{6} \left( \frac{2}{r/R_p - 4} + \frac{2}{(r/R_p)^2} + \ln \left( 1 - \frac{1}{(r/R_p)^2} \right) \right) \tag{2.9}
\]

\[
V_{\text{El}}(r) = \frac{4\pi\varepsilon_0\varepsilon_r R_p^2}{r/R_p} \ln \left( 1 + \exp \left( -\kappa R_p \left( \frac{r}{R_p} - 2 \right) \right) \right) \tag{2.10}
\]

In the above equations, \( A_H \) is the Hamaker constant, \( \varepsilon_0 \) is the vacuum permittivity, \( \varepsilon_r \) is the relative dielectric constant of the medium, \( \psi_0 \) is the protein surface potential and \( \kappa \) is the Debye parameter, also called inverse Debye length.

As most of the proteins have similar compositions, it can be assumed that proteins in aqueous solutions share similar Hamaker constants, which is expected to be in the order of \( 3 - 5k_B T \) [127, 128], and was set here to \( A_H = 3k_B T \).

The Debye parameter was computed as:

\[
\kappa = \sqrt{\frac{2e^2 N_a I}{\varepsilon_r \varepsilon_0 k_B T}} \tag{2.11}
\]

where \( N_a \) is the Avogadro number, \( k_B \) is the Boltzmann constant, \( T \) is the temperature and \( I \) is the solution ionic strength defined as:

\[
I = \frac{1}{2} \sum_j z_j^2 c_j^b \tag{2.12}
\]

where \( z_j \) and \( c_j^b \) are respectively the charge and bulk concentration of the various ionic species present in solution.
Assuming that the distribution of all the ionic species present in the system can be described by the Poisson-Boltzmann equation, the protein surface potential can be estimated from the protein surface charge density $\sigma_0$ by solving the following equation [126]:

$$\sigma_0 = -\sqrt{\frac{2F_\epsilon r_0 k_B T}{e} \sum_j c_j^b \left( \exp \left( -\frac{z_j e \psi_0}{k_B T} \right) - 1 \right)}$$

(2.13)

where $F$ is the Faraday constant.

In this work, the protein surface charge density was computed from the value of the protein net charge estimated from the zeta potential measurements.

The Fuchs stability ratio was then computed from the total interaction potential $V_T$ (defined as the sum of $V_{dW}$ and $V_{El}$) according to Equation 2.7.

### 2.3 Results

The aggregation kinetics of the two antibodies were performed at neutral pH under thermal stress (pH 6.5, 70 °C) in the protein concentration range 1 - 5 g/L. The high temperature of 70 °C was selected in order to induce protein unfolding, which promotes the aggregation process. Protein thermal stability was investigated by CD and DLS, as presented in Figure 2.1. DLS experiments showed that significant aggregation occurs above 70 °C for both antibodies (Figure 2.1(b)), which is consistent with the notable changes in protein structure observed by CD at this temperature (Figure 2.1(a)).

After characterizing the protein thermal stability, the monomer depletion, the dimer and trimer formation, as well as the aggregate weight-average molecular weight were monitored by SEC-MALS at 70 °C. Moreover, the increase in average hydrodynamic radius of the non-fractionated population was followed by DLS in situ. The experimental monomer conversion has been analyzed to characterize the apparent reaction order, while experiments on diluted samples have been
performed to probe for aggregate reversibility. Then, aggregation kinetic schemes consistent with the experimental observations were developed for each mAb. The proposed kinetic models have been validated by fitting the model parameters at a reference protein concentration (2 g/L) and by predicting the aggregation kinetics at two different protein concentrations (1 and 5 g/L). In the following, the results are reported for mAb-1 first, and then for mAb-2.

2.3.1 mAb-1

2.3.1.1 Experimental observations

The measured concentrations of monomer, dimer and trimer, the aggregate weight-average molecular weight and the average hydrodynamic radius of the overall population at protein concentrations of 1, 2 and 5 g/L are shown in Figure 2.2.

The analysis of the concentration dependence of the monomer depletion kinetics provides relevant information on the rate-limiting step of the aggregation mechanism under investigation. For a first order kinetic, the depletion of the monomer concentration $M$ can be described by the equation $\frac{dM}{dt} = -k_{app}M$, which can be integrated to give $\ln\left(\frac{M}{M_0}\right) = -k_{app}t$, with $M_0$ corresponding to the initial protein concentration and $k_{app}$ being independent of $M_0$. Figure 2.3(a) presents the logarithm of the residual monomer normalized by the total protein concentration as a function of time for three initial protein concentrations (1 g/L, 2 g/L and 5 g/L).
Figure 2.2: Comparison between model simulations (lines) and experimental data (symbols) for mAb-1 aggregation. The parameters were determined at the protein concentration of 2 g/L based on the fittings to (a) the monomer depletion and oligomer formation, to (b) the aggregate weight-average molecular weight, and to (c) the average hydrodynamic radius. The parameters determined at the protein concentration of 2 g/L were used to predict the data measured at 1 g/L (d-f) and 5 g/L (g-i). The parameter values are summarized in Table 2.1.

All the experimental points are aligned on a single straight line, indicating that the monomer consumption follows a first order process. This observation suggests that the monomer consumption of mAb-1 is rate-limited by an unimolecular reaction step, i.e. protein unfolding from a native state to a non-native aggregation-prone conformational state.
Figure 2.3: Experimental observations for mAb-1. (a) Linearization of the monomer concentration data, highlighting that the monomer depletion kinetics of mAb-1 follows a first order process. (b) Dilution experiments, showing that mAb-1 aggregates are irreversible. (c) SLS experiments, providing the fractal dimension value of $d_f = 1.85$ for mAb-1 aggregates.

The aggregate reversibility was then investigated by dilution experiments. Figure 2.3(b) shows the monomer and aggregate content of a 5 g/L sample of mAb-1 that was incubated at 70 °C for 15 min and analyzed immediately after incubation. These values are compared to the results obtained after a ten times dilution. The ten times diluted sample was analyzed both immediately after dilution and after one day of storage in the refrigerator. It can be seen that the monomer concentration and the aggregate content were not affected by dilution, thus indicating that the formed aggregates are irreversible.
The aggregate morphology was studied by SLS at various aggregation time points. Figure 2.3(c) shows the structure factor $S(q)$ as a function of $q(R_g)$ for five average aggregate sizes that were obtained by incubating a sample of mAb-1 at 70 °C at the protein concentration of 1 g/L for different times. The overlapping of the curves reveals that aggregates exhibit self-similarity, and the power-law regime provides a fractal dimension value equal to $d_f = 1.85 \pm 0.02$.

### 2.3.1.2 Kinetic model development

The experimental characterization indicate that (i) the monomer depletion of mAb-1 is rate-limited by protein unfolding; and that (ii) mAb-1 aggregates are irreversible. Based on these observations, the aggregation kinetic scheme presented in Figure 2.4 is proposed.

\[ (\text{M1-1}) \text{ Monomeric unfolding} \quad N \xrightarrow{k_U} U \]

\[ (\text{M1-2}) \text{ Oligomer formation} \quad \begin{cases} U + U \xrightarrow{k_{11}} U_2 \\ U + U_2 \xrightarrow{k_{12}} U_3 \end{cases} \]

\[ (\text{M1-3}) \text{ Aggregate growth} \quad \begin{cases} U + U_i \xrightarrow{k_{i1}} U_{i+1} \\ U_i + U_j \xrightarrow{k_{ij}} U_{i+j} \end{cases} \quad i \geq 3, \quad i, j \geq 2 \]

**Figure 2.4:** Aggregation mechanism proposed for mAb-1 aggregation.

First, the monomer in its native form $N$ unfolds to form $U$, which is a denatured conformational state of the monomeric protein. This step, denoted as (M1-1) in the kinetic scheme, is regarded as irreversible in this study. Indeed, $U$ can be considered as an intermediate reactive species which is depleted by irreversible aggregation before it can re-fold. Therefore, aggregation is faster with respect to the possible backward reaction of unfolding, and the reversibility of the unfolding step can be neglected. The aggregation-prone form of the protein, $U$, can then aggregate to form oligomers, according to step (M1-2). Finally, aggregates grow irreversibly either by monomer addition or by cluster-cluster aggregation, as depicted in step (M1-3).
The PBE describing mAb-1 aggregation are shown below:

\[
\begin{align*}
\frac{dN}{dt} & = -k_U N \\
\frac{dU}{dt} & = +k_U N - U \sum_{j=1}^{\infty} k_{1,j} U_j \\
\frac{dU_{i \geq 2}}{dt} & = \frac{1}{2} \sum_{j=1}^{i-1} k_{j,i-j} U_j U_{i-j} - U_i \sum_{j=1}^{\infty} k_{i,j} U_j
\end{align*}
\] (2.14)

Note that the formation term of the dimer is given by \(\frac{1}{2} k_{1,1} U_1 U_1\) (where \(U_1\) is by definition equal to \(U\)). The factor 1/2 prevents overcounting since the two monomers are indistinguishable (as compared for example to the case of monomer-dimer aggregation, where the two colliding species are distinguishable) [129].

To account for the differences in reactivity of the various species present in solution, different Fuchs stability ratios are considered. In particular, the unfolded aggregation-prone monomer \(U\) is an unstable intermediate which has a very high reactivity compared to other aggregates species. Therefore, three types of aggregation events were identified: monomer-monomer, monomer-aggregate and aggregate-aggregate. Accordingly, three Fuchs stability ratios are defined: \(W_{11}\), \(W_{1i}\) and \(W_{ij}\), which describe oligomer formation, aggregate growth by monomer addition and aggregate growth by cluster-cluster aggregation, respectively. It is worth precising that in order to reflect the high reactivity of the dimer, which is suggested by its low and nearly constant concentration, it is assumed that \(W_{12} = W_{11}\). The aggregation rate constants for each aggregation event are then computed according to Equation 2.6.

\[2.3.1.3 \quad \text{Model validation}\]

The implementation of the proposed kinetic scheme requires the estimation of several parameters. The fractal dimension has been measured by SLS and the power law factor \(\lambda\) appearing in the aggregation kernel has been estimated as \(\lambda \approx 1 - 1/d_f\). The remaining parameters (i.e., the unfolding rate constant and the three Fuchs ratios) have been fitted to describe the experimental data at the
reference protein concentration of 2 g/L. Nevertheless, since the concentration of $U$ is low and nearly constant due to its high reactivity, the value of unfolding rate constant $k_U$ is very close to $k_{app}$, which is determined from the linearization of the experimental monomer depletion. The parameters used for the simulations are summarized in Table 2.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k_U$</th>
<th>$d_f$</th>
<th>$\lambda$</th>
<th>$W_{11}$</th>
<th>$W_{1j}$</th>
<th>$W_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>$1 \times 10^{-3}$ s$^{-1}$</td>
<td>1.85</td>
<td>0.5</td>
<td>$8.5 \times 10^6$</td>
<td>$4 \times 10^7$</td>
<td>$8 \times 10^8$</td>
</tr>
<tr>
<td>Source</td>
<td>Fit</td>
<td>SLS Exp.</td>
<td>$1 - 1/d_f$</td>
<td>Fit</td>
<td>Fit</td>
<td>Fit</td>
</tr>
</tbody>
</table>

**Table 2.1:** Parameters used for the simulations of the kinetics of aggregation of mAb-1 in the protein concentration range from 1 to 5 g/L.

In Figure 2.2(a-c) it can be seen that the simulations are in excellent agreement with the experiments, indicating that the proposed kinetic model can successfully describe the aggregation of mAb-1 under the investigated conditions.

To further validate the proposed kinetic scheme, the kinetics of aggregation were simulated at protein concentrations of 1 and 5 g/L using the same set of values reported in Table 2.1, with no additional parameters. In Figure 2.2(d-i), it can be seen that the model predictions agree very well with all experimental results, proving that the model is capable of predicting the concentration effect on the aggregation kinetics of mAb-1 in the concentration range from 1 to 5 g/L.

The simulated concentrations of the two monomeric states of mAb-1, $N$ and $U$, are presented in Figure 2.5, showing the exponential depletion of $N$ and the formation of $U$, which is then consumed by aggregation. The total monomer

![Figure 2.5](image)

**Figure 2.5:** Simulations of the time evolution of the monomeric species of mAb-1 at protein concentration of (a) 1 g/L and (b) 5 g/L.
concentration $M = N + U$ is the quantity detected by SEC. Simulations are presented at protein concentrations of 1 and 5 g/L, showing that the relative concentration of $U$ decreases when the mAb concentration is increased.

The comparison between the different fitted Fuchs stability ratios provides information on the relative reactivity of the various species involved in the aggregation process. Considering the values reported in Table 2.1, it can be noticed that $W_{1j} \ll W_{ij}$. This means that, for this system, aggregate growth by monomer addition prevails over aggregate growth by cluster-cluster aggregation, probably due to the high reactivity of the unfolded monomer.

### 2.3.2 mAb-2

#### 2.3.2.1 Experimental observations

After characterizing the aggregation mechanism of mAb-1, we apply the same analysis to mAb-2. In Figure 2.6, the measured concentrations of monomer, dimer and trimer, as well as the aggregate weight-average molecular weight, and the average hydrodynamic radius as a function of time at protein concentrations of 1, 2 and 5 g/L are reported.

In analogy with mAb-1, we evaluated the apparent reaction order from the concentration dependence of the experimental monomer depletion. This value has been found equal to 2, as shown in Figure 2.7(a), where the monomer conversion data are plotted in the linearized form $\frac{1}{M} - \frac{1}{M_0} = k_{app}t$, which is the integrated solution of the mass balance $\frac{dM}{dt} = -k_{app}M^2$. All experimental points fall on a single straight line passing through the origin, showing that the monomer depletion of mAb-2 follows a second order process. This observation suggests that the monomer consumption of mAb-2 is rate-limited by bimolecular collisions, in contrast with the situation observed with mAb-1, where monomer consumption is rate-limited by a monomolecular event, i.e. protein unfolding.
Figure 2.6: Comparison between model simulations (lines) and experimental data (symbols) for mAb-2 aggregation. The parameters were determined at the protein concentration of 2 g/L based on the fittings to (a) the monomer depletion and oligomer formation, to (b) the aggregate weight-average molecular weight, and to (c) the average hydrodynamic radius. The parameters determined at the protein concentration of 2 g/L were used to predict the data measured at 1 g/L (d-f) and 5 g/L (g-i). The parameter values are summarized in Table 2.2.

In Figure 2.7(b), we show the reversibility analysis by comparing the monomer and aggregate contents of a 5 g/L sample of mAb-2 that was incubated at 70 °C for 5 min and analyzed immediately after incubation with the results obtained with the same sample after a ten times dilution. The ten times diluted sample was analyzed both immediately after dilution and after one day of storage in
the refrigerator. The results show that the monomer concentration is affected by dilution, thus suggesting that the oligomer formation of mAb-2 is reversible. It can be noticed that not only the dimer and trimer concentrations are reduced upon dilution, but also the concentration of larger oligomers. Since all aggregates larger than trimer elute from SEC in a single unresolved peak, the size of the largest reversible aggregate cannot be determined with accuracy. In the absence of precise information on the reversible oligomer formation, tetramer was assumed to be the largest reversible aggregate.

Figure 2.7: Experimental observations for mAb-2. (a) Linearization of the monomer concentration data, highlighting that the monomer depletion kinetics of mAb-2 follows a second order process. (b) Dilution experiments, showing that mAb-2 aggregates are irreversible. (c) SLS experiments, providing the fractal dimension value of $d_f = 2.05$ for mAb-2 aggregates.

<table>
<thead>
<tr>
<th>Time [min]</th>
<th>Monomer</th>
<th>Dimer</th>
<th>Trimer</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
\ln\left(\frac{M}{M_0}\right) = \frac{1}{2} \cdot t
\]

\[d_f = 2.05\]

\[\langle R_g \rangle\]

\[S(q)\]

\[q \langle R_g \rangle [-]\]
Figure 2.7(c) shows the structure factor $S(q)$ as a function of $q\langle R_g \rangle$ for four average aggregate sizes. The overlapping of the curves reveals that, as in the case of mAb-1, aggregates exhibit self-similarity. The fractal dimension can be estimated from the power law regime equal to $d_f = 2.05 \pm 0.02$.

2.3.2.2 Kinetic model development

Based on the above experimental observations, the mechanism presented in Figure 2.8 is proposed for mAb-2 aggregation.

\begin{align*}
&\text{(M2-1) Monomeric unfolding} \\
&M : N \leftrightarrow U \\
&\begin{cases}
M + M \xrightleftharpoons[\kappa_1]{\kappa_1} M_2 \\
M + M_2 \xrightleftharpoons[\kappa_{12}]{\kappa_{12}} M_3 \\
M + M_3 \xrightleftharpoons[\kappa_{13}]{\kappa_{13}} M_4 \\
M_2 + M_2 \xrightleftharpoons[\kappa_{22}]{\kappa_{22}} M_4
\end{cases}
\end{align*}

\begin{align*}
&\text{(M2-2) Oligomer formation} \\
&M_4 \xrightarrow{\kappa_{\text{nucl}}} M_4.
\end{align*}

\begin{align*}
&\text{(M2-3) Structural rearrangement} \quad M_4 \xrightarrow{\kappa_{\text{nucl}}} M_4.
\end{align*}

\begin{align*}
&\text{(M2-4) Aggregate growth} \\
&M_4^* + M_i \xrightarrow{\kappa_{4i}^{-1}} M_{i+4} \quad i \geq 1 \\
&M_i + M_j \xrightarrow{\kappa_{ij}} M_{i+j} \quad i \geq 5, j \geq 1
\end{align*}

Figure 2.8: Aggregation mechanism proposed for mAb-2 aggregation.

At the considered high temperature, a population of partially unfolded conformational states of the antibody is likely present. However, since the aggregation process is limited by bimolecular collisions, the collected set of experimental data does not provide information on the kinetics and thermodynamics of the unfolding step. It is therefore not possible to determine which monomeric form is involved in the aggregation process. For the sake of generality, we consider reversible monomeric changes from the native state ($N$) to a non-native state of the protein ($U$), as schematized in step (M2-1) of the reaction scheme, and we introduce a generic aggregation-prone intermediate ($M$) in the aggregation kinetic scheme.
This generic intermediate is able to represent in a coarse-grained approach all the possible unfolding scenarios. Two limiting cases consistent with the observed second order kinetics of the monomer depletion could occur: the protein could react (i) in its native form \( (M = N) \); or (ii) in a denaturated state \( (M = U) \) provided that the generation of \( U \) is fast compared to its consumption by aggregation. Another intermediate situation could occur where only a fraction \( f \) of the monomer would be prone to aggregate. In this case, the dimerization and monomer addition rate constants can be computed as apparent rate constants \( k_{11}^{\text{app}} = f^2 \times k_{11} \) and \( k_{1j}^{\text{app}} = f \times k_{1j} \).

The reactive monomer can form a nucleus by reversible oligomerization followed by irreversible structural rearrangement, according to the steps (M2-2) and (M2-3) in Figure 2.8, respectively. As mentioned above, tetramer has been assumed to be the largest reversible species. The complete description of reversible tetramer formation requires eight parameters (four forward and four backward reactions). The reversibility experiments showed that the aggregates are not equally reversible, and in particular trimer appears to be the least reversible species. In order to reduce the number of fitting parameters, we assumed that oligomer formation can be described by a single Fuchs stability ratio, i.e. \( W_{11} = W_{12} = W_{13} = W_{22} \). In addition, we assumed that each bound inside the tetramer can break with equal probability, and therefore \( k_{13}^r \approx 2k_{22}^r \), since the probability that a tetramer breaks into one monomer and one trimer is twice higher than the probability to form two dimers. Consequently, the number of parameters describing reversible oligomerization is reduced from eight to four \( \left( W_{11}, k_{11}^r, k_{12}^r, k_{13}^r \right) \).

Finally, the step (M2-4) in the reaction scheme represents aggregate growth, which can occur either by monomer addition or by cluster-cluster aggregation.

In order to account for the differences in reactivity of the various populations present in solution, three different Fuchs ratios have been considered: one for oligomer formation \( (W_{11}) \), one for nucleus consumption \( (W_{1*}) \), and one for cluster-cluster aggregation \( (W_{ij}) \).
The population balance equations describing mAb-2 aggregation are shown below:

\[
\begin{align*}
\frac{dM}{dt} &= -M \sum_{j=1}^{3} k_{1,j} M_j - k_{1,4^*} M M_{4^*} - M \sum_{j=5}^{\infty} k_{1,j} M_j \\
&\quad + 2k'_{11} M_2 + k'_{12} M_3 + k'_{13} M_4 \\
\frac{dM_2}{dt} &= \frac{1}{2} k_{1,1} M M - k_{1,2} M M_2 - k_{2,2} M_2 M_2 - k_{2,4^*} M_2 M_{4^*} \\
&\quad - M_2 \sum_{j=5}^{\infty} k_{2,j} M_j + 2k'_{r_{11}} M_4 - k'_{r_{11}} M_2 - k'_{r_{12}} M_3 \\
\frac{dM_3}{dt} &= k_{1,2} M M_2 - k_{1,3} M M_3 - k_{3,4^*} M_3 M_{4^*} - M_3 \sum_{j=5}^{\infty} k_{3,j} M_j \\
&\quad + k'_{13} M_4 - k'_{12} M_3 \\
\frac{dM_4}{dt} &= k_{1,3} M M_3 + \frac{1}{2} k_{2,2} M_2 M_2 - k'_{r_{22}} M_4 - k_{nuc} M_4 \\
\frac{dM_{4^*}}{dt} &= k_{nuc} M_4 - M_{4^*} \sum_{j=1}^{\infty} k_{4^*j} M_j \\
\frac{dM_{i \geq 5}}{dt} &= k_{4^*,i-4} M_{4^*} M_{i-4} + \frac{1}{2} \sum_{j=5}^{i-1} k_{j,i-j} M_j M_{i-j} - M_i \sum_{j=1}^{3} k_{i,j} M_j \\
&\quad - k_{4^*,i} M_{4^*} M_i - M_i \sum_{j=5}^{\infty} k_{i,j} M_j 
\end{align*}
\]

(2.15)

2.3.2.3 Model validation

As in the case of mAb-1, the fractal dimension was measured by independent SLS experiments and \( \lambda \) was estimated as \( \lambda \approx 1 - 1/d_f \). All the other kinetic parameters (\( W_{11}, W_{4^*j}, W_{ij}, k'_{11}, k'_{12}, k'_{13}, k_{nuc} \)) were estimated by fitting to five independent sets of experimental data (monomer, dimer and trimer concentrations as well as aggregate weight-average molecular weight and average hydrodynamic radius of the overall population) at the reference protein concentration of 2 g/L.

In Figure 2.6(a-c), we present the comparison between experimental and simulated kinetics of aggregation of mAb-2. It can be seen that the simulations are in very close agreement with all experimentally accessible quantities. The parameter values used for the simulations are summarized in Table 2.2.
Table 2.2: Parameters used for the simulations of the kinetics of mAb-2 aggregation in the protein concentration range from 1 to 5 g/L.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$d_f$</th>
<th>$\lambda$</th>
<th>$W_{11}$</th>
<th>$W_{4j}^*$</th>
<th>$W_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>2.05</td>
<td>0.5</td>
<td>$2.7 \times 10^7$</td>
<td>$10^6$</td>
<td>$4 \times 10^8$</td>
</tr>
<tr>
<td>Source</td>
<td>SLS</td>
<td>1 $- 1/d_f$</td>
<td>Fit</td>
<td>Fit</td>
<td>Fit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k_{r11}^r$</th>
<th>$k_{r12}^r$</th>
<th>$k_{r13}^r$</th>
<th>$k_{nuc}^r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>$4.4 \times 10^{-3}$</td>
<td>$1.4 \times 10^{-4}$</td>
<td>$1.3 \times 10^{-2}$</td>
<td>$3.7 \times 10^{-3}$</td>
</tr>
<tr>
<td>Source</td>
<td>Fit</td>
<td>Fit</td>
<td>Fit</td>
<td>Fit</td>
</tr>
</tbody>
</table>

With the parameters determined at the reference protein concentration of 2 g/L, the aggregation kinetics were then simulated at 1 and 5 g/L, as shown in Figure 2.6(d-i). It is seen that the model predictions of the monomer depletion, dimer and trimer formation, as well as the aggregate weight-average molecular weight and average hydrodynamic radius of the non-fractionated population are in good agreement with the experimental data, thus confirming the validity of the proposed kinetic model.

As in the case of mAb-1, the ranking of the Fuchs stability ratios provides information on the relative reactivity of the different species present in the system. Considering the values of $W$ reported in Table 2.2, the various species involved in the kinetic mechanism can be classified according to their reactivity in the following order: nucleus $>$ oligomers $>$ large aggregates. This observation is consistent with the definition of nucleus as the least stable species present in the system.

## 2.4 Discussion

### 2.4.1 Comparison between the two mAbs

The kinetics of aggregation of two monoclonal antibodies, which differ in terms of IgG subclass, glycosylation and net charge, were followed at neutral pH under thermal stress. The comparison between experimental and simulated kinetic
data allowed the identification of the key steps involved in the aggregation mechanism of the two proteins. The two mAbs were found to exhibit different aggregation behaviors under the same operating conditions. In Figure 2.9, a scheme of the aggregation mechanisms proposed for each mAb is presented, highlighting the identified rate-limiting step for monomer consumption in each aggregation pathway. The kinetic mechanism proposed for mAb-1 includes protein unfolding, which is the rate-limiting step, followed by irreversible oligomer formation and aggregate growth, which is dominated by monomer addition and accompanied by cluster-cluster aggregation. The mechanism of mAb-1 aggregation significantly differs from the aggregation pathway identified for mAb-2, which was found to be rate-limited by bimolecular collisions and includes reversible oligomer formation, irreversible tetramer rearrangement and aggregate growth.

(a) mAb-1

\begin{itemize}
  \item[I.] Monomeric conformational changes
  \item[II.] Irreversible oligomerization
  \item[III.] Irreversible aggregate growth
\end{itemize}

(b) mAb-2

\begin{itemize}
  \item[I.] Monomeric conformational changes
  \item[II.] Reversible oligomerization
  \item[III.] Irreversible rearrangement
  \item[IV.] Irreversible aggregate growth
\end{itemize}

\textbf{r.l.s.: rate-limiting step for monomer depletion}

\textbf{Figure 2.9:} Scheme of the mechanisms of mAb-1 and mAb-2 aggregation.
It would be of great relevance to correlate the observed aggregation behaviors with antibody properties (such as IgG subclass, glycosylation pattern or protein charge). However, such an attempt is extremely challenging since it would require to analyze a large number of antibodies to improve statistical relevance. Regarding the two antibodies under examination in this study, we summarized in Table 2.3 the main differences between the two molecules in terms of both structures and aggregation behaviors. Although correlating antibody structure with aggregation behaviors is beyond the scope of this study, the two following remarks can be formulated.

<table>
<thead>
<tr>
<th>Property</th>
<th>mAb-1</th>
<th>mAb-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibody subclass</td>
<td>IgG1</td>
<td>IgG2</td>
</tr>
<tr>
<td>Glycosylation</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Zeta potential</td>
<td>5.71 ± 3.51 mV</td>
<td>0.53 ± 0.75 mV</td>
</tr>
<tr>
<td>Protein Charge</td>
<td>+4</td>
<td>+0.4</td>
</tr>
<tr>
<td>Apparent order</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rate-limiting step</td>
<td>Monomeric unfolding</td>
<td>Bimolecular aggregation</td>
</tr>
<tr>
<td>Reversibility</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Fractal dimension</td>
<td>1.85 ± 0.02</td>
<td>2.05 ± 0.02</td>
</tr>
</tbody>
</table>

Table 2.3: Comparison between the two antibodies in terms of their structure and aggregation behaviors at pH 6.5 and 70 °C.

First, it can be noted that the antibody which is the more resistant to aggregation (i.e. mAb-1, whose monomer depletion kinetics is rate-limited by protein unfolding) is also the one which is glycosylated and which carries the highest net charge.

Second, it can be observed that the antibody carrying a larger net charge (i.e. mAb-1) forms relatively open aggregates characterized by a fractal dimension value of $d_f = 1.85$. On the other hand, the antibody carrying a lower net charge (i.e. mAb-2) forms more compact aggregates characterized by a fractal dimension of $d_f = 2.05$. This observation might be due to the fact that protein clusters can interpenetrate more when the energy barrier between protein molecules is low.
2.4.2 DLVO theory

The kinetic approach used in this study allows the quantification of the inter-molecular interaction potentials, expressed in terms of the Fuchs stability ratio \( W_{11} \), from the reaction rate constants estimated by fitting to the experimental data. The Fuchs stability ratio quantifies the fraction of collisions which are successful in forming larger aggregates, providing a measure of the extent to which the aggregation kinetics is delayed with respect to diffusion-limited conditions.

The Fuchs stability ratio can be directly related to the protein-protein interaction potential, as shown in Equation 2.7. It quantifies protein-protein interactions in a coarse-grained manner, averaging protein surface heterogeneities over space. It is of great interest to compare the Fuchs stability value obtained by fitting to experimental kinetics of aggregation to the theoretical value computed with Equation 2.7. As illustrated in the next paragraph, this comparison allows to identify some gaps between theories dealing with colloidal stability (such as the DLVO theory) and experimental observations. In the following, the values of the Fuchs stability ratio evaluated by fitting to experimental kinetics of aggregation are denoted by \( W_{11} \), while the values computed theoretically in the frame of the DLVO theory are referred to as \( W_{11}^{DLVO} \).

The values of the Fuchs stability ratio estimated by fitting model simulations to experimental data are in the order of \( W_{11} \sim 10^7 \) (Table 2.1 and Table 2.2). This large value indicates the presence of a high energy barrier that particles must overcome before colliding, which reduces the collision efficiency and thus delays the aggregation process with respect to diffusion-limited conditions.

On the other hand, the very low values of zeta potential of the antibody solutions (see Table 2.3) suggest that, at the considered pH of 6.5, the antibody net charge is close to zero. Net electrostatic stabilization is thus expected to be small.

Figure 2.10 shows the computed contribution of the Van der Waals attraction and of the electrostatic repulsion to the total protein-protein interaction potential
as a function of the intermolecular distance, as computed from Equations 2.9 and 2.10 for mAb-1 and mAb-2. Based on these calculations, $W_{DLVO}^{11} \sim 1$ was computed for the two antibodies from Equation 2.7. These low computed values of the Fuchs stability ratio, which correspond to the DLCA regime, reflect the low net charge of the antibody solutions under these conditions, and are several orders of magnitude smaller than the values estimated by kinetic analysis.

![Diagram](image)

**Figure 2.10:** Computed protein-protein interaction potential ($V$) of (a) mAb-1 and (b) mAb-2 in the frame of the DLVO theory as a function of intermolecular distance. The Fuchs stability ratios computed from the integration of the interaction potentials are close to 1 for both mAbs.

This difference between the measured and the calculated Fuchs stability ratio values is striking, and it is not affected by potential inaccuracies in the assessment of the protein net charge from zeta potential measurements (partly due to the fact that zeta potential measurements were carried out at room temperature, while aggregation kinetics were performed under thermal stress). Indeed, an increase in the protein net charge of mAb-1 by a factor of three would result in a change in the computed Fuchs ratio from $W_{DLVO}^{11} \sim 1$ to $W_{DLVO}^{11} = 2.3$, which is still much lower than the value obtained by kinetic analysis. Moreover, to obtain a value of the Fuchs stability ratio of $W_{DLVO}^{11} = 8.5 \times 10^6$, which is equal to the value obtained by fitting to the experimental data in the case of mAb-1, an unreasonable change in the protein charge from $z = +4$ to $z = +43$ would be required.

Therefore, our results strongly indicate that the colloidal stability of antibody solutions cannot be rationalized only in the frame of the DLVO theory, which considerably underestimates the repulsive energy barrier between two approaching
molecules. The DLVO theory may underestimate the colloidal stability of antibody solutions for several reasons including:

(i) the low collision efficiency due to the heterogeneous protein surface, which may provide only a limited number of aggregation-prone patches and thus requires specific geometrical orientation of the two aggregating molecules, in contrast with sticking homogeneous spherical colloids;

(ii) additional repulsive non-DLVO repulsion forces, such as hydration forces, which further contribute to the interaction potential.

2.5 Conclusion

We investigated the aggregation behavior under thermal stress of two monoclonal antibodies belonging to the IgG1 and IgG2 subclass.

The combination of experimental data and theoretical kinetic analysis allowed us to quantify the contribution of the single elementary steps on the global aggregation rate. We demonstrated that the two mAbs exhibit different aggregation mechanisms under the same operating conditions. In particular, it was found that the monomer depletion kinetics of mAb-1 is rate-limited by unimolecular protein unfolding, while bimolecular aggregation was identified as the rate-limiting step for the monomer depletion of mAb-2. In the next chapter, the kinetic models developed here will be used to investigate the impact of cosolutes on the elementary events that contribute to the global aggregation rate.

Moreover, for each mAb, the aggregate growth steps were described by using the RLCA kernel. Several values of the Fuchs stability ratio were considered to characterize the reactivity of the various sub-population of species (including monomer, oligomers and larger aggregates). Remarkably, the proposed models were able to predict the concentration dependence of the aggregation kinetics in the protein concentration range from 1 to 5 g/L.
Finally, from the measurement of the kinetic rate constants, we quantified the protein-protein interaction potentials expressed in terms of the Fuchs stability ratio, and we showed that the DLVO theory considerably underestimates the colloidal stability of the antibody solutions under investigation. This result is possibly due to the reduced efficiency of collisions related to the presence of a limited number of specific aggregation-prone patches on the heterogeneous protein surface, as well as to the contribution of additional repulsive non-DLVO forces to the protein-protein interaction potential, such as hydration forces.

Chapter 3

Fragmentation mechanisms of insulin fibrils

3.1 Introduction

In this chapter, we apply an approach similar to the one introduced in Chapter 2 (i.e. we combine experimental data with PBE simulations) to gain insights into the kinetics of insulin fibril formation. In addition to its crucial importance as a therapeutic protein, insulin also represents a good model system to study amyloid fibril formation.

Indeed, just like insulin, a wide variety of proteins and peptides share the common feature to readily self-assemble into highly organized fibrillar aggregates that are characterized by supramolecular β-sheet structures, known as amyloid or amyloid-like fibrils. These protein fibrillar aggregates have received great attention over the past decades due to their connection with neurodegenerative disorders, including Alzheimer’s and Parkinson’s diseases [130, 131]. It was discovered later on that, besides their implication in pathological phenomena, amyloid fibrils also perform beneficial biological functions [130, 132]. For example, some bacteria take advantage of the peculiar properties of amyloid-like fibrils to form protective or
adhesive biofilms [132]. Moreover, the remarkable strength and rigidity of amyloid fibrils, which are conferred by the dense hydrogen-bonding network, open up the possibility to produce novel high-performance bionanomaterials with diverse technological applications [133, 134]. The perspective of both combatting protein aggregation diseases and of creating sophisticated materials requires to unravel the complex mechanisms underlying the process of amyloid fibril formation.

In spite of the differences in the protein amino acid sequence and of the broad range of operating conditions explored during in vitro studies, both experimental data and computer simulations suggest the existence of a common pathway for fibril formation [32, 66, 135, 136]. Indeed, amyloid fibril formation has almost universally been shown to follow a nucleated polymerization mechanism, which consists of a nucleation step followed by monomer addition (also referred to as fibril elongation) events.

Moreover, there is increasing evidence that fibril fragmentation (also called breakage) significantly contributes to the kinetics of fibril growth as it produces new reactive fibril ends prone to elongation [68, 87, 137]. Along with the formation of new reactive seeds, fibril fragmentation heavily reshapes the fibril length distribution, which may have crucial implications on biological functions and on biomaterial properties. The recent discovery that fibril fragmentation enhances amyloid cytotoxicity sparked interest in understanding the mechanisms of fibril breakage [138, 139].

In this frame, deterministic kinetic models based on mass-action laws coupled with proper experimental characterization have been proven useful to get insights into the mechanisms of formation and breakage of protein aggregates [62, 63, 66]. In particular, the comparison between simulations from population balance equations (PBE) with the experimental time evolution of the fibril length distribution followed by atomic force microscopy (AFM) provides fundamental knowledge on the amyloid fibril fragmentation process [68, 88].

However, the large number of monomeric units constituting amyloid fibrils implies solving numerous (typically several tens of thousands) coupled ordinary
differential equations, thus requiring an extensive computational effort. One common strategy to significantly reduce the simulation time involves solving the moment equations and then reconstructing the fibril length distribution by assuming a suitable distribution type [68]. Although this procedure dramatically eases the computational work, it suffers from one severe limitation: only average properties can be reliably determined.

In this chapter, we employ a methodology which offers the advantage of capturing accurately all the details of the fibril length distribution, while still being extremely efficient in terms of simulation time. Briefly, the discretization technique is based on the approximation of the fibril distribution as the sum of only a small number of Gaussian functions. This numerical method, which was shown to be effective in studying polymer and colloidal systems [140, 141], is applied here to amyloid fibrils.

The aim of this chapter is to deepen the mechanism of insulin fibril fragmentation as a function of temperature and agitation rate. In particular, we intend to clarify the dependence of the fibril breakage rate constant both on the fibril length as well as on the position of fragmentation along the fibril longitudinal axis. In this context, we use the concept of daughter distribution, widely used in colloid science, which describes the probability of breakage along the fibril longitudinal axis. We are interested in discriminating between three types of breakage mechanisms: central mechanism (i.e. breakage occurs preferentially at the center of the fibril), erosion mechanism (i.e. breakage occurs preferentially at the ends of the fibril), and random mechanism (i.e. breakage occurs with the same likelihood at any position).

To do so, we compare the time evolution of insulin fibril distributions, followed by atomic force microscopy, with simulations from a kinetic model based on population balance equations. Moreover, we extend our analysis to other protein systems by considering literature data obtained with β-lactoglobulin [68] and β2-microglobulin [88].
3.2 Materials and methods

3.2.1 Sample preparation and experimental conditions

Purified human insulin was kindly donated by the healthcare company Novo Nordisk (Bagsvaerd, Denmark). Insulin solutions at protein concentration of 10 g/L were freshly prepared prior to each experiment by dissolving insulin powder in a solution of 20 wt % acetic acid in Milli-Q water (Acetic acid puriss. p.a., ≥ 99.8 %, Sigma-Aldrich). The protein solutions were filtered by using syringe filters with 0.20 μm pore size (Millex-LG, 0.20 μm, Hydrophilic, PTFE, Merck Millipore) shortly before performing aggregation experiments.

