ROYAL SOCIETY OF CHEMISTRY

REVIEW

Check for updates

Stabilization challenges and aggregation in protein-based therapeutics in the pharmaceutical industry

Mahdie Rahban, [ID] [a] Faizan Ahmad,[b] Mieczyslaw A. Piatyszek, [ID] [c] Thomas Haertlé [ID] [d] Luciano Saso[e] and Ali Akbar Saboury [ID] *[f]

Protein-based therapeutics have revolutionized the pharmaceutical industry and become vital components in the development of future therapeutics. They offer several advantages over traditional small molecule drugs, including high affinity, potency and specificity, while demonstrating low toxicity and minimal adverse effects. However, the development and manufacturing processes of protein-based therapeutics presents challenges related to protein folding, purification, stability and immunogenicity that should be addressed. These proteins, like other biological molecules, are prone to chemical and physical instabilities. The stability of protein-based drugs throughout the entire manufacturing, storage and delivery process is essential. The occurrence of structural instability resulting from misfolding, unfolding, and modifications, as well as aggregation, poses a significant risk to the efficacy of these drugs, overshadowing their promising attributes. Gaining insight into structural alterations caused by aggregation and their impact on immunogenicity is vital for the advancement and refinement of protein therapeutics. Hence, in this review, we have discussed some features of protein aggregation during production, formulation and storage as well as stabilization strategies in protein engineering and computational methods to prevent aggregation.

[a]Neuroscience Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran. E-mail: mrohban@ut.ac.ir

[b]Department of Biochemistry, School of Chemical & Life Sciences, Jamia Hamdard, New Delhi-110062, India. E-mail: faizanahmad@jamiahamdard.ac.in

[c]Biotechnology Consultant, Gainesville, FL, USA. E-mail: piatyszekm@gmail.com

[d]National Institute of Agronomic Research, Nantes, France. E-mail: tom@haertle.fr

[e]Department of Physiology and Pharmacology "Vittorio Erspamer", Sapienza University, Rome, Italy. E-mail: luciano.saso@uniroma1.it

[f]Institute of Biochemistry and Biophysics, University of Tehran, Tehran 1417614335, Iran. E-mail: saboury@ut.ac.ir; Fax: +9821 66404680; Tel: +9821 66956984

Mahdie Rahban

Mahdie Rahban earned her PhD in biophysics from the Institute of Biochemistry and Biophysics at the University of Tehran. Currently serving as a senior researcher at the Neuroscience Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences. Her research focuses on protein engineering, enzyme kinetics, drug discovery, and computational biology, with a particular emphasis on cancer research.

Faizan Ahmad

Faizan Ahmad is a Full Professor, which completed his PhD in Biophysical Chemistry in 1974. During his stay in Jamia Millia Islamia, Ahmad taught both undergraduate and post graduate courses. His major research work has been in the area of Protein Folding with special interest in the development of methods for the determination of protein stability and understanding of mechanism of protein stabilization by osmolytes under stress conditions. He has guided 53 PhD students and published over 330 research papers. He has received several awards and honors such as Fellow of Indian Science Academy, Fellow of National Academy of Science, Pride of Delhi, DAAD Fellow.

1. Introduction

In the last three decades, protein-based therapeutics have emerged as a significant category of pharmaceuticals. Protein-based drugs encompass therapeutic agents consisting of proteins or peptides engineered to interact with specific targets within the body for the treatment or management of diseases.[1–3] The human proteome contains an estimated 20 000 proteins.[4] Proteins are highly versatile biomolecules playing crucial roles in various biological processes. They act as catalysts, scaffolds for structural integrity, signalling molecules, molecular transporters and receptors, among other functions.[5,6] Their versatility enables proteins to contribute to the essential functions and maintenance of cellular and tissue integrity.[7] Protein-based therapeutics provide a multitude of advantages relative to traditional small-molecule drugs. Notably, their intricate three-dimensional structures confer the ability to interact with specific receptors or molecules in the body with remarkable selectivity and affinity, with minimal side effects.[8–10] The therapeutic proteins can be derived from natural sources or produced by genetic engineering.[11–13] The approval and market introduction of human insulin, produced by recombinant technique in *Escherichia coli* (*E. coli*), in 1982, represented an important milestone in the field of protein therapeutics.[14] Following this achievement, significant progress was made in the cloning and expression of other protein-based drugs, including human growth hormone, interferon β and α, and monoclonal antibodies.[15] These advancements contributed further to the development of recombinant therapeutics and vaccines.[16] The production of recombinant proteins has gradually emerged as a vital component within the biopharmaceutical industry.[17,18] The application of genetic

Mieczyslaw A. Piatyszek

Mieczyslaw A. Piatyszek is an independent biotechnology consultant. His primary areas of expertise include cancer research, aging, and drug discovery. Dr Piatyszek is the author of numerous publications elucidating the involvement of human telomerase in cancer and cellular immortality. During his two decades career in the biotechnology industry, he made significant contributions to collaborative drug discovery programs with Pharmacia, Kyowa Hakko, Ltd, and Pfizer. Dr Piatyszek is listed as an inventor on seven US-issued patents related to drug discovery and his publications have generated over 16 000 citations.

Luciano Saso

Luciano Saso is Full Professor of the Faculty of Pharmacy and Medicine, Sapienza University of Rome. He is author of more than 350 original scientific articles published in peer reviewed international journals with impact factor (H-index Google Scholar = 63) working mainly in the field of oxidative stress. In 2022 he was nominated highly cited scientist (https://clarivate.com/highly-cited-researchers). He coordinated several international research projects and has been Editor and Guest Editor of Special Issues of different international journals.

Thomas Haertlé

Thomas Haertle completed his PhD in Biochemistry in 1976. Retired from Institute National de la Recherche Agronomique, Nantes, France is Adjunct Professor in Institute of Biochemistry and Biophysics, University of Teheran. Studied Chemistry, Biochemistry and Physico-Chemistry of proteases, food proteins, use of lactic acid and extremophile bacteria, the antimicrobial activities of bacteriocins and of modified dairy proteins. Author of 470 publications, over 190 communications, 90 conferences, 2 books, 3 book chapters, 5 patents, several expertises. His WoS H-Index is 53. Promotor of over 30 PhDs, Host to over 100 post-doctoral visits. Head of few dozens of multilateral and national research projects.
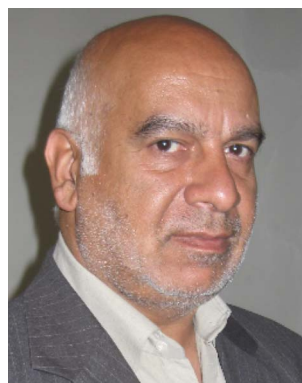
Ali Akbar Saboury

Ali Akbar Saboury completed his PhD in Biophysical Chemistry in 1996. He is Distinguished Professor in the Institute of Biochemistry and Biophysics, University of Tehran, Iran. He conducts research in the field of Biophysics, focusing on protein ligand interaction and stability using spectroscopy and microcalorimetry techniques. He oversees research projects that have resulted in the completion of 150 PhD and MSc degrees. To date, he has authored over 580 articles in international journals, 80 articles in national journals, and has written 26 books. He is a highly cited scientist (https://clarivate.com/highly-cited-researchers).

© 2023 The Author(s). Published by the Royal Society of Chemistry

engineering techniques allows for the efficient production of large quantities of target proteins with high consistency and cost-effectiveness compared to traditional methods. This advancement facilitates enhanced availability of therapeutic proteins, thereby improving accessibility for patients requiring these treatments.[19] By employing recombinant DNA technology, the production of protein-based drugs can be achieved using well-controlled and scalable systems. This allows for the generation of high-quality, consistent and purified protein products.[20,21] Fig. 1 exhibits schematically the process of recombinant biopharmaceutical products.

The progress in genetic engineering technology and a deeper comprehension of diseases provide a promising opportunity in the design and development of new biomolecular entities with desired biopharmaceutical features.[22] However, several challenges arise in this field. These challenges include: protein aggregation, denaturation, degradation, concomitant loss of activity, immunogenicity, non-specific distribution, rapid clearance from the body and potential toxicity.[23–25] These issues raise significant concerns that need to be resolved for the

successful development and application of protein-based therapeutics.[26] Recent advances in protein engineering and the ability to intentionally introduce chemical and structural modifications have created a paradigm shift in the fine-tuning of protein properties.[18] These advancements have enabled researchers to modify proteins at the molecular level, allowing precise control over their functional properties.[27] This capability has opened new avenues for the design of protein-based therapeutics with enhanced efficacy and stability, optimized pharmacokinetics and reduced immunogenicity. The ability to manipulate and fine-tune protein properties broadens avenues for the development of next-generation biopharmaceuticals.

