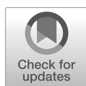

Chapter 4

Assessment of Therapeutic Antibody Developability by Combinations of In Vitro and In Silico Methods

Adriana-Michelle Wolf Pérez, Nikolai Lorenzen, Michele Vendruscolo, and Pietro Sormanni

Abstract

Although antibodies have become the fastest-growing class of therapeutics on the market, it is still challenging to develop them for therapeutic applications, which often require these molecules to withstand stresses that are not present in vivo. We define developability as the likelihood of an antibody candidate with suitable functionality to be developed into a manufacturable, stable, safe, and effective drug that can be formulated to high concentrations while retaining a long shelf life. The implementation of reliable developability assessments from the early stages of antibody discovery enables flagging and deselection of potentially problematic candidates, while focussing available resources on the development of the most promising ones. Currently, however, thorough developability assessment requires multiple in vitro assays, which makes it labor intensive and time consuming to implement at early stages. Furthermore, accurate in vitro analysis at the early stage is compromised by the high number of potential candidates that are often prepared at low quantities and purity. Recent improvements in the performance of computational predictors of developability potential are beginning to change this scenario. Many computational methods only require the knowledge of the amino acid sequences and can be used to identify possible developability issues or to rank available candidates according to a range of biophysical properties. Here, we describe how the implementation of in silico tools into antibody discovery pipelines is increasingly offering time- and cost-effective alternatives to in vitro experimental screening, thus streamlining the drug development process. We discuss in particular the biophysical and biochemical properties that underpin developability potential and their trade-offs, review various in vitro assays to measure such properties or parameters that are predictive of developability, and give an overview of the growing number of in silico tools available to predict properties important for antibody development, including the CamSol method developed in our laboratory.

Key words CamSol method, Developability, Immunogenicity, In silico, Prediction, Stability, Therapeutic antibody

1 Introduction

Antibody-based therapies are revolutionizing the treatment of major human conditions, including cancer, inflammatory and autoimmune diseases. At the time of writing, about 90 antibody drugs

Gunnar Houen (ed.), *Therapeutic Antibodies: Methods and Protocols*, Methods in Molecular Biology, vol. 2313,
https://doi.org/10.1007/978-1-0716-1450-1_4, © Springer Science+Business Media, LLC, part of Springer Nature 2022

are available on the market, and many more are currently in late-stage clinical studies or under regulatory review [1]. The development of antibody drugs, however, remains a slow and expensive process, which requires expert knowledge in many areas, including the discovery, characterization, and optimization of antibody molecules, manufacturing, pharmacokinetic and clinical studies, and regulatory affairs. Because of the challenges in the development and manufacturing of antibody drugs, as well as for specific commercial reasons, antibody-based therapies are typically substantially more expensive than small-molecule and peptide therapies, with a mean annual price of nearly $100,000 in the US [2].

In addition to the challenge of achieving the required biological activity, there are many other obstacles on the route of antibody candidates toward clinical reality. These include: (1) manufacturability risks such as poor expression, instability during virus inactivation or elution, and stickiness to purification columns; (2) formulation risks such as chemical and conformational instability, self-association, high viscosity and aggregation; and (3) in vivo issues such as lack of specificity, immunogenicity, precipitation upon administration, and rapid clearance. Taken together, these factors determine the likelihood of the successful development of an antibody into a stable, safe, and effective drug, which is known as developability.

In the last decade, the in vitro screening for these characteristics has become routine in most industrial pipelines [3] (Fig. 1). The implementation of developability screenings reduces the risk of late-stage failures by aiding the selection of those antibodies

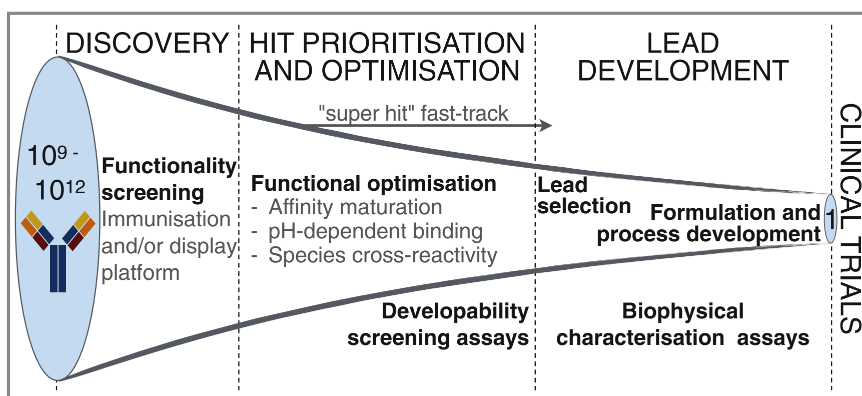

Fig. 1 Simplified pipeline of therapeutic antibody discovery and preclinical development. Schematic showing the three main phases of preclinical antibody development: antibody discovery, hit prioritization and optimization, and lead development. The specific procedures carried out at each phase may vary substantially from company to company and among different projects. At each phase, the number of antibody candidates is greatly reduced. Candidates that are taken forward are primarily selected based on functionality, but measurements of developability potential and of biophysical properties are increasingly employed to rule out potentially problematic candidates

embodying the best compromise between biological activity and developability potential at the early stages of a project. With the growing competition within the antibody pharmaceutical space, and the targeting of increasingly complex antigens, an optimization phase is often required after antibody discovery. During hit optimization, engineering campaigns are carried out most often with the aim of improving binding affinity, and sometimes also with the aim of introducing pH dependent target binding or species cross reactivity [4]. Engineering of affinity is a delicate maneuver, which typically requires careful counter screenings with developability assays, as mutations that improve affinity have often been reported to negatively affect other essential properties, such as stability, solubility, or specificity [5, 6].

Early developability screenings are often challenged by the large number of antibody candidates and by the preparation quality that are available at the early stages of preclinical development. Preliminary hits from antibody discovery campaigns are commonly prepared in minute amounts, low concentrations, and relatively low purity [7]. These factors may cause large errors and issues of measurement reproducibility [3]. While properties like conformational stability can now be measured with high accuracy and throughput using small amounts of material, other essential properties such as solubility, viscosity, and specificity remain much more challenging to assess in a quantitative way. For this reason, most of the developability assays that are usually implemented at the early stages of preclinical development are the result of a compromise between accuracy, throughput and material requirement. In particular, these screening methods do not attempt to directly measure properties such as solubility, aggregation, or viscosity, but they rather determine parameters that are easier to measure and considered to be predictive of these properties [7]. Because of the challenges associated with accurately measuring directly some biophysical properties, many developability assays probing self-association and solubility in different ways have been developed. Currently, comprehensive developability assessments require the combination of several in vitro screening assays, since no single assay appears to be fully predictive. This aspect makes developability assessment particularly demanding in terms of both time and resources [8].

Overall, the current limitations and challenges of experimental assessments of developability potential expose a pressing need for more effective, cheaper, and faster ways to assess antibody developability. Given the promising advances in the computational prediction of solubility, aggregation, and viscosity that have been made in the last decade [9–17], in silico predictors are emerging as a convenient alternative to experimental approaches due to their rapidity and lack of materials requirement. This review focuses on various aspects of antibody developability, with a focus on in silico predictions and their interplay with in vitro assays.

2 Requirements for Therapeutic Antibody Development

Antibodies for therapeutic applications are required to endure and survive a wide range of stresses related to manufacturing, development, shipping, storage, and administration in order to become safe and effective therapeutic agents approved for usage in patients [18–21] (Table 1).

2.1 Expression and Purification

Upon recombinant expression, antibodies may aggregate as a result of overexpression and can accumulate as inclusion bodies in cells or as aggregates in culture media, which may hinder yield and manufacturability [22]. Purification represents another challenge as antibodies are required to bind to different chromatography resins and must often endure low pH conditions. For example, during affinity purification using Protein A, which is a common unit operation during antibody purification, elution typically occurs at around

Table 1
List of common physical and biological stresses that therapeutic antibodies encounter during manufacture, development, storage, and administration

Stress	Manufacturing	Development and administration
Agitation/stirring	In fermentation tanks, during pumping and product transportation	Sample handling
pH stress	Low pH during Protein A purification and virus inactivation	Formulation, blood pH
Salt stress	Purification	Formulation, blood salt concentration
Very high concentration	Purification	Storage and subcutaneous administration
Shear stress	Ultrafiltration, diafiltration, pumping	Filtration, injection
Air water interface	Vial filling, storage	Vial filling, storage, transport
Solid interface	Steel tanks, membrane filters, purification columns, vial or syringe interphases	Vial or syringe interphases, biological membranes
Freeze/thaw cycles and temperature changes	Storage, transport	Storage (4 °C or frozen) vs injection (37 °C)
Prolonged storage		2 years stability of liquid formulation (4 °C), often including 2–4 weeks of room temperature stability

pH 3–4 followed by immediate pH adjustment to around pH 7–8 [23]. In addition, to ensure viral safety of antibody products destined for therapeutic usage, virus inactivation at low pH is commonly performed [24]. Moreover, antibodies need to withstand varying salt conditions during purification, as some buffers contain high amounts of salts. In theory, highly soluble and hydrophilic antibodies should feature a better tolerance against salt stress, while more hydrophobic antibodies may encounter problems such as aggregation or precipitation during purification [25, 26].

2.2 Sample Preparation and Transport

After purification, various filtration processes are performed, to remove residual contaminants, for product concentration, or buffer exchange. Extensive membrane contact has been associated to product loss and aggregation [20]. Once impurities are removed and a formulation selected, antibodies are collected into bags, bottles, stainless steel tanks or vials and may be frozen for long-term storage. Freezing prevents microbial growth in liquid formulations and eliminates foaming during transport. However, repeated freeze-thaw cycles can be associated with product deterioration and are an additional stress factor on the antibody [20].

During their lifetime, antibody drugs are exposed to a variety of interfaces that they would never encounter while performing their normal function in living organisms. These interfaces can be generally divided into air, solid, and liquid interfaces. Adsorption to these interfaces can impact the conformational stability and hence the activity of the drug [21, 27]. For example, during storage of liquid formulations, proteins will encounter an air–water interface in the headspace of vials. This interface adsorption is likely exacerbated by agitation caused by the transportation of the pharmaceutical product [28, 29]. As the air–water interface is relatively hydrophobic, the structure of adsorbed proteins may be slightly perturbed, thereby inducing aggregation [30, 31]. In addition, syringes are commonly treated with hydrophobic silicone oil, which may also induce aggregation [32, 33]. Finally, glass, plastic, and stainless steel are ubiquitous surface materials in therapeutic protein production, storage and administration equipment. Under high shear conditions or agitation, microparticles of these materials can contaminate protein formulations, which have also been reported to induce antibody aggregation [34, 35].

2.3 Storage and Administration

Most proteins are highly unstable in the gastro-intestinal tract, because of their susceptibility to enzymatic degradation and the low pH of the environment. Furthermore, large proteins like monoclonal antibodies have very low membrane permeability and therefore low oral bioavailability [36]. For these reasons, the vast majority of protein-based drugs are administered through parenteral (intravenous or subcutaneous) administration [37]. In

particular, antibody therapeutics that require frequent administration, such as those for chronic diseases, are typically administered subcutaneously [18]. This route of administration enables the rapid administration also outside the hospital, which is convenient for patients. However, highly concentrated antibody solutions ($\geq$100 mg/ml) are usually needed to achieve the relevant therapeutic dosing in small injection volumes (<1.5 ml). Therefore, therapeutic antibodies destined to subcutaneous injection must be formulated at concentrations much higher than those at which antibodies are typically produced in living organisms, and they must remain active at those concentrations over the shelf life of the product (typically $\geq$1 year), which is far longer than the typical in vivo half-life of antibodies from the immune system (~20 days). Formulation screenings aid to determine conditions (e.g., pH, ionic strength, and excipients) at which therapeutic antibodies are maintained in a stable, soluble state with low viscosity for a long time [18, 37, 38]. Nonetheless, these screenings can be resource-intensive and time-consuming, and are highly constrained by the requirements that the formulation must be safe for patients, and must also guarantee product stability upon delivery into the subcutaneous space until absorption [18, 38, 39].

Another challenge encountered upon administration is shear stress, which can occur during injection and may cause antibody aggregation [40, 41]. It has been suggested that aggregation-triggering factors are not only magnitude and duration of shear exposure, but also surface interactions (e.g., air–water interphase or solid-solid interactions). Surface adsorption and associated conformational perturbations may act synergistically in the presence of shear stress which could exacerbate antibody unfolding and aggregation [40, 41]. A further concern is the possibility of product precipitation at the injection site and/or altered pharmacokinetics resulting from the complexity of the human subcutaneous region [42]. Strategies to mitigate the risk of precipitation and increase the bioavailability of monoclonal antibody formulations are reviewed in Refs. 39, 42.

Overall, the stringent requirements of therapeutics applications, and in particular the constraints imposed by manufacturing, development, shipping, storage, and administration, imply that most biophysical properties of antibodies must be optimized beyond their typical natural levels [6]. Therefore, biophysical properties including thermodynamic stability and solubility, but also chemical liabilities like oxidation and deamidation, play a key role in determining the success of therapeutic antibody development [8, 15, 43].

3 Biophysical and Biochemical Determinants of Developability

The clinical success of an antibody candidate is determined by its biological activity as well as by many additional extrinsic and intrinsic factors. Extrinsic factors include administration dosage and frequency, administration route, formulation components and concentration, and selected manufacturing and storage conditions, while intrinsic factors are biophysical and biochemical properties of the antibody itself, which ultimately determine its developability potential. Highly optimized biophysical properties often translate into high expression, high conformational and colloidal stability, low viscosity, and often even low immunogenicity and polyspecificity [3, 8]. In this section we briefly discuss the key biochemical and biophysical properties that underpin developability potential of antibodies.

3.1 Chemical Stability

Deamidation, isomerization, oxidation, and improper glycosylation are examples of chemical liabilities that may occur in antibodies and that can be linked to safety and efficacy issues [44]. Deamidation of asparagine, succinimide formation, and isomerization of aspartate residues are some of the most common chemical liabilities encountered during manufacture, storage and after in vivo administration [3]. Deamidation and isomerization reactions are highly dependent on the solution pH and the storage temperature, as well as on the structural context in which the residues are found. In particular, asparagine and aspartic acid in more flexible regions, such as the CDR loops, are more prone, respectively, to deamidation and isomerization than those found in rigid regions, such as the framework. As CDR loops are often critical for binding, this fact translates into an increased chance of chemical liabilities impacting functionality. Asparagine residues followed by certain sequence motifs including glycine, serine, threonine, aspartate, and histidine have been reported as degradation hotspots [3, 9, 44]. Experimentally, degradation propensities can be tested by subjecting samples to temperature and pH stresses like slightly acidic (e.g., 2 weeks at 40 °C at pH 5–6) and mildly basic (e.g., 2 weeks at 40 °C at pH 7.4) conditions. Strategies to mitigate these reactions include re-engineering of the antibody hotspot and/or formulation optimizations to reduce the risk of deamidation and isomerization during storage [3]. A recent analysis of 131 clinical-stage mAbs found a relatively high frequency of deamidation and isomerization under low and high pH stress, highlighting the key role that formulation is likely to play in preventing the occurrence of these liabilities in the pharmaceutical products assessed [45].

Oxidation of methionine and tryptophan residues represents another chemical liability. Reactive oxygen species can be generated in antibody formulations in the presence of light [46], surfactants

[47] and metal ions and surfaces [48, 49] which may cause oxidation and the formation of aggregates [49]. Antibody formulations can be subjected to forced oxidation reactions (e.g., with 2,2-azobis(2-amidinopropane)dihydrochloride) to probe for oxidation propensities of methionine and tryptophan hotspots [50]. The addition of antioxidants, such as methionine, to formulations may stabilize the antibodies and mitigate oxidation [51].

The production of homogenous antibodies with consistent human-like glycosylation profiles can also be challenging [52]. Consistent glycosylation patterns are important as the position of the F_c region carbohydrate can impact the functional properties of an antibody [53] and determine in vivo clearance rates [54]. Potential *N*-glycosylation sites can be identified by the consensus motifs NXS or NXT (X can be any amino acid except proline) in the sequence [3]. Lysine residues are known to be subjectable to glycation. Glycation is a post-translational modification that occurs when the amine group of the lysine reacts with reducing sugars that can be present during manufacturing in cell media, storage, or in vivo after administration [3]. Experimentally, the glycation susceptibility can be determined by forced glycation assays, where antibodies are incubated in the presence of reducing sugars and analyzed using mass spectrometry methods [3]. Lastly, unpaired cysteine residues may under certain conditions lead to disulfide scrambling, leading to covalently linked dimers and oligomers, which may constitute an efficacy as well as a safety issue [55].

To mitigate the occurrence of chemical liabilities antibody sequences harboring unpaired cysteine, methionine and tryptophan residues at non-conserved positions, or other known liability hotspots are often excluded during lead selection. Alternatively, mutagenesis is carried out to remove such liabilities, which entails reassessing the activity and structural integrity of the mutated antibodies [56–59].

3.2 Conformational Stability and its Link to Aggregation

Conformational stability is an essential property of antibodies to ensure efficacy and safety during manufacturing, formulation, refrigerated storage, distribution, and administration [20, 60]. Furthermore, various studies [8, 61, 62], have reported a correlation between conformational stability and recombinant expression yield from yeast and mammalian cells. The thermodynamic stability of an antibody determines the equilibrium between the unfolded and the native states (Fig. 2). Fully or partially unfolded antibodies not only lose their activity, but can readily aggregate, as the solubility of the unfolded state is typically lower than that of the native state [63]. Aggregation can be irreversible, and aggregated antibodies have been reported to induce immunogenicity upon injection [18, 64, 65], which can be fatal in some cases [66]. Although it is still not clear which types of aggregates mediate immunogenicity,

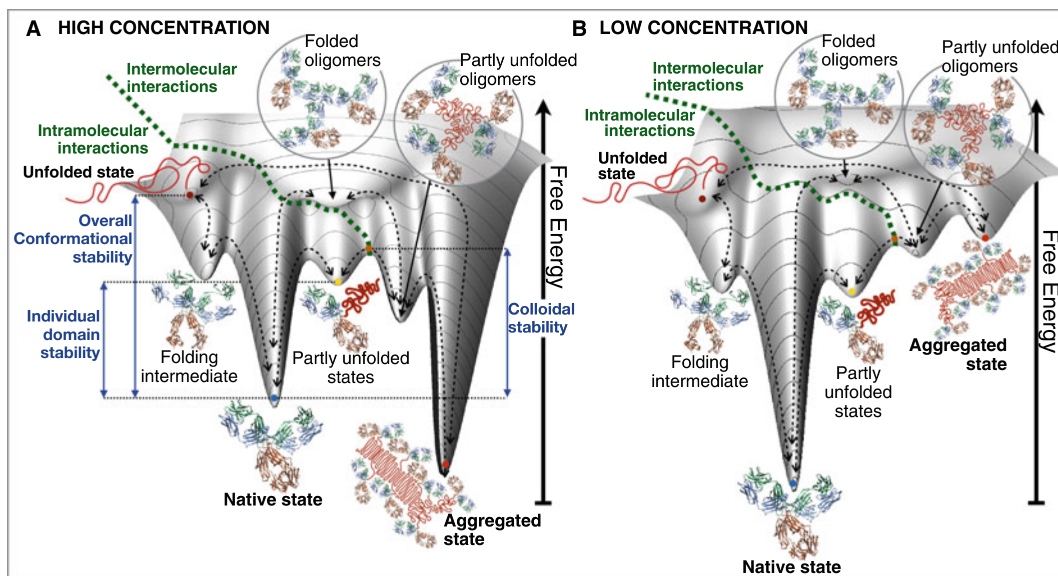

Fig. 2 Competition between folding, unfolding, and aggregation in antibody formulations. (**A**, **B**) The different states that antibodies can populate in formulations are concisely represented through free-energy surfaces. The green dotted line separates the region of the landscape where the enthalpic contributions to the free energy come from intramolecular interactions (folding and unfolding), from that where these contributions come also from intermolecular interactions (self-association and aggregation). The roughness of the free-energy landscape reflects the highly dynamical nature of antibodies, and it results in the presence of a variety of conformational states that populate the different local minima. At equilibrium, these conformations interconvert with each other (following for example the gray dashed arrows) with rates that depend on the height of the free-energy barriers that separate different minima, while the relative populations of the various states depend on the free-energy differences between the minima. Some of these conformations, such as the aggregated state in panel **A**, are more kinetically trapped than others, as to escape their local minimum and reach other states they need to climb up high free-energy barriers. The vertical blue arrows in panel **A** highlight some particularly important free-energy differences, which correspond to biophysical properties that are commonly measured and that underpin developability potential. For instance, the conformational stability is the free-energy difference between the native and the unfolded state. Conversely, colloidal stability is a kinetic property that depends on the height of the free-energy barrier that separates the native state from the aggregated states (the free energy needed to cross the green line). The overall shape of this landscape, and consequently all these biophysical properties, depend on the amino acid sequence of an antibody, but also on extrinsic properties such as protein concentration, formulation composition, and temperature. Panel **B** is a schematic landscape of a low-concentration formulation, or alternatively of a highly optimized antibody that remains very soluble even at high concentration. The ideal landscape of a pharmaceutical formulation should look more like that in panel B and should retain, even at very high concentrations, one single deep minimum corresponding to the native state, with few other competing states at much higher free energies

regulatory agencies are cautious on this aspect, and require formulations with a minimum amount of aggregates at the end of formulation shelf life to grant market approval [67–69].