Isothermal aggregation kinetics were performed at the temperature of 60 °C. To do so, insulin samples of 1.5 mL were placed in 2 mL glass HPLC vials containing a small Teflon stirring bar (Micro 5 × 2 mm, VWR) and incubated in an oil bath heated by means of a hot plate stirrer. Aggregation experiments were performed under stirring conditions, with two different agitation speeds: either 50 or 250 rpm.

In order to follow the kinetics of fibril formation, small aliquots were taken at different incubation times and placed for a few minutes in an ice-water bath to quench the reaction. The samples were then analyzed immediately after by size exclusion chromatography (SEC), thioflavin T (ThT) fluorescence, or deposited on mica disks to allow atomic force microscopy (AFM) measurements.

In order to investigate the impact of temperature on the breakage rate constant, an experiment was performed at room temperature. For this purpose, an insulin solution was incubated at 60 °C under agitation at 250 rpm during 6 h as described above, and then the temperature was reduced to 25 °C while continuing stirring at 250 rpm.
3.2.2 Size exclusion chromatography

The monomer conversion was monitored by SEC with a Superdex Peptide GL, 10 mm × 300 mm size-exclusion column (GE Healthcare, Uppsala, Sweden) assembled on an Agilent series HPLC unit (Santa Clara, CA, USA). The samples were diluted five times in 20 wt % acetic acid solution and filtered with a syringe filter with 0.20 µm pore size (Millex-LG, Hydrophilic, PTFE, Merck Millipore) before analysis. Samples of 50 µL were injected in the SEC column and were eluted for 60 min at a constant flow rate of 0.5 mL/min using as mobile phase a solution of 20 wt % acetic acid. The eluting species were detected by UV absorbance at 280 nm. SEC chromatograms obtained at different incubation time points showed a main peak centered at 21.3 min, and a negligible peak corresponding to insulin dimer at 18.7 min. The monomer content was determined by integrating the area below the main peak by using the Agilent ChemStation software. The points and error bars reported in the plots correspond to the average and standard deviation of measurements performed on at least two independent samples.

3.2.3 ThT fluorescence

In order to follow the kinetics of fibril formation, ThT fluorescence was measured off-line by using an EnSpire 2300 Multilabel Plate Reader (Perkin Elmer, Boston, MA, USA). At each selected incubation times, two aliquots of 147 µL of 5-fold diluted samples were placed in a standard 96-well black plate with a transparent bottom (ProxiPlate-96, Perking Elmer). Then, 3 µL of a 500 µM ThT solution was added to each sample. The dye was excited at the wavelength of 440 nm and the emission was measured at 480 nm in five different points of each well. Measurements were carried out at room temperature and three repetitions were performed per well. For each sample, average values of the 30 measurements were considered (2 wells, 3 repetitions of 5 measurements per well). The points and error bars reported in the plots correspond to the average and standard deviation of measurements performed on at least two independent samples.
3.2.4 Atomic force microscopy

The time evolution of the fibril length distribution was monitored with AFM by using an Asylum Cypher Scanning Probe Microscope (Asylum Research, an Oxford Instruments Company, Santa Barbara, CA, USA) operating in tapping mode. Aliquots of 30 µL of 5 to 1000-fold diluted samples were spotted on a freshly cleaved mica surface for 30 sec before washing gently with Milli-Q water and drying under nitrogen flux. The mica surface was scanned at room temperature at a frequency of 4 Hz with a high resonance frequency silicon cantilever with resonance frequency of 1600 kHz and tip radius of 7 nm (Olympus, Japan). For each mica disk, at least 10 pictures with a scan size of 10 × 10 µm were taken in order to analyze at least 300 fibrils. The recorded pictures were processed (plane correction and background removal) with Gwyddion. The length distribution was determined manually by using the software ImageJ.

3.2.5 Kinetic scheme

The most accepted reaction mechanism for protein fibrillation involves three events, namely nucleation, fibril elongation and fibril breakage [32, 66, 87, 135, 136], as shown in Figure 3.1, and schematically illustrated in Figure 3.2(a).

<table>
<thead>
<tr>
<th>Process</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleation</td>
<td>( n_CM \xrightarrow{k_N} f(n_C, t) )</td>
</tr>
<tr>
<td>Elongation</td>
<td>( f(x, t) + M \xrightarrow{k_E} f(x + 1, t) )</td>
</tr>
<tr>
<td>Fragmentation</td>
<td>( f(x, t) \xrightarrow{k_F(x, y)} f(x - y, t) + f(y, t) )</td>
</tr>
</tbody>
</table>

**Figure 3.1:** General reaction scheme used to describe amyloid fibril formation.

The concentration of monomeric protein is denoted \( M \), while the fibril mass distribution \( f(x, t) \) is defined such that \( f(x, t) \) represents the concentration of fibrils consisting of \( x \) to \( x + dx \) monomeric proteins at time \( t \). The nucleus size is referred to as \( n_C \) (with \( n_C \geq 1 \)), while \( k_N \) and \( k_E \) are the rate constants of nucleation and fibril elongation, respectively. Finally, \( k_F(x, y) \) represents the fragmentation rate constant of a fibril of size \( x \) at the position \( y \). As highlighted in
Figure 3.2(a), we are particularly interested in clarifying the dependence of the fibril breakage rate constant on the position of fragmentation.

![Reaction Scheme and Daughter Distribution](image)

**Figure 3.2:** (a) General reaction scheme used to describe amyloid fibril formation, highlighting the three mechanisms of fibril fragmentation considered in this study. (b) Examples of daughter distributions for the fragmentation mechanisms under investigation.

The main assumptions of the present model, commonly employed in describing amyloid fibrillation [66, 68], are the following: (i) the kinetic constant of elongation $k_E$ is assumed to be fibril-length independent; (ii) fibril-fibril association (both lateral and end-to-end) is neglected; and (iii) fibrils with size $x \leq n_C$ can be formed by breakage only.

To best describe the breakage process, two quantities typically employed in colloid science, namely $\beta(x)$ and $\Gamma(x, y)$, are introduced [86, 142]. In this framework, $\beta(x)$ symbolizes the breakage rate of an $x$-sized fibril:

$$\beta(x) = \int_0^x k_F(x, y)dy$$  \hspace{1cm} (3.1)

On the other hand, $\Gamma(x, y)$ is the daughter distribution function describing the probability of breakage of an $x$-sized fibril at a given position $y$, with respect to all other possible breakage events an $x$-sized fibril undergoes:

$$\Gamma(x, y) = \frac{k_F(x, y)}{\int_0^x k_F(x, y)dy} = \frac{k_F(x, y)}{\beta(x)}$$  \hspace{1cm} (3.2)
3.2.6 Choice of the breakage kernel

One of the first kernels employed to describe amyloid fibril breakage was only size dependent [68]:

\[ k_F(x, y) = k_{F_1} x^{k_{F_2}} \] (3.3)

It was shown that with only two parameters, this kernel describes well the time evolution of the average fibril length of β-lactoglobulin [68]. An interesting evolution was then suggested by Xue and Radford [88], who introduced a 5-parameter, size- and position-dependent breakage kernel:

\[
k_F(x, y) = k_{F_1} x^{k_{F_2}} \times \left[ 1 + k_{F_3} \left( \frac{y - x/2}{x/2} \right)^2 + k_{F_4} \left( \frac{y - x/2}{x/2} \right)^4 + k_{F_5} \left( \frac{y - x/2}{x/2} \right)^6 \right] \] (3.4)

This kernel was proven to capture very well the time-evolution of the complete fibril mass distribution in the case of β2-microglobulin [88]. On the other hand, it can be seen from the expression above that when \( k_{F_3}, k_{F_4}, \) or \( k_{F_5} \) are much larger than unity, they become strongly correlated to \( k_{F_1} \), which complicates the search of the optimal set of parameters.

In order to reduce the number of fitting parameters, we set here the parameters \( k_{F_4} \) and \( k_{F_5} \) to zero. Moreover, with a view to overcoming the aforementioned issue of parameter interdependency, we normalized the breakage kernel presented in Equation 3.4 by the following quantity:

\[
\int_0^x \left( 1 + k_{F_3} \left( \frac{y - x/2}{x/2} \right)^2 \right) dy = x \left( 1 + \frac{k_{F_3}}{3} \right)
\] (3.5)

The breakage kernel that we employed in this work was therefore the following:

\[ k_F(x, y) = k_{F_1} x^{k_{F_2}} \times \frac{1 + k_{F_3} \left( \frac{y - x/2}{x/2} \right)^2}{x \left( 1 + \frac{k_{F_3}}{3} \right)} \] (3.6)
Notably, such empirical kernel involves only 3 parameters ($k_{F_1}$, $k_{F_2}$ and $k_{F_3}$) and guarantees the independence of $k_{F_1}$ and $k_{F_3}$. Indeed, the kernel actually becomes independent of $k_{F_3}$ for $k_{F_3} \gg 1$.

Using the latter kernel (Equation 3.6) to calculate the breakage rate $\beta(x)$ and the daughter distribution function $\Gamma(x, y)$ introduced in Equations 3.1 and 3.2, one gets:

$$\beta(x) = k_{F_1} x^{k_{F_2}} \quad (3.7)$$

$$\Gamma(x, y) = \frac{1 + k_{F_3} \left(\frac{y-x/2}{x/2}\right)^2}{x \left(1 + k_{F_3}/3\right)} \quad (3.8)$$

In order to highlight the versatility of the employed breakage kernel, qualitative daughter distribution functions are shown in Figure 3.2(b) for different values of $k_{B_3}$ (namely -1, 0 and 3). In particular, three different breakage mechanisms can be described depending on the sign of $k_{B_3}$: (i) $k_{B_3} > 0$ leads to an erosion mechanism, i.e. fibrils are more likely to be broken close to their ends (red curve); (ii) $k_{B_3} = 0$ leads to a random mechanism, i.e. breakage occurs with the same likelihood at any position (black curve); and (iii) $k_{B_3} < 0$ leads to a central mechanism, i.e. fibrils are preferentially broken in their center (blue curve).

### 3.2.7 Population balance equations

Given the aforementioned kinetic scheme (Figure 3.2) and the definitions of $\beta(x)$ and $\Gamma(x, y)$, it is possible to write the following PBE in continuous form:

$$\frac{df(x,t)}{dt} = \delta_{nc} \frac{1}{n_C!} k_N M(t)^{n_C} - \delta_1 \left(\frac{n_C}{n_C!} k_N M(t)^{n_C}ight) + 2 k_E M(t) \left(\int_{0}^{\infty} f(y, t)dy - M(t)\right) + \delta_{L_2} 2 k_E M(t)f(x-1, t) - \delta_{L_1} 2 k_E M(t)f(x, t) + 2 \int_{x}^{\infty} \Gamma(z, x) \beta(z) f(z, t)dz - \delta_{L_1} \beta(x) f(x, t) \quad (3.9)$$
The term $1/n_C!$ accounts for the correct amount of permutations resulting from the association of $n_C$ monomeric proteins to form a nucleus [143]. The prefactor 2 in the breakage term accounts for the fact that a fibril of a given size $z$ can break in two positions to form a fibril of size $x < z$, namely in position $x$ or in position $z - x$. In the elongation terms instead, the pre-factor 2 accounts for the fact that both ends of a fibril can each incorporate a monomeric protein. Notably, some terms are multiplied by Kronecker deltas, as not all species undergo the same reactions: for instance, only the monomeric protein is involved in nucleation.

This ordinary differential equations (ODE) system has to be coupled with the initial conditions that $f(x, t = 0)$ is equal to the initial monomeric protein concentration for $x = 1$, and to zero for $x > 1$.

The comparison between the fibril mass distributions simulated by solving the PBE with the fibril length distributions measured experimentally by AFM requires an estimate of the mass per unit length fibril, which can be obtained by scanning transmission electron microscopy [144].

### 3.2.8 Numerical solution

Considering that amyloid fibrils typically consist of at least $10^4$ monomeric protein units, a corresponding number of ODE has to be solved [68, 88]. Such a large ODE system would lead to unacceptably large computational times, especially considering the necessity of performing an optimization to discriminate the occurring breakage mechanism. In order to overcome this problem, a possible strategy is to employ the method of moments, which considerably decreases the number of equations to be solved at the price of losing information on the shape of the distribution.
Here, a discretization method based on Gaussian basis functions [140, 141], which allows the comparison between the experimental and simulated fibril length distributions without assuming any shape of the distribution, has been implemented. The key idea of this discretization method is to approximate the actual fibril mass distribution \( f(x, t) \) with a sum of \( N_G \) Gaussians:

\[
 f(x, t) \simeq \sum_{i=1}^{N_G} \alpha_i(t) \exp \left( -s_i (x - x_i)^2 \right) \tag{3.10}
\]

where \( \alpha_i \) are the time-dependent coefficients of the Gaussians, while \( x_i \) and \( s_i \) are fixed parameters, which determine the positions and the widths of the Gaussians, respectively. In fact, differently shaped distributions can be obtained upon changing the \( \alpha_i \) values, as illustrated in Figure 3.3.

Figure 3.3: Distributions (continuous lines) reconstructed from their respective Gaussians (dashed lines). Despite the Gaussians are always placed in the same positions, different distributions can be described by changing the heights of the Gaussians. For the sake of clarity, only 14 out of 80 Gaussians are shown.

In other words, when approximating the fibril mass distribution with a sum of Gaussians, the coefficients \( \alpha_i \) become the new unknowns: once they are determined, the fibril mass distribution can be reconstructed at any time. The number of ODE to be solved is thus significantly reduced as it is equal to the number of Gaussians chosen to properly approximate the distribution. Numerical experiments show that a very good approximation of the full distribution is obtained with \( N_G \approx 100 \) [140].
To obtain the discretized balances it is sufficient to plug the approximation shown in Equation 3.10. in the PBE presented in Equation 3.9. For the sake of brevity, the detailed derivation of the discretized equations is reported in Appendix B.

3.2.9 Optimization procedure

The general idea behind the optimization procedure is the minimization of the residual sum of squares (RSS), defined as follows:

$$RSS = \sum_{i=1}^{N_{exp}} (y_i^{exp} - y_i^{sim})^2$$  \hspace{1cm} (3.11)

where $y_i^{exp}$ is the $i^{th}$ experimental data point, $y_i^{sim}$ is the corresponding value simulated with the model, and $N_{exp}$ is the total number of experimental points considered.

The quantities $y$ have been defined in terms of residual monomer concentration (accessible experimentally by SEC), aggregate content (corresponding to the normalized ThT fluorescence signal), average fibril length, and fibril fraction populating a given bin of the fibril length distribution (accessible experimentally by AFM measurements). It is worth mentioning that the average fibril length has been computed from the ratio between the first and the zero-th order moments of the fibril length distribution. Moreover, this value has been divided by the maximum average fibril length to render the quantity $y$ dimensionless.

It must be precised that the nucleus size for insulin fibril formation has been determined according to the methodology proposed by Dovidchenko et al. [145]. A nucleus size of 1 has been estimated, in agreement with the findings of Selivanova et al. [146]. In the case of β-lactoglobulin, literature data have been used for the parameters corresponding to the nucleation and elongation steps [68], so that only the parameters corresponding to breakage have been determined in this study. In the case of β2-microglobulin, whose fibril length distributions have been extracted
from the data of Xue and Radford [88], breakage experiments were performed under conditions where no aggregation occurs. Accordingly, only the parameters describing fibril breakage have been evaluated by minimizing the RSS.

3.3 Results

The aim of this study is to investigate the mechanisms of amyloid fibril breakage under various conditions of temperature and agitation. To this end, we used insulin (INS) as a model system and followed with AFM the time evolution of the fibril distribution at high temperature (where both fibril formation and fibril breakage occur), as well as at low temperature (where only fibril breakage occurs). The impact of agitation was instead evaluated by performing experiments at two agitation speeds.

In order to assess the generality of our findings, we also re-analyzed previous results obtained with β-lactoglobulin (bLAC) at high temperature under stagnant conditions [68], and the results of Xue and Radford obtained with β2-microglobulin (b2MGB) under mechanical stress at room temperature [88]. The different systems and their corresponding experimental conditions are summarized in Table 3.1. Note that the fibril distribution is governed by nucleation, elongation and breakage in cases 1 to 3, and by breakage only in cases 4 and 5, as pre-formed fibrils were employed in the latter cases.

<table>
<thead>
<tr>
<th>Case</th>
<th>Protein</th>
<th>Temp. [°C]</th>
<th>Agitation [rpm]</th>
<th>pH</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>β-lactoglobulin</td>
<td>80</td>
<td>0</td>
<td>2.0</td>
<td>Arosio et al. [68]</td>
</tr>
<tr>
<td>2</td>
<td>insulin</td>
<td>60</td>
<td>250</td>
<td>2.0</td>
<td>This work</td>
</tr>
<tr>
<td>3</td>
<td>insulin</td>
<td>60</td>
<td>50</td>
<td>2.0</td>
<td>This work</td>
</tr>
<tr>
<td>4</td>
<td>insulin</td>
<td>25</td>
<td>250</td>
<td>2.0</td>
<td>This work</td>
</tr>
<tr>
<td>5</td>
<td>β2-microglobulin</td>
<td>25</td>
<td>1000</td>
<td>2.0</td>
<td>Xue &amp; Radford [88]</td>
</tr>
</tbody>
</table>

Table 3.1: Data sets and respective conditions considered in this study.
The increase in fibril content with time was monitored by ThT fluorescence, the monomer depletion kinetics by SEC, whereas the time evolution of the fibril length distribution was followed by AFM.

The results are shown in Figures 3.4 and 3.5 for β-lactoglobulin (case 1), and insulin at high temperature and high agitation speed (case 2), respectively.

![β-lactoglobulin (case 1)](image)

**Figure 3.4:** Comparison between experimental results [68] (symbols or histograms) and model simulations (solid lines) for β-lactoglobulin (case 1 in Table 3.1) in terms of (a) ThT fluorescence, (b) fibril average length, (c-f) fibril length distributions at several time points obtained by AFM.

As the results of case 3 (i.e. insulin at high temperature and lower agitation speed) are qualitatively comparable with those shown in Figure 3.5, the full fibril distributions are not shown here. Nevertheless, the monomer depletion kinetics and the time evolution of the average length are presented in the Discussion section.

In Figure 3.6 and Figure 3.7 instead, the results for insulin at room temperature (case 4) and for β2-microglobulin (case 5) are shown, respectively.

It is interesting to observe that in the cases where both fibril formation and breakage occur (i.e. in the first three cases of Table 3.1), the fibril average length exhibits a maximum as a function of time (cf. Figures 3.4(b), and 3.5(c)).
Figure 3.5: Comparison between experimental results (symbols or histograms) and model simulations (solid lines) for insulin at 60 °C and 250 rpm (case 2 in Table 3.1) in terms of (a) ThT fluorescence, (b) monomer content measured by SEC, (c) fibril average length, (d-f) fibril length distributions at several time points obtained by AFM.

A first phase, fibril elongation dominates over fibril breakage, thus leading to an increase in the average fibril length. In a second phase, when most of the monomer is consumed (as seen from ThT fluorescence in Figures 3.4(a) and 3.5(a), as well as from SEC data in Figure 3.5(b)), fibril breakage dominates, leading to a decrease in the average fibril length with time. No such maximum is found in the experiments corresponding to cases 4 and 5, where breakage was studied at low temperature with pre-formed fibrils. In these conditions, fibril elongation is negligible, and a monotonic decrease in the average length is observed (Figure 3.6(a)).

Notably, when comparing the shapes of the measured fibril distributions, it turns out that in all cases under investigation, the distributions are right-tailed. In other words, a predominance of shorter fibrils, as compared to longer ones, is observed. However, the level of asymmetry varies from one case to the other. Indeed, in the cases of β-lactoglobulin (case 1 - Figure 3.4) and insulin at high temperature (case 2 - Figure 3.5), the distributions are strongly skewed to the right. In contrast, the level of asymmetry is much lower in the case of β2-microglobulin (case 5 - Figure 3.7), where the distributions exhibit a maximum, resembling...
Figure 3.6: Comparison between experimental results (symbols or histograms) and model simulations (solid lines) for insulin at 25 °C and 250 rpm (case 4 in Table 3.1) in terms of (a) fibril average length, (b-f) fibril length distributions at several time points obtained by AFM.

almost bell-shaped curves. Such differently shaped distributions are expected to arise as a result of distinct fragmentation mechanisms.

With a view to determining the fibril breakage mechanisms, we then analyzed the five sets of experimental data in the frame of a kinetic model. To do so, we fitted the model simulations to the experimental results by minimizing the residual sum of squares (RSS) as defined in Equation 3.11. In addition to the parameters quantifying fibril formation, the description of fibril breakage requires to determine three parameters ($k_{F_1}$, $k_{F_2}$ and $k_{F_3}$), as shown in Equation 3.6. While $k_{F_1}$ and $k_{F_2}$ characterize the dependence of the breakage rate constant on the fibril length (cf. Equation 3.7), $k_{F_3}$ is related to the preferential position of breakage along the fibril axis (cf. Equation 3.8). In particular, a positive value of $k_{F_3}$ indicates an erosion mechanism, while a negative value of $k_{F_3}$ indicates a central breakage mechanism. The estimated values of the kinetic parameters can be found in Table 3.2.

The breakage kernels estimated for the five cases listed in Table 3.1 are presented in Figure 3.8(a-c) in terms of $\beta(x)$ and $\Gamma(x, y)$, as defined in Equations 3.7 and 3.8, respectively. While $\beta(x)$ quantifies the dependence of the breakage rate
constant on the fibril length, the daughter distribution \( \Gamma(x, y) \) describes the preferential points of breakage along the fibril length.

Interestingly, it can be observed in Figure 3.8(a-b) that the breakage rate constant strongly increases with the fibril size in all cases investigated in this study.

The main objective of this work was to determine the impact of operating parameters on the type of fibril breakage mechanism (erosion, random or central),

<table>
<thead>
<tr>
<th>Case</th>
<th>( \text{n}_C )</th>
<th>( k_N )</th>
<th>( k_E )</th>
<th>( k_{F1} )</th>
<th>( k_{F2} )</th>
<th>( k_{F3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>( 1.1 \times 10^{-67} )</td>
<td>( 5.0 \times 10^{-21} )</td>
<td>( 2.3 \times 10^{-16} )</td>
<td>3</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>( 2.0 \times 10^{-9} )</td>
<td>( 6.0 \times 10^{-22} )</td>
<td>( 1.0 \times 10^{-25} )</td>
<td>5</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>( 3.0 \times 10^{-10} )</td>
<td>( 4.0 \times 10^{-22} )</td>
<td>( 2.5 \times 10^{-26} )</td>
<td>5</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>( 2.0 \times 10^{-14} )</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>( 3.3 \times 10^{-11} )</td>
<td>1.7</td>
<td>–1</td>
</tr>
</tbody>
</table>

Table 3.2: Parameter values for the five investigated systems. For case 1, the values of the parameters \( \text{n}_C \), \( k_N \), and \( k_E \), and two additional describing protein hydrolysis were taken from reference [68]. The breakage parameters estimated for case 5 are consistent with literature values [88].
Figure 3.8: (a) Length dependence of the breakage rate constant for insulin (cases 2 to 4). (b) Length dependence of the breakage rate constant for β-lactoglobulin (case 1) and β2-microglobulin (case 5). (c) Daughter distributions for the 5 cases. (d) Relative minimal RSS as a function of the parameter $k_{F3}$, i.e. best $RSS$ value versus $k_{F3}$ obtained by leaving the other kinetic parameters free. The results are normalized by the $RSS$ value of the global minimum. The definition of $RSS$ is given in Equation 3.11.

which is captured here by the sign of the parameter $k_{F3}$. In Figure 3.8(d), the relative minimal RSS is plotted as a function of $k_{F3}$. In other words, for each $k_{F3}$, the best $RSS$ value obtained by varying all the other parameters is considered. To allow a better comparison between the different cases, the $RSS$ values are scaled by their minimum.
In the case of experiments performed at high temperature (cases 1 to 3), the best results are obtained for large positive values of $k_{F_3}$. It is worth mentioning that given the breakage kernel employed in this study, the breakage rate constant becomes independent of $k_{F_3}$ for large $k_{F_3}$ values, resulting in the observed plateau in the plot of the RSS (cf. Figure 3.8(d)).

In contrast, clear minima are obtained in cases 4 and 5, corresponding to $k_{F_3}$ equal to +1 and −1, respectively. The different values of $k_{F_3}$ result in distinct shapes of the daughter distribution function, as shown in Figure 3.8(c). In particular, it can be observed that fibril breakage occurs preferentially at the fibril end for all the cases under investigation, except for $\beta$2-microglobulin (case 5) where fibrils are more likely to be broken at their center.

Model simulations are compared to experimental data in Figures 3.4, 3.5, 3.6, and 3.7. In all cases, it can be seen that the agreement between model simulations and experimental results is satisfactory.

### 3.4 Discussion

We investigated the impact of operating parameters on the breakage mechanisms of insulin fibrils by combining experimental data from AFM measurements with model simulations based on population balance equations. The effect of temperature and agitation rate on the kinetics of monomer depletion and on the time evolution of the average length of insulin fibrils can be appreciated in Figure 3.9 (cases 2 to 4).

#### 3.4.1 Impact of agitation

Considering first the data obtained at 60 °C (cases 2 and 3), where both fibril growth and fibril fragmentation occur, it is seen that an increase in the agitation rate induces an slight acceleration of the monomer depletion (Figure 3.9(a)) and a decrease in the height of the maximum average fibril length (Figure 3.9(b)).
The latter observation indicates that mechanical agitation accelerates the kinetics of breakage with respect to the kinetics of fibril elongation, as further evidenced by kinetic analysis. Indeed, a rise in the agitation rate from 50 to 250 rpm was found to cause an increase in the breakage constant \( (k_{B_1}) \) by 400 %, whereas the elongation rate constant was increased by 50 % only (cf. Table 3.2). The method proposed in this study allows unravelling the dependence of the breakage rate constant both on the fibril length \( \beta(x) \), and on the fibril position \( \Gamma(x,y) \). Interestingly, it is found that an increase in the agitation rate leads to higher values of \( \beta(x) \) without affecting \( \Gamma(x,y) \), i.e. the observed erosion mechanism is unchanged upon a variation in the agitation rate (cf. Figure 3.8(a,c)). This means that mechanical agitation accelerates the kinetics of breakage without impacting the preferential position of fibril fragmentation, at least in the range of agitation rates considered.

### 3.4.2 Impact of temperature

In order to assess the impact of temperature on the fragmentation of insulin fibrils, an experiment was performed at 25 °C and 250 rpm agitation on fibrils produced beforehand at high temperature (case 4). It is seen in Figure 3.9 that lowering the temperature from 60 °C to 25 °C at constant agitation speed almost completely
suppressed fibril elongation (Figure 3.9(a)), and significantly slowed down fibril fragmentation (Figure 3.9(b)).

Following the same analysis conducted to study the impact of agitation on fibril breakage, we investigated the effect of temperature on the two contributions of the breakage kernel, i.e. \( \beta(x) \) and \( \Gamma(x,y) \), as shown in Figure 3.8(a) and Figure 3.8(c), respectively. It is observed that the fibril breakage rate \( \beta(x) \) is dramatically affected by temperature, indicating that thermal breakage plays a key role in the fibril fragmentation process. In fact, when the thermal energy \( k_B T \) of the system is larger than the energy linking two monomeric proteins together, the hydrogen bonding network is ruptured, leading to fibril fragmentation.

Moreover, it is observed that besides strongly reducing the magnitude of fibril breakage (Figure 3.8(a)), a decrease in temperature leads to a change in the daughter distribution function (Figure 3.8(c)). Indeed, the preferential breakage at the fibril ends observed under heat stress is much weaker at room temperature. This suggests that a decrease in temperature shifts the mechanism of fibril breakage from an erosion towards a random breakage mechanism, implying that insulin fibrils may break in different ways according to the environmental conditions they are exposed to.

### 3.4.3 Comparison with literature data

In order to extend our analysis to other protein systems, we analyzed literature data obtained with \( \beta \)-lactoglobulin (case 1), whose kinetics of fibril formation and breakage was followed at high temperature (80 °C) under stagnant conditions [68]. The corresponding breakage kernel was identified in this study by comparing model simulations with detailed experimental data and is presented in Figure 3.8(b,c). It is remarkable that, just as in the case of insulin fibrils, the breakage of \( \beta \)-lactoglobulin fibrils at high temperature occurs preferentially at fibril ends. This suggests that elevated temperatures favor fibril fragmentation by an erosion mechanism, regardless of the applied agitation.
To clarify the role of fibril breakage at lower temperature under mechanical agitation, we compared the results obtained with insulin (case 4 - 250 rpm agitation) to the literature data of β2-microglobulin (case 5 - 1000 rpm agitation) [88]. Notably, the two systems exhibit different breakage mechanisms despite the same temperature conditions were employed. This indicates that mechanical aspects (either related to the agitation rate or to the fibrils properties) play a decisive role in determining the breakage mechanism at low temperature. The fragmentation of insulin fibrils was found to occur by a weak erosion mechanism (close to a random breakage mechanism), while a central breakage mechanism was identified in the case of β2-microglobulin fibrils.

With a view to better understanding these differences, we analyzed qualitatively the morphology of the protein filaments. In Figure 3.10, representative AFM pictures taken at several incubation times are presented for the five cases under examination.

**Figure 3.10:** Representative AFM pictures for the five cases of Table 3.1, showing that β2-microglobulin fibrils (case 5) are straight, as compared to the fibrils produced in the other conditions, which appear curved.

Notably, β2-microglobulin fibrils appear very straight, implying that their contour length is almost equal to their end-to-end distance. As a result, the persistence length of β2-microglobulin fibrils is much larger than their contour length, indicating that β2-microglobulin fibrils are stiff [17]. It is worth pointing out that such stiffness is not related to specific intrinsic properties of β2-microglobulin, but rather to the selected environmental conditions, since this protein has been shown to form various types of fibrils (including flexible fibrils) depending on the operating parameters (e.g. varying pH, salt type, salt concentration) [147].
Insulin fibrils are instead more curved and thus more flexible, as can be inferred from Figure 3.10. This suggests that the distinct breakage mechanisms observed between insulin and β2-microglobulin fibrils under similar experimental conditions might be attributed to the difference in fibril stiffness. The fact that stiff β2-microglobulin fibrils are found to preferentially break in their center is in agreement with the theory of Hill developed for the fragmentation of stiff rods [148]. Nevertheless, more experimental data, both in terms of protein type and agitation conditions, would be necessary to confirm the role of fibril stiffness in the preferential position of breakage.

The results of this study are summarized in the schematic diagram of Figure 3.11, where it appears that there is no universal fragmentation mechanism of amyloid fibrils. Indeed, depending on the system under investigation, either a central, random or erosion breakage mechanism is observed. In particular, it is seen from the insulin data obtained at different temperatures that one given protein system can undergo different fibril breakage mechanisms depending on the environmental conditions the fibrils are exposed to. The observation that amyloid fibrils can undergo different breakage mechanisms suggests that the intermolecular forces underlying the aggregation process are strongly affected by the environmental parameters and/or by the protein type.

![Figure 3.11: Summary of the fibril breakage mechanisms observed for the five systems under investigation.](image)

In this frame, Brownian dynamics simulations represent a powerful tool to complement experimental characterization and population balance equations modeling. Brownian simulations at the molecular level indeed open the possibility to...
gain fundamental insights into the connection between the fibril breakage mechanism and the type and strength of intermolecular bonds [149].

3.5 Conclusion

We investigated the impact of operational parameters on the breakage mechanism of insulin fibrils by comparing the time evolution of the fibril length distribution followed by AFM with PBE simulations. Following the approach used in colloid science, we used the concept of daughter distributions to describe the probability of fibril fragmentation along the fibril longitudinal axis.

Our results show that temperature is a key parameter regulating both the rate of fibril breakage, as well as the preferential position of fragmentation along the fibril axis. Under thermal stress, significant fibril breakage occurs and an erosion mechanism is favored, regardless of the agitation applied. At lower temperatures instead, the magnitude of fibril breakage is significantly reduced and mechanical aspects (such as the agitation rate or the fibril stiffness) become prominent, resulting in a random or a central breakage mechanism.

To conclude, different fibril fragmentation mechanisms occur depending on the environmental parameters. These findings may be of particular interest for the production of amyloid-based biomaterials, as they show the possibility to tune the fibril distribution by controlling operative parameters.

The content of this chapter has been partially published in the following article: L. Nicoud, S. Lazzari, D. Balderas, Barragán, and M. Morbidelli. Fragmentation of amyloid fibrils occurs in preferential positions depending on the environmental positions. J. Phys. Chem. B, 119(13):4644-4652, 2015
Chapter 4

Role of cosolutes in the kinetics of mAb aggregation

4.1 Introduction

In this chapter, we aim at gaining knowledge on the impact of cosolutes on the aggregation kinetics of monoclonal antibodies. Improving our understanding of the impact of cosolutes on protein stability is indeed of fundamental relevance in the bioprocessing of therapeutic proteins since excipients such as salts and sugars are commonly used in drug formulation to provide physiological osmolality or hinder protein aggregation [9, 11, 12]. Accelerated studies at high temperature are widely performed to screen a large amount of formulation buffers in an attempt to optimize protein stability under storage conditions. Nevertheless, the proper extrapolation of stability studies from high to low temperature requires the mechanistic understanding of the impact of cosolutes on the aggregation process, which is challenging to achieve.

Indeed, as highlighted in Chapter 1, protein aggregation is a complex multistep process which involves several elementary steps such as protein unfolding, nucleation and aggregate growth. In particular, two contributions can be distinguished in the aggregation propensity of a protein: (i) the protein colloidal stability, i.e.
the energy barrier that two protein molecules need to overcome upon collision in order to aggregate; (ii) the protein conformational stability, i.e. the energy barrier preventing protein molecules from unfolding [10, 13]. These two contributions are highly sensitive to the presence of cosolutes since they are regulated by the delicate balance between inter- and intramolecular protein interactions, as well as protein-solvent interactions, which are strongly affected by the solution composition. Moreover, the protein colloidal stability and the protein conformational stability are strongly interconnected, since changes in the protein structure lead to changes in the intermolecular interactions, making these two contributions extremely difficult to analyze independently experimentally.

In this chapter, we build on the power of chemical kinetics analysis to investigate the effect of cosolutes on the single elementary steps which contribute to the global aggregation mechanism. Chemical kinetics is indeed increasingly emerging as a key tool to gain insights into the mechanisms of protein aggregation from macroscopic measurements [62, 63, 116]. We illustrate this general procedure by considering the effect of NaCl and sorbitol on the aggregation behavior of two monoclonal antibodies, whose aggregation mechanisms in the absence of cosolute have been indentified in Chapter 2.

\section*{4.2 Materials and methods}

The two model monoclonal antibodies used in this study are the same as used in Chapter 2, i.e. a glycosylated IgG1 and a non-glycosylated IgG2. In the following, they will be referred to as mAb-1 and mAb-2, respectively, in agreement with the notation used in Chapter 2.

The isothermal aggregation kinetic experiments were performed on dialyzed samples at the temperature of 70 °C and at the protein concentration of 1 g/L. The pH was set to 6.5 by using a 20 mM histidine buffer. Details about the sample preparation and experimental characterization are reported in Chapter 2.
4.3 Results

4.3.1 Impact of cosolutes on aggregation thermal stability, aggregate morphology and electrostatic interactions

We first evaluated the impact of NaCl and sorbitol on the aggregation thermal stability (i.e. the aggregation propensity upon temperature increase) of the antibody solutions by recording the increase in the solution scattered intensity as a function of temperature with DLS. An increase in the solution scattered intensity is indeed an indication of the increase in size and/or number of aggregates. Figure 4.1(a) shows the DLS intensity as a function of temperature for mAb-1 in the absence and in the presence of NaCl and sorbitol. The results show that NaCl and sorbitol have a destabilizing and a stabilizing effect, respectively, on the aggregation propensity of mAb-1. Similar results were obtained in the case of mAb-2 (data not shown).

The aggregate morphology was then investigated by SLS, as shown in Figure 4.1(b) for mAb-1. The aggregates were found to exhibit fractal geometry, and a fractal dimension value could be estimated from the power-law regime of the plot of the structure factor \( S(q) \) as a function of the \( q \) vector. The results indicate that the presence of NaCl leads to an increase in the fractal dimension, i.e. salt leads to the formation of more compact aggregates. One possible explanation is that the presence of salt makes the protein surface more uniform due to screening of heterogeneous charges. The value of \( d_f = 2.05 \) in the presence of 10 mM of salt is indeed consistent with fractal dimension values usually reported in the case of uniformly charged spheres aggregating under reaction limited conditions, while the lower value of \( d_f = 1.85 \) in the absence of salt may indicate some patchiness of the protein surface, as proposed by Wu et al. [150]. An alternative explanation would be that the presence of NaCl screens the repulsive energy barrier between protein molecules, thus favoring cluster interpenetration. Unfortunately, the impact of salt on the fractal dimension value could not be investigated in the case of mAb-2 due to the precipitation of large aggregates occurring during the cooling
Figure 4.1: Impact of NaCl and sorbitol on mAb-1 (a) aggregation thermal stability and (b) aggregate fractal dimension. Impact of (c) NaCl and (d) sorbitol on the zeta potential values of mAb-1 and mAb-2.

required by the off-line scattering experiments. The presence of sorbitol, instead, does not impact the aggregate fractal dimension, neither in the case of mAb-1 ($d_f = 1.85$, Figure 4.1(b)), nor in the case of mAb-2 ($d_f = 2.05$, data not shown).

The impact of cosolutes on the effective protein charge was then assessed by zeta potential measurements. Figure 4.1(c) and Figure 4.1(d) show the protein zeta potential as a function of NaCl and sorbitol concentration, respectively, for both antibodies. The zeta potential measurements quantify the electric potential
at the double layer, which is related to the protein surface charge and the non-diffusive counterions located in the layer around the protein. Charged cosolutes may affect the zeta potential value by binding to the opposite charged groups on the surface of the protein and by modifying the electrostatic interactions in the double layer. By adding salt, the zeta potential of mAb-1 decreases from 5.7 mV to a value close to zero (Figure 4.1(c)), while the zeta potential value of mAb-2 is already close to zero in the absence of salt, and is therefore less sensitive to the addition of NaCl. It can be observed in Figure 4.1(d) that the presence of sorbitol does not significantly affect the zeta potential values of both antibodies.

It is interesting to observe that changes in aggregate structure correlate with changes in electrostatic interactions: the presence of salt decreases the net repulsive electrostatic interactions and induces the formation of more compact aggregates characterized by a larger fractal dimension value. On the other hand, the presence of sorbitol does not affect the electrostatic interactions and does not impact aggregate morphology (Figure 4.1).

4.3.2 Impact of cosolutes on the mAb aggregation kinetics

After these preliminary experiments, we analyze the impact of cosolutes on the elementary processes of the aggregation mechanisms described in Chapter 2. For both antibodies, the monomer depletion, dimer and trimer formation as well as the aggregate weight-average molecular weight were followed by SEC coupled with an inline MALS detector. The experimental data were then compared with model simulations to extract information on the impact of cosolutes on the elementary reaction rates. The obtained results are discussed in the following, with reference first to mAb-1, and then to mAb-2.

4.3.2.1 mAb-1 - Impact of NaCl

Figure 4.2 shows the experimental results on the kinetics of aggregation of mAb-1 in the absence of salt as well as in the presence of 5 mM and 10 mM NaCl.
Figure 4.2: Comparison between simulated (lines) and experimental (symbols) aggregation kinetics of mAb-1 incubated in the absence, and presence of 5 mM and of 10 mM NaCl. (a) Monomer, (b) dimer, (c) trimer concentrations, and (d) aggregate weight-average molecular weight versus time, (e) aggregate weight-average molecular weight versus monomer conversion, defined as the non-aggregated fraction of the initial protein content. Parameter values used in the simulations are summarized in Table 4.1.

It can be observed that the salt significantly accelerates the increase in aggregate molecular weight during time (Figure 4.2(d)), while it has only a slight impact on the monomer depletion kinetics (Figure 4.2(a)). In Figure 4.2(e), the average aggregate molecular weight is shown as a function of monomer conversion. It can be seen that at a given monomer conversion, the presence of NaCl induces the formation of larger aggregates, with respect to the situation without salt. This experimental observation suggests that the presence of salt decreases the characteristic time of aggregate growth, while it does not change the characteristic time of monomer depletion, which has been shown in Chapter 2 to be rate-limited by monomeric conformational changes.

Therefore, the analysis of the impact of NaCl on the aggregation kinetics of
mAb-1 reveals that the unfolding rate constant is not affected by the presence of salt, while the aggregation rate constants increase upon salt addition, indicating a decrease of the net repulsive intermolecular forces, in agreement with the changes in zeta potential and aggregate morphology described in the previous paragraph (Figure 4.1).

We can quantify this behavior with the kinetic model developed in Chapter 2, which considers several Fuchs stability ratios to account for the different reactivities of the two populations present in solution, i.e. the unfolded monomer, which is a highly reactive aggregation-prone intermediate, and the aggregates. Accordingly, \( W_{11}, W_{1j}, W_{ij} \) have been introduced to describe dimer formation, monomer addition and cluster-cluster aggregation, respectively.

In order to simplify the fitting procedure, we assumed that salt has the same impact on all the aggregation rate constants. Thus, we introduced a parameter \( \alpha_1 \), which was defined as the ratio between each aggregation rate constant in the presence of cosolute divided by the corresponding value in the absence of cosolute. Since the aggregation rate constants are inversely proportional to the Fuchs stability ratios, \( \alpha_1 \) is defined as:

\[
\alpha_1 = \frac{W_{11}^{-CS}}{W_{11}^{+CS}} = \frac{W_{1j}^{-CS}}{W_{1j}^{+CS}} = \frac{W_{ij}^{-CS}}{W_{ij}^{+CS}} \tag{4.1}
\]

where the superscripts \(-CS\) and \(+CS\) denote the situations without and with cosolute, respectively. Parameter values in the absence of cosolute have been determined in Chapter 2.

At each salt concentration, the parameter \( \alpha_1 \) was evaluated by fitting the model simulations to the experimental data. Regarding the other parameters, the fractal dimension was measured by independent SLS experiments and the power law factor \( \lambda \) was estimated as \( \lambda \approx 1 - 1/d_f \). The unfolding rate constant \( k_U \) was assumed to be unchanged as compared to the situation without cosolute. The parameter values used for the simulations are summarized in Table 4.1. In particular, we found \( \alpha_1 = 3.5 \) in the presence of 5 mM NaCl and and \( \alpha_1 = 7 \) in
the presence of 10 mM NaCl. These values correspond to a decrease in the Fuchs stability ratio by a factor of 3.5 and 7, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$d_f$</th>
<th>$\lambda$</th>
<th>$k_U$ (s$^{-1}$)</th>
<th>$\alpha_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 mM NaCl</td>
<td>1.94</td>
<td>0.48</td>
<td>$10 \times 10^{-4}$</td>
<td>3.5</td>
</tr>
<tr>
<td>10 mM NaCl</td>
<td>2.05</td>
<td>0.51</td>
<td>$10 \times 10^{-4}$</td>
<td>7</td>
</tr>
<tr>
<td>250 mM sorbitol</td>
<td>1.85</td>
<td>0.46</td>
<td>$6.7 \times 10^{-4}$</td>
<td>1</td>
</tr>
<tr>
<td>500 mM sorbitol</td>
<td>1.85</td>
<td>0.46</td>
<td>$3.5 \times 10^{-4}$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.1: Parameters used for the simulations of mAb-1 aggregation in the presence of cosolutes. The parameter $\alpha_1$ is defined in Equation 4.1.

In Figure 4.2, model simulations are compared to experimental data. It can be seen that the model is capable of describing well the time evolution of the different sets of experimental data. In particular, it is worth noticing that the slight acceleration in the monomer depletion kinetics in the presence of salt (Figure 4.2(a)) can be explained by a decrease in the value of the Fuchs stability only (i.e. without changes in the unfolding rate constant), as proven by model simulations. Moreover, the fact that one single fitting parameter $\alpha_1$ is capable of describing the decrease in protein stability upon salt addition suggests that NaCl affects the reactivity of the reactive monomer and of the clusters to the same extent.

In summary, our results show that, in the investigated range of salt concentration, the presence of NaCl increases the aggregation kinetics of mAb-1 only by promoting the aggregation events, i.e. by reducing the repulsive energy barrier between two colliding particles, without impacting the kinetics of formation of the aggregation prone monomer, i.e. the unfolding step.

### 4.3.2.2 mAb-1 - Impact of sorbitol

Figure 4.3 shows experimental kinetics data of mAb-1 in the absence of polyol sugar and in the presence of 250 mM and 500 mM sorbitol.

It can be seen that sorbitol significantly delays the monomer depletion while it has almost no impact on the increase in the aggregate molecular weight. Therefore,
Chapter 4. Impact of cosolutes on mAb aggregation

Figure 4.3: Comparison between simulated (lines) and experimental (symbols) aggregation kinetics of mAb-1 incubated in the absence, and presence of 250 mM and of 500 mM sorbitol. (a) Monomer, (b) dimer, (c) trimer concentrations, and (d) aggregate weight-average molecular weight versus time, (e) aggregate weight-average molecular weight versus monomer conversion, defined as the non-aggregated fraction of the initial protein content. Parameter values used in the simulations are summarized in Table 4.1.

At a given monomer conversion, the presence of sorbitol induces the formation of larger aggregates compared to the situation without polyol sugar, as can be seen in Figure 4.3(e). This result suggests that the presence of sorbitol increases the characteristic time of monomer depletion, which corresponds to the characteristic time of protein unfolding, to a larger extent than the characteristic time of aggregate growth. Moreover, it was reported that sorbitol neither impacts the protein net charge nor the aggregate morphology of mAb-1 (Figure 4.1), suggesting that sorbitol does not impact protein reactivity.

In agreement with these experimental observations, the values of the Fuchs stability ratios estimated in the absence of cosolute were considered in the model simulations. Moreover, the fractal dimension was measured by independent SLS
experiments and the power law factor $\lambda$ was estimated as $\lambda \approx 1 - 1/d_f$. The only fitting parameter was therefore the unfolding rate constant $k_U$. The parameters used for the simulations are summarized in Table 4.1.

It can be seen in Figure 4.3. that the results of the model simulations are in very close agreement with the experimental data. In particular, it is worth noticing that the slight delay observed in the kinetics of aggregate growth in the presence of sorbitol (Figure 4.3 (d)) can be explained by a decrease in the unfolding rate constant only (i.e. without changes in the Fuchs stability ratio), as proven by model simulations.

This analysis reveals that sorbitol delays the kinetics of mAb-1 aggregation only by delaying the kinetics of formation of the aggregation-prone monomer without impacting protein reactivity, i.e. sorbitol slows down protein unfolding, while it has a negligible impact on the energy barrier that colliding molecules must overcome in order to aggregate.

Sorbitol is known to favor a compact native conformation of the protein, which exposes a small surface area to the solvent, with respect to more open unfolded structures. Sorbitol has indeed been shown to impart thermodynamic stability to protein solutions due to a preferential exclusion effect [151, 152]. Here, we show that sorbitol also increases the kinetic stability of the mAb solution under investigation by delaying the kinetics of formation of the partially unfolded aggregation-prone monomer.

4.3.2.3 mAb-2 - Impact of NaCl

The same approach used for mAb-1 was applied to investigate the impact of cosolutes on the aggregation mechanism of mAb-2. Figure 4.4 shows the experimental data of mAb-2 aggregation in the absence of salt as well as in the presence of 10 mM and 50 mM NaCl.

Larger NaCl concentrations were used for mAb-2 with respect to mAb-1 in order to observe an appreciable impact of the salt on the aggregation kinetics.
The lower zeta potential value of mAb-2 with respect to mAb-1 indeed suggests that the protein net charge is lower. Therefore, mAb-2 molecules experience less electrostatic repulsion and the charge screening by NaCl addition might thus have less impact on the aggregation kinetics of mAb-2.

In Figure 4.4, it can be observed that NaCl accelerates both the kinetics of monomer depletion and of aggregate growth. In addition, at a given monomer conversion, the aggregate weight-average molecular weight (Figure 4.4(e)), the dimer and trimer concentrations (data not shown) are independent of salt concentration. This important observation indicates that the aggregate distribution obtained at a given monomer conversion is independent of salt concentration. Since specific promotion of nucleation or growth events would affect in a different way the aggregate
distribution at a given monomer conversion, the analysis provides evidence that NaCl does not impact specific steps of the reaction scheme, but rather accelerates all the elementary processes of the aggregation process, i.e. nucleation and growth to a similar extent.

Based on these considerations, we introduced a parameter $\alpha_2$ defined as the ratio of the value of each kinetic rate constant (corresponding either to an aggregation, dissociation or nucleation event) in the presence of cosolute divided by the corresponding value in the absence of cosolute (which are reported in Table 4.2). Since the aggregation rate constants are inversely proportional to the Fuchs stability ratios, $\alpha_2$ is defined as:

$$
\alpha_2 = \frac{W_{11}^{-CS}}{W_{11}^{+CS}} = \frac{W_{4j}^{-CS}}{W_{4j}^{+CS}} = \frac{W_{ij}^{-CS}}{W_{ij}^{+CS}} = \frac{k_{r11}^{+CS}}{k_{r11}^{-CS}} = \frac{k_{r12}^{+CS}}{k_{r12}^{-CS}} = \frac{k_{r13}^{+CS}}{k_{r13}^{-CS}} = \frac{k_{nuc}^{+CS}}{k_{nuc}^{-CS}}
$$

(4.2)

where $-CS$ and $+CS$ denote the situations without and with cosolute, respectively.

At each salt concentration, the parameter $\alpha_2$ was evaluated by fitting model simulations to the set of experimental data. Due to sample precipitation at large aggregate sizes, the fractal dimension of mAb-2 aggregates in the presence of salt could not be measured experimentally, and was thus assumed to be the same as the one estimated in the absence of salt. The parameter values used for the simulations are summarized in Table 4.2. In particular, we obtained $\alpha_2 = 1.2$ and $\alpha_2 = 1.6$ at 10 mM and 50 mM NaCl, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$d_f$</th>
<th>$\lambda$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM NaCl</td>
<td>2.05</td>
<td>0.51</td>
<td>1.20</td>
</tr>
<tr>
<td>50 mM NaCl</td>
<td>2.05</td>
<td>0.51</td>
<td>1.60</td>
</tr>
<tr>
<td>250 mM sorbitol</td>
<td>2.05</td>
<td>0.51</td>
<td>0.60</td>
</tr>
<tr>
<td>500 mM sorbitol</td>
<td>2.05</td>
<td>0.51</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 4.2: Parameters used for the simulations of mAb-2 aggregation in the presence of cosolutes. The parameter $\alpha_2$ is defined in Equation 4.2.
As shown in Figure 4.4, the model simulations are in very good agreement with the experimental results. Therefore, the kinetic analysis confirms that NaCl accelerates (in a concentration dependent manner) all the elementary steps involved in the aggregation scheme to the same extent.

4.3.2.4 mAb-2 - Impact of sorbitol

Finally, the impact of sorbitol on the aggregation kinetics of mAb-2 was investigated. Experimental data are presented in Figure 4.5, showing that sorbitol is capable of delaying both the monomer depletion and the aggregate growth kinetics.

![Figure 4.5](image)

**Figure 4.5:** Comparison between simulated (lines) and experimental (symbols) aggregation kinetics of mAb-2 incubated in the absence, and presence of 250 mM and of 500 mM sorbitol. (a) Monomer, (b) dimer, (c) trimer concentrations, and (d) aggregate weight-average molecular weight versus time, (e) aggregate weight-average molecular weight versus monomer conversion, defined as the non-aggregated fraction of the initial protein content. Parameter values used in the simulations are summarized in Table 4.2.
As in the case of NaCl, the aggregate weight-average molecular weight (Figure 4.5(e)) as well as the dimer and trimer concentrations (data not shown) obtained at a given monomer conversion are not affected by the presence of sorbitol. This indicates that the presence of sorbitol does not affect the aggregate distribution at a given monomer conversion, suggesting that sorbitol does not impact specific steps of the reaction scheme, but rather delays all the elementary steps of the aggregation process to the same extent.

Similarly to the approach used in the case of NaCl, we performed the simulations of the inhibited aggregation kinetics in the presence of sorbitol by considering the set of kinetic rate constants evaluated in the absence of sorbitol multiplied by a constant factor $\alpha_2$, which was determined by fitting to the experimental data. Moreover, the fractal dimension was measured by SLS and $\lambda$ was estimated by $\lambda \approx 1 - 1/d_f$. The parameters used for the simulations are summarized in Table 4.2. In particular, we obtained $\alpha_2 = 0.6$ and $\alpha_2 = 0.35$ at 250 mM and 500 mM sorbitol, respectively.

As shown in Figure 4.5, the model simulations describe very well the set of experimental data, thus confirming that sorbitol delays (in a concentration dependent manner) all the elementary steps that contribute to the global aggregation process to the same extent.

4.4 Discussion

We investigated the impact of NaCl and sorbitol on the aggregation kinetics of two monoclonal antibodies under thermal stress. The two cosolutes were selected for their relevance in drug formulation and for their opposite effects on protein stability: while salt addition accelerates the kinetics of protein aggregation, the presence of polyol sugar delays the aggregation process.

By applying kinetic analysis, we could discriminate the impact of the two cosolutes on the elementary steps that contribute to the global protein aggregation
Figure 4.6: Scheme of the aggregation mechanism of (a) mAb-1 and (b) mAb-2 highlighting the impact of NaCl and sorbitol on the elementary steps. In the case of mAb-1, NaCl accelerates the aggregation process without impacting the unfolding step, while sorbitol inhibits protein unfolding without affecting the aggregation process. In the case of mAb-2, NaCl accelerates and sorbitol inhibits all the elementary steps to the same extent.

rate. Our results are summarized in Figure 4.6, which presents a schematic drawing of the kinetic mechanisms of the two antibodies under investigation, and highlights the impact of NaCl and sorbitol on the elementary events.

The impact of NaCl and sorbitol on the aggregation kinetics of mAb-1 is further illustrated with the energy diagram shown in Figure 4.7(a), where the global aggregation pathway is schematically represented as an unimolecular unfolding event followed by a bimolecular collision step leading to the formation of aggregates. By applying kinetic analysis, it was found that, in the investigated range of cosolute concentrations, NaCl promotes aggregate growth without impacting
protein unfolding, while sorbitol delays protein unfolding without affecting aggregate growth. We can therefore conclude that NaCl reduces the activation energy of aggregation without impacting the activation energy of protein unfolding, while sorbitol increases the activation energy of protein unfolding without impacting the activation energy of aggregation.

The decrease in the energy barrier that colliding molecules must overcome before aggregating in the presence of NaCl can be partly attributed to the reduction of electrostatic repulsion induced by salt addition. Since it has been shown in Chapter 2 that the contribution of electrostatics to the colloidal stability is small, it is also likely that the presence of salt modifies the aggregation kinetics of mAb-1 by impacting other intermolecular forces, such as hydration forces. Moreover, it is worth highlighting that the impact of salt on mAb aggregation kinetics was studied in the millimolar range. At higher salt concentrations, additional effects such as specific ion binding may occur, possibly affecting the kinetics of protein unfolding.

**Figure 4.7:** (a) Impact of NaCl and sorbitol on the aggregation pathway of mAb-1. (b) Impact of NaCl and sorbitol on the characteristic times of protein unfolding and dimer formation of mAb-1. The characteristic times of dimer formation were computed at the reference protein concentration of 1 g/L. The dashed black line defines the border between a process which is unfolding rate limited and a process which is aggregation rate limited.
On the other hand, the finding that sorbitol increases the energy barrier of protein unfolding can be related to the preferential exclusion of sugars at the protein surface. Indeed, sugars and polyol sugars are known to impart thermodynamic stability (i.e. increase the Gibbs free energy difference between the folded and unfolded states) to protein solutions due to unfavorable protein-solvent interactions that favor compact protein conformations with respect to more open unfolded structures [151, 152]. In this work, it was shown that the preferential exclusion of sorbitol from the protein surface induces also a kinetic stabilization since sorbitol is found to be capable of delaying protein unfolding.

The impact of cosolutes on the elementary events involved in the aggregation process was also quantified in terms of characteristic times, as shown in Figure 4.7(b). As protein unfolding is a monomolecular event described by a first order kinetic, its characteristic time was simply computed as \( \tau_{\text{Unf}} = 1/k_U \). Dimerization, instead, arises from a bimolecular collision event and is described by a second order kinetics. Its characteristic time at a reference protein concentration \( M_0 \) was thus computed as \( \tau_{\text{Dim}} = 1/(k_{11}M_0) \), where \( k_{11} \) is the aggregation rate constant of dimer formation (computed from \( W_{11} \)). It can be seen that in the absence of cosolutes, the characteristic time of protein unfolding is in the order of 20 min, while the characteristic time of dimer formation is in the order of 1 min (at the protein concentration of 1 g/L), consistent with the previous finding that mAb-1 aggregation is unfolding rate-limited. It is worth noticing that both cosolutes increase the ratio between the characteristic times of unfolding and aggregation (Figure 4.7(b)). This important observation allows us to conclude that the monomer depletion remains rate-limited by protein unfolding upon addition of NaCl or sorbitol. Moreover, it is speculated that the monomer depletion of mAb-1 would become rate-limited by aggregation upon addition of a cosolute which is either increasing the characteristic time of aggregation, or reducing the characteristic time of protein unfolding, such as urea or guanidinium chloride.

While NaCl and sorbitol were shown to impact specific elementary steps in the aggregation mechanism of mAb-1, the addition of cosolutes in the case of mAb-2 was found to impact all the elementary steps (i.e. reversible oligomerization,
oligomer rearrangement and aggregate growth) to the same extent in the investigated range of cosolute concentrations. All the elementary steps that contribute to the global aggregation process were found to be accelerated (respectively delayed) to the same extent upon NaCl (respectively sorbitol) addition.

It is worth stressing that the observed behaviors are specific to the proteins examined, and do not identify the general properties of the respective classes IgG1 and IgG2. Indeed, due to the complex chemical structure of the antibodies, a wide variety of aggregation behaviours has been observed even among members of the same class [84], and robust conclusions about the general properties of the different classes would require the analysis of a large number of molecules.

### 4.5 Conclusion

We investigated the impact of NaCl and sorbitol on the aggregation kinetics of two model antibodies, and IgG1 and an IgG2. By combining experimental characterization with theoretical kinetic analysis we quantified the impact of cosolutes on the elementary events that contribute to the global aggregation rate.

In the case of the IgG1, it was found that NaCl accelerates the kinetics of aggregation by promoting aggregation events without impacting the kinetics of protein unfolding, which is rate-limiting the kinetics of monomer depletion. Conversely, sorbitol is found to delay the kinetics of aggregation by inhibiting specifically protein unfolding, without impacting the aggregation events. In the case of the IgG2, whose monomer consumption kinetics is rate-limited by dimer formation, both cosolutes are found to impact all the elementary steps to the same extent.

In the next chapter, the stabilization effect of mAb-1 by polyols of various sizes will be deepened.

---

Chapter 5

Stabilization of monoclonal antibodies by polyols

5.1 Introduction

In this chapter, we focus on the impact of polyols (including sugars and polyol sugars) on the aggregation kinetics of a model monoclonal antibody. Polyols such as sorbitol are indeed widely used as excipients in drug formulation in order to hinder protein aggregation [9–11].

The impact of such cosolutes on the thermodynamic properties of proteins in aqueous solutions has been extensively investigated. Pioneering work by Timasheff and Arakawa indicated the preferential exclusion of sugars at the protein surface as a major mechanism of action [151, 153–156]. They provided evidence that unfavorable protein-solvent interactions favor native compact protein conformations with respect to more open unfolded structures, as illustrated in Figure 5.1.

Although several works examined the effect of sugars and polyol sugars on the thermodynamics of protein unfolding (i.e. on the difference in the free energy between the native and unfolded states in the absence and presence of polyols), only a few studies were dedicated to the impact of these cosolutes on the kinetics of protein aggregation.
In this chapter, we combine experimental characterization with theoretical modeling to investigate the impact of several polyols (glycerol, threitol, sorbitol, sucrose, maltitol) on the heat-induced aggregation of a model monoclonal antibody, whose monomer depletion kinetics is limited by the rate of protein unfolding, as determined in Chapter 4. Although these conditions do not represent strictly the situation observed during long-term storage, this type of mechanistic study can provide fundamental insights on the effect of stabilizers on protein aggregation, which could be relevant also at lower temperatures.

We propose two mechanistic schemes describing the stabilization effect observed in the presence of polyols. In the first scheme, the mAb stabilization is described by a decrease of the unfolding rate constant, while in the second scheme, the stabilization effect is attributed to a change in the conformational equilibrium constant. The latter scheme involves a shift of the native-state ensemble towards a more compact protein conformation, in line with results proposed in the literature [157–159]. After applying these two schemes to the case of sorbitol, we investigate the effect of polyol size on the mAb stabilization.
5.2 Materials and methods

The mAb used for this study was a glycosylated IgG1, which was previously denoted as mAb-1 in Chapters 2 and 4.

The isothermal aggregation kinetic experiments were performed on dialyzed samples at the temperature of 70 °C and at the protein concentration of 1 g/L. The pH was set to 6.5 by using a 20 mM histidine buffer. Details about the sample preparation and experimental characterization are reported in Chapter 2.

The chemicals (glycerol, threitol, sorbitol, sucrose, maltitol, PEG200 and PEG300) have been purchased from Sigma-Aldrich at the highest purity available. The molecular weights of PEG200 and PEG300 are in the range 190-210 g/mol and 285-315 g/mol, respectively.

5.3 Results and discussion

The mechanism of the heat-induced aggregation of the mAb under investigation in this chapter has been identified in the absence of cosolute in Chapter 2. Briefly, the aggregation mechanism was described as follows: a native folded conformation of the monomer $N$ unfolds with a rate constant $k_U$ to form an aggregation-prone monomer $U$, which is then depleted by aggregation. In addition, it was shown in Chapter 2 that protein unfolding is the rate-limiting step for monomer depletion, which implies that the concentration of $U$ is small compared to that of $N$.

In Chapter 4, the impact of sorbitol on the elementary events of the multistep aggregation mechanism of mAb-1 has been investigated. It has been shown that the presence of sorbitol delays the kinetics of mAb-1 aggregation by specifically inhibiting the unfolding step, without impacting the aggregation events. This finding is consistent with literature data showing that sugars and other polyols reduce mAb aggregation propensity by improving their conformational stability [39]. According to the mechanism proposed by Timasheff and Arakawa, the stabilizing
effect of these cosolutes can be attributed to unfavorable protein-solvent interactions which favor compact native protein conformations with respect to more open unfolded structures [151–156], as illustrated in Figure 5.1.

It is worth mentioning here that the kinetics of monomer depletion remains limited by the rate of protein unfolding also in the presence of sorbitol. Indeed, aggregation experiments performed at various protein concentrations in the presence of sorbitol revealed that the kinetics of monomer depletion follows a first order process, as shown in Figure 5.2. Moreover, kinetic analysis indicated that sorbitol increases the characteristic time of unfolding without impacting the characteristic time of aggregation (see Chapter 4), thus confirming that mAb aggregation remains unfolding rate-limited in the presence of sorbitol.

![Figure 5.2: Linearization of the decrease of the total monomer concentration (M = N + U) with time at 500 mM sorbitol for three protein concentrations, showing that the kinetics of monomer depletion follows a first order process.](image)

### 5.3.1 Models proposed

**Model 1**

In the model developed in Chapter 4, and which is here summarized as Model 1 in Figure 5.3, the stabilization effect of sorbitol was captured by modifying the value of the unfolding rate constant $k_U$ at different sorbitol concentrations, with $k_U$ decreasing upon an increase in sorbitol concentration.
The deceleration of protein unfolding in the presence of the polyol is attributed to the increase in the activation energy of unfolding due to the greater exclusion of the polyol molecules from the expanded transition state with respect to the native protein state [11]. Indeed, the increased surface area of the expanded transition state leads to higher unfavorable interactions between the protein surface and the polyol molecules as compared to the native state (see Figure 5.1).

The impact of the presence of polyol on the activation energy of unfolding can be quantified within the transition state theory by using the following expressions [160, 161]:

\[
\begin{align*}
    k_U &= A \exp \left( \frac{-\Delta H_U^\#}{RT} \right) \\
    A &= \kappa_T k_B T \frac{\Delta S_U^\#}{R} \exp \left( \frac{\Delta S_U^\#}{R} \right)
\end{align*}
\]  

(5.1)

where $\Delta H_U^\#$ and $\Delta S_U^\#$ are the enthalpy and entropy contributions to the free energy of activation, respectively. $T$ is the temperature, $R$, $k_B$, and $h$ are the gas, Boltzmann, and Planck constants, respectively. $\kappa_T$ is the transmission coefficient.
Over a narrow range of temperatures, $\Delta H^\#_U$ and the prefactor $A$ can be assumed independent of the temperature and their values can be determined from the slope and y-intercept, respectively, of an Arrhenius plot. The free energy of activation can then be computed as:

$$\Delta G^\#_U(T) = \Delta H^\#_U - T \Delta S^\#_U$$ (5.2)

Model 2

In this chapter, we also describe the stabilization by polyols in the frame of a second scheme, denoted as Model 2 in Figure 5.3, where two native monomeric conformations are considered: $N_U$, which is subject to unfolding, and $N_S$, which is a more compact conformation of the native-state ensemble and which is characterized by a negligible unfolding rate constant. The presence of polyol shifts the conformational equilibrium towards the more compact state [157–159], which is poorly populated in the absence of cosolute. This structural conversion reduces the effective concentration of native monomer which can undergo unfolding, therefore delaying the kinetics of unfolding, and consequently the kinetics of aggregation.

It is worth highlighting that within Model 2, the unfolding rate constant $k^0_U$ is not affected by the presence of polyol and is equal to the value estimated in the absence of cosolute.

In the following, we connect mathematically the two proposed models. The formation of the aggregation-prone monomer $U$ in Model 1 and in Model 2 can be expressed as follows:

$$\left( \frac{dU}{dt} \right)_{1,+} = k_U N$$ (5.3)

$$\left( \frac{dU}{dt} \right)_{2,+} = k^0_U N_U$$ (5.4)

where the concentration of native folded monomer ($N$) in Model 1 is equal to the sum of the concentrations of native monomer subject to unfolding ($N_U$) and native
monomer stabilized by the presence of polyol \((N_S)\):

\[
N = N_U + N_S
\]  (5.5)

Since the formation of \(U\) must be the same in the two models, we can write:

\[
k_U N = k_U^0 N_U
\]  (5.6)

At equilibrium, the constant \(K_{eq}\) in Model 2 is defined as:

\[
K_{eq} = \frac{N_U}{N_S}
\]  (5.7)

Then, Equation 5.5 reads:

\[
N = N_U \times \left(1 + \frac{1}{K_{eq}}\right)
\]  (5.8)

Finally, Equations 5.6 and 5.8 can be combined to provide a relation between the unfolding rate constant of Model 1 and the equilibrium constant of Model 2:

\[
k_U = \frac{k_U^0}{1 + 1/K_{eq}}
\]  (5.9)

where \(k_U^0\) corresponds to the value of \(k_U\) in the absence of polyol.

This relation shows that, from a mathematical point of view, the stabilization by polyols can be described equivalently either by a decrease in the unfolding rate constant of Model 1, or by a decrease in the conformational equilibrium constant of Model 2.

The Wyman linkage equation states that the equilibrium constant varies with the presence of cosolute according to [152, 162]:

\[
\frac{\partial \log (K_{eq})}{\partial (\gamma_p P)} = \Delta \nu
\]  (5.10)

where \(P\) is the concentration of polyol, \(\gamma_p\) is the activity coefficient, and \(\Delta \nu\) is the preferential binding parameter, i.e. the difference in ligand binding between the
two native states involved in the equilibrium.

The preferential binding parameter is a thermodynamic quantity which can be either positive or negative depending on the type of protein-cosolute interactions [151, 152]. In the case of unfavorable protein-cosolute interactions (as for example with sugars and other polyols), $\Delta \nu$ is negative, reflecting the preferential exclusion of the cosolute from the protein surface, or in other words the preferential hydration of the protein. In contrast, positive $\Delta \nu$ values correspond to favorable protein-cosolute interactions, which are typically encountered with denaturants such as urea or guanidine hydrochloride.

The activity coefficients of several polyol solutions, in a concentration range comparable to that investigated here, were reported to be comprised between 1 and 1.1 [163, 164]. Therefore, the activity coefficients of the polyols were approximated to 1 in this study.

The integration of Equation 5.10 then provides the following relation:

$$K_{eq} = K_{eq,ref} \times P^{\Delta \nu} \tag{5.11}$$

where the equilibrium constant $K_{eq,ref}$ is defined at the reference polyol concentration of 1 mol/L.

The parameters $\Delta \nu$ and $K_{eq,ref}$ can be easily estimated from the $K_{eq}$ values obtained at several polyol concentrations by using the linearized form:

$$\log (K_{eq}) = \Delta \nu \times \log (P) + \log (K_{eq,ref}) \tag{5.12}$$

It is worth noticing that the parameters $k_0^U$, $\Delta \nu$ and $K_{eq,ref}$ are independent of the polyol concentration. Therefore, once these three parameters are determined, the monomer depletion kinetics can be predicted at any polyol concentration lying in the range of parameter validity.
5.3.2 Stabilization by sorbitol

In the following, we apply the two models described previously to quantify the stabilization induced by the presence of sorbitol.

On this purpose, we determine first the values of the unfolding rate constant $k_U$ by fitting numerical simulations of Model 1 to experimental data acquired at several concentrations of sorbitol (0, 250, 500 and 1000 mM). The mass balance equations which were implemented to simulate the aggregation kinetics can be found in Chapter 2. The model simulations together with the experimental data are shown in Figure 5.4(a), and the obtained $k_U$ values are shown in Figure 5.4(b), where it appears that at zero sorbitol concentration $k_U^0 = 10^{-3}$ s$^{-1}$.

![Figure 5.4](image-url)
The impact of sorbitol on the activation energy of unfolding was then estimated within the transition state theory. For this purpose, the monomer depletion kinetics was monitored by SEC at several temperatures ranging between 65 and 70 °C (data shown in Figure 5.5(a)). The Arrhenius plots of the unfolding rate constant in the absence and in the presence of sorbitol are shown in Figure 5.5(b). The energies of activation of unfolding were determined at the reference temperature $T_m=70$ °C by using Equations 5.1 and 5.2, assuming $\kappa_T = 1$.

![Figure 5.5](image)

**Figure 5.5:** (a) Kinetics of the monomer depletion in the absence of polyol at the protein concentration of 1 g/L at various temperatures. (b) Arrhenius plot of the unfolding rate constant, in the absence and presence of 250 mM and 500 mM sorbitol. (c) Enthalpic and entropic contributions to the free energy of activation of unfolding as a function of sorbitol concentration evaluated from the Arrhenius plots of (b) at the reference temperature $T_m = 70$ °C.

It can be seen in Figure 5.5(c) that the presence of sorbitol increases both $\Delta H_U^\#$ and $\Delta S_U^\#$, but has only a little impact on $\Delta G_U^\#$. This observation suggests an enthalpy-entropy compensation mechanism, which is characteristic of many reactions in solutions [165–167]. The dehydration of the protein surface in the transition state is accompanied by an unfavorable increase in enthalpy related to the lower affinity of the sorbitol molecules for the protein surface with respect
to water molecules; this unfavorable increase in enthalpy is only partially com-
pensated by the favorable increase in entropy associated to the release of water
molecules absorbed on the protein surface. Overall, the enthalpy increase domi-
nates, leading to an increase in the free activation energy.

In a second step, we quantified the impact of sorbitol in the frame of Model 2
by computing the equilibrium constant $K_{eq}$ with Equation 5.9 at each sorbitol
concentration, as shown in Figure 5.4(c). The linearization provides $\Delta \nu = -2.0$
and $K_{eq,ref} = 0.13$ by using Equation 5.12.

Finally, we used the parameters $k_0^0$, $\Delta \nu$ and $K_{eq,ref}$ previously determined to
predict the monomer depletion at an additional sorbitol concentration (750 mM)
by using Model 2. In Figure 5.4(d), it can be seen that the prediction of the
monomer depletion kinetics at 750 mM sorbitol is in very good agreement with
the experimental data.

### 5.3.3 Impact of polyol size on mAb stabilization

In the following, we apply this modeling framework to other polyols (glycerol,
threitol, sucrose, maltitol) with the aim to understand the effect of cosolute size
on mAb stabilization. The cosolutes used in this study are listed in Table 5.1,
together with some of their most meaningful physical properties.

<table>
<thead>
<tr>
<th>Name</th>
<th>Chemical formula</th>
<th>Molecular weight [g/mol]</th>
<th>Molar volume [mL/mol]</th>
<th># -OH groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycerol</td>
<td>C$_3$H$_8$O$_3$</td>
<td>92.1</td>
<td>73.5 [168]</td>
<td>3</td>
</tr>
<tr>
<td>Threitol</td>
<td>C$<em>4$H$</em>{10}$O$_4$</td>
<td>122.1</td>
<td>86.7 [168]</td>
<td>4</td>
</tr>
<tr>
<td>Sorbitol</td>
<td>C$<em>6$H$</em>{14}$O$_6$</td>
<td>182.2</td>
<td>118.2 [168]</td>
<td>6</td>
</tr>
<tr>
<td>Sucrose</td>
<td>C$<em>{12}$H$</em>{22}$O$_{11}$</td>
<td>342.3</td>
<td>210.5 [169]</td>
<td>8</td>
</tr>
<tr>
<td>Maltitol</td>
<td>C$<em>{12}$H$</em>{24}$O$_{11}$</td>
<td>344.3</td>
<td>215.9 [170]</td>
<td>9</td>
</tr>
</tbody>
</table>

**Table 5.1:** List of the polyols considered in this study together with some
selected meaningful physical parameters.

The experimental data from size exclusion chromatography together with the
model simulations are presented in Figure 5.6.
In Figure 5.7(a), the unfolding rate constant $k_U$ evaluated in the frame of Model 1 is plotted as a function of cosolute concentration for all the cosolutes mentioned above. It is seen that, at a given cosolute molarity, the unfolding rate constant decreases with an increasing cosolute size. This result indicates that large molecules have a stronger stabilization effect as compared to small molecules.

The values of the conformational equilibrium constant $K_{eq}$ (computed from Equation 5.9) are shown as a function of cosolute concentration in Figure 5.7(b). The parameters $\Delta \nu$ and $K_{eq,ref}$ of Model 2 were determined by using Equation 5.12 and are presented in Figure 5.7(c) and Figure 5.7(d), respectively.

Interestingly, the value of $\Delta \nu$ is close to $-2$ for the all polyols, whereas $K_{eq,ref}$ is largely affected by the cosolute size. This means that the impact of polyol concentration on the mAb stability is similar for all the polyols investigated here, while the stabilization at a reference polyol concentration is polyol specific.
Figure 5.7: (a) Unfolding rate constant $k_U$ evaluated in the frame of Model 1 as a function of polyol concentration for polyols of increasing size. (b) Conformational equilibrium constant evaluated in the frame of Model 2, which provides the values of $\Delta \nu$ and $K_{eq,ref}$ by using Equation 5.12. (c) $\Delta \nu$ and (d) $K_{eq,ref}$ as a function of the molar volume of the polyols.

In particular, the stabilization effect is found to be enhanced with polyol size until a critical cosolute size corresponding to sorbitol (Figure 5.7(d)) is reached. In contrast, the stabilization effect becomes almost independent on the polyol size for cosolute molecules larger than sorbitol. For example, sucrose (342 g/mol) and sorbitol (182 g/mol) are found to have similar $K_{eq,ref}$ values, close to 0.1.

Moreover, it is worth noticing that even though sucrose and maltitol have similar molar volumes, the stabilization by maltitol is somewhat more effective than that by sucrose (Figure 5.7). This is possibly correlated to the number of -OH groups (and thus to the possible number of hydrogen bonds formed) of the two molecules. Indeed, maltitol carries nine -OH groups, while sucrose carries only eight -OH groups, due to its bi-cyclic structure. This observation suggests that the stabilization effect is not related exclusively to the volume occupied by the polyol molecules but also to the cosolute chemical structure.
To confirm this hypothesis, we plot the unfolding rate constant as a function of the volume fraction occupied by the polyols in Figure 5.8(a). It can be clearly seen that the results obtained with the various polyols do not overlap on one single curve. In particular, sorbitol appears to be the most stabilizing cosolute among the molecules investigated in this study, while glycerol is found to have a very weak stabilization effect.

Similar conclusions are reached when plotting the conformational equilibrium constant as a function of the occupied volume fraction (figure not shown) since the values of $K_{eq}$ of Model 2 are directly correlated to the values of $k_U$ computed within Model 1 (Equation 5.9). These results show that the stabilization effect does not depend only on the volume fraction occupied by the polyol molecules, thus suggesting that cosolute chemistry plays a central role in the mAb stabilization.

![Figure 5.8:](image)

**Figure 5.8:** (a) Unfolding rate constant as a function of the volume fraction occupied by the polyols for polyols of various sizes. (b) Kinetics of monomer depletion without PEG, with 250 mM PEG200 and with 250 mM PEG300.

The effect of cosolute chemistry on mAb stability is also evidenced in Figure 5.8(b), which shows the kinetics of monomer depletion in the absence and presence of 250 mM polyethylenglycol (abbreviated as PEG). PEG200 and PEG300 exhibit comparable sizes with respect to the polyols used for this study, but they exert an opposite effect on the kinetics of monomer depletion: while polyols stabilize the antibody, PEG has a destabilizing effect, as commonly observed for several protein systems [171].
5.4 Conclusion

In this chapter, we coupled experimental characterization with theoretical kinetic modeling to investigate the stabilization effect of polyols on the heat-induced aggregation of a monoclonal antibody whose monomer depletion kinetics is rate-limited by protein unfolding.

We interpreted the results within the frame of two mechanistic schemes. In the first scheme, the stabilization effect was described from a kinetic point of view, i.e. by an increase in the energy barrier preventing the protein from unfolding. In the second scheme, the presence of the polyol was interpreted from a thermodynamic point of view, i.e. by a shift of the native-state ensemble towards a more compact configuration.

This framework was applied to investigate the impact of polyol size on antibody stability. It was found that the relative increase in the protein stability (expressed in terms of unfolding rate constant in Model 1, or equilibrium constant in Model 2) with increasing polyol concentration is similar for the different polyol molecules. However, at constant polyol concentration, the stabilization effect is highly polyol specific. In particular, we found that the stabilization effect is enhanced with increasing polyol size for polyol molecules smaller than a critical size, corresponding to sorbitol. In contrast, for polyol molecules larger than sorbitol, the aggregation inhibition becomes almost independent of the polyol size. These results show that the mAb stabilization does not depend only on the volume fraction occupied by the polyol, but is also affected by the polyol chemical structure.

The content of this chapter has been partially published in the following article: L. Nicoud, N. Colurs, P. Arosio, E. Norrant, and M. Morbidelli. Effect of polyol sugars on the stabilization of monoclonal antibodies. *Biophys. Chem.*, 197:40-46, 2015
Chapter 6

Impact of aggregation on the viscosity of mAb solutions

6.1 Introduction

In the previous chapters, mAb aggregation studies were performed at low protein concentration, in the range from 1 to 5 g/L. Although investigations performed under such dilute environments represent a simplified situation as compared to the concentrated conditions that are generally required for drug administration, they provide a good basis to gain knowledge on the mechanisms of mAb aggregation (see Chapter 2), as well as on the effects of cosolutes (see Chapters 4 and 5). In this chapter, we move progressively from low towards high protein concentration, in the range from 20 to 60 g/L, which corresponds to occupied volume fractions up to approximately 20%.

Formulating protein-based drugs at high concentration is challenging, not only due to the strong aggregation propensity of proteins [3], but also due to the potentially high viscosity of protein solutions, which can heavily complicate product handling and delivery by injection [4]. The viscosity of protein solutions has been shown to be highly sensitive to the protein amino acid sequence and to the buffer composition [172–175], as well as to the presence of protein aggregates [69, 95, 176].
In this chapter, we aim at gaining a deeper understanding on the impact of mAb aggregation on the solution viscosity.

Since excessively high solution viscosities would hinder the commercialization of potential drug candidates, it is paramount to characterize the solution viscosity during the early stages of product formulation. Due to the limited amount of material available during drug development, effort is put towards the design of new methods for measuring solution viscosity at the microliter scale [177, 178]. Moreover, the analysis of traditional rheological data of protein solutions is a rather delicate task and requires some caution due to the complex behavior of proteins under shear [179, 180].

In this work, we follow the methodology proposed by He et al. [178], which consists in measuring the diffusion coefficient of tracer nanoparticles immersed in concentrated protein solutions with dynamic light scattering (DLS). While the experiments of the initial work were carried out under native conditions, we extend the technique to denaturating conditions with a view to quantifying the impact of aggregation on the solution viscosity. This method, which offers the advantages of being relatively time and material saving, allows one to measure the increase in solution viscosity in situ without shearing the sample, which may potentially induce rearrangement or breakage of loose aggregates.

Even though the irreversible aggregation and the reversible self-association of protein molecules have already been reported to increase the solution viscosity [69, 95, 176], a theoretical framework allowing the prediction of the impact of aggregate formation on the solution viscosity is still lacking. In this chapter, we correlate the increase in the solution viscosity during mAb aggregation with the increase in the aggregate occupied volume fraction, which has been proven useful to study the rheology of aggregating colloidal dispersions [98, 99]. The occupied volume fraction depends not only on the aggregate size and concentration, but also on the aggregate compactness, which is quantified here in the frame of the fractal scaling.
6.2 Materials and methods

6.2.1 Materials

The mAb used for this study was a glycosylated IgG1, which was denoted as mAb-1 in the previous chapters.

Prior to each experiment, the antibody stock solution was dialyzed against a 20 mM histidine buffer at pH 6.5 as described in Chapter 2. All the samples for the aggregation studies (performed in the protein concentration range from 20 to 60 g/L) were prepared by diluting the dialyzed stock solution to the targeted concentration with a 20 mM histidine buffer at pH 6.5.

For the viscosity measurements of the monomeric protein solution performed at room temperature at concentrations larger than 60 g/L, the protein solution was concentrated by centrifugal ultrafiltration with a 30 kDa cut-off molecular weight membrane (Vivaspin 500, VS0121, Sartorius).

6.2.2 Hydrodynamic radius measurement

Isothermal aggregation kinetic measurements were performed under thermal stress at the temperature of 70 °C. The increase in the average aggregate hydrodynamic radius with time was monitored by dynamic light scattering measurements performed both in batch mode by using a Zetasizer Nano (Malvern), and after elution in a size exclusion chromatography column by using a Dawn-Heleos II device equipped with a DLS module (Wyatt, Santa Barbara, CA, USA) and assembled on an Agilent series HPLC unit (Santa Clara, CA, USA). The points and error bars reported in the plots correspond to the average and standard deviation of measurements performed on at least two independent aggregate samples.
6.2.2.1 Batch mode

The samples analyzed by batch DLS were incubated directly in the DLS instrument in a quartz cuvette (ZEN 2112, Malvern), and then quenched on ice for a few minutes before measuring the aggregate size at room temperature. In order to prevent evaporation, a custom made plastic cap was added in the cuvette to reduce the air volume on top of the sample.

The aggregate size was estimated from the extrapolation to zero protein concentration by performing successive dilutions, as further explained in the Results section.

DLS measurements were performed at the fixed angle $\theta = 173^\circ$ by using a laser with wavelength $\lambda = 633$ nm. Briefly, the fitting of the autocorrelation function with the method of cumulants allows the determination of the average diffusion coefficient $D_0$, which is connected to the average hydrodynamic radius $\langle R_h \rangle$ by the Stokes-Einstein equation:

$$D_0 = \frac{k_B T}{6 \pi \eta_0 \langle R_h \rangle} \quad (6.1)$$

where $k_B$ is the Boltzmann constant, $T$ the temperature and $\eta_0$ the solvent viscosity.

The mean aggregate size provided by the methods of cumulants is the so-called $z$-average. It is mathematically stable and relatively insensitive to noise, making it the preferred parameter to estimate aggregate size from DLS measurements. Therefore, unless otherwise stated, the reported values of the hydrodynamic radius correspond to the $z$-average. Nevertheless, the volume and the number-average size, which are computed from the volume and number aggregate distributions respectively, were also considered with a view to assessing the impact of the selected averaging method on the estimated mean aggregate size (see Discussion part).
6.2.2.2 Flowing mode

The samples analyzed by size exclusion chromatography with inline light scattering were incubated in a block-heater (Rotilabo H 250, Roth, Karlsruhe) in hermetically sealed HPLC vials, as described in Chapter 2. The eluting species were detected by UV absorbance at 280 nm and by a DLS detector ($\lambda = 658$ nm, $\theta = 100.3^\circ$). The aggregate hydrodynamic radius was determined with the Astra software (Wyatt, Santa Barbara, CA, USA), by averaging all the hydrodynamic radius values measured across the aggregate peak. A representative SEC chromatogram and an example of a correlation function measured inline are presented in Appendix A.

It was verified by this technique that the non-heated antibody does not associate in the range of concentrations investigated in this study (i.e., up to 160 g/L).

6.2.3 Fractal dimension measurement

The aggregate morphology was investigated by static and dynamic light scattering measurements. The aggregate fractal dimension $d_f$ and the scaling prefactor $k_f$ were estimated from the correlation between the aggregate weight-average molecular weight $\langle MW_{Agg}^w \rangle$ and the aggregate average hydrodynamic radius $\langle R_{Agg}^h \rangle$ according to:

$$\frac{\langle MW_{Agg}^w \rangle}{MW_p} = k_f \left( \frac{\langle R_{Agg}^h \rangle}{R_p} \right)^{d_f}$$

(6.2)

where $MW_p = 150$ kDa and $R_p = 6$ nm are the molecular weight and the hydrodynamic radius of the monomeric protein, respectively.

Aggregate samples of different sizes were produced by heating protein samples at 70 °C in a block heater (Rotilabo H 250, Roth, Karlsruhe) for various incubation times. The samples were analyzed through a SEC column equipped with an inline multi-angle light scattering detector, as described in Chapter 2. The aggregate weight-average molecular weight and the aggregate average hydrodynamic radius were determined from the SLS and DLS, respectively, by using the Astra
software (Wyatt, Santa Barbara, CA, USA), as further detailed in Appendix A. The log linearization of Equation 6.2 for aggregates of different sizes allows the determination of the scaling parameters $d_f$ and $k_f$.

It was verified that these measurements give results that are consistent with the static light scattering data reported in Chapter 2, where it was shown that $d_f = 1.85$ at the protein concentration of 1 g/L from the dependence of structure factor on the q-vector.

### 6.2.4 Viscosity measurement with tracer nanoparticles

**Principle**

The viscosity of protein solutions was estimated by measuring the diffusion coefficient of the tracer nanoparticles with dynamic light scattering, both in the solvent ($D_{0,T}$) and in the concentrated protein solution ($D_T$). According to the Stokes-Einstein equation applied to the tracer particle of hydrodynamic radius $R_{NP}$, both in solvent and in the protein solution:

\[
\begin{align*}
D_{0,T} &= \frac{k_B T}{6\pi \eta_0 R_{NP}} \\
D_T &= \frac{k_B T}{6\pi \eta_{\infty} R_{NP}}
\end{align*}
\]  \( (6.3) \)

It follows that the solution macroscopic viscosity can be estimated from:

\[
\eta_{\infty} = \frac{\eta_0 D_{T,0}}{D_T}
\]  \( (6.4) \)

where $\eta_0$ is the viscosity of the solvent and $\eta_{\infty}$ is the macroscopic viscosity of the concentrated protein solution.

The subscript $\infty$ indicates that the tracer nanoparticles are large as compared to the surrounding protein molecules. Indeed, insufficiently large tracer nanoparticles would probe an effective viscosity comprised between the solvent viscosity
and the macroscopic viscosity of the concentrated protein solution, as described in Chapter 7.

This method was used to measure both the viscosity of monomeric protein solutions at 25 °C, and to evaluate the increase in solution viscosity due to aggregate formation at the elevated temperature of 70 °C. The aggregation experiments were performed in situ in a quartz cuvette covered with a custom made plastic cap to limit sample evaporation. The total sample volume was 65 µL and the volume fraction of tracer nanoparticles was $1.4 \times 10^{-3}$. The dilution of the protein by addition of the nanoparticles was accounted for during sample preparation. The points and error bars reported in the plots correspond to the average and standard deviation of measurements performed on at least two independent samples.

**Tracer nanoparticle synthesis**

Poly(methyl methacrylate) (PMMA) based nanoparticles were synthetized through bulk emulsion polymerization in a three-neck round-bottom flask equipped with a reflux condenser and a thermocouple. Stripped deionized water was introduced in the reactor and heated to 70 °C with an oil bath placed on a hot-plate stirrer. The reactor atmosphere was kept inert by flushing the reaction set-up with nitrogen. The surfactant (hexadecyltrimethylammonium bromide, abbreviated as CTAB) and the monomers (methyl methacrylate and 2-methacryloyloxy ethyl trimethylammonium, abbreviated as MMA and META, respectively) were then added in the reactor. Finally, the initiator (2,2'-bis(diazen-1,2-diyl) bis (2-methylpropanimidamide) dihydrochloride, abbreviated as INI) was injected to initiate the polymerization. The reaction volume was 100 mL, with the following mass concentrations: $w_{MMA}/w_{TOT} = 5\%$, $w_{META}/w_{TOT} = 0.06\%$, $w_{CTAB}/w_{MMA} = 0.0125\%$, $w_{INI}/w_{MMA} = 0.02\%$. The reaction time was 2 h.

The synthetized nanoparticles had a final hydrodynamic radius $R_{NP} = 99$ nm and a zeta potential of $49.7 \pm 0.9$ mV, as measured with a Zetasizer Nano (Malvern). Positively charged nanoparticles were produced with the purpose of
reducing attractive interactions with the antibody molecules which carry a positive net charge, as evidenced by the positive value of the zeta potential measured in Chapter 2 \((5.7 \pm 3.5 \text{ mV})\).

**Validity of the method**

The main limitation of the DLS technique is related to the interactions between the tracer nanoparticles and the protein molecules, which can potentially lead to nanoparticle aggregation \([178, 181]\). We proved experimentally that the tracer nanoparticles are stable under the conditions of interest, and that they are not affecting the kinetics of mAb aggregation by successively verifying that: (i) the elevated temperature required to induce mAb aggregation does not destabilize the nanoparticles (Figure C.1(a)); (ii) the nanoparticles are stable in the presence of mAb molecules under native conditions (Figure C.1(a)); (iii) the size of the nanoparticles is unchanged after mAb aggregation has been induced at high temperature (Figure C.1(b)); (iv) the measured viscosity is independent of the concentration of tracer particles (Figure C.2). These data are reported in Appendix C.

Moreover, the accurate measurement of the solution viscosity from the diffusion coefficient of tracer nanoparticles relies on the assumption that the nanoparticles do not interact via long-range electrostatic interactions with the protein molecules \([182]\). The validity of this assumption can be assessed from the comparison between the Debye length, which quantifies the thickness of the diffusive layer, and the size of the nanoparticles. To do so, the Debye length can be computed from the inverse of the Debye parameter, which is defined in Chapter 2. In a buffer of 20 mM histidine, it is computed that the Debye length is equal to 3.8 and to 4.1 nm, at the temperatures of 25 and 70 °C, respectively. The Debye length is thus much smaller than the tracer particle size, implying that the electrostatic interparticle interactions are short-ranged. Therefore, it can reasonably be assumed that the viscosity measurements are not affected by the electrostatic repulsion between the nanoparticles and the protein molecules.
6.2.5 Rheological measurements

Rheological measurements were carried out with a Physica MCR 300 rheometer (Paar Physica) by using a cone-and-plate geometry (12.5 mm radius, 2° angle). The gap was set to 0.05 mm and the sample volume was 160 µL. The temperature was maintained at 25 °C by a Peltier element (TEK 150 PA-C). A solvent trap was used to prevent solvent evaporation. The shear rate was increased gradually from 10 to 1000 s⁻¹, and viscosity values at 1000 s⁻¹ were considered, in agreement with protocols proposed in the literature [95, 175, 181, 183]. The points and error bars reported in the graph correspond to the average and standard deviation of measurements performed on three independent samples.

6.3 Results

6.3.1 Viscosity of monomeric protein solutions

Before following the viscosity increase during mAb aggregation, the experimental method was tested on monomeric protein solutions. To do so, the viscosity of several mAb solutions estimated from the measurement of the diffusion coefficient of tracer nanoparticles by DLS was compared to results obtained from rheology. In Figure 6.1 it can be seen that the two techniques provide very similar results, thus validating the technique proposed in this chapter.

Moreover, the measured viscosity values are compared to the theoretical expression proposed by Gillespie for a suspension of hard spheres [184]:

\[ \eta_{\infty}(\Phi_0) = \eta_0 \frac{1 + \Phi_0/2}{(1 - \Phi_0)^2} \]  

(6.5)

where \( \eta_0 \) is the solvent viscosity and \( \Phi_0 \) is the volume fraction occupied by the protein molecules, which is directly proportional to the protein concentration.
Chapter 6. Impact of aggregation on viscosity

Figure 6.1: Comparison between viscosity results obtained from the diffusion coefficient of tracer nanoparticles with DLS, and data obtained from rheology for monomeric protein solutions. The measurements were performed at 25 °C to avoid aggregation. The dotted line corresponds to theoretical viscosity values computed by using Equation 6.5 considering a protein diameter of 12 nm.

It can be observed that the experimental data are well described by Equation 6.5. Nevertheless, it must be underlined that this observation can hardly be generalized, since the viscosity of protein solutions has been shown to be very sensitive to the protein amino acid sequence as well as to the buffer composition [172–175]. To take interparticle interactions into account, it has been suggested that $\Phi_0$ can be replaced by an effective volume fraction which depends on the interaction potential [184].

6.3.2 Viscosity of aggregated protein solutions

Aggregation kinetics experiments were performed at neutral pH under thermal stress (pH 6.5, 70 °C) at three mAb concentrations (20, 40, 60 g/L). The increase in solution viscosity with time was estimated from the diffusion coefficient of tracer nanoparticles measured with DLS in situ. As explained in Section 6.2.4, preliminary experiments have been performed in order to verify the validity of this technique. In particular, it has been demonstrated that the nanoparticles are stable during mAb aggregation, and that the measurements of the rise in solution viscosity are independent of the concentration of tracer particles (data are shown in Appendix C).
Figure 6.2(a) shows the kinetics of viscosity increase for the three considered protein concentrations. It can be seen that, as expected, the higher the protein concentration, the faster is the rise in solution viscosity.

In a second type of experiments, the growth of protein aggregates with time was monitored (in the absence of nanoparticles) with DLS. Figure 6.2(b) shows the increase in hydrodynamic radius for the three investigated protein concentrations measured both in batch and in flowing modes.

6.3.3 Interaction parameter

It is worth mentioning that the accurate determination of the protein aggregate size requires performing measurements under dilute conditions. Indeed, the value of the diffusion coefficient $D_c$ measured by DLS is affected both by protein-protein interactions and by hydrodynamics effects (including viscosity effects), which become significant at high protein concentration. These non-ideal effects can be quantified by the interaction parameter $k_D$, which is defined as the first order concentration dependence of the diffusion coefficient [38, 50]:

$$D_c = D_0(1 + k_D c + \ldots)$$  \hspace{1cm} (6.6)
where $c$ is the protein concentration, and $D_0$ is the extrapolation of the diffusion coefficient to zero protein concentration. The subscript $c$ in $D_c$ stands for collective diffusion (i.e. the overall movement of solutes under a concentration gradient), which differs from the self-diffusion (i.e. the random motion of a labeled molecule) [185].

For a solute of molecular weight $MW$ and of partial specific volume $\nu$, the interaction parameter $k_D$ can be related to the second virial coefficient $\overline{B}$ (defined in Equation 1.1) and to the first order concentration dependence of the sedimentation coefficient $k_S$ according to [186]:

$$k_D = 2N_a \times \overline{B} \times MW - k_S - \nu$$  \hspace{1cm} (6.7)

where $N_a$ is the Avogadro number and $\overline{B}$ is in L.

Samples analyzed by SEC-DLS are extensively diluted during elution in the chromatography column and thus reach the light scattering detector at sufficiently low protein concentration. Samples analyzed by batch DLS instead require manual dilution before analysis, which is performed at room temperature to avoid further aggregation. It is worth mentioning here that the mAb aggregates were shown to be irreversible upon dilution in Chapter 2. In Figure 6.3(a), we show the impact of protein concentration on the apparent hydrodynamic radius measured by DLS (considering the solvent viscosity) for a monomeric sample and for four aggregated samples of various sizes. It can be seen that the measured hydrodynamic radius dramatically increases upon sample dilution.

In Figure 6.3(b), these results are presented in terms of diffusion coefficients normalized by the value under dilute conditions. Interestingly, the data points from the five independent samples follow the same linear trend. This shows that, at least for the system under investigation, one single interaction parameter $k_D$ is sufficient to describe both the monomeric and the aggregated protein solutions. It is also interesting to note that no deviation from the linear trend is observed up to the high concentration of 60 g/L, suggesting that higher order terms in Equation 6.6 can be neglected in the investigated range of protein concentrations.
Chapter 6. Impact of aggregation on viscosity

Figure 6.3: Impact of sample concentration on batch DLS measurements. (a) Measured hydrodynamic radius as a function of protein concentration for a monomeric sample (⋆) and for four aggregated samples incubated at: 20 g/L for 20 min (■), 20 g/L for 40 min (♦), 40 g/L for 30 min (●), 60 g/L for 20 min (▲). (b) Normalized diffusion coefficient as a function of protein concentration, providing the interaction parameter $k_D$.

It is worth highlighting that the batch and SEC-DLS measurements give access to conceptually different quantities: while batch experiments provide the average hydrodynamic radius of the whole sample, SEC-DLS experiments provide the average hydrodynamic radius of the aggregate population only, i.e., excluding the monomer.

However, in this specific case, most of the monomer is already consumed at the incubation time points where DLS analysis is carried out, as will be shown in Figure 8.4. Moreover, as the scattered intensity increases with the scatterer radius to the power of 6, the contribution of the residual monomer to the total scattered intensity is negligible as compared to the contribution of the aggregates. Therefore, both batch and SEC-DLS measurements provide very similar values of the average hydrodynamic radius for the system under investigation, as can be seen in Figure 6.2(b). The good agreement between the two sets of data also supports the validity of the SEC-DLS analysis, ruling out possible artifacts due to aggregate breakage or interactions with the column matrix.
6.3.4 Occupied volume fraction

We now correlate the increase in solution viscosity with the increase in aggregate size. In Figure 6.4(a), the measured solution viscosity is plotted as a function of the aggregate hydrodynamic radius for each protein concentration and incubation time. It appears clearly from this set of experimental data that the solution viscosity increases both with the aggregate size and with the protein concentration.

![Graph showing viscosity increase as a function of aggregate size and protein concentration](image)

**Figure 6.4:** (a) Viscosity increase as a function of aggregate size for three protein concentrations: 20 g/L (●), 40 g/L (■), 60 g/L (▲). (b) Viscosity increase as a function of the estimated volume fraction (defined by Equation 6.11) for the same three protein concentrations. All the data from (a) collapse on one single master curve.

In the following, we aim at rationalizing the increase in solution viscosity with time by introducing an effective volume fraction occupied by the aggregates, which accounts for the two contributions identified experimentally: the aggregate size and the protein concentration. Accordingly, we define the occupied volume fraction \( \Phi \) as:

\[
\Phi = \frac{4}{3} \pi \sum_i R_{h,i}^3 M_i
\]  

(6.8)

where \( M_i \) and \( R_{h,i} \) are the number concentration and hydrodynamic radius, respectively, of the aggregates containing \( i \) primary particles.

By approximating the aggregate population to a monodisperse population of average radius \( \langle R_h \rangle \) and average number concentration \( \langle M \rangle \), we can estimate an approximate volume fraction \( \phi \) with:

\[
\phi = \frac{4}{3} \pi \langle R_h \rangle^3 \langle M \rangle
\]  

(6.9)
The average concentration of aggregates can be estimated from the initial concentration of primary particles (i.e. the initial number protein concentration $M_0$) divided by the average number of primary particles per aggregate (which is equal to the ratio between the average aggregate molecular weight $\langle MW_{Agg} \rangle$ and the molecular weight of the monomeric protein $MW_p$):

$$\langle M \rangle = \frac{M_0}{\langle MW_{Agg} \rangle / MW_p}$$  \hspace{1cm} (6.10)

Moreover, the average aggregate molecular weight can be estimated from the aggregate radius by using the fractal scaling introduced in Equation 6.2. Finally, the occupied volume fraction can be evaluated from the following equation:

$$\phi = 4\pi \frac{M_0}{3k_f R_p^{d_f}} \frac{R_f}{\langle R_h \rangle}^{3-d_f}$$  \hspace{1cm} (6.11)

The above relation shows that the impact of aggregation on the occupied volume fraction strongly depends on the aggregate morphology through the fractal dimension $d_f$ and the scaling prefactor $k_f$. For a system characterized by a fractal dimension smaller than 3, the increase in aggregate size with time leads to an increase in the occupied volume fraction. The more open are the aggregates, i.e. the lower is the fractal dimension, the more pronounced is this effect.

It is worth highlighting that an estimate of the occupied volume fraction can be obtained with Equation 6.11 only for fractal aggregates. Nevertheless, the key concept of occupied volume fraction (as defined by Equation 6.8) is general enough to be applied to any kind of aggregate geometry. Although defining the occupied volume of non-fractal aggregates would require more theoretical work than the approximation proposed in Equation 6.11, the general approach of correlating viscosity increase with the occupied volume fraction should remain valid even for non-fractal aggregates.
6.3.5 Aggregate fractal morphology

Considering Equation 6.11, it can be seen that the proper estimation of the occupied volume fraction requires to characterize the aggregate morphology, and in particular to evaluate the parameters $d_f$ and $k_f$. These two parameters can be assessed experimentally directly from the correlation between the aggregate molecular weight (measured by SLS) and the aggregate hydrodynamic radius (measured by DLS) by using Equation 6.2. Figure 6.5(a) shows the correlation between aggregate mass and aggregate radius, as determined from a multi-angle light scattering device, for four protein concentrations: 1, 20, 40 and 60 g/L.

Figure 6.5(b) shows the aggregate fractal dimension obtained from the power law fitting of the experimental data with Equation 6.2. It is observed that the fractal dimension increases with the protein concentration, i.e. denser aggregates are produced at high concentration. One possible reason is that in concentrated systems the colliding clusters are likely to be entangled, thus favoring the formation of bonds in the core rather than at the tips of the clusters. The impact of the initial volume fraction $\Phi_0$ on the fractal dimension $d_f$ has been quantitatively investigated by means of Monte Carlo simulations by González et al. [76], who report the following square root type relation under diffusion-limited conditions:

$$d_f = 1.8 + 0.9\sqrt{\Phi_0}$$  \hspace{1cm} (6.12)

The experimental $d_f$ values obtained in this work follow a fairly similar trend, as shown in Figure 6.5(b).

The power law fitting of the experimental data presented in Figure 6.5(a) provides not only the fractal dimension $d_f$, but also the scaling prefactor $k_f$. In Figure 6.5(c), the scaling prefactor is plotted as a function of the aggregate fractal dimension. It is seen that the prefactor values are close to 1 and decrease with increasing fractal dimensions. This reverse correlation between $k_f$ and $d_f$ has already been reported in the literature [74, 121, 187], and Ehrl et al. proposed the
Figure 6.5: Impact of protein concentration on aggregate morphology. (a) Correlation between aggregate molecular weight and aggregate hydrodynamic radius from light scattering experiments. Lines correspond to the power law fitting of the experimental data. (b) The fractal dimension of aggregates increases with protein concentration. Experimental data (symbols) are compared to the correlation (dashed line) proposed by Gonzáles et al. (Equation 6.12). (c) The scaling prefactor decreases with the aggregate fractal dimension. Experimental data (symbols) are compared to the correlation (dashed line) proposed by Ehrl et al. (Equation 6.13). Error bars in (b) and (c) represent the 90 % confidence interval for the parameter determination from the regressions shown in (a).

following empirical fitting [188]:

\[ k_f = 4.46d_f^{-2.08} \]  

(6.13)
The experimental values of the prefactor obtained in this work can be well described by the above correlation, as shown in Figure 6.5(c).

6.3.6 Master curve

Knowing the values of the aggregate fractal dimension $d_f$ and of the prefactor $k_f$, it is possible to estimate the occupied volume fraction for a given aggregate size and protein concentration through Equation 6.11. In Figure 6.4(b), the solution viscosity is plotted as a function of the occupied volume fraction for the three investigated protein concentrations. Most remarkably, all the data points, which were collected at several incubation times and several protein concentrations collapse on one single master curve. This demonstrates that the impact of aggregate formation on the increase in viscosity of protein solutions can be successfully rationalized by using the key concept of occupied volume fraction. It appears in Figure 6.4(b) that the occupied volume fraction reaches values slightly above unity. This may be either due to cluster spatial overlapping (i.e. the volume of porous clusters is counted twice), or due to the approximate expression used to compute $\phi$ (this point will be further discussed in the next section).

6.4 Discussion

We studied the impact of irreversible aggregate formation on the viscosity increase of a mAb solution exposed to thermal stress in the protein concentration range from 20 to 60 g/L. The rise in viscosity was monitored by measuring the diffusion coefficient of tracer nanoparticles by using dynamic light scattering in situ, while the kinetics of aggregate growth was followed by off-line DLS measurements performed both in batch and in flowing modes. The two types of measurements provided similar values of the aggregate hydrodynamic radius provided that the samples analyzed by batch DLS were diluted to a sufficiently low protein concentration (i.e. at around 1 g/L).
6.4.1 Protein-protein interactions

Interestingly, we found that the diffusion coefficient of aggregates of different sizes have a similar dependence on the protein concentration than the monomer. This suggests that protein interactions can be estimated in a coarse-grained manner by using a single interaction coefficient \( k_D \), both in a monomeric and in an aggregated solution. In the following, we attempt to deepen this finding and its implications on the second virial coefficient between protein aggregates. In order to do so, we simplify the expression of \( k_D \) given by Equation 6.7 by estimating the contribution of the various terms for this specific system.

First, the partial specific volume \( \nu \), which is in the order of 0.7 mL/g [189], is neglected with respect to \( k_D \), which was evaluated here to \( k_D = 20.8 \) mL/g. Second, it is assumed that the sedimentation coefficient \( k_S \) can be approximated by the diffusion coefficient \( k_D \). Indeed, \( k_S \) was reported to be of the same order of magnitude than \( k_D \) for several monoclonal antibodies characterized by \( k_D \) values in the order of 20 mL/g, which are values typically encountered at low ionic strength [50, 173]. These two simplifications imply that Equation 6.7 reduces to:

\[
k_D = 2N_a \times \bar{B} \times MW
\]  

(6.14)

Since it was shown experimentally that the diffusion coefficient \( k_D \) of several aggregated samples is similar to the one of the monomeric protein, it follows that the second virial coefficient between aggregates of molecular weight \( MW_{Agg} \) (denoted as \( \bar{B}_{Agg} \)) can be roughly estimated from the second virial coefficient between monomeric molecules of molecular weight \( MW_p \) (denoted as \( \bar{B} \)) according to the following equation:

\[
\bar{B}_{Agg} = \bar{B} \frac{MW_{Agg}}{MW_p}
\]  

(6.15)

Nevertheless, it must be underlined once more that the above equation is valid only under the assumptions of \( k_S \approx k_D \) and \( \nu \ll k_D \).
6.4.2 Master curve of the macroscopic viscosity

The increase in volume fraction during time was then computed from the protein concentration, aggregate size and fractal dimension, which was measured by light scattering. We showed that, when plotted as a function of the occupied volume fraction, the kinetics of viscosity increase obtained at three different protein concentrations collapse on one single master curve (Figure 6.4(b)).

It is then interesting to compare this master curve, which was constructed by analyzing aggregated samples of various sizes and protein concentrations, with a similar curve obtained with monomeric samples only, which were prepared at several protein concentrations. In Figure 6.6, data obtained with aggregated samples under thermal stress (black circles) are compared to viscosity measurements performed on monomeric samples at room temperature (back stars). The results are presented in terms of normalized viscosity in order to remove the impact of temperature on the solvent viscosity.

![Figure 6.6](image)

**Figure 6.6:** Normalized solution viscosity as a function of the occupied volume fraction for a monomeric protein solution at 25 °C (stars) and for aggregated samples at 70 °C (circles). The occupied volume fraction of the aggregated samples was computed by using the z-average (black circles), the volume average (grey circles) or the number average (white circles) aggregate radius. The data computed from the volume average and number average aggregate radius correspond to the batch DLS measurements only.

It can be observed that the two curves do not collapse. Indeed, the curve obtained with monomeric samples lies above the curve constructed with aggregated samples. Three main possible reasons can be put forth to explain this observation:
(i) The evaluated occupied volume fraction is overestimated because clusters interpenetrate due to the high protein concentration.

(ii) The evaluated occupied volume fraction is overestimated because the average value of the hydrodynamic radius which was considered to compute the occupied volume fraction is biased towards large aggregate sizes. Indeed, for a given aggregate size distribution, several values of the average size can be computed, and the z-average, which is commonly evaluated in DLS experiments, gives prominence to large aggregate sizes.

(iii) At a given occupied volume fraction, the viscosity of a monomeric solution is higher than the viscosity of an aggregated sample due to the polydispersity (i.e. the broadness) of the aggregate distribution. Polydispersity is indeed known to decrease the viscosity of colloidal dispersions [190–193].

It can be observed in Figure 6.6 that even at the early-stages of the aggregation process (let say for $\phi$ values up to 0.3), where cluster overlapping can be considered negligible, the viscosity of the monomeric protein solutions is already larger than the viscosity of aggregated samples of similar volume fractions. This suggests that assumption (i) alone is unlikely to explain the observed results.

In order to assess the likelihood of explanation (ii), we estimated the impact of the choice of the definition used to compute the average aggregate size on the computed occupied volume fraction. In order to do so, we compared the results obtained with the z-average to those obtained with the number-average and the volume-average aggregate radius for the batch DLS measurements. In Figure 6.6, it can be seen that the selected type of average strongly impacts the value of the estimated volume fraction. In particular, for a given aggregate population, the volume fraction computed based on an average aggregate size scales in this order: number-average < volume-average < z-average. Nevertheless, it is worth highlighting that for each type of average, a master curve is obtained when the kinetic profiles acquired at different protein concentrations are normalized based on the occupied volume fraction. This shows that, even though the absolute values of $\phi$ must be taken with some caution due to the difficulty to estimate the occupied
volume fraction from experimental data, the concept of occupied volume fraction is key to rationalize the viscosity increase during aggregation. Moreover, its is demonstrated in Appendix D that the volume fractions computed from the number and volume-average aggregate radius provide, respectively, a lower and an upper boundary for the real occupied volume fraction. Since the master curve obtained when considering the number-average hydrodynamic radius (which underestimates the real occupied volume fraction) lies below the curve obtained with monomeric protein solutions, explanation (ii) can be ruled out.

Therefore, we conclude from this analysis that aggregate polydispersity plays a prominent role in decreasing the solution viscosity at constant volume fraction.

6.5 Conclusion

In this chapter, the effect of irreversible aggregate formation on the change in viscosity of a concentrated monoclonal antibody solution was investigated. To do so, the increase in solution viscosity was monitored by following the reduction in the diffusion coefficient of tracer nanoparticles with DLS in situ, while the aggregate size was followed by light scattering.

It was shown that the rise in solution viscosity with time monitored at three different protein concentrations collapse on one single master curve when the reaction profiles are normalized based on the effective volume fraction, which depends on aggregate concentration, size and morphology. Importantly, it was found that the viscosity of an aggregated protein solution is lower than the viscosity of a monomeric protein solution of similar occupied volume fraction due to the polydispersity of the aggregate distribution.

In the next chapter, we will study the impact of an increase in the solution macroscopic viscosity on the aggregation kinetics of a colloidal system.

Chapter 7

Viscosity scaling in concentrated dispersions and its impact on colloidal aggregation

7.1 Introduction

In Chapter 6, we investigated how aggregation leads to an increase in the viscosity of concentrated protein solutions. Here, we are interested in understanding how an increase in solution viscosity results in a reduction of the aggregation rate. All in all, we aim at gaining insights into the interplay between aggregate formation and viscosity increase in concentrated colloidal systems.

As a first step, we focus on quantifying the effective viscosity that is experienced by a tracer particle immersed in a colloidal dispersion. It is worth pointing out that a comprehensive understanding of tracer diffusion in crowded environments is of fundamental relevance, not only in the estimation of aggregation rates in colloidal systems [19], but also in other research fields covering a wide range of applications, such as molecular transport during polymerization reactions [194], drug release from hydrogels [195], and protein mobility in living cell organisms [196]. Such interest has inspired a wealth of experimental
studies [101, 197–201], computer simulations [201–204], and theoretical investigations [205–209] on tracer diffusion over the past decades.

It has been evidenced that the diffusion of a tracer object in a crowded solution depends on a variety of factors, such as the filled volume fraction [101, 197, 199, 204], the interactions between the tracer and its environment [202], the size and shape of the crowders [201, 203], as well as the size and shape of the tracer itself [101, 197, 198, 200, 204]. In this work, we intend to clarify the complex dependence of the tracer diffusion coefficient on the tracer size in various host dispersions.

Fluorescence correlation spectroscopy (FCS) studies revealed that the viscosity experienced by small probes at the nanoscale in crowded solutions can be much lower than the macroscopic viscosity measured with a standard rheometer [101, 197]. Indeed, small tracers are mainly surrounded by solvent molecules, and thus experience a viscosity close to the solvent viscosity, which can be several orders of magnitude lower than the viscosity of the crowded solution. This observation introduces the concept of scale-dependent viscosity, which ranges from the solvent viscosity at the nanoscale to the viscosity of the complex solution or dispersion at the macroscale. The notion of scale-dependent viscosity has been evidenced by FCS experiments carried out with fluorescent probes of various sizes in polymer and micellar solutions [101, 197]. It has also been endorsed based on theoretical arguments in the particular cases of polymer melts [210], and of dilute colloidal dispersions of hard spheres with hydrodynamic interactions [205–207]. However, a general theoretical framework allowing the description of the variations of the effective viscosity at different length scales in a crowded environment is still lacking.

In this study, we focus on the case of a spherical tracer particle immersed in a concentrated dispersion of host spherical particles, and we quantify the effective viscosity as a function of the tracer particle size by means of Brownian dynamics simulations. This situation represents a well-defined system as compared to the aforementioned polymer and micellar solutions that form complex entangled
networks at high concentration [101, 197]. After investigating the dependence of the viscosity experienced by a tracer particle with the tracer particle size in a given host dispersion, we examine successively the impact of varying the host particle size, the volume fraction, and the polydispersity of the host particle size distribution on the viscosity scaling.

As a second step, we apply our results in the frame of the aggregation of concentrated colloidal systems by including the scaling of the effective viscosity in the classical diffusion and reaction-limited aggregation kernels. We quantify the impact of viscosity effects on the kinetics of aggregate growth as well as on the shape of the aggregate distribution by means of population balance equations simulations.

### 7.2 Methods

#### 7.2.1 Brownian dynamics simulations

The effective viscosity experienced by a spherical tracer particle of radius $a_T$ immersed in a dispersion of spherical host particles was computed by means of Brownian dynamics simulations, as schematically represented in Figure 7.1(a). The mean radius and the variance of the host particle size distribution are denoted $a_H$ and $\sigma_H^2$, respectively, while the total occupied volume fraction is referred to as $\Phi_0$.

A virtual simulation box was generated and 500 particles including one tracer particle were initially positioned inside the box in a random manner avoiding particle overlapping. The size of the box was defined in order to meet the target total volume fraction, which corresponds to the volume fraction of the host dispersion since the volume fraction of the tracer particle is negligible in the range of size ratios explored (comprised between 0.2 and 2). In addition, periodic boundary conditions were applied to mimic an infinite system.
Figure 7.1: (a) Schematic representation of the tracer and host particles in a Brownian dynamics simulation. (b) Representative examples of mean square displacement curves of a tracer particle with $a_T = 10$ nm obtained in pure solvent (continuous line) as well as in a dispersion of host particles using $a_H = 10$ nm, $\Phi_0 = 15\%$, $\sigma_H^2 = 0$ nm$^2$ (dashed line). The slope of such curves provides the diffusion coefficient of the tracer particle in both environments (denoted $D_T$ and $D_{0,T}$, respectively). The mean square displacement curves were obtained by averaging results of 500 simulations.

The motion of the particles was described by the overdamped Langevin equation neglecting inertia according to:

$$F_{El,i} + F_{Br,i} + F_{Dr,i} = 0 \quad (7.1)$$

where $F_{El,i}$, $F_{Br,i}$, and $F_{Dr,i}$ are, respectively, the total force associated to the particle-particle electrostatic interaction potential, the Brownian force, and the drag force acting on the particle $i$.

More explicitly, the expression for $F_{El,i}$ is as follows:

$$F_{El,i} = -\sum_{j \neq i} \frac{\partial V_{El,ij}}{\partial r_{ij}} \quad (7.2)$$

where $V_{El,ij}$ is the electrostatic interaction potential between the particles $i$ and $j$ and $r_{ij}$ is the interparticle distance.

It was assumed that the particles interact only via a short-range electrostatic potential with a characteristic Debye length equal to 1% of the smallest particle diameter and with a surface potential of 10 mV. In this way, the summation in Equation 7.2 is only carried out for those particles that are nearest neighbors.
of \(i\). The modulus of the electrostatic interaction potential between two differently sized particles was computed by using the analytical formula proposed by Sader et al. [211] considering equal surface potentials:

\[
V_{El,ij}(r_{ij}) = 4\pi\epsilon_0\epsilon_r\psi_0^2 a_i a_j \frac{r_{ij}}{r_{ij}} \ln (1 + \exp(-\kappa h_{ij})) \tag{7.3}
\]

where \(h_{ij}\) is the distance of closest approach between the particles \(i\) and \(j\), \(\kappa\) is the Debye length, \(\psi_0\) is the surface potential, \(a_i\) and \(a_j\) are the particle radii, \(\epsilon_0\) is the vacuum permittivity, and \(\epsilon_r\) is the relative dielectric constant of the medium.

The presence of a repulsive potential was necessary to prevent overlap among particles during simulations. It was nevertheless verified that small changes in the surface potential and in the characteristic length of electrostatic interactions do not affect the tracer particle diffusion coefficient.

The drag force was computed by using the Stokes law:

\[
F_{Dr,i} = -6\pi\eta_0 a_i^3 u_i \tag{7.4}
\]

where \(u_i\) is the velocity of the \(i^{th}\) particle.

Finally, \(F_{Br,i}\) is a stochastic term describing random motion of the particles, and was computed from a vector \(X_i\) containing Gaussian distributed random numbers with zero mean and a variance equal to unity.

Introducing these three contributions in Equation 7.1, the increments in particles position \(\Delta r_i\) were computed at each time step \(\Delta t\) according to the following equation:

\[
\Delta r_i = \frac{F_{El,i}}{6\pi\eta_0 a_i^3} \Delta t + X_i \sqrt{\frac{k_B T}{3\pi\eta_0 a_i^3} \Delta t} \tag{7.5}
\]

The first term in Equation 7.5 is proportional to the time step, corresponding to the deterministic contribution given by the repulsive forces exerted by the other particles. The second term is proportional to the square root of the time step, corresponding to the random motion contribution [212].
Chapter 7. Viscosity scaling

The initial time step was chosen such that the deterministic part of the displacement of a particle could not be higher than 1% of the average interparticle surface to surface spacing during one integration step. The implemented repulsive potential prevents particles to aggregate or overlap provided that the particles properties (e.g. position, speed) are recomputed at small enough time intervals. The initially chosen time step was reduced if necessary during the simulation run. In typical simulations, the time step is comprised between $10^{-9}$ and $10^{-8}$ seconds, depending on the particle size.

7.2.2 Effective viscosity computation

The mean square displacement of the tracer particle was computed from its space coordinates $x$, $y$, $z$ at each integration step with the following formula:

$$MSD(t) = (x(t) - x_0)^2 + (y(t) - y_0)^2 + (z(t) - z_0)^2$$

(7.6)

where $x_0$, $y_0$ and $z_0$ are the initial coordinates of the tracer particle.

For each case under investigation, the tracer mean square displacement was averaged over 500 simulations, and the tracer diffusion coefficient $D_T$ was estimated from the slope of the averaged mean square displacement as a function of time, as shown in Figure 7.1(b).

$$\overline{MSD}(t) = 6D_T \times t$$

(7.7)

The hindered diffusive mobility of the tracer particle in the colloidal host dispersion was then expressed in terms of an effective viscosity according to:

$$\eta_{eff} = \frac{k_B T}{6 \pi a_T D_T}$$

(7.8)

Simulations were performed in water at the temperature of 298 K, and the solvent viscosity was accordingly set to $\eta_0 = 8.9 \times 10^{-4}$ Pa.s.
The data points and error bars presented in the graphs correspond to the average values and standard deviations, respectively, computed based on at least two batches of simulations, each consisting of 500 runs.

7.3 Results and discussion

7.3.1 Viscosity scaling in concentrated dispersion

7.3.1.1 Reference case

At first, the reference case of a host dispersion consisting of monodisperse particles of 10 nm radius at the occupied volume fraction of 15 % was considered. As an example, the average mean square displacement curve of tracer particles of 10 nm radius is shown in Figure 7.1(b). It is observed that the slope of the mean square displacement is reduced in the host dispersion as compared to the solvent, reflecting the hindered mobility of the tracer particle due to the crowded environment. Moreover, it is worth highlighting that the mean square displacement still increases linearly with time under the range of concentrations investigated in this work, indicating normal diffusion in the concentrated dispersion. By using Equations 7.7 and 7.8, an effective viscosity of \(1.21 \times 10^{-3}\) Pa.s was computed for the situation presented in Figure 7.1(b).

In Figure 7.2(a), the viscosity experienced by a tracer particle in the reference host dispersion \((a_T = 10\ \text{nm}, \ \Phi_0 = 15\ %, \ \sigma_{H}^2 = 0\ \text{nm})\) is plotted as a function of the tracer size (purple circles). It is seen that the effective viscosity is close to the solvent viscosity \((\eta_0 = 8.9 \times 10^{-4}\ \text{Pa.s})\) for a tracer particle of 2 nm radius, and then progressively increases until it reaches a plateau slightly above \(1.2 \times 10^{-3}\) Pa.s for tracer particles larger than 10 nm. This viscosity value is in fairly good agreement with the macroscopic viscosity of a dispersion of weakly interacting spheres computed with the modified Einstein equation \((\eta_\infty = 1.3 \times 10^{-3}\ \text{Pa.s})\) at
the volume fraction of 15 %), which is given by [184]:

\[ \eta_{\infty}(\Phi_0) = \eta_0 \frac{1 + \Phi_0/2}{(1 - \Phi_0)^2} \]  

(7.9)

In the following, we extend this analysis to several host dispersions having different particle sizes, volume fractions and polydispersity values. The cases under investigation are summarized in Table 7.1, where the reference case previously described corresponds to case 1.

<table>
<thead>
<tr>
<th>Case</th>
<th>( a_H [\text{nm}] )</th>
<th>( \Phi_0 [\phi] )</th>
<th>( \sigma^2_H [\text{nm}^2] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 7.1: Summary of the host dispersions considered for the Brownian dynamics simulations.

### 7.3.1.2 Impact of host particle size

The impact of the host particle size was investigated at the constant volume fraction of 15 % by considering host dispersions with particle sizes of 5, 10 and 20 nm (cases 1 to 3 in Table 7.1). Results are presented in Figure 7.2(a), where it is observed that similar plateau values are reached in the three cases under investigation, which correspond to the macroscopic viscosity of a dispersion at 15 % volume fraction. The curves obtained with \( a_H = 5 \text{ nm} \) and \( a_H = 20 \text{ nm} \) are shifted to the left and to the right, respectively, as compared to the reference case of \( a_H = 10 \text{ nm} \), showing that the crossover from the nano- to the macroviscosity occurs at tracer particle sizes comparable to the size of the host particles.

This is further evidenced in Figure 7.2(b), where it is shown that the viscosity results obtained with different host particle sizes collapse on one single curve when
Chapter 7. Viscosity scaling

Figure 7.2: Impact of host particle size (cases 1, 2, 3 of Table 7.1). (a) Effective viscosity as a function of tracer size for monodisperse host dispersions with various particle sizes (5, 10 and 20 nm in radius) at the constant occupied volume fraction of 15%. (b) Same data as (a) plotted as a function of the size ratio between the tracer and host particles (see Equation 7.10) showing curve overlapping.

the tracer particle size is normalized by the host particle size, i.e. when the data are plotted as a function of the size ratio $sr$ defined as:

$$sr = \frac{a_T}{a_H}$$

(7.10)

7.3.1.3 Impact of volume fraction

In the following, a host dispersion of monodisperse particles of 10 nm radius is considered, and the volume fraction is varied (cases 1, 4, 5 in Table 7.1). The increase in the effective viscosity with the tracer size is plotted in Figure 7.3(a) for three volume fractions: 10, 15 and 20%. As expected, the macroscopic viscosity reached at large tracer sizes increases with the occupied volume fraction. Nevertheless, the trends of the increase in the effective viscosity with the tracer particle size appear similar for the three volume fractions under examination.

With a view to identifying a general scaling of the viscosity, we introduce a normalized viscosity defined as:

$$\eta_{norm}(a_T) = \frac{\eta_{eff}(a_T, \Phi_0) - \eta_0}{\eta_{\infty}(\Phi_0) - \eta_0}$$

(7.11)

where $\eta_0$ and $\eta_{\infty}$ are the solvent and macroscopic viscosities, respectively.
Figure 7.3: Impact of volume fraction (cases 1, 4, 5 of Table 7.1). (a) Effective viscosity as a function of tracer size for monodisperse host dispersions with particle size of 10 nm radius, at various volume fractions (10, 15 and 20 %). (b) Same data as (a) plotted in terms of normalized viscosity (see Equation 7.11) showing curve overlapping.

In Figure 7.3(b), it is seen that the results obtained with three different volume fractions collapse on one single curve when the data are plotted in terms of the normalized viscosity.

7.3.1.4 Impact of host particle polydispersity

Finally, the impact of the polydispersity of the host particle size distribution was studied at the volume fraction of 15 % and mean host particle size of $a_H = 10$ nm (cases 1 and 6 in Table 7.1). To do so, a Gaussian distribution with variance $\sigma^2_H = 5$ nm$^2$ was considered, as represented in Figure 7.4(a). The increase in effective viscosity with tracer particle size is plotted in Figure 7.4(b), where it appears that the host particle polydispersity leads to smaller viscosity values. Polydispersity is indeed known to decrease the viscosity of colloidal dispersions at constant occupied volume fraction, as it increases the maximum packing volume fraction [190–193].

As for simulations performed at various volume fractions, it is observed that the two sets of data collapse on one single curve when the results are plotted in terms of the normalized viscosity defined in Equation 7.11, which eliminates the differences due to the distinct macroscopic viscosities (figure not shown).
7.3.1.5 Master curve of the effective viscosity

In summary, it was shown that:

(i) curves of the effective viscosity as a function of the tracer particle size obtained with different host particle sizes tend to collapse on one single curve when plotted as a function of the size ratio between the tracer and host particles (defined in Equation 7.10);

(ii) data obtained with host dispersions characterized by distinct macroscopic viscosities (due to the different volume fractions or to the distribution polydispersity) approximately fall on one single curve when the viscosity is normalized between 0 and 1 according to Equation 7.11.

In Figure 7.5, we consider all the data points acquired in this study and plot the normalized viscosity as a function of the ratio between the tracer particle size and the mean host particle size. A master curve is obtained, starting at $\eta_{norm} = 0$ (i.e. $\eta_{eff} = \eta_0$) for very small size ratios, and then increasing almost linearly with the size ratio until reaching $\eta_{norm} = 1$ (i.e. $\eta_{eff} = \eta_{\infty}$) at size ratios larger than 1.

The following empirical expression is proposed to describe the master curve:

$$\eta_{norm}(sr) = 1 - \exp \left( -\alpha \times (sr)^\beta \right)$$  \hspace{1cm} (7.12)
By fitting the Brownian dynamics simulation data, we obtain the following values of the parameters: $\alpha = \beta = 2$

### 7.3.2 Interplay between viscosity and aggregation

In the following, we use the viscosity scaling shown in Equation 7.12 to characterize the diffusive mobility of a fractal cluster immersed in an aggregating colloidal system. In this frame, we aim at gaining understanding on the interplay between aggregate formation and viscosity increase in concentrated colloidal systems.

#### 7.3.2.1 Impact of aggregation on the macroscopic viscosity of colloidal dispersions

In the absence of coalescence, aggregation leads to the formation of porous fractal clusters, which occupy more space than their single constituting primary particles. This leads to an increase in the occupied volume fraction, which is defined as:

$$\Phi(t) = \sum_i M_i(t) \frac{4}{3} \pi R_i^3$$

(7.13)

where $M_i$ and $R_i$ are the number concentration and radius, respectively, of clusters constituted of $i$ primary particles. Note that the choice of the radius definition (gyration, hydrodynamic, collision...) will affect the value of the occupied volume fraction.
fraction. In this chapter, \( R_i \) was simply computed as \( R_i = i^{1/d_f}/k_f \). At initial time, \( \Phi(0) \equiv \Phi_0 = M_0 \times (4/3)\pi R_p^3 \), where \( R_p \) and \( M_0 \) are the primary particle radius and the initial particle number concentration, respectively.

This increase in the occupied volume fraction results in an increase in the macroscopic solution viscosity, which can be described by Equation 7.9 assuming that the aggregate polydispersity has a limited impact on \( \eta_\infty \).

### 7.3.2.2 Impact of viscosity on the aggregation kernel

It is essential to quantify the effective viscosity experienced by the aggregates of various sizes, as it directly affects the cluster diffusive mobility and therefore the aggregation rate. To this end, we extend the empirical expression presented in Equation 7.12, which was obtained with spherical particles, to the case of fractal aggregates by: (i) making use of the fractal scaling to relate the aggregate radius with the aggregate mass through the fractal dimension \( d_f \); (ii) describing the host dispersion by the weight-average mass of the aggregate distribution \( n_W \). It is worth recalling that it was shown in the previous section that the polydispersity of the host distribution affects only the absolute values of the viscosity, but does not impact the scaling of the normalized viscosity (Figure 7.5 and Equation 7.12). Accordingly, we compute the normalized viscosity experienced by an aggregate of size \( i \) immersed in an aggregate population of weight-average mass \( n_W \) as:

\[
\eta_{\text{norm}}(sr_i, t) = 1 - \exp \left( -2 \times \left( \frac{i}{n_W(t)} \right)^{2/d_f} \right) \tag{7.14}
\]

where the size ratio \( sr_i \) is defined as:

\[
sr_i(t) = \left( \frac{i}{n_W(t)} \right)^{1/d_f} \tag{7.15}
\]

This normalized viscosity is easily rescaled in an effective viscosity by using Equation 7.11, where the time dependent macroscopic viscosity can be computed from the estimation of the occupied volume fraction (Equations 7.9 and 7.13).
In this section, we focus on the effect of viscosity on the reduction of the aggregation rate constant. For that purpose, we consider the kernels classically used to describe diffusion-limited cluster aggregation (DLCA) and reaction-limited cluster aggregation (RLCA).

In the case where aggregation is limited by diffusion, the aggregation rate constant results from two contributions: the diffusive mobility of the aggregates and their collision cross section. The aggregation rate constant between two clusters of size $i$ and $j$ under dilute conditions is then given by [85]:

$$k_{i,j} = \frac{8k_BT}{3\eta_0} \times \frac{1}{4} \left( i^{-1/d_f} + j^{-1/d_f} \right) \times \left( i^{1/d_f} + j^{1/d_f} \right) \times \left( \text{aggregate diffusion} \right) \times \left( \text{collision section} \right) \tag{7.16}$$

In order to include viscosity effects occurring at high concentration, we propose to replace the solvent viscosity appearing in the diffusion term of the classical kernel by the effective viscosity that the aggregates of various sizes experience:

$$k_{i,j}(t) = \frac{8k_BT}{3} \times \frac{1}{4} \left( \frac{i^{-1/d_f}}{\eta_{eff}(sr_i, \Phi)} + \frac{j^{-1/d_f}}{\eta_{eff}(sr_j, \Phi)} \right) \times \left( i^{1/d_f} + j^{1/d_f} \right) \tag{7.17}$$

Note that both $sr_i$ and $\Phi$ are functions of time, which results in a time-dependent aggregation kernel.

The impact of introducing the effective viscosity in the DLCA kernel can be appreciated in Figure 7.6(a) and (b), where values of the aggregation rate constant are shown as a function of the size of the two colliding clusters for both the classical and modified kernels. Since the modified kernel depends on the average aggregate size and on the macroscopic viscosity, values of aggregation rate constants are shown in Figure 7.6 only for a representative example, considering $d_f = 1.85$, $n_W = 500$, and $\eta_\infty = 5\eta_0$. 
Figure 7.6: Impact of viscosity on the aggregation rate constants under DLCA conditions. For the sake of clarity, aggregation rate constants have been normalized by $k_s$, which is defined in Equation 7.18. (a) Classical DLCA kernel without viscosity correction. (b) Modified DLCA kernel accounting for viscosity effects, considering the following parameters: $d_f = 1.85$, $u_W = 500$, $\eta_\infty = 5\eta_0$. (c) Rate constant for monomer-cluster aggregation as a function of cluster size for both the classical and modified DLCA kernels, considering the parameters reported in (b). (d) Aggregation rate constant for equally sized clusters as a function of cluster size for both the classical and modified DLCA kernels, considering the parameters reported in (b).

For the sake of clarity, the values of the aggregation rate constants have been normalized by $k_s$, the prefactor in Equation 7.16:

$$k_s = \frac{8k_BT}{3\eta_0}$$  \hspace{1cm} (7.18)

It can be observed that viscosity effects reduce only slightly the rate constant for monomer-cluster aggregation. This result is reasonable, considering that the small size of the monomer allows its fast diffusion towards a cluster, which is larger in size and acts as a ‘target’. The same picture can be appreciated from a mathematical point of view, considering Equation 7.17, where the diffusion term which includes the effective viscosity is dominated by the contribution of the monomer. Since the monomer experiences a viscosity that is close to the solvent viscosity,
very similar values of the rate constant for monomer-cluster aggregation are obtained with or without viscosity corrections (Figure 7.6(e)).

On the other hand, the rate constant for cluster-cluster aggregation is significantly impacted by the increase in the dispersion viscosity. Crucially, while the aggregation rate constant of equally sized clusters is independent of the cluster size when the solvent viscosity is considered, it becomes strongly decreasing with cluster size when introducing the viscosity scaling in the aggregation kernel (Figure 7.6(d)).

Considering now the case where aggregation occurs under reaction-limited conditions, the aggregation kernel is given by [85]:

Classical RLCA (solvent viscosity):

\[
 k_{i,j} = \frac{8k_BT}{3\eta_0} \times \frac{1}{4} \left( i^{-1/d_f} + j^{-1/d_f} \right) \times \left( i^{1/d_f} + j^{1/d_f} \right) \times \frac{(ij)^\lambda}{W} \tag{7.19}
\]

where \( W \) is the so-called Fuchs stability ratio, which quantifies the energy barrier that the colliding particles must overcome before aggregating, and \( \lambda \) is an exponent in the order of \( 1 - 1/d_f \) [85].

Following the approach used under DLCA conditions, we propose to replace the solvent viscosity appearing in the diffusion term by the effective size dependent viscosity experienced by the aggregates.

Modified RLCA (viscosity scaling):

\[
 k_{i,j}(t) = \frac{8k_BT}{3} \left( \frac{i^{-1/d_f}}{\eta_{eff}(sr_i, \Phi)} + \frac{j^{-1/d_f}}{\eta_{eff}(sr_j, \Phi)} \right) \left( i^{1/d_f} + j^{1/d_f} \right) \frac{(ij)^\lambda}{W} \tag{7.20}
\]

The impact of this modification on the aggregation kernel can be appreciated in Figure 7.7. It is worth noticing that, just as in the DLCA case, the introduction of an effective viscosity in the aggregation rate constant has only a slight impact on the rate constant for monomer-cluster aggregation, but strongly reduces the rate constant for cluster-cluster aggregation.
Chapter 7. Viscosity scaling

Figure 7.7: Impact of viscosity on the aggregation rate constants under RLCA conditions. For the sake of clarity, aggregation rate constants have been normalized by $k_s \times W$, where $k_s$ is defined in Equation 7.18. (a) Classical RLCA kernel without viscosity correction. (b) Modified RLCA kernel accounting for viscosity effects, considering the following parameters: $d_f = 2.05$, $\lambda = 0.5$, $n_W = 500$, $\eta_\infty = 5\eta_0$. (c) Rate constant for monomer-cluster aggregation as a function of cluster size for both the classical and modified RLCA kernels, considering the parameters reported in (b). (d) Aggregation rate constant for equally sized clusters as a function of cluster size for both the classical and modified RLCA kernels, considering the parameters reported in (b).

However, it must be emphasized that in the RLCA case, even though the absolute aggregation rate constants between equally sized clusters are reduced when accounting for the viscosity increase, the relative trend is unchanged. Indeed, the aggregation rate constant increases almost linearly with cluster size, both in the cases where the solvent viscosity and the effective viscosity are considered (Figure 7.7(d)). In the DLCA case instead, it was found that viscosity effects suppress the size-independence of the aggregation rate constant between equally sized clusters, and induce a decrease of $k_{i,i}$ with cluster size (Figure 7.7(d)).
7.3.2.3 Role of viscosity in the time evolution of the cluster mass distribution

The reduction in the aggregation rate constant previously described will eventually lead to a delay in the kinetics of aggregate growth. This effect can be evaluated by performing population balance equations (PBE) simulations, i.e. by solving the following set of ordinary differential equations (ODE):

\[
\frac{dM_i(t)}{dt} = \frac{1}{2} \sum_{j=1}^{i-1} k_{j,i-j}(t) M_j(t) M_{i-j}(t) - M_i(t) \sum_{j=1}^{\infty} k_{i,j}(t) M_j(t)
\]  

(7.21)

The description of aggregation phenomena in colloidal systems typically requires to cover several orders of magnitudes in aggregate size, which results in a large number of ODE to be solved. In order to reduce the computational time, a discretization technique based on Gaussian basis functions has been employed [140, 141], as in Chapter 3. Some details about this numerical method are given in Appendix E.

In what follows, we report successively PBE simulations performed under diffusion-limited and under reaction-limited conditions. In this frame, it is convenient to introduce the characteristic time for rapid aggregation, defined as:

\[
t_{RC} = \frac{2}{k_s M_0}
\]

(7.22)

In Figure 7.8(a), the increase in the average aggregate size simulated by using the modified DLCA kernel (Equation 7.17) is shown (red dashed line). This situation corresponds to a time and size dependent effective viscosity, and is compared to two limiting cases. In the first limiting case, it is assumed that the viscosity increase during aggregation can be neglected. This situation corresponds to the classical DLCA kernel shown in Equation 7.16, where the solvent viscosity is considered (black solid line). In the second limiting case instead, it is assumed that all the aggregates experience the macroscopic viscosity, whose increase with time is computed from the occupied volume fraction by using Equations 7.9 and 7.13.
In other words, the solvent viscosity appearing in Equation 7.16 is replaced by the time dependent macroscopic viscosity (blue dotted line).

As expected, the rise in the average aggregate size computed by using the modified aggregation kernel is comprised between the two aforementioned limiting cases. Indeed, in the case of the modified aggregation kernel, the aggregates experience an effective viscosity which is comprised between the solvent and the

![Figure 7.8](image)

**Figure 7.8**: PBE simulations performed under DLCA conditions by considering the cases where the aggregates experience: (i) the solvent viscosity ($\eta = \eta_0$), (ii) the time and size dependent effective viscosity ($\eta = \eta_{eff}$), (iii) the time dependent macroscopic viscosity ($\eta = \eta_\infty$). (a) Weight-average aggregate size as a function of the normalized time $\tau = t/t_{RC}$, where $t_{RC}$ is the characteristic time for rapid coagulation defined in Equation 7.22. (b) Polydispersity index as a function of aggregate size. (c) Cluster mass distribution at the average aggregate size $n_W = 500$. The simulations were performed in water at 298 K with the following parameter values: $d_f = 1.85$, $t_{RC} = 6.8 \times 10^{-5}$ s, $\Phi_0 = 1 \%$. 
macroscopic viscosity. While aggregates larger than the average aggregate size experience the macroscopic viscosity, smaller aggregates experience a lower viscosity, whose magnitude depends on the aggregate size.

In Figure 7.8(b), the impact of viscosity effects on the polydispersity index (PDI) of the aggregate distribution is assessed as a function of the average aggregate size. Interestingly, it is observed that in the two limiting cases where all the aggregates experience the same viscosity (either the solvent or the macroscopic viscosity), the PDI increases first with increasing aggregate sizes, and then reaches a constant value slightly below 2. Strikingly, in the case where a time and size dependent effective viscosity is considered, the PDI decreases above a certain average size. Therefore, the self-preserving behavior of cluster mass distributions typically observed under DLCA conditions [129] turns into a self-sharpening behavior when accounting for the viscosity scaling.

This effect is further highlighted in Figure 7.8(c), where aggregate distributions characterized by the same average aggregate size ($n_W = 500$) are shown for the three cases under investigation. Note that these three distributions were obtained at three different simulation times, due to differences in the kinetics of aggregation of the considered systems. It is striking that the aggregate distribution obtained by using the viscosity scaling is narrower as compared to the distributions obtained in the two limiting cases, where all the aggregates feel the same viscosity. This self-sharpening effect of the aggregate distribution is rooted in the inhibition of aggregate growth by cluster-cluster aggregation as compared to aggregate growth by monomer-cluster aggregation when accounting for viscosity scaling.

Finally, PBE simulations performed under RLCA conditions are shown in Figure 7.9 for the same three cases as those previously described (i.e. the aggregates experience either the solvent, effective, or macroscopic viscosity). Once more it can be observed that the simulation of the increase in the average aggregate size performed with the modified kernel lies in between the two limiting cases, where all the aggregates experience either the solvent or the macroscopic viscosity (see Figure 7.9(a)).
Figure 7.9: PBE simulations performed under RLCA conditions by considering the cases where the aggregates experience: (i) the solvent viscosity ($\eta = \eta_0$), (ii) the time and size dependent effective viscosity ($\eta = \eta_{\text{eff}}$), (iii) the time dependent macroscopic viscosity ($\eta = \eta_{\infty}$). (a) Weight-average aggregate size as a function of the normalized time $\tau' = t/t_{RC}/W$, where $t_{RC}$ is the characteristic time for rapid coagulation defined in Equation 7.22. (b) Polydispersity index as a function of aggregate size. (c) Cluster mass distribution at the average aggregate size $n_W = 500$. The simulations were performed in water at 298 K with: $d_f = 2.05$, $\lambda = 0.5$, $t_{RC} = 6.8 \times 10^{-5}$ s, $\Phi_0 = 1\%$.

As previously observed under diffusion-limited conditions, introducing the viscosity scaling in the aggregation kernel leads to smaller PDI values (Figure 7.9(b)). However, this effect is much less pronounced in the RLCA regime, as further evidenced in Figure 7.9(c). Indeed, aggregate distributions characterized by the same average size are found to almost overlap for the three cases under investigation.
This difference observed between the DLCA and RLCA regimes can be better understood from the plots of the aggregation kernels shown in Figure 7.6 and Figure 7.7. It is seen that the trend of the aggregation rate constant with the aggregate size in the RLCA case is similar for both the classical and modified kernels: the aggregation rate constant increases almost linearly with the aggregate size, just the slope is reduced in the case of the modified kernel (Figure 7.7(d)). In the DLCA case instead, the trend of the aggregation rate constant with the aggregate size is altered upon introduction of the viscosity scaling in the kernel: the aggregation rate constant is independent of the aggregate size when the solvent viscosity is considered, while it decreases with the aggregate size when accounting for the viscosity scaling (Figure 7.6(d)). This modification in the ranking of aggregate reactivity introduces self-sharpening of the aggregate distribution.

It is important to point out that the role played by the viscosity, leading to a reduction in the overall aggregation rate, is only one of the peculiar effects appearing at high concentration. The fact that, in average, particles and clusters need to diffuse over shorter distances is typically leading to an increase in the overall rate of aggregation, at least under diffusion-limited conditions, as several simulation works indicate [213–215]. Nevertheless, it is interesting to note that the introduction of the viscosity dependence on the cluster size leads to a change in the aggregate size distribution, which should be compared with results from both simulation and experimental studies.

7.4 Conclusion

The effective viscosity experienced by a spherical tracer particle immersed in colloidal dispersions of spherical particles was quantified by means of Brownian dynamics simulations. It was found that the crossover between the nano- and macro-viscosity occurs at a tracer particle size comparable to the average size of the host particles. For tracer particle sizes smaller than the average size of the host particles, the effective viscosity is found to increase almost linearly with the tracer
size from the solvent to the dispersion macroscopic viscosity. For tracer particles larger than the average host particle size, the effective viscosity reaches a plateau corresponding to the macroscopic viscosity of the dispersion.

It was shown that curves obtained with different host particle sizes, distinct occupied volume fractions and with a polydisperse host particle size distribution all collapse on one single master curve, when the normalized effective viscosity is plotted as a function of the ratio between the tracer particle size and the mean host particle size.

Finally, these results were applied to characterize the size dependent effective viscosity experienced by a fractal aggregate in a colloidal system undergoing aggregation. We proposed a modification of the classical DLCA and RLCA kernels to account for the scaling of the effective viscosity experienced by the aggregates of different sizes. It was found that the viscosity scaling has a major impact on the aggregation process. In particular, viscosity reduces significantly the rate of cluster-cluster aggregation as compared to the rate of monomer-cluster aggregation. As a consequence, viscosity does not only slow down the aggregation process, but also changes the shape of the cluster mass distribution. Indeed, it was shown by means of PBE simulations that the scaling of effective viscosity with aggregate size leads to the formation of narrower distributions as compared to the case where the solvent viscosity is considered. This sharpening behavior was found to be considerably more pronounced under DLCA than under RLCA conditions.

In the next chapter, results from this chapter (i.e. the theoretical impact of the viscosity scaling on aggregation rates) and from the previous chapter (i.e. the experimental characterization of the increase in the macroscopic viscosity) will be combined to gain understanding on the interplay between aggregate formation and viscosity increase in concentrated antibody solutions. Moreover, additional non-idealities due to protein-protein interactions will be examined.

---

The content of this chapter has been partially published in the following article: L. Nicoud, M. Lattuada, S. Lazzari, M. Morbidelli. Viscosity scaling in concentrated dispersions and its impact on colloidal aggregation. *Phys. Chem. Chem. Phys.*, 17(37):24392-24402, 2015
Chapter 8

Complex phenomena arising at high mAb concentration

8.1 Introduction

Chapters 6 and 7 were dedicated to the study of the interplay between aggregate formation and viscosity increase in concentrated colloidal systems. In this chapter, we go beyond viscosity effects and analyse which additional complex phenomena impact the kinetics of protein aggregation under concentrated conditions.

It is worth pointing out that this topic is of broad interest, and is not limited to the case of mAb aggregation during manufacturing. Improving our knowledge of protein stability in crowded environments is indeed of fundamental relevance in several research fields, such as cell biology [216] or biomedical science [217, 218], and also has practical applications in various industrial sectors, including the production of food products [82], and biomaterials [219].

There is currently a broad consensus that highly concentrated mAb formulations present notable advantages in terms of patience compliance and convenience of delivery (e.g. via subcutaneous route) over more diluted formulations. However, such concentrated formulations give rise to a certain number of issues related to...
the poor protein solubility, the high solution viscosity, and the formation of protein aggregates, which may reduce drug potency or incite unwanted immune-system responses [3, 4]. In addition, the use of highly concentrated drug formulations poses some analytical challenges [4]. For example, dilution to lower concentrations is generally required prior to analysis, what may impact the results of the assays due to the potential reversibility of loose protein aggregates.

Since protein concentration critically affects the kinetics of aggregation, understanding the impact of mAb concentration on mAb stability is a key requisite to optimize final product formulation. Moreover, due to the limited amount of material available during the early-stages of drug development, there is a high interest in assessing formulation stability by performing accelerated studies under dilute conditions. Attempts to extrapolate results from these accelerated studies to conditions relevant for storage require to unravel the role of mAb concentration (as well as temperature or any other accelerating factor) on the kinetics of aggregation [14, 15]. Gaining a deep understanding of protein aggregation mechanisms under concentrated conditions is nevertheless extremely challenging. Indeed, on top of the complex features affecting the protein aggregation behavior under dilute conditions (such as the protein surface non-uniformity, or the strong interconnection between the conformational and colloidal protein stabilities), additional intricate phenomena arise at high protein concentration. For example, multibody interactions [220], excluded volume effects [53, 221], short-range interactions [172], activity coefficients [222], positional correlations [223], as well as ion binding [224] become of significant importance in concentrated systems.

One particular challenge in studying protein aggregation under concentrated conditions is to evaluate the distinct contributions of the various non-ideal effects on the aggregation rate. For example, an increase in protein concentration leads on the one hand to a larger excluded volume (expected to accelerate bimolecular aggregation of aggregation-prone species), and on the other hand to an increase in the solution viscosity (expected to slow down aggregation) [225]. The interplay between these effects makes predictions of the trend of the aggregation rate with the protein concentration extremely difficult.
In this chapter, we combine detailed experimental data and kinetic analysis to investigate the heat-induced aggregation of a model monoclonal antibody from low to high protein concentration. At low protein concentration, we describe the aggregate growth with the classical reaction-limited cluster aggregation kernel, while at high protein concentration, we propose to modify this aggregation rate constant by including two effects: (i) solution non-idealities due to protein-protein interactions, which arise even under stable conditions; and (ii) the increase in solution viscosity during the aggregation process (which has been investigated in Chapter 7). Experiments are carried out with four buffer solutions, with different pH values and salt content, with a view to studying the impact of protein-protein interactions on the aggregation rate.

8.2 Materials and methods

8.2.1 Sample preparation

The mAb used for this study was a glycosylated IgG1, which was denoted as mAb-1 in the previous chapters.

The theoretical isoelectric point (pI) of mAb-1 lies between 8 and 9.2. Moreover, the mAb pI was determined experimentally at UCB Pharma both by isoelectric focusing (using an ICE3 system from ProteinSimple at the protein concentration of 2 g/L) and by pH titration (using the Wyatt Mobius system at the protein concentration of 10 g/L). The two measurements yielded pI values of 8.7-8.8 and 8.1-8.2, respectively.

Prior to aggregation studies, the stock solution was dialyzed against a 20 mM histidine buffer at selected pH (5.5, 6.5 or 7.4) as described in Chapter 2. The samples were then prepared by diluting the dialyzed solution with appropriate buffer solutions to reach the targeted protein concentration (and NaCl concentration in case salt was added) in 20 mM histidine buffers at selected pH values. The pH
range used for this study (5.5-7.4) corresponds to the effective pH range for the histidine buffer.

8.2.2 Isothermal aggregation kinetics

Isothermal aggregation kinetics were performed by incubating antibody samples in hermetically sealed HPLC vials at 70 °C in a block-heater, as previously described in Chapter 2.

8.2.3 Size exclusion chromatography with inline light scattering

Monomer depletion and aggregate growth were monitored by SEC coupled with a multi-angle light scattering detector, as described in Chapter 2.

8.2.4 Aggregate fractal dimension

The aggregate fractal dimension $d_f$ and the scaling prefactor $k_f$ were estimated experimentally from the correlation between the aggregate weight-average molecular weight $\langle MW_{w}^{Agg} \rangle$ and the aggregate average hydrodynamic radius $\langle R_{h}^{Agg} \rangle$ according to:

$$\frac{\langle MW_{w}^{Agg} \rangle}{MW_p} = k_f \left( \frac{\langle R_{h}^{Agg} \rangle}{R_p} \right)^{d_f} \quad (8.1)$$

where $MW_p = 150$ kDa and $R_p = 6$ nm are the molecular weight and the hydrodynamic radius of the monomeric protein, respectively.

The aggregate weight-average molecular weight and the aggregate average hydrodynamic radius were determined from the static and dynamic light scattering measurements, respectively, as further detailed in Chapter 6.
8.2.5 Viscosity measurements

The viscosity increase during protein aggregation has been monitored in situ by measuring the diffusion coefficient of polymeric tracer nanoparticles with dynamic light scattering (DLS). This technique has been introduced and validated previously in Chapter 6.

8.2.6 Protein-protein interactions

8.2.6.1 Dynamic light scattering

Protein-protein interactions were investigated under stable conditions by DLS using a Zetasizer Nano (Malvern, Worcestershire, UK) at the fixed angle of 173°. To do so, the mAb collective diffusion coefficient $D_c$ was measured at various protein concentrations $c$ ranging from 1 to 60 g/L. The interaction parameters $k_D$ and $k'_D$ were then estimated from the following expansion:

$$D_c(c) = D_0(1 + k_Dc + k'_Dc^2)$$ (8.2)

where $D_0$ is the diffusion coefficient at infinite dilution, which can be computed from the Stokes-Einstein equation (see Equation 1.6).

Note that such experiments are typically conducted under dilute conditions (till around 10 g/L) in order to access to the value of the interaction parameter $k_D$. In this study, a higher order term was included in order to extend the description of the diffusion coefficient to the concentrated regime.

Experiments were performed at several temperatures, ranging from 25 to 50 °C, which was determined as the highest temperature where aggregation does not occur during the time frame of the experiment (data not shown). Data presented in the graph correspond to the average values obtained from at least two independent repetitions.
8.2.6.2 Static light scattering

To complement the results obtained by DLS, protein-protein interactions between monomeric species were also assessed with static light scattering (SLS) by using a BI-200SM goniometer (Brookhaven Instruments, Holtsville, NY, USA). A solid-state laser Ventus LP532 (Laser Quantum Manchester, U.K.), with a wavelength of $\lambda_0 = 532$ nm was used as a light source. The scattered intensity $I_s$ was measured at the angle of 90° for a series of samples at various protein concentrations $c$, ranging from 1 to 65 g/L. The excess Rayleigh ratio $R_{90}$ was then computed according to:

$$R_{90} = (I_s - I_0) \frac{R_{tol}}{I_{tol}} \left( \frac{n}{n_{tol}} \right)^2$$

(8.3)

where $n$ and $n_{tol}$ represent the refractive index of water and toluene, equal to $n = 1.333$ and $n_{tol} = 1.496$, respectively. Moreover, $I_0$ and $I_{tol}$ denote the intensity of the solvent and of toluene, respectively. $R_{tol}$ is the Rayleigh ratio of toluene, which was set to $1.98 \times 10^{-5}$ cm$^{-1}$ [226].

Under dilute conditions, the protein-protein interactions were studied through the measure of the second virial coefficient $B$ by constructing a Debye plot:

$$\frac{Kc}{R_{90}} = \frac{1}{MW_{app}} (1 + 2Bc)$$

(8.4)

where $MW_{app}$ is the apparent molecular weight of the protein, $B$ has the unit of a volume divided by a mass, and $K$ is an optical constant defined as:

$$K = 4\pi^2 n^2 \left( \frac{dn}{dc} \right)^2 N_a^{-1} \lambda_0^{-4}$$

(8.5)

where $N_a$ is the Avogadro number and $dn/dc$ is the sample refractive index increment, which was set to 1.90 [227].

Formally, the second virial coefficient is defined as the integral over space of the orientation-averaged protein-protein interaction potential:

$$\overline{B} = -2\pi \int_0^\infty \left( \exp \left( -\frac{V_T(r)}{k_BT} \right) - 1 \right) r^2 dr$$

(8.6)
where $\overline{B} = B \times MW_p/N_a$ has the unit of a volume, $V_T$ is the total protein-protein interaction potential that accounts for various types of contributions (such as hard sphere, electrostatics, Van der Waals, or hydration forces) and $r$ denotes the distance between the centers of mass of proteins. Note that, as compared to Equation 1.1, protein surface heterogeneities (reflected in the orientation dependence of the interaction potential) have already been lumped in Equation 8.6.

Under concentrated conditions, protein-protein interactions were examined following the method proposed by Blanco and co-workers [53]. Briefly, the Kirkwood-Buff integral $\overline{G}$ was defined as the orientation-averaged protein-protein correlation function $\overline{g}(r)$:

$$\overline{G} = 4\pi \int_{0}^{\infty} (\overline{g}(r) - 1) r^2 dr \quad (8.7)$$

In the limit where the protein concentration approaches 0, $\overline{g}(r)$ can be replaced by the expression of $\exp \left( -V_T(r)/(k_B T) \right)$ in the dilute regime, and thus:

$$\overline{G}(c \to 0) = -2\overline{B} \quad (8.8)$$

A local Taylor series expansion was computed over small concentration windows to determine the dependence of $G = \overline{G} \times N_a/MW_p$ on $c$:

$$\frac{R_{00}}{K \times MW_{app}} = c \times (1 + G(c) \times c) \quad (8.9)$$

Protein concentrations ranging from 1 to 10 g/L were used to determine the values of the second virial coefficient $B$ and of the apparent molecular weight $MW_{app}$ from the Debye plot (Equation 8.4). Protein concentrations up to 65 g/L were used to determine $G$ values with the local Taylor series approach (Equation 8.9). The values of $MW_{app}$ determined at low protein concentration for each buffer solution were used in Equation 8.9.

SLS measurements were performed at 50°C, which was determined as the highest temperature where aggregation does not occur during the time frame of the experiment, with at least two independent repetitions per condition.
8.2.7 Aggregation rate constant

At low protein concentration, the classical reaction-limited cluster aggregation kernel introduced in Chapter 1 was used. It is worth recalling that, as compared to the case of traditional polymer colloids, three types of aggregation events were identified: monomer-monomer, monomer-aggregate, and aggregate-aggregate. Accordingly three Fuchs stability ratios were defined in Chapter 2: $W_{11}$, $W_{1j}$, and $W_{ij}$, which describe dimer formation, aggregate growth by monomer addition, and aggregate growth by cluster-cluster aggregation, respectively.

Solution non-idealities under stable conditions

At high mAb concentration, protein-protein interactions influence the rate of protein diffusion, and significant deviations from the Stokes-Einstein value (see Equation 1.6) may be observed, implying that the concentration dependent self-diffusion coefficient must be used in Equation 1.5 instead of $D_0$. By using Stefan-Maxwell diffusion equations for multicomponent systems, we derived an expression for the self-diffusion coefficient of a monomeric protein solution valid under concentrated conditions [228–230]. As compared to the Fick’s law for diffusion, Stefan-Maxwell equations indeed allow to account for thermodynamic non-idealities as well as for the friction between the diffusing species. Details of the derivation are given in Appendix F, while the final result is reported here:

$$D_s(c) = \frac{RT}{(1 - c\nu)^2} \frac{\eta_0}{\eta_{eff}(1, 0)} \left( \frac{\partial \Pi}{\partial c} \right)^{-1} \times D_c(c) \quad (8.10)$$

where $\nu$ denotes the protein partial specific volume, so that $c\nu$ is equal to the volume fraction occupied by the protein. Moreover, $\eta_{eff}(1, 0)$ is the effective viscosity experienced by the monomeric protein molecules at time 0, in agreement with the notation used in Chapter 7.

Interestingly, Equation 8.10 relates the collective diffusion coefficient $D_c$ (which quantifies the diffusion of solutes under a concentration gradient) to the self-diffusion coefficient $D_s$ (which corresponds to the random motion of one labeled
molecule) [185]. Under dilute conditions, the collective and self-diffusion coefficient are equal to the Stokes-Einstein diffusion coefficient given by Equation 1.6. However, under concentrated conditions, non-idealities arise, implying that \( D_c \) and \( D_s \) differ from \( D_0 \) [185].

The collective diffusion coefficient can be computed from Equation 8.2, while the derivative of the osmotic pressure with respect to the protein concentration is given by [55]:

\[
\frac{\partial \Pi}{\partial c} = \frac{RT}{1 + cG(c)}
\]  

(8.11)

It results from Equations 8.2, 8.10 and 8.11 that:

\[
D_s(c) = D_0 \times \alpha_{PI}(c)
\]  

(8.12)

with:

\[
\alpha_{PI}(c) = \frac{(1 + k_D c + k'_D c^2)(1 + G(c) \times c)}{(1 - \alpha)^2} \frac{\eta_0}{\eta_{eff}(1,0)}
\]  

(8.13)

where the subscript PI stands for ‘Protein Interactions’.

**Viscosity increase during aggregation**

When aggregation occurs under concentrated conditions, the diffusion of clusters is affected not only by protein-protein interactions, but also by the increase in the macroscopic viscosity of the solution with time, which slows down aggregation.

The effective viscosity experienced by a given aggregate depends both on the time as well as on the aggregate size, as described in Chapter 7. The concept of scale dependent viscosity is illustrated here in Figure 8.1. Briefly, aggregates much larger than the average aggregate size experience the macroscopic solution viscosity (\( \eta_\infty \), which increases with time), while aggregates much smaller than the average aggregate size experience a viscosity close to that of the solvent (\( \eta_0 \)). The aggregates of intermediate sizes feel an effective viscosity comprised between \( \eta_0 \) and \( \eta_\infty \), whose mathematical expression is given by Equation 7.14.
The experimental values of the macroscopic viscosity, estimated from the diffusion coefficient of tracer particles with DLS, were given as inputs for the model simulations. To do so, the time evolution of the increase in the macroscopic viscosity was fitted for each investigated condition by the following empirical function:

\[ \eta_\infty(t) = \eta_0 \times \left( \exp \left( \frac{t}{C_1} - C_2 \right) + C_3 \right) \]  

(8.14)

where \( C_1, C_2 \) and \( C_3 \) are three positive real numbers reported in Appendix C.

**Proposed aggregation kernel**

The aggregation rate constant \( k_{i,j} \) given in Equation 1.7 for the dilute case can now be modified to account for solution non-idealities and for the increase in the solution viscosity during aggregation.

To do so, it is necessary to discuss the size dependence of the correction factor introduced in Equation 8.13. It has been shown in Chapter 6 with DLS experiments that the interaction coefficient \( k_D \) is independent of the aggregate size. This implies that \( \overline{B_\text{Agg}}/\overline{B} = MW_\text{Agg}/MW_p \), where \( \overline{B_\text{Agg}} \) and \( MW_\text{Agg} \) denote the second virial coefficient and the aggregate molecular weight of the aggregates, respectively (Equation 6.15). Since \( B = \overline{B} \times N_a/MW_p \) and \( B_{\text{Agg}} = \overline{B_\text{Agg}} \times N_a/MW_\text{Agg} \), it results that \( B_{\text{Agg}} = B \), implying that the mass second virial coefficient is independent of the aggregate size (at least in the case under investigation). Since \( G \) and \( B \) are equivalent under dilute conditions (Equation 8.8), it is assumed
that \( G \) is independent of the aggregate size. On the other hand, the effective viscosity has been shown to depend on the aggregate size in Chapter 7. Therefore, the self-diffusion coefficient of an aggregate of size \( i \) can be written as \( D_{s,i} = D_{0,i} \times \alpha_{PI} \times \eta_{eff}(1,0)/\eta_{eff}(i,t) \). One easily verifies that the latter equation simplifies to Equation 8.12 in the case of a stable monomeric protein solution.

Finally, by introducing the self-diffusion coefficient in the diffusion term of the classical RLCA kernel, the following aggregation kernel is proposed to describe the aggregation kinetics at high protein concentration:

\[
k_{i,j}(t) = \alpha_{PI} \frac{k_s}{4} \left( i^{-1/df} \frac{\eta_{eff}(1,0)}{\eta_{eff}(i,t)} + j^{-1/df} \frac{\eta_{eff}(1,0)}{\eta_{eff}(j,t)} \right) \left( i^{1/df} + j^{1/df} \right) \frac{(ij)^{\lambda}}{W_{i,j}}
\]

where \( \alpha_{PI}, \eta_{eff} \) and \( k_s \) are given by Equations 8.13, 7.14, and 1.8, respectively.

It can be noted that, at low protein concentration, \( \alpha_{PI} \) tends to 1 and the solution viscosity tends to the solvent viscosity, so that Equation 8.15 becomes equivalent to the classical RLCA kernel introduced in Chapter 1.

### 8.3 Results

Aggregation experiments were performed at 70 °C in a 20 mM histidine buffer. In order to vary the protein net charge, experiments were carried out at three different pH values (5.5, 6.5 and 7.4) below the protein pI. Moreover, in the case of pH 5.5, an additional experiment was performed in the presence of 10 mM NaCl with a view to assessing the impact of charge screening on the aggregation kinetics.

For each of the investigated conditions, experiments were performed at four different mAb concentrations (1, 20, 40 and 60 g/L) and analyzed in the frame of a kinetic model. We report first the results obtained under dilute conditions (i.e. at protein concentration of 1 g/L), and then the results obtained under concentrated conditions (i.e. in the range 20-60 g/L).
8.3.1 Aggregation under dilute conditions

The time evolution of the concentrations of monomer, dimer and trimer followed
by SEC are shown in Figure 8.2(a-c) for the four buffer conditions under investigation. Moreover, the increase in the average aggregate molecular weight monitored by inline light scattering is presented in Figure 8.2(d). It is observed that the aggregation kinetics of mAb-1 is strongly affected by the solution conditions. In particular, aggregate growth is found to be much faster at pH 7.4 as compared to the other conditions (Figure 8.2(d)).

![Figure 8.2: Kinetics of aggregation of mAb-1 at protein concentration of 1 g/L under various buffer conditions, as indicated in the legend. Model simulations (lines) are compared to experimental data (symbols) in terms of (a) monomer depletion, (b) dimer concentration, (c) trimer concentration, and (d) aggregate weight-average molecular weight.](image)

In order to quantify the impact of buffer conditions on the elementary steps that contribute to the global aggregation rate (i.e. protein unfolding and aggregation events), these experimental results were interpreted within the kinetic scheme...
presented in Figure 2.4. The reaction scheme includes an unfolding event, which has been shown to be the rate-limiting step for monomer depletion in Chapter 2, followed by irreversible aggregate formation. Under dilute conditions, aggregate growth is described by using the aggregation kernel shown in Equation 2.6.

The implementation of this kinetic model required the estimation of some parameters, which were determined as follows:

- the fractal dimension $d_f$ was assessed from light scattering experiments as shown in Figure 8.3(a),
- the exponent $\lambda$ in the aggregation kernel was approximated by $\lambda = 1 - 1/d_f$,
- the remaining parameters (i.e., $k_U$, $W_{11}$, $W_{ij}$) were estimated from the fitting of model simulations (see population balance equations shown in Equation 2.14) to the set of experimental data shown in Figure 8.2.

The estimated parameter values are reported in Figure 8.3(b-d), whereas the agreement between model simulations and experimental results can be appreciated in Figure 8.2.

In Figure 8.3(b), it is observed that the aggregate fractal dimension increases when the solution pH is increased towards the pI of the mAb, thus indicating that more compact aggregates are formed when the protein net charge is reduced. Interestingly, the addition of 10 mM NaCl at pH 5.5 also leads to an increase in the aggregate fractal dimension. These data suggest that a reduction in the electrostatic repulsion (induced either by a change in pH, or by charge screening due to the addition of salt) leads to the formation of denser aggregates, possibly due to the fact that the interpenetration of clusters is favored when the electrostatic barrier is reduced.

The values of the unfolding rate constant, which quantifies the protein conformational stability, are shown in Figure 8.3(c) as a function of the solution pH. It is seen that the highest value of the unfolding rate constant is obtained at pH 5.5.
In order to interpret these results, it is worth recalling that the kinetics of protein unfolding is strongly affected by protein intramolecular and protein-solvent interactions. In particular, it is expected that the protein structure is destabilized at pH values far removed from the mAb pI due to charge repulsion within the protein molecule [10], as it is observed at pH 5.5. In addition to these non-specific electrostatic effects, specific charge interactions such as ion pairing between oppositely charged residues can also affect the protein stability [10]. These specific electrostatic effects may lead to a stabilization of the folded state at higher protein charge density. This might explain why $k_U$ is slightly lower at pH 6.5 as compared to pH 7.4, and why it is slightly larger when charge screening is induced by the addition of 10 mM NaCl at pH 5.5.
While the protein conformational stability is reflected by the value of the unfolding rate constant, the protein colloidal stability is described by the value of the Fuchs stability ratio, which quantifies the energy barrier that protein molecules need to overcome before aggregating. As explained in Chapter 2, three values of the Fuchs stability ratio are required to describe the aggregation of mAb-1: $W_{11}$, $W_{ij}$ and $W_{ij}$, which characterize monomer-monomer, monomer-cluster and cluster-cluster aggregation, respectively. It is observed in Figure 8.3(d) that the ranking between the three $W$ values is similar for all the investigated conditions, namely $W_{11} < W_{ij} < W_{ij}$. This ranking can be explained by considering the high reactivity of the unfolded monomer as compared to the aggregates, whose aggregation-prone regions are already partially covered. It is worth noticing that the concentration of dimer is significantly lower at pH 6.5 and 7.4 as compared to pH 5.5 (Figure 8.2(b)), suggesting a higher reactivity of the dimer species at higher pH values. This observation was reflected in the kinetic model by setting the value of $W_{12}$ equal to the value of $W_{11}$ (instead of $W_{ij}$) at pH 6.5 and 7.4.

In Figure 8.3(d) it is seen that the three Fuchs stability ratio values follow a monotonic decrease with the solution pH. This decrease in colloidal stability can be attributed to the decrease of the protein net charge when the solution pH is increased towards the pI of the protein. Moreover, it is observed that the presence of 10 mM NaCl leads to lower $W$ values at pH 5.5 as compared to the situation in the absence of salt. This effect can be explained by the screening of the protein charge in the presence of salt, as already mentioned in Chapter 4.

To summarize, by combining experimental data with kinetic modeling, we could discriminate the impact of pH on the conformational and colloidal stabilities of the mAb under investigation. While decreasing the pH below the pI destabilizes the protein structure (i.e. promotes monomolecular protein unfolding), it increases protein intermolecular repulsion (i.e. reduces bimolecular protein aggregation). These two competing effects lead to an optimal pH value for which monomer depletion occurs at the slowest rate. It is indeed seen in Figure 8.3(a) that monomer depletion is slower at pH 6.5 as compared to pH 5.5 and 7.4.
8.3.2 Aggregation under concentrated conditions

In this second part, we move progressively towards higher protein concentrations (in the range 20-60 g/L) under the same four buffer solution conditions. As under dilute conditions, the kinetics of aggregation was followed experimentally by SEC-MALS. The results of the monomer depletion kinetics are presented in Figure 8.4.

![Figure 8.4: Monomer depletion kinetics of mAb-1 at various protein concentrations and in several solution conditions. Experimental data (symbols) are compared to model simulations (lines). The results have been normalized by the initial protein concentration.](image)

To allow a better assessment of the impact of protein concentration on the rate of monomer depletion, the data were normalized by the initial protein concentration. Since monomer depletion has been shown to be rate-limited by monomeric protein unfolding, these normalized data are expected to overlap on one single curve for each of the investigated solution conditions. It can be observed that the concentration profile of monomer obtained at 1 g/L is systematically above those
obtained at larger protein concentrations. This can be attributed to the accumulation of $U$ at low protein concentration (see Figure 2.5). SEC analysis indeed gives access only to the sum of the concentrations of the native and unfolded conformations. When the protein concentration increases, the concentration of $U$ becomes negligible, since its disappearance by dimer formation (bimolecular process) is accelerated as compared to its formation by protein unfolding (monomolecular process). Curve overlapping at high protein concentration is observed at pH 5.5 (both in the absence and in the presence of salt), while slight deviations from this ideal behavior are observed at pH 6.5 and 7.4. Possible reasons for this observation will be discussed later on.

The increase of the weight-average aggregate molecular weight with time is shown in Figure 8.5 for the four solution conditions. It is observed that an increase in protein concentration significantly accelerates the rate of aggregate growth.

**Figure 8.5:** Aggregate growth kinetics of mAb-1 at various protein concentrations and in several solution conditions. Experimental data (symbols) are compared to model simulations (lines).
In addition to monitoring the kinetics of monomer depletion and aggregate growth, the increase in the solution viscosity was followed at high protein concentration. To do so, the diffusion coefficient of tracer nanoparticles was measured with DLS, as described in Chapter 6. Results are presented in Figure 8.6, where it can be seen that an increase in protein concentration significantly speeds up the rate of viscosity increase. This effect is attributed to the increase of the occupied volume fraction, as demonstrated in Chapter 6.

![Figure 8.6: Kinetics of the increase in solution viscosity at various protein concentrations and in several buffer solutions. Symbols correspond to experimental data and lines to fittings with Equation 8.14.](image)

Then, similarly to the method used at low protein concentration, we analyzed these data in the frame of the kinetic mechanism presented in Figure 2.4. Nevertheless, as compared to the dilute case, the aggregation kernel shown in Equation 8.15 was used. This kernel includes two effects arising at high protein concentration: solution non-idealities due to protein-protein interactions, and the increase in solution viscosity due to aggregate formation.
The model parameters were estimated as follows:

- the fractal dimension \( d_f \) was assessed by light scattering experiments from the mass scaling between aggregate mass and aggregate size shown in Equation 8.1,

- the exponent \( \lambda \) in the aggregation kernel was approximated by \( \lambda = 1 - 1/d_f \),

- the increase in the solution macroscopic viscosity shown in Figure 8.6 was used to compute the effective viscosity appearing in the aggregation kernel shown in Equation 8.15,

- the factor \( \alpha_{PI} \) that accounts for protein-protein interactions was evaluated at each mAb concentration and at each buffer conditions by using DLS (to evaluate \( k_D \) and \( k_D' \)) and SLS (to evaluate \( G \)) according to Equation 8.13,

- the unfolding rate constant \( k_U \), which characterizes a monomolecular event, is expected to be independent of protein concentration and was thus set at the value estimated under dilute conditions (shown in Figure 8.3(c)),

- the Fuchs stability ratios \( W_{11}, W_{1j} \) and \( W_{ij} \) were set to the values evaluated under dilute conditions (shown in Figure 8.3(d)). This approximation relies on the assumption that the energy barrier that colliding protein molecules must overcome to aggregate once they are at very short distances is independent of the protein concentration.

It is worth stressing that with this approach, there are no fitting parameters. All model parameters are indeed estimated experimentally or assumed to be equal to those assessed under dilute conditions.

The aggregate fractal dimension is plotted as a function of protein concentration in Figure 8.7(a). It is observed that for all the conditions under investigation, \( d_f \) increases with the protein concentration, indicating that denser aggregates are formed under concentrated conditions. This observation, which was already made in Chapter 6 at pH 6.5 is extended here to three additional solution conditions. One possible reason is that in concentrated systems, the colliding clusters are likely to be partially overlapped, thus favoring contact in the core of the clusters.
rather than at their tips. As in Chapter 6, we compare our experimental results to the square root correlation proposed by Gonzáles et al. [76], which was obtained from Monte Carlo simulations (see dashed line in Figure 8.7(a)). Interestingly, the experimental fractal dimension data follow a similar trend with protein concentration, with values at low protein concentration which depend on the buffer conditions (see Figure 8.3(b)).

![Figure 8.7: (a) Aggregate fractal dimension as a function of protein concentration. The dashed line corresponds to the correlation proposed by Gonzáles et al. [76] (b) Prefactor of the fractal scaling as a function of the fractal dimension. The dashed line corresponds to the correlation proposed by Ehrl et al. [188] Error bars represent the 90 % confidence interval for the parameter estimation from the regression shown in Equation 8.1.](image)

For the sake of completeness, the values of the prefactor $k_f$ obtained by regressing the light scattering data with Equation 8.1 are plotted as a function of the fractal dimension in Figure 8.7(b). It is observed that $k_f$ values are close to unity and tend to decrease with increasing $d_f$ values, in agreement with the empirical correlation proposed by Ehrl et al. [188], which is represented as a dashed line in Figure 8.7(b).

With a view to estimating the parameter $\alpha_{PI}$ defined in Equation 8.13, protein-protein interactions were quantified with light scattering techniques. Measurements were performed at 50 °C, which was identified as the highest temperature where aggregation does not occur in the time frame required for the experiments.

Figure 8.8(a) shows the collective diffusion coefficient measured with DLS as a function of the protein concentration. These data were regressed with Equation 8.2 in order to obtain the interaction parameters $k_D$ and $k_D'$. The so obtained
$k_D$ values are shown in Figure 8.8(b), which also contains data obtained at temperatures lower than 50 °C in order to assess the temperature dependence of the $k_D$ parameter.

It is seen in Figure 8.8(b) that $k_D$ is positive and decreases upon the addition of NaCl. This indicates a decrease of the strength of repulsive interactions due to charge screening. Moreover, it is observed that $k_D$ increases when the solution pH is increased from 5.5 to 7.4.

It is worth highlighting that an increase in the solution pH towards the mAb pI leads on the one hand to a decrease in the protein net charge, and on the other hand to a decrease in the ionic strength of the buffer solutions, which is around 15, 5 and 1 mM at pH 5.5, 6.5 and 7.4, respectively. Indeed, the higher amount of HCl and of positive charges on the histidine molecules lead to a higher ionic strength at acidic pH. It was verified that $k_D$ correlates with the protein net charge (i.e. decreases when the solution pH is increased towards the mAb pI) when the ionic strength is kept constant at 15 mM by adding 10 and 14 mM NaCl at pH 6.5 and 7.4, respectively (see Figure G.1 of Appendix G).

With regards to the second order interaction parameter, slightly negative values of $k'_D$ are found, with absolute values increasing when the pH is increased towards the mAb pI: −0.08, −0.14 and −0.20 at pH 5.5, 6.5, and 7.4, respectively. At pH 5.5 in the presence of salt, $k'_D$ is found to be close to 0.

The results of static light scattering experiments are shown in Figure 8.8(c) where it can be seen that the Rayleigh ratio increases with the protein concentration till around 40 g/L, and then flattens out. In the range from 1 to 10 g/L, these data were used to determine the second virial coefficient and the apparent molecular weight by using Equation 8.4 (see Debye plot in Figure G.2 in Appendix G). It is observed that the interaction parameter $k_D$ (measured with DLS) correlates with the second virial coefficient $B$ (measured with SLS), as shown in Figure 8.8(e). Similar empirical correlations between $k_D$ and $B$ were reported in the literature [38, 39, 173], and are represented as lines in Figure 8.8(e).
Figure 8.8: Protein-protein interactions from low to high protein concentration under various solution conditions. (a) Concentration dependence of the collective diffusion coefficient measured with DLS at 50 °C. Lines correspond to fittings with Equation 8.2. (b) Temperature dependence of the first order interaction coefficient determined with Equation 8.2. Lines, which correspond to linear fits, are guides for the eyes. (c) Concentration dependence of the Rayleigh ratio measured with SLS at 50 °C. (d) Concentration dependence of the Kirwood-Buff integral at 50 °C determined with Equation 8.9. (e) Correlation between $k_D$ and $B$ at 50 °C. Lines correspond to correlations found in the literature: Lehermayr et al. [38] (solid line), Connolly et al. [173] (dashed line), Saito et al. [39] (dotted line). (f) Correction factor $\alpha_{PI}$ as a function of protein concentration calculated with Equation 8.13 at 50 °C. Error bars in panels (b), (c), (d) and (e) represent the 80 % confidence intervals.

The SLS data shown in Figure 8.8(c) were then regressed over small concentration windows to determine the concentration dependence of the Kirwood-Buff integral $G$ with Equation 8.9. The results are shown in Figure 8.8(d), where it is
seen that for all the tested conditions $G$ is negative and increases with protein concentration, indicating that protein-protein interactions are overall repulsive, and that their magnitude decreases with protein concentration, as already observed in the literature for another protein system [53].

Finally, the values of $k_D$, $k'_D$ and $G(c)$ that were measured at 50 °C for each investigated condition were used to estimate the parameter $\alpha_{PI}$ defined in Equation 8.13. Results are shown in Figure 8.8(f), where it is seen that $\alpha_{PI}$, which is proportional to the self-diffusion coefficient, decreases with the protein concentration with a trend that depends on the solution condition.

The model parameters were then used to predict the kinetics of mAb aggregation under concentrated conditions. The model simulations of the monomer depletion kinetics are compared to the experimental data in Figure 8.4. It is seen that the model describes extremely well the experimental results at pH 5.5, both in the absence and presence of salt. However, it is observed that the model predictions tend to underestimate the rate of monomer consumption when the pH of the buffer solution is increased towards the protein pI. Two main reasons may underlie these results:

(i) The unfolding rate constant increases with protein concentration because surrounding protein molecules exhibit a destabilizing effect on the protein structure. Although this effect can hardly be corroborated with the set of experimental techniques used in this study, it cannot be ruled out.

(ii) A small portion of the monomeric protein aggregates in a native conformation. Experimental and computational studies indeed revealed that molecular crowding may induce a change in the shape of the native protein, potentially leading to the exposure of aggregation-prone patches [231]. Moreover, it can be speculated that native monomer aggregation is favored at pH close to the pI of the protein, where electrostatic repulsion is very weak, possibly explaining larger discrepancies between model simulations and experimental data at pH 7.4.
The model predictions of the time evolution of the aggregate molecular weight are shown in Figure 8.5. A comparison of these model predictions with model predictions obtained by (i) neglecting solution non-idealities (i.e. considering $\alpha_{PI} = 1$ in Equation 8.15), and (ii) neglecting both solution non-idealities and viscosity effects (i.e. using the classical RLCA kernel valid under dilute conditions given in Equation 1.7) is shown in Figure 8.9.

![Figure 8.9: Experimental data (symbols) and model simulations (lines) of the aggregate growth. Solid lines correspond to simulations performed with the aggregation kernel proposed in Equation 8.15, which accounts for solution non-idealities and for the viscosity increase with time. Dashed lines correspond to model simulations performed by considering the rise in viscosity only, and neglecting solution non-idealities. Dotted lines correspond to model simulations performed without taking into account any of the two corrections.](image-url)
It is seen in Figure 8.5 that the model describes very well the set of experimental data acquired at pH 5.5 and 6.5. The discrepancies between model simulations and experimental results observed at pH 5.5 in the presence of NaCl are regarded as relatively acceptable considering that there is no fitting parameter. At pH 7.4 instead, the model fails to describe the kinetics of aggregate growth. In particular, the simulated aggregate molecular weight is significantly overestimated at the three protein concentrations under investigation. Two main reasons can be proposed to explain this observation:

(i) The strength of protein-protein interactions significantly changes between 50 °C and 70 °C, implying that the parameter $\alpha_{PI}$ estimated at 50 °C cannot be used to describe aggregation kinetics at 70 °C. Indeed, it can be observed in Figure 8.8(b) that the $k_D$ parameter is strongly impacted by temperature at pH 7.4, as compared to the other solution conditions.

(ii) Protein surface non-uniformity and anisotropic interactions become critical when the solution pH approaches the pI of the protein [232, 233], possibly explaining the failure of coarse-grained molecular potentials (quantified through $k_D$, $k'_D$ and $G$) to describe orientation dependent protein-protein interactions.

8.4 Discussion

8.4.1 Relation between the second virial coefficient and aggregation kinetics under dilute conditions

With a view to achieving a rapid commercialization of potential drug candidates, there is a high interest in assessing product shelf-life within a limited timeframe. To do so, one strategy consists in assessing protein-protein interactions via the measurement of the interaction parameter ($k_D$) or the second virial coefficient ($B$) [50, 234]. Nevertheless, such approach relies on the assumption of a clear correlation between measurable protein-protein interaction parameters and the protein aggregation rate. In this section, we investigate the quantitative correlation
between the second virial coefficient $B$ (measured with SLS) and the Fuchs stability ratio $W_{11}$ (estimated from the kinetics of mAb aggregation).

Although $B$ and $W_{11}$ are both defined from the protein-protein interaction potential (see Equations 8.6 and 2.7, respectively), they are not sensitive to the same interaction potential characteristics [235]. To better visualize this, let us consider the simple energy profile constituted of a repulsive square barrier (of height $E_b$ and width $d_b$) and of an attractive square well (of height $E_w$ and width $d_w$) shown in Figure 8.10(a).

The thermodynamic parameter $B$ is very sensitive to the energy well, which characterizes the most stable state of the system, while the kinetic parameter $W_{11}$ is strongly dependent on the energy barrier that proteins need to overcome before reaching equilibrium (i.e. the aggregated state).

Calculations of $B$ and $W_{11}$ for the potential profile introduced in Figure 8.10(a) are presented in Figure 8.10(b). Lines were obtained by varying the energy barrier $E_b$ for a given set of $E_w$, $d_w$, $d_b$ parameters.

Under the assumptions of a deep narrow well ($E_w \gg k_B T$, $d_w \ll 2R_p$) and of a sufficiently large energy barrier ($E_b \gg k_B T$), it can be shown that:

$$\bar{B} = K_1(R_p, d_b, E_w, d_w) - \frac{K_2(R_p, d_b)}{W_{11}} \quad (8.16)$$

where $K_1(R_p, d_b, E_w, d_w)$ and $K_2(R_p, d_b)$ are two functions of the parameters indicated in brackets that are defined in Appendix H. Figure 8.10(b) shows that Equation 8.16 provides a good approximation of the $B - W_{11}$ correlation provided that $W_{11} \gtrsim 5$, which is always the case in practice.

Importantly, Equation 8.16 shows that $B$ correlates with $W_{11}$ only for a given set of $E_w$, $d_w$, $d_b$ parameters. This implies that the highest $B$ value does not necessarily corresponds to the highest $W_{11}$ value, as can be visualized by considering the three black points $F_1$, $F_2$, $F_3$ in Figure 8.10(b), which correspond to three theoretical mAb formulations. Although the second virial coefficient values follow the
Chapter 8. Complex phenomena at high mAb concentration

Figure 8.10: Correlation between the second virial coefficient $B$ and the Fuchs stability ratio $W$. (a) Simple energy profile considered for the simulations. (b) Simulation results obtained by varying the value of the energy barrier $E_b$. ‘Ref’ corresponds to $d_b = R_p$, $d_w = 0.1R_p$, and $E_w = 2k_BT$, while variations to this reference case are indicated in the graph. Solid and dotted lines were obtained with exact (see Appendix H) and simplified (Equation 8.16) expressions of $W_{11}$ and $B$, respectively. (c) Experimental results, where $B$ and $W_{11}$ were estimated at 50 and 70 °C, respectively.

Order $B(F_3) > B(F_2) > B(F_1)$, the opposite trend could potentially be observed for $W_{11}$ values.

In addition, it is worth noting that for a given set of $E_w, d_w, d_b$ parameters, $B$ becomes almost independent of $W_{11}$ when the energy barrier is high (see Equation 8.16 and Figure 8.10(b)). This suggests that to one given value of the second virial coefficient may correspond different values of the aggregation rate constant. In other words, mAb solution presenting similar second virial coefficient values may lead to very different aggregation rates.

Experimental results are shown in Figure 8.10(c), where it can be seen that there is no clear trend between $B$ and $W_{11}$. This observation may be explained by
the theoretical arguments presented above and / or by the difference in temperature at which $B$ and $W_{11}$ were measured (50 °C and 70 °C, respectively). It is worth underlining that $B$ measurements were performed with the native protein, while $W_{11}$ values were estimated for the unfolded protein.

To conclude, the analysis presented in this section suggests that the correlation between $B$ and $W_{11}$ is not straightforward, implying that measuring the second virial coefficient does not necessarily provide relevant information on the kinetics of protein aggregation.

### 8.4.2 Aggregation kernel under concentrated conditions

The classical RLCA kernel derived under dilute conditions accounts for two main features of the aggregating system: the *energy barrier* that the colliding proteins must overcome before aggregating, and the *aggregate compactness*. In this work, we propose a new aggregation kernel, as indicated in Equation 8.15, in order to extend the classical expression to the concentrated regime. This model includes two additional contributions as compared to the dilute case:

(i) A factor $\alpha_{PI}$ reflecting the textithydrodynamic and *thermodynamic non-idealities* of the concentrated protein solution under *stable conditions*.

(ii) A term taking into account the *viscosity increase during aggregation*.

#### 8.4.2.1 Protein solution non-idealities

The hydrodynamic and thermodynamic solution non-idealities that are quantified by the correction factor $\alpha_{PI}$ arise due to protein-protein interactions. In the following, we discuss how the kind and strength of intermolecular protein-protein interactions are affected by an increase in the protein concentration from dilute to concentrated conditions.

First, a *crowding* or *excluded volume effect* arises at high concentration due to the reduction in the free volume available to proteins [53, 221].
Chapter 8. Complex phenomena at high mAb concentration

Moreover, since protein molecules are on average closer from each other at high concentration, both long and short-range electrostatic interactions are of significant importance under concentrated conditions [172].

Increasing the protein concentration also results in an increase in the solution ionic strength due to [223, 236]: (i) the increase in the concentration of counter-ions necessary to maintain solution electroneutrality, (ii) the reduction of the volume available to the counter-ions [236]. In the absence of ion specific effects, an increase in the ionic strength is expected to screen electrostatic interactions. Nonetheless, counter-ion binding to the protein surface, which possibly becomes relevant at high protein concentration [224], may affect intermolecular interactions in a complex manner [237].

It has reported in the literature that attractive interactions may also be affected by a change in protein concentration [223, 238–240]. A reduction of the strength of attractive interactions with increasing protein concentration was indeed evidenced with small angle X-ray scattering studies performed on globular proteins [223, 238, 239]. Nevertheless, this experimental observation whose underlying reasons are still unclear, might be an artifact due to many-body interactions.

Indeed, under concentrated conditions, the pair-wise additivity of intermolecular potentials breaks down and many-body interactions become non-negligible. In particular, an attractive three-body interaction was identified, both for a protein system [220] and for charged colloidal particles [241].

8.4.2.2 Increase in solution viscosity during aggregation

In addition to protein-protein interactions, which affect the rate of protein diffusion even under stable conditions, the rise in solution viscosity during aggregation further hinders protein mobility. The increase in the macroscopic viscosity of the solution is attributed to the increase in the aggregate occupied volume fraction, and can be quantified experimentally, for example from the measure of the diffusion of tracer nanoparticles with DLS, as described in Chapter 6.
The impact of a rise in solution viscosity on the aggregation rate constant depends on the aggregate size: aggregates much larger than the average aggregate size experience the macroscopic viscosity of the dispersion, while aggregates much smaller than the average aggregate size experience the solvent viscosity. The aggregates of intermediate sizes experience an effective viscosity comprised between the solvent and the macroscopic viscosities, as discussed in Chapter 7.

8.4.2.3 Increase in aggregate compactness

Finally, for all the conditions under investigation, it was observed with light scattering experiments that more compact aggregates are formed at high protein concentration. This is possibly rooted in the fact that colliding clusters are likely to be partially overlapped under concentrated conditions. This effect does not require the addition of a special correction term in the aggregation kernel, since the aggregate compactness is already taken into account through the value of the fractal dimension $d_f$.

A change in the morphology of aggregates impacts the aggregation rate in a complex way. Indeed, for a given aggregate dimensionless mass $i$, an increase in the aggregate fractal dimension $d_f$ leads to a smaller collision radius ($\sim i^{1/d_f}$), a larger diffusive mobility ($\sim i^{-1/d_f}$), and a larger number of possible contact points at the external surface of the cluster ($\sim i^{1-1/d_f}$). While the former contribution leads to a decrease in aggregate reactivity, the latter two lead to an increase in aggregate reactivity. Numerical evaluation of the reaction-limited aggregation kernel show that, overall, an increase in the aggregate compactness at high concentration leads to an increase in the aggregation rate constant.

8.4.3 Stability ranking from low to high concentration

The magnitude of the aforementioned effects on the aggregation rate depends on the system under investigation. For example, the increase of the ionic strength with protein concentration may become particularly important when the mAb
molecules are highly charged or the background ion concentration is low [236]. Moreover, many-body interactions are expected to be stronger at low ionic strength, where double layers overlap [242]. Therefore, the concentration dependence of the aggregation rate is likely to be formulation specific, possibly resulting in a change of the stability ranking when moving from dilute to concentrated conditions.

This idea is illustrated in Figure 8.11, where the fraction of residual monomer and the aggregate molecular weight at 15 min incubation measured in the previous sections are plotted as a function of the solution condition, at protein concentrations of 1 and 60 g/L. Considering first the data of the residual monomer (Figure 8.11(a)), it is seen that the stability ranking is unchanged when moving from 1 to 60 g/L. This observation is in line with the fact that, in the case under investigation, the kinetics of monomer depletion is rate limited by protein unfolding, which is only weakly affected by a change in protein concentration. Let us consider now the results of the aggregate molecular weight, which is affected by both protein unfolding and bimolecular aggregation. It is seen in Figure 8.11(b) that the most stable condition at low protein concentration (pH 5.5) does not correspond to the most stable condition at high protein concentration (pH 6.5). Similarly, the least stable condition at low protein concentration (pH 7.4) does not correspond to the least stable condition at high concentration (pH 5.5 + 10 mM NaCl).

![Figure 8.11](image)

**Figure 8.11:** (a) Fraction of residual monomer and (b) aggregate molecular weight experimentally measured at 15 min incubation as a function of the solution conditions at protein concentration of 1 and 60 g/L. The stability ranking at high concentration differs from the one at low concentration.
NaCl). Moreover, it is interesting to note that the destabilizing effect of NaCl is much more pronounced at 60 g/L than at 1 g/L. These experimental observations illustrate the fact that the concentration dependence of the aggregation rate constant depends on the solution conditions. These considerations are particularly relevant in the context of drug formulation since they imply that identifying the most stable formulation under dilute conditions does not guarantee to select the most stable conditions under concentrated conditions.

8.5 Conclusion

The heat-induced aggregation of a model monoclonal antibody was investigated from low to high protein concentration under different buffer solution conditions, with varying pH and ionic strength. To do so, experimental characterization was combined with a kinetic model including protein unfolding and aggregation events.

Under dilute conditions, the aggregate growth was successfully described by using the classical reaction-limited aggregation kernel, which depends on two main factors: the aggregate compactness (quantified by the aggregate fractal dimension) and the energy barrier preventing the colliding protein molecules from aggregating (quantified by the Fuchs stability ratio).

It was observed that decreasing the solution pH below the protein pI leads on the one hand to an increase in the colloidal stability (reflected by an increase in the Fuchs stability ratio), and on the other hand to a decrease in the conformational stability (reflected by an increase in the unfolding rate constant). This implies that there is an optimum pH that minimizes the rate of monomer depletion.

Interestingly, it was observed that the second virial coefficient values obtained under various solution conditions could not be used to predict the ranking of the Fuchs stability ratios.
Under concentrated conditions, we proposed to modify the classical aggregation rate constant to account for two effects: (i) hydrodynamic and thermodynamic non-idealities, that can be evaluated under stable conditions with light scattering techniques; (ii) the rise in solution viscosity during the aggregation process, that can be accounted for by using a size and time dependent effective viscosity.

At pH far below the pI, this model was shown to describe well the set of experimental data, both in terms of monomer depletion and aggregate growth. When the pH was increased towards the mAb pI, discrepancies were observed between model simulations and experimental results. Indeed, model simulations tend to: (i) slightly underestimate the rate of monomer depletion, possibly due to an increase in the unfolding rate constant or to native monomer aggregation, (ii) considerably overestimate the rate of aggregate growth, possibly due to the strong temperature dependence of protein-protein interactions (which make the estimation of protein-protein interactions under stable conditions non applicable to aggregating conditions at pH close to the pI), or to strong anisotropic interactions (that make coarse-grained molecular potentials fail to describe protein-protein interactions at pH close to the pI).

Importantly, the finding that the concentration dependence of the aggregation rate constant is specific to the buffer solution implies that the most stable formulation under dilute conditions is not necessarily the most stable formulation under concentrated conditions.

9.1 Approaches to assess protein stability

The aggregation of therapeutic proteins is a major hurdle for the biopharmaceutical industry, hindering rapid commercialization of potential drug candidates. Indeed, protein aggregates are regarded as impurities that may affect product quality and safety. Therefore, the development of strategies seeking to predict long-term protein stability is an important focus of both pharmaceutical industries and academic institutions. Four main types of approaches aiming at assessing protein stability can be identified:

$(i)$ Molecular dynamics simulations, which provide detailed information on the location of protein hydrophobic regions, on the surface electrostatic potential as well as on the protein spatial conformation.

$(ii)$ Experimental characterization of the protein colloidal stability, for example with the measurements of the interaction parameter ($k_D$) or of the second-virial coefficient ($B$) with light scattering, which quantify in a coarse-grained manner the type and strength of protein-protein interactions.
(ii) Experimental characterization of the protein conformational stability, for example with the measurements of the melting temperature with DSC or CD, which describe the propensity of a protein to undergo unfolding.

(iv) Kinetic models developed in the frame of forced-degradation studies, which allow to determine the aggregation mechanism and the aggregation rate constant under accelerated conditions.

It must be underlined that the aforementioned strategies probe different aspects of therapeutic protein solutions. They have their own advantages and limitations, and are thus best used in a complementary manner. Due to a lack of fundamental understanding of the aggregation phenomena, the correlation between experimentally measurable parameters (such as the second virial coefficient or the melting temperature) and the aggregation rate is still unclear. Consequently, aggregation inhibition is still largely empirical in practice, requiring the use of high-throughput strategies for formulation screening [243]. Kinetic models are highly valuable to gain insights into the aggregation process, and therefore constituted the core of this thesis. Indeed, the comparison between experimental data and simulation results provides deep knowledge on the mechanisms of protein aggregation, as well as on the role of operational parameters (e.g. mAb concentration, pH, cosolutes) on the aggregation kinetics.

As illustrated in Figure 9.1, the experimental investigations of this thesis were performed under accelerated conditions, moving from low towards high protein concentration. The estimation of the product shelf-life from force degradation studies would be one of the core objectives of a research plan aiming at continuing this project. To do so, it would be necessary to study the impact of temperature on the aggregation rate at high protein concentration (see Figure 9.1). Protein aggregation kinetics are often reported to exhibit a non-Arrhenius behavior, possibly due to either an intrinsically temperature-dependent activation energy, or a temperature-dependent aggregation pathway [244, 245]. As a matter of fact, protein aggregation is complex phenomena involving several elementary steps (such as unfolding and aggregation), which are likely to be affected by temperature in
different ways. It is for example expected that native protein aggregation plays a more significant role at low temperature than under thermal stress [15].

![Figure 9.1: Outline of the research project.](image)

### 9.2 Input from colloid science

Concepts and techniques developed in the frame of colloid science represent powerful tools to gain understanding on the complex phenomena underlying the protein aggregation process, and were used throughout the thesis:

"Coarse-grained molecular potentials" have been quantified through the value of the *Fuchs stability ratio* ($W$), which is a kinetic parameter related to the energy barrier that protein molecules need to overcome before aggregating. In particular, it was shown in Chapters 4 and 8 that the analysis of the trend of $W$ in various environmental conditions provides a quantification of the impact of operational parameters (cosolutes and solution pH, respectively) on protein-protein interaction potentials. In Chapter 2, it has been found that predictions from the DLVO theory considerably underestimate the value of $W$ evaluated from the fitting of kinetic models to experimental data, indicating that the stability of protein solutions cannot be attributed to electrostatic repulsion only.
Coarse-grained molecular potentials have also been analyzed through the value of the second virial coefficient \( (B) \), which is a thermodynamic parameter characterizing the type (net attractive or repulsive) and strength of protein-protein interactions. In Chapter 8, it was shown that the correlation between \( W \) (which depends mainly on the height of the energy barrier preventing proteins from colliding) and \( B \) (which depends mainly on the deepness of the energy well where proteins fall to form an aggregate) is not straightforward.

The notion of fractal scaling was used to describe the aggregate compactness, revealing in Chapter 2 that different antibodies can form aggregates of different morphologies under the same operating conditions. Moreover, it was shown in Chapters 4 and 8 that variations in protein-protein interactions (for example mediated by the presence of salt or a change in the solution pH) affect the fractal dimension of aggregates. Finally, it was found in Chapter 8 that an increase in protein concentration leads to the formation of more compact aggregates.

The Fuchs stability ratio and the fractal dimension were implemented in the reaction limited cluster aggregation kernel, which was used to simulate the mAb aggregation kinetics by means of population balance equations. In Chapters 2, 4 and 8, the comparison between PBE simulations and experimental data provided quantitative information on the importance of the various elementary steps that contribute to the global aggregation process, i.e. protein unfolding and aggregation events per se, which reflect the protein conformational stability and the protein colloidal stability, respectively. In Chapter 3, a similar approach was applied to investigate the mechanisms of insulin fibril fragmentation. To do so, breakage kernels relying on the notion of daughter distributions were implemented in PBE simulations.

The concept of occupied volume fraction was used to rationalize the increase in solution viscosity during protein aggregation. Indeed, it was shown in Chapter 6 that profiles of the increase in viscosity with time fall on one single master curve when plotted as a function of the occupied volume fraction, which depends on the aggregate size, concentration and compactness.
Brownian dynamics simulations were performed in Chapter 7 to quantify the scale-dependent effective viscosity experienced by a tracer particle immersed in a concentrated colloidal dispersion. Importantly, it was found that the crossover between the nano- and the macroscopic viscosity occurs at tracer sizes comparable to the size of the crowders. Then, it was shown by means of PBE simulations that viscosity effects do not only reduce the rate of aggregate growth, but also lead to a sharpening of the aggregate distribution, especially under DLCA conditions.

Finally, in Chapter 8, an expression for the self-diffusion coefficient was derived under concentrated conditions, and introduced in the aggregation kernel. This correction, which accounts for both thermodynamic and hydrodynamic solution non-idealities due to protein-protein interactions, was evaluated from light scattering experiments.

In addition to the aforementioned concepts and techniques, which are directly borrowed from colloid science, additional features become relevant when dealing with the specific case of proteins.

In particular, as opposed to classical colloids, elementary events such as protein unfolding, or reversible oligomer formation are of primary importance in the description of protein aggregation phenomena. As shown in Chapter 2, such additional events can easily be included in the aggregation scheme, which is reflected in population balance equations. This approach was especially useful to discriminate the impact of operational parameters on the protein conformational and colloidal stabilities. For one antibody, it was shown in Chapter 4 that the presence of salt specifically accelerates aggregation events (i.e. reduces the protein colloidal stability), while the presence of polyol specifically inhibits protein unfolding (i.e. increases the protein conformational stability). Notably, it was found in Chapter 5 that the mAb stabilization effect does not depend only on the volume fraction filled by the polyol molecules, but also on the polyol chemical structure. In Chapter 8, it was evidenced that decreasing the pH below the pI of the mAb the leads on the one hand to an increase in the protein colloidal stability, and on the other hand
to a decrease in the protein conformational stability, thus resulting in an optimal pH minimizing the rate of monomer consumption.

Moreover, as compared to classical colloids, the protein reactivity strongly depends on the protein spatial conformation and on the exposition of aggregation-prone patches. To account for this, several values of the Fuchs stability ratio were introduced in Chapter 2 to characterize the different reactivities of the various sub-populations present in solution.

Finally, the protein surface heterogeneities and anisotropic interactions may play a crucial role in protein stability, especially at pH values close to the protein pI, as suggested in Chapter 8. In this frame, the use of patchy colloids as simplified model systems may open up the possibility to further investigate the role of surface patchiness in the aggregation behavior of proteins [246].

To conclude, colloid science offers a solid framework which can be complemented in a fruitful way to better address the challenging topic of protein stability.
Appendix A

Details about SEC-MALS experiments

Representative SEC chromatograms

A representative SEC chromatogram is shown in Figure A.1.

**Figure A.1:** Representative chromatogram observed by SEC, for a sample of mAb-1 that was incubated for 15 min at 70 °C and protein concentration of 2 g/L in a buffer of 20 mM histidine at pH 6.5. (a) Peak deconvolution was performed with OriginPro. (b) SEC chromatogram and corresponding weight average molecular weight determined from inline MALS.

This appendix complements Chapters 2 and 6.
The monomer, dimer and trimer elute at 25.1 min, 21.5 min and 19.4 min, as determined from a chromatogram containing a mixture of these three species only. The monomer, dimer and timer peaks correspond to peak 1, peak 2 and peak 3, respectively. Aggregates larger than trimer could not be deconvoluted with enough precision due to the exclusion limit of the column. Therefore, aggregates larger than trimer were described by two partially overlapped Gaussian peaks (denoted as peak 4 and peak 5 in the figure), whose position, height and width were adjusted to fit the experimental chromatogram. The information extracted from such a chromatogram is consequently the concentration of monomer, dimer, trimer and aggregates larger than trimer. The sum of these four concentrations is equal to the initial protein concentration.

The molecular weight of aggregates was determined by averaging all molecular weight values across all aggregate peaks, i.e. from 16.2 to 23.2 min.

**Raw light scattering data**

Some representative examples of dynamic and static light scattering data obtained with the Astra software are shown in Figure A.2.

![Figure A.2: Representative examples of raw light scattering results, obtained for a sample of mAb-1 that was incubated 35 min at the protein concentration of 40 g/L in a buffer of 20 mM histidine at pH 6.5. (a) Correlation function measured by inline DLS at the elution time of 16 min. The fit of the autocorrelation function (line) is compared to experimental data (crosses) (b) Results from SLS measurements at the same elution time.]

It is worth precising that for the analysis of the DLS results, the minimum and maximum thresholds for data processing were set to 5 and 300 nm, respectively.
Moreover, the minimum fit delay time was set to $2 \times 10^{-5} \text{s}^{-1}$ in order to remove the contribution from arginine, which is present at a high content in the elution buffer.

**Comparison between off-line and in situ measurements**

In order to assess the impact of lowering the temperature before analyzing the aggregated samples, the time evolution of the average hydrodynamic radius reconstructed from the population analyzed by off-line SEC-MALS analysis was compared with the values measured by DLS in-situ, as shown in Figure A.3 for both mAb-1 and mAb-2 samples. The two sets of data are in very close agreement, thereby proving that the aggregated samples of the systems under investigation were not significantly impacted by a change in temperature during the off-line analysis.

![Comparison between the average hydrodynamic radius measured from DLS in situ and the one reconstructed from the off-line SEC-MALS results for (a) mAb-1 and (b) mAb-2. The samples were incubated at 70 °C and protein concentration of 2 g/L in a buffer of 20 mM histidine at pH 6.5.](image-url)
Appendix B

PBE discretization with Gaussian functions (fibril breakage)

Change of basis matrix

In the frame of the discretization technique employed, the fibril mass distribution is approximated by a sum of $N_G$ Gaussian functions:

$$f(x, t) \simeq \sum_{i=1}^{N_G} \alpha_i(t) \exp\left(-s_i (x - x_i)^2\right) \quad (B.1)$$

where $\alpha_i$ are the time-dependent coefficients of the Gaussians, which are placed in the fixed grid positions $x_i$. The parameters $s_i$ instead characterize the width of the Gaussian functions, and have been defined as follows in order to ensure proper overlapping between adjacent Gaussian functions:

$$s_i = \frac{1}{(x_{i+1} - x_i)^2} \quad (B.2)$$

This appendix complements Chapter 3.
For numerical stability reasons, the grid points were spaced linearly up to \(i = 20\), and then logarithmically up to \(i = 5 \times 10^4\) with a total of 90 Gaussian functions.

For the sake of clarity, we define the function \(\varphi_i(x)\) such as:

\[
\varphi_i(x) = \exp \left( -s_i(x - x_i)^2 \right)
\]

(B.3)

In order to reduce the size of the problem, the concentrations are evaluated only at the \(N_G\) grid points by using the approximation shown in Equation B.1:

\[
\begin{bmatrix}
    f(x_1, t) \\
    f(x_2, t) \\
    \vdots \\
    f(x_{N_G})
\end{bmatrix}
\approx
\begin{bmatrix}
    \sum_{i=1}^{N_G} \alpha_i(t) \varphi_i(x_1) \\
    \sum_{i=1}^{N_G} \alpha_i(t) \varphi_i(x_2) \\
    \vdots \\
    \sum_{i=1}^{N_G} \alpha_i(t) \varphi_i(x_{N_G})
\end{bmatrix}
\]

(B.4)

Equation B.4 can be re-written in a more compact form as follows:

\[
f(t) \approx \mathbb{C} \alpha(t)
\]

(B.5)

where \(f(t)\) and \(\alpha(t)\) are vectors of length \(N_G\), which contain the fibril concentrations and the heights of the Gaussians at a given time \(t\), respectively. The matrix \(\mathbb{C}\) is the so-called change of base matrix, of size \([N_G \times N_G]\), allowing to ‘move’ from the concentration space to the Gaussian one. It is defined as follows:

\[
\mathbb{C} =
\begin{bmatrix}
    \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_{N_G}(x_1) \\
    \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_{N_G}(x_2) \\
    \vdots & \vdots & \ddots & \vdots \\
    \varphi_1(x_{N_G}) & \varphi_2(x_{N_G}) & \cdots & \varphi_{N_G}(x_{N_G})
\end{bmatrix}
\]

(B.6)
Appendix B. PBE discretization (fibril breakage)

Writing the left-hand side of the PBE

The left-hand side of the PBE can simply be treated by using the approximation shown in Equation B.5, and recalling that the matrix $C$ is constant in time since the coefficients $x_i$ and $s_i$ are time independent.

\[
\frac{df(t)}{dt} \simeq C \frac{d\alpha(t)}{dt}
\]  

(B.7)

Writing the right-hand side of the PBE

Let us introduce the following quantities, which appear in the balance shown in Equation 3.9:

\[
g(x_i, t) = \beta(x_i) f(x_i, t) \quad \text{(B.8)}
\]

\[
h(x_i, t) = f(x_i - 1, t) \quad \text{(B.9)}
\]

It is worth noting that the distribution $g(x_i, t)$ is simply a weighted form of the fibril distribution function $f(x_i, t)$, while $h(x_i, t)$ is the very same distribution $f(x_i, t)$, just shifted by 1.

The distributions $g(x_i, t)$ and $h(x_i, t)$ can be estimated from:

\[
h(t) \simeq D \alpha(t)
\]  

(B.10)

\[
g(t) \simeq C \alpha^W(t)
\]  

(B.11)

where the matrix $D$ and the weighted coefficients $\alpha^W$ are computed as follows:

\[
D = \begin{bmatrix}
\varphi_1(x_1 - 1) & \varphi_2(x_1 - 1) & \cdots & \varphi_N_G(x_1 - 1) \\
\varphi_1(x_2 - 1) & \varphi_2(x_2 - 1) & \cdots & \varphi_N_G(x_2 - 1) \\
\vdots & \vdots & \ddots & \vdots \\
\varphi_1(x_{N_G} - 1) & \varphi_2(x_{N_G} - 1) & \cdots & \varphi_N_G(x_{N_G} - 1)
\end{bmatrix}
\]  

(B.12)
Appendix B. PBE discretization (fibril breakage)

\[ \alpha^W(t) = C^{-1} E C \alpha(t) \]  
(B.13)

\[
E = \begin{bmatrix}
\beta(x_1) & 0 & \cdots & 0 \\
0 & \beta(x_2) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \beta(x_{NG})
\end{bmatrix}
\]  
(B.14)

**PBE in discretized form**

Given the quantities introduced above, the mass balance to be solved can be rewritten as:

\[
\frac{d\alpha(t)}{dt} = C^{-1} \left[ \delta_{nc} \frac{k_N}{n_C!} M(t)^{nc} - \delta_1 \left( k_N \frac{n_C}{n_C!} M(t)^{nc} + 2k_E M(t) (\mu_0(t) - M(t)) \right) \right.
\]
\[
+ \delta_{L2} 2k_E M(t) D \alpha(t) - \delta_{L1} 2k_E M(t) C \alpha(t)
\]
\[
+ 2\mu_\Gamma(t) - \delta_{L1} C \alpha^W(t) \right]
\]  
(B.15)

where the Krönecker deltas have been introduced in Equation 3.9, while the quantity \( \mu_0 \) and the elements of the vector \( \mu_\Gamma \) are defined as follows:

\[
\mu_0(t) = \int_0^\infty f(y, t) dy
\]  
(B.16)

\[
\mu_\Gamma(x_i, t) = \int_{x_i}^{\infty} \Gamma(z, x_i) \beta(z) f(z, t) dz
\]  
(B.17)

Note that \( \mu_0 \) and \( \mu_\Gamma \) are evaluated at every integration step through a numerical integration, as the vector \( f \) is known at every integration step by using Equation B.5.

The same holds for the concentration of monomeric protein \( M(t) \), i.e. the concentration of fibril of mass 1, which can be estimated at every time step using
the following relation:

\[ M(t) \simeq \sum_{i=1}^{N_G} C(1, i) \alpha_i(t) \]  (B.18)

It is worth mentioning that in the case of insulin fibril formation, a slight adjustment of the PBE shown in Equation B.15 is necessary since the nucleus size was found to be equal to 1. Namely, a distinction between the native and non-native form of the monomer needs to be introduced.
Appendix C

Details about viscosity measurements

Stability of the nanoparticles

First, it was verified that the nanoparticles are not destabilized, neither by the presence of the mAb molecules (under conditions where mAb aggregation does not occur), nor by the elevated temperature that is required to induce mAb aggregation (in the absence of mAb molecules) (Figure C.1(a)).

Then, it was verified that the nanoparticles are stable during a mAb aggregation experiment. To do so, the size of the nanoparticles was measured after mAb aggregation has been induced at high temperature. Measurements were performed at 25 °C on the quenched sample and at various dilutions. Indeed, a high solution viscosity would lead to an apparent hydrodynamic radius which is larger than the actual particle size. Sufficient dilution is thus necessary to reach the solvent viscosity, which is required for accurate size measurements (Figure C.1(b)).

This appendix complements Chapters 6 and 8.
Figure C.1: (a) Apparent hydrodynamic radius of the tracer nanoparticles (Nps) as a function of time in the buffer solution at 70 °C, and in a 40 g/L mAb solution at 25 °C. (b) Apparent hydrodynamic radius of the tracer nanoparticles as a function of the dilution factor for a mAb sample that was incubated in the presence of the nanoparticles at 70 °C and at the protein concentration of 40 g/L. It is seen that when the sample is sufficiently diluted, a hydrodynamic radius corresponding to the tracer particle size is reached.

Impact of nanoparticles concentration

Moreover, it was checked that the viscosity measurements are independent of the concentration of nanoparticles, as shown in Figure C.2.

Figure C.2: Kinetics of the increase in viscosity measured with different nanoparticle concentrations (as indicated in the legends) at protein concentrations of (a) 20 g/L, (b) 40 g/L, (c) 60 g/L.

Viscosity results

The time evolution of the increase in the macroscopic viscosity was fitted for each investigated condition by the following empirical function:

\[ \eta_\infty(t) = \eta_0 \times \left( \exp \left( \frac{t}{C_1} - C_2 \right) + C_3 \right) \]  

\[ \text{(C.1)} \]
where $C_1$, $C_2$ and $C_3$ are three positive real numbers, which are reported in Table C.1.

<table>
<thead>
<tr>
<th>Solution</th>
<th>c [g/L]</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH 5.5</td>
<td>20</td>
<td>3030</td>
<td>0.97</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>610</td>
<td>2.11</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>324</td>
<td>2.42</td>
<td>1.80</td>
</tr>
<tr>
<td>pH 6.5</td>
<td>20</td>
<td>833</td>
<td>4.61</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>404</td>
<td>4.67</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>222</td>
<td>4.61</td>
<td>1.93</td>
</tr>
<tr>
<td>pH 7.4</td>
<td>20</td>
<td>416</td>
<td>6.76</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>273</td>
<td>5.79</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>278</td>
<td>2.98</td>
<td>1.81</td>
</tr>
<tr>
<td>pH 5.5 + 10 mM NaCl</td>
<td>20</td>
<td>531</td>
<td>2.88</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>219</td>
<td>3.94</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>166</td>
<td>3.66</td>
<td>1.91</td>
</tr>
</tbody>
</table>

Table C.1: Coefficients describing the experimental viscosity increase with Equation C.1.
Appendix D

Boundaries for the approximate volume fraction

In this appendix, we identify boundaries for the approximate volume fraction \( \phi \) estimated by:

\[
\phi = \frac{4}{3} \pi \frac{M_0}{k_f} R_p^{d_f} \langle R_h \rangle^{3-d_f}
\]  

(D.1)

where \( R_p \) denotes the primary particle radius, while \( \langle R_h \rangle \) stands for the average hydrodynamic radius. The parameters \( d_f \) and \( k_f \) are the aggregate fractal dimension and scaling prefactor, respectively. \( M_0 \) denotes the initial protein concentration and is also equal to \( M_0 = \sum_i i M_i \).

We recall that the exact volume fraction is defined as:

\[
\Phi = \sum_i \frac{4}{3} \pi R_{h,i}^3 M_i
\]  

(D.2)

where \( M_i \) and \( R_{h,i} \) are the number concentration and hydrodynamic radius, respectively, of the aggregates containing \( i \) primary particles.

This appendix complements Chapter 6.
Lower bound

First, we show that the approximate volume fraction \( \phi \) estimated from Equation D.1 provides a lower bound of the exact volume fraction \( \Phi \) defined by Equation D.2 when the number average radius is considered.

The initial volume fraction is:

\[
\Phi_0 = \frac{4}{3} \pi R_p^3 M_0 
\]  
(D.3)

It can be noticed that:

\[
\phi = \Phi_0 \times \frac{1}{k_f} \left( \frac{\langle R_h \rangle}{R_p} \right)^{3-d_f} 
\]  
(D.4)

We define the number, surface and volume average hydrodynamic radii respectively as:

\[
\begin{align*}
\langle R_N \rangle &= \frac{\sum_i M_i R_{h,i}}{\sum_i M_i} \\
\langle R_S \rangle &= \frac{\sum_i M_i R_{h,i}^2}{\sum_i M_i R_{h,i}} \\
\langle R_V \rangle &= \frac{\sum_i M_i R_{h,i}^3}{\sum_i M_i R_{h,i}^2}
\end{align*}
\]  
(D.5)

Moreover, we introduce \( r_i \) and \( m_i \) defined as:

\[
\begin{align*}
\begin{cases} 
  r_i &= \frac{R_{h,i}}{\langle R_N \rangle} \\
  m_i &= \frac{M_i}{\sum_i M_i}
\end{cases}
\]  
(D.6)

By combining the above equations, it follows that:

\[
\Phi = \Phi_0 \frac{\langle R_N \rangle \langle R_S \rangle \langle R_V \rangle}{R_p^3} \times \frac{\sum_i M_i}{\sum_i iM_i} 
\]  
(D.7)
Appendix D. Boundaries for the approximate volume fraction

We try to find an upper bound for the quantity $A$ defined as:

$$A = \frac{\sum_i i M_i}{\sum_i M_i} = \sum_i i m_i$$  \hspace{1cm} (D.8)

According to the fractal scaling, the aggregate mass is connected to the aggregate radius by:

$$i = k_f \left( \frac{R_{h,i}}{R_p} \right)^{d_f}$$  \hspace{1cm} (D.9)

Therefore:

$$A = \sum_i k_f \left( \frac{R_{h,i}}{R_p} \right)^{d_f} m_i$$  \hspace{1cm} (D.10)

Let us recall the theorem of norm monotonicity. For a series of numbers $x_i$ weighted by coefficients $w_i$ such as $\sum_i w_i = 1$ and for two non-zero real numbers $p$ and $q$ such as $p \leq q$:

$$\left( \sum_i w_i x_i^p \right)^{1/p} \leq \left( \sum_i w_i x_i^q \right)^{1/q}$$  \hspace{1cm} (D.11)

Since $d_f \leq 3$ and $\sum_i m_i = 1$, it follows:

$$\sum_i m_i r_i^{d_f} \leq \left( \sum_i m_i r_i^3 \right)^{d_f/3}$$  \hspace{1cm} (D.12)

Hence:

$$\sum_i m_i r_i^{d_f} \leq \sum_i m_i r_i^3$$  \hspace{1cm} (D.13)

$$\sum_i m_i R_{h,i}^{d_f} \leq \left( \sum_i m_i R_{h,i}^3 \right) \langle R_N \rangle^{d_f-3}$$  \hspace{1cm} (D.14)

$$\sum_i m_i R_{h,i}^{d_f} \leq \langle R_N \rangle \langle R_S \rangle \langle R_V \rangle \langle R_N \rangle^{d_f-3}$$  \hspace{1cm} (D.15)

$$A \leq \frac{k_f}{R_p^{d_f}} \langle R_N \rangle \langle R_S \rangle \langle R_V \rangle \langle R_N \rangle^{d_f-3}$$  \hspace{1cm} (D.16)
Therefore:

$$\Phi \geq \Phi_0 \frac{\langle R_N \rangle \langle R_S \rangle \langle R_V \rangle}{R_p^3} \times \frac{1}{k_f \frac{\langle R_N \rangle \langle R_S \rangle \langle R_V \rangle}{R_p^3}^{d_f - 3}} \quad (D.17)$$

$$\Phi \geq \Phi_0 \times \frac{1}{k_f} \times \left( \frac{\langle R_N \rangle}{R_p} \right)^{3-d_f} \quad (D.18)$$

Equations D.4 and D.18 show that $\Phi \leq \phi$ provided that $\langle R_h \rangle = \langle R_N \rangle$. In other words, the estimated occupied volume fraction is underestimated when computed from the number average hydrodynamic radius.

**Upper bound**

Using a similar procedure, it is possible to provide an upper boundary to the volume fraction. In order to obtain the required result, we can proceed as follows. We introduce $r_i$ and $m_i$ defined now as:

$$\begin{cases} 
    r_i = \frac{R_{h,i}}{\langle R_V \rangle} \\
    m_i = \frac{M_i}{\sum M_i}
\end{cases} \quad (D.19)$$

This time, we will find a lower boundary for the quantity $A$, defined by Equation D.10.

By applying Equation D.11 to $r_i$ between 1 and $d_f$, it follows:

$$\sum_i m_i r_i^{d_f} \geq \left( \sum_i m_i r_i \right)^{d_f} \geq \sum_i m_i r_i \quad (D.20)$$

This implies that:

$$\sum_i m_i R_{h,i}^{d_f} \geq \left( \sum_i m_i R_{h,i} \right) \langle R_V \rangle^{d_f - 1} \quad (D.21)$$
Appendix D. **Boundaries for the approximate volume fraction**

\[ \sum_i m_i R_{h,i}^{d_f} \geq \langle R_N \rangle \langle R_V \rangle^{d_f-1} \] \hspace{1cm} (D.22)

Therefore:

\[ A \geq \frac{k_f}{R_p^{d_f}} \langle R_N \rangle \langle R_V \rangle^{d_f-1} \] \hspace{1cm} (D.23)

By inserting the above inequality in Equation D.7, we obtain:

\[ \Phi \leq \frac{\Phi_0}{k_f R_p^{3-d_f}} \langle R_S \rangle \langle R_V \rangle^{2-d_f} \] \hspace{1cm} (D.24)

Finally, the surface average size is smaller than the volume average size. Therefore:

\[ \Phi \leq \Phi_0 \times \frac{1}{k_f} \times \left( \frac{\langle R_V \rangle}{R_p} \right)^{3-d_f} \] \hspace{1cm} (D.25)

To conclude, the real value of the volume fraction is comprised between the one evaluated using the number-average size and the one evaluated using the volume average size.
Appendix E

PBE discretization with Gaussian functions (effective viscosity)

In the following, we show how to discretize the PBE describing aggregation under RLCA conditions using a size and time dependent effective viscosity. The discretization technique is based on the approximation of the cluster mass distribution by a sum of Gaussian functions, as was done in the case of amyloid fibril breakage (see Appendix B).

The population balance equation in its continuous form reads:

\[
\frac{dM(x,t)}{dt} = \frac{1}{2} \int_0^x k(x-y,y,t)M(x-y,t)M(y,t) \, dy - M(x,t) \int_0^\infty k(x,y,t)M(y,t) \, dy \tag{E.1}
\]

where \(M(x,t)\) is defined such that \(M(x,t)\, dt\) represents the concentration of aggregates of mass \(x\) to \(x + dx\) at time \(t\), and \(k(x,y,t)\) is the aggregation rate constant between two clusters of mass \(x\) and \(y\) at time \(t\), given by:

\[
k(x,y,t) = \frac{2k_BT}{3W} \left( \frac{x^{-1/d_f}}{\eta_{eff}(x,t)} + \frac{y^{-1/d_f}}{\eta_{eff}(y,t)} \right) \left( x^{1/d_f} + y^{1/d_f} \right) (xy)^\lambda \tag{E.2}
\]

This appendix complements Chapter 7.
where \( \eta_{\text{eff}}(x, t) \) represents the effective viscosity experienced by an aggregate of mass \( x \) at time \( t \). Note that since \( sr_i \) and \( \phi \) are monotonic functions of time, defining \( \eta_{\text{eff}}(x, t) \) is mathematically equivalent to defining \( \eta_{\text{eff}}(sr_i, \phi) \), as done in the Chapter 7.

The discretized PBE in the DLCA case can be simply obtained by setting the Fuchs stability ratio and the power law exponent to \( W = 1 \) and \( \lambda = 0 \), respectively.

The treatment of the left-hand side of the PBE was already presented in Appendix B (see Equation B.7, where the vector \( f(t) \) is replaced here by \( M(t) \)).

In order to treat the right hand side, let us first re-write the aggregation kernel as follows:

\[
k(x, y, t) = \frac{2k_B T}{3W} \times \left( (xy)^\lambda + \frac{x^{\lambda+d-1} y^{\lambda-d-1}}{\eta_{\text{eff}}(x, t)} + \frac{x^{\lambda-d-1} y^{\lambda+d-1}}{\eta_{\text{eff}}(y, t)} + \frac{(xy)^\lambda}{\eta_{\text{eff}}(x, t)} \right)
\]  

(E.3)

The loss term is given by:

\[
\mathcal{C} \left( \frac{d\alpha}{dt} \right) = -\frac{2k_B T}{3W} \left( \mathcal{C} \alpha^{w_E}(t) \sum_{j=1}^{N_G} \alpha^{w_F}_j(t) \int_0^\infty \varphi_j(y)dy \right.
\]

\[
+ \mathcal{C} \alpha^{w_G}(t) \sum_{j=1}^{N_G} \alpha^{w_H}_j(t) \int_0^\infty \varphi_j(y)dy
\]

\[
- \mathcal{C} \alpha^{w_F}(t) \sum_{j=1}^{N_G} \alpha^{w_E}_j(t) \int_0^\infty \varphi_j(y)dy
\]

\[
+ \mathcal{C} \alpha^{w_H}(t) \sum_{j=1}^{N_G} \alpha^{w_G}_j(t) \int_0^\infty \varphi_j(y)dy
\]

\[
\right)
\]  

(E.4)

where the weighted coefficients are defined as:

\[
\alpha^{w_E}(t) = \mathcal{C}^{-1} W_E(t) \mathcal{C} \alpha(t)
\]  

(E.5)

\[
\alpha^{w_F}(t) = \mathcal{C}^{-1} W_F(t) \mathcal{C} \alpha(t)
\]  

(E.6)

\[
\alpha^{w_G}(t) = \mathcal{C}^{-1} W_G(t) \mathcal{C} \alpha(t)
\]  

(E.7)
Appendix E. PBE discretization (effective viscosity)

\[ \alpha_{\text{eff}}(t) = C^{-1} W_H(t) C \alpha(t) \]  

(E.8)

with the following matrices:

\[
W_E(t) = \begin{bmatrix}
\frac{x_1}{\eta_{\text{eff}}(x_1,t)} & 0 & \cdots & 0 \\
0 & \frac{x_2}{\eta_{\text{eff}}(x_2,t)} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{x_{NG}}{\eta_{\text{eff}}(x_{NG},t)}
\end{bmatrix}
\]  

(E.9)

\[
W_F = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix}
\]

(E.10)

\[
W_G(t) = \begin{bmatrix}
\frac{x_1^{\lambda+d^{-1}}}{x_1} & 0 & \cdots & 0 \\
0 & \frac{x_2^{\lambda+d^{-1}}}{x_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{x_{NG}^{\lambda+d^{-1}}}{x_{NG}}
\end{bmatrix}
\]

(E.11)

\[
W_H(t) = \begin{bmatrix}
\frac{x_1^{\lambda-d^{-1}}}{\eta_{\text{eff}}(x_1,t)} & 0 & \cdots & 0 \\
0 & \frac{x_2^{\lambda-d^{-1}}}{\eta_{\text{eff}}(x_2,t)} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{x_{NG}^{\lambda-d^{-1}}}{\eta_{\text{eff}}(x_{NG},t)}
\end{bmatrix}
\]

(E.12)

Let us recall that:

\[
\int_0^\infty \exp(-s_i(y - x_i)^2)dy = \frac{1}{2} \sqrt{\frac{\pi}{s_i}} \left( 1 + \text{erf} \left( x_i \sqrt{s_i} \right) \right)
\]

(E.13)
Since the selected grid ensures that \( x_i \sqrt{s_i} \) for every \( i > 2 \), we apply the following approximation:

\[
\int_0^\infty \exp(-s_i(y - x_i)^2)dy \approx \sqrt{\frac{\pi}{s_i}} \quad (E.14)
\]

The loss term then becomes:

\[
C \left( \frac{d\alpha}{dt} \right)_- = - \frac{2k_B T}{3W} \left( C' \alpha^{wE}(t) \sum_{j=1}^{NG} \alpha^{WF}_j(t) \sqrt{\frac{\pi}{s_j}} \right.
\]

\[
C' \alpha^{wG}(t) \sum_{j=1}^{NG} \alpha^{wH}_j(t) \sqrt{\frac{\pi}{s_j}}
\]

\[
C' \alpha^{wH}(t) \sum_{j=1}^{NG} \alpha^{wG}_j(t) \sqrt{\frac{\pi}{s_j}}
\]

\[
C' \alpha^{wF}(t) \sum_{j=1}^{NG} \alpha^{wE}_j(t) \sqrt{\frac{\pi}{s_j}} \quad (E.15)
\]

By using the weighting coefficient previously introduced, the production term of the PBE can be treated as follows:

\[
C' \left( \frac{d\alpha}{dt} \right)_+ = - \frac{k_B T}{3W} \left( \int_0^x \left( \sum_{i=1}^{NG} \alpha_i^{wE}(t) \phi_i(x - y) \sum_{j=1}^{NG} \alpha_j^{wF}(t) \phi_j(y) \right) dy 
\]

\[
\int_0^x \left( \sum_{i=1}^{NG} \alpha_i^{wG}(t) \phi_i(x - y) \sum_{j=1}^{NG} \alpha_j^{wH}(t) \phi_j(y) \right) dy 
\]

\[
\int_0^x \left( \sum_{i=1}^{NG} \alpha_i^{wH}(t) \phi_i(x - y) \sum_{j=1}^{NG} \alpha_j^{wG}(t) \phi_j(y) \right) dy 
\]

\[
\int_0^x \left( \sum_{i=1}^{NG} \alpha_i^{wF}(t) \phi_i(x - y) \sum_{j=1}^{NG} \alpha_j^{wE}(t) \phi_j(y) \right) dy \quad (E.16)
\]

Let us recall the definition of the convolution of two functions \( f \) and \( g \) on the domain of definition \( D \):

\[
(f * g)(x) = \int_D f(y)g(x - y)dy \quad (E.17)
\]
Applying the above definition and using the matrix form, one obtains:

$$C \left( \frac{d\alpha}{dt} \right)_+ = -\frac{2k_BT}{3W} \left( C\alpha^{wE}(t) \ast C\alpha^{wF}(t) + C\alpha^{wG}(t) \ast C\alpha^{wH}(t) \right)$$ (E.18)

The convolution of two Gaussians results in another Gaussian whose position and width can be computed from the parameters of the two convoluted Gaussians. Accordingly, we define the convolution quantities $\alpha^{c,EF}$, $\alpha^{c,GH}$, and $\varphi^c$ such as:

$$C\alpha^{wE}(t) \ast C\alpha^{wF}(t) = \varphi^{c} \alpha^{c,EF}(t)$$ (E.19)

$$C\alpha^{wG}(t) \ast C\alpha^{wH}(t) = \varphi^{c} \alpha^{c,GH}(t)$$ (E.20)

where:

$$\varphi^{c}_{i,j}(x) = \exp \left( -s^{c}_{i,j}(x - x^{c}_{i,j})^2 \right) \quad \text{with} \quad \begin{cases} x^{c}_{i,j} = x_i + x_j \\ s^{c}_{i,j} = s_i + s_j \end{cases}$$ (E.21)

$$\alpha^{c,EF}_{i,j} = \alpha^w_i(t)\alpha^w_j(t) \sqrt{\frac{\pi}{s_i + s_j}}$$ (E.22)

$$\alpha^{c,GH}_{i,j} = \alpha^w_i(t)\alpha^w_j(t) \sqrt{\frac{\pi}{s_i + s_j}}$$ (E.23)

Finally, the discretized balance to be solved reads:

$$\frac{d\alpha}{dt} = \left( \frac{-2k_BT}{3W} + \varphi^{c} \alpha^{c,EF}(t) + \varphi^{c} \alpha^{c,GH}(t) \right)$$ (E.24)
Appendix F

Collective and self-diffusion coefficients

General expressions

The diffusion of solutes in a multicomponent system can be described by the Stefan-Maxwell equations [228, 230]:

\[ \sum_{j \neq i} \xi_{ij} c_j^m (u_i - u_j) = F_i^m \]  \hspace{1cm} (F.1)

where \( \xi_{ij} \) is the friction coefficient between the species \( i \) and \( j \), \( u_i \) is the average velocity of the species \( i \), \( c_i^m \) is the molar concentration of \( i \), and \( F_i^m \) is the driving force acting on 1 mole of \( i \).

According to the Onsager reciprocal relations:

\[ \xi_{ij} = \xi_{ji} \] \hspace{1cm} (F.2)

This appendix complements Chapter 8.
Let us denote $J_i$ the flux of the component $i$ defined such as:

$$J_i = c_i^m u_i \quad \text{(F.3)}$$

Equation F.1 then becomes:

$$\sum_{j \neq i} \xi_{ij} \left( \frac{c_j^m}{c_i^m} J_i - J_j \right) = F_i^m \quad \text{(F.4)}$$

For a volumetric flux conserving system, we have the additional relation [247]:

$$\sum_i J_i \nu_i = 0 \quad \text{(F.5)}$$

where $\nu_i$ is the partial molar volume of $i$, which according to the conservation of mass satisfies the following relation:

$$\sum_i c_i^m \nu_i = 1 \quad \text{(F.6)}$$

**Collective diffusion**

Let us start to derive an expression for the collective diffusion coefficient, which describes the diffusion of protein molecules under a concentration gradient [185].

To do so, we consider the binary system solvent (denoted by $s$) - protein (denoted by $p$), where the driving force is given by a gradient of the chemical potential $\mu$:

$$F_s = -\text{grad} \mu_s \quad \text{(F.7)}$$

We aim at finding $D_c$ defined such as:

$$J_p = -D_c \times \text{grad} c_p^m \quad \text{(F.8)}$$
By applying Equations F.4 and F.7, we find:

\[ \xi_{sp} \left( \frac{c_p}{c_s} J_s - J_p \right) = -\text{grad} \mu_s \quad (F.9) \]

Moreover, by using Equations F.5 and F.6, we obtain:

\[ J_p = \frac{c_p^m \nu}{\xi_{sp}} \text{grad} \mu_s \quad (F.10) \]

The chemical potential can be expressed from the osmotic pressure \( \Pi \) as:

\[ \mu_s - \mu_s^0 = -\nu \Pi \quad (F.11) \]

It follows that:

\[ J_p = -\frac{(1 - \frac{c_p^m \nu}{c_s})^2}{\xi_{sp} c_s^m} \frac{\partial \Pi}{\partial c_p^m} \text{grad} c_p^m \quad (F.12) \]

Finally, it results from Equation F.8 that:

\[ D_c = \frac{(1 - \frac{c_p^m \nu}{c_s})^2}{\xi_{sp} c_s^m} \frac{\partial \Pi}{\partial c_p^m} \quad (F.13) \]

The term \( \xi_{sp} c_s^m \), i.e. the friction coefficient between the protein and the solvent, can be estimated from Equation F.1 since the driving force equilibrates with the frictional force, which is given by Stokes’ law [228]. For a protein of radius \( R_p \), one gets:

\[ \xi_{sp} c_s^m (u_p - u_s) = 6\pi \eta_0 R_p (u_p - u_s) N_a \quad (F.14) \]

which leads to:

\[ \xi_{sp} c_s^m = 6\pi \eta_0 R_p N_a \quad (F.15) \]

**Self-diffusion**

Let us now focus on the case of protein self-diffusion, where there are no gradients in the total protein and solvent concentrations, but only in the fraction of labelled
protein molecules as compared to the unlabelled ones [185]. Accordingly, we consider a labelled protein molecule (denoted as \( p^* \)) immersed in a mixture of solvent and unlabelled protein. The system is then constituted of three components: \( s, p \) and \( p^* \).

We aim at finding \( D_s \) defined such as:

\[
J_{p^*} = -D_s \times \text{grad} c_{p^*} \quad \text{(F.16)}
\]

According to Equation F.4:

\[
\xi_{sp} \left( \frac{c_m}{c_s} J_s - J_p \right) + \xi_{sp^*} \left( \frac{c_m}{c_{p^*}} J_s - J_{p^*} \right) = -\text{grad} \mu_s \quad \text{(F.17)}
\]

\[
\xi_{sp} \left( \frac{c_m}{c_p} J_p - J_s \right) + \xi_{sp^*} \left( \frac{c_m}{c_{p^*}} J_p - J_{p^*} \right) = -\text{grad} \mu_p \quad \text{(F.18)}
\]

The condition of self-diffusion involves the following relations [92, 230]:

\[
\text{grad} \mu_s = 0 \quad \text{(F.19)}
\]

\[
\xi_{pp} = \xi_{pp^*} \quad \text{(F.20)}
\]

\[
\xi_{sp} = \xi_{sp^*} \quad \text{(F.21)}
\]

\[
\nu_p = \nu_{p^*} \quad \text{(F.22)}
\]

Equation F.17 then becomes:

\[
J_s = \frac{c_m}{c_p + c_{p^*}} (J_p + J_{p^*}) \quad \text{(F.23)}
\]

The condition of volumetric flux conservation of Equation F.5 reads:

\[
J_s \nu_s + (J_p + J_{p^*}) \nu_p = 0 \quad \text{(F.24)}
\]
Equations F.23 and F.24 imply that:

\[ J_s = 0 \]  \hspace{1cm} (F.25)

\[ J_p + J_{p*} = 0 \]  \hspace{1cm} (F.26)

According to the Gibbs-Duhem relation at constant temperature and pressure:

\[ c_s^m \text{grad}\mu_s + c_p^m \text{grad}\mu_p + c_p^m \text{grad}\mu_{p*} = 0 \]  \hspace{1cm} (F.27)

By using Equations F.18, F.19, F.25, F.26 and F.27, it follows:

\[ J_{p*} = -\frac{c_p^m}{\xi_{sp} c_s^m + \xi_{pp} (c_p^m + c_{p*}^m)} \text{grad}\mu_{p*} \]  \hspace{1cm} (F.28)

According to Vink [230], \text{grad}\mu_{p*} can be computed as follows:

\[ \text{grad}\mu_{p*} = RT \frac{c_p^m}{c_{p*}^m} \]  \hspace{1cm} (F.29)

Equation F.28 therefore reads:

\[ J_{p*} = -\frac{RT}{\xi_{sp} c_s^m + \xi_{pp} (c_p^m + c_{p*}^m)} \text{grad}c_p^m \]  \hspace{1cm} (F.30)

It results from Equation F.16 that:

\[ D_s = \frac{RT}{\xi_{sp} c_s^m + \xi_{pp} (c_p^m + c_{p*}^m)} \]  \hspace{1cm} (F.31)

One now needs to estimate the frictional term \( \xi_{sp} c_s^m + \xi_{pp} (c_p^m + c_{p*}^m) \). By applying Equation F.1 to \( p^* \) and to \( p \) for the Stokes’ drag force, and recalling that \( u_s = 0 \) and \( c_p^m u_{p*} = -c_p^m u_p \) (see Equations F.25 and F.26), one gets:

\[ (\xi_{sp} c_s^m + 2\xi_{pp*} (c_p^m + c_{p*}^m)) u_{p*} = 6\pi \eta_p R_p u_{p*} N_a \]  \hspace{1cm} (F.32)
Collective and self-diffusion coefficients

\[
(\xi_{sp}c_s^m + 2\xi_{pp}(c_p^m + c_{p*}^m))u_p = 6\pi\eta_p R_p u_p N_a \tag{F.33}
\]

where \(\eta_p\) is the viscosity experienced by a protein molecule under concentrated conditions. This viscosity is however extremely difficult to access experimentally as it would require to track one labelled protein diffusing in the protein-solvent mixture. Nevertheless, it is possible to assess the viscosity experienced by tracer polymeric nanoparticles, as described in Chapter 6. Since the polymeric particles are large compared to the protein molecules, they experience the macroscopic viscosity of the protein solution. By using Brownian dynamics simulations, we identified in Chapter 7 how the effective viscosity scales with the tracer particle size. Overall, we can estimate the effective viscosity of a tracer particle having the same size than a protein molecule, which we denote \(\eta_{eff}\) in agreement with the notation of Chapter 8.

By summing Equations F.32 and F.33, and assuming \(\eta_p \approx \eta_{eff}\) it results:

\[
\xi_{sp}c_s^m + \xi_{pp}(c_p^m + c_{p*}^m) = 6\pi\eta_{eff} R_p N_a \tag{F.34}
\]

Relation between collective and self-diffusion coefficients

By dividing Equations F.13 and F.31, it results:

\[
\frac{D_s}{D_c} = \frac{RT}{(1 - c_p^m / T_p)^2} \frac{\xi_{sp}c_s^m}{(\xi_{sp}c_s^m + \xi_{pp}c_p^m)} \left( \frac{\partial \Pi}{\partial c_p^m} \right)^{-1} \tag{F.35}
\]

In this equation, \(c_p^m\) represents the total protein concentration, which corresponds to \(c_p^m + c_{p*}^m\) (unlabelled + labelled) in the section on the self-diffusion case.

The difficulty is now to estimate the friction coefficient ratio:

\[
\xi_R = \frac{\xi_{sp}c_s^m}{\xi_{sp}c_s^m + \xi_{pp}c_p^m} \tag{F.36}
\]
Appendix F. Collective and self-diffusion coefficients

As a first approximation, it can be assumed that $\xi_{pp} \ll \xi_{sp}$, so that $\xi_R$ is close to unity in the case where the protein concentration $c_m$ is not too large [248].

Another strategy consists in assuming that the value $\xi_R$ can be computed from Equations F.15 and F.34, so that:

$$\xi_R \approx \frac{\eta_0}{\eta_{eff}} \quad (F.37)$$

which tends to unity at low protein concentration.

We finally obtain:

$$\frac{D_s}{D_c} = \frac{RT}{(1 - c_m^{eq})^2} \frac{\eta_0}{\eta_{eff}} \left( \frac{\partial \Pi}{\partial c_m} \right)^{-1} \quad (F.38)$$
Appendix G

Supplementary light scattering data

Figure G.1 shows that the interaction parameter $k_D$ correlates with the protein net charge provided that the $k_D$ measurements are performed at constant ionic strength.

![Graphs showing Zeta potential and $k_D$ vs pH](image)

**Figure G.1:** (a) Zeta potential measured at 25 °C and at the protein concentration of 1 g/L. (b) $k_D$ measurements performed at 25 °C and at the constant ionic strength of 15 mM. To do so, 10 and 14 mM of NaCl were added to the buffer solutions at pH 6.5 and 7.4, respectively.

This appendix complements Chapter 8.
Figure G.2 shows the determination the second virial coefficient $B$ and the apparent molecular weight $MW_{app}$ under dilute conditions. As explained in Chapter 8, $B$ and $MW_{app}$ can be obtained under dilute conditions from a Debye plot according to:

$$\frac{K_c}{R_{00}} = \frac{1}{MW_{app}} + 2Bc \quad \text{(G.1)}$$

**Figure G.2:** Static light scattering data under dilute protein conditions at 50 °C. (a) Debye plot used to determine the second virial coefficient and the apparent molecular weight. (b) Second virial coefficient normalized by the hard-sphere value $B^{HS} = \frac{2}{3}\pi(2R_p)^3 N_a/MW_p$ as a function of solution pH. A protein hard-sphere radius of 5.5 nm was considered. (c) Apparent molecular weight normalized by the true protein molecular weight as a function of solution pH. Open symbols in (b) and (c) correspond to the case where 10 mM NaCl was added.
Appendix H

Relation between Fuchs ratio and second virial coefficient

We aim at deriving a relation between the Fuchs stability ratio and the second virial coefficient considering the following interaction potential:

\[
V_T = \begin{cases} 
\infty & \text{for } r < 2R_p \\
E_w & \text{for } 2R_p \leq r < 2R_p + d_w \\
E_b & \text{for } 2R_p + d_w \leq r < 2R_p + d_w + d_b \\
0 & \text{for } 2R_p + d_w + d_b \leq r
\end{cases}
\]  

(H.1)

As illustrated in Chapter 8, this potential profile is constituted of a square energy well and a square energy barrier.

The Fuchs stability ratio can be computed as follows:

\[
W_{11} = 2R_p \int_{2R_p}^{\infty} \exp \left( \frac{V_T(r)}{k_B T} \right) \frac{dr}{r^2}
\]  

(H.2)

This appendix complements Chapter 8.
W_{11} = 2R_p \left[ \int_{2R_p}^{2R_p+d_w} \exp \left( \frac{V_T(r)}{k_B T} \right) \frac{dr}{r^2} \right. \\
+ \int_{2R_p+d_w}^{2R_p+d_w+d_b} \exp \left( \frac{V_T(r)}{k_B T} \right) \frac{dr}{r^2} \\
\left. + \int_{2R_p+d_w+d_b}^{\infty} \exp \left( \frac{V_T(r)}{k_B T} \right) \frac{dr}{r^2} \right] 

\text{(H.3)}

W_{11} = 2R_p \left[ \exp \left( \frac{-E_w}{k_B T} \right) \left[ -\frac{1}{2R_p+d_w} + \frac{1}{2R_p} \right] \\
+ \exp \left( \frac{E_b}{k_B T} \right) \left[ -\frac{1}{2R_p+d_w+d_b} + \frac{1}{2R_p} \right] \\
\left. + \frac{1}{2R_p+d_w+d_b} \right] \right] 

\text{(H.4)}

Under the assumptions that $E_b \gg k_B T$, $E_w \gg k_B T$ and $d_w \ll 2R_p$, we obtain:

\[ W_{11} \approx \frac{d_b}{2R_p + d_b} \exp \left( \frac{E_b}{k_B T} \right) \]

\text{(H.5)}

On the other hand, the second virial coefficient can be computed as follows:

\[ \overline{B} = \overline{B}^{HS} - 2\pi \int_{2R_p}^{\infty} \left[ \exp \left( \frac{-V_T(r)}{k_B T} \right) - 1 \right] r^2 dr \]

\text{(H.6)}

where $\overline{B}^{HS} = \frac{2}{3} \pi (2R_p)^3$ is the hard-sphere contribution to the second virial coefficient.

\[ \overline{B} = \overline{B}^{HS} - 2\pi \int_{2R_p}^{2R_p+d_w} \left[ \exp \left( \frac{-V_T(r)}{k_B T} \right) - 1 \right] r^2 dr \\
- 2\pi \int_{2R_p+d_w+d_b}^{\infty} \left[ \exp \left( \frac{-V_T(r)}{k_B T} \right) - 1 \right] r^2 dr \]

\text{(H.7)}
\[ \overline{B} = \overline{B}^{\text{HS}} + \frac{2\pi}{3} \left[ 1 - \exp \left( \frac{E_w}{k_B T} \right) \right] \times \left[ (2R_p + d_w)^3 - (2R_p)^3 \right] \\
\left. \right. + \frac{2\pi}{3} \left[ 1 - \exp \left( -\frac{E_b}{k_B T} \right) \right] \times \left[ (2R_p + d_w + d_b)^3 - (2R_p + d_w)^3 \right] \tag{H.8} \]

Under the assumptions that \( d_w \ll 2R_p \), and using a first order Taylor expansion we have:

\[ (2R_p + d_w)^3 - (2R_p)^3 \approx 3d_w(2R_p)^2 \tag{H.9} \]

\[ (2R_p + d_w + d_b)^3 - (2R_p + d_w)^3 \approx 3d_b(2R_p)^2 \tag{H.10} \]

It results that:

\[ \overline{B} = \overline{B}^{\text{HS}} + 2\pi d_w(2R_p)^2 \left[ 1 - \exp \left( \frac{E_w}{k_B T} \right) \right] \\
\left. \right. + 2\pi d_b(2R_p)^2 \left[ 1 - \exp \left( -\frac{E_b}{k_B T} \right) \right] \tag{H.11} \]

By introducing Equation H.11 into Equation H.11 and using the expression of \( \overline{B}^{\text{HS}} \) we finally obtain the following equation relating the second virial coefficient to the Fuchs stability ratio:

\[ \overline{B} = K_1(R_p, E_w, d_w, d_b) - K_2(R_p, d_b) \times \frac{1}{W_1} \tag{H.12} \]

with the following parameters:

\[ K_1(R_p, E_w, d_w, d_b) = \overline{B}^{\text{HS}} \left[ 1 + \frac{3d_w}{2R_p} \left[ 1 - \exp \left( \frac{E_w}{k_B T} \right) \right] + \frac{3d_b}{2R_p} \right] \tag{H.13} \]

\[ K_2(R_p, d_b) = \frac{3\overline{B}^{\text{HS}}}{1 + d_b/(2R_p)} \left( \frac{d_b}{2R_p} \right)^2 \tag{H.14} \]

Note that \( \overline{B} \) and \( W_1 \) both depend on \( R_p, E_b, E_w, d_w, d_b \). The parameter dependence is not indicated in Equation H.12 for the sake of clarity.
Bibliography


References


[230] H. Vink. Mutual diffusion and self-diffusion in the frictional formalism of

[231] D. Homouz, M. Perham, A. Samiotakis, M. S. Cheung, and P. Wittung-
Stafshede. Crowded, cell-like environment induces shape changes in aspher-

wicker, and R. Curtis. The role of electrostatics in protein-protein interac-


Behavior of monoclonal antibodies: Relation between the second virial coef-
cient ($B_2$) at low concentrations and aggregation propensity and viscosity

Salt-induced aggregation of a monoclonal human Immunoglobulin G1. *J.


2009.

Thermodynamics of protein aqueous solutions: From the structure factor to


Abbreviations

AFM  Atomic Force Microscopy
CD   Circular Dichroism
DLCA Diffusion Limited Cluster Aggregation
DLS  Dynamic Light Scattering
DLVO Derjaguin Landau Verwey Overbeek
DSC  Differential Scanning Calorimetry
HPLC High-Performance Liquid Chromatography
IgG  Immunoglobulin G
mAb  monoclonal Antibody
MALS Multi Angle Light Scattering
ODE  Ordinary Differential Equations
PBE  Population Balance Equations
PEG  Polyethylene Glycol
pI   Isoelectric point
PDI  Polydispersity Index
RLCA Reaction Limited Cluster Aggregation
RSS  Residual Sum of Squares
SEC  Size Exclusion Chromatography
SLS  Static Light Scattering
ThT  Thioflavin T
UV   Ultraviolet
# Physical Constants

<table>
<thead>
<tr>
<th>Physical Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avogadro number $N_a$</td>
<td>$6.022 \times 10^{23}$ mol$^{-1}$</td>
</tr>
<tr>
<td>Boltzmann constant $k_B$</td>
<td>$1.381 \times 10^8$ m$^2$ kg s$^{-2}$ K$^{-1}$</td>
</tr>
<tr>
<td>Electron charge $e$</td>
<td>$1.602 \times 10^{-19}$ C</td>
</tr>
<tr>
<td>Faraday constant $\mathcal{F}$</td>
<td>$9.649 \times 10^4$ C mol$^{-1}$</td>
</tr>
<tr>
<td>Planck constant $h$</td>
<td>$6.626 \times 10^{-34}$ m$^2$ kg s$^{-1}$</td>
</tr>
<tr>
<td>Vacuum permittivity $\varepsilon_0$</td>
<td>$8.854 \times 10^{-12}$ F m$^{-1}$</td>
</tr>
</tbody>
</table>
Symbols

Roman symbols

*a*<sub>H</sub> host particles mean radius \( \text{m} \)

*a*<sub>T</sub> tracer particle radius \( \text{m} \)

*A*<sub>H</sub> Hamaker constant \( \text{J} \)

*B* mass second-virial coefficient \( \text{L g}^{-1} \)

*B* second-virial coefficient \( \text{L} \)

*c* massic protein concentration \( \text{g L}^{-1} \)

*c*<sub>m</sub> molar concentration of species \( i \) \( \text{mol L}^{-1} \)

*d*<sub>b</sub> width of the energy barrier (square potential) \( \text{m} \)

*d*<sub>f</sub> aggregate fractal dimension \( - \)

*d*<sub>w</sub> width of the energy well (square potential) \( \text{m} \)

*d*n/*dc refractive index increment \( \text{L g}^{-1} \)

*D*<sub>c</sub> collective diffusion coefficient \( \text{m}^2 \text{s}^{-1} \)

*D*<sub>s</sub> self-diffusion coefficient \( \text{m}^2 \text{s}^{-1} \)

*D*<sub>T</sub> tracer diffusion coefficient \( \text{m}^2 \text{s}^{-1} \)

*D*<sub>0</sub> diffusion coefficient under dilute conditions \( \text{m}^2 \text{s}^{-1} \)

*E*<sub>b</sub> energy barrier (square potential) \( \text{J} \)

*E*<sub>w</sub> energy well (square potential) \( \text{J} \)

*f*<sub>i</sub> number concentration of fibrils of mass \( i \) \( \text{N} \)

*F*<sub>Br</sub> force associated to Brownian motion \( \text{N} \)

*F*<sub>Dr</sub> drag force \( \text{N} \)
<table>
<thead>
<tr>
<th>Symbols</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{EI}$</td>
<td>force associated to the electrostatic potential</td>
<td>N</td>
</tr>
<tr>
<td>$F_i^m$</td>
<td>driving force acting on 1 mole of $i$</td>
<td>N mol$^{-1}$</td>
</tr>
<tr>
<td>$\bar{g}$</td>
<td>pair-correlation function</td>
<td>L</td>
</tr>
<tr>
<td>$G$</td>
<td>massic Kirkwood-Buff integral</td>
<td>L g$^{-1}$</td>
</tr>
<tr>
<td>$\bar{G}$</td>
<td>Kirkwood-Buff integral</td>
<td>L</td>
</tr>
<tr>
<td>$J_i$</td>
<td>flux of species $i$</td>
<td>mol L m$^{-2}$ s$^{-1}$</td>
</tr>
<tr>
<td>$k_{app}$</td>
<td>apparent aggregation rate constant</td>
<td>L$^{n_{app} - 1}$ s$^{-1}$</td>
</tr>
<tr>
<td>$k_D$</td>
<td>DLS interaction parameter</td>
<td>L g$^{-1}$</td>
</tr>
<tr>
<td>$k_E$</td>
<td>fibril elongation rate constant</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>$k_f$</td>
<td>prefactor fractal scaling</td>
<td>-</td>
</tr>
<tr>
<td>$k_F$</td>
<td>fibril fragmentation rate constant</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>$k_{i,j}$</td>
<td>aggregation rate constant of clusters of mass $i$ and $j$</td>
<td>m$^{-3}$ s$^{-1}$</td>
</tr>
<tr>
<td>$k_N$</td>
<td>fibril nucleation rate constant</td>
<td>L$^{nc - 1}$ s$^{-1}$</td>
</tr>
<tr>
<td>$k_s$</td>
<td>Smoluchowski rate constant</td>
<td>m$^{-3}$ s$^{-1}$</td>
</tr>
<tr>
<td>$k_U$</td>
<td>unfolding rate constant</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>$k_U^0$</td>
<td>unfolding rate constant in the absence of polyol</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>$K$</td>
<td>optical constant</td>
<td>m$^2$ mol g$^{-1}$</td>
</tr>
<tr>
<td>$K_{eq}$</td>
<td>equilibrium constant between $N_U$ and $N_S$</td>
<td>-</td>
</tr>
<tr>
<td>$K_{eq,ref}$</td>
<td>equilibrium constant $K_{eq}$ at 1 mol/L</td>
<td>-</td>
</tr>
<tr>
<td>$i$</td>
<td>aggregate dimensionless mass</td>
<td>-</td>
</tr>
<tr>
<td>$I$</td>
<td>solution ionic strength</td>
<td>mol L$^{-1}$</td>
</tr>
<tr>
<td>$I_s$</td>
<td>scattered intensity</td>
<td>cps</td>
</tr>
<tr>
<td>$I_{tol}$</td>
<td>toluene intensity</td>
<td>cps</td>
</tr>
<tr>
<td>$I_0$</td>
<td>solvent intensity</td>
<td>cps</td>
</tr>
<tr>
<td>$M$</td>
<td>concentration of monomeric species</td>
<td>mol L$^{-1}$</td>
</tr>
<tr>
<td>$M_i$</td>
<td>number concentration of aggregates containing $i$ units</td>
<td>L$^{-1}$</td>
</tr>
<tr>
<td>$M_0$</td>
<td>initial concentration of monomeric species</td>
<td>mol L$^{-1}$</td>
</tr>
<tr>
<td>$MW_{app}$</td>
<td>protein apparent molecular weight</td>
<td>g mol$^{-1}$</td>
</tr>
</tbody>
</table>
\[
\langle MW_{agg} \rangle \quad \text{aggregate weight-average molecular weight} \quad \text{mol}^{-1}
\]

\[n\quad \text{solvent refractive index} \quad -
\]

\[n_{app}\quad \text{apparent reaction order} \quad -
\]

\[n_C\quad \text{nucleus size for fibril formation} \quad -
\]

\[n_N\quad \text{dimensionless number-average aggregate mass} \quad -
\]

\[n_W\quad \text{dimensionless weight-average aggregate mass} \quad -
\]

\[N\quad \text{concentration of native monomer} \quad \text{mol L}^{-1}
\]

\[N_G\quad \text{number of Gaussian functions} \quad -
\]

\[N_S\quad \text{concentration of native monomer protected by polyol} \quad \text{mol L}^{-1}
\]

\[N_U\quad \text{concentration of native monomer undergoing unfolding} \quad \text{mol L}^{-1}
\]

\[P\quad \text{concentration of polyol} \quad \text{mol L}^{-1}
\]

\[q\quad \text{scattering vector} \quad \text{m}^{-1}
\]

\[r\quad \text{intermolecular distance} \quad \text{m}
\]

\[R_{c,i}\quad \text{collision radius of aggregate of mass } i \quad \text{m}
\]

\[R_{g,i}\quad \text{gyration radius of aggregate of mass } i \quad \text{m}
\]

\[\langle R_g \rangle\quad \text{average gyration radius} \quad \text{m}
\]

\[R_{h,i}\quad \text{hydration radius of aggregate of mass } i \quad \text{m}
\]

\[\langle R_h \rangle\quad \text{average hydrodynamic radius} \quad \text{m}
\]

\[R_p\quad \text{radius of monomeric protein} \quad \text{m}
\]

\[RSS\quad \text{Residual sum of squares} \quad -
\]

\[R_{tol}\quad \text{toluene Rayleigh ratio} \quad \text{m}^{-1}
\]

\[R_\theta\quad \text{Rayleigh ratio at angle } \theta \quad \text{m}^{-1}
\]

\[sr\quad \text{size ratio} \quad -
\]

\[S\quad \text{structure factor} \quad -
\]

\[t\quad \text{time} \quad \text{s}
\]

\[t_{RC}\quad \text{characteristic time for rapid coagulation} \quad \text{s}
\]

\[T\quad \text{temperature} \quad \text{K}
\]

\[u_i\quad \text{average velocity of species } i \quad \text{m s}^{-1}
\]
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U$</td>
<td>concentration of unfolded monomer</td>
<td>mol L(^{-1})</td>
</tr>
<tr>
<td>$V_{El}$</td>
<td>electrostatic interaction potential</td>
<td>J</td>
</tr>
<tr>
<td>$V_T$</td>
<td>total interaction potential</td>
<td>J</td>
</tr>
<tr>
<td>$V_{VdW}$</td>
<td>Van der Waals interaction potential</td>
<td>J</td>
</tr>
<tr>
<td>$W$</td>
<td>Fuchs stability ratio (indexes indicate collision type)</td>
<td>-</td>
</tr>
<tr>
<td>$x_i$</td>
<td>position of the (i^{th}) Gaussian</td>
<td>-</td>
</tr>
<tr>
<td>$z$</td>
<td>protein net charge</td>
<td>-</td>
</tr>
</tbody>
</table>

**Greek symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>coefficient of the (i^{th}) Gaussian</td>
<td>-</td>
</tr>
<tr>
<td>$\alpha_{PI}$</td>
<td>correction factor for protein interactions</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>activity coefficient of species (i)</td>
<td>-</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Kronecker delta</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta G_U^#$</td>
<td>Gibbs free energy of activation of unfolding</td>
<td>J</td>
</tr>
<tr>
<td>$\Delta H_U^#$</td>
<td>enthalpy of activation of unfolding</td>
<td>J</td>
</tr>
<tr>
<td>$\Delta S_U^#$</td>
<td>entropy of activation of unfolding</td>
<td>J K(^{-1})</td>
</tr>
<tr>
<td>$\Delta \nu$</td>
<td>preferential binding parameter</td>
<td>-</td>
</tr>
<tr>
<td>$\epsilon_r$</td>
<td>relative medium dielectric constant</td>
<td>-</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>zeta potential</td>
<td>V</td>
</tr>
<tr>
<td>$\eta_{eff}$</td>
<td>effective viscosity</td>
<td>Pa s</td>
</tr>
<tr>
<td>$\eta_{norm}$</td>
<td>normalized viscosity</td>
<td>-</td>
</tr>
<tr>
<td>$\eta_0$</td>
<td>solvent viscosity</td>
<td>Pa s</td>
</tr>
<tr>
<td>$\eta_\infty$</td>
<td>macroscopic viscosity</td>
<td>Pa s</td>
</tr>
<tr>
<td>$\theta$</td>
<td>light scattering detector angle</td>
<td>-</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>inverse Debye length</td>
<td>m(^{-1})</td>
</tr>
<tr>
<td>$\kappa_T$</td>
<td>transmission coefficient in transition state theory</td>
<td>-</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>exponent in aggregation kernel</td>
<td>-</td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>wavelength of laser beam</td>
<td>m</td>
</tr>
</tbody>
</table>
### Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Π</td>
<td>osmotic pressure</td>
<td>Pa</td>
</tr>
<tr>
<td>σ_H</td>
<td>variance of the host particle size distribution</td>
<td>m^{-2}</td>
</tr>
<tr>
<td>σ_0</td>
<td>protein (or particle) surface charge density</td>
<td>C m^{-2}</td>
</tr>
<tr>
<td>μ_i</td>
<td>chemical potential of species (i)</td>
<td>J mol^{-1}</td>
</tr>
<tr>
<td>μ_{el}</td>
<td>electrophoretic mobility</td>
<td>m^2 s^{-1} V^{-1}</td>
</tr>
<tr>
<td>ν</td>
<td>protein partial specific volume</td>
<td>L g^{-1}</td>
</tr>
<tr>
<td>(\overline{v}_i)</td>
<td>partial molar volume of species (i)</td>
<td>L mol^{-1}</td>
</tr>
<tr>
<td>ξ_{ij}</td>
<td>friction coefficient between species (i) and (j)</td>
<td>N m^2 s mol^{-2}</td>
</tr>
<tr>
<td>Φ</td>
<td>occupied volume fraction (exact)</td>
<td>-</td>
</tr>
<tr>
<td>Φ_0</td>
<td>initial occupied volume fraction</td>
<td>-</td>
</tr>
<tr>
<td>(\phi)</td>
<td>occupied volume fraction (approximate)</td>
<td>-</td>
</tr>
<tr>
<td>ψ_0</td>
<td>protein (or particle) surface potential</td>
<td>V</td>
</tr>
</tbody>
</table>

### Conversion of units

- \(J = \text{kg} \, \text{m}^2 \, \text{s}^{-2} = \text{C} \, \text{V}\)
- \(N = \text{J} \, \text{m}^{-1}\)
- \(\text{Pa} = \text{J} \, \text{m}^{-3}\)
Curriculum Vitae

Education

02/2012 - 12/2015 PhD studies at ETH Zurich (Switzerland) under the supervision of Prof. Morbidelli

02/2010 - 06/2011 Master thesis at ETH Zurich (Switzerland) in the group of Prof. Morbidelli

Thesis topic: Kinetic study of β-lactoglobulin amyloid fibril formation

09/2010 - 02/2011 Exchange semester at ETH Zurich (Switzerland) in the department of Chemistry and Applied Biosciences

09/2008 - 06/2011 Master studies in chemical engineering at École Nationale Supérieure des Industries Chimiques (Nancy, France)

09/2006 - 07/2008 Preparation for the entrance examinations to the French Grandes Écoles (Nancy, France)

06/2006 French baccalaureate

Work experience

06/2011 - 12/2011 Internship at UCB Pharma (Braine, Belgium) at the Cell Culture Process Development department

Internship topic: Aeration and agitation strategies of a bioreactor at the laboratory scale

07/2009 - 08/2009 Internship at Solvay (Rheinberg, Germany) at the Quality Control department


Conference presentations