Protein-based therapy faces various obstacles that can significantly impact its effectiveness and safety. Overcoming these challenges requires a combination of scientific, technological, and regulatory approaches.[17] Continuous research in protein engineering, advanced drug delivery systems, and manufacturing technologies contributes to addressing these obstacles and enhancing the efficacy and safety of protein-based therapies.[27] Table 1 outlines the specific challenges in
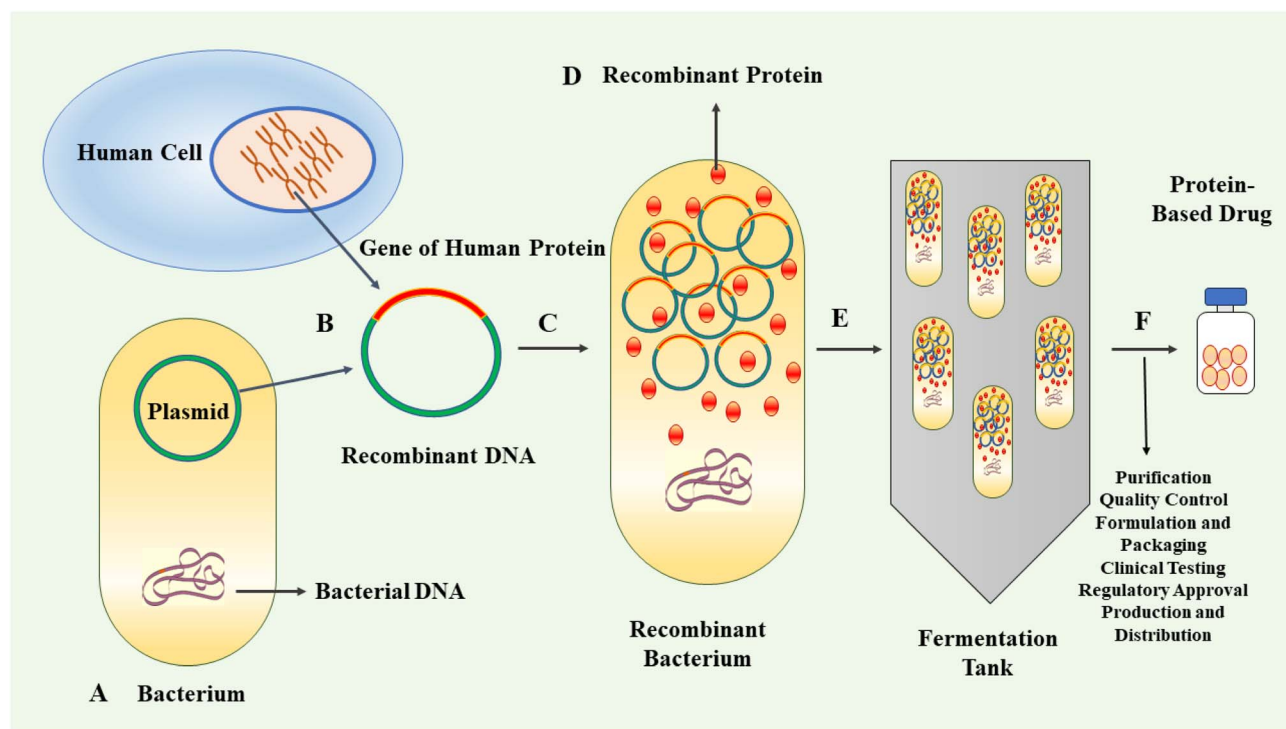

Fig. 1 Figure exhibited the process of producing recombinant biopharmaceutical products involves using genetic engineering techniques to create therapeutic proteins. (A) The first step is to select a suitable host organism, often a microorganism like bacteria, yeast, or mammalian cells. The choice of host organism depends on factors such as protein complexity and desired post-translational modifications. (B) The gene encoding the desired therapeutic protein is isolated and cloned into a vector. The vector also contains regulatory elements to control gene expression. (C) In the case of bacteria, the vector with the cloned gene is introduced into the host cells through a process called transformation. (D) Once the gene is inside the host cells, it is transcribed and translated, leading to the production of the therapeutic protein. (E) The cells are then grown in bioreactors (fermentation tanks) under controlled conditions to maximize protein yield. (F) After production, the biopharmaceutical product needs to be separated and purified from the host cell components. This is typically done through a series of chromatography and filtration steps, which isolate the protein of interest. The purified protein undergoes rigorous quality control testing to ensure it meets safety, efficacy, and purity standards. This includes tests for identity, potency, sterility, and absence of contaminants. The purified protein is formulated to the desired concentration and stability. It may also be mixed with excipients to improve shelf-life and administration. Before receiving regulatory approval, the biopharmaceutical product goes through extensive clinical trials to assess its safety and efficacy in humans. Once clinical trials are successful, the product can be submitted for regulatory approval by health authorities. If approved, the biopharmaceutical is produced at a larger scale and distributed to healthcare providers for patient use.

Table 1 A summary of the main obstacles in protein-based therapy and the corresponding strategies used to overcome these challenges

Obstacles in protein-based therapy	Strategies to overcome
Protein stability	Formulation optimization, stabilizers, protein engineering for enhanced stability
Immunogenicity	Protein engineering to reduce immunogenic epitopes, immunomodulatory agents
Delivery and targeting	Advanced delivery systems (nanoparticles, liposomes), specific targeting methods
Production costs	Advances in biotechnology, improved manufacturing processes
Dosing precision	Tailoring dosages, controlled-release systems for precision
Regulatory hurdles	Clear documentation, rigorous testing to meet regulatory standards
Short half-life	Modification (*e.g.*, PEGylation) to extend the therapeutic effect, altering formulation
Storage and handling	Proper storage conditions, suitable packaging for preservation

protein-based therapy and the corresponding methods to overcome them.

Our aim is to bridge the gap between the broader understanding of protein stability and its specific applications in therapeutic contexts by addressing the obstacles encountered in protein-based therapy due to protein instability.

2. Stability challenges of protein-based therapy

Stability plays a vital role in the development of protein-based therapy, encompassing aspects such as optimal storage conditions and *in vivo* performance.[28] Proteins are susceptible to various destabilizing factors, such as aggregation, degradation and denaturation, all of which have the potential to downgrade their intended functions.[28,29] Ensuring stability is essential to preserve the efficacy and functionality of protein-based therapies. The stability of protein-based drugs is of utmost importance during the entire process of manufacturing, storage, and delivery.[30] Structural instability, which can arise from

Fig. 2 Figure depicts the issues associated with protein-based therapies' stabilities, notably influencing their efficacy and safety attributes. This figure illustrates the critical role of stability in the development of protein-based therapy. Ensuring stability is crucial to maintain the efficacy and functionality of protein-based therapies throughout manufacturing, storage, and delivery processes. Structural instability, arising from misfolding, unfolding, various modifications, and aggregation, can impair drug efficacy. Proteins and peptides exhibit a limited stability range influenced by factors such as concentration, ionic strength, temperature, and pH often referred to as "marginal stability". Furthermore, proteins face additional challenges caused by proteases, immune responses and human physiological conditions that can impact their stability.

misfolding, unfolding, and different non-covalent or covalent modifications as well as aggregation, may deteriorate drug efficacy and their promising curative properties.[30,31] Proteins and peptides demonstrate a restricted range of stabilities, influenced by factors such as concentration, ionic strength, temperature and pH.[32–36] This narrow range is commonly referred to as "marginal stability".[37,38] Moreover, many recombinant proteins are inherently unstable under the conditions in which they are expressed, leading to a loss of correct folding or increased susceptibility to proteolytic digestion. This instability poses a significant challenge to obtaining adequately folded proteins for therapeutic applications.[39] Fig. 2 illustrates the stability hurdles confronting protein-based therapies, with profound implications for their effectiveness, stability, and safety profiles.

Maintaining the appropriate temperature conditions (below the midpoint of denaturation) helps to preserve protein stability and prevent undesirable changes in structure and function over time.[40,41] Strategies such as cold chain management, using temperature-controlled storage facilities and employing suitable freeze–thaw protocols are employed to ensure the integrity of therapeutic proteins.[41] Maintenance of proper pH of the environment plays a critical role in preserving the protein structural integrity and prevents undesirable chemical reactions that can lead to degradation.[30] Due to precise control and optimization of pH conditions, proteins can maintain their native conformations, preserving their biological activities and therapeutic efficacies throughout the entire lifecycle of the product.[42] The pH significantly influences intramolecular folding and protein–protein interactions by playing a critical role. At various pH values, specific amino acid residues can undergo protonation or deprotonation, thereby impacting their capacity to form hydrogen bonds, form electrostatic and hydrophobic interactions. Alterations in pH can disturb the charge balance within the protein resulting in modifications of its conformation and stability.[37,42,43] Additionally, certain proteins may interact favourably with surfaces of various vessels, leading to adsorption and decreasing the concentration of the active ingredient available for therapeutic action.[30] Therefore, a proper formulation encompasses factors such as buffer systems, excipients and stabilizers, which contribute to preserving the protein native structure and preventing undesired modifications. These formulations are designed to create optimal conditions, including appropriate pH, ionic strength
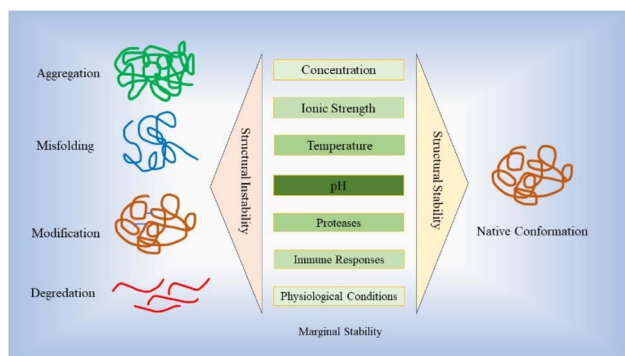
and osmolarity, in order to safeguard the stability of therapeutic proteins.[44–46]