Under physiological conditions monoclonal antibodies typically have a thermodynamic stability corresponding to a Gibbs free energy of unfolding (ΔG^u) in the range of 10–20 kcal/mol,

depending on the antibody under scrutiny as well as buffer and temperature [70, 71]. In particular, as monoclonal antibodies are multi-domain proteins, the stability of individual domains may be significantly lower than the total net apparent stability [70, 71], and partial unfolding is often enough to trigger aggregation [72] (Fig. 2). A ΔG^{u} of 10 kcal/mol corresponds to an equilibrium constant $K = [D]/[N] = e^{-\Delta G/RT}$ of approximately 10^{-8} at 4 °C, [73] chosen as the typical temperature of refrigerated storage. This number indicates that unfolding is rare, as only one of every 10^{8} antibody molecules is in the unfolded state when the solution is at equilibrium, or, in other words, the native state population is 99.999999%. Nonetheless, a solution containing 50 mg of monoclonal antibodies, which is at the lower end of the spectrum typical of pharmaceutical formulations, contains 2×10^{17} molecules (MW ~ 150 kDa), which means that more than 2 billion molecules are in the unfolded state at any given time at equilibrium, and can act as seeds of aggregation, or create a variety of other problems including, for example, eliciting an immunogenic response. However, given that the equilibrium constant depends exponentially on the thermodynamic stability, a ΔG^{u} of 20 kcal/mol instead of 10 brings the number of unfolded antibody molecules in every 50 mg of formulation from more than 2 billions to only about 30. The difference between these two numbers highlights the importance of selecting therapeutic candidates with the highest possible intrinsic stability, in order to maximize their developability potential. In summary, the presence of unfolded molecules in solution is not only linked to loss of function, but also to aggregation, immunogenicity, and ultimately safety [18, 64, 65]. Moreover, the intrinsic stability of an antibody translates into its ability to withstand physical or chemical stresses (Table 1), while retaining its structural conformation and activity.

3.3 Solubility

Owing to the stringent requirements of therapeutic applications, which demand antibody drugs to be formulated at very high concentrations and to remain soluble and active for the shelf life of the product, solubility is a key biophysical property underpinning developability potential [18, 20, 74]. Poor solubility is a major bottleneck for manufacturing [19, 20], quality, and safety of pharmaceutical formulations [18, 19, 43]. The solubility of an antibody is dependent on its amino acid sequence, its net charge and spatial charge distribution, as well as on other extrinsic factors, like storage temperature and buffer composition [75]. However, the solubility of complex macromolecules like monoclonal antibodies cannot be defined in absolute terms, which makes quantitative assessments highly problematic. The thermodynamic solubility of a substance is an equilibrium property defined as the value of the concentration—termed critical concentration—at which the soluble and insoluble states are in equilibrium. While this definition is rigorous, it only

applies directly to substances that have just two relevant states, a soluble state (the liquid phase) and an insoluble state (the solid phase) [75, 76]. The vast majority of proteins, including antibodies, are not substances of this type. Depending on the concentration, most proteins populate a variety of states, including dimers, various oligomers, large amorphous aggregates and precipitates, and occasionally fibrils. Given this heterogeneity, the boundary between the soluble and insoluble states is ultimately arbitrary and operationally dependent on the method used to separate the two phases (e.g., on centrifugation speed or filter size). This aspect, and the fact that some pathways may lead to irreversible aggregation in the timescales relevant for therapeutic formulations, complicate the definition of protein solubility as an absolute quantity, which poses important limitations to our ability to measure solubility as an absolute value [7, 63]. Despite this problem, it is possible to measure solubility differences among antibodies or among different formulations of the same antibody, or at least differences in the propensity to self-associate, precipitate, or populate aggregated states, which are common proxies for solubility across the literature [25, 75, 77–80]. However, these relative measurements are not comparable across different experimental assays, or across different conditions, including buffer composition, temperature, incubation times, protein concentration, centrifugation speed and time, size of filters, type of precipitant or chromatography columns. Moreover, measurements of solubility difference, especially when carried out at the early stages of preclinical development, when the material available is little and at low purity, may have poor reproducibility and can often be significantly affected by impurities [3].

3.4 Colloidal Stability and its Link to Self-Association, Aggregation, and Viscosity

The overall stability of high-concentration formulations on the time scale of the product shelf-life is not only defined by conformational stability and solubility, but also by colloidal stability. Colloidal stability strongly depends on intermolecular interactions, and describes the free-energy barrier between the native and aggregated states [81] (Fig. 2). While solubility is defined as the critical concentration observed when soluble and insoluble phases are in equilibrium, colloidal stability is defined by the long-term integrity of a formulation, and hence by the time it takes for aggregation to occur. In some instances, the timescales of these processes can be very long, so that in practice an antibody may be formulated around its critical concentration and still constitute a viable clinical product.

Antibodies are generally stable at low concentrations, but their behavior changes at high concentrations (50–150 mg/ml). Decreased intermolecular distances between antibodies at high concentrations lead to an increased likelihood of molecular collisions and reduce the entropic penalty of association [82], so that attractive short-range interactions may overcome repulsive longer-

range interactions, such as repulsion between equal charges [77]. Consequently, the formation of reversibly associating intermolecular complexes may occur [83], and these may act as seeds for the formation of kinetically trapped aggregates [84].

Both reversible self-association and kinetically trapped aggregation can further result in dramatic increases in viscosity, which negatively impacts manufacturing and patient administration [19, 77, 85]. High viscosity formulations can cause pain upon injection [85], and can also be a manufacturing challenge by blocking processing pumps [18]. For example, attractive electrostatic interactions e.g., between Fab-Fab domains [77] are often implicated in governing self-association and viscosity behaviors of antibodies. Also, an increased positive charge in the CDRs was observed to correlate with an increased risk of high self-association of antibodies [86]. For instance, it has been shown that multiple consecutive arginine residues in an antibody sequence can facilitate nonspecific antibody interactions [86, 87]. On the other hand, also highly hydrophobic CDRs have been reported to be involved in aggregation, viscosity and poor specificity [9, 88–90].

Overall, while solubility and conformational stability may be regarded as intrinsic properties of an antibody, colloidal stability and viscosity directly depend on the antibody concentration in the formulation under scrutiny. Besides concentration—surface hydrophobicity, charge heterogeneity, and net charge—are factors that contribute to the colloidal stability of an antibody formulation [91–94]. Controlling the colloidal stability is crucial in order to mitigate antibody self-association and reduce viscosity and aggregation. Approaches to confer antibodies with resistance toward self-association include formulation optimization [85] and mutagenesis [56–59].

3.5 Specificity

A fundamental requirement for antibodies is the ability to specifically bind their targets while avoiding interactions with other molecules [95, 96]. However, some antibodies have been reported to nonspecifically interact with a variety of molecular targets besides their intended one [97–100]. For example, antibodies generated by immature B-lymphocytes tend to be less specific than those originating from mature ones [101, 102]. Similarly, antibodies derived from synthetic libraries have sometimes been found to be more problematic in the development phase than immune-system-derived ones also because of poor specificity [8, 103–106]. For example, phage-derived antibodies have been reported to often feature solubility issues [103, 105, 106] and substantial nonspecific interactions [104], while antibodies resulting from immunization techniques are commonly more specific due to the selection carried out in vivo by the immune system [106].

Nonspecificity has been linked to poor solubility [96] and low expression [96], but also to poor antibody pharmacokinetics, including impaired bioavailability [99], faster clearance rates [97, 107], and toxicity [108]. Therefore, the avoidance of nonspecific interactions is a crucial parameter underpinning developability potential.

3.6 Immunogenicity

Immunogenicity depends on extrinsic factors including administration concentration, frequency of dosing, formulation, and patient medical background, but also on sequence composition and several biophysical properties [109]. For example, conformationally unstable, unfolded and aggregated antibodies can trigger an immunogenic response upon injection by eliciting a humoral or cell-mediated immune response [64]. Aggregated antibodies are often more immunogenic than their monomeric form mainly due to the exposure of epitopes that mimic pathogenic patterns [64].

Another common source of immunogenicity is the exposure of non-human (non-endogenous) sequence patterns. Mouse-derived antibodies have been observed to harbor high immunogenicity risks in humans, which have been ameliorated by using chimeric, humanized or fully human antibodies [110]. Efforts to overcome this limitation of antibodies obtained from animal immunization include the development of transgenic mice that produce fully human antibodies following immunization [111, 112], and of in vitro display methods to screen fully human antibody libraries [113–116].

Immunogenicity is commonly induced through the exposure to T-helper cells of epitopes of non-human origin, or that are usually buried in the protein structure but become surface exposed in partially or fully unfolded species. Upon T-helper cell activation, cytokine release initiates an immune response that leads to the generation of B-cell-derived anti-drug antibodies [65]. Patients may experience side-effects, which can range from local skin inflammation at the injection site and fever, to acute anaphylaxis and systemic inflammatory response syndrome, which may be fatal [66]. Becoming immune to an antibody therapeutic can be a major problem, when there are no other treatments available for patients that are resistant to a given drug [43]. In rare cases, patients can also develop immunity toward endogenous proteins as reported for antibody-mediated pure red cell aplasia [117]. Overall, immunogenicity is an important property to determine the clinical success of an antibody drug candidate.

3.7 Biochemical and Biophysical Properties are Closely Interlinked

The balance between the unfolding and folding rates determines conformational stability of antibodies. The folding process is typically a first-order reaction that depends on the formation of intramolecular interactions, and it is therefore concentration independent [118]. In contrast, protein aggregation relies on

intermolecular contacts, and therefore is a second or higher order reaction, which is strongly dependent on protein concentration [72, 119] (Fig. 2). A key driving force of folding is hydrophobicity, which leads to the collapse of hydrophobic motifs in the sequence forming the so-called folding nucleus, where they are no longer exposed to the solvent [120, 121]. However, in fully or partly unfolded states such motifs can remain solvent-exposed, and particularly in high-concentration formulations, where intermolecular interactions become much more likely, they can readily trigger aggregation instead of folding.

Therefore, conformational stability, colloidal stability, and solubility are closely intertwined in biopharmaceutical formulations. The presence of partly of fully unfolded states significantly lowers the free-energy barrier that determines colloidal stability and favors the formation of partly unfolded oligomers and larger aggregates. These misfolded aggregates are typically more stable and kinetically trapped than folded oligomers, which can readily dissociate back to the native state (Fig. 2). In turn, the fact that when conformational stability is low, misfolded aggregates can readily form via partly unfolded intermediates lowers the critical concentration of the system, and hence the solubility. In other words, the concentration needed for intermolecular interactions, which drive the formation of aggregated states, to become dominant over intramolecular interactions, which drive the refolding, is much lower for a poorly stable antibody than for a highly stable one, as the presence of partly or fully unfolded states lowers the free-energy barrier to reach the aggregated states.

This complex picture describing the interplay among different biophysical properties emerges from the fact that the fundamental forces that drive folding, such as hydrophobic and polar interactions, are the same ones that determine protein aggregation, as well as antibody binding and specificity. It is therefore highly challenging to identify mutations that selectively affect one biophysical property while leaving the others unaffected. Amino acid substitutions that improve conformational stability have been observed to negatively impact solubility, probably because increased hydrophobicity stabilizes the native state, but also facilitates self-assembly [122]. Conversely, because increased conformational stability usually increases the free-energy barrier to reach the aggregated state, solubility and conformational stability have often been reported to correlate [123, 124]. Moreover, surface mutations that would in principle improve solubility can sometimes cause an increase in conformational dynamics leading to the transient exposure of otherwise buried hydrophobic patches, which may actually elicit aggregation [125–127]. In the case of antibodies, binding often requires long and irregular CDRs, and the presence of solvent-exposed hydrophobic residues, which may negatively impact both stability and solubility [128–130].

Regarding binding affinity and specificity, a recent analysis of over 400 antibody–antigen complexes confirmed previous reports that paratope residues including Tyr, Trp, Ser, Asn, Arg, and Gly contribute substantially to the interactions between antibody and antigen, thus driving binding affinity [131]. However, while binding affinity and specificity are typically correlated, the enrichment of Trp, Arg, Gly, and Val has also been associated with the occurrence of poor specificity [87, 132]. Interestingly, Tyr, Ser, and Trp are prevalent in antigen-contacting residues in germline antibodies, but not among mutations introduced by the somatic hypermutation mechanism carried out by the immune system, thus suggesting a role of this process in suppressing nonspecific interactions [104]. Generally speaking, complementary electrostatic interactions and the formation of antibody–antigen hydrogen bonds have been associated with the occurrence of specificity, while aromatic and hydrophobic interactions, mostly occurring with the epitope main chain atoms and side-chain carbons, with binding affinity [133].

Taken together, these apparently contrasting findings, and the existence of many cases contradicting the expected average behavior, show that it is very challenging to extrapolate universally applicable rules to explain the molecular basis of the balance between antibody biophysical properties. Overall, the determinants of stability, solubility, and interaction affinity and specificity appear to be highly context-dependent, and rather specific to individual antibody–antigen complexes [6]. This fact greatly complicates the experimental assessment of developability potential, as well as the development of effective in silico tools to predict it or rationally improve it with mutations.

4 Experimental Measurements of Developability Potential

Until recently, biophysical and biochemical properties of antibody candidates have been assessed at the late stages of the preclinical development process, as some assays can be highly material-demanding and time-consuming [3]. Therefore, those lead candidates with excellent biological activity but poor developability would only be identified late in the pipeline, when a significant amount of time and resources had already been deployed toward their characterization [43]. The high attrition rate experienced following this strategy has prompted and is still fuelling the development of cost-effective, de-risking approaches to reduce late-stage failures. In the last two decades, many novel in vitro assays have been introduced that are readily applicable at the early stages of antibody discovery campaigns. These assays require little material and are typically relatively high throughput to enable the screening of a large number of hits.

In this section, we divide in vitro assays to measure developability potential into "biophysical characterization methods" and "developability screening methods." Biophysical characterization methods are generally traditional biophysical assays that are low throughput and resource-intensive, but directly measure biophysical properties of interest. These methods are typically used in the late stages of preclinical development in the therapeutic antibody pipeline (Fig. 1). Conversely, developability screening methods usually require small amounts of material and have higher throughput, but often measure parameters that are predictive of some underlying biophysical property, and in some cases, especially when the antibody purity is not very high, may have low reproducibility. Quite generally, the implementation of developability screening assays in a pipeline enables to rule out potentially troublesome antibody candidates already at the early stages of preclinical development [3, 8, 10, 134–136].

4.1 Biophysical Characterization Assays

4.1.1 Assays Probing Conformational Stability

Many biophysical methods are available to assess the conformational stability of proteins and antibodies. For example, the integrity of the folded state can be assessed by looking at secondary structure changes with far-UV circular dichroism (CD) spectroscopy or Fourier transform infrared (FTIR) spectroscopy [137–139], while tertiary structural changes may be assessed by near-UV CD [140, 141]. These methods are quite material-demanding and low throughput, as they require relatively large volumes and/or high concentrations and need very high optical quality, which challenges the use of multi-well microplate readers [140, 141]. CD and other techniques like differential scanning calorimetry (DSC) can be used to determine thermal stability. With CD and DSC, the signal is monitored as a function of increasing temperature, and the melting temperature T_m is defined as the midpoint of the unfolding transition at which 50% of the antibody is unfolded. The higher the T_m value, the higher the required energy to unfold an antibody. Higher T_m values therefore mean higher conformational stability, albeit the correlation is not perfect when comparing widely different molecules [142, 143]. Common methods to measure directly the conformational stability (ΔG_u) rely on isothermal chemical denaturation induced by urea or guanidine hydrochloride (GuHCl). Either tryptophan fluorescence or CD spectra are recorded at increasing amount of denaturant to monitor the unfolding process, and a suitable model (e.g., two-state, three-state) is fitted to the data to obtain an apparent free energy of unfolding ΔG_u [73].

4.1.2 Assays Probing Colloidal Stability

Antibody–antibody interactions can be quantified by the non-ideality parameters osmotic second virial coefficient B_{22} and the diffusion interaction parameter k_D. For example, B_{22} describes the magnitude and type of self-interaction (attractive or repulsive)

between two protein molecules in solution [144, 145] and can be determined by analytical ultracentrifugation (AUC) [146] and dynamic light scattering (DLS) [147]. The B_{22} value has been shown to be predictive of protein solubility [148–150], aggregation, and viscosity issues at high concentrations [146]. B_{22} measurements, however, remain material-demanding and time-consuming [151]. k_D determinations are usually conducted by DLS. The measurements are based on the fact that the diffusion of an antibody in solution is impacted by size and intermolecular interactions between antibodies. The diffusion coefficient in strongly interacting systems, such as high-concentration formulations, is referred to as mutual diffusion coefficient (D_m) [82, 152, 153]. D_m is determined at increasing antibody concentrations (typically from 1 to 20 mg/ml) and the k_D values are obtained by a linear fit of the measured D_m as a function of concentration [151, 154]. A negative k_D value translates to an increase in attractive intermolecular interactions, while a positive k_D indicates repulsive protein–protein interactions [153, 155]. k_D has been reported to be predictive of solubility [76, 148] and viscosity issues of antibodies in some cases [151, 155]. However, k_D may be a poor predictor of aggregation behavior in scenarios, where the hydrophobic interaction may become the governing force in solution and outbalance electrostatic interactions [84].

4.1.3 Assays Probing Aggregation and Aggregation Propensity

Aggregates can arise from non-covalently, or more rarely covalently (e.g., intermolecular disulfide bonds), linked species, and typically vary widely in size and structural properties. Different types of aggregates may be more or less kinetically trapped (*see* Fig. 2), which means that aggregation may effectively become irreversible on relevant timescales. Those aggregates that do not sediment upon centrifugation or cannot be removed by filtration are often termed soluble aggregates, albeit this definition is ultimately dependent on time and speed of centrifugation or filter size, whereas insoluble aggregates sediment and can be cleared from the solution [156]. The exact mechanism of aggregation and the type of aggregates formed in high-concentration formulations strongly depends on the antibody, concentration, formulation conditions (e.g., pH and excipients), and other external stresses (e.g., agitation, interphases, freeze–thaw cycles, and temperature) [157–160]. Common methods to detect and quantify aggregates include turbidity measurements (optical density at 340 nm), size exclusion chromatography (SEC), AUC and DLS. Antibodies are often quite stable towards aggregation, and therefore it typically takes a long time (weeks to months) before significant amounts of aggregates are formed. As aggregation reactions are highly concentration-dependent, studies of antibody aggregation should ideally be performed at formulation-relevant concentrations which are normally

in the range of 50–150 mg/ml. Therefore, all of these methods are material-demanding and time-consuming. Besides, some of these methods, including SEC, involve the dilution of protein samples, which prevents the detection of highly reversible aggregates [78, 158, 161, 162].

Therapeutic antibody candidates are often subjected to stress tests during formulation development in order to determine their stability toward different stresses and to predict their propensity to aggregate upon long-term storage [163–165]. Some of the most common stress tests include thermal stress, mechanical stress, and freeze–thaw cycles [165, 166]. For example, a few milligrams of antibody material are often incubated at elevated temperatures (e.g., 40 °C) for several weeks. Increasing temperature facilitates conformational changes in the protein, such as local unfolding, which may induce aggregation. Concomitantly, elevated temperatures increase the diffusion of proteins in solution thus increasing the rate of collisions, which ultimately leads to self-association, aggregation and chemical degradation reactions [164, 165]. Many studies have attempted to predict the shelf life based on extrapolation of stability data obtained from elevated temperature experiments to actual storage conditions (1–2 years refrigerated). However, antibodies are complex and large proteins—and so are their aggregation pathways. Different protein conformations have been reported to be populated at different temperatures, pH, and ionic strength conditions, which may lead to different types of aggregation [167–169]. Ultimately, the aggregation of multi-domain proteins like antibodies may not follow simple kinetics over a wide temperature range but rather exhibit non-Arrhenius behavior [164, 165, 168]. This means that the rate constant of aggregation is non-linear with temperature, which has been shown for various antibodies [170]. Also, the melting temperature (T_m) has been used in several studies to predict long-term storage of therapeutic proteins. Some studies suggest a correlation between T_m and stability upon storage [165, 171]. However, T_m measurements may not be predictive of long-term storage in cases where the predominant aggregation process or the rate-limiting step are not related to partial or full unfolding of a protein but rather on the formation of native oligomers [165].

4.1.4 Assays Probing Solubility

Solubility may be experimentally determined as the maximum concentration of antibody remaining in solution without forming precipitates, gels, crystals, or soluble aggregates, for example following harsh centrifugation (e.g., 30,000 × *g* for 30 min) or filtration [156]. Nonetheless, as discussed in Subheading 3.7, the solubility of complex macromolecules is ultimately poorly defined and therefore not measurable as an absolute quantity. For example, in centrifugation assays performed at increasing total concentration, the

supernatant concentration often does not reach a clear plateau when plotted as a function of the total concentration, but rather gradually deviates from the unity line [172]. In addition to these intrinsic challenges, solubility measurements are also impractical, because of the need of large amounts of purified protein (tens to hundreds of milligrams), and of experimental difficulties resulting from the high heterogeneity of the aggregated states [78]. Nonetheless, ultrafiltration or ultracentrifugation assays are commonly employed to rank different antibodies or to compare different formulations in the late preclinical development stage. In up-concentration experiments, solubility is measured as the maximal protein concentration that can be achieved in ultrafiltration devices with a molecular cut-off filter. Disadvantages of ultrafiltration are high material consumption, adsorption losses and shear-induced degradation. Another issue is that up-concentrated protein solutions can exhibit high viscosity. Viscosity can become problematic when membrane pore clogging occurs, which can limit reaching high concentration. Gel formation can also impede solubility measurements [18, 78]. Another method to measure solubility is by addition of lyophilized protein to a solution until it becomes saturated and its solubility limit is reached [75]. Insoluble protein is removed by centrifugation or filtration [173], and the maximum protein concentration in the soluble supernatant measured [156]. Alternatively, the apparent protein solubility can be determined by protein precipitation assays, which rely on additives like ammonium sulfate or polyethylene glycol (PEG) [25, 79, 80, 174]. These assays have significantly lower material requirement than the aforementioned methods (but still require a few milligrams), and are more amenable to high-throughput automation [80, 175]. There also exist several reports which suggest that measurement of colloidal stability (described previously) can give an indirect measure of solubility and be used for ranking of molecules [148–150].

4.1.5 Assays Probing Viscosity

Viscosity is highly relevant for drug manufacture and administration, since highly concentrated antibody solutions are typically pushed through narrow tubing and needles as previously discussed. In the pharmaceutical setting, dynamic viscosity is the most common parameter to measure solution viscosity, as it reflects directly how a solution resists a flow when an external force is applied. Dynamic viscosity increases with protein concentration, and it is governed by the volume exclusion effect of antibodies and formulation excipients [176, 177].

The most common method to measure viscosity is the cone-and- plate rheology method, which enables high accuracy measurements [9, 151, 178, 179]. Viscosity measurements should be carried out at varying antibody concentrations [9, 151], up to ideally

150 mg/ml, in order to enable reliable ranking of different antibodies and formulations at pharmaceutically relevant conditions [9, 151], which makes these measurements very resource-intensive. An expedient that has been proposed in order to lower material requirements is to exploit the strong temperature dependence of viscosity measurements. Geoghegan et al. [178] have demonstrated that robust viscosity ranking of different antibodies at 70 mg/ml can be obtained by simply lowering the measurement temperature to 4 °C.

Microfluidic capillary viscometers and microrheology technologies (reviewed in Ref. 180) also represent low-throughput methods, but may require less material than cone-and-plate rheology techniques. Another way to measure viscosity of antibody formulations is to use DLS as an indirect method. Here, polystyrene beads are added to antibody solutions and from the diffusion coefficient of the beads, which are much larger than antibodies, the solution viscosity can be calculated [181–183]. Due to the need of high amounts of antibody material and of the labor-intensive process of protein concentration, predictive methods are very often used to evaluate whether a molecule may be associated with viscosity risks. For example, the determination of protein–protein interactions using e.g., the diffusion interaction parameter (k_D) described previously is commonly used as a predictive proxy for viscosity [184].

4.2 Developability Screening Assays

Several in vitro methods are now available to aid the prediction of the likelihood of clinical success (i.e., developability) of drug candidates at the early stages. These methods indirectly measure biophysical properties using minute amounts of material (usually a few μg), run in little time (few minutes to a couple of hours), and are often high throughput. Therefore, their implementation in the pipeline holds the promise of streamlining the selection of the fittest from the very beginning of a drug discovery campaign [3, 8, 10, 134–136]. However, these methods most often do not directly measure solubility, aggregation, and viscosity, but rather determine surrogate parameters reported to be predictive of these properties [3, 8, 135]. Antibody candidates are then ranked according to these parameters, which can be helpful to exclude biophysically challenging candidates from further characterization.

Overall, early developability assessments aim to bridge the gap between drug discovery and drug development, which are traditionally not fully connected processes. For example, manufacturability and safety constraints are per se not considered in the discovery process, while these are of utmost importance for development and progression into clinical trials. Currently, there is no single in vitro assay that can fully predict the developability potential of an antibody candidate, as developability is a broad concept that encompasses several interlinked biophysical and biochemical properties. Therefore, at present only the combination of many

methods, whose implementation may significantly complicate the pipeline, can provide valuable comprehensive estimates of the developability potential of a drug candidate [8, 15].

4.2.1 High-Throughput, Low-Material Stability Measurements

As protein folding depends only on intramolecular interactions, and it is therefore concentration independent, protein stability is perhaps the easiest biophysical property to measure in highly quantitative ways using very little material. Measurements of thermal stability only require monitoring the unfolding of a sample with increasing temperature, which can be done using small sample volumes and low protein concentrations. High-throughput methods include the protein-thermal shift assay, which can be implemented using small volumes in PCR microplates by exploiting fluorescent dyes whose brightness increases upon binding to exposed hydrophobic patches [185]. Another popular assay is differential scanning fluorimetry (DSF), which measures the change of tryptophan and tyrosine intrinsic fluorescence under a temperature gradient and determines the T_m of the antibody under scrutiny. Such methods require very low amounts of protein and are highly applicable for high-throughput screening [186].

4.2.2 Assays Probing for Nonspecific Interactions

Poor specificity, or generic stickiness, has been linked to off-target effects and consequently to poor antibody pharmacokinetics, as well as to poor solubility and high aggregation propensity [97–100]. Quantifying specificity is challenging due to the high number of possible off-targets and the wide range of affinity values. At present, specificity can be determined by assays that measure antibody interaction with selected surfaces and biomolecules, including proteins and DNA [96, 97, 107, 187, 188].

Cross-interaction chromatography (CIC [8, 187]), and stand-up monolayer adsorption chromatography (SMAC [188]) are widely used chromatographic methods to assess antibody specificity. With CIC, specificity is measured by coupling polyclonal antibodies from human serum onto a column. Increased retention times of mAb candidates are indicative of interactions with such antibodies and therefore of low specificity. Correlations of CIC readouts have been reported with solubility measurements [187], and in vivo clearance rates in mice [189]. In contrast, SMAC probes specificity by using a non-biological surface (Zenix™ column). SMAC retention times have been reported to correlate with colloidal stability, and with CIC itself [188]. Overall, these chromatographic methods typically require little material and are amenable to automation to run multiple samples. Nonetheless, as samples still run one at a time through the column, and cleaning steps are often required among samples, the throughput remains quite low. Hence, these methods are not ideal to process the large numbers of hits commonly generated during antibody library panning, which often are expressed transiently in microplates and

purified to relatively low purity [190]. With higher throughput are enzyme-linked immunosorbent assay (ELISA) formats, which are carried out in multi-well plates. The baculovirus particle (BV) ELISA detects non-specificity of protein therapeutics using a complex mixture of various biomolecules. BV particles are composed of phospholipids, carbohydrates, nucleic acids, glycoproteins and other molecules that are immobilized as antigens on an ELISA plate. It was found that antibodies, rapidly clearing in cynomolgus monkeys and humans, bind nonspecifically in this assay [97]. Another ELISA-based method has been developed that uses insulin, dsDNA, and ssDNA, and was initially employed to study self-reactivity and poly-specificity of natural antibody repertoires over the course of B-cell maturation [101]. This method has been adapted in the context of pharmacokinetic risk screening. Non-specific binding of antibodies to negatively charged substrates such as insulin or DNA resulted in faster clearance in mice [107]. Fluorescence-activated cell sorting (FACS) has also been employed to assess the specificity of antibodies [96, 191]. A soluble biotinylated membrane protein reagent termed poly-specificity reagent (PSR) is prepared from homogenized mammalian cells using a mild surfactant. Binding of PSR to antibody libraries displayed on the surface of yeast is assessed by FACS, which allows detecting low binding variants. A correlation with CIC and BV ELISA has been established, suggesting that this assay may also be correlated with antibody clearance rates [96, 191]. One study, which compared the PSR with CIC on over 400 mAbs, observed that PSR may be more sensitive to assess specificity [191]. Protein arrays and biochips represent another type of non-specificity screening tools in which different sets of proteins are spotted onto an array, which can then be probed with candidate therapeutic antibodies. Such arrays can contain many different proteins immobilized on a nitrocellulose-coated glass slide, and off-target effects are most commonly quantified with fluorescence or chemiluminescence readouts [192–195]. At variance with other methods, protein arrays not only provide information about the degree of specificity of the antibody under scrutiny, but also about which and how many other targets have been identified.

4.2.3 Assays Probing Self-Association and Aggregation Propensity

High self-association propensity of biopharmaceutics has been reported to correlate with low solubility and high viscosity [77, 85, 94, 151, 191, 196, 197], as well as with unfavorable in vivo pharmacokinetics [9, 198]. Self-interaction chromatography (SIC) measures the retention time of an antibody as it flows across a column conjugated with the same antibody. Therefore, longer retention times result from stronger interactions of the antibody with itself, and usually correlate with lower solubility [199]. The disadvantages of this technique include the need to

set up a separate column for each antibody of interest, and the relatively high amount of material required for column coupling, which make this technique relatively low throughput [200–202]. Another assay probing for self-association is clone self-interaction bio-layer interferometry [203]. Here, the Fc region of the antibody of interest is immobilized on a biosensor, and its interaction with the same antibody in solution is measured, which requires very little material [203]. The method has been reported to correlate with CIC and SIC [203], albeit the use of detergent is often needed to suppress nonspecific interactions with the biosensor itself, which may limit the sensitivity of the assay. Affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS) is another commonly used method to detect self-association [191, 196]. In this assay, gold nanoparticles are coated with anti-Fc mAbs which capture the antibody candidate of interest. Self-association of the target antibody results in reduced inter-particle distances, which can be quantified via the change in the wavelength of maximal absorbance (plasmon wavelength) [191]. AC-SINS has been reported to predict viscosity, solubility issues [100, 178, 190, 204], colloidal stability [86, 191], CIC [191] and in vivo clearance [107]. Besides low protein material consumption (tens of μg/ml), AC-SINS is in principle compatible directly with cell-culture supernatants, making it a promising tool for early stage screenings even before purification [204]. In a recent work, however, no correlation was observed between AC-SINS carried out on supernatants and more resource-intensive measurements of self-interaction (k_D) and relative solubility (PEG precipitation) carried out on the corresponding purified antibodies [205]. Other possible caveats of this approach include the fact that mAb binding to nanoparticle-conjugated antibodies may impact the observed self-association, and that the nanoparticles may be prone to aggregate by themselves under certain conditions [204]. Finally, specific physicochemical properties, such as solvent-exposed hydrophobic patches, have been linked to self-association. Hydrophobic interaction chromatography (HIC) is used to characterize the hydrophobicity of folded proteins [26, 206]. Estimates of the apparent hydrophobicity have been shown to correlate in some cases with aggregation [8, 88], viscosity and impaired in vivo clearance [9]. SGAC-SINS evaluates self-association at high ammonium sulfate concentrations using the AC-SINS method [190]. Measurements of protein self-association at high ionic strength were correlated to HIC data. The major advantage of this method is the high throughput compared to traditional HIC [190].

4.3 Discussion on Developability Screening Assays

Developability screening assays offer the opportunity to readily measure parameters that are important to determine crucial properties of antibody drug candidates using small amounts of material and often in a high-throughput way. Nonetheless, no single assay

has been found to be fully predictive of long-term storage, pharmacokinetics, and in vivo specificity [8, 207]. Therefore, many developability assays must be used in combination to evaluate the developability potential of antibody candidates. Holistic approaches [7, 8] are based on the combination of in vitro assays to assess the developability potential of antibodies in a thorough manner. However, implementing multiple assays at the early stages of an antibody discovery pipeline makes developability assessment highly resource-intensive and time-consuming.

Furthermore, there are no well-established guidelines for the efficient selection of biotherapeutic molecules with desired features from developability readouts. While the implementation of a set of thresholds beyond which an antibody candidate should be discarded has been suggested [8], varying formulation conditions can largely determine whether a molecule is above or below the threshold [207], and formulation screening with a dozen assays on many candidates remains highly unpractical.

Another factor that challenges developability assessment is the quality and amount of antibody material available at the early stages of preclinical development. The first hits are often expressed in multi-well plates in minute amounts and relatively low purity, which can induce large errors and issues of reproducibility in developability measurements [7, 8]. Likewise, some assays (e.g., CIC) rely on polyclonal antibodies, which make it difficult to reproduce experiments [208], and nearly all assays that probe nonspecific interactions and self-association require the presence of surfaces. The presence of surfaces creates interfaces that may introduce unwanted artifacts resulting from the adsorption of some molecules, and it results in limited capacity to explore different formulations, as detergents or high salt may be needed to prevent nonspecific antibody–surface interactions [209]. Some surfaces (e.g., Zenix column for SMAC assay) are also proprietary, thus making it difficult to rationalize experiments.

Finally, there is a lack of developability assays that directly and quantitatively probe for aggregation. For example, Jain et al. [8] performed accelerated temperature studies where they incubated 1 mg/ml at 40 °C for 30 days and took time points which they consequently analyzed by SEC. Their analysis showed the outcome of this experiment did not correlate with any of the other developability assays. As previously discussed, the aggregation of large, complex proteins does not necessarily show a linear temperature dependence, and thus thermally accelerated aggregation may not be an optimal way to predict aggregation propensity at storage condition. Similarly, a recent work found that no single developability readout correlates with protein aggregation during storage, and suggests the implementation of risk scores based on multiple experiments in different solution conditions [207].

4.4 Microfluidics for Developability Assessment

The simultaneous determination of multiple biophysical parameters in one assay at high throughput and low-material consumption would greatly facilitate in vitro developability assessments. Toward this goal, promising advances have been recently made using microfluidic approaches [210–212], which offer little material consumption, fast experimental run-time, and limited dilution of the sample during analysis. In addition, surface chemistries are highly tuneable in these devices, enabling for example the study of the effects of interfaces on protein stability. Therefore, advances in microfluidics may open up the exciting possibility to screen large numbers of drug candidates under a wide range of conditions at the early stages of antibody characterization using negligible amounts [210, 211]. Several successful applications of microfluidic devices for the characterization of protein sizing, separation of impurities, viscosity, and thermal stability have already been reported [210, 211]. These devices can implement both label-mediated (e.g., fluorescent dyes or tracers) or label-free (e.g., intrinsic fluorescence) strategies to detect proteins in solution.

In a microfluidic approach, for example, one can assess the extent of antibody oligomerization by sizing antibody species, including monomers and oligomer assemblies, according to their hydrodynamic radii by determining their diffusion coefficients in laminar flow. The challenge of sizing heterogeneous protein solutions has been overcome by acquiring multiple diffusion profiles in space and time [211]. A microfluidic device has indeed been developed to size antibodies even in high-concentration formulations (150 mg/ml), thus opening a complementary approach to methods such as SEC and DLS [183]. Devices used for sizing may also be exploited for determining intermolecular interactions including aggregation, phase separation and crystallization [210, 212]. In perspective, the translation of these findings to an industrial scale will require to address challenges such as automation and reproducibility of produced devices [211]. Tackling these challenges may result in powerful novel tools to assess the developability of therapeutic proteins.

5 In Silico Predictions of Developability

The development of in silico predictors to screen for the developability potential of biotherapeutics is a rapidly growing field. Computational approaches hold the promise of quickly and accurately screening thousands of sequences almost instantaneously, and, crucially, without material requirements [88, 89, 128, 213]. Moreover, some of these tools enable the identification of potential liabilities arising from specific sequence patterns within lead antibody candidates, thereby greatly facilitating the rational engineering toward improved developability [13, 15, 172].

Although the potential of computational approaches to replace resource-intensive experiments has been recognized decades ago, it is only in recent years that these methods have started to become sufficiently accurate and reliable to be competitive with experimental approaches in some practical applications [3, 88, 89, 128, 172, 213–215]. To date, in silico tools have been successfully applied to several problems, including the screening of biopharmaceuticals for their aggregation propensity [13, 89, 213] or developability potential [7, 128, 216], enhancing the solubility of proteins [13, 172, 215, 217] and investigating the determinants of solubility and aggregation in whole proteomes [63, 218, 219]. A list of available predictors of aggregation propensity, solubility, or more generally developability potential is provided in Table 2 with a short description of each method and relevant references. These methods are classified into "Amino acid composition-based algorithms," "Sliding window/pattern-based algorithms," "Consensus predictors," or "Tertiary/quaternary structure-based algorithms."

Amino acid composition-based algorithms. The methods use the frequency of individual amino acid and/or of short peptide motifs (typically di- or tri-peptides) to calculate an overall solubility or aggregation propensity score [213, 218–221, 225, 228]. Some of these algorithms have been shown to perform well in predicting the aggregation propensity of short disordered peptides and their mutational variants [251, 252]. Short peptides are highly dynamic and populate structural ensembles, where most residues are close to each other in at least some conformations. Therefore, the solubility of short peptides may be more directly dependent on global biophysical properties of the sequence and less on the presence of specific sequence patterns or structured motifs [251, 252]. However, because of its construction, this score is often insensitive to the order of the residues in the sequence, and thus residue permutations may not affect the outcome of the predictions. In addition, some of these algorithms, rather than resulting in a numerical score that may be quantitatively compared with measurements of aggregation rates or of relative solubility, provide a binary classification (e.g., soluble/insoluble), which is primarily useful to predict the likelihood of successful recombinant expression [218, 253].

Sliding window/pattern-based algorithms. These techniques typically employ a more sophisticated framework than amino acid composition ones, so that the order and location of the individual amino acids are important [172, 222, 227, 230–237, 240, 243]. These algorithms build on the realization that the presence of a single aggregation-promoting sequence pattern can readily lead to the aggregation of otherwise soluble polypeptide sequences. This observation was historically made for amyloidogenic proteins [254, 255], but has been shown to hold true also for the aggregation of biopharmaceutical antibodies [216, 256]. The vast majority of these algorithms yield a sequence profile, which is most typically one number per residue that reflects the contribution of each

Table 2
Available in silico tools to predict or design the developability of biotherapeutics. The central column provides a concise description of each method. Methods are divided into categories and sorted according to the date of first publication

Method [Ref] Year	Approach	Type of predictor
Peptide and protein solubility		
SolPro [220] 2009	A two-stage support vector machine (SVM) to predict the propensity of a protein to be soluble upon overexpression	Amino acid composition-based algorithms Machine learning
SCM [221] 2012	Solubility predictions based on a dipeptide solubility scoring matrix	Amino acid composition-based algorithms Machine learning
PROSO II [218] 2012	Protein solubility predictor using a two-layer architecture of SVM and naive Bayes classifiers	Amino acid composition-based algorithms Machine learning
CCSOL [219] 2012	SVM trained to discriminate between soluble and insoluble protein expression in *E. coli*. Training parameters include coil/disorder, hydrophobicity, hydrophilicity, β-sheet and α-helix propensities	Amino acid composition-based algorithms Machine learning
ESPRESSO [222] 2013	Binary classification of sequences using predicted secondary structural properties and sequence pattern-based methods to predict protein solubility and expression in *E. coli*	Sliding window/pattern-based algorithms
Chan-Warwicker [223] 2013	Solubility predictor using a correlation between positively charged surface patches and insoluble expression, particularly when the patch is enriched in arginine relative to lysine	Tertiary/quaternary structure-based algorithms
Obrezanova et al. [213] 2015	Prediction of intrinsic aggregation propensities of antibodies by using statistical modelling and machine learning techniques using sequence input. The tool is constructed and validated on over 500 sequences	Amino acid composition-based algorithms Machine learning
Intrinsic CamSol method [172] 2015	This algorithm uses a sliding window average of solubility propensity scores that are adjusted for physicochemical properties, gatekeepers and alternating patterns of hydrophobic and hydrophilic residues	Sliding window/pattern-based algorithms
Structurally corrected CamSol method [172] 2015	Sequence-based solubility predictions (247) projected onto a 3D structure and adjusted for solvent exposure and the influence of other residues within an 8 Å radius	Tertiary/quaternary structure-based algorithms

(continued)

Table 2
(continued)

Method [Ref] Year	Approach	Type of predictor
Schaller-Middleberg [224] 2015	Prediction of soluble protein expression using parameters obtained from high-temperature protein unfolding molecular dynamics simulations in an SVM classifier	Tertiary/quaternary structure-based algorithms
PON-Sol [225] 2016	Prediction of the effects of amino acid substitutions on solubility using a three-class (solubility increasing, decreasing, or unchanged), binary random forest classifier	Amino acid composition-based algorithms Machine learning
Solubis [12] 2016	An automated pipeline to identify mutations that reduce protein aggregation, while retaining conformational stability. Uses TANGO and FoldX	Tertiary/quaternary structure-based algorithms
SODA [226] 2017	Predicts changes in protein solubility upon mutation based on several physicochemical properties of the protein and intrinsic disorder. Uses the PASTA algorithm	Sliding window/pattern-based algorithms And also tertiary/quaternary structure-based algorithms
Protein-Sol [227] 2017	Data from *E. coli* protein solubility system used. 35 sequence-based properties are considered and two other properties (fold propensity and net segment charge) are profiled in a sliding window fashion	Sliding window/pattern-based algorithms
Aggregation propensity		
Chiti-Dobson [228] 2003	The aggregation rates of protein mutants are predicted by the weighted sum of change in hydrophobicity, secondary structure propensity, and net charge	Amino acid composition-based algorithms
Dubay et al. [229] 2004	The Chiti-Dobson approach is extended to predict the aggregation rates of whole proteins	Amino acid profile-based algorithm
TANGO [230] 2004	Sequence-based protein aggregation predictions based on physicochemical parameters involved in β-sheet structure	Sliding window/pattern-based algorithms
AGGRESCAN [231] 2007	Sequence-based evaluation of aggregation hotspots in polypeptide chains. It uses a sliding window average of aggregation propensity scores for amino acids derived from measurements of intracellular aggregation by Aβ42 mutants	Sliding window/pattern-based algorithms
Zyggregator [232] 2008	Based on DuBay et al. 2004, uses a linear combination of several physicochemical properties, which are adjusted for gatekeeping residues and alternating patterns of hydrophobic and hydrophilic residues	Sliding window/pattern-based algorithms

(continued)

Table 2 (continued)

Method [Ref] Year	Approach	Type of predictor
SAP [13, 88] 2009	Spatial aggregation propensity (SAP) identifies hotspots for aggregation based on dynamically exposed hydrophobic surface patches which are determined by structural analysis and short MD simulations	Tertiary/quaternary structure-based algorithms
FoldAmyloid [233] 2010	Predicts positions of amylogenic regions in proteins by using sliding window average of amino acid packing density and hydrogen bond probability scores	Sliding window/pattern-based algorithms
3D Profile [234] ZipperDB [235] 2006, 2010	Calculates fibrillogenic propensities of proteins using a structure-based algorithm. This algorithm assesses the compatibility of sequence segments with the conformation adopted by the NNQQNY hexapeptide in cross-β fibrils	Sliding window/pattern-based algorithms
Waltz [236] 2010	Allows distinction between amyloid sequences and amorphous β-sheet aggregates and identification of amyloid-forming regions in functional amyloids. It uses a position-specific scoring matrix derived from amyloidogenic hexapeptides combined with physicochemical properties, and structural modelling using amyloid backbone structures	Sliding window/pattern-based algorithms
AmyloidMutants [237] 2011	Combination of statistical mechanics approach and super-secondary structure prediction that quantifies the effects of sequence mutation on fibril conformation and stability	Sliding window/pattern-based algorithms
Amylpred2 [238] 2013	A consensus algorithm that combines eleven aggregation predictors (including waltz, Tango, AGGRESCAN, and AmyloidMutants) that identify aggregation-prone stretches in proteins and compare their results	Consensus predictor
MetAmyl [239] 2013	A consensus algorithm that combines eleven aggregation predictors that identify amyloid aggregation hot spots based on a logistic regression model and provides weighted predictions	Consensus predictor
PASTA [240, 241] 2006 PASTA 2.0 [242] 2014	Predicts amyloid fibril regions from protein sequences by statistical analysis of residue pairings between adjacent β-strands. The extension of PASTA includes new features, e.g., comparison of aggregation propensity of wild-type and mutated proteins and is validated on a larger dataset of globular protein domains	Sliding window/pattern-based algorithms Machine learning (only for structural prediction)

(continued)

Table 2 (continued)

Method [Ref] Year	Approach	Type of predictor
FISH Amyloid [243] 2014	Machine learning method for the identification of amyloidogenic segments in amino acid sequences using a discriminative pattern of site-specific co-occurrences of residue pairs in known amyloidogenic hexapeptides	Sliding window/pattern-based algorithms
AGGRESCAN [215] 2015	AGGRESCAN sequence-based predictions projected onto a 3D structure and adjusted for solvent exposure and the influence of other residues within a 10 A radius; optional simulation of dynamic exposure using CABS-flex	Tertiary/quaternary structure-based algorithms
Aggrescan3D 2.0 [244] 2019	Extension of AGGRESCAN3D 2015 with three main differences: (1) protein flexibility simulations using new CABS-flex standalone package. (2) protein stability calculations using the FoldX force field, to account for the impact of amino acid mutations on the overall conformational stability. (3) Option that suggests mutations with optimized solubility	Tertiary/quaternary structure-based algorithms
Viscosity		
Sharma et al., [9] 2014	Combines sequence-based and molecular dynamics simulations (Fv domain) to predict viscosity (experimental dataset of 14 mAbs), chemical degradation (22 mAbs) and fast in vivo clearance (61 mAbs). Sequence-based parameters are net charge at given pH, charge symmetry and hydrophobicity index	Tertiary/quaternary structure-based algorithms
Argawal et al., [245] 2015	Phenomenological, electrostatics-based viscosity predictor. It determines the spatial summation of residue charge of surface-exposed residues of the Fv domain of antibodies	Tertiary/quaternary structure-based algorithms
Tomar et al. [246] 2017	Predicts concentration-dependent viscosity behaviors of mAbs using sequence-structural attributes, parameters were fitted on experimental data of 16 mAbs	Tertiary/quaternary structure-based algorithms
Immunogenicity		
EpiMatrix [247] 2002	T-cell epitope predictor based on allotype-specific HLA-associated peptides (MHC ligands) derived from 133 proteins	Sliding window/pattern-based algorithms
SEPIa [248] 2017	Sequence-based B-cell epitope predictor based on 13 parameters and two different classifiers	Sliding window/pattern-based algorithms

(continued)

Table 2 (continued)

Method [Ref] Year	Approach	Type of predictor
Developability predictors that consider more than one biophysical propensity		
SAP DI [89] 2012	Uses SAP in combination with the antibody's net charge of the full structure to determine a developability index (DI) of antibody candidates. The algorithm was validated on a set of 12 mAbs, which were subjected to accelerated aggregation assays at 2 different temperatures	Tertiary/quaternary structure-based algorithms
AGGRESCAN3D + FoldX [249] 2018	Updated version of the aggregation predictor AG GRESCAN3D. Calculates the stability of protein of interest using FoldX and reports energy difference of wild-type and mutants in kcal/mol	Tertiary/quaternary structure-based algorithms
Intrinsic CamSol method [172] 2018	The sequence-based solubility predictor was shown to correlate with a set of experimental developability assays probing for nonspecificity for a library of 17 mAbs	Sliding window/pattern-based algorithms
Therapeutic Antibody Profiler (TAP) [10] 2019	Developability predictor (of primarily expression or aggregation issues) that builds homology models of variable domain sequences first and then tests them against five guidelines: CDR total length, surface hydrophobicity, positive charge and negative charge in the CDRs, and asymmetry in the net heavy- and light-chain surface charges. The predictor was validated on 242 post-phase-I antibody therapeutics	Tertiary/quaternary structure-based algorithms
Protein-Sol developability extension [250] 2019	Structure-based extension of the solubility predictor Protein-Sol that looks at hydrophobicity and charge patches on the protein surface pH and ionic strength variation can be taken into account	Tertiary/quaternary structure-based algorithms

sequence region to the biophysical property that the algorithm predicts (e.g., solubility, aggregation propensity, amyloidogenicity, etc.), and several also calculate an overall sequence score. An overall score for the whole sequence may be used to compare different variants, as it enables to readily rank sequences according to the predicted biophysical property [128, 253]. Sliding window/pattern-based methods have been reported to base their predictions for example on combinations of physicochemical properties of amino acids (e.g., hydrophobicity, charge, and secondary structure propensity), statistical analysis of residue pairings between adjacent β-strands in known structures, and/or patterns of residue distribution of amyloid hexapeptides [253].

Consensus algorithms. These methods run a number of different methods in order to compute a consensus prediction based on the outcome of each individual algorithm. The rationale behind the development of consensus methods is that, as the algorithms in Table 2 consider different types of protein aggregation (e.g., amyloid propensity, amorphous aggregation, etc.) and typically perform best on different protein or peptide classes, a combination of their outputs might minimize method-associated biases toward the overprediction of certain aggregation behavior.

Tertiary/quaternary structure-based algorithms. These algorithms are based on the knowledge of the sequence as well as its specific native three-dimensional arrangement. These predictors are often employed to detect surface-exposed patches that might impact the biophysical properties of antibodies. For example, hydrophobic patches (e.g., SAP [13, 88]), aggregation-inferring patches (e.g., AGGRESCAN3D [215]) and patches of poor solubility (e.g., CamSol structurally corrected [172]) can be identified with these predictors. The identification of these liabilities can be used to guide protein engineering efforts of lead candidates with promising biological activity but otherwise problematic biophysical properties. Compared to random mutagenesis and screening, these predictors offer a rational, safer, and usually faster route to optimize developability potential. Some of these methods implement additionally short local molecular dynamics simulations to take into account the transient exposure of otherwise buried patterns that may trigger aggregation [13, 88, 215]. However, because of challenges in conformational sampling and force field accuracy, sometimes the implementation of these simulations may not contribute significant performance improvements [89].

5.1 Machine Learning-Based Algorithms

Some of the algorithms listed in Table 2 rely on a machine learning architecture to produce a developability prediction. Machine learning, sometimes referred to as artificial intelligence, refers to a broad class of computer algorithms that can have widely different architectures, but have in common the ability to learn from data without being explicitly programmed. The learning or training phase typically consists in the optimization of a large number of free parameters, such as connection weights in the case of neural network architectures. These parameters are optimized in such a way that from known inputs (e.g., protein sequences) the algorithm reproduces the corresponding known outputs (e.g., the measured developability parameter). When trained on a large enough amount of diverse data, these algorithms have the ability to generalize that is to correctly predict the outputs of never-seen-before inputs. The growing amount of available data, including measurements of biophysical properties, is increasingly enabling applications of machine learning algorithms to address biological problems. Although

machine learning-based methods lack the physical transparency of other approaches rooted on physicochemical principles or physical laws, their frank pragmatism can be remarkably successful.

However, as the number of free parameters often ranges in the thousands, these algorithms require a large number of experimental data points for a reliable training. As discussed above, properties like solubility and aggregation propensity remain very challenging to measure accurately, and measurements are typically not comparable across different experimental setups or conditions. This fact poses severe limitations to the current applicability of machine learning algorithms for the prediction of solubility or aggregation propensity, as it is challenging to compile large enough experimental datasets that can be correctly compared with each other. Most typically, algorithms are trained and tested on one or few datasets of experimental measurements, employing validation procedures such as ten-fold-cross-validation, which ensure that a subset of sequences diverse enough from those used for parameter optimization (i.e., the training) is kept aside to validate the performance. A common issue in machine learning is overfitting, whereby the trained algorithm, rather than capturing the biophysical properties underlying the data, maps mathematically inputs and outputs with limited ability of generalizing. Cross-validation approaches are common ways to avoid overfitting, as overfitting characteristically results in very high performance on the data used for training, but lower performance on those set aside for testing. Nonetheless, the measurements employed in training and testing are often of the same kind and may be affected by the same errors, which sometimes are systematic and dependent on the employed experimental assay and condition. Besides, some parameters that define the architecture of the machine learning algorithm (e.g., number of hidden neurons, learning rate, etc.) are in practice optimized on the sequences in the validation dataset. Taken together, all of these facts, and first and foremost the current unavailability of large numbers of high quality, quantitative measurements of solubility or aggregation taken under identical conditions for diverse antibodies and proteins, limit the successful application of machine learning approaches for developability prediction. This situation may change in the near future as the throughput and accuracy of available experimental assays increases, which may ultimately enable to compile large datasets of highly accurate measurements for diverse proteins taken at the same experimental conditions.

5.2 Structure-Based and Sequence-Based Computational Methods

In silico predictors of developability potential can be divided into sequence-based and structure-based predictors. Sequence-based predictors typically only require the amino acid sequence as input, and in some cases formulation-related information such as the pH value. Conversely, structure-based predictors, which are termed

tertiary/quaternary structure-based algorithms in Table 2, also require the protein structure as input whether experimentally determined or computationally modeled.

Usually, sequence-based predictors are faster and more generally applicable than structure-based ones, as the amino acid sequences of antibodies are typically readily available by sequencing hits from screening campaigns. These algorithms are thus most useful for ranking many molecules according to biophysical properties of interest. Conversely, structure-based methods can be used only when either the structure or a good-quality model are available. Therefore, these methods are typically more amenable to small-scale rankings or may be used at later stages of the antibody preclinical development to facilitate lead engineering. Indeed, most of these methods enable the identification of potential biophysical liabilities within antibody molecules, and may provide insights for their molecular origins [253]. Some of these tools can also suggest mutation sites, or even specific amino acid substitutions to carry out in order to improve solubility or stability [172, 249].

Sequence-based predictors either neglect, or only consider to a minimal extent [242], the effect of the folded structure on the developability parameter they are predicting. For example, it is common for stable globular proteins, such as antibodies, to possess relatively large aggregation-promoting, poorly soluble regions in their sequence. These regions drive the folding process and are usually buried in the native state but may elicit aggregation when exposed on the protein surface. Despite this potential limitation, sequence-based methods have been shown to perform really well when predicting the developability potential of antibodies, but only in cases where all antibodies under scrutiny were known to be well folded and stable [7, 128].

Mutations can cause aggregation through two main pathways. In the first, the mutations destabilize the native state, so that poorly soluble regions usually buried in the hydrophobic core become transiently exposed to the solvent thus triggering aggregation. In the second case, mutations may occur on the antibody surface or in regions where they do not substantially alter the conformational fluctuations of the native structure. The effects on solubility or aggregation propensity of mutations in this second class can readily be predicted with sequence-based methods, which explains the high performance observed in predicting antibody developability potential [7, 128]. However, the effect of mutations in the first class cannot readily be predicted by looking only at the change in predicted solubility or aggregation propensity, as the aggregation is triggered by the fact that the solubility of non-native states is typically much lower than that of the native state, disregarding the impact on intrinsic solubility of the specific mutations under scrutiny. Nevertheless, hits from antibody discovery campaigns are

most often well folded as otherwise they would not be functional and would not be selected during library panning. Therefore, sequence-based predictors may be implemented in antibody discovery pipeline just after the sequencing of the hits in a relatively safe way, as all mutations between the different hits are unlikely to have caused drastic destabilization and unfolding.

5.3 Structure-Based Predictions Using Structural Models

An important aspect to consider when using modeled structures as input for structure-based in silico tools is how sensitive these tools are to errors in the atomic coordinates. Some structure-based methods may perform well as long as residue solvent exposure and relative distances are correctly modeled, while others may be highly sensitive to the atomic coordinates of sidechains and their relative orientation. For example, Fig. 3 shows a comparison between the predictions of two unrelated structure-based methods run on two slightly different input models of the same antibody. The CamSol structurally corrected method is only sensitive to residue solvent exposure and their relative distance, and in fact its prediction does not change substantially depending on which model is chosen as input (Fig. 3b, e). Conversely, the FoldX prediction of stability change upon mutation [258] can vary depending on which model is selected as input (Fig. 3c, f). FoldX has ranked among the best predictors of stability change upon mutation when assessed on experimentally determined crystal structures [253]. Nonetheless, its calculations are exquisitely sensitive to the atomistic distances between sidechains, and therefore even small differences in the input atomic coordinates can lead to large differences in the outcome of the prediction. Some of the in silico tools listed in Table 2 for the design of mutations that decrease aggregation propensity, such as Aggrescan3D 2.0 or Solubis, implement FoldX calculations to account for the impact of the suggested substitution on conformational stability.

Owing to the high degree of homology between variable domains from different antibodies, and to the fact that most CDR loops can be clustered into canonical conformations [259–261], the vast majority of the Fv region can now be modeled with high accuracy [262]. The most challenging parts to model remain the heavy-chain CDR3 loop (CDR-H3), and the relative orientation between the VH and VL domains, but significant progress has recently been made also for these regions [263, 264]. However, as the example in Fig. 3 shows, care must be taken when running structure-based predictors of developability potential on modeled antibody structures, and the robustness of the prediction should ideally be assessed by building multiple models with different templates or modelling techniques.

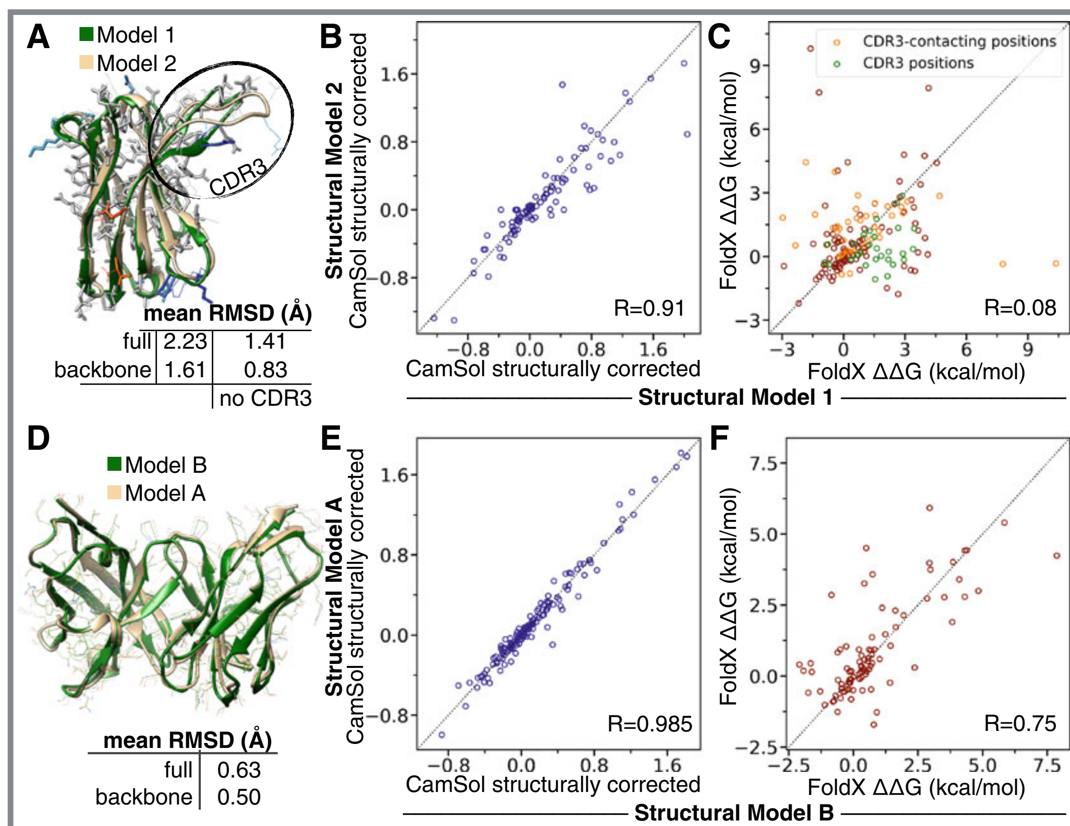

Fig. 3 Analysis of structure-based in silico methods using structural models. (**A**) Two models of the same single-domain antibody (VHH) were constructed using the online web server ABodyBuilder [257] (Model 1, green), and the homology-model software Modeller (version 9.16 on template PDB 3b9v; Model 2, tan). The table (inset) reports the root mean square deviation (RMSD) between the two models calculated using all atom coordinates (full) or only mainchain backbone atoms (backbone), also by excluding the CDR3 (second column), which is the region showing the highest discrepancy between the two models (circled in the structure). (**B**, **C**) Scatter plots of the results of structure-based predictors ran on the two models (on *x*- and *y*- axis respectively); the dotted line is the unity line ($y = x$). (**B**) The CamSol structurally corrected prediction of solubility for all residues in the structure. (**c**) FoldX predictions of stability change upon mutation for 167 mutations carried out on the two models at sites of poor solubility as identified from the CamSol prediction, or of low conservation as identified from a multiple-sequence alignment of similar sequences. The command *FoldX Optimize* was performed on each input model before running the prediction of stability change, and each point represents the average of three runs. Mutations carried out at CDR3 positions are in green, at sites in contact with the CDR3 in orange, and at other sites in red. The Pearson's coefficient of correlation (*R*) is also reported for each plot. (**D**, **E**, **F**) Like **A**, **B**, **C** respectively, but this time the models are of the Fv region of the same antibody, comprising both heavy and light chain. Here, both models are constructed with ABodyBuilder [257]. One is obtained using as template for the framework region and for the heavy–light chain orientation the PDB structure 5veb (framework sequence identity 99%), while the other also uses 5veb for the heavy–light chain orientation, but it uses 5hi4 and 3eo9 as templates for heavy and light chain respectively (framework sequence identity 100%)

5.4 Empirical Algorithms for the Identification of Developability Liabilities

The field of small-molecule drug discovery can rely on Lipinski's rule of five and its more recent adaptations to guide the selection of molecules with biophysical properties suitable for drug development [265]. In the quest for an analogue for antibody therapeutics, a recent study systematically compared the physicochemical properties of clinical-stage antibodies (post-phase I clinical trial) with those from subsets of the human antibody repertoire, and proposed a set of rules to profile therapeutic candidates based on five computational developability guidelines, which are encoded in the Therapeutic Antibody Profiler [10]. These guidelines include the parameters surface hydrophobicity, charge of CDRs, length of CDRs and symmetry of net charge of the heavy and light chains. A new antibody candidate is considered acceptable if its properties fall in the range of those observed for clinical-stage antibodies [10].

These empirical statistical approaches have great potential, as their accuracy is likely to increase as the number of clinical-stage antibodies grows and as new guidelines are added to the list. However, depending on formulation, administration route, frequency, and dosage, each new antibody will have different requirements. Furthermore, while these approaches can readily flag antibodies likely to contain liabilities, they do not provide a quantitative ranking for the selection of the best candidates among those available, and advances in process development and formulation may soon redefine the limits of permissible values. Other algorithms in this group include those that can identify potentially immunogenic motifs (Table 2), as well as those that analyze the primary sequence or the three-dimensional structure to identify potential chemical liabilities (e.g., deamidation sites, isomerization sites, etc.) [15, 45].

5.5 The CamSol Method of Predicting Solubility and Developability Potential

The pioneering work by Chiti and colleagues [251] showed that a linear combination of the biophysical properties of the individual amino acids comprising peptide sequences could accurately predict changes in aggregation rates caused by mutations. This approach was then extended to predict the absolute aggregation rates of the whole sequences of disordered proteins [229], and to enable the identification of aggregation-prone regions within protein or peptides sequences, with a particular focus of predicting amyloidogenic regions within disease-related proteins [232, 266, 267].

By building on these advances, in 2015 we introduced the CamSol method for the prediction of solubility changes upon mutation in proteins and antibodies, and for the rational design of solubility-improving mutations [172]. Shortly after, we generalized the method [268] to enable accurate solubility predictions of more distantly related proteins, including multi-chain complexes such as monoclonal antibodies differing by up to 32 mutations [128] and proteome-wide applications [63].

To obtain accurate predictions, CamSol first calculates an intrinsic solubility profile from the physicochemical properties of the amino acids in the input sequence. This profile consists of a score for each residue, which depends on both the identity of the residue itself and the local environment in which this residue is found (i.e., the neighboring residues) [172]. Then a solubility score for the whole sequence is calculated from this profile, accounting for the contributions of both poorly and highly soluble regions [128]. Therefore, this solubility score is sensitive to the overall amino acid composition of the input sequence as well as to the presence of specific patterns of residues, which typically correspond to regions of low or high solubility in the profile.

At the core of the CamSol calculation, there is a phenomenological combination of amino acid physicochemical properties to add up in a linear manner including hydrophobicity, charge, α-helix and β-sheet propensity, with additional corrections for the presence of patterns of alternating hydrophilic and hydrophobic residues and for the gatekeeping effect contributed by nearby charges of the same sign [172]. The coefficients of this linear combination, as well as some free parameters in the equation used to calculate the solubility score from the intrinsic profile [128] were fitted on experimental data. At variance with machine learning methods that have hundreds, and sometimes thousands, of free parameters to be optimized (e.g., the connection weights in neural network), CamSol only has six parameters in the linear combination [172], and six others in the solubility score calculation [128]. Moreover, many of these parameters, such as the thresholds that define poorly or highly soluble regions in the intrinsic profile, can only sensibly vary in a narrow range. Therefore, relatively few quantitative experimental data are enough to determine these coefficients, which is particularly advantageous as accurate and quantitative solubility measurements are scarce in the literature. Parameters were fitted as described in Refs. [128, 172], using measurements of aggregation rates from the literature [229], and a dataset of non-aggregating and aggregating peptides and proteins [269], which contains totally unrelated sequences rather than mutational variants of the same protein. After the fitting of these parameters, the performance of the method was initially tested with a qualitative validation on solubility measurements from the literature [172] (55/56 correctly predicted solubility changes when using CamSol version 2 [128]), and in many other quantitative benchmarks [7, 128, 172, 217].

When a structure is available, the CamSol method can also be used to identify surface-exposed aggregation hotspots and to predict solubility-improving mutations. This result is accomplished by computing a structurally corrected solubility profile, which is calculated like the intrinsic profile but also accounts for the solvent exposure and the relative distance of residues in the 3D space

[172]. This profile can be color-coded on the surface of the protein and it can be used to identify solvent-exposed regions of low solubility that may act as aggregation hotspots. In the example in Fig. 4d, the amino acids responsible for the increased self-association of mAb2 with respect to mAb1 were experimentally identified with a structural proteomics approach based on hydrogen-deuterium exchange [256] as W30, F31, L561, in quantitative agreement with the CamSol prediction [128].

Mutations may then be carried out in silico to replace residues within the identified aggregation hotspots. The resulting protein variants can readily be ranked according to their contributions to solubility by using the sequence-based calculation, so that modelling the structure of each mutant is not required. As the CamSol intrinsic calculations can process about 200 sequences per second on a single computing core of a standard laptop, this procedure can be applied to sample very large mutational spaces, including combinations of mutations at multiple identified sites [7]. We applied this pipeline to identify aggregation hotspots and predict solubilizing amino acid substitutions and insertions in the case of a problematic gammabody [172] (Fig. 4a), a monoclonal antibody [7], and of disease-related human proteins [127, 217].

The ability of CamSol to rank the solubility of biopharmaceutical antibodies was assessed on a phage-display-derived library from MedImmune [128]. The mAbs analyzed differed by up to 32 mutations in the Fv region, and a strong correlation was observed between calculated scores and corresponding solubility measurements (Fig. 3b). Similarly, a statistically significant correlation between CamSol predictions and solubility measurements was also reported for mutational variants of a troublesome mAb [205].

Recent advances in deep mutational scanning are beginning to enable the experimental investigation of the effect of mutations in living cells at an unprecedented scale [271]. A combination of a yeast dihydrofolate reductase (DHFR) aggregation assay with deep mutational scanning was recently used to measure the cytosolic relative solubility of 791/798 single mutations at all residue positions of the Amyloid-β42 peptide (Aβ42), and of 99 double mutants [272]. As Aβ42 is an intrinsically unstructured peptide, it is devoid of a native state. Therefore, mutations impact solubility and colloidal stability without the additional complication of the trade-off with conformational stability [63], thus making this dataset particularly suitable to assess the performance of the sequence-based CamSol prediction. Besides, this dataset exhaustively covers the whole mutational space, as it contains 99.1% of all possible point mutations across the sequence, therefore representing a strongly unbiased benchmark opportunity. The average Pearson's coefficient of correlation between the outcomes of two replicates of this experiment under identical conditions is reported to be 0.83 [273], which may be regarded as an upper limit for the

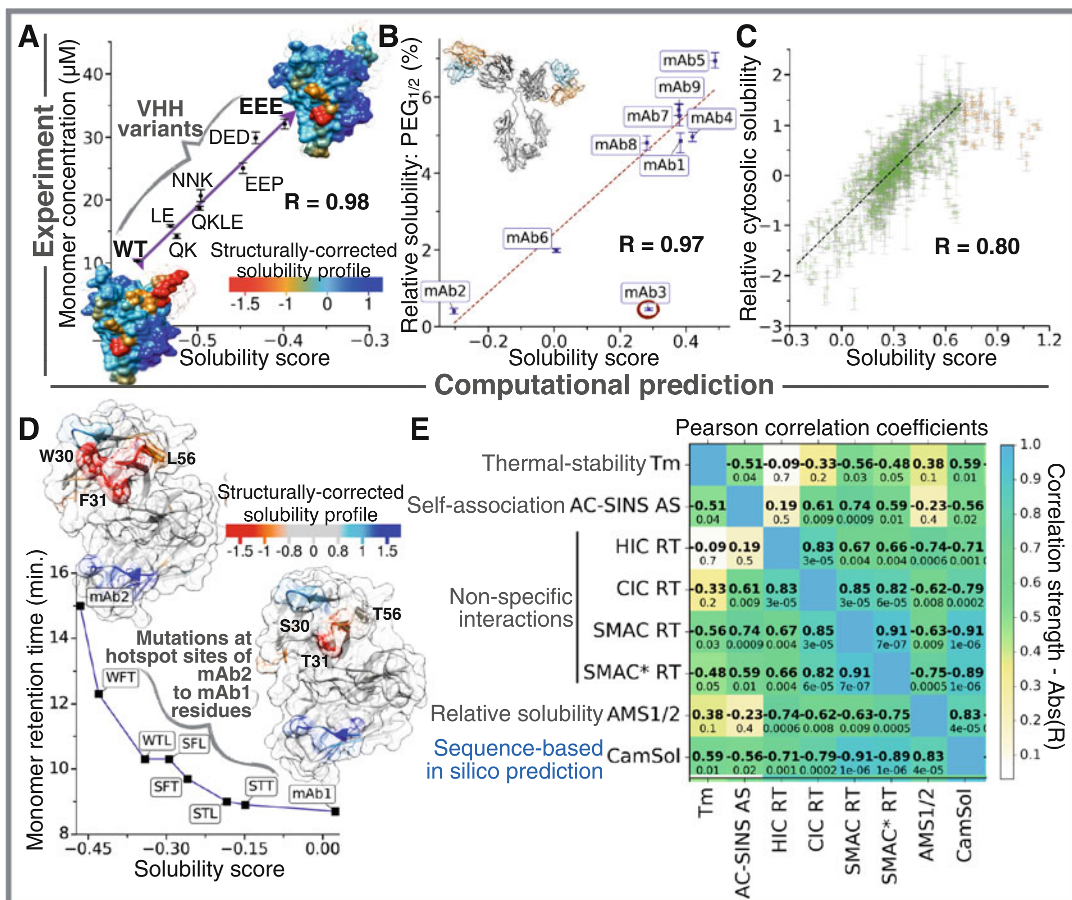

Fig. 4 Assessment of the performance of the CamSol method of predicting protein solubility. (**A**–**C**) Scatter plots of different experimental readouts of solubility of different molecules (*y*-axis) as a function of their CamSol solubility score calculated from the sequence alone (*x*-axis). (**A**) Monomer concentration determined with SEC after incubation of rationally designed mutational variants of single-domain antibody (WT). The structurally corrected CamSol profile is color-coded (see color-bar) on the surface of homology models of the WT and of the most soluble variant (the triple-mutant EEE, adapted from Ref. 172). (**B**) Apparent solubility expressed as the value of $PEG_{1/2}$ of a mAb library [128, 172] obtained with phage display. Regression line and reported Pearson's coefficients of correlation (*R*) are calculated by excluding the outlier point circled in red (mAb3) (adapted from Ref[201]). (**C**) Yeast cytosol solubility assessed with a DHFR assay of 791 different single mutants and 99 double mutants of the Aβ42 peptide [270]. The points in orange in the plateau are excluded from the calculation of the regression line and reported correlation coefficient (see text). (**D**) The structurally corrected solubility profile is color-coded (see color-bar) on the surface of homology models of the Fv region of mAb2 (top left) and mAb1 (lower right). Aggregation-promoting regions are in orange/red, whereas aggregation-protecting regions are in light blue/blue. The labeled residue positions on mAb2 (W30, F31, L57) are those that have been experimentally identified as aggregation hotspots [256]. The plot shows the measured high performance SEC monomer retention time [256] for various mAb variants as a function of their solubility score. mAb2 has the residue types W, F, and L at the hotspot positions 30, 31, and 57, respectively, while mAb1 has S, T, and T. The six variants between mAb2 and mAb1 are named according to which mAb2 positions have been mutated to the corresponding mAb1 amino acid (e.g., WFT is mAb2 L57T). The line serves as a visual guide (adapted from Ref. 128). (**E**) Matrix of Pearson's coefficients of correlation calculated comparing all combinations of in vitro developability measurements and CamSol sequence-based predictions for a library of 17 mAbs. The corresponding *p*-values are reported below the coefficients, and the matrix is color-coded according to the correlation strength (see color-bar, adapted from Ref. 7). Experimental data used in this figure were gathered at the University of Cambridge in our laboratory (**A**), Medimmune UK (**B**), Medimmune UK and University of Leeds (**D**), the University of Washington (**C**), and Novo Nordisk Denmark (**E**)

predictability of these data since a higher correlation would suggest one is also predicting the experimental noise. The correlation with the sequence-based CamSol prediction is $R = 0.77$. The scatter plot in Fig. 4c further reveals a plateau at high experimental solubility values, suggesting that the experiment may not be able to distinguish the solubility of those variants, as they are soluble enough to avoid aggregation altogether at the concentration at which they are expressed in yeast cells. Conversely, the computational prediction is not limited by experimental constraints and hence possesses a broader dynamic range. Removing the points in the plateau further increases the correlation between CamSol and the experiment to $R = 0.8$, which indicates that the errors in the CamSol predictions are comparable with those between experimental replicates (Fig. 4c). This is very encouraging in terms of the long-term strategy discussed in this review article of replacing experimental measurements with computational predictions when possible.

Finally, in recent study on a library of 17 mAbs we quantitatively compared commonly used in vitro developability assays with each other, with one resource-intensive solubility measurement of ammonium sulfate precipitation, and with the sequence-based CamSol predictions as well as other in silico predictors [7]. The results showed that the in silico predicted solubility scores correlate closely with the experimental relative solubility from the ammonium sulfate precipitation assay, and also with many developability assays with the expected exception of stability readouts such as the melting temperature. In particular, the strongest correlations were observed with assays that measure nonspecific interactions. Overall, the resulting correlations observed between CamSol and these experimental readouts were at par, and in some instances better, than those among the different in vitro assays with each other [7] (Fig. 4e). This study demonstrates how the selection and design of lead candidates with a high developability potential can be facilitated by rapid computational predictions that only require the amino acid sequence as an input.

In summary, CamSol predictions rely on a combination of physicochemical properties of amino acids. These include charge, hydrophobicity, and propensity to form secondary structure elements, which are first considered at the individual residue level, then averaged locally across sequence regions, and finally considered globally to yield a solubility score. In particular, while the calculation of the structurally corrected profile is necessary to identify suitable mutation or insertion sites, the solubility prediction is performed using only the amino acid sequence. This aspect not only makes in silico solubility screening significantly faster, but it means that the method is readily applicable to the screening of antibody libraries without the need of structural modelling, and thus it is fully independent from model accuracy.

The results that we have obtained over the years using CamSol show strong quantitative correlations with experimental measurements of relative solubility or of developability potential (Fig. 4). Notably, these measurements were carried out with different experimental techniques, on widely different molecules ranging from monoclonal antibodies to unstructured peptides, and in different laboratories. Taken together, these strong correlations suggest that CamSol predictions can greatly facilitate the screening of developability potential and more specifically of solubility. In particular, at the initial stages of antibody discovery campaigns, when numbers of candidates can be very high while yield and purity are often low, such predictions may entirely replace experiments.

6 Conclusions: Incorporation of Silico Tools in Antibody Discovery Pipelines

Therapeutic proteins and antibodies must not only bind to their targets and elicit the required biological effects, but must also be free from developability issues, which include poor stability or solubility, high levels of aggregation, various chemical liabilities, poor specificity, or immunogenicity. Therefore, not every antibody with the desired biological activity can be developed into a drug, as, unlike natural antibodies, therapeutic antibodies must endure many stresses during production, purification, shipping, high-concentration storage, and administration. The stringent requirements of therapeutic applications imply that several biophysical properties of biotherapeutics must be optimized beyond typical natural levels, a task that poses substantial challenges in the processes of candidate selection and optimization.

In this review, we have first discussed the key biophysical properties that underpin antibody developability and their interplay. We have then described the wide range of in vitro assays that are now available to measure biophysical properties or parameters that are predictive of developability potential, and the challenges associated with rigorous developability assessments, especially at the early stages of preclinical development. Next, we have given an overview of the growing number of computational methods available to predict properties that are relevant for antibody development, and we have discussed in more depth the CamSol method that we have developed.

At present, no single in vitro developability readout seems to be fully predictive of key properties for antibody development, such as stability and aggregation during long-term storage [8, 207]. The implementation of multiple experimental assays in the antibody discovery phase is challenged by the high number of candidates to screen, and by the fact that these are typically produced in minute amounts and low purity, which may affect quality and reliability of experimental measurements. Conversely, in silico predictors offer

an inexpensive alternative that can readily be implemented for as many antibodies as needed, provided that the sequence is known.

After decades of research, some in silico predictors are now able to accurately recapitulate biophysical properties of interest in many cases, a feature that makes them highly competitive with experiments. As we discussed in this review article, two types of in silico methods may be particularly suitable for implementation in antibody discovery pipelines. Those that identify liabilities, and therefore can flag potentially problematic candidates, and those that can quickly rank the candidates according to one or more biophysical property of interest. Liabilities include immunogenic regions [270, 273, 274], sites of chemical instability (e.g., deamidation, etc.) [3, 9, 44, 45], as well as other potentially problematic motifs identified using empirical algorithms [10]. Antibody candidates predicted to contain one or more of such liabilities may readily be excluded from further progression along the pipeline (Fig. 5).

The vast majority of available approaches to discover novel antibodies for a target of interest rely on screening procedures, which typically yield a readout proportional to the strength of antigen binding (e.g., enrichment ratios from deep sequencing, ELISA absorbance values, etc.) [116, 275]. This knowledge can readily be combined with in silico predictions of relevant biophysical properties, such as aggregation propensity or solubility, so that only those candidates embodying the best compromise between binding strength and predicted developability potential may be taken forward. The implementation of this computer-aided screening can readily lead to the selection of a shortlist of lead candidates. This shortlist will contain a manageable number of antibodies that can be produced to high purity and subjected to further functional characterization as well as comprehensive in vitro developability screening to select the final leads to be taken forward for preclinical and clinical studies (Fig. 5).

In perspective, notwithstanding the promising advances in the computational predictions of developability potential, there is still a need for improved algorithms that can reliably calculate the impact of formulation and storage conditions. Indeed, there is a lack of effective computational methods that assess the compatibility of candidate protein drugs with different buffers, excipients and process conditions [135]. Most in silico predictors compare best with experimental data obtained at near-physiological conditions, but not all biopharmaceuticals are formulated at neutral pH or at salt concentrations that are isotonic with blood plasma [43]. Despite this limitation, the ranking of different antibody candidates should be relatively robust to changes in formulation conditions, at least within a pharmaceutically relevant range. Most exceptions to this rule typically result from molecules getting too close to, or crossing their isoelectric point (pI) [205], which can also be calculated quite accurately in silico thus allowing to flag those molecules for which

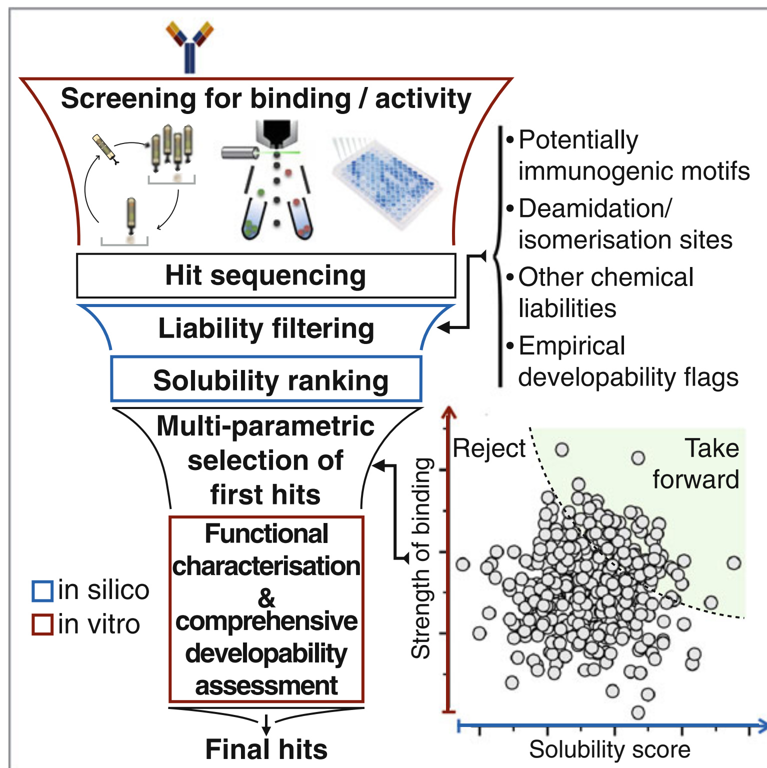

Fig. 5 Incorporation of in silico prediction methods into antibody discovery pipelines. Schematic of a generic funneled antibody discovery pipeline with added steps of computational analysis. The width of the funnel at different positions is representative of the number of candidates going through the pipeline. The initial library (top) may be obtained from animal immunization, DNA laboratory construction, or other ways. The library is then screened in vitro (steps with red edge) for antigen binding or for biological activity and, once the first hits are sequenced, in silico (steps with blue edge) to aid the selection of the final hits (bottom). In silico steps are used to exclude molecules with potential liabilities, such as those listed in the bullet points, and to rank the remaining molecules according to one or more biophysical property. The example in the figure is for the prediction of solubility. The scatter plot is an example of multiparametric lead selection, whereby those candidates taken forward are those embodying the best compromise between in vitro readouts (e.g., antigen binding strength in the *y*-axis of the plot) and in silico prediction (e.g., solubility on the *x*-axis). The incorporation of in silico tools enables the selection of a shortlist containing a manageable number of hits, which can thus be produced to high purity for further functional characterization and comprehensive in vitro developability assessment, so that only the very best hits are then taken forward for preclinical and clinical studies

the developability prediction is less likely to be accurate at the formulation pH [276]. More importantly, however, computational methods of predicting developability potential currently offer little insights for formulation optimization, which is another key aspect for the successful development of biotherapeutics. As the throughput of experimental method improves, and more data become available for different molecules in different formulation conditions, we anticipate that the next generation of in silico tools will be able to accurately tackle this issue.

In summary, the use of in silico tools from the early stages of antibody discovery can provide considerable advantages in terms of high-throughput, short analysis time, and low costs. Since they have no material requirements, these methods can easily be implemented at any point of the biopharmaceutical pipeline, as long as at least the sequence of the candidates is known. Additionally, some of these tools may help to build a preliminary understanding of certain properties of lead candidates that would otherwise be available only much later in the development phase [135, 253]. The increasing performance of computational predictors of developability potential observed in the last few years indicates that these tools can reduce or eliminate the need of carrying out laborious experiments for large numbers of antibody candidates. We believe that it is now particularly timely to incorporate in silico tools as integrated components of antibody discovery pipelines, as they can significantly facilitate the rapid and early selection of antibodies with optimal developability potential. We anticipate that, in the future, computational methods will play an increasingly important role in the discovery and optimization of therapeutic antibodies [6].

References

1. Kaplon H, Muralidharan M, Schneider Z, Reichert JM (2020) Antibodies to watch in 2020. mAbs 12:1703531
2. Hernandez I, Bott SW, Patel AS, Wolf CG, Hospodar AR, Sampathkumar S, Shrank WH (2018) Pricing of monoclonal antibody therapies: higher if used for cancer? Am J Manag Care 24:109–112
3. Jarasch A, Koll H, Regula JT, Bader M, Papadimitriou A, Kettenberger H (2015) Developability assessment during the selection of novel therapeutic antibodies. J Pharm Sci 104:1885–1898
4. Igawa T, Ishii S, Tachibana T, Maeda A, Higuchi Y, Shimaoka S, Moriyama C, Watanabe T, Takubo R, Doi Y, Wakabayashi T, Hayasaka A, Kadono S, Miyazaki T, Haraya K, Sekimori Y, Kojima T, Nabuchi Y, Aso Y, Kawabe Y, Hattori K (2010) Antibody recycling by engineered pH-dependent antigen binding improves the duration of antigen neutralization. Nat Biotechnol 28:1203–1207
5. Rabia LA, Desai AA, Jhajj HS, Tessier PM (2018) Understanding and overcoming trade-offs between antibody affinity, specificity, stability and solubility. Biochem Eng J 137:365–374
6. Sormanni P, Aprile FA, Vendruscolo M (2018) Third generation antibody discovery methods: in silico rational design. Chem Soc Rev 47:9137–9157
7. Wolf Pérez A-M, Sormanni P, Andersen JS, Sakhnini LI, Rodriguez-Leon I, Bjelke JR, Gajhede AJ, De Maria L, Otzen DE, Vendruscolo M, Lorenzen N (2018) In vitro and in silico assessment of the developability of a designed monoclonal antibody library. mAbs 11(2):388–400

8. Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, Sharkey B, Bobrowicz B, Caffry I, Yu Y, Cao Y, Lynaugh H, Brown M, Baruah H, Gray LT, Krauland EM, Xu Y, Vásquez M, Wittrup KD (2017) Biophysical properties of the clinical-stage antibody landscape. Proc Natl Acad Sci U S A 114:944–949
9. Sharma VK, Patapoff TW, Kabakoff B, Pai S, Hilario E, Zhang B, Li C, Borisov O, Kelley RF, Chorny I, Zhou JZ, Dill KA, Swartz TE (2014) In silico selection of therapeutic antibodies for development: viscosity, clearance, and chemical stability. Proc Natl Acad Sci U S A 111:18601–18606
10. Raybould MIJ, Marks C, Krawczyk K, Taddese B, Nowak J, Lewis AP, Bujotzek A, Shi J, Deane CM (2019) Five computational developability guidelines for therapeutic antibody profiling. Proc Natl Acad Sci U S A 116:4025–4030
11. Pallarès I, Ventura S (2016) Understanding and predicting protein misfolding and aggregation: insights from proteomics. Proteomics 16:2570–2581
12. Van Durme J, De Baets G, Van Der Kant R, Ramakers M, Ganesan A, Wilkinson H, Gallardo R, Rousseau F, Schymkowitz J (2016) Solubis: a webserver to reduce protein aggregation through mutation. Protein Eng Des Sel 29(8):285–289
13. Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL (2009) Design of therapeutic proteins with enhanced stability. Proc Natl Acad Sci U S A 106:11937–11942
14. Kuroda D, Shirai H, Jacobson MP, Nakamura H (2012) Computer-aided antibody design. Protein Eng Des Sel 25:507–521
15. Seeliger D, Schulz P, Litzenburger T, Spitz J, Hoerer S, Blech M, Enenkel B, Studts JM, Garidel P, Karow AR (2015) Boosting antibody developability through rational sequence optimization. MAbs 7(3):505–515
16. Tomar DS, Singh SK, Li L, Broulidakis MP, Kumar S (2018) In silico prediction of diffusion interaction parameter (kD), a key indicator of antibody solution behaviors. Pharm Res 35:193
17. Norman RA, Ambrosetti F, Bonvin AMJJ, Colwell LJ, Kelm S, Kumar S, Krawczyk K (2019) Computational approaches to therapeutic antibody design: established methods and emerging trends. Brief Bioinform 208:1673–1619
18. Shire SJ, Shahrokh Z, Liu J (2004) Challenges in the development of high protein concentration formulations. J Pharm Sci 93:1390–1402
19. Roberts CJ (2014) Protein aggregation and its impact on product quality. Curr Opin Biotechnol 30:211–217
20. Vázquez-Rey M, Lang DA (2011) Aggregates in monoclonal antibody manufacturing processes. Biotechnol Bioeng 108:1494–1508
21. Bee JS, Randolph TW, Carpenter JF, Bishop SM, Dimitrova MN (2011) Effects of surfaces and Leachables on the stability of biopharmaceuticals. J Pharm Sci 100:4158–4170
22. Rajan RS, Illing ME, Bence NF, Kopito RR (2001) Specificity in intracellular protein aggregation and inclusion body formation. Proc Natl Acad Sci U S A 98:13060–13065
23. Hober S, Nord K, Linhult M (2007) Protein a chromatography for antibody purification. J Chromatogr B Analyt Technol Biomed Life Sci 848:40–47
24. Roberts PL (2014) Virus elimination during the purification of monoclonal antibodies by column chromatography and additional steps. Biotechnol Prog 30:1341–1347
25. Yamniuk AP, Ditto N, Patel M, Dai J, Sejwal P, Stetsko P, Doyle ML (2013) Application of a kosmotrope-based solubility assay to multiple protein therapeutic classes indicates broad use as a high-throughput screen for protein therapeutic aggregation propensity. J Pharm Sci 102:2424–2439
26. Haverick M, Mengisen S, Shameem M, Ambrogelly A (2014) Separation of mAbs molecular variants by analytical hydrophobic interaction chromatography HPLC: overview and applications. mAbs 6:852–858
27. Pinholt C, Hartvig RA, Medlicott NJ, Jorgensen L (2011) The importance of interfaces in protein drug delivery—why is protein adsorption of interest in pharmaceutical formulations? Expert Opin Drug Deliv 8:949–964
28. Clarkson JR, Cui ZF, Darton RC (2000) Effect of solution conditions on protein damage in foam. Biochem Eng J 4:107–114
29. Smith C, Li Z, Holman R, Pan F, Campbell RA, Campana M, Li P, Webster JRP, Bishop S, Narwal R, Uddin S, van der Walle CF, Lu JR (2017) Antibody adsorption on the surface of water studied by neutron reflection. mAbs 9:466–475
30. Mahler H-C, Müller R, Friess W, Delille A, Matheus S (2005) Induction and analysis of aggregates in a liquid IgG1-antibody formulation. Eur J Pharm Biopharm 59:407–417
31. Liu L, Qi W, Schwartz DK, Randolph TW, Carpenter JF (2013) The effects of excipients on protein aggregation during agitation: an interfacial shear rheology study. J Pharm Sci 102:2460–2470

32. Jones LS, Kaufmann A, Middaugh CR (2005) Silicone oil induced aggregation of proteins. J Pharm Sci 94:918–927
33. Krayukhina E, Tsumoto K, Uchiyama S, Fukui K (2015) Effects of syringe material and silicone oil lubrication on the stability of pharmaceutical proteins. J Pharm Sci 104:527–535
34. Bee JS, Davis M, Freund E, Carpenter JF, Randolph TW (2010) Aggregation of a monoclonal antibody induced by adsorption to stainless steel. Biotechnol Bioeng 105:121–129
35. Hoehne M, Samuel F, Dong A, Wurth C, Mahler H-C, Carpenter JF, Randolph TW (2011) Adsorption of monoclonal antibodies to glass microparticles. J Pharm Sci 100:123–132
36. Frokjaer S, Otzen DE (2005) Protein drug stability: a formulation challenge. Nat Rev Drug Discov 4:298–306
37. Daugherty AL, Mrsny RJ (2006) Formulation and delivery issues for monoclonal antibody therapeutics. Adv Drug Deliv Rev 58:686–706
38. Bye JW, Platts L, Falconer RJ (2014) Biopharmaceutical liquid formulation: a review of the science of protein stability and solubility in aqueous environments. Biotechnol Lett 36:869–875
39. Richter WF, Bhansali SG, Morris ME (2012) Mechanistic determinants of biotherapeutics absorption following SC administration. AAPS J 14:559–570
40. Bee JS, Stevenson JL, Mehta B, Svitel J, Pollastrini J, Platz R, Freund E, Carpenter JF, Randolph TW (2009) Response of a concentrated monoclonal antibody formulation to high shear. Biotechnol Bioeng 103:936–943
41. Tyagi AK, Randolph TW, Dong A, Maloney KM, Hitscherich C, Carpenter JF (2009) IgG particle formation during filling pump operation: a case study of heterogeneous nucleation on stainless steel nanoparticles. J Pharm Sci 98:94–104
42. Collins DS, Kourtis LC, Thyagarajapuram NR, Sirkar R, Kapur S, Harrison MW, Bryan DJ, Jones GB, Wright JM (2017) Optimizing the bioavailability of subcutaneously administered biotherapeutics through Mechanochemical drivers. Pharm Res 34:2000–2011
43. Roberts CJ (2014) Therapeutic protein aggregation: mechanisms, design, and control. Trends Biotechnol 32:372–380
44. Sydow JF, Lipsmeier F, Larraillet V, Hilger M, Mautz B, Mølhøj M, Kuentzer J, Klostermann S, Schoch J, Voelger HR, Regula JT, Cramer P, Papadimitriou A, Kettenberger H (2014) Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions. PLoS One 9:e100736
45. Lu X, Nobrega RP, Lynaugh H, Jain T, Barlow K, Boland T, Sivasubramanian A, Vásquez M, Xu Y (2018) Deamidation and isomerization liability analysis of 131 clinical-stage antibodies. mAbs 262:1–13
46. Adem YT, Molina P, Liu H, Patapoff TW, Sreedhara A, Esue O (2014) Hexyl glucoside and hexyl Maltoside inhibit light-induced oxidation of tryptophan. J Pharm Sci 103:409–416
47. Lam XM, Lai WG, Chan EK, Ling V, Hsu CC (2011) Site-specific tryptophan oxidation induced by autocatalytic reaction of Polysorbate 20 in protein formulation. Pharm Res 28:2543–2555
48. Torosantucci R, Mozziconacci O, Sharov V, Schöneich C, Jiskoot W (2012) Chemical modifications in aggregates of recombinant human insulin induced by metal-catalyzed oxidation: covalent cross-linking via Michael addition to tyrosine oxidation products. Pharm Res 29:2276–2293
49. Torosantucci R, Schöneich C, Jiskoot W (2014) Oxidation of therapeutic proteins and peptides: structural and biological consequences. Pharm Res 31:541–553
50. Ji JA, Zhang B, Cheng W, Wang YJ (2009) Methionine, tryptophan, and histidine oxidation in a model protein, PTH: mechanisms and stabilization. J Pharm Sci 98:4485–4500
51. Anraku M, Tsurusaki Y, Watanabe H, Maruyama T, Kragh-Hansen U, Otagiri M (2004) Stabilizing mechanisms in commercial albumin preparations: octanoate and N-acetyl-l-tryptophanate protect human serum albumin against heat and oxidative stress. Biochim Biophys Acta 1702:9–17
52. Grebenau RC, Goldenberg DM, Chien-Hsing C, Koch GA, Gold DV, Kunz A, Hansen HJ (1992) Microheterogeneity of a purified IgG1, due to asymmetric fab glycosylation. Mol Immunol 29:751–758
53. Coloma MJ, Trinh RK, Martinez AR, Morrison SL (1999) Position effects of variable region carbohydrate on the affinity and in vivo behavior of an anti-(1→6) dextran antibody. J Immunol 162:2162–2170
54. Huang L, Biolsi S, Bales KR, Kuchibhotla U (2006) Impact of variable domain glycosylation on antibody clearance: an LC/MS characterization. Anal Biochem 349:197–207
55. Wang Y, Lu Q, Wu S-L, Karger BL, Hancock WS (2011) Characterization and comparison of disulfide linkages and scrambling patterns

in therapeutic monoclonal antibodies - using LC-MS with electron transfer dissociation. Anal Chem 83:3133–3140
56. Dudgeon K, Rouet R, Kokmeijer I, Schofield P, Stolp J, Langley D, Stock D, Christ D (2012) General strategy for the generation of human antibody variable domains with increased aggregation resistance. Proc Natl Acad Sci U S A 109:10879–10884
57. Perchiacca JM, Ladiwala ARA, Bhattacharya M, Tessier PM (2012) Aggregation-resistant domain antibodies engineered with charged mutations near the edges of the complementarity-determining regions. Protein Eng Des Sel 25:591–601
58. Perchiacca JM, Lee CC, Tessier PM (2014) Optimal charged mutations in the complementarity-determining regions that prevent domain antibody aggregation are dependent on the antibody scaffold. Protein Eng Des Sel 27:29–39
59. Perchiacca JM, Bhattacharya M, Tessier PM (2011) Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions. Proteins 79:2637–2647
60. Kayser V, Chennamsetty N, Voynov V, Helk B, Trout BL (2011) Conformational stability and aggregation of therapeutic monoclonal antibodies studied with ANS and thioflavin T binding. mAbs 3:408–411
61. Kowalski JM, Parekh RN, Mao J, Wittrup KD (1998) Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. J Biol Chem 273:19453–19458
62. Kowalski JM, Parekh RN, Wittrup KD (1998) Secretion efficiency in *Saccharomyces cerevisiae* of bovine pancreatic trypsin inhibitor mutants lacking disulfide bonds is correlated with thermodynamic stability. Biochemistry 37:1264–1273
63. Sormanni P, Vendruscolo M (2019) Protein solubility predictions using the CamSol method in the study of protein homeostasis. Cold Spring Harb Perspect Biol 11(12): a033845
64. Moussa EM, Panchal JP, Moorthy BS, Blum JS, Joubert MK, Narhi LO, Topp EM (2016) Immunogenicity of therapeutic protein aggregates. J Pharm Sci 105:417–430. https://doi.org/10.1016/j.xphs.2015.11.002
65. De Groot AS, Scott DW (2007) Immunogenicity of protein therapeutics. Trends Immunol 28:482–490
66. Hansel TT, Kropshofer H, Singer T, Mitchell JA, George AJT (2010) The safety and side effects of monoclonal antibodies. Nat Rev Drug Discov 9:325–338
67. Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJA, Middaugh CR, Winter G, Fan Y-X, Kirshner S, Verthelyi D, Kozlowski S, Clouse KA, Swann PG, Rosenberg A, Cherney B (2009) Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. J Pharm Sci 98:1201–1205
68. Singh SK, Afonina N, Awwad M, Bechtold-Peters K, Blue JT, Chou D, Cromwell M, Krause H-J, Mahler H-C, Meyer BK, Narhi L, Nesta DP, Spitznagel T (2010) An industry perspective on the monitoring of subvisible particles as a quality attribute for protein therapeutics. J Pharm Sci 99:3302–3321
69. Wang W, Singh S, Zeng DL, King K, Nema S (2007) Antibody structure, instability, and formulation. J Pharm Sci 96:1–26
70. Lazar KL, Patapoff TW, Sharma VK (2010) Cold denaturation of monoclonal antibodies. mAbs 2:42–52
71. Temel DB, Landsman P, Brader ML (2016) Orthogonal methods for characterizing the unfolding of therapeutic monoclonal antibodies. In: Methods in Enzymology. Elsevier, Amsterdam, pp 359–389
72. Dobson CM (2003) Protein folding and misfolding. Nature 426:884–890
73. Kazlauskas R (2018) Engineering more stable proteins. Chem Soc Rev 47:9026–9045
74. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS (2010) Stability of protein pharmaceuticals: an update. Pharm Res 27:544–575
75. Trevino SR, Scholtz JM, Pace CN (2008) Measuring and increasing protein solubility. J Pharm Sci 97:4155–4166
76. Pindrus M, Shire SJ, Kelley RF, Demeule B, Wong R, Xu Y, Yadav S (2015) Solubility challenges in high concentration monoclonal antibody formulations: relationship with amino acid sequence and intermolecular interactions. Mol Pharm 12:3896–3907
77. Kanai S, Liu J, Patapoff TW, Shire SJ (2008) Reversible self-association of a concentrated monoclonal antibody solution mediated by Fab-Fab interaction that impacts solution viscosity. J Pharm Sci 97:4219–4227
78. Wu S-J, Gilliland GL, Feng Y (2014) Solubility and early assessment of stability for protein therapeutics. In: Biophysical methods for

biotherapeutics. John Wiley & Sons Ltd, Hoboken, New Jersey, pp 65–91

79. Gibson TJ, Mccarty K, McFadyen IJ, Cash E, Dalmonte P, Hinds KD, Dinerman AA, Alvarez JC, Volkin DB (2011) Application of a high-throughput screening procedure with PEG-induced precipitation to compare relative protein solubility during formulation development with IgG1 monoclonal antibodies. J Pharm Sci 100:1009–1021
80. Toprani VM, Joshi SB, Kueltzo LA, Schwartz RM, Middaugh CR, Volkin DB (2016) A micro–polyethylene glycol precipitation assay as a relative solubility screening tool for monoclonal antibody design and formulation development. J Pharm Sci 105:2319–2327
81. Nicoud L, Owczarz M, Arosio P, Morbidelli M (2015) A multiscale view of therapeutic protein aggregation: a colloid science perspective. Biotechnol J 10:367–378
82. Mahler H-C, Friess W, Grauschopf U, Kiese S (2009) Protein aggregation: pathways, induction factors and analysis. J Pharm Sci 98:2909–2934
83. Arora J, Hu Y, Esfandiary R, Sathish HA, Bishop SM, Joshi SB, Middaugh CR, Volkin DB, Weis DD (2016) Charge-mediated fab-fc interactions in an IgG1 antibody induce reversible self-association, cluster formation, and elevated viscosity. mAbs 8:1561–1574
84. Kumar V, Dixit N, Zhou LL, Fraunhofer W (2011) Impact of short range hydrophobic interactions and long range electrostatic forces on the aggregation kinetics of a monoclonal antibody and a dual-variable domain immunoglobulin at low and high concentrations. Int J Pharm 421:82–93
85. Liu J, Nguyen MDH, Andya JD, Shire SJ (2005) Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. J Pharm Sci 94:1928–1940
86. Alam ME, Geng SB, Bender C, Ludwig SD, Linden L, Hoet R, Tessier PM (2018) Biophysical and sequence-based methods for identifying monovalent and bivalent antibodies with high colloidal stability. Mol Pharm 15:150–163
87. Tiller KE, Li L, Kumar S, Julian MC, Garde S, Tessier PM (2017) Arginine mutations in antibody complementarity-determining regions display context-dependent affinity/ specificity trade-offs. J Biol Chem 292:16638
88. Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL (2010) Prediction of aggregation prone regions of therapeutic proteins. J Phys Chem B 114:6614–6624
89. Lauer TM, Agrawal NJ, Chennamsetty N, Egodage K, Helk B, Trout BL (2012) Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. J Pharm Sci 101:102–115
90. Jain T, Boland T, Lilov A, Burnina I, Brown M, Xu Y, Vásquez M (2017) Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning. Bioinformatics 33:3758–3766
91. Lawrence MS, Phillips KJ, Liu DR (2007) Supercharging proteins can impart unusual resilience. J Am Chem Soc 129:10110–10112
92. Miklos AE, Kluwe C, Der BS, Pai S, Sircar A, Hughes RA, Berrondo M, Xu J, Codrea V, Buckley PE, Calm AM, Welsh HS, Warner CR, Zacharko MA, Carney JP, Gray JJ, Georgiou G, Kuhlman B, Ellington AD (2012) Structure-based design of supercharged, highly thermoresistant antibodies. Chem Biol 19:449–455
93. Trevino SR, Scholtz JM, Pace CN (2007) Amino acid contribution to protein solubility: asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. J Mol Biol 366:449–460
94. Yadav S, Shire SJ, Kalonia DS (2012) Viscosity behavior of high-concentration monoclonal antibody solutions: correlation with interaction parameter and electroviscous effects. J Pharm Sci 101:998–1011
95. Jain D, Salunke DM (2019) Antibody specificity and promiscuity. Biochem J 476:433–447
96. Xu Y, Roach W, Sun T, Jain T, Prinz B, Yu T-Y, Torrey J, Thomas J, Bobrowicz P, Vásquez M, Wittrup KD, Krauland E (2013) Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. Protein Eng Des Sel 26:663–670
97. Hötzel I, Theil F-P, Bernstein LJ, Prabhu S, Deng R, Quintana L, Lutman J, Sibia R, Chan P, Bumbaca D, Fielder P, Carter PJ, Kelley RF (2012) A strategy for risk mitigation of antibodies with fast clearance. mAbs 4:753–760
98. Vugmeyster Y, Guay H, Szklut P, Qian MD, Jin M, Widom A, Spaulding V, Bennett F, Lowe L, Andreyeva T, Lowe D, Lane S, Thom G, Valge-Archer V, Gill D, Young D, Bloom L (2010) In vitro potency, pharmacokinetic profiles and pharmacological activity of optimized anti-IL-21R antibodies in a mouse model of lupus. mAbs 2:335–346

99. Wu H, Pfarr DS, Johnson S, Brewah YA, Woods RM, Patel NK, White WI, Young JF, Kiener PA (2007) Development of Motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. J Mol Biol 368:652–665
100. Geng SB, Wittekind M, Vigil A, Tessier PM (2016) Measurements of monoclonal antibody self-association are correlated with complex biophysical properties. Mol Pharm 13:1636–1645
101. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC (2003) Predominant autoantibody production by early human B cell precursors. Science 301:1374–1377
102. Warszawski S, Katz AB, Lipsh R, Khmelnitsky L, Nissan GB, Javitt G, Dym O, Unger T, Knop O, Albeck S, Diskin R, Fass D, Sharon M, Fleishman SJ (2019) Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. PLoS Comput Biol 15:e1007207
103. Spencer S, Bethea D, Raju TS, Giles-Komar J, Feng Y (2012) Solubility evaluation of murine hybridoma antibodies. mAbs 4:319–325
104. Burkovitz A, Sela-Culang I, Ofran Y (2014) Large-scale analysis of somatic hypermutations in antibodies reveals which structural regions, positions and amino acids are modified to improve affinity. FEBS J 281:306–319
105. Pepinsky RB, Silvian L, Berkowitz SA, Farrington G, Lugovskoy A, Walus L, Eldredge J, Capili A, Mi S, Graff C, Garber E (2010) Improving the solubility of anti-LINGO-1 monoclonal antibody Li33 by isotype switching and targeted mutagenesis. Protein Sci 19:954–966
106. Wu S-J, Luo J, O'Neil KT, Kang J, Lacy ER, Canziani G, Baker A, Huang M, Tang QM, Raju TS, Jacobs SA, Teplyakov A, Gilliland GL, Feng Y (2010) Structure-based engineering of a monoclonal antibody for improved solubility. Protein Eng Des Sel 23:643–651
107. Avery LB, Wade J, Wang M, Tam A, King A, Piche-Nicholas N, Kavosi MS, Penn S, Cirelli D, Kurz JC, Zhang M, Cunningham O, Jones R, Fennell BJ, McDonnell B, Sakorafas P, Apgar J, Finlay WJ, Lin L, Bloom L, O'Hara DM (2018) Establishing in vitro in vivo correlations to screen monoclonal antibodies for physicochemical properties related to favorable human pharmacokinetics. mAbs 10:244–255
108. Finlay WJJ, Coleman JE, Edwards JS, Johnson KS (2019) Anti-PD1 'SHR-1210′ aberrantly targets pro-angiogenic receptors and this polyspecificity can be ablated by paratope refinement. mAbs 11:26–44
109. Ratanji KD, Derrick JP, Dearman RJ, Kimber I (2014) Immunogenicity of therapeutic proteins: influence of aggregation. J Immunotoxicol 11:99–109
110. Nelson AL, Dhimolea E, Reichert JM (2010) Development trends for human monoclonal antibody therapeutics. Nat Rev Drug Discov 9:767–774
111. Lonberg N, Taylor LD, Harding FA, Trounstine M, Higgins KM, Schramm SR, Kuo CC, Mashayekh R, Wymore K, McCabe JG (1994) Antigen-specific human antibodies from mice comprising four distinct genetic modifications. Nature 368:856–859
112. Jakobovits A, Amado RG, Yang X, Roskos L, Schwab G (2007) From XenoMouse technology to panitumumab, the first fully human antibody product from transgenic mice. Nat Biotechnol 25:1134–1143
113. Ponsel D, Neugebauer J, Ladetzki-Baehs K, Tissot K (2011) High affinity, developability and functional size: the holy grail of combinatorial antibody library generation. Molecules 16:3675–3700
114. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR (1994) Making antibodies by phage display technology. Annu Rev Immunol 12:433–455
115. Smith GP, Petrenko VA (1997) Phage display. Chem Rev 97:391–410
116. Bradbury ARM, Sidhu S, Dübel S, McCafferty J (2011) Beyond natural antibodies: the power of in vitro display technologies. Nat Biotechnol 29:245–254
117. Macdougall IC (2005) Antibody-mediated pure red cell aplasia (PRCA): epidemiology, immunogenicity and risks. Nephrol Dial Transplant 20(Suppl 4):iv9–iv15
118. Anfinsen CB (1972) The formation and stabilization of protein structure. Biochem J 128:737–749
119. Knowles TPJ, Waudby CA, Devlin GL, Cohen SIA, Aguzzi A, Vendruscolo M, Terentjev EM, Welland ME, Dobson CM (2009) An analytical solution to the kinetics of breakable filament assembly. Science 326:1533–1537
120. Fersht AR (1999) Structure and mechanism in protein science. Macmillan, New York City
121. Vendruscolo M, Paci E, Dobson CM, Karplus M (2001) Three key residues form a critical

contact network in a protein folding transition state. Nature 409:641–645

122. Broom A, Jacobi Z, Trainor K, Meiering EM (2017) Computational tools help improve protein stability but with a solubility tradeoff. J Biol Chem 292:14349–14361
123. Goldenzweig A, Fleishman S (2018) Principles of protein stability and their application in computational design. Annu Rev Biochem 87:annurev-biochem-062917-012102
124. Espargaró A, Castillo V, de Groot NS, Ventura S (2008) The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: the SH3 case. J Mol Biol 378:1116–1131
125. Thakkar SV, Sahni N, Joshi SB, Kerwin BA, He F, Volkin DB, Middaugh CR (2013) Understanding the relevance of local conformational stability and dynamics to the aggregation propensity of an IgG1 and IgG2 monoclonal antibodies. Protein Sci 22:1295–1305
126. Trainor K, Gingras Z, Shillingford C, Malakian H, Gosselin M, Lipovšek D, Meiering EM (2016) Ensemble modeling and intracellular aggregation of an engineered immunoglobulin-like domain. J Mol Biol 428:1365–1374
127. Stenvang M, Schafer NP, Malmos KG, Pérez A-MW, Niembro O, Sormanni P, Basaiawmoit RV, Christiansen G, Andreasen M, Otzen DE (2018) Corneal dystrophy mutations drive pathogenesis by targeting TGFBIp stability and solubility in a latent amyloid-forming domain. J Mol Biol 430:1116–1140
128. Sormanni P, Amery L, Ekizoglou S, Vendruscolo M, Popovic B (2017) Rapid and accurate in silico solubility screening of a monoclonal antibody library. Sci Rep 7:8200
129. Julian MC, Lee CC, Tiller KE, Rabia LA, Day EK, Schick AJ, Tessier PM (2015) Co-evolution of affinity and stability of grafted amyloid-motif domain antibodies. Protein Eng Des Sel 28:339–350
130. Julian MC, Li L, Garde S, Wilen R, Tessier PM (2017) Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability. Sci Rep 7:45259
131. Nguyen MN, Pradhan MR, Verma C, Zhong P (2017) The interfacial character of antibody paratopes: analysis of antibody-antigen structures. Bioinformatics 33:2971–2976
132. Kelly RL, Le D, Zhao J, Wittrup KD (2018) Reduction of nonspecificity motifs in synthetic antibody libraries. J Mol Biol 430:119–130
133. Peng H-P, Lee KH, Jian J-W, Yang A-S (2014) Origins of specificity and affinity in antibody-protein interactions. Proc Natl Acad Sci U S A 111:E2656–E2665
134. Xu Y, Wang D, Mason B, Rossomando T, Li N, Liu D, Cheung JK, Xu W, Raghava S, Katiyar A, Nowak C, Xiang T, Dong DD, Sun J, Beck A, Liu H (2019) Structure, heterogeneity and developability assessment of therapeutic antibodies. mAbs 11:239–264
135. Zurdo J, Arnell A, Obrezanova O, Smith N, Gómez de la Cuesta R, Gallagher TRA, Michael R, Stallwood Y, Ekblad C, Abrahmsén L, Höidén-Guthenberg I (2015) Early Implementation of QbD in Biopharmaceutical Development: A Practical Example. Biomed Res Int 2015:605427
136. Norman RA, Ambrosetti F, Bonvin AMJJ, Colwell LJ, Kelm S, Kumar S, Krawczyk K (2020) Computational approaches to therapeutic antibody design: established methods and emerging trends. Brief Bioinform 21 (5):1549–1567
137. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1:2876–2890
138. Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, Fielden PR, Fogarty SW, Fullwood NJ, Heys KA, Hughes C, Lasch P, Martin-Hirsch PL, Obinaju B, Sockalingum GD, Sulé-Suso J, Strong RJ, Walsh MJ, Wood BR, Gardner P, Martin FL (2014) Using Fourier transform IR spectroscopy to analyze biological materials. Nat Protoc 9:1771–1791
139. Filipe V, Hawe A, Carpenter JF, Jiskoot W (2013) Analytical approaches to assess the degradation of therapeutic proteins. Trac-Trend Anal Chem 49:118–125
140. Ranjbar B, Gill P (2009) Circular dichroism techniques: biomolecular and Nanostructural analyses- a review. Chem Biol Drug Des 74:101–120
141. Fiedler S, Cole L, Keller S (2013) Automated circular dichroism spectroscopy for medium-throughput analysis of protein conformation. Anal Chem 85:1868–1872
142. Schermeyer M-T, Wöll AK, Kokke B, Eppink M, Hubbuch J (2017) Characterization of highly concentrated antibody solution - a toolbox for the description of protein long-term solution stability. mAbs 9:1169–1185
143. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. Biophys J 80:2093–2109
144. Saluja A, Kalonia DS (2008) Nature and consequences of protein-protein interactions in

high protein concentration solutions. Int J Pharm 358:1–15

145. Curtis RA, Prausnitz JM, Blanch HW (1998) Protein-protein and protein-salt interactions in aqueous protein solutions containing concentrated electrolytes. Biotechnol Bioeng 57:11–21
146. Saito S, Hasegawa J, Kobayashi N, Kishi N, Uchiyama S, Fukui K (2012) Behavior of monoclonal antibodies: relation between the second virial coefficient (B2) at low concentrations and aggregation propensity and viscosity at high concentrations. Pharm Res 29:397–410
147. Blanco MA, Perevozchikova T, Martorana V, Manno M, Roberts CJ (2014) Protein–protein interactions in dilute to concentrated solutions: α-Chymotrypsinogen in acidic conditions. J Phys Chem B 118:5817–5831
148. Ruppert S, Sandler SI, Lenhoff AM (2001) Correlation between the osmotic second virial coefficient and the solubility of proteins. Biotechnol Prog 17:182–187
149. Le Brun V, Friess W, Bassarab S, Mühlau S, Garidel P (2010) A critical evaluation of self-interaction chromatography as a predictive tool for the assessment of protein–protein interactions in protein formulation development: a case study of a therapeutic monoclonal antibody. Eur J Pharm Biopharm 75:16–25
150. Haas C, Drenth J, Wilson WW (1999) Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution. J Phys Chem B 103:2808–2811
151. Connolly BD, Petry C, Yadav S, Demeule B, Ciaccio N, Moore JMR, Shire SJ, Gokarn YR (2012) Weak interactions govern the viscosity of concentrated antibody solutions: high-throughput analysis using the diffusion interaction parameter. Biophys J 103:69–78
152. Harding SE, Johnson P (1985) The concentration-dependence of macromolecular parameters. Biochem J 231:543–547
153. Saluja A, Badkar AV, Zeng DL, Nema S, Kalonia DS (2007) Ultrasonic storage modulus as a novel parameter for analyzing protein-protein interactions in high protein concentration solutions: correlation with static and dynamic light scattering measurements. Biophys J 92:234–244
154. He F, Woods CE, Becker GW, Narhi LO, Razinkov VI (2011) High-throughput assessment of thermal and colloidal stability parameters for monoclonal antibody formulations. J Pharm Sci 100:5126–5141
155. Yadav S, Shire SJ, Kalonia DS (2010) Factors affecting the viscosity in high concentration solutions of different monoclonal antibodies. J Pharm Sci 99:4812–4829
156. Schein CH (1990) Solubility as a function of protein structure and solvent components. Biotechnology 8:308–317
157. Tomar DS, Kumar S, Singh SK, Goswami S, Li L (2016) Molecular basis of high viscosity in concentrated antibody solutions: strategies for high concentration drug product development. mAbs 8:216–228
158. Wang W, Nema S, Teagarden D (2010) Protein aggregation--pathways and influencing factors. Int J Pharm 390:89–99
159. Wang W (1999) Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 185:129–188
160. Chi EY, Krishnan S, Randolph TW, Carpenter JF (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm Res 20:1325–1336
161. Hong P, Koza S, Bouvier ESP (2012) A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. J Liq Chromatogr Relat Technol 35:2923–2950
162. Liu J, Andya JD, Shire SJ (2006) A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. AAPS J 8: E580–E589
163. Hofmann M, Gieseler H (2018) Predictive screening tools used in high-concentration protein formulation development. J Pharm Sci 107:772–777
164. Weiss WF, Young TM, Roberts CJ (2009) Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. J Pharm Sci 98:1246–1277
165. Hawe A, Wiggenhorn M, van de Weert M, Garbe JHO, Mahler H-C, Jiskoot W (2012) Forced degradation of therapeutic proteins. J Pharm Sci 101:895–913
166. Joubert MK, Luo Q, Nashed-Samuel Y, Wypych J, Narhi LO (2011) Classification and characterization of therapeutic antibody aggregates. J Biol Chem 286:25118–25133
167. Sahin E, Grillo AO, Perkins MD, Roberts CJ (2010) Comparative effects of pH and ionic strength on protein-protein interactions, unfolding, and aggregation for IgG1 antibodies. J Pharm Sci 99:4830–4848
168. Wang W, Roberts CJ (2013) Non-Arrhenius Protein Aggregation. AAPS J 15:840–851

169. Kameoka D, Masuzaki E, Ueda T, Imoto T (2007) Effect of buffer species on the unfolding and the aggregation of humanized IgG. J Biochem 142:383–391
170. Kayser V, Chennamsetty N, Voynov V, Helk B, Forrer K, Trout BL (2011) Evaluation of a non-Arrhenius model for therapeutic monoclonal antibody aggregation. J Pharm Sci 100:2526–2542
171. Goldberg DS, Bishop SM, Shah AU, Sathish HA (2011) Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: role of conformational and colloidal stability. J Pharm Sci 100:1306–1315
172. Sormanni P, Aprile FA, Vendruscolo M (2015) The CamSol method of rational Design of Protein Mutants with enhanced solubility. J Mol Biol 427:478–490
173. Cromwell MEM, Hilario E, Jacobson F (2006) Protein aggregation and bioprocessing. AAPS J 8:E572–E579
174. Li L, Kantor A, Warne N (2013) Application of a PEG precipitation method for solubility screening: a tool for developing high protein concentration formulations. Protein Sci 22:1118–1123
175. Chai Q, Shih J, Weldon C, Phan S, Jones BE (2019) Development of a high-throughput solubility screening assay for use in antibody discovery. mAbs 11:747–756
176. Laue T (2012) Proximity energies: a framework for understanding concentrated solutions. J Mol Recognit 25:165–173
177. Minton AP (2006) Macromolecular crowding. Curr Biol 16:R269–R271
178. Geoghegan JC, Fleming R, Damschroder M, Bishop SM, Sathish HA, Esfandiary R (2016) Mitigation of reversible self-association and viscosity in a human IgG1 monoclonal antibody by rational, structure-guided Fv engineering. mAbs 8:941–950
179. Tilegenova C, Izadi S, Yin J, Huang CS, Wu J, Ellerman D, Hymowitz SG, Walters B, Salisbury C, Carter PJ (2019) Dissecting the molecular basis of high viscosity of monospecific and bispecific IgG antibodies. mAbs 12 (1):1692764
180. Zhang Z, Liu Y (2017) Recent progresses of understanding the viscosity of concentrated protein solutions. Curr Opin Chem Eng 16:48–55
181. He F, Becker GW, Litowski JR, Narhi LO, Brems DN, Razinkov VI (2010) High-throughput dynamic light scattering method for measuring viscosity of concentrated protein solutions. Anal Biochem 399:141–143
182. Arosio P, Hu K, Aprile FA, Müller T, Knowles TPJ (2016) Microfluidic diffusion viscometer for rapid analysis of complex solutions. Anal Chem 88:3488–3493
183. Kopp MRG, Villois A, Capasso Palmiero U, Arosio P (2018) Microfluidic diffusion analysis of the size distribution and microrheological properties of antibody solutions at high concentrations. Ind Eng Chem Res 57:7112–7120
184. Neergaard MS, Kalonia DS, Parshad H, Nielsen AD, Møller EH, van de Weert M (2013) Viscosity of high concentration protein formulations of monoclonal antibodies of the IgG1 and IgG4 subclass—prediction of viscosity through protein-protein interaction measurements. Eur J Pharm Sci 49:400–410
185. Wright TA, Stewart JM, Page RC, Konkolewicz D (2017) Extraction of thermodynamic parameters of protein unfolding using parallelized differential scanning Fluorimetry. J Phys Chem Lett 8:553–558
186. Johnson CM (2013) Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys 531:100–109
187. Jacobs SA, Wu S-J, Feng Y, Bethea D, O'Neil KT (2009) Cross-interaction chromatography: a rapid method to identify highly soluble monoclonal antibody candidates. Pharm Res 27:65
188. Kohli N, Jain N, Geddie ML, Razlog M, Xu L, Lugovskoy AA (2015) A novel screening method to assess developability of antibody-like molecules. mAbs 7:752–758
189. Kelly RL, Sun T, Jain T, Caffry I, Yu Y, Cao Y, Lynaugh H, Brown M, Vásquez M, Wittrup KD, Xu Y (2015) High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. mAbs 7:770–777
190. Estep P, Caffry I, Yu Y, Sun T, Cao Y, Lynaugh H, Jain T, Vásquez M, Tessier PM, Xu Y (2015) An alternative assay to hydrophobic interaction chromatography for high-throughput characterization of monoclonal antibodies. mAbs 7:553–561
191. Liu Y, Caffry I, Wu J, Geng SB, Jain T, Sun T, Reid F, Cao Y, Estep P, Yu Y, Vásquez M, Tessier PM, Xu Y (2014) High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. mAbs 6:483–492

192. Lueking A, Possling A, Huber O, Beveridge A, Horn M, Eickhoff H, Schuchardt J, Lehrach H, Cahill DJ (2003) A nonredundant human protein Chip for antibody screening and serum profiling. Mol Cell Proteomics 2:1342–1349
193. Feyen O, Lueking A, Kowald A, Stephan C, Meyer HE, Göbel U, Niehues T (2008) Off-target activity of TNF-α inhibitors characterized by protein biochips. Anal Bioanal Chem 391:1713–1720
194. Frese K, Eisenmann M, Ostendorp R, Brocks B, Pabst S (2013) An automated immunoassay for early specificity profiling of antibodies. mAbs 5:279–287
195. Ramos-López P, Irizarry J, Pino I, Blackshaw S (2018) Antibody specificity profiling using protein microarrays. Methods Mol Biol 1785:223–229
196. Sule SV, Dickinson CD, Lu J, Chow C-K, Tessier PM (2013) Rapid analysis of antibody self-association in complex mixtures using immunogold conjugates. Mol Pharm 10:1322–1331
197. Yadav S, Laue TM, Kalonia DS, Singh SN, Shire SJ (2012) The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. Mol Pharm 9:791–802
198. Dobson CL, Devine PWA, Phillips JJ, Higazi DR, Lloyd C, Popovic B, Arnold J, Buchanan A, Lewis A, Goodman J, van der Walle CF, Thornton P, Vinall L, Lowne D, Aagaard A, Olsson L-L, Wollberg AR, Welsh F, Karamanos TK, Pashley CL, Iadanza MG, Ranson NA, Ashcroft AE, Kippen AD, Vaughan TJ, Radford SE, Lowe DC (2016) Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. Sci Rep 6:1–14
199. Tessier PM, Vandrey SD, Berger BW, Pazhianur R, Sandler SI, Lenhoff AM, IUCr (2002) Self-interaction chromatography: a novel screening method for rational protein crystallization. Acta Crystallogr D Biol Crystallogr 58:1531–1535
200. Patro SY, Przybycien TM (1996) Self-interaction chromatography: a tool for the study of protein–protein interactions in bioprocessing environments. Biotechnol Bioeng 52:193–203
201. Tessier PM, Lenhoff AM, Sandler SI (2002) Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. Biophys J 82:1620–1631
202. Johnson DH, Parupudi A, Wilson WW, DeLucas LJ (2008) High-throughput self-interaction chromatography: applications in protein formulation prediction. Pharm Res 26:296
203. Sun T, Reid F, Liu Y, Cao Y, Estep P, Nauman C, Xu Y (2013) High throughput detection of antibody self-interaction by bio-layer interferometry. mAbs 5:838–841
204. Wu J, Schultz JS, Weldon CL, Sule SV, Chai Q, Geng SB, Dickinson CD, Tessier PM (2015) Discovery of highly soluble antibodies prior to purification using affinity-capture self-interaction nanoparticle spectroscopy. Protein Eng Des Sel 28:403–414
205. Shan L, Mody N, Sormani P, Rosenthal KL, Damschroder MM, Esfandiary R (2018) Developability assessment of engineered monoclonal antibody variants with a complex self-association behavior using complementary analytical and in silico tools. Mol Pharm 15:5697–5710
206. Fekete S, Veuthey J-L, Beck A, Guillarme D (2016) Hydrophobic interaction chromatography for the characterization of monoclonal antibodies and related products. J Pharm Biomed Anal 130:3–18
207. Gentiluomo L, Svilenov HL, Augustijn D, El Bialy I, Greco ML, Kulakova A, Indrakumar S, Mahapatra S, Morales MM, Pohl C, Roche A, Tosstorff A, Curtis R, Derrick JP, Nørgaard A, Khan TA, Peters GHJ, Pluen A, Rinnan Å, Streicher WW, van der Walle CF, Uddin S, Winter G, Roessner D, Harris P, Frieß W (2020) Advancing therapeutic protein discovery and development through comprehensive computational and biophysical characterization. Mol Pharm 17:426–440
208. Bradbury A, Plückthun A (2015) Reproducibility: standardize antibodies used in research. Nat News 518:27
209. Tsumoto K, Ejima D, Senczuk AM, Kita Y, Arakawa T (2007) Effects of salts on protein–surface interactions: applications for column chromatography. J Pharm Sci 96:1677–1690
210. Charmet J, Arosio P, Knowles TPJ (2018) Microfluidics for protein biophysics. J Mol Biol 430:565–580
211. Kopp MRG, Arosio P (2018) Microfluidic approaches for the characterization of therapeutic proteins. J Pharm Sci 107:1228–1236
212. Arosio P, Müller T, Rajah L, Yates EV, Aprile FA, Zhang Y, Cohen SIA, White DA, Herling TW, De Genst EJ, Linse S, Vendruscolo M,

Dobson CM, Knowles TPJ (2016) Microfluidic diffusion analysis of the sizes and interactions of proteins under native solution conditions. ACS Nano 10:333–341
213. Obrezanova O, Arnell A, de la Cuesta RG, Berthelot ME, Gallagher TR, Zurdo J, Stallwood Y (2015) Aggregation risk prediction for antibodies and its application to biotherapeutic development. mAbs 7:352–363
214. Meric G, Robinson AS, Roberts CJ (2017) Driving forces for nonnative protein aggregation and approaches to predict aggregation-prone regions. Annu Rev Chem Biomol Eng 8:139–159. https://doi.org/10.1146/annurev-chembioeng-060816-101404
215. Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S (2015) AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. Nucleic Acids Res 43:W306–W313
216. Shan L, Mody N, Sormanni P, Rosenthal KL, Damschroder MM, Esfandiary R (2018) Developability assessment of engineered monoclonal antibody variants with a complex self-association behavior using complementary analytical and in silico tools. Mol Pharm 15(12):5697–5710
217. Camilloni C, Sala BM, Sormanni P, Porcari R, Corazza A, Rosa MD, Zanini S, Barbiroli A, Esposito G, Bolognesi M, Bellotti V, Vendruscolo M, Ricagno S (2016) Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. Sci Rep 6:1–11
218. Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D (2012) PROSO II – a new method for protein solubility prediction. FEBS J 279:2192–2200
219. Agostini F, Vendruscolo M, Tartaglia GG (2012) Sequence-based prediction of protein solubility. J Mol Biol 421:237–241
220. Magnan CN, Randall A, Baldi P (2009) SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics 25:2200–2207
221. Huang H-L, Charoenkwan P, Kao T-F, Lee H-C, Chang F-L, Huang W-L, Ho S-J, Shu L-S, Chen W-L, Ho S-Y (2012) Prediction and analysis of protein solubility using a novel scoring card method with dipeptide composition. BMC Bioinformatics 13:S3
222. Hirose S, Noguchi T (2013) ESPRESSO: a system for estimating protein expression and solubility in protein expression systems. Proteomics 13:1444–1456
223. Chan P, Curtis RA, Warwicker J (2013) Soluble expression of proteins correlates with a lack of positively-charged surface. Sci Rep 3:3333
224. Schaller A, Connors NK, Oelmeier SA, Hubbuch J, Middelberg APJ (2015) Predicting recombinant protein expression experiments using molecular dynamics simulation. Chem Eng Sci 121:340–350
225. Yang Y, Niroula A, Shen B, Vihinen M (2016) PON-sol: prediction of effects of amino acid substitutions on protein solubility. Bioinformatics 32:2032–2034
226. Paladin L, Piovesan D, Tosatto SCE (2017) SODA: prediction of protein solubility from disorder and aggregation propensity. Nucleic Acids Res 45:W236–W240
227. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J (2017) Protein–sol: a web tool for predicting protein solubility from sequence. Bioinformatics 33:3098–3100
228. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide andprotein aggregation rates. Nature 424:805–808
229. DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. J Mol Biol 341:1317–1326
230. Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22:1302–1306
231. Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. BMC Bioinformatics 8:65
232. Tartaglia GG, Vendruscolo M (2008) The Zyggregator method for predicting protein aggregation propensities. Chem Soc Rev 37:1395–1401
233. Garbuzynskiy SO, Lobanov MY, Galzitskaya OV (2010) FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. Bioinformatics 26:326–332
234. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D (2006) The 3D profile method for identifying fibril-forming segments of proteins. Proc Natl Acad Sci U S A 103:4074–4078

235. Goldschmidt L, Teng PK, Riek R, Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci U S A 107:3487–3492
236. Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JWH, Rousseau F (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods 7:237–242
237. O'Donnell CW, Waldispühl J, Lis M, Halfmann R, Devadas S, Lindquist S, Berger B (2011) A method for probing the mutational landscape of amyloid structure. Bioinformatics 27:i34–i42
238. Tsolis AC, Papandreou NC, Iconomidou VA, Hamodrakas SJ (2013) A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. PLoS One 8: e54175
239. Emily M, Talvas A, Delamarche C (2013) MetAmyl: a METa-predictor for AMYLoid proteins. PLoS One 8:e79722
240. Trovato A, Seno F, Tosatto SCE (2007) The PASTA server for protein aggregation prediction. Protein Eng Des Sel 20:521–523
241. Trovato A, Chiti F, Maritan A, Seno F (2006) Insight into the structure of amyloid fibrils from the analysis of globular proteins. PLoS Comput Biol 2:e170
242. Walsh I, Seno F, Tosatto SCE, Trovato A (2014) PASTA 2.0: an improved server for protein aggregation prediction. Nucleic Acids Res 42:W301–W307
243. Gasior P, Kotulska M (2014) FISH amyloid—a new method for finding amyloidogenic segments in proteins based on site specific co-occurence of aminoacids. BMC Bioinformatics 15:54
244. Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S (2019) Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. Nucleic Acids Res 47: W300–W307
245. Agrawal NJ, Helk B, Kumar S, Mody N, Sathish HA, Samra HS, Buck PM, Li L, Trout BL (2015) Computational tool for the early screening of monoclonal antibodies for their viscosities. mAbs 8(1):43–48
246. Tomar DS, Li L, Broulidakis MP, Luksha NG, Burns CT, Singh SK, Kumar S (2017) In-silico prediction of concentration-dependent viscosity curves for monoclonal antibody solutions. mAbs 9:476–489
247. De Groot AS, Sbai H, Aubin CS, McMurry J, Martin W (2002) Immuno-informatics: mining genomes for vaccine components. Immunol Cell Biol 80:255–269
248. Dalkas GA, Rooman M (2017) SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. BMC Bioinformatics 18:95
249. Gil-Garcia M, Bañó-Polo M, Varejão N, Jamroz M, Kuriata A, Díaz-Caballero M, Lascorz J, Morel B, Navarro S, Reverter D, Kmiecik S, Ventura S (2018) Combining structural aggregation propensity and stability predictions to redesign protein solubility. Mol Pharm 15:3846–3859
250. Hebditch M, Warwicker J (2019) Web-based display of protein surface and pH-dependent properties for assessing the developability of biotherapeutics. Sci Rep 9:1969
251. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424:805–808
252. Fowler SB, Poon S, Muff R, Chiti F, Dobson CM, Zurdo J (2005) Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. Proc Natl Acad Sci U S A 102:10105–10110
253. Trainor K, Broom A, Meiering EM (2017) Exploring the relationships between protein sequence, structure and solubility. Curr Opin Struct Biol 42:136–146
254. Pastor MT, Esteras-Chopo A, Serrano L (2007) Hacking the code of amyloid formation: the amyloid stretch hypothesis. Prion 1:9–14
255. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366
256. Dobson CL, Devine PWA, Phillips JJ, Higazi DR, Lloyd C, Popovic B, Arnold J, Buchanan A, Lewis A, Goodman J, van der Walle CF, Thornton P, Vinall L, Lowne D, Aagaard A, Olsson L-L, Ridderstad Wollberg A, Welsh F, Karamanos TK, Pashley CL, Iadanza MG, Ranson NA, Ashcroft AE, Kippen AD, Vaughan TJ, Radford SE, Lowe DC (2016) Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. Sci Rep 6:38644
257. Dunbar J, Krawczyk K, Leem J, Marks C, Nowak J, Regep C, Georges G, Kelm S, Popovic B, Deane CM (2016) SAbPred: a

structure-based antibody prediction server. Nucleic Acids Res 44:W474–W478

258. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. Nucleic Acids Res 33:W382–W388
259. Adolf-Bryfogle J, Xu Q, North B, Lehmann A, Dunbrack RL (2015) PyIgClassify: a database of antibody CDR structural classifications. Nucleic Acids Res 43: D432–D438
260. Di Rienzo L, Milanetti E, Lepore R, Olimpieri PP, Tramontano A (2017) Superposition-free comparison and clustering of antibody binding sites: implications for the prediction of the nature of their antigen. Sci Rep 7:45053
261. Al-Lazikani B, Lesk AM, Chothia C (1997) Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 273:927–948
262. Krawczyk K, Dunbar J, Deane CM (2017) Computational tools for aiding rational antibody design. Methods Mol Biol 1529:399–416
263. Regep C, Georges G, Shi J, Popovic B, Deane CM (2017) The H3 loop of antibodies shows unique structural characteristics. Proteins 85:1311–1318
264. Marks C, Deane CM (2017) Antibody H3 structure prediction. Comput Struct Biotechnol J 15:222–231
265. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
266. Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350:379–392
267. Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M (2008) Prediction of aggregation-prone regions in structured proteins. J Mol Biol 380:425–436
268. Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, Ciryam P, Morimoto RI, Dobson CM, Vendruscolo M, Mann M, Hartl FU (2015) Widespread proteome remodeling and aggregation in Aging C. elegans. Cell 161:919–932
269. Família C, Dennison SR, Quintas A, Phoenix DA (2015) Prediction of peptide and protein propensity for amyloid formation. PLoS One 10:e0134679
270. Flower DR (2003) Towards in silico prediction of immunogenic epitopes. Trends Immunol 24:667–674
271. Fowler DM, Fields S (2014) Deep mutational scanning: a new style of protein science. Nat Methods 11:801–807
272. Gray VE, Sitko K, Kameni FZN, Williamson M, Stephany JJ, Hasle N, Fowler DM (2019) Elucidating the molecular determinants of Aβ aggregation with deep mutational scanning. G3 9:3683–3689
273. Walle IV, Gansemans Y, Parren PW, Stas P, Lasters I (2007) Immunogenicity screening in protein drug development. Expert Opin Biol Ther 7:405–418
274. De Groot AS, McMurry J, Moise L (2008) Prediction of immunogenicity: in silico paradigms, ex vivo and in vivo correlates. Curr Opin Pharmacol 8:620–626
275. Kennedy PJ, Oliveira C, Granja PL, Sarmento B (2018) Monoclonal antibodies: technologies for early discovery and engineering. Crit Rev Biotechnol 38:394–408
276. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7:525–537