Chem Soc Rev



View Article Online

REVIEW ARTICLE

Check for updates

Cite this: Chem. Soc. Rev., 2018, 47, 9137

Received 25th June 2018 DOI: 10.1039/c8cs00523k

rsc.li/chem-soc-rev

Introduction

Antibodies and antibody fragments

Antibodies are proteins produced by the immune system to target and neutralise pathogens and toxins. The binding of antibodies to their target molecules, referred to as antigens, triggers their elimination by activating downstream processes in the immune response. Furthermore, when it occurs to functionally relevant regions, the binding can also neutralise the antigen directly. Antibodies are tetrameric proteins with a characteristic Y-shaped structure, consisting of two pairs of

Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. E-mail: ps589@cam.ac.uk, mv245@cam.ac.uk



Third generation antibody discovery methods: *in silico* rational design

Pietro Sormanni, 🕑 * Francesco A. Aprile ២ and Michele Vendruscolo 🕑 *

Owing to their outstanding performances in molecular recognition, antibodies are extensively used in research and applications in molecular biology, biotechnology and medicine. Recent advances in experimental and computational methods are making it possible to complement well-established *in vivo* (first generation) and *in vitro* (second generation) methods of antibody discovery with novel *in silico* (third generation) approaches. Here we describe the principles of computational antibody design and review the state of the art in this field. We then present Modular, a method that implements the rational design of antibodies in a modular manner, and describe the opportunities offered by this approach.

> heavy and light chains (Fig. 1). The tips of the Y are formed by the variable domains of the heavy (VH) and light (VL) chains, and contain the binding regions (paratopes). The paratopes are usually located within six binding loops, three per variable domain, which are referred to as the complementarity determining regions (CDRs), as they mediate the interaction with the target regions of the antigens (epitopes). The base of the antibody, called the crystallisable fragment (Fc) region, regulates the communication with other parts of the immune system, and it contains conserved glycosylation sites that modulate these interactions. By exploiting sophisticated processes of genetic recombination followed by somatic hypermutation and clonal selection, the adaptive immune system generates novel antibodies against new antigens to great effect and staggering



Pietro Sormanni

Dr Pietro Sormanni is a Borysiewicz Fellow at the University of Cambridge. His research focuses on the development of innovative data-driven technologies of rational antibody design, to obtain antibodies against targets that have been challenging to access using conventional approaches. At the Centre for Misfolding Diseases, he is applying these technologies to generate novel opportunities for research, diagnostics, and eventually treatment of neuro-

degenerative disorders, such as Alzheimer's and Parkinson's diseases. Prior to pursuing his postdoctoral research, he obtained a PhD in Chemistry from the University of Cambridge, and an MSc in Theoretical Physics from the University of Milan.



Francesco A. Aprile

Dr Francesco A. Aprile is a Senior Research Fellow of the Alzheimer's Society at the Centre of Misfolding Diseases, University of Cambridge, where he develops antibodies as research and diagnostic tools and as potential therapeutics against neurodegenerative diseases. Dr Aprile completed his PhD in Biology at the University of Milano-Bicocca (Italy), where he studied the mechanisms of formation and toxicity of amyloid aggregates as well as the protec-

tive role of molecular chaperones in neurodegeneration. He was awarded the 2017 Early Career Outstanding Contribution to Dementia Research by the Alzheimer's Society.

Published on 09 October 2018. Downloaded by University of Cambridge on 12/18/2018 7:55:08 AM.



Fig. 1 Structures of an antibody and of common antibody fragments. An antibody structure is shown on the left with the heavy (H) and light (L) chains in blue and green, respectively. Complementarity determining regions (CDRs) containing the paratopes are coloured in red, and the heavy and light variable domains (VH and VL) are labelled. The antigen-binding fragment (Fab) region is responsible for recognising the target, while the crystallisable fragment (Fc) region for immune function and lysosome escape. Two commonly used fragment formats are shown: a Fab fragment, which retains monovalent antigen binding while lacking the effector function, and a single-domain antibody, which corresponds to an autonomously folding VH domain. The three CDR loops are highlighted in red on this domain.

diversity, thus making antibodies the most versatile among the currently known classes of binding molecules.^{1–3}

Because of their ability to bind a large variety of molecular targets with high affinity and specificity, antibodies have been the focus of a wide range of technological developments aimed at isolating, producing, and optimising these molecules for specific targets of interest. Antibodies have thus become key tools in research and diagnostics, and represent the fastest-growing class of biological therapeutics on the market. The therapeutic antibody market is projected to steadily rise for the next several years, and it is expected to reach \$125 billions by $2020.^{4-6}$

Different fragments derived from the immunoglobulin structure retain binding and biological activity (Fig. 1). For example,



Michele Vendruscolo

Prof. Michele Vendruscolo is the co-director of the Centre of Misfolding Diseases (University of Cambridge). He obtained a PhD in Condensed Matter Physics at the International School for Advanced Studies (Trieste, Italy) in 1996, and since 2010 he is Professor of Biophysics at the Department of Chemistry, University of Cambridge. He was awarded the Soft Matter & Biophysical Chemistry Award by the Royal

Society of Chemistry (UK) in 2013, and the Giuseppe Occhialini Prize jointly by the Institute of Physics (UK) and the Italian Physical Society (Italy) in 2017 for his research on the fundamental principles of protein aggregation, solubility and homeostasis. Fab fragments retain the antigen-binding function while lacking the Fc-mediated effector function, single-chain antibodies (scFv) consist in the VH and VL domains connected by a linker, and single-domain antibodies (known as sdAb, V_{HH} or nanobodies) have been isolated from camelids and some species of sharks, as well as engineered from human VHs.⁷⁻⁹ These smaller antibody fragments have gained momentum for some applications, as they have thermodynamic stabilities comparable to those of full-length antibodies but can usually be expressed in prokaryotic systems. In particular, their small size may enable better target engagement by allowing for deeper penetration into tissues and active sites. The Fab and scFv formats, for example, are widely used in the recombinant display and engineering technologies that are extensively applied to isolate new antibodies.¹⁰

Third generation antibody discovery methods

Well-established procedures for the discovery of novel antibodies can be broadly classified into in vivo and in vitro approaches, which are often used in synergy. In vivo (first generation) approaches harness the power of an immune system to generate the required antibodies, traditionally through animal immunisation, and more recently also from human patients.^{11,12} By contrast, *in vitro* (second generation) approaches rely on the laboratory construction of large libraries of antibody sequences mimicking the diversity achieved by the immune system, and thus likely to contain some binding molecules for each given antigen. In both cases extensive laboratory screenings or deep-sequencing and advanced analysis¹³ are needed to isolate those antibodies actually binding to the intended target. Synthetic libraries can be built from human DNA sequences, and can now routinely generate antibodies with binding affinities down to the pM range, and for some targets even fM.¹⁴ However, as these antibodies do not undergo the in vivo selection process carried out by the immune system, they often possess poorer biophysical properties than those

derived from mammalian platforms.¹⁵ Conversely, antibodies of non-human origins may elicit adverse immune responses when used as therapeutics.¹⁶ Much effort is being devoted to overcome these limitations, for example advances in *in vivo* DNA manipulation have been exploited to genetically engineer mice able to produce human antibodies, which can thus be obtained through immunisation.^{17–20} The wide range of established technologies of these types that are available to scientists for the discovery of novel antibodies has recently been extensively reviewed.^{21–23} Here we focus instead on the emerging field of *in silico* (third generation) discovery methods.

Novel therapeutic applications push the boundaries of antibody discovery

Despite the great success of existing antibody discovery methods, some classes of hard targets still remain, including some membrane proteins, proteins within highly homologous families, epitopes that agonize or antagonize a biological pathway, aggregation-prone peptides, and disease-related short-lived protein aggregates. Although the diversity of the human antibody repertoire has been estimated to a staggering $\sim 10^{11}$ unique sequences,²⁴ it presents certain limitations, such as biases and redundancy in representing the mutational and conformational space available in principle to antibodies.²⁵ In addition, in vitro approaches that rely on the laboratory screening of large libraries can be time consuming and costly,²⁶ and may become particularly challenging when one is interested in targeting a specific epitope within the target antigen.²⁷⁻²⁹ In fact, these methods usually select for the tightest binders, which typically occur for immunodominant epitopes, thus precluding the discovery of antibodies with lower affinities but binding to functionally relevant sites. Finally, the lessons learned from these approaches do not easily generalise to inform the development of other systems, and screening campaigns sometimes yield antibodies with favourable binding but otherwise poor expression yield, stability, or solubility, which may hinder many applications.^{30–32}

In addition to its biological activity, another key property of a promising therapeutic antibody candidate is the likelihood of its successful development into a stable, safe, and effective drug, which is known as 'developability'.³³ 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 >1 year).³¹ 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.^{32,34,35}

The exquisite binding specificity of antibodies may actually create practical issues for some therapeutic programs due to the lack of species cross-reactivity. This property refers to the ability of an antibody to bind and functionally interact with the orthologous proteins from different animal species that are used as models for the evaluation of *in vivo* efficacy,

pharmacokinetic, and safety. The limitations in the study of therapeutic antibody candidates in rodent models imply that these properties must often be assessed in costlier and lengthier studies involving primates, thus slowing down progress, and increasing development costs and risk of late-stage failures.³⁶ A common strategy to overcome these problems is to screen for antibody candidates that functionally interact with both the human target and its relevant orthologs, which however complicates screening and lead selection. An appealing alternative is to carry out a simple bioinformatics analysis to identify epitopes conserved among the orthologs of interest, and raise antibodies specific for these epitopes, but this route is often highly challenging with existing technologies. The ability to target conserved epitopes is also of great relevance to obtain broadly neutralizing antibodies, which are able to inactivate a wide spectrum of genetic variants of a given pathogen (usually a virus).28

In order to function effectively in the crowded cellular or extracellular environments, antibodies should exhibit a balance of different biophysical traits, including stability, solubility, interaction affinity and selectivity. However, the stringent requirements of therapeutic applications imply that in most cases all these traits must be optimised far beyond their typical natural levels. This task is highly problematic, as these traits are often conflicting from an evolutionary perspective, in the sense that mutations that improve one of them tend to worsen the others.^{37,38} Therefore, there is a need to develop fundamentally new approaches to enable the simultaneous optimisation of all these traits.

Computational approaches offer a promising avenue to generate such technologies, as they could drastically reduce time and costs of antibody discovery, and in principle allow for a highly controlled parallel screening of multiple biophysical properties. Moreover, rational design inherently allows targeting specific epitopes of interest, which can be particularly daunting using available techniques. It is not surprising therefore that antibody engineers have long sought to design customised antibodies from scratch.³⁹

There is plenty of room for improving the biophysical properties of antibodies

The random mutations that are the engine of molecular evolution, on average, tend to destabilise proteins and to decrease their solubility, but only those few that instead improve these properties are typically selected.^{40,41} However, once natural selection has made these molecules stable and soluble enough to function optimally at their physiological concentrations and conditions, there is no further evolutionary pressure to improve their stability and solubility. The balance of the contrasting forces of random mutational drift and natural selection has effectively placed proteins and antibodies on an edge of biophysical fitness. It is increasingly recognized that natural proteins are just stable and soluble enough to optimally perform their function at physiological conditions,42 but with little or no safety margin besides the buffering effect provided by molecular chaperones and other components of the protein homeostasis system.40 This argument has recently been

corroborated by a range of proteomic observations, including the fact that *in vivo* thermal stability⁴³ and solubility⁴⁴ strongly correlate with protein abundance. Highly abundant proteins must be more soluble than their low-abundance counterparts to avoid aggregation, yet if all natural proteins had evolved for maximum solubility no correlation with abundance would be expected.⁴⁵ Similarly, highly expressed proteins are more exposed to random translational errors, which if not welltolerated by means of high conformational stability can lead to the accumulation of toxic misfolded species.46 Taken together, these observations imply that there is room to optimize the biophysical fitness of proteins and antibodies far beyond their natural levels, as also indicated by the many successes of in vitro directed evolution, which for example has been applied to improve antibody binding-affinities by several order of magnitudes over typical natural values.¹⁴

The existence of this large biophysical space left unexplored by evolution is a crucial requirement for the successful development of effective methods of rational antibody design aimed at producing optimal antibodies for biomedical applications.

Quantitative parameters for the systematic antibody optimisation

Antibody design refers to methods aimed at producing antibodies with specific properties. These properties include solubility, stability, effector functions, and binding affinity and specificity, which are closely intertwined and usually determine other essential properties, such as biological activity, *in vivo* half-life, protease susceptibility, solution viscosity, chemical stability, pharmacokinetics and in some instances also immunogenicity.^{47–50} In this section we review the state of the art in the rational design of these properties.

Solubility

In thermodynamics the solubility of a substance is a property defined as the specific value of the concentration - termed the critical concentration - at which the soluble and insoluble phases are in equilibrium. While this definition is rigorous, it only applies directly to substances that have well-defined soluble and insoluble phases. However, depending on the concentration, the vast majority of proteins can populate a variety of states, including monomers, dimers, small and large oligomers, amorphous aggregates and sometimes fibrils. Thus, the boundary between soluble and insoluble species is ultimately arbitrary and operationally dependent on the method used to separate them, for example on the centrifugation speed or filter size, which hampers the possibility of defining and therefore measuring absolute solubility values. However, it is possible to measure solubility differences among different antibodies, or at least differences in their propensity to self-associate, precipitate, or populate aggregated states, which are common proxies for solubility across the literature.

Solubility differences can be measured both *in vivo* and *in vitro*. *In vivo* measurements are generally used as preliminary

screenings and include protein yield determination and analysis of the composition of the inclusion bodies from expressing bacterial strains. Recently, new and more quantitative methods based on automated immune-detection,⁵¹ split beta-galactosidase, and split GFP,⁵² have been developed. PEG-induced precipitation is for example a common way to estimate protein solubility in vitro. In this set up, the antibody is incubated in the presence of increasing concentrations of a precipitant, such as PEG or ammonium sulphate, and the amount of precipitated protein is estimated by measurements of turbidity or retained protein concentration after filtering.⁵³ The apparent solubility may then be extrapolated from the data points in the linear range of the precipitation sigmoid,⁵⁴ or, if the variants under scrutiny have similar sigmoid slope, it is possible to use PEG_{1/2} as a solubility proxy (i.e. the concentration of PEG at which the antibody is 50% precipitated), which can be estimated with greater accuracy.³⁰ Other common measurements of solubility differences include quantifications of the monomeric population carried out with analytical size exclusion,⁵⁵ or aggregation rates from measurements of turbidity, and dynamic or static light scattering. However, measurements of solubility differences are not comparable across different experimental conditions or assays, thus limiting their applicability to train or test methods of solubility prediction and design.

Given that insufficient solubility is often the main reason behind the high development costs of antibodies,^{31,56} and that low solution viscosity is a key requirement for subcutaneous injection, there in a strong pressure to carry out solubilityrelated screenings as early as possible in antibody development pipelines in order to avoid expensive late-stage failures. Therefore, in vitro developability assays are increasingly employed to predict the clinical success of antibodies, by screening for desirable biophysical properties in the early development phase, when samples are usually available in limited amount and low concentration and purity. These approaches typically measure parameters considered to be predictive of solubility or solution viscosity, such as reversible self-association, or non-specific interaction with different materials. Examples include crossinteraction chromatography (CIC), standup monolayer adsorption chromatography (SMAC), hydrophobic interaction chromatography (HIC), affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS), as well as emerging microfluidic-based approaches,^{57–59} which have all been shown to correlate with varying extent with viscosity, solubility and sometimes in vivo clearance.32,33

Conformational stability

The stability of a protein is defined as the free energy difference between the native state and the unfolded state. In practical terms, the stability of an antibody translates into its ability to retain its structural conformation and activity when subjected to physical or chemical stress (*e.g.* temperature, pH and denaturants). Antibody stability is generally evaluated *in vitro* by means of biophysical approaches based on spectroscopic or calorimetric measurements. In particular, one can use steady-state or time-resolved circular dichroism or intrinsic

fluorescence, and differential scanning calorimetry or fluorimetry (DSC and DSF respectively), to accurately monitor the denaturation of an antibody over time, a range of temperatures, or different concentrations of denaturants (e.g. urea or guanidinium). In this way, one can determine the kinetics of folding and unfolding, and thermodynamic stability of a protein either as the melting temperature $(T_{\rm m})$ or the free energy difference (ΔG) between the folded and unfolded states, or between the folded and the aggregated states. The vast majority of antibodies generated with existing discovery technologies are typically stable enough to be fully active at physiologically-relevant temperatures and pH values. However, it is highly desirable to select or design extremely stable antibodies. In fact, high conformational stability is an important determinant of heterologous-expression yield, proteolytic resistance, and *in vivo* half-life.^{43,60-62} Furthermore, a higher native-state stability corresponds at equilibrium to decreased population of partially of fully unfolded states, whose presence may elicit protein-aggregation at highly concentrated storage condition.

Binding affinity

Binding affinity refers to the strength of antigen binding, and it is typically reported as the dissociation constant K_d , which is defined as the product of the concentrations of antibody and ligand divided by the concentration of their complex. K_d is therefore measured in units of concentration, with smaller values denoting stronger binding.

Various techniques have been developed to determine binding affinities and kinetic constants (on- and off-rates). Experimental setups generally consist of titration experiments or kinetic experiments where the readout is proportional to the amount of formed complex. The most used techniques include nuclear magnetic resonance (NMR) chemical shifts, isothermal calorimetry,⁶³ surface plasmon resonance (SPR),⁶⁴ microscale thermophoresis (MST),65 and various types of spectroscopic measurements broadly relying on the use of fluorescently labelled proteins or ligands (i.e. fluorescence intensity, polarisation, or fluorescence resonance energy transfer). In the context of affinity maturation/optimisation protocols, high-throughput (but less quantitative) techniques such as enzyme-linked immunosorbent assay (ELISA) and immune-precipitation assays or cell two-hybrid screens are routinely used⁶⁶ Recently, more quantitative techniques such as bio-layer interferometry (BLI),⁶⁷ and micro-diffusion measurements^{68,69} have been proved to be particularly suitable for binding screenings.

Binding specificity

Binding specificity is generally understood as the relative weakness of off-target interactions. The fewer ligands an antibody can bind to, the greater its specificity. Protein arrays currently represent the gold standard to probe the specificity of an antibody, as they can provide a direct readout of which and how many off-target interactions exist. Following this approach, one incubates a chip containing an array of several hundred immobilised proteins in the presence of a fluorescently labelled variant of the antibody of interest.⁷⁰

Fluorescence emissions at each position of the array are then used to determine the amount of antibody that is bound to the corresponding immobilised protein. Alternative methods to probe specificity are for example SPR or BLI measurements with the immobilised antibody of interest against a cell-lysate or other relevant protein mixtures in the absence of the target protein, where the expectation for highly specific antibodies is to observe no signal. Ultimately, most antibody development campaign for therapeutic applications rely directly on *in vivo* target engagement and absence of side effects in animal models and eventually in patients as a comprehensive readout of specificity and pharmacokinetics.

The trade-off between properties is the greatest challenge for rational design

There is a delicate balance between antibody stability, solubility and binding affinity and specificity, as mutations that improve one of these traits often negatively impact the others. For example, increasing the conformational stability of a protein typically increases its rigidity, thus affecting its functionality, and vice versa.⁷¹ For antibodies, more rigid scaffolds and CDR loops that are pre-disposed in the bound conformation in the absence of the antigen have been associated with better affinities, as there should be less entropy loss upon binding.^{72,73} Yet, sometimes significant conformational rearrangements occur upon binding, probably because the antibody requires a different conformation to be stable in the absence of the antigen,⁷⁴ and the relationship between paratope rigidification and affinity maturation remains controversial.75 Similarly, amino acid substitutions that improve conformational stability have been seen to negatively impact solubility,⁷⁶ while these two properties have often been reported to correlate.77-79 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.^{80–82} 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.^{30,83,84} However, at least for nanobodies, the length of the CDR3 loop was recently shown to positively correlate with thermal stability, thus challenging this view.⁸⁵ 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.⁸⁶ 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.^{87,88} Interestingly, Tyr, Ser, and Trp are prevalent in antigen-contacting residues in germline antibodies, but not among mutations introduced by the somatic hypermutation mechanism, thus suggesting a role of this process in suppressing non-specific interactions.⁸⁹ 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.⁹⁰

Taken together, these apparently contrasting findings, and the existence of many cases contradicting the expected average behaviour, show that it is very difficult 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 specify appear to be highly context-dependent, and rather specific to individual antibody–antigen complexes. Therefore, the accurate modelling of this delicate equilibrium is probably the greatest challenge that rational antibody design has to face in order to offer practically useful antibodies.

General approaches to protein design

Since the pioneering work of Regan, DeGrado and co-workers in the late '80s,^{91,92} a multitude of different methods to rationally modify protein molecules have been proposed. However, it is only in the last few years that reports of successful design of proteins with practical applications and uses have started to emerge.⁹³ Quite generally, protein-design methods can be classified into four broadly defined groups, with the most successful design strategies typically combining approaches from different groups. While this section refers to protein design in general, the approaches described are all applicable to antibodies.

Energy-based methods

The largest class of protein design methods are those that directly exploit the laws of statistical thermodynamics, according to which a protein is preferentially found in the conformation corresponding to its free energy minimum. To date, the majority of successfully applied computational design methods have exploited this principle, and have therefore implemented modelling strategies aimed at directly maximising the free energy difference between the target state, such as the native state or a bound state, and all other possible states.⁹⁴

In principle achieving this goal requires accurate macromolecular representation, correct modelling of the interactions among the many atoms comprising proteins, solvent, and ligands, and exhaustive sampling of the conformational ensemble to estimate the free-energy landscape. All of these steps should be repeated for each tested design. In practice, however, accurate ensemble determination remains daunting even for small proteins for which high-quality experimental data are available to guide computer simulations,95,96 and methods of proteindesign have the additional complication that they must also sample the mutational space, for example by trying as many mutations as possible to optimize the traits of interest. Therefore, in most energy-based approaches the sampling of the conformational space is either not performed or it is limited to small fluctuations about the target structure, which are used to adjust the backbone to accommodate different side-chains from candidate mutations.^{97,98} Consequently, rather than calculating the free-energy landscape of the system, most energy-based

approaches in practice calculate the energy of the target structure (*i.e.* one calculates the energy *E*, rather than the free energy *E-TS*, where *S* is the entropy and *T* is the temperature). The goal then becomes to minimise this energy to the point that all other alternative conformations are likely to be higher in free energy, at least within a relevant range of temperatures and protein concentrations. Sometimes these approaches are combined with elements of negative design, which are particularly relevant for the generation of specific interactions (e.g. substrate recognition or interface assembly).99,100 These elements of negative design may consist in known alternative conformations that are explicitly considered in the simulations so that they are energetically disfavoured,^{101–103} or heuristics, whose occurrence is known or suspected to bring about undesired molecular traits that may give rise to alternative free-energy minima (e.g. aggregation-prone segments or specific structural motifs).^{104,105} Overall, a common feature of all systems that have been successfully designed with energy-based approaches is that the actual free-energy difference between the target state and all other possible states is larger than the error introduced by the approximations and assumptions employed.^{93,94} Energy-based methods, especially when combined with appropriate experimental validations,⁹³ and more recently with high-throughput manufacturing and screening technologies,^{60,106} have revealed many fundamental features of protein structure, stability and interactions.

Combinatorial design methods

Combinatorial design methods directly mimic some combinatorial aspects of natural evolution. It is increasingly recognised that proteins can evolve by repurposing existing folds or fold fragments.¹⁰⁷ This evolutionary mechanism can be imitated to rationally design new proteins or new functionalities through the combination of fragments from existing proteins.¹⁰⁸ These combinatorial methods are not necessarily affected by the accuracy of energy functions, and they usually only require relatively small-scale sampling of the different combinations of the available fragments, and therefore are not particularly computationally demanding. Conversely, these methods are less general than their energy-based counterparts, as rather than on general physical laws they rely on custom rules that determine how and which fragments should be combined. This set of rules and the database of fragments employed directly affect the reliability of a combinatorial method and may ultimately limit its applicability. For example, a method recently developed in our laboratory to rationally design antibodies binding to specific epitopes within disordered regions of target proteins is based on combinatorial design.¹⁰⁹ In this approach, which will be explained in more detail below, protein fragments found to interact with components of the target epitope within native structures of proteins are combined together with a set of rules to build antibody paratopes.¹⁰⁹

Empirical methods

Empirical methods of protein design directly exploit efforts of rationalising biological observations or measurements, by implementing empirical functions that recapitulate these observations. The most common examples are knowledge-based potentials, which are statistical force-fields derived from databases of known protein properties that empirically capture aspects of the physical chemistry of protein structure and function.^{110,111} These potentials have played important roles in protein design and structure prediction, by improving the accuracy of energy-based models of interatomic interactions, as well as the computational efficiency of the design process.^{112–114} Another area where empirical methods have been widely successful is the rationalisation and prediction of the aggregation propensity of peptides^{115,116} and proteins.^{117–119} For example, the CamSol method, which we have developed in our laboratory to predict and rationally design antibody solubility, is a method of this type.^{30,55}

Machine-learning-based methods

The growing amount of available biological data, including macromolecular structures and measurements of biophysical properties, is increasingly enabling applications of machinelearning algorithms to address biological problems. Machine learning, sometimes referred to as "artificial intelligence", refers to a broad class of computer algorithms that have the ability to "learn" from data, without being explicitly programmed. The learning or training phase typically consists in the optimisation of a large number of free parameters, so that from known inputs (e.g. protein sequences) the algorithm reproduces the corresponding known outputs (e.g. the corresponding native structures). When properly trained on a largeenough amount of diverse data, these algorithms have the ability to generalise, that is to correctly predict the outputs of never-seen-before inputs. Following pioneering applications of these algorithms to the prediction of protein secondarystructure,120,121 machine-learning has been widely employed in protein science. Its applications include predictions of conformational dynamics,^{122,123} protein tertiary structure,¹²⁴ and binding sites,¹²⁵ including those of antibodies,^{126,127} while direct applications to protein design are beginning to emerge.^{128,129} Although machine-learning-based methods lack the physical transparency of other approaches, their frank pragmatism is remarkably successful. Therefore, given that the amount of available "training-data" across biological databases is rapidly increasing, and that machine-learning algorithms are constantly improving,¹³⁰ these methods are destined to play key roles in shaping the future of protein design.

De novo antibody design

In this section we review ongoing efforts to design antibodies *de novo*, considering those methods for which some extent of experimental validation has been presented. Furthermore, we discuss the approaches that can, at least in principle, take as input only an epitope within an antigen, and predict one or more antibody sequences capable of binding it. We therefore do not cover studies aimed at developing methods of antibody docking and of *in silico* affinity maturation.^{131–134}

Structure-based methods of antibody design

Structural modelling of antibodies. Structure-based methods of antibody design (Fig. 2) involve the modelling of the structure of the antibody, as well as that of the antibody–antigen complex. The quality of this modelling is a key determinant of the ability of these methods to correctly predict novel interactions.

Owing to the high degree of structural homology between variable domains from different antibodies, and in particular to the fact that most CDR loops can be clustered into canonical conformations,¹³⁵⁻¹³⁷ the vast majority of the Fv region can now be modelled reliably and with high accuracy.^{131,138} None-theless, it remains challenging to obtain accurate models of the conformations of the heavy-chain CDR3 loop (CDR-H3), and also of the relative orientation between the VH and VL domains, which are often the two most important features in determining binding.¹³⁹⁻¹⁴¹ Additionally, as the CDR-H3 is often quite dynamical, its solution conformations can differ significantly from those observed in crystal structures, which may be stabilized by crystal packing effects, and may also change upon binding,^{74,142} thus further complicating both the modelling and the assessments of modelling quality.¹⁴³

Then, even when an accurate model is obtained, correctly docking the antibody and the antigen remains an arduous task. For example, for protein–protein or protein–peptide interactions, in a recent round of Community-Wide Critical Assessment of Protein Interactions (CAPRI) there was not a single model, among the 20 670 submitted, that ranked as 'acceptable' in its quality (*i.e.* identifying correctly at least 50% of the interface contacts) for six out of 20 complexes assessed.¹⁴⁴ Similarly, a recent analysis of antibody docking carried out on 17 complexes, found that when using modelled antibody structures as a starting point, the correct antibody–antigen pair was the top-ranking model in only two cases.¹⁴⁵

Taken together, these findings suggest that successful efforts of designing novel antibodies that depend the structural modelling of the complexes with their targets cannot rely on one model only. Consequently, hundreds, and sometimes thousands of different models for different sequences should be built *in silico*. These models are typically further refined computationally by carrying out point mutations predicted to improve the interaction energy. This optimization step, however, has usually limited performance, as highlighted by a recent study comparing the predicted and observed effect of mutations on antibody binding affinity.¹⁴⁶ Finally, a high number of diverse, top-scoring models is typically taken forward for experimental validation hoping that at least one binds the target.

OptMAVEn and OptCDR. Maranas and colleagues have developed the OptCDR method¹⁴⁷ and the OptMAVEn method,¹⁴⁸ which extends the applicability of the earlier OptCDR approach to the design of the whole Fv region. A recent version of the OptMAVEn design pipeline has been applied to design scFvs binding to a hydrophobic heptameric peptide with a repetitive sequence (FYPYPYA).¹⁴⁹ Similarly, OptCDR managed to yield scFvs binding to the FLAG-tag tetrapeptide.¹⁵⁰



Fig. 2 Structure-based antibody design. Schematic representations of the key steps most commonly employed in approaches of structure-based *de novo* antibody design. Step 0: pre-computation stage, consisting in the compilation of a fragment library and, if employed later, the associated sequence constraints. Step 1: the target epitope is selected and positioned in a simulation box, if the target epitope is a dynamical peptide, multiple conformations are generated and each one is used as a starting point for the next step. Step 2: various combinations of fragments from step 0 are sampled ('germline' models) and docked against the target epitope, each combination is scored with an energy and/or geometrical function. Step 3: the best-ranking models from the previous step are subjected to sequence optimization by carrying out point mutations (blue to purple side-chains) predicted to improve binding energy and/or total energy (*i.e.* antibody stability).

OptCDR generates thousands of random antigen positions,¹⁵⁰ and for each one it generates the backbone of the binding loops by sampling from clusters of canonical CDR conformations using a geometric score.¹⁴⁷ Sidechains are then placed according to (i) sequence preferences within each CDR cluster, (ii) a rotamer search from a backbone-dependent rotamer library, and (iii) a CHARMM-based energy function.^{147,150} Finally, an *in silico* affinity maturation is carried out to introduce point mutations predicted to improve binding affinity.¹⁵¹

OptMAVEn consists in a combinatorial-design procedure that essentially reproduces the immune-system V(D)J recombination, followed by an energy-based in silico affinity maturation inspired by the somatic hypermutation mechanism. The OptMAVEn pipeline can be split into four steps.^{148,149} The first step is the sampling of the antigen positions into a predefined antibody binding site, which is defined as a rectangular box covering all mean epitope coordinates determined from the analysis of 750 antibody-antigen complexes.¹⁴⁹ Antigen-positioning is followed by the assignment of the best V, (D), and J antibody modular parts selected from the MAPs database,¹⁵² thus leading to the assembly of germline-antibody models for the target of interest. The third step consists in all-atom, explicit-solvent, 100 ns molecular dynamics simulations carried out for each model. These simulations are used to rule out those models in which the antigen dissociates rapidly, and to refine the antibody-antigen interface of the remaining models. The fourth and final step is an in silico affinity maturation, during which mutations in the antibody are selected to improve the binding energy. In this step, mutations are

also attempted outside antigen-contacting sites, as these may still mediate antigen interactions and contribute to the stability of the complex.

OptMAVEn was used to design five scFv antibodies against the peptide FYPYPYA, starting from the structure of an existing antibody bound to a dodecameric peptide containing this sequence (scFv-2D10, PDB 4hoh). All five scFvs had designed sequences reasonably different from that of the parental scFv-2D10, and three of the designs successfully bound the target. The K_d-s of these antibodies were in the low/mid nM range, all slightly higher than the K_d of scFv-2D10, which remained the best binder.¹⁴⁹ OptCDR was instead used to design 50 scFvs predicted to bind to the minimal FLAG peptide (DYKD). Of these about half resulted in detectable scFv expression, and four were able to bind to the FLAG sequence.¹⁵⁰ Even in these successful cases, the melting temperature of the designed antibodies in Fab format was about 10 °C lower than that of the parental Fab scaffold.¹⁵⁰ These findings highlight the challenges of designing antibody affinity and specificity while not compromising other essential biophysical properties.

AbDesign. Fleishman and coworkers have introduced the AbDesign algorithm,¹⁵³ which follows an approach similar to OptMAVEn combining V regions and CDR3 regions by analogy to V(D)J recombination. AbDesign segments the antibody backbone using junctions of high structure conservation (*e.g.* the disulphide-bonded cysteines and conserved positions at the end of CDR3), as opposed to genomic-recombination sites,

which may help generate structurally compatible framework-CDR interactions. Also, the backbone fragments employed are extracted from all antibodies in the structure database, including highly homologous ones, thus increasing the potential coverage. However, because the CDR1 and CDR2 are grafted together in the latest implementation of the algorithm,¹⁵⁴ AbDesign has limited flexibility in terms of setting which CDRs to design and what CDR lengths or combination of conformations to sample.

Specifically, a set of antibody conformations representing all canonical conformations for a chosen starting Fv scaffold is generated from these backbone fragments, docked against the target surface, and designed for optimal binding affinity. As implemented in the Rosetta Software Suite, the sequence design is carried out from a position-specific scoring matrix (PSSM) of aligned antibody sequences. This PSSM effectively constrains the mutational space of each residue position in the Fv scaffold to amino acids that are actually found at that site in natural antibodies. Finally, a filter is applied using a fuzzy-logic design that aims at finding the optimal compromise between antibody binding and antibody stability.¹⁵⁵

AbDesign was recently employed to generate novel antibodies against insulin and mycobacterial acyl-carrier protein (ACP).¹⁵⁴ In this endeavour five design/experiment cycles were performed, leading to several improvements in the algorithm. In particular, to prevent the emergence of destabilizing features during the design, the backbone was segmented in only two parts per chain (the CDR3 and the rest of the variable domain), and conformation-dependent sequence constraints (PSSMs) were used to guide Rosetta design choices. Throughout the five design/experiment cycles, 114 insulin-targeting designs were custom-synthetized and tested, from which one insulin binder was identified, although cycle 5 consistently produced highexpression antibodies. Similarly, two binders were identified among the 79 designs tested for ACP. All the three identified binders required manually-inserted mutations to improve binding and stability, as 32 and 5 mutations for the ACP and 6 for the insulin scFvs were inserted with FoldIt and/or to improve charge complementarity with the antigen.¹⁵⁴ Finally, in vitro affinity maturation was carried out to increase the binders' affinity for their designated targets, by using errorprone PCR (one to four random mutations in the scFv-coding gene) followed by yeast display. Affinity in isolated clones increased by approximately an order of magnitude thus reaching the mid nM range for two of the three designed antibodies.¹⁵⁴

Rosetta antibody design (RAbD). Dunbrack and colleagues have recently presented the RosettaAntibodyDesign method, which like AbDesign employs the Rosetta Modelling suite to carry out a structural-bioinformatic-based combinatorial design.¹⁵⁶ RAbD exploits the PyIgClassify database,¹³⁵ which contains CDR sequences and cluster identifications for all antibodies in the Protein Data Bank (PDB) to obtain antibody fragments. These fragments are then grafted onto an antibody framework and sequence design is performed according to PSSMs of each cluster. CDR backbones are sampled with a flexible-backbone design protocol that includes cluster-based CDR constraints. RAbD can handle both λ and κ light chains, and, in addition to the six canonical CDRs, it can also engineer the DE loop, sometimes called L4, which is a short loop between strands D and E of the VL domain.

While RAbD could in principle be applied to design antibodies *de novo* starting from a modelled structure, the currently available experimental validation consisted in an *in silico* affinity maturation performed starting from two experimentally-determined antibody–antigen structures.¹⁵⁶ Of the 30 designs generated from an antibody that binds to hyaluronidase with a 9.2 nM K_d (PDB: 2j88), 20 retained some degree of binding with affinities at least in the μ M range. Of those 20, 3 exhibited a binding affinity better than that of the parent antibody, with the best design yielding a 12-fold improvement. Similarly, of the 27 designs made from the antibody CH103 that binds to HIV gp120 (PDB 4jan), 21 could be purified and tested, and 6 could bind one or more of seven gp120 variants from different strains of HIV. One designed antibody had an improved binding affinity to some of the gp120 variants tested.

Overall, as RAbD samples CDR fragments rather than whole Fv regions and requires an antibody–antigen structure or model as a starting point, its present implementation could be expected to be particularly suitable for affinity maturation or redesign applications, for example to target homolog antigens that can be modelled onto existing antibody–antigen structures.

Discussion on the state-of-the-art of structure-based methods. The structure-based antibody-design methods discussed above combine combinatorial design of antibody fragments and energy-based design. The energy function is used to rank the various combinations of fragments, sometimes together with empirical scores,¹⁴⁷ and also to design affinity-improving mutations along the backbone, often guided by PSSMs.^{148,153,156}

Typically in these methods a high number of designs is generated (for example RAbD recommend between 1000 and 10000)¹⁵⁶ to then select the best candidates. In addition, one should consider the vast size of the combinatorial fragment space and associated mutational space that must be sampled during the design process. In particular, those methods that model the whole Fv region rather than just the CDR loops,^{148,153} and that involve all-atom explicit solvent molecular dynamics simulations for all generated candidates,¹⁴⁸ should be particularly resource-demanding. Overall, therefore, the gain in time and resources offered by these in silico approaches over standard laboratory methods may not be substantial at present, especially in those cases where the *de novo* designs require additional optimisation before becoming practically useful (e.g. in vitro directed evolution).¹⁵⁴ Nevertheless, all these methods in principle enable to target pre-selected epitopes of choice, which can often be daunting with purely experimental methods. Furthermore, as computers become more powerful, energy functions improve, and fragment libraries grow in size and diversity, the competitiveness of these computer-based technologies could be expected to increase.

The successes obtained by these approaches demonstrate that it is now possible to implement methods of *de novo* antibody design able to yield novel antibodies binding to targets of interest. However, the low success rates, and the quality of most *de novo* designed antibodies, which were often found to be poorly stable, weak binders, or both, indicate that further developments are required for these methods to become competitive with traditional antibody-discovery technologies for most applications.

In particular, the five cycles of design and experiments carried out in the validation of AbDesign provided important insights into those features of antibodies that are difficult to model with this approach, and yet are crucial for antibody folding and expressibility.¹⁵⁴ Importantly, antibody-expressibility was improved throughout design/experiment cycles by gradually drifting away from energy-based design towards more conservative combinatorial approaches. Expression levels comparable to that of a natural scFv were obtained by employing conformationdependent sequence constraints to guide Rosetta design choices, thus greatly restricting the available mutational space, and only once the Fv backbone was segmented in merely two parts per chain: the whole framework region, comprising CDR1 and 2, and the CDR3 loop, which is well-known to be the most versatile loop in length, amino acid composition, and conformation.^{136,140,157,158} This finding, together with the relatively low success rates of the energy-based approaches discussed here, suggest that the combinatorial component of these structure-based strategies plays a key role in yielding functional antibodies. On the contrary, even the most advanced energy functions (e.g. Rosetta) are still unable to fully capture on their own the complex molecular trade-offs that control antibody binding and stability. Generally speaking, restricting the sampling space to pre-determined fragments and sequences is not a desirable feature, since it limits the possibility to sample sequences and conformations that are unlikely to be present in an animal immune system or in an antibody display library, thus decreasing the appeal of a computational approach. Nevertheless, the fact that these methods were able to produce diverse antibodies for unrelated targets demonstrate that the available sampling space is still large enough to provide novel antibodies, which were consistently found to be different in terms of both sequence and conformations from those present in the PDB.

Finally, both OptMAVEn and OptCDR were applied to generate novel antibodies targeting linear epitopes (i.e. short peptide), while AbDesign was used to design antibodies targeting conformational epitopes (i.e. non-sequential epitopes of welldefined 3D structure). These two types of epitopes are different and thus present different challenges for computational design. On the one hand, modelling antibodies to target conformational epitopes may be regarded as more difficult. The paratope shape tends to be convex, and antibody loops and side-chains must be accurately built around relatively rigid molecular structures containing irregularities, and whose surface properties may be difficult to model (e.g. hydrophobicity and electrostatic potential). On the other hand, the many conformations that peptides populate in solution, entail that no well-defined antigen structure can be used as a starting point for the design, which often must be carried out starting from a wide range of possible conformations.¹⁵⁰ In addition, depending on the biophysical properties of the target peptide, there may be a large entropy

loss associated with its binding to the antibody. Such entropy loss, if present, must be compensated by an even larger enthalpy gain that requires accurate modelling of the antibody-peptide interactions, albeit the concave paratope shape allowed by short linear epitopes may facilitate this task.¹⁵⁹

Empirical approaches of antibody design

The growing amount of high-quality structural and biological information available on antibodies, and also on protein interactions more generally, is increasingly being exploited to devise knowledge-based strategies to obtain antibodies for targets of interest. Pre-defined interaction patterns, which may be known from structural or biological (i.e. protein-protein interaction) data, can be grafted onto the CDRs of antibodies to produce lead antibodies binding to epitopes of interest. This strategy has been referred to as hotspot-centric design, and it has been applied to several reports of successful design of proteinprotein interactions,^{160,161} including recently introduced massively-parallel approaches.¹⁰⁶ These methods are not universally applicable, as their success strictly depends on the pre-existing knowledge of an interaction pattern for the target epitope. Moreover, for applications to antibodies, such patterns must be, or must contain a fragment, suitable for CDR grafting, and this interaction motif must retain enough binding affinity once grafted onto the antibody scaffold. Despite these limitations, strategies of this type have been successful in generating novel antibodies for challenging targets.

For example, a strategy of this type has been used for the structure-guided design of an anti-dengue antibody directed to a non-immunodominant, yet highly conserved, epitope within the E protein of the dengue virus.¹⁶² Starting from a known antibody with modest binding affinity (100 nM) an epitopeparatope connectivity network was computed to identify putative affinity-enhancing mutation sites, and a site-saturation combinatorial library was generated and screened with yeastdisplay. The resulting antibody retained the intended epitope, which it bound with high-affinity ($K_d < 0.1$ nM for most serotypes), and was able to broadly neutralize multiple genotypes within all four serotypes.¹⁶² While this example relied on the existence a known binding antibody used as a starting point, in another application of hotspot-centric antibody design, epitope-specific antibodies binding Keap1, an important component of the antioxidant response, have been obtained by grafting structural interaction patterns directly from Keap1 native binding partner, Nrf2, the master regulator of such response. Key binding residues were grafted onto geometrically matched positions of a set of antibody scaffolds.¹⁶³ Following in silico redesign of the CDR-H3 sequence, the affinity of one of these antibodies was taken from $K_d > 100$ nM to the low nM range, and the optimized antibody was able to effectively compete with the binding of Nrf2 itself.163

Empirical approaches have been particularly successful in generating antibodies binding to disease-related amyloidogenic proteins, both in their monomeric and aggregated forms. Following a proof-of-concept on the prion protein,¹⁶⁴ Tessier and co-workers have generalized this idea by introducing

the gammabodies (for Grafted AMyloid-Motif AntiBODIES). Gammabodies are human-derived single-domain antibodies¹⁶⁵ that contain in their CDR3 the sequence of an amyloidogenic motif (typically of about 10 residues), as found within disease-related amyloidogenic proteins or peptides.¹⁶⁶ Gammabodies carrying different amyloid-motifs have been shown to be effective inhibitors of the aggregation of their parent proteins in the case of α -synuclein, islet amyloid polypeptide (IAPP) and A β .¹⁶⁶ Moreover, A β -gammabodies were shown to be sequence- and conformation-specific depending on which amyloid motif was grafted in their CDR3.¹⁶⁷ Gammabodies have been optimised for improved solubility,^{168,169} subjected to *in vitro* directed-evolution to improve affinity and stability,^{83,84} and extended to the scFv format.⁸⁸

Overall, these empirical approaches that rely on pre-defined interaction motifs offer an expedite route for the generation of antibodies towards pre-determined epitopes. Their success rate is at the present stage considerably higher than that of other *de novo* design methods that rely on the minimisation of interaction energies. Nonetheless, they remain applicable only to cases for which at least one suitable pre-defined interaction motif is known.

Modular: a method of rational antibody design

The concept of modular design

Modular design, which separates the design of different properties into different modules, is particularly well suited to address the challenge of designing antibodies with a range of different biophysical properties (Fig. 3a), since these properties can require conflicting design principles, as for instance in the case of solubility and affinity.

To implement this strategy, we recently developed the Modular method, in which we combined together different modules to optimise individual molecular traits. In particular,



Fig. 3 Illustration of the rational design strategy of Modular. (a) The CamSol method (left) is used to optimise the solubility of the designed antibodies by applying it to engineer the stem of the CDR3, where the complementary peptides are grafted, and to rank the complementary peptides themselves (table, left column). The scatter plot shows an example of correlation between predicted (*x*-axis) and measured (*y*-axis) solubility.⁵⁵ The Cascade method (right) is used for the sequence-based design of complementary peptides. The example shows the target epitope of DesAb-F on the NAC region of α -synuclein in red, with dashed lines representing predicted hydrogen bonds with the complementary peptide (in green). The circles show the fragments comprising the complementary peptide (in green) interacting with fragments comprising the target epitope (in red) within different protein structures from the PDB database. The table in the middle contain four candidate complementary peptides for this epitope, together with two statistical scores from the Cascade procedure (complementarity scores to the left and mean enrichments to the right, see ref. 109). The peptide embodying the best compromise between predicted solubility and various statistical scores (see ref. 109) is then grafted onto the CDR3 of a stable antibody scaffold. (b) Micro-scale thermophoresis (MST) of fluorescently-labelled monomeric α -synuclein in the presence of increasing concentrations (*x*-axis) of DesAb-F (red) and a DesAb with two designed binding loops (blue). (c) Seeded aggregation assay (3% seeds) of 70 μ M α -synuclein alone (green) and in the presence of the DesAbs from (b) (see legend) at 1:2 DesAbs: synuclein-monomers ratio. Error bars are SEs over three replicates.

to design the antibody paratope we employ the Cascade method, as we discuss in detail in the next section. This method enables the design of linear motifs, called complementary peptides, which are predicted to bind to a target epitope and are grafted onto the CDR of an antibody scaffold. We then employ the CamSol method of predicting protein solubility to reduce the tendency of the antibody scaffold to self-associate, as well as to rank different designed paratopes to select highly soluble binding regions. The CamSol method was shown to be highly quantitative in recapitulating experimental measurements of solubility changes upon mutations,^{55,82,170} as well as in ranking monoclonal antibody libraries according to their solubility.³⁰

The optimisation of antibody solubility is a crucial requirement for a successful antibody design strategy, as antibodies perform their function in crowded molecular environments, and need to interact specifically with their target while avoiding non-specific interactions and self-association, in particular under storage conditions. As binding interfaces are on average enriched with aggregation-promoting patches,¹⁷¹ neglecting the design of solubility is likely to yield poorly soluble binding sites, which may favour antibody aggregation and compromise expression yields. In addition, enrichment of sticky hydrophobic residues on antibody paratopes has been associated with poor specificity.^{87,90} Therefore, controlling for the solubility of designed CDR may be an expedient to also reduce the emergence of non-specific interactions.

The grafting of an artificial sequence onto a CDR is expected to be a destabilizing feature of the designed antibody. There are many tools available to design stability-improving mutations that could be employed within the modular design process without requiring computationally expensive simulations.76-78,97 However, a systematic assessment in 2009 quantitatively compared predicted and experimentally-observed stability changes for 6 of such methods, finding a median Pearson's coefficient of correlation of R = 0.52, which indicates rather poor predictions at a quantitative level.¹⁷² A more recent (late 2017) similar analysis on 11 methods reported a median coefficient of R = 0.54.⁷⁶ Since over the same period of time the number of protein structures in the PDB has more than doubled, the marginal improvement in the performance of the methods highlights the challenges of accurately predicting stability changes with computationally-efficient structure-based approaches. Therefore, to bypass this issue, in the design strategy of Modular we exploited the availability of highly stable antibody and protein scaffolds that were reported to be relatively tolerant of insertions in the CDR loops. Indeed, the key advantages of this modular approach are that it allows the use of independent techniques to design different molecular traits, and also that different designed or available components can be combined in different ways to obtain antibodies with tailored characteristics.

Combinatorial sequence-based antibody design: the Cascade method

We recently introduced a combinatorial method to rationally design novel antibodies targeting specific epitopes within intrinsically disordered proteins (IDPs).¹⁰⁹ IDPs carry out their

functions while being devoid of a well-defined three-dimensional structure, constitute about one-third of eukaryotic proteomes, and play key roles in regulation and signalling.¹⁷³ Importantly, these proteins are closely linked with a wide range of human diseases, including cancer and neurodegeneration, and mostly classify as undruggable according to the lock-and-key paradigm.¹⁷⁴

On average, traditional antibody discovery methods are able to yield antibodies for disordered antigens with about the same likelihood of ordered antigens.¹⁵⁹ Besides, in spite of the entropic cost associated with IDP immobilisation upon binding, the median binding affinity for disordered epitopes was found to be only five-fold lower than that for structured ones.159 However, some disordered regions have been reported to contain relatively few MHC class I and II binding peptides, thus suggesting that the immune response against IDPs may have specific characteristics distinct from those against structured proteins,¹⁷⁵ and may be restricted towards fewer epitopes within disordered antigens, which appear to be immunodominant. For example, in the case of disordered human α -synuclein, nanobodies raised in camelids were all directed towards the immunodominant C-terminus of the protein.^{176,177} However, different regions along the sequence of IDPs typically mediate different interactions regulating different processes, including pathological ones such as neurotoxic aggregation.^{173,178} It is therefore important to be able to obtain antibodies targeting specific epitopes within disordered regions, ideally with singleresidue resolution.

To address this requirement, we recently introduced a peptide design and grafting procedure, whereby peptides designed to bind to linear epitopes of interest are grafted onto antibody CDRs.¹⁰⁹ These complementary peptides are designed with a combinatorial strategy, named Cascade method. Since disordered regions lack a specific structure by definition, this method only requires the amino acid sequence of the target epitope as input. The identification of complementary peptides is based on the analysis of the interactions between amino acid sequences from the PDB. With this choice, the affinity and the specificity of the interactions exploited are already proven in a biological context (*i.e.* the native structures of proteins). The Cascade procedure starts by collecting from the PDB all fragments facing in a β -strand any subsequence of the target epitope of at least three residues. Then, starting from the longest of these fragments, complementary peptides are grown to the length of the target epitope by joining some of the other fragments following three rules: (i) fragments generating the same complementary peptide must come from β -strands of the same type (*i.e.*, parallel or antiparallel), (ii) they must partly overlap with their neighbouring fragments, and (iii) the overlapping regions must be identical in terms of both sequence and hydrogen-bonding pattern (Fig. 3a).¹⁰⁹ The choice of β -strands is motivated by the fact that these are linear in nature, and therefore directly amenable to sequence-based design. Furthermore, protein aggregates formed by IDPs and underpinning protein-misfolding diseases are enriched in β-strand structures.¹⁷⁸ We therefore reasoned that the designed antibodies (DesAbs) obtained in this way would preferentially bind to pathological aggregated species, as the entropic cost of binding should be smaller, and aggregates will have greater avidity. Finally, while β -strand-mediated binding is not the typical binding mode of natural antibodies, the structure of a camelid nanobody binding its antigen with this mechanism was recently solved.¹⁷⁹

Given the large number of protein structures deposited in the PDB, the Cascade approach was shown to be generally applicable, as for the vast majority of potential target epitopes it is able to generate a number of complementary peptides to choose from.¹⁰⁹ The best peptides are then selected according to statistical scores reflecting how often the fragments comprising the designed peptide are found interacting with the corresponding ones within the target epitope, as opposed to with other sequences, and to a solubility score from the CamSol method (Fig. 3a).⁵⁵ This solubility prediction is particularly important because β -strands are notoriously enriched with aggregationprone motifs that drive protein folding, but whose occurrence should be minimized on antibody paratopes.

The Cascade method was initially experimentally validated by designing five human-derived single-domain antibodies with the CDR3 loop engineered to target different epitopes within three disease-related IDPs: α -synuclein, A β 42 and IAPP, whose aggregations are respectively hallmarks of Parkinson's disease, Alzheimer's disease and type-II diabetes. All designed antibodies (DesAbs) could be effectively expressed in E. coli, were correctly folded and stable, and successfully bound their targets. Encouraged by these results, we measured the K_d value of one antibody designed for the central region of α -synuclein (DesAb-F), and we quantitatively probed its ability to inhibit α -synuclein aggregation *in vitro*. DesAb-F bound α -synuclein monomers with a K_d value of about 15 μ M, and was a potent inhibitor of its aggregation even at sub-stoichiometric concentrations,109 at variance with immunisation-obtained nanobodies targeting the C-terminus of α-synuclein^{176,177} (Fig. 3b and c). In a later study, DesAb-F was also shown to inhibit α-synuclein aggregation *in vivo* in a *C. elegans* model of Parkinson's disease, and to rescue the disease phenotype.¹⁸⁰

Finally, as single-domain antibodies bind the antigen with up to three CDRs, and these validation DesAbs only had one engineered CDR, a novel DesAb was designed with a second complementary peptide grafted in the CDR2 with the aim of increasing the affinity for the monomeric antigen. The K_d value of this two-loop design was about 45 nM, corresponding to a two/three orders of magnitude improvement over the one-loop design targeting the same region of α -synuclein. However, expression yields and stability of the domain scaffold were affected by the insertion of a second extended loop, indicating that further work is required to make this type of design generally applicable.

New possibilities in sequence-activity relationship studies

Understanding which parts of a protein are key for its function is an important requirement for the development of new therapeutics, as it allows the identification of regions to target pharmacologically. Methods that enable such identification are globally referred as "sequence–activity relationship studies" or "scanning", as they are based on the functional analysis of the primary sequence of a protein in a systematic manner. These scanning strategies often rely on site-directed mutagenesis, and include for example the widely-employed alanine scanning.^{181,182} This method consists in the functional screening of a library of mutational variants, each carrying an alanine substitution at a different site. The identification of the variants with the lowest activity thus enable the mapping of those residues that are most important in determining function.¹⁸² However, it is often difficult to determine whether a mutation selectively affects function over structure and stability, and therefore the structural integrity of each mutant must be assessed separately.¹⁸³

An alternative technique for the identification of functional parts of a protein is antibody scanning, also called epitopemapping.^{184,185} Rather than mutagenesis, this strategy employs a library of antibodies, in which each antibody targets a different epitope of the protein under investigation. The activity of the protein is then measured in the presence of each antibody individually, and the ones causing the strongest inhibition are those that target an important functional region. The strength of the antibody scanning approach relies on the antibody library that is used for the analysis. To obtain reliable results, this library has to provide exhaustive coverage of the target protein, and there should be no ambiguity about where each antibody actually binds. With standard antibody-discovery techniques, it can be very challenging to obtain such a pool of antibodies.

In this regard, the ability to quickly generate antibodies with essentially single-residue resolution in the choice of epitope has opened up novel avenues for the rapid production of antibody libraries to be used in scanning studies. In a recent work, we generated a library of antibodies designed to bind successively along the whole sequence of the A β 42 peptide (Fig. 4a),¹⁸⁶ whose aggregation is an hallmark of Alzheimer's disease.

The aim of the study was to determine which regions of the A β sequence should be targeted to inhibit the production of the toxic oligomeric species, which form during the aggregation of AB and are widely regarded as one of the major factors triggering the disease.^{178,187–190} The formation of A β oligomers is a complex process that depends on several interconnected steps. In particular, once a critical concentration of early oligomers has formed through the self-assembly of AB monomers as a result of primary nucleation, and converted into fibrils, surface-catalysed secondary nucleation becomes the dominant mechanism of oligomer generation, as fibril surfaces can promote the formation of new oligomeric species.^{188,191} These oligomers can then grow and convert into additional fibrils, thus providing a positive feedback mechanism that results in rapid aggregate proliferation (Fig. 4b). The inhibition of primary and secondary nucleation events may respectively delay or decrease toxicity, whereas the inhibition of fibril elongation may actually lead to an overall increase in toxicity.¹⁹² Therefore, effective therapeutic strategies must be aimed at targeting precise microscopic steps



Fig. 4 Antibody scanning applied to the Aβ peptide. (a) Schematic representation of the panel of designed antibodies (DesAbs) developed and characterised in ref. 186 and corresponding epitopes on the sequence of Aβ42 (in the middle). (b) Cartoon illustrating the aggregation mechanism of Aβ42. Monomers, oligomer and fibrils are shown. (c) Effect of the DesAbs on primary (red) and secondary (blue) nucleation, and elongation (black) steps in the aggregation of Aβ42. (d) Numerical estimates of the change in the population of on-pathway oligomers in the absence and presence of the DesAbs. Reproduced from ref. 186 with permission from the American Association for the Advancement of Science, copyright 2017.

during the A β aggregation process, and the knowledge of which regions to target will be essential to develop a truly rational drug discovery program. The availability of this small but maximallydiverse library of DesAbs enabled to investigate systematically the different fundamental molecular-level events in A β aggregation, both *in vitro* and *in vivo* in a *C. elegans* model.¹⁸⁶

All five DesAbs in the library were folded, stable, and able to bind A β 42 monomers, and showed an epitope-dependent binding to amyloid fibrils in agreement with the known solvent-exposure of their epitopes in fibril structures.¹⁸⁶ All antibodies inhibited the *in vitro* aggregation of A β 42, and they did so in distinct manners, which are fully compatible with the corresponding suppression of toxicity in a *C. elegans* model of A β -overexpression. The analysis revealed that the central (residues 18–24) and c-terminal (29–36) regions of the peptide are important to selectively target primary and secondary nucleation, respectively, while the DesAb targeting the n-terminus (3–9) inhibited both processes, albeit to a weaker extent (Fig. 4c and d).¹⁸⁶ Furthermore, targeting regions 13–19 and 36–42 also resulted in a strong inhibition of fibril elongation, which should be avoided in therapeutic development.^{178,192} In this way, we successfully mapped the contribution of the primary sequence to the microscopic events that underpin the aggregation of Aβ42, which is very relevant for drug discovery.

This work demonstrate that approaches of rational antibody design can be exploited to generate powerful research tools, which enable to address basic research questions as well as provide novel insights for drug development programmes.

Discussion on the opportunities offered by the Modular method. In summary, the design strategy that we implemented in the Modular platform involves a peptide-design and grafting procedure, which, in synergy with an accurate prediction of solubility, enables the design of novel antibodies targeting linear epitopes of interest. As complementary peptides are

designed almost independently of the antibody scaffold, this strategy is highly modular and applicable to the design of protein-protein interactions more generally (Fig. 3). The modularity of the design also allows overcoming, at least in part, the complexity of modelling contrasting molecular traits discussed in the introduction. For instance, for our studies we relied on a single-domain antibody scaffold that was engineered for high stability,165 we designed solubility-improving mutations on the stem of the CDR loops,⁵⁵ and we exploited the rest of the loops to design affinity and specificity, while also screening for solubility.¹⁰⁹ In this modular framework different components of the antibodies are designed to modulate a single molecular trait, in essence yielding the correct balance trough compensatory effects. For example, the grafting of an artificial complementary peptide is almost certainly a destabilising feature, which is however compensated by the high stability of the scaffold and therefore very well tolerated. This framework also inherently enable the use of unrelated techniques (e.g. CamSol and the Cascade method) to design different components, and different components can then be combined in a number of ways to obtain protein binders for specific tasks, as shown for instance by our work on the rational design of molecular chaperones.^{193,194} This flexibility will enable in the future to readily extend the potential and the applicability of the Modular approach by incorporating new "modules", for example for the design of thermodynamic stability or of low immunogenicity. The latter in particular is a crucial property for the development of therapeutic antibodies, and the combinatorial assembly of designed antibodies may in principle give rise to neo-epitopes, which may elicit an adverse immune response. The increasing number of human antibody sequences available across public databases can be exploited to ensure that designed antibodies are as human-like as possible, for example by combining only amino acid fragments also found within human antibodies and proteins. A similar and complementary approach would be to predict and consequently remove candidate T-cell epitopes.195

This modular combinatorial design strategy has proven to be highly successful, as all eleven DesAbs that were tested in experimental validation were stable and active.^{109,186} As only the best scoring designs were chosen for laboratory production, we do not expect that all complementary peptides generated by the Cascade procedure will yield functional antibodies. Nonetheless, hundreds of different peptides are typically generated for a given epitope,¹⁰⁹ and our experimental validation on three unrelated antigens,¹⁰⁹ together with the antibody scanning of the A β sequence,¹⁸⁶ as well as other applications to molecular chaperones^{193,194} and solubility design,^{30,55,170} demonstrate the broad applicability of this strategy.

In addition, thanks to a pre-computed library of interacting fragments and the lack of complex energy calculations, the Cascade method runs in a few seconds on standard laptops, and CamSol solubility predictions are even faster. This feature makes the computational approach highly competitive with laboratory-based methods of antibody discovery, at least in terms of time and resources necessary for implementation. However, further method development is required to optimise the rational design of high affinity DesAbs for monomeric proteins. For example, while the proof-of-concept two-loop DesAb was able to bind its target with a K_d of ~ 45 nM, stability and expression yield of the antibody were affected. Moreover, the current implementation of the algorithm is limited to unstructured linear epitopes, albeit the overarching idea of combining protein fragments known to interact within protein structures can be generalized to structured targets.

Overall, the high success rate of this purely combinatorial sequence-based design, together with that of the empirical hotspot-centric approaches discussed in the previous section, indicate that structural databases provide an invaluable resource for the design of novel antibodies. By contrast, the lower success rates observed with energy-based approaches suggest that the available energy functions and sampling strategies are not yet fully able to recapitulate the complex rules that govern antibody binding, stability and solubility.

Conclusions

The development of the hybridoma technology in the mid '70s opened up the possibility of producing monoclonal antibodies against specific antigens,¹⁹⁶ and paved the way for the establishment of a first generation of antibody discovery methods based on animal immunisation. Subsequently, in the early '90s, breakthroughs in biotechnology lead to the development of a second generation of antibody discovery methods based on DNA-library construction and screening, for example with phage display and related techniques.^{197,198} These in vitro methods offer several advantages over in vivo approaches, including increased control over antigen presentation and the possibility of creating and screening fully-human antibody libraries. In addition, these second-generation technologies are highly complementary to immunisation approaches, as for instance scFv libraries can be generated from spleen cells of immunized animals and then screened with phage-display. For these reasons, in vitro approaches quickly gained popularity and are now widely employed in industry and academia.

In this review we have discussed the emerging endeavour of de novo rational antibody design. We believe that it is now particularly timely to take on the challenge to develop and establish a third generation of technologies of antibody discovery and optimisation, based on in silico rational design (Fig. 5a). The last two decades have seen transformational advances in the biomedical sciences, to the extent that the 21st century has been heralded as the century of biology, following the extraordinary developments of the 20th century in physics and the 19th century in chemistry.¹⁹⁹ In particular, the Human Genome Project has triggered the development of next-generation sequencing technologies, which are enriching biological databases with millions of sequences of proteins and antibodies from myriad different sources. Furthermore, improvements in the pace and accuracy of protein structure determination techniques are contributing unprecedented



Fig. 5 Towards the advent of *in silico* antibody discovery. (a) Pictorial timeline of the three generations of antibody discovery technologies, which consists in *in vivo* methods based on animal immunisation (blue), *in vitro* methods of library construction and screening (yellow) and *in silico* methods of rational design (green). These approaches are not mutually exclusive but are highly complementary to each other. A non-comprehensive list of some of the advances that enabled the development of each generation of technologies is presented below the *x*-axis. (b and c) Numbers of protein (b) and antibody (c) structures in public repositories, which are rapidly growing in time (sources: RCSB and SAbDab databases).

amounts of high-quality structural data, comprising large numbers of antibody-antigen complexes (Fig. 5b and c).^{135,200}

The increasing use of quantitative methods in biology has gradually transformed the way biological observations are made, and it is now possible to assemble large datasets of highly accurate measurements of antibody biophysical properties, including binding affinity, stability and solubility.^{55,146} These data, together with structural and dynamical information (e.g. from NMR spectroscopy), are also enabling the development of increasingly accurate energy functions, which are employed to simulate the conformational dynamics of proteins and protein complexes underlying many approaches of rational design.⁹⁵ Finally, computers able to quickly perform complex calculations are now available, and the cost of computing power is going down at a staggering rate. Taken together, these advances enable to address questions that were essentially intractable a decade ago, including the rational design of antibodies.

While still in its infancy, *de novo* antibody design has already produced some success stories, and it has already facilitated novel types of experiments, like the antibody scanning approach that we have discussed here. We anticipate that, in combination with well-established *in vitro* directed-evolution techniques, and with emerging high-throughput DNA-library manufacturing technologies, *in silico* antibody discovery will make it possible to generate novel opportunities to understand the fundamental principles of molecular recognition, and to develop effective antibodies for research, and for diagnostic and therapeutic applications.

In our vision, *in silico* rational design has the potential of becoming the preferred procedure for the generation of antibodies for many applications, as this approach offers unique opportunities to accelerate discoveries in the biomedical sciences. Not only it promises to provide a cheaper and faster route to antibody discovery, but also to enable a uniquely precise control over the properties of the resulting antibodies. We envision a future in which it may become possible to use computational methods to swiftly and reliably obtain antibodies with customised stability, solubility and binding affinity, targeting pre-determined epitopes of interests that would be challenging to access using current approaches.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge the Centre for Misfolding Diseases for providing funding for this work. P. S. is also supported by a Borysiewicz Fellowship from the University of Cambridge, and F. A. A. by a Senior Research Fellowship award from the Alzheimer's Society, UK (grant number 317, AS-SF-16-003). We are grateful to Mr Oded Rimon for a critical reading of the manuscript.

References

Published on 09 October 2018. Downloaded by University of Cambridge on 12/18/2018 7:55:08 AM.

- 1 S. Tonegawa, Nature, 1983, 302, 575-581.
- 2 P. C. Wilson, O. de Bouteiller, Y. J. Liu, K. Potter, J. Banchereau, J. D. Capra and V. Pascual, *J. Exp. Med.*, 1998, **187**, 59–70.
- 3 E. Market and F. N. Papavasiliou, PLoS Biol., 2003, 1, E16.
- 4 D. M. Ecker, S. D. Jones and H. L. Levine, *mAbs*, 2015, 7, 9–14.
- 5 J. G. Elvin, R. G. Couston and C. F. van der Walle, *Int. J. Pharm.*, 2013, **440**, 83–98.
- 6 H. Kaplon and J. M. Reichert, *mAbs*, 2018, 10, 183–203.
- 7 K. Famm, L. Hansen, D. Christ and G. Winter, *J. Mol. Biol.*, 2008, **376**, 926–931.
- 8 P. Holliger and P. J. Hudson, Nat. Biotechnol., 2005, 23, 1126-1136.
- 9 L. Jespers, O. Schon, K. Famm and G. Winter, *Nat. Bio*technol., 2004, 22, 1161–1165.
- 10 J. McCafferty, A. D. Griffiths, G. Winter and D. J. Chiswell, *Nature*, 1990, **348**, 552–554.
- E. Traggiai, S. Becker, K. Subbarao, L. Kolesnikova, Y. Uematsu, M. R. Gismondo, B. R. Murphy, R. Rappuoli and A. Lanzavecchia, *Nat. Med.*, 2004, **10**, 871–875.
- 12 J. Wrammert, K. Smith, J. Miller, W. A. Langley, K. Kokko, C. Larsen, N.-Y. Zheng, I. Mays, L. Garman, C. Helms, J. James, G. M. Air, J. D. Capra, R. Ahmed and P. C. Wilson, *Nature*, 2008, 453, 667–671.
- 13 S. T. Reddy, X. Ge, A. E. Miklos, R. A. Hughes, S. H. Kang, K. H. Hoi, C. Chrysostomou, S. P. Hunicke-Smith, B. L. Iverson, P. W. Tucker, A. D. Ellington and G. Georgiou, *Nat. Biotechnol.*, 2010, 28, 965–969.
- 14 E. T. Boder, K. S. Midelfort and K. D. Wittrup, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 10701–10705.
- 15 S. Spencer, D. Bethea, T. S. Raju, J. Giles-Komar and Y. Feng, *mAbs*, 2012, 4, 319–325.
- 16 P. T. Jones, P. H. Dear, J. Foote, M. S. Neuberger and G. Winter, *Nature*, 1986, **321**, 522–525.
- 17 N. Lonberg, L. D. Taylor, F. A. Harding, M. Trounstine,
 K. M. Higgins, S. R. Schramm, C. C. Kuo, R. Mashayekh,
 K. Wymore and J. G. McCabe, *Nature*, 1994, 368, 856–859.
- 18 A. Jakobovits, R. G. Amado, X. Yang, L. Roskos and G. Schwab, *Nat. Biotechnol.*, 2007, 25, 1134–1143.
- 19 E. Traggiai, L. Chicha, L. Mazzucchelli, L. Bronz, J.-C. Piffaretti, A. Lanzavecchia and M. G. Manz, *Science*, 2004, 304, 104–107.
- 20 A. J. Murphy, L. E. Macdonald, S. Stevens, M. Karow, A. T. Dore, K. Pobursky, T. T. Huang, W. T. Poueymirou,

L. Esau, M. Meola, W. Mikulka, P. Krueger, J. Fairhurst, D. M. Valenzuela, N. Papadopoulos and G. D. Yancopoulos, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 5153–5158.

- 21 P. J. Kennedy, C. Oliveira, P. L. Granja and B. Sarmento, *Crit. Rev. Biotechnol.*, 2018, **38**, 394–408.
- 22 A. R. M. Bradbury, S. Sidhu, S. Dübel and J. McCafferty, *Nat. Biotechnol.*, 2011, **29**, 245–254.
- 23 A. Zhao, M. R. Tohidkia, D. L. Siegel, G. Coukos and Y. Omidi, *Crit. Rev. Biotechnol.*, 2016, **36**, 276–289.
- 24 J. Glanville, W. Zhai, J. Berka, D. Telman, G. Huerta,
 G. R. Mehta, I. Ni, L. Mei, P. D. Sundar, G. M. R. Day,
 D. Cox, A. Rajpal and J. Pons, *Proc. Natl. Acad. Sci. U. S. A.*,
 2009, 106, 20216–20221.
- 25 A. Bradbury and A. Plückthun, Nature, 2015, 518, 27-29.
- 26 K. E. Tiller and P. M. Tessier, *Annu. Rev. Biomed. Eng.*, 2015, 17, 191–216.
- 27 J. B. Spangler, J. Tomala, V. C. Luca, K. M. Jude, S. Dong,
 A. M. Ring, P. Votavova, M. Pepper, M. Kovar and
 K. C. Garcia, *Immunity*, 2015, 42, 815–825.
- 28 D. Corti and A. Lanzavecchia, Annu. Rev. Immunol., 2013, 31, 705–742.
- 29 K. C. Ohaegbulam, A. Assal, E. Lazar-Molnar, Y. Yao and X. Zang, *Trends Mol. Med.*, 2015, 21, 24–33.
- 30 P. Sormanni, L. Amery, S. Ekizoglou, M. Vendruscolo and B. Popovic, *Sci. Rep.*, 2017, 7, 8200.
- 31 J. M. Perchiacca and P. M. Tessier, Annu. Rev. Chem. Biomol. Eng., 2012, 3, 263–286.
- 32 T. Jain, T. Sun, S. Durand, A. Hall, N. R. Houston, J. H. Nett, B. Sharkey, B. Bobrowicz, I. Caffry, Y. Yu, Y. Cao, H. Lynaugh, M. Brown, H. Baruah, L. T. Gray, E. M. Krauland, Y. Xu, M. Vásquez and K. D. Wittrup, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 944–949.
- 33 A. Jarasch, H. Koll, J. T. Regula, M. Bader, A. Papadimitriou and H. Kettenberger, J. Pharm. Sci., 2015, 104, 1885–1898.
- 34 C. J. Roberts, Trends Biotechnol., 2014, 32, 372-380.
- 35 D. Seeliger, P. Schulz, T. Litzenburger, J. Spitz, S. Hoerer, M. Blech, B. Enenkel, J. M. Studts, P. Garidel and A. R. Karow, *mAbs*, 2015, 7, 505–515.
- 36 F. Sewell, K. Chapman, J. Couch, M. Dempster, S. Heidel, L. Loberg, C. Maier, T. K. Maclachlan, M. Todd and J. W. van der Laan, *mAbs*, 2017, 9, 742–755.
- 37 D. U. Ferreiro, E. A. Komives and P. G. Wolynes, *Q. Rev. Biophys.*, 2014, **47**, 285–363.
- 38 S. Gianni, C. Camilloni, R. Giri, A. Toto, D. Bonetti, A. Morrone, P. Sormanni, M. Brunori and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14141–14146.
- 39 G. Winter and C. Milstein, Nature, 1991, 349, 293-299.
- 40 N. Tokuriki and D. S. Tawfik, *Curr. Opin. Struct. Biol.*, 2009, **19**, 596–604.
- 41 L. Rockah-Shmuel, A. Toth-Petroczy and D. S. Tawfik, *PLoS Comput. Biol.*, 2015, **11**, e1004421.
- 42 G. G. Tartaglia, S. Pechmann, C. M. Dobson and M. Vendruscolo, *Trends Biochem. Sci.*, 2007, **32**, 204–206.
- 43 P. Leuenberger, S. Ganscha, A. Kahraman, V. Cappelletti,
 P. J. Boersema, C. von Mering, M. Claassen and P. Picotti, *Science*, 2017, 355, eaai7825.

Published on 09 October 2018. Downloaded by University of Cambridge on 12/18/2018 7:55:08 AM.

- 44 D. M. Walther, P. Kasturi, M. Zheng, S. Pinkert, G. Vecchi,
 P. Ciryam, R. I. Morimoto, C. M. Dobson, M. Vendruscolo,
 M. Mann and F. U. Hartl, *Cell*, 2015, 161, 919–932.
- 45 P. Ciryam, R. Kundra, R. I. Morimoto, C. M. Dobson and M. Vendruscolo, *Trends Pharmacol. Sci.*, 2015, 36, 72–77.
- 46 D. A. Drummond and C. O. Wilke, *Cell*, 2008, **134**, 341–352.
- 47 M. Sauerborn, V. Brinks, W. Jiskoot and H. Schellekens, *Trends Pharmacol. Sci.*, 2010, **31**, 53–59.
- 48 K. L. S. Cleary, H. T. C. Chan, S. James, M. J. Glennie and M. S. Cragg, *J. Immunol.*, 2017, **198**, 3999–4011.
- 49 Y. Tang, J. Lou, R. K. Alpaugh, M. K. Robinson, J. D. Marks and L. M. Weiner, *J. Immunol.*, 2007, **179**, 2815–2823.
- 50 J. Li, W. Zhao, R. Akbani, W. Liu, Z. Ju, S. Ling, C. P. Vellano, P. Roebuck, Q. Yu, A. K. Eterovic, L. A. Byers, M. A. Davies, W. Deng, Y. N. V. Gopal, G. Chen, E. M. von Euw, D. Slamon, D. Conklin, J. V. Heymach, A. F. Gazdar, J. D. Minna, J. N. Myers, Y. Lu, G. B. Mills and H. Liang, *Cancer Cell*, 2017, **31**, 225–239.
- 51 R. Vincentelli, S. Canaan, J. Offant, C. Cambillau and C. Bignon, Anal. Biochem., 2005, 346, 77–84.
- 52 S. Cabantous and G. S. Waldo, Nat. Methods, 2006, 3, 845-854.
- 53 T. J. Gibson, K. Mccarty, I. J. McFadyen, E. Cash, P. Dalmonte, K. D. Hinds, A. A. Dinerman, J. C. Alvarez and D. B. Volkin, *J. Pharm. Sci.*, 2011, **100**, 1009–1021.
- 54 D. H. Atha and K. C. Ingham, J. Biol. Chem., 1981, 256, 12108-12117.
- 55 P. Sormanni, F. A. Aprile and M. Vendruscolo, *J. Mol. Biol.*, 2015, **427**, 478–490.
- 56 Z. Hamrang, N. J. W. Rattray and A. Pluen, *Trends Biotechnol.*, 2013, 31, 448–458.
- 57 M. R. G. Kopp and P. Arosio, J. Pharm. Sci., 2018, 107, 1228–1236.
- 58 M. R. G. Kopp, A. Villois, U. Capasso Palmiero and P. Arosio, *Ind. Eng. Chem. Res.*, 2018, 57, 7112–7120.
- 59 P. Arosio, K. Hu, F. A. Aprile, T. Muller and T. P. J. Knowles, *Anal. Chem.*, 2016, 88, 3488–3493.
- 60 G. J. Rocklin, T. M. Chidyausiku, I. Goreshnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith and D. Baker, *Science*, 2017, 357, 168–175.
- 61 D. A. Parsell and R. T. Sauer, J. Biol. Chem., 1989, 264, 7590–7595.
- 62 J. Willuda, A. Honegger, R. Waibel, P. A. Schubiger, R. Stahel, U. Zangemeister-Wittke and A. Pluckthun, *Cancer Res.*, 1999, **59**, 5758–5767.
- 63 J. R. Livingstone, Nature, 1996, 384, 491-492.
- 64 S. Hearty, P. Leonard and R. O'Kennedy, *Methods Mol. Biol.*, 2012, **907**, 411–442.
- 65 C. J. Wienken, P. Baaske, U. Rothbauer, D. Braun and S. Duhr, *Nat. Commun.*, 2010, 1, 100.
- 66 M. Visintin, E. Tse, H. Axelson, T. H. Rabbitts and
 A. Cattaneo, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, 96, 11723-11728.
- 67 Y. Abdiche, D. Malashock, A. Pinkerton and J. Pons, *Anal. Biochem.*, 2008, 377, 209–217.

- 68 P. Arosio, T. Muller, L. Rajah, E. V. Yates, F. A. Aprile, Y. Zhang, S. I. A. Cohen, D. A. White, T. W. Herling, E. J. De Genst, S. Linse, M. Vendruscolo, C. M. Dobson and T. P. J. Knowles, *ACS Nano*, 2016, **10**, 333–341.
- 69 T. W. Herling, D. J. O'Connell, M. C. Bauer, J. Persson, U. Weininger, T. P. J. Knowles and S. Linse, *Biophys. J.*, 2016, **110**, 1957–1966.
- 70 D. J. Cahill, J. Immunol. Methods, 2001, 250, 81-91.
- 71 N. Tokuriki, F. Stricher, L. Serrano and D. S. Tawfik, *PLoS Comput. Biol.*, 2008, 4, e1000002.
- 72 J. D. Dimitrov, S. V. Kaveri and S. Lacroix-Desmazes, *Trends Biochem. Sci.*, 2014, **39**, 221–226.
- 73 P. Koenig, C. V. Lee, B. T. Walters, V. Janakiraman, J. Stinson, T. W. Patapoff and G. Fuh, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E486–E495.
- 74 I. Sela-Culang, S. Alon and Y. Ofran, J. Immunol., 2012, 189, 4890–4899.
- 75 J. R. Jeliazkov, A. Sljoka, D. Kuroda, N. Tsuchimura, N. Katoh, K. Tsumoto and J. J. Gray, *Front. Immunol.*, 2018, 9, 413.
- 76 A. Broom, Z. Jacobi, K. Trainor and E. M. Meiering, J. Biol. Chem., 2017, 292, 14349–14361.
- 77 A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman,
 P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck,
 J. Prilusky, R. L. Lieberman, A. Aharoni, I. Silman,
 J. L. Sussman, D. S. Tawfik and S. J. Fleishman, *Mol. Cell*, 2016, 63, 337–346.
- 78 D. Bednar, K. Beerens, E. Sebestova, J. Bendl, S. Khare, R. Chaloupkova, Z. Prokop, J. Brezovsky, D. Baker and J. Damborsky, *PLoS Comput. Biol.*, 2015, **11**, e1004556.
- 79 A. Espargaró, V. Castillo, N. S. de Groot and S. Ventura, J. Mol. Biol., 2008, 378, 1116–1131.
- 80 S. V. Thakkar, N. Sahni, S. B. Joshi, B. A. Kerwin, F. He, D. B. Volkin and C. R. Middaugh, *Protein Sci.*, 2013, 22, 1295–1305.
- 81 K. Trainor, Z. Gingras, C. Shillingford, H. Malakian, M. Gosselin, D. Lipovšek and E. M. Meiering, *J. Mol. Biol.*, 2016, 428, 1365–1374.
- 82 M. Stenvang, N. P. Schafer, K. G. Malmos, A.-M. W. Pérez, O. Niembro, P. Sormanni, R. V. Basaiawmoit, G. Christiansen, M. Andreasen and D. E. Otzen, *J. Mol. Biol.*, 2018, 430, 1116–1140.
- 83 M. C. Julian, C. C. Lee, K. E. Tiller, L. A. Rabia, E. K. Day,
 A. J. Schick and P. M. Tessier, *Protein Eng., Des. Sel.*, 2015, 28, 339–350.
- 84 M. C. Julian, L. Li, S. Garde, R. Wilen and P. M. Tessier, *Sci. Rep.*, 2017, 7, 45259.
- 85 P. Kunz, K. Zinner, N. Mücke, T. Bartoschik, S. Muyldermans and J. D. Hoheisel, *Sci. Rep.*, 2018, 8, 7934.
- 86 M. N. Nguyen, M. R. Pradhan, C. Verma and P. Zhong, *Bioinformatics*, 2017, 33, 2971–2976.
- 87 R. L. Kelly, D. Le, J. Zhao and K. D. Wittrup, J. Mol. Biol., 2018, 430, 119–130.
- 88 K. E. Tiller, L. Li, S. Kumar, M. C. Julian, S. Garde and P. M. Tessier, J. Biol. Chem., 2017, 292, 16638–16652.
- 89 A. Burkovitz, I. Sela-Culang and Y. Ofran, *FEBS J.*, 2014, 281, 306–319.

- 90 H.-P. Peng, K. H. Lee, J.-W. Jian and A.-S. Yang, Proc. Natl. Acad. Sci. U. S. A., 2014, 111, E2656–65.
- 91 S. P. Ho and W. F. DeGrado, J. Am. Chem. Soc., 1987, 109, 6751–6758.
- 92 L. Regan and W. F. DeGrado, Science, 1988, 241, 976-978.
- 93 P.-S. Huang, S. E. Boyken and D. Baker, *Nature*, 2016, 537, 320–327.
- 94 S. J. Fleishman and D. Baker, Cell, 2012, 149, 262-273.
- 95 M. Bonomi, G. T. Heller, C. Camilloni and M. Vendruscolo, *Curr. Opin. Struct. Biol.*, 2017, 42, 106–116.
- 96 P. Sormanni, D. Piovesan, G. T. Heller, M. Bonomi, P. Kukic, C. Camilloni, M. Fuxreiter, Z. Dosztányi, R. V. Pappu, M. M. Babu, S. Longhi, P. Tompa, A. K. Dunker, V. N. Uversky, S. C. E. Tosatto and M. Vendruscolo, *Nat. Chem. Biol.*, 2017, **13**, 339–342.
- 97 J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano, *Nucleic Acids Res.*, 2005, **33**, W382–W388.
- 98 A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, *Methods Enzymol.*, 2011, **48**7, 545–574.
- 99 S. J. Fleishman, T. A. Whitehead, E.-M. Strauch, J. E. Corn, S. Qin, H.-X. Zhou, J. C. Mitchell, O. N. A. Demerdash, M. Takeda-Shitaka, G. Terashi, I. H. Moal, X. Li, P. A. Bates, M. Zacharias, H. Park, J.-S. Ko, H. Lee, C. Seok, T. Bourquard, J. Bernauer, A. Poupon, J. Azé, S. Soner, S. K. Ovali, P. Ozbek, N. B. Tal, T. Haliloglu, H. Hwang, T. Vreven, B. G. Pierce, Z. Weng, L. Pérez-Cano, C. Pons, J. Fernández-Recio, F. Jiang, F. Yang, X. Gong, L. Cao, X. Xu, B. Liu, P. Wang, C. Li, C. Wang, C. H. Robert, M. Guharoy, S. Liu, Y. Huang, L. Li, D. Guo, Y. Chen, Y. Xiao, N. London, Z. Itzhaki, O. Schueler-Furman, Inbar, V. Potapov, M. Cohen, G. Schreiber, Υ. Y. Tsuchiya, E. Kanamori, D. M. Standley, H. Nakamura, K. Kinoshita, C. M. Driggers, R. G. Hall, J. L. Morgan, V. L. Hsu, J. Zhan, Y. Yang, Y. Zhou, P. L. Kastritis, A. M. J. J. Bonvin, W. Zhang, C. J. Camacho, K. P. Kilambi, A. Sircar, J. J. Gray, M. Ohue, N. Uchikoga, Y. Matsuzaki, T. Ishida, Y. Akiyama, R. Khashan, S. Bush, D. Fouches, A. Tropsha, J. Esquivel-Rodríguez, D. Kihara, P. B. Stranges, R. Jacak, B. Kuhlman, S.-Y. Huang, X. Zou, S. J. Wodak, J. Janin and D. Baker, J. Mol. Biol., 2011, 414, 289-302.
- 100 G. Schreiber and A. E. Keating, *Curr. Opin. Struct. Biol.*, 2011, **21**, 50–61.
- 101 J. Ashworth, J. J. Havranek, C. M. Duarte, D. Sussman, R. J. Monnat, B. L. Stoddard and D. Baker, *Nature*, 2006, 441, 656–659.
- 102 G. Grigoryan, A. W. Reinke and A. E. Keating, *Nature*, 2009, 458, 859–864.
- 103 S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. DiMaio,

J. H. Pereira, B. Sankaran, G. Seelig, P. H. Zwart and D. Baker, *Science*, 2016, **352**, 680–687.

- 104 A. R. Thomson, C. W. Wood, A. J. Burton, G. J. Bartlett, R. B. Sessions, R. L. Brady and D. N. Woolfson, *Science*, 2014, 346, 485–488.
- 105 P. Gainza, H. M. Nisonoff and B. R. Donald, *Curr. Opin.* Struct. Biol., 2016, **39**, 16–26.
- 106 A. Chevalier, D.-A. Silva, G. J. Rocklin, D. R. Hicks, R. Vergara, P. Murapa, S. M. Bernard, L. Zhang, K.-H. Lam, G. Yao, C. D. Bahl, S.-I. Miyashita, I. Goreshnik, J. T. Fuller, M. T. Koday, C. M. Jenkins, T. Colvin, L. Carter, A. Bohn, C. M. Bryan, D. A. Fernandez Velasco, L. Stewart, M. Dong, X. Huang, R. Jin, I. A. Wilson, D. H. Fuller and D. Baker, *Nature*, 2017, **550**, 74–79.
- 107 C. A. Orengo and J. M. Thornton, Annu. Rev. Biochem., 2005, 74, 867–900.
- 108 B. Höcker, Curr. Opin. Struct. Biol., 2014, 27, 56-62.
- 109 P. Sormanni, F. A. Aprile and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9902–9907.
- 110 A. M. Poole and R. Ranganathan, *Curr. Opin. Struct. Biol.*, 2006, **16**, 508–513.
- 111 B. Li, M. Fooksa, S. Heinze and J. Meiler, *Crit. Rev. Biochem. Mol. Biol.*, 2018, **53**, 1–28.
- 112 K. Sankar, K. Jia and R. L. Jernigan, *Proc. Natl. Acad. Sci.* U. S. A., 2017, **114**, 2928–2933.
- 113 D. Xu and Y. Zhang, Proteins, 2012, 80, 1715-1735.
- 114 P. Xiong, M. Wang, X. Zhou, T. Zhang, J. Zhang, Q. Chen and H. Liu, *Nat. Commun.*, 2014, **5**, 5330.
- 115 F. Chiti, M. Stefani, N. Taddei, G. Ramponi and C. M. Dobson, *Nature*, 2003, **424**, 805–808.
- 116 S. B. Fowler, S. Poon, R. Muff, F. Chiti, C. M. Dobson and J. Zurdo, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10105–10110.
- 117 R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik and S. Ventura, *Nucleic Acids Res.*, 2015, **43**, W306–W313.
- 118 A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz and L. Serrano, *Nat. Biotechnol.*, 2004, **22**, 1302–1306.
- 119 G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti and M. Vendruscolo, *J. Mol. Biol.*, 2008, 380, 425-436.
- 120 H. Bohr, J. Bohr, S. Brunak, R. M. Cotterill, B. Lautrup, L. Nørskov, O. H. Olsen and S. B. Petersen, *FEBS Lett.*, 1988, 241, 223–228.
- 121 B. Rost and C. Sander, J. Mol. Biol., 1993, 232, 584-599.
- 122 P. Sormanni, C. Camilloni, P. Fariselli and M. Vendruscolo, J. Mol. Biol., 2015, 427, 982–996.
- 123 E. Cilia, R. Pancsa, P. Tompa, T. Lenaerts and W. F. Vranken, *Nat. Commun.*, 2013, 4, 2741.
- 124 L. Wei and Q. Zou, Int. J. Mol. Sci., 2016, 17, 2118.
- 125 B. Alipanahi, A. Delong, M. T. Weirauch and B. J. Frey, *Nat. Biotechnol.*, 2015, 33, 831–838.
- 126 E. Liberis, P. Velickovic, P. Sormanni, M. Vendruscolo and P. Lio, *Bioinformatics*, 2018, **273**, 927.
- 127 K. Krawczyk, T. Baker, J. Shi and C. M. Deane, *Protein Eng.*, *Des. Sel.*, 2013, **26**, 621–629.
- 128 J. Wang, H. Cao, J. Z. H. Zhang and Y. Qi, *Sci. Rep.*, 2018, 8, 6349.

- 129 A. Paladino, F. Marchetti, S. Rinaldi and G. Colombo, WIREs Comput. Mol. Sci., 2017, 7, e1318.
- 130 Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444.
- 131 K. Krawczyk, J. Dunbar and C. M. Deane, *Methods Mol. Biol.*, 2017, **1529**, 399–416.
- 132 Y. S. Choong, Y. V. Lee, J. X. Soong, C. T. Law and Y. Y. Lim, *Adv. Exp. Med. Biol.*, 2017, **1053**, 221–243.
- 133 W. D. Lees, L. Stejskal, D. S. Moss and A. J. Shepherd, *Front. Immunol.*, 2017, 8, 143.
- 134 B. D. Weitzner, J. R. Jeliazkov, S. Lyskov, N. Marze, D. Kuroda, R. Frick, J. Adolf-Bryfogle, N. Biswas, R. L. Dunbrack and J. J. Gray, *Nat. Protoc.*, 2017, **12**, 401–416.
- 135 J. Adolf-Bryfogle, Q. Xu, B. North, A. Lehmann and R. L. Dunbrack, *Nucleic Acids Res.*, 2015, **43**, D432–8.
- 136 B. Al-Lazikani, A. M. Lesk and C. Chothia, *J. Mol. Biol.*, 1997, **273**, 927–948.
- 137 L. Di Rienzo, E. Milanetti, R. Lepore, P. P. Olimpieri and A. Tramontano, *Sci. Rep.*, 2017, 7, 45053.
- 138 A. M. Sevy and J. Meiler, *Microbiol. Spectrum*, 2014, 2, 173–190.
- 139 J. C. Almagro, A. Teplyakov, J. Luo, R. W. Sweet, S. Kodangattil, F. Hernandez-Guzman and G. L. Gilliland, *Proteins*, 2014, 82, 1553–1562.
- 140 C. Regep, G. Georges, J. Shi, B. Popovic and C. M. Deane, *Proteins*, 2017, **85**, 1311–1318.
- 141 C. Marks and C. M. Deane, *Comput. Struct. Biotechnol. J.*, 2017, **15**, 222–231.
- 142 O. Keskin, BMC Struct. Biol., 2007, 7, 31.
- 143 H. Nishigami, N. Kamiya and H. Nakamura, *Protein Eng.*, *Des. Sel.*, 2016, **29**, 477–484.
- 144 M. F. Lensink, S. Velankar and S. J. Wodak, *Proteins*, 2017, 85, 359–377.
- 145 K. P. Kilambi and J. J. Gray, Sci. Rep., 2017, 7, 8145.
- 146 S. Sirin, J. R. Apgar, E. M. Bennett and A. E. Keating, *Protein Sci.*, 2016, **25**, 393–409.
- 147 R. J. Pantazes and C. D. Maranas, *Protein Eng., Des. Sel.*, 2010, 23, 849–858.
- 148 T. Li, R. J. Pantazes and C. D. Maranas, *PLoS One*, 2014, 9, e105954.
- 149 V. G. Poosarla, T. Li, B. C. Goh, K. Schulten, T. K. Wood and C. D. Maranas, *Biotechnol. Bioeng.*, 2017, **114**, 1331–1342.
- 150 K. C. Entzminger, J.-M. Hyun, R. J. Pantazes, A. C. Patterson-Orazem, A. N. Qerqez, Z. P. Frye, R. A. Hughes, A. D. Ellington, R. L. Lieberman, C. D. Maranas and J. A. Maynard, *Sci. Rep.*, 2017, 7, 10295.
- 151 R. J. Pantazes, M. J. Grisewood, T. Li, N. P. Gifford and C. D. Maranas, *J. Comput. Chem.*, 2015, 36, 251–263.
- 152 R. J. Pantazes and C. D. Maranas, *BMC Bioinf.*, 2013, 14, 168.
- 153 G. D. Lapidoth, D. Baran, G. M. Pszolla, C. Norn, A. Alon,
 M. D. Tyka and S. J. Fleishman, *Proteins*, 2015, 83, 1385–1406.
- 154 D. Baran, M. G. Pszolla, G. D. Lapidoth, C. Norn, O. Dym, T. Unger, S. Albeck, M. D. Tyka and S. J. Fleishman, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 10900–10905.

- 155 S. Warszawski, R. Netzer, D. S. Tawfik and S. J. Fleishman, *J. Mol. Biol.*, 2014, **426**, 4125–4138.
- 156 J. Adolf-Bryfogle, O. Kalyuzhniy, M. Kubitz, B. D. Weitzner, X. Hu, Y. Adachi, W. R. Schief and R. L. Dunbrack, *PLoS Comput. Biol.*, 2018, 14, e1006112.
- 157 C. Chothia, A. M. Lesk, A. Tramontano, M. Levitt, S. J. Smith-Gill, G. Air, S. Sheriff, E. A. Padlan, D. Davies, W. R. Tulip, P. M. Colman, S. Spinelli, P. M. Alzari and R. J. Poljak, *Nature*, 1989, 342, 877–883.
- 158 Y. Tsuchiya and K. Mizuguchi, Protein Sci., 2016, 25, 815–825.
- 159 C. A. MacRaild, J. S. Richards, R. F. Anders and R. S. Norton, *Structure*, 2016, 24, 148–157.
- 160 S. J. Fleishman, T. A. Whitehead, D. C. Ekiert, C. Dreyfus, J. E. Corn, E.-M. Strauch, I. A. Wilson and D. Baker, *Science*, 2011, 332, 816–821.
- 161 S. Liu, S. Liu, X. Zhu, H. Liang, A. Cao, Z. Chang and L. Lai, Proc. Natl. Acad. Sci. U. S. A., 2007, 104, 5330–5335.
- 162 L. N. Robinson, K. Tharakaraman, K. J. Rowley, V. V. Costa, K. R. Chan, Y. H. Wong, L. C. Ong, H. C. Tan, T. Koch, D. Cain, R. Kirloskar, K. Viswanathan, C. W. Liew, H. Tissire, B. Ramakrishnan, J. R. Myette, G. J. Babcock, V. Sasisekharan, S. Alonso, J. Chen, J. Lescar, Z. Shriver, E. E. Ooi and R. Sasisekharan, *Cell*, 2015, **162**, 493–504.
- 163 X. Liu, R. D. Taylor, L. Griffin, S.-F. Coker, R. Adams, T. Ceska, J. Shi, A. D. G. Lawson and T. Baker, *Sci. Rep.*, 2017, 7, 41306.
- 164 G. Moroncini, N. Kanu, L. Solforosi, G. Abalos, G. C. Telling, M. Head, J. Ironside, J. P. Brockes, D. R. Burton and R. A. Williamson, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, 101, 10404–10409.
- 165 P. A. Barthelemy, H. Raab, B. A. Appleton, C. J. Bond, P. Wu, C. Wiesmann and S. S. Sidhu, *J. Biol. Chem.*, 2008, 283, 3639–3654.
- 166 A. R. A. Ladiwala, M. Bhattacharya, J. M. Perchiacca, P. Cao, D. P. Raleigh, A. Abedini, A. M. Schmidt, J. Varkey, R. Langen and P. M. Tessier, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 19965–19970.
- 167 J. M. Perchiacca, A. R. A. Ladiwala, M. Bhattacharya and P. M. Tessier, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 84–89.
- 168 C. C. Lee, M. C. Julian, K. E. Tiller, F. Meng, S. E. DuConge, R. Akter, D. P. Raleigh and P. M. Tessier, *J. Biol. Chem.*, 2016, **291**, 2858–2873.
- 169 J. M. Perchiacca, A. R. A. Ladiwala, M. Bhattacharya and P. M. Tessier, *Protein Eng., Des. Sel.*, 2012, **25**, 591–601.
- 170 C. Camilloni, B. M. Sala, P. Sormanni, R. Porcari, A. Corazza, M. De Rosa, S. Zanini, A. Barbiroli, G. Esposito, M. Bolognesi, V. Bellotti, M. Vendruscolo and S. Ricagno, *Sci. Rep.*, 2016, 6, 25559.
- 171 S. Pechmann, E. D. Levy, G. G. Tartaglia and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10159–10164.
- 172 V. Potapov, M. Cohen and G. Schreiber, *Protein Eng., Des. Sel.*, 2009, **22**, 553–560.
- 173 J. Habchi, P. Tompa, S. Longhi and V. N. Uversky, *Chem. Rev.*, 2014, **114**, 6561–6588.

Published on 09 October 2018. Downloaded by University of Cambridge on 12/18/2018 7:55:08 AM.

- 174 G. T. Heller, P. Sormanni and M. Vendruscolo, *Trends Biochem. Sci.*, 2015, **40**, 491–496.
- 175 A. J. Guy, V. Irani, C. A. MacRaild, R. F. Anders, R. S. Norton, J. G. Beeson, J. S. Richards and P. A. Ramsland, *PLoS One*, 2015, **10**, e0141729.
- 176 E. J. De Genst, T. Guilliams, J. Wellens, E. M. O'Day, C. A. Waudby, S. Meehan, M. Dumoulin, S.-T. D. Hsu, N. Cremades, K. H. G. Verschueren, E. Pardon, L. Wyns, J. Steyaert, J. Christodoulou and C. M. Dobson, *J. Mol. Biol.*, 2010, **402**, 326–343.
- 177 T. Guilliams, F. El-Turk, A. K. Buell, E. M. O'Day, F. A. Aprile, E. K. Esbjörner, M. Vendruscolo, N. Cremades, E. Pardon, L. Wyns, M. E. Welland, J. Steyaert, J. Christodoulou, C. M. Dobson and E. J. De Genst, *J. Mol. Biol.*, 2013, 425, 2397–2411.
- 178 F. Chiti and C. M. Dobson, Annu. Rev. Biochem., 2017, 86, 27-68.
- M. B. Braun, B. Traenkle, P. A. Koch, F. Emele, F. Weiss,
 O. Poetz, T. Stehle and U. Rothbauer, *Sci. Rep.*, 2016, 6, 19211.
- 180 M. Perni, F. A. Aprile, S. Casford, B. Mannini, P. Sormanni, C. M. Dobson and M. Vendruscolo, *Sci. Rep.*, 2017, 7, 15045.
- 181 B. C. Cunningham, P. Jhurani, P. Ng and J. A. Wells, *Science*, 1989, 243, 1330–1336.
- 182 K. L. Morrison and G. A. Weiss, *Curr. Opin. Chem. Biol.*, 2001, 5, 302–307.
- 183 F. F. Vajdos, C. W. Adams, T. N. Breece, L. G. Presta, A. M. de Vos and S. S. Sidhu, *J. Mol. Biol.*, 2002, **320**, 415–428.
- 184 T. Toyoda, K. Masunaga, Y. Ohtus, K. Hara, N. Hamada, T. Kashiwagi and J. Iwahashi, *Curr. Protein Pept. Sci.*, 2000, 1, 303–308.
- 185 J. Bergh, P. Zetterström, P. M. Andersen, T. Brännström, K. S. Graffmo, P. A. Jonsson, L. Lang, J. Danielsson, M. Oliveberg and S. L. Marklund, *Proc. Natl. Acad. Sci.* U. S. A., 2015, 201419228.

- 186 F. A. Aprile, P. Sormanni, M. Perni, P. Arosio, S. Linse, T. P. J. Knowles, C. M. Dobson and M. Vendruscolo, *Sci. Adv.*, 2017, 3, e1700488.
- 187 M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson and M. Stefani, *Nature*, 2002, **416**, 507–511.
- 188 T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland and C. M. Dobson, *Science*, 2009, 326, 1533–1537.
- 189 C. Haass and D. J. Selkoe, *Nat. Rev. Mol. Cell Biol.*, 2007, 8, 101–112.
- 190 A. Jan, O. Adolfsson, I. Allaman, A.-L. Buccarello, P. J. Magistretti, A. Pfeifer, A. Muhs and H. A. Lashuel, *J. Biol. Chem.*, 2011, 286, 8585–8596.
- 191 S. I. A. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson and T. P. J. Knowles, *Proc. Natl. Acad. Sci.* U. S. A., 2013, **110**, 9758–9763.
- 192 P. Arosio, M. Vendruscolo, C. M. Dobson and T. P. J. Knowles, *Trends Pharmacol. Sci.*, 2014, **35**, 127–135.
- 193 F. A. Aprile, P. Sormanni and M. Vendruscolo, *Biochemistry*, 2015, **54**, 5103–5112.
- 194 M. N. Bongiovanni, F. A. Aprile, P. Sormanni and M. Vendruscolo, *Int. J. Mol. Sci.*, 2018, **19**, 1443.
- 195 P. Kar, L. Ruiz-Perez, M. Arooj and R. L. Mancera, *Pept. Sci.*, 2018, **110**, e24046.
- 196 G. Köhler and C. Milstein, Nature, 1975, 256, 495-497.
- 197 G. Winter, A. D. Griffiths, R. E. Hawkins and H. R. Hoogenboom, *Annu. Rev. Immunol.*, 1994, **12**, 433–455.
- 198 G. P. Smith, Science, 1985, 228, 1315-1317.
- 199 The scientific century: securing our future prosperity, RS Policy Document, 2010.
- 200 J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs,
 G. Georges, J. Shi and C. M. Deane, *Nucleic Acids Res.*,
 2014, 42, D1140.