Furthermore, by serving as protein folding assistants, chaperones enhance the yield of properly folded proteins, mitigating the formation of inclusion bodies during recombinant protein production.[47] Inclusion bodies are dense aggregates of misfolded or aggregated proteins that accumulate within the cytosol, nucleus, or periplasm of cells during recombinant protein production.[48] Integrating chaperones into the production process can significantly improve the overall quality and functionality of the final recombinant protein product.[49] Molecular chaperones, whether in the cytoplasm or the periplasm, play a crucial role in facilitating the production of properly folded recombinant proteins. In the cytoplasm, chaperones like Hsp100 (Clp), Hsp90, Hsp70 (DnaK/DnaJ/GrpE), Hsp60 (GroEL/GroES), and small heat shock proteins (sHsps) are involved in aiding the correct folding of recombinant proteins.[50] Periplasmic chaperones in cells are crucial for properly folding recombinant proteins by aiding in disulfide bond formation, ensuring their soluble form.[51] The efficiency of this process varies based on the protein type and is improved by proteins like FkpA and SurA, aiding in critical peptidyl–prolyl bond isomerization. Moreover, DsbA and DsbC are essential for forming disulfide bridges, preventing misfolded proteins and guiding recombinant proteins toward their functional form in the periplasmic space.[50] Certain chemical additives in culture media enhance the expression of soluble recombinant proteins by inducing osmotic stress and triggering the production of osmolytes like betaine and trehalose in *E. coli* cells. These osmolytes serve as "chemical chaperones," boosting the stability of native proteins and aiding in refolding unfolded polypeptides. Notably, D-sorbitol, in combination with betaine, is widely employed to reduce the formation of inclusion bodies in recombinant proteins.[52]

In the complex environment of the human body, proteins face additional challenges caused by proteases, immune responses and physiological conditions that can impact their stability. Understanding and addressing these stability concerns are critical for ensuring the effectiveness and safety of protein therapeutics.[53,54] To address these stability concerns, various structural modifications such as site-specific mutations, post transcriptional modifications (PTMs) and PEGylation, have been explored to improve protein solubility and stability. These modifications aim to enhance the overall performance and longevity of protein therapeutics.[55–59]

In addition to stability, protein-based therapeutics need to possess suitable pharmacokinetic and pharmacodynamic properties to ensure optimal functionality.[60] Pharmacokinetics encompasses absorption, distribution, metabolism and excretion of the therapeutic proteins within the body. It is important for the protein to exhibit appropriate bioavailability, distribution in the target site and elimination from the body.[61] On the other hand, pharmacodynamics relates to how the proteins interact with their targets and produce desired therapeutic effects. This includes factors such as binding affinity, receptor activation and downstream signalling pathways.[25,60] Achieving the desired pharmacokinetic and pharmacodynamic profiles is

crucial for the efficacy, safety and dosing regimen optimization of protein-based therapeutics. It requires careful design, optimization and characterization of the protein molecules, to ensure that they properly interact well with the therapeutic target and achieve the desired therapeutic effects with none or negligible off-target effects.[25,60,62–64]

2.1 Immunogenicity and protein aggregation

One of the primary challenges associated with the production, distribution and storage of protein-based drugs is the risk of protein aggregation. Aggregation refers to the process in which proteins associate and form clusters or aggregates, which can negatively impact the therapeutic efficacy.[65] Aggregated proteins may be recognized by the immune system as foreign or abnormal, triggering an immune response that can result in the production of antibodies against the therapeutic protein.[66] The immunogenic potential of all aggregate types remains uncertain and subject to debate within the scientific community. It is not yet fully understood whether all aggregates have the potential to elicit an immune response. Furthermore, the impact of additional clinical factors on immunogenicity is a complex area requiring further investigation.[67,68] Immunogenicity refers to the development of serum anti-drug antibodies (ADAs) that specifically target and bind to the proteins of interest, such as a therapeutic proteins or drugs. This immune response can have significant implications for the efficacy and safety of the protein products and can lead to reduced their effectiveness.[69,70] When ADAs bind to the therapeutic protein, they can neutralize its activity, prevent it from reaching its target or accelerate its clearance from the body. This can result in diminished therapeutic effects and reduced clinical benefits to the patients. In more severe cases, the formation of ADAs can pose serious health concerns.[71,72] The immune response triggered by ADAs can lead to adverse immune reactions, including hypersensitivity reactions, allergic responses or immune-mediated diseases. Examples include the formation of ADAs leading to reduced efficacy or life-threatening conditions like pure red cell aplasia (PRCA) with Eprex.[73] Thorlaksen *et al.*[74] showed that various aggregate populations are more immunogenic compared to the native human insulin, and the degree of immune activation depends on specific aggregate features.

Aggregates containing flexible, micron-sized particles with altered secondary structure were found to be highly immunogenic.[75] Compact micron-sized particles with native-like structures also exhibited immunogenicity *in vivo*. Subvisible were more immunogenic than submicron particles or soluble oligomers[76] which, is discussed in the next section. Chemical modifications of the insulin molecule did not significantly impact immunogenicity.[77] The correlation between *in vitro* and *in vivo* immunogenicity assessments suggests that distinct aggregate types activate different immunological pathways, contributing to overall immunogenicity.[74] A study demonstrated that changes in protein structure resulting from aggregation can alter the range of antigens targeted by the immune response. Protein aggregation increases the number of epitopes presented to major histocompatibility complex class II (MHC II)
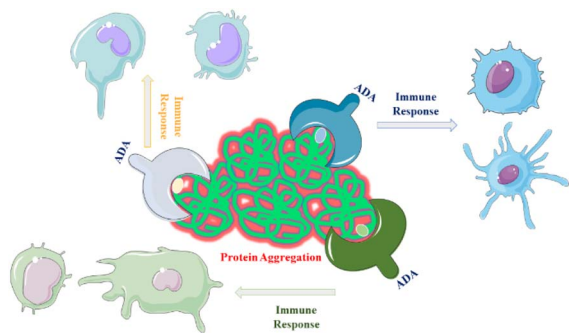
Fig. 3 The illustration depicts how protein aggregates can enhance the immunogenicity of a therapeutic protein by triggering the formation of serum anti-drug antibodies (ADAs) that bind to the protein. The distinct aggregate types activate different immunological pathways.

molecules.[78] For example, an assay comparing non-aggregated and aggregated formulations of two human IgGs subjected to heat and shake-induced aggregation demonstrated a significant increase in the number of epitopes presented on monocyte-derived dendritic cells (MoDCs).[23] This suggests that structural changes induced by aggregation may cause immune responses targeting antigens that are chemically modified or hidden in the native fold of the proteins (Fig. 3). This hypothesis is further supported by research using humanized single-chain variable antibody fragments (scFv).[23,79]

Peptide mapping and enzyme-linked immunosorbent assay (ELISA) identified a peptide sequence, which is highly targeted by antibodies induced by protein aggregates but not by monomers in wild-type mice.[23] Molecular dynamics (MD) simulations, protein structure prediction software and simulated annealing revealed that this peptide sequence was located in a hydrophobic domain of the scFv, which tends to become exposed due to partial unfolding.[80] Additionally, while native aggregates increased anti-drug antibody (ADA) levels, non-native aggregates induced ADA levels similar to, or lower than those induced by the monomer.[23] This suggests that the native conformation of the protein is important for immunogenicity. Understanding the structural changes induced by aggregation and their effects on immunogenicity is crucial for the development and optimization of protein therapeutics.[23,80]

Analytical techniques and immunogenicity assessments are utilized to detect and evaluate protein aggregates. Monitoring and managing immunogenicity is crucial, and it involves preclinical and clinical testing, assay development, and strategies to minimize immune responses. By addressing immunogenicity, the efficacy and safety of protein products can be optimized for better therapeutic outcomes.[81,82]

3. Aggregation in protein-based therapeutic products

Proteins strive to maintain their native conformations, driven by thermodynamics and the lowest energy states.[83] In the folded protein, hydrophobic amino acid residues are buried in the protein avoiding contact with water.[84] However, it is important to note that not all hydrophobic residues are buried within the protein structure. Water molecules can form loose structures (cages) around exposed hydrophobic regions. Consequently, folding, surface adsorption or collapse of hydrophobic patches in proteins can result in the release of structured (caged) water into the surrounding solvent, thereby increasing the overall system entropy.[29,85,86] Hydrophilic residues (uncharged and charged amino acids) would prefer to be on the surface of the folded protein. In fact, burial of charged side chains is very rare.[87] In contrast, buried uncharged polar side chains are very common. These buried uncharged polar side chains are involved in hydrogen bond formation; it is very rare to find unpaired hydrogen bond donors and acceptors. These buried hydrogen bonds are more stable than those exposed to polar solvent water. In addition to hydrophobic interactions and hydrogen bonds, other noncovalent interactions that contribute to the thermodynamic stability are van der Waals interactions and salt bridges.[88]

Monomeric folded proteins are marginally stable. This low thermodynamic stability may facilitate conformational changes within the protein structure, which can ultimately lead to protein instability and the subsequent aggregation.[85] Protein aggregation is a complex process that can occur through three main mechanisms, depending on the seeding entity involved, *i.e.*, the native monomers, denatured proteins and pre-existing aggregates.[89,90] Two or more native monomers, which exist in folded state, can self-associate and form oligomers. This self-association can be driven by complementary charge–charge interactions between different monomers or through the formation of noncovalent linkages between hydrophilic and hydrophobic residues on the protein surface. These interactions can lead to the assembly of larger aggregates.[91] The size of aggregates increases over time and they become more resistant to get back to their native states. It is commonly observed that protein products contain a fraction of denatured proteins, which have a higher propensity for irreversible aggregation. While partially unfolded proteins may possess the ability to refold under specific conditions, the thermodynamics and kinetics generally favour the aggregation process over refolding.[92–94] Another mechanism is that the native protein monomers can aggregate by adhering to pre-existing protein oligomers, contaminants or to the surfaces of containers. This adherence can initiate a nucleation process in which the aggregates rapidly grow in size. This expansion of aggregates further contributes to protein aggregation.[30]

Soluble aggregates have low molecular mass and may exhibit reversibility, while insoluble aggregates exceed the solution solubility limit and precipitate out of the solution.[69] In biological products, a small amount of soluble aggregates (5–10%) is generally considered acceptable. Insoluble aggregates are evaluated based on the size of particles detected upon reconstitution. Particles as small as 150 μm in diameter can be visually detected in injectable products and particles below 150 μm are more likely to induce immune reactions. Subvisible particles with a hydrodynamic radius of ≤50–100 μm are deemed acceptable, while particles ≥10 μm can obstruct blood flow. The U.S. Pharmacopeia (USP) sets the limit for particulate matter in

a container of ≤100 mL to be 6000 particles ≥10 μm and 600 particles ≥25 μm. However, recent regulatory scrutiny by the U.S. Food and Drug Administration (FDA) has emphasized the importance of controlling aggregates below 10 μm in biological products. Aggregates of this size pose a greater risk to product stability and immunogenicity.[76,85]

Colloidal stability of proteins relies on the intermolecular interactions among protein molecules. These interactions involve various forces, including short-range attractions, steric repulsions and electrostatic interactions. Short-range attractions arise from van der Waals forces between protein molecules when they come within a certain distance of each other. Steric repulsion occurs when protein molecules approach each other closely, leading to repulsive forces due to overlapping electron clouds or steric hindrance.[95] Electrostatic interactions involve the attraction or repulsion of charged protein molecules based on their electrostatic charges. These intermolecular forces play a crucial role in determining the colloidal stabilities of proteins.[96] The balance between these forces can either promote or hinder protein aggregation. For example, a strong steric repulsion and electrostatic repulsion between protein molecules can prevent their close association and aggregation.[97] Otherwise, if the attractive forces outweigh the repulsive forces, protein molecules may associate and form aggregates. Therefore, understanding and manipulating these intermolecular interactions are important in controlling the colloidal stability of proteins.[98]

Aggregate morphology, influenced by the protein and stress conditions, plays a crucial role in the surface area and subsequent immunogenicity of aggregates. Comparisons between globular and filamentous aggregates of human IgG models revealed that filamentous aggregates induced higher expression of immune activation markers (CD86, CD80, CD83) on MoDCs. Filamentous aggregates were predominantly formed under heat and shaking stress, while shear stress or freeze–thawing induced more globular aggregates.[23,78] Significantly, it has been observed that globular aggregates have a lower surface area relative to their mass/volume ratio compared to elongated filamentous aggregates. This observation implies that an increased surface area could potentially contribute to enhanced immune activation.[99]

In addition, oxidation induced by UV light exposure or metal contamination as well as temperature, pH, ionic conditions, mechanical agitation and freeze–thawing pressure, are among the key factors that can impact protein aggregation.[100–104] The pH of a solution has been found to exert diverse effects on protein aggregation. Extreme pH conditions or significant pH shifts from the protein isoelectric point (pI) can disrupt the electrostatic interactions and hydrogen bonds within protein. This disruption often results in a decrease in protein solubility accompanied by conformational changes that ultimately lead to aggregation.[105–107] Moreover, alterations in ionic conditions of solvents such as changes in the concentration of salts or divalent cations, can also influence protein solubility and facilitate aggregation.[108] For instance, in the context of monoclonal antibodies (mAbs), a recent study indicated that mAbs formulated in a sodium acetate buffer demonstrated more

conformational and colloidal stabilities compared to those formulated in a sodium citrate buffer.[109,110]

Protein and peptide aggregation depends on protein concentration, and the presence of metal ions can affect the rate of aggregation. However, the relationship between metal ions and protein aggregation is complex and not fully understood. Under normal physiological conditions, aggregation typically does not occur, but it can be induced by certain metal ions.[111,112] Metal ions like Zn^{2+}, Cu^{2+} and Pb^{2+} can inhibit or accelerate aggregation. Overall, understanding the precise impact of cations on protein aggregation remains a challenge, and further research is needed to establish a consensus.[37,111,113,114]

Chemical modifications, such as deamidation and oxidation, can have a significant impact on the stability of protein-based therapeutics, including antibodies. Deamidation involves the conversion of asparagine to aspartic acid or iso-aspartic acid, while oxidation involves the addition of oxygen or reactive oxygen species to amino acid residues. Both modifications can lead to conformational changes, reduce solubility and increased aggregation of antibodies. These modifications can occur spontaneously or be induced by various factors.[115,116] In a study by Alam et al.,[116] it was demonstrated that methionine oxidation in antibodies leads to a significant reduction in conformational stability. However, this oxidation has a limited effect on antibody aggregation, except under extreme conditions such as low pH and elevated temperature. Conversely, tryptophan oxidation and asparagine deamidation have minimal impact on antibody conformational stability but promote aggregation over the broader range of solution conditions.[116]

Furthermore, various factors contribute to protein aggregation. Mechanical agitation such as stirring, pumping or filtration, can disrupt protein structures and facilitate aggregation by inducing unfolding and exposing aggregation-prone regions. Freeze–thaw cycles encountered during storage, can exert pressure on proteins, resulting in conformational changes, unfolding and aggregation.[69,117] Exposure to ultraviolet (UV) light can trigger chemical reactions and oxidation, altering protein structures and promoting aggregation. Temperature plays a crucial role as elevated temperatures enhance molecular motions and unfold proteins, whereas extremely low temperatures can induce conformational changes and increase the risk of aggregation.[91]

4. Stabilization strategies to prevent aggregation of protein-based therapy

To prevent aggregation of proteins, it is crucial to consider and manage the environmental factors during the manufacturing and formulation development processes. During the manufacturing, processing, storage and delivery of protein therapeutics, various stresses can arise leading to protein aggregation and the occurrence of covalent modifications. These modifications encompass oxidation, deamidation, formation of disulfide bridges and cross-linking, which alter the native conformation of proteins.[27] Methionine, cysteine,
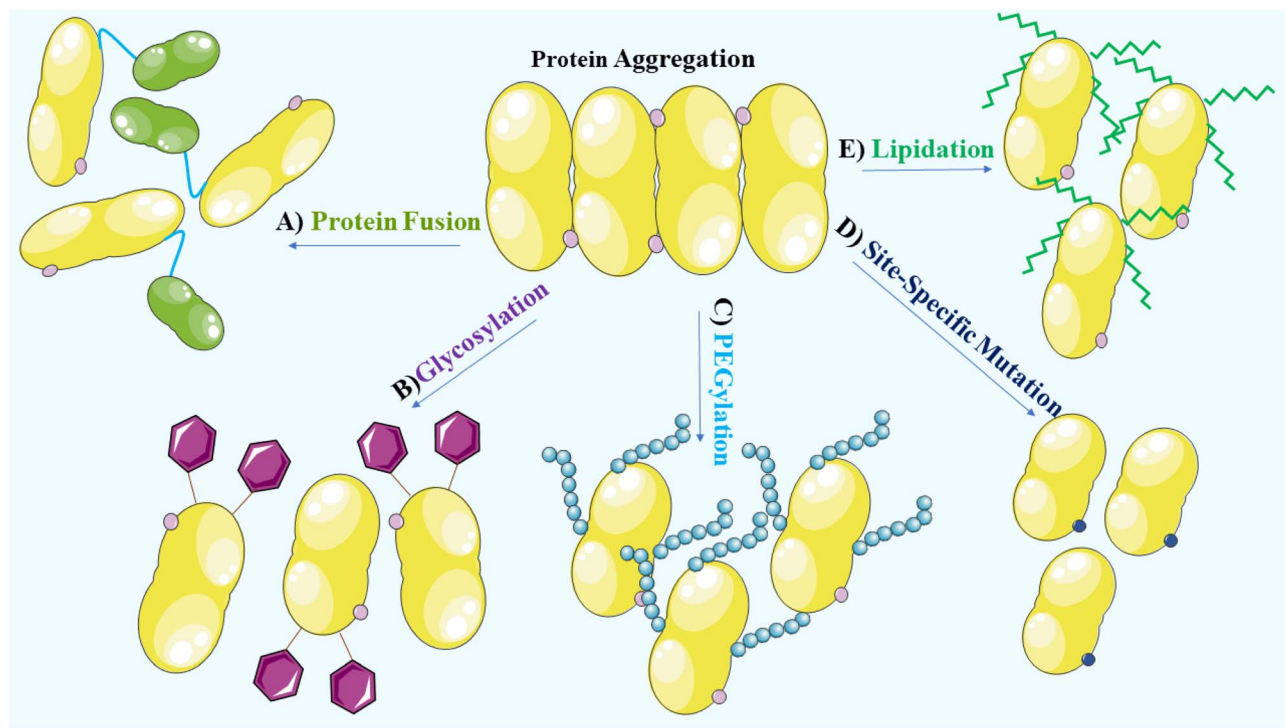
Fig. 4 A schematic representation demonstrates the utilization of various strategies to enhance the solubility and stability of protein aggregates, vital considerations in the development of protein-based therapy. These strategies include: (A) protein fusion, (B) glycosylation, (C) PEGylation, (D) structural modifications and site-specific mutation, and (E) lipidation.

histidine, tryptophan, tyrosine and phenylalanine are vulnerable to oxidation and promote aggregation.[23,118] In addition, deamidation of asparagine and glutamine, disulfide bridge formation involving two cysteine residues and cross-linking due to the formation of covalent bonds between protein molecules, contribute to the aggregate formation. Consideration of these factors is, therefore, vital for maintaining the stability and effectiveness of protein therapeutics.[27] Structural modifications, site-specific mutations,[27] glycosylation,[119] PEGylation,[120] lipidation[121] and protein fusion[122,123] that are widely used to enhance the solubility and stability of proteins, are vital considerations in the development of protein-based therapeutics (Fig. 4).

PEGylation is a widely employed method to enhance the stability of biotherapeutics. It involves the covalent attachment of polyethylene glycol (PEG) chains to the target protein.[124,125] PEGylation has significant pharmacokinetic advantages that include prolonged circulation time and reduced clearance. The addition of PEG increases the hydrodynamic size of the protein, resulting in decreased renal filtration and an extended half-life. This leads to improved therapeutic efficacy. PEGylation also provides a steric barrier that protects the protein from proteolytic degradation and immune recognition. Furthermore, it improves solubility and stability, preventing aggregation and denaturation under various environmental conditions.[56,126] The selection of an appropriate PEGylation strategy, including site of attachment and PEG chain length, is essential for achieving desired stability enhancements. Care must be taken to balance the benefits of PEGylation with potential drawbacks, such as

altered protein activity and potential immunogenicity of the PEGylated biotherapeutics.[124,127]

Optimizing the freeze-drying or lyophilization process and formulation parameters is crucial for minimizing aggregation. The choice of freezing procedure should consider the specific protein stability requirements to prevent conformational changes and subsequent aggregation, ultimately reducing the risk of immunogenicity.[128] During the freezing step in lyophilization, the solution is cooled to initiate ice nucleation and the growth of ice crystals. The freezing temperature is selected below the formulation glass transition temperature, and the freezing rate is chosen based on the desired ice crystal size and protein stability. Sufficient holding time at the freezing temperatures ensures thermal equilibrium. An annealing step can follow freezing to promote ice crystal growth and excipient crystallization. Annealing temperature is typically above the glass transition temperature but below the eutectic melting temperature.[129,130] Optimization of these parameters is essential for achieving efficient drying and uniform freeze-dried products.[131] Stabilizers play a crucial role in freeze-dried protein formulations to protect proteins during the freezing process. Cryoprotectants such as sucrose and trehalose are commonly used stabilizers. These cryoprotectants work by selectively excluding themselves from the protein surface, thereby raising the energy barrier for protein unfolding during freezing. This mechanism preserves effectively the native state of proteins, enhancing their stability throughout the freeze-drying process.[132–134] Furthermore, proteins tend to adsorb at different interfaces, including the air–liquid interface during

mixing or the ice–liquid or ice–air interface during freeze-drying. To counteract this phenomenon and enhance protein stability in such scenarios, surfactants are incorporated into protein formulations. Non-ionic surfactants such as Polysorbate 20, Polysorbate 80 and Poloxamer 188 are frequently employed in both liquid and freeze-dried protein formulations. These surfactants assist in preventing protein aggregation and adsorption at interfaces, thereby preserving the integrity and functionality of the proteins throughout the formulation and manufacturing processes.[133,135] A study of a model protein, myoglobin, known to be susceptible to cold denaturation, emphasized the significance of employing a fast-freezing technique to mitigate aggregation and maintain the activity. The hypothesis proposed that rapid freezing reduces the exposure of the protein to cold conditions, where denaturation can occur. However, this fast freezing approach leads to the generation of small ice crystals, which, in turn, enhances the surface area available for interactions between ice and water molecules.[128] Although the fast-freezing procedure proves beneficial for proteins prone to surface-induced conformational instability, it may not be suitable for all proteins. To illustrate this point, lactate dehydrogenase, another model protein, was employed as an example. The study demonstrated that the same fast-freezing method that effectively preserved myoglobin stability had adverse effects on lactate dehydrogenase. It led to protein aggregation and a subsequent loss of enzymatic activity.[136] This study emphasizes the significant variation in protein stabilities and the importance of adjusting environmental parameters accordingly to prevent potentially immunogenic aggregate formation.

5. Protein engineering strategy

5.1 Site-specific mutation

Protein engineering is a crucial component of protein-based therapy, enabling the creation of tailored proteins with improved therapeutic properties. Through computational modelling and structural analysis, protein engineers modify amino acid sequences, introduce site-specific mutations and incorporate functional motifs to optimize protein stability, binding affinity and enzymatic activity.[137,138] Furthermore, protein engineers have the ability to introduce or manipulate post-translational modifications (PTMs) in therapeutic proteins. PTMs such as glycosylation, phosphorylation, or lipidation exert substantial effects on protein stability, pharmacokinetics and immune recognition. These protein engineering strategies enhance the therapeutic potential of proteins and contribute to the development of more effective protein-based therapies.[139,140]

Site-specific mutations involve targeted alterations to specific amino acids within a protein sequence, exerting significant effects on its properties including solubility and stability.[141] These mutations are designed to address issues such as protein aggregation, susceptibility to proteolytic degradation and unfavourable interactions with container surfaces. By strategically introducing mutations at precise locations, researchers can disrupt aggregation-prone regions,

enhance resistance to proteolysis and mitigate undesirable adsorption, thereby improving the protein solubility and stability.[27]

Here are some specific examples of insulin variants that have been developed through site-specific mutagenesis to achieve different kinetics of action.

● Rapid-acting insulin analogues such as lispro (Humalog): lispro was developed through site-specific mutagenesis by exchange of proline and lysine amino acids at positions 28 and 29 in the insulin B chain. This modification enables lispro to be quickly absorbed after subcutaneous injection, leading to a faster onset of action than regular human insulin. Consequently, lispro closely imitates the rapid insulin secretion observed after meals, offering improved postprandial glucose control for individuals with diabetes.[142]

● Long-acting insulin analogues (Lantus, Toujeo): insulin glargine was engineered through site-specific mutagenesis by replacing asparagine at position 21 of the insulin A chain with glycine and adding two arginine residues at the C-terminus of the B chain. These modifications influence the solubility and stability of the insulin molecule, leading to the formation of micro-precipitates following injection. These micro-precipitates enable a gradual release of insulin over an extended duration, providing basal insulin coverage for up to 24 hours.[143,144]

5.2 Computational methods

The combination of computational methods and experimental approaches accelerates the development and optimization of protein-based therapy by offering valuable insights into protein structure, dynamics, interactions and design. These methods contribute to the rational design of therapeutic proteins, identification of potential drug candidates and better understanding of the underlying mechanisms of protein function and disease.[145–147] Computational methods such as molecular dynamics (MD) simulations and machine learning algorithms play a crucial role in predicting protein stability for protein-based therapy. These techniques analyse protein structures and sequences to identify regions that are prone to instability.[148] Furthermore, they can predict the impact of mutations or modifications on protein stability, facilitating the design of more stable protein variants.[84,149] These computational methods provide valuable insights into protein stability and contribute to the development of more robust and effective protein-based therapies. MD simulations are computational tools that simulate the movement of atoms and molecules over time. In the protein-based drugs, MD simulations provide insights into protein dynamics, stability and behaviour.[150] They assess protein stability under various conditions and predict how proteins will behave. MD simulations also offer a way to study protein-ligand interactions, helping researchers to understand the binding affinities and to optimize drug candidates.[151–153] Moreover, machine learning algorithms play a pivotal role in the protein-based therapy by leveraging large datasets of protein stability information to train predictive models. These models analyse protein sequences, structures and other pertinent features, enabling accurate predictions of
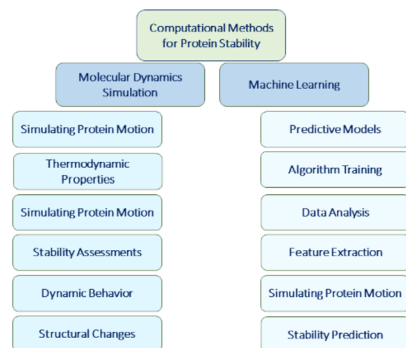
Fig. 5 This diagram represents primary branches of computational methods used in protein stability prediction.

protein stability.[154–156] By harnessing the power of machine learning, researchers can streamline the screening and prioritization of protein candidates based on their predicted stabilities. This approach minimizes the need for laborious and expensive experimental screening methods, allowing researchers to allocate resources toward the most promising candidates for further development. Machine learning expedites the identification and optimization of stable protein candidates, facilitating drug discovery of protein-based therapeutics (Fig. 5).[157,158]

5.3 Rational design

Rational protein engineering is an approach utilized to enhance protein stabilities by strategically introducing targeted modifications based on a thorough understanding of the protein properties.[159] Through a comprehensive analysis of the protein structure, dynamics and interactions, specific amino acid substitutions or structural modifications can be implemented to improve stability.[84] Computational tools, such as molecular modelling and energy calculations, are instrumental in guiding these rational design efforts.[160] By utilizing these computational tools, researchers can make rational decisions on which modifications are the most likely to enhance protein stabilities, leading ultimately to the development of more robust and effective proteins for protein-based therapeutics.[161]

Deep learning is applied to unlabelled amino acid sequences to extract essential protein features in rational protein engineering. Alley et al.[162] used unsupervised learning with a recurrent artificial neural network called UniRep (Unified Representation). This model was trained on the UniRef50 dataset, which consists of 24 million amino acid sequences. UniRep generates fixed-length vectors that summarize the information in protein sequences, independent of structural or evolutionary data. These vectors contain rich information and have excellent generalization ability, making them suitable for protein stability prediction. UniRep also significantly improves efficiency of protein engineering tasks, making it a versatile tool for protein engineering informatics.[162]

Wang et al.[163] developed an innovative computational method named BayeStab, specifically designed for predicting changes in protein thermostability caused by mutations. This

method serves as a valuable tool for researchers evaluating the impact of amino acid substitutions on protein stability. By utilizing BayeStab, scientists can make accurate predictions regarding the thermodynamic stability of mutated proteins, facilitating the rational design and optimization of proteins for various applications in protein-based therapeutics.[163]

Furthermore, the development of DeepDDG predicts changes in protein stability caused by point mutations. By utilizing neural network-based techniques and comprehensive training on a diverse dataset, DeepDDG enhances the ability to understand and manipulate protein stability, facilitating protein engineering efforts and advancing the understanding of missense mutations in proteins. Input features of DeepDDG demonstrated that the solvent-accessible surface area (SASA) of the mutated residue is the main character for the determination of protein stability.[164]

6. Prediction of protein aggregation by computational tools

Successful prediction of protein aggregation requires the adaption of computational schemes tailored to the specific properties of the protein being studied. However, this task can be challenging for untrained users due to the requirement of in-depth knowledge about the available computational tools. Understanding the intricacies of these tools is crucial for effectively applying them to analyze protein aggregation. SAP (Spatial Aggregation Propensity) is a technology used to identify regions within therapeutic antibodies that are prone to aggregation. By employing full antibody atomistic molecular dynamics simulations, researchers can pinpoint these aggregation-prone regions. This knowledge allows them to perform targeted mutations and engineer antibodies with improved stability. The application of SAP has significantly enhanced in the stability of therapeutic antibodies compared to the wild type. This technology has the potential to be integrated into the screening and discovery phase of antibody development, enabling the development of more stable and effective antibodies for different diseases.[165]

In a study by De Baets et al., a comprehensive analysis was performed on 611 protein sequences and their respective lifetimes using the aggregation predictor known as TANGO. The results of the study revealed an interesting correlation between protein lifetime and aggregation propensity. It was observed that proteins with shorter lifetimes exhibited a higher tolerance for aggregation.[166] Phenomenological algorithms utilize experimental data to understand the factors that contribute to protein aggregation and establish empirical aggregation scales. These algorithms, such as AGGRESCAN[167] and Zyggregator,[168] are based on the rationalization of experimentally determined factors that influence protein aggregation. Incorporating these factors allows the algorithms, to predict the protein aggregation, more effectively.[169]

The second class of computational approaches is centered around the theoretical assessment of specific sequence properties associated with protein aggregation. Algorithms within

this category such as TANGO,[166] PASTA 2.0,[170] Amyloid Mutants,[171] FoldAmyloid[172] and Waltz[173] analyze various characteristics of protein sequences. These include the propensity of a sequence to adopt a defined β-enriched conformation, the packing density of proteins, the composition and patterning of residues and the ability to adopt topologically restricted conformations commonly observed in amyloid-like states. By considering these sequence properties, these algorithms offer insights into the potential for protein aggregation, enabling researchers to predict and understand aggregation tendencies more effectively.[145]

Studies of protein aggregation are increasingly relying on the development of machine learning methods. These methods harness the capabilities of artificial neural networks to detect sequential features that exhibit a strong correlation with protein aggregation. Remarkably, machine learning approaches have achieved comparable, and in some cases even superior, performance compared to the traditional predictors.[174] Algorithms such as APPNN,[175] FISH Amyloid[176] and netCSSP[177] are noteworthy examples of these machine learning-based approaches. These algorithms effectively utilize the power of neural networks to accurately identify and predict protein aggregation, offering researchers valuable tools for studying and understanding this important phenomenon.

7. Conclusions

The stability of protein-based drugs is a crucial factor in ensuring their functional integrity for therapeutic applications. Challenges related to proper conditions during manufacturing, storage and transportation as well as immune responses, proteases and physiological conditions in the human body, significantly influence protein stabilities. Protein aggregation poses a substantial risk to therapeutic efficacy as it can induce immunogenicity and diminish the effectiveness of protein products. It is important to note that the choice of stabilization strategies may depend on the specific characteristics of the protein-based drugs, including its size, complexity and intended application. Each protein may require a tailored approach to optimize stability and ensure its therapeutic efficacy and safety. The integration of multiple computational and experimental methods is necessary to address the various challenges in protein-based therapeutic development. In this context, computational methodologies provide value by accelerating experimental processes for design and optimization of therapeutic proteins. Furthermore, this integration of computational and experimental methods provides a more efficient way of reaching a deeper understanding of drug discovery processes and their Overall, the combined efforts of computational modeling and experimental approaches hold promise in advancing therapeutic protein design and implementation in clinical settings.

Author contributions

The manuscript's evolution was a collaborative effort among the authors. Mahdie Rahban initially drafted the manuscript based on insights received from Ali Akbar Saboury. Faizan Ahmad meticulously reviewed the first draft, offering insightful comments for improvement. Subsequently, Mieczyslaw A. Piatyszek, Thomas Haertlé, and Luciano Saso independently reviewed the manuscript, providing valuable corrections and comments. Finally, Mahdie Rahban and Ali Akbar Saboury amalgamated all the corrections and comments from the co-authors, shaping the final manuscript. This version underwent confirmation and approval by all authors before submission.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 X. Yang and M. G. Bartlett, Glycan analysis for protein therapeutics, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2019, **1120**, 29–40.

2 A. G. Mukherjee, U. R. Wanjari, A. V. Gopalakrishnan, P. Bradu, A. Biswas, R. Ganesan, *et al.*, Evolving strategies and application of proteins and peptide therapeutics in cancer treatment, *Biomed. Pharmacother.*, 2023, **163**, 114832.

3 J. Wu, J. K. Sahoo, Y. Li, Q. Xu and D. L. Kaplan, Challenges in delivering therapeutic peptides and proteins: A silk-based solution, *J. Controlled Release*, 2022, **345**, 176–189.

4 B. Dale, M. Cheng, K.-S. Park, H. Ü. Kaniskan, Y. Xiong and J. Jin, Advancing targeted protein degradation for cancer therapy, *Nat. Rev. Cancer*, 2021, **21**(10), 638–654.

5 M. Rahban, P. Norouzi, Z. Moosavi-Movahedi and A. A. Moosavi-Movahedi, Study of catalase reversibility during multiple injections of H2O2 using online measurement by FFT continuous cyclic voltammetry, *Mol. Catal.*, 2023, **545**, 113214.

6 G. Kustatscher, T. Collins, A.-C. Gingras, T. Guo, H. Hermjakob, T. Ideker, *et al.*, Understudied proteins: opportunities and challenges for functional proteomics, *Nat. Methods*, 2022, **19**(7), 774–779.

7 J. Zhu, N. Avakyan, A. Kakkis, A. M. Hoffnagle, K. Han, Y. Li, *et al.*, Protein Assembly by Design, *Chem. Rev.*, 2021, **121**(22), 13701–13796.

8 A. P. Davenport, C. C. G. Scully, C. de Graaf, A. J. H. Brown and J. J. Maguire, Advances in therapeutic peptides targeting G protein-coupled receptors, *Nat. Rev. Drug Discovery*, 2020, **19**(6), 389–413.

9 E. V. Gurevich and V. V. Gurevich, Therapeutic potential of small molecules and engineered proteins, *Handb. Exp. Pharmacol.*, 2014, **219**, 1–12.

10 D. J. Craik, D. P. Fairlie, S. Liras and D. Price, The future of peptide-based drugs, *Chem. Biol. Drug Des.*, 2013, **81**(1), 136–147.

11 H. Stutz, Advances and applications of capillary electromigration methods in the analysis of therapeutic and diagnostic recombinant proteins – A Review, *J. Pharm. Biomed. Anal.*, 2023, **222**, 115089.

12 X. Feng and Q. She, Chapter Eleven – Recombinant protein expression in Sulfolobus islandicus, in *Methods in Enzymology*, ed. Z. Kelman and W. B. O'Dell, Academic Press, 2021, vol. 659, pp. 275–295.

13 D. L. Hacker and F. M. Wurm. Recombinant DNA Technology for Production of Protein Therapeutics in Cultured Mammalian Cells, *Reference Module in Life Sciences*, Elsevier, 2017.

14 N. A. Baeshen, M. N. Baeshen, A. Sheikh, R. S. Bora, M. M. M. Ahmed, H. A. I. Ramadan, et al., Cell factories for insulin production, *Microb. Cell Factories*, 2014, **13**(1), 141.

15 E. Moorkens, N. Meuwissen, I. Huys, P. Declerck, A. G. Vulto and S. Simoens, The Market of Biopharmaceutical Medicines: A Snapshot of a Diverse Industrial Landscape, *Front. Pharmacol*, 2017, **8**, 314.

16 A. D. Riggs, Making, Cloning, and the Expression of Human Insulin Genes in Bacteria: The Path to Humulin, *Endocr. Rev.*, 2021, **42**(3), 374–380.

17 C. Matanguihan and P. Wu, Upstream continuous processing: recent advances in production of biopharmaceuticals and challenges in manufacturing, *Curr. Opin. Biotechnol.*, 2022, **78**, 102828.

18 K. K. Dubey, A. Kumar, A. Baldia, D. Rajput, S. Kateriya, R. Singh, et al., Biomanufacturing of glycosylated antibodies: Challenges, solutions, and future prospects, *Biotechnol. Adv.*, 2023, **69**, 108267.

19 N. K. Tripathi and A. Shrivastava, Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development, *Front. Bioeng. Biotechnol.*, 2019, **7**, 420.

20 U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, Engineering the Third Wave of Biocatalysis, *Nature*, 2012, **485**(7397), 185–194.

21 J. Nazor, J. Liu and G. Huisman, Enzyme Evolution for Industrial Biocatalytic Dascades, *Curr. Opin. Biotechnol.*, 2021, **69**, 182–190.

22 S. Khan, M. W. Ullah, R. Siddique, G. Nabi, S. Manan, M. Yousaf, et al., Role of Recombinant DNA Technology to Improve Life, *Int. J. Genomics*, 2016, **2016**, 2405954.

23 M. L. E. Lundahl, S. Fogli, P. E. Colavita and E. M. Scanlan, Aggregation of Protein Therapeutics Enhances their Immunogenicity: Causes and Mitigation Strategies, *RSC Chem. Biol.*, 2021, **2**(4), 1004–1020.

24 D. S. Pisal, M. P. Kosloski and S. V. Balu-Iyer, Delivery of Therapeutic Proteins, *J. Pharm. Sci.*, 2010, **99**(6), 2557–2575.

25 Y. Vugmeyster, X. Xu, F. P. Theil, L. A. Khawli and M. W. Leach, Pharmacokinetics and Toxicology of Therapeutic Proteins: Advances and Challenges, *World J. Biol. Chem.*, 2012, **3**(4), 73–92.

26 P. Zhang, F. Sun, C. Tsao, S. Liu, P. Jain, A. Sinclair, et al., Zwitterionic Gel Encapsulation Promotes Protein Stability, Enhances Pharmacokinetics, and Reduces Immunogenicity, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**(39), 12046–12051.

27 S. B. Ebrahimi and D. Samanta, Engineering Protein-Based Therapeutics Through Structural and Chemical Design, *Nat. Commun.*, 2023, **14**(1), 2411.

28 J. Schuster, A. Koulov, H. C. Mahler, P. Detampel, J. Huwyler, S. Singh, et al., In Vivo Stability of Therapeutic Proteins, *Pharm. Res.*, 2020, **37**(2), 23.

29 R. Qing, S. Hao, E. Smorodina, D. Jin, A. Zalevsky and S. Zhang, Protein Design: From the Aspect of Water Solubility and Stability, *Chem. Rev.*, 2022, **122**(18), 14085–14179.

30 M. Akbarian and S. H. Chen, Instability Challenges and Stabilization Strategies of Pharmaceutical Proteins, *Pharmaceutics*, 2022, **14**(11), 2533.

31 R. Zaman, R. A. Islam, N. Ibnat, I. Othman, A. Zaini, C. Y. Lee, et al., Current Strategies in Extending Half-Lives of Therapeutic Proteins, *J. Controlled Release*, 2019, **301**, 176–189.

32 A. K. Bordbar, A. A. Saboury, M. R. Housaindokht and A. A. Moosavi-Movahedi, Statistical Effects of the Binding of Ionic Surfactant to Protein, *J. Colloid Interface Sci.*, 1997, **192**(2), 415–419.

33 A. A. Saboury, M. Miroliae, M. Nemat-Gorgani and A. A. Moosavi-Movahedi, Kinetics Denaturation of Yeast Alcohol Dehydrogenase and the effect of temperature and trehalose. An isothermal microcalorimetry Study, *Thermochim. Acta*, 1999, **326**(1), 127–131.

34 N. S. Sarraf, A. A. Saboury, B. Ranjbar and A. A. Moosavi-Movahedi, Structural and Functional Changes of Bovine Carbonic Anhydrase as a Consequence of Temperature, *Acta Biochim. Pol.*, 2004, **51**(3), 665–671.

35 A. Divsalar, A. A. Saboury, H. Mansoori-Torshizi, M. I. Moghaddam, F. Ahmad and G. H. Hakimelahi, Comparative Studies on the Interaction Between Bovine β-lacto-globulin Type A and B and a New Designed Pd(II) Complex with Anti-tumor Activity at Different Temperatures, *J. Biomol. Struct. Dyn.*, 2009, **26**(5), 587–597.

36 E. T. Lemeski, G. R. Behbehani, A. A. Saboury, M. Monajjemi, R. Z. Mehrabian, M. A. Golsefidi, et al., Thermodynamic Study of the Binding of Mercury Ion to Human Growth Hormone at Different Temperatures, *J. Solution Chem.*, 2011, **40**(4), 575–586.

37 B. G. Poulson, K. Szczepski, J. I. Lachowicz, L. Jaremko, A. H. Emwas and M. Jaremko, Aggregation of Biologically Important Peptides and Proteins: Inhibition or Acceleration Depending on Protein and Metal Ion Concentrations, *RSC Adv.*, 2019, **10**(1), 215–227.

38 K. L. Zapadka, F. J. Becher, A. L. Gomes Dos Santos and S. E. Jackson, Factors Affecting the Physical Stability (Aggregation) of PeptideTherapeutics, *Interface Focus*, 2017, **7**(6), 20170030.

39 C. Whitfield, Packing in the Proteins, *Nat. Rev. Chem*, 2021, **5**(10), 670.

40 A. Schön, B. R. Clarkson, M. Jaime and E. Freire, Temperature Stability of Proteins: Analysis of Irreversible Denaturation Using Isothermal Calorimetry, *Proteins*, 2017, **85**(11), 2009–2016.

41 M. M. Zhu, M. Mollet, R. S. Hubert, Y. S. Kyung and G. G. Zhang, Industrial Production of Therapeutic Proteins: Cell Lines, Cell Culture, and Purification, *Handbook of Industrial Chemistry and Biotechnology*, 2017, pp. 1639–1669, DOI: 10.1007/978-3-319-52287-6_29.

42 W. Wang, S. Nema and D. Teagarden, Protein Aggregation Pathways and Influencing Factors, *Int. J. Pharm.*, 2010, **390**(2), 89–99.

43 Y. Chen, T. T. Mutukuri, N. E. Wilson and Q. T. Zhou, Pharmaceutical Protein Solids: Drying Technology, Solid-State Characterization and Stability, *Adv. Drug Delivery Rev.*, 2021, **172**, 211–233.

44 A. J. Castañeda Ruiz, M. A. Shetab Boushehri, T. Phan, S. Carle, P. Garidel, J. Buske, *et al.*, Alternative Excipients for Protein Stabilization in Protein Therapeutics: Overcoming the Limitations of Polysorbates, *Pharmaceutics*, 2022, **14**(12), 2575.

45 S. Bandopadhyay, N. Bandyopadhyay, P. K. Deb, C. Singh and R. K. Tekade, Chapter 12 – Preformulation Studies of Drug Substances, Protein, and Peptides: Role in Drug Discovery and Pharmaceutical Product Development, in *Dosage Form Design Considerations*, ed. R. K. Tekade, Academic Press, 2018, pp. 401–433.

46 S. J. Shire, 4 – Formulation of Proteins and Monoclonal Antibodies (mAbs), in *Monoclonal Antibodies*, ed. S. J. Shire, Woodhead Publishing, 2015, pp. 93–120.

47 K. Fatima, F. Naqvi and H. Younas, A Review: Molecular Chaperone-mediated Folding, Unfolding and Disaggregation of Expressed Recombinant Proteins, *Cell Biochem. Biophys.*, 2021, **79**(2), 153–174.

48 A. Beygmoradi, A. Homaei, R. Hemmati and P. Fernandes, Recombinant Protein Expression: Challenges in Production and Folding Related Matters, *Int. J. Biol. Macromol.*, 2023, **233**, 123407.

49 J. Jomrit, S. Suhardi and P. Summpunn, Effects of Signal Peptide and Chaperone Co-Expression on Heterologous Protein Production in Escherichia coli, *Molecules*, 2023, **28**(14), 5594.

50 M. Mamipour, M. Yousefi and M. Hasanzadeh, An Overview on Molecular Chaperones Enhancing Solubility of Expressed Recombinant Proteins with Correct folding, *Int. J. Biol. Macromol.*, 2017, **102**, 367–375.

51 M. Schlapschy and A. Skerra, Periplasmic Chaperones Used to Enhance Functional Secretion of Proteins in E. coli, *Methods Mol. Biol.*, 2011, **705**, 211–224.

52 A. Bhatwa, W. Wang, Y. I. Hassan, N. Abraham, X.-Z. Li and T. Zhou, Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications, *Front. Bioeng. Biotechnol.*, 2021, **9**, 630551.

53 R. Dingman and S. V. Balu-Iyer, Immunogenicity of Protein Pharmaceuticals, *J. Pharm. Sci.*, 2019, **108**(5), 1637–1654.

54 C. Ross, K. M. Clemmesen, M. Svenson, P. S. Sørensen, N. Koch-Henriksen, G. L. Skovgaard, *et al.*, Immunogenicity of Interferon-Beta in Multiple Sclerosis Patients: Influence of Preparation, Dosage, Dose Frequency, and Route of Administration. Danish Multiple Sclerosis Study Group, *Ann. Neurol.*, 2000, **48**(5), 706–712.

55 K. Tamshen, Y. Wang, S. M. F. Jamieson, J. K. Perry and H. D. Maynard, Genetic Code Expansion Enables Site-Specific PEGylation of a Human Growth Hormone Receptor Antagonist through Click Chemistry, *Bioconjugate Chem.*, 2020, **31**(9), 2179–2190.

56 J. K. Dozier and M. D. Distefano, Site-Specific PEGylation of Therapeutic Proteins, *Int. J. Mol. Sci.*, 2015, **16**(10), 25831–25864.

57 A. Fontana, B. Spolaore, A. Mero and F. M. Veronese, Site-Specific Modification and PEGylation of Pharmaceutical Proteins Mediated by Transglutaminase, *Adv. Drug Delivery Rev.*, 2008, **60**(1), 13–28.

58 B. Ma, X. Guan, Y. Li, S. Shang, J. Li and Z. Tan, Protein Glycoengineering: An Approach for Improving Protein Properties, *Front. Chem.*, 2020, **8**, 622.

59 S. Ramazi and J. Zahiri, Posttranslational Modifications in Proteins: Resources, Tools and Prediction Methods. Database, *Database: J. Biol. Databases Curation*, 2021, baab012.

60 D. K. Shah, Pharmacokinetic and Pharmacodynamic Considerations for the Next Generation Protein Therapeutics, *J. Pharmacokinet. Pharmacodyn.*, 2015, **42**(5), 553–571.

61 J. Tibbitts, D. Canter, R. Graff, A. Smith and L. A. Khawli, Key Factors Influencing ADME Properties of Therapeutic Proteins: A Need for ADME Characterization in Drug Discovery and Development, *mAbs*, 2016, **8**(2), 229–245.

62 L. Diao and B. Meibohm, Tools for Predicting the PK/PD of Therapeutic Proteins, *Expert Opin. Drug Metab. Toxicol.*, 2015, **11**(7), 1115–1125.

63 C. Li, S. C. Chen, Y. Chen, S. Girish, M. Kaagedal, D. Lu, *et al.*, Impact of Physiologically Based Pharmacokinetics, Population Pharmacokinetics and Pharmacokinetics/Pharmacodynamics in the Development of Antibody-Drug Conjugates, *J. Clin. Pharmacol.*, 2020, **60**, 105.

64 W. Wang and H. Zhou, Pharmacological Considerations for Predicting PK/PD at the Site of Action for Therapeutic Proteins, *Drug Discovery Today: Technol.*, 2016, **21–22**, 35–39.

65 C. Mueller, U. Altenburger and S. Mohl, Challenges for the Pharmaceutical Technical Development of Protein Coformulations, *J. Pharm. Pharmacol.*, 2018, **70**(5), 666–674.

66 W. Wang, S. K. Singh, N. Li, M. R. Toler, K. R. King and S. Nema, Immunogenicity of Protein Aggregates Concerns and Realities, *Int. J. Pharm.*, 2012, **431**(1–2), 1–11.

67 A. Srivastava, K. M. G. Mallela, N. Deorkar and G. Brophy, Manufacturing Challenges and Rational Formulation Development for AAV Viral Vectors, *J. Pharm. Sci.*, 2021, **110**(7), 2609–2624.

68 M. D. Swanson, S. Rios, S. Mittal, G. Soder and V. Jawa, Immunogenicity Risk Assessment of Spontaneously

Occurring Therapeutic Monoclonal Antibody Aggregates, *Front. Immunol.*, 2022, **13**, 915412.

69 C. J. Roberts, Therapeutic Protein Aggregation: Mechanisms, Design, and Control, *Trends Biotechnol.*, 2014, **32**(7), 372–380.

70 E. M. Jahn and C. K. Schneider, How to Systematically Evaluate Immunogenicity of Therapeutic Proteins – Regulatory Considerations, *New Biotechnol.*, 2009, **25**(5), 280–286.

71 J. S. Bee, T. J. Goletz and J. A. Ragheb, The Future of Protein Particle Characterization and Understanding its Potential to Diminish the Immunogenicity of Biopharmaceuticals: a Shared Perspective, *J. Pharm. Sci.*, 2012, **101**(10), 3580–3585.

72 I. Büttel, K. Völler and C. Schneider, Immunogenicity and Its Impact on Benefit/Risk Considerations in the Authorisation of Biopharmaceuticals, *Curr. Drug Saf.*, 2010, **5**(4), 287–292.

73 G. Shankar, E. Shores, C. Wagner and A. Mire-Sluis, Scientific and Regulatory Considerations on the Immunogenicity of Biologics, *Trends Biotechnol.*, 2006, **24**(6), 274–280.

74 C. Thorlaksen, H. S. Schultz, S. K. Gammelgaard, W. Jiskoot, N. S. Hatzakis, F. S. Nielsen, *et al.*, In Vitro and In Vivo Immunogenicity Assessment of Protein Aggregate Characteristics, *Int. J. Pharm.*, 2023, **631**, 122490.

75 M. Ahmadi, C. J. Bryson, E. A. Cloake, K. Welch, V. Filipe, S. Romeijn, *et al.*, Small Amounts of Sub-Visible Aggregates Enhance the Immunogenic Potential of Monoclonal Antibody Therapeutics, *Pharm. Res.*, 2015, **32**(4), 1383–1394.

76 G. Kijanka, J. S. Bee, S. A. Korman, Y. Wu, L. K. Roskos, M. A. Schenerman, *et al.*, Submicron Size Particles of a Murine Monoclonal Antibody Are More Immunogenic Than Soluble Oligomers or Micron Size Particles Upon Subcutaneous Administration in Mice, *J. Pharm. Sci.*, 2018, **107**(11), 2847–2859.

77 T. Uchio, M. Baudyš, F. Liu, S. C. Song and S. W. Kim, Site-Specific Insulin Conjugates with Enhanced Stability and Extended Action Profile, *Adv. Drug Delivery Rev.*, 1999, **35**(2), 289–306.

78 V. Rombach-Riegraf, A. C. Karle, B. Wolf, L. Sordé, S. Koepke, S. Gottlieb, *et al.*, Aggregation of Human Recombinant Monoclonal Antibodies Influences the Capacity of Dendritic Cells to Stimulate Adaptive T-cell Responses In Vitro, *PLoS One*, 2014, **9**(1), e86322.

79 A. Karle, S. Spindeldreher and F. Kolbinger, Secukinumab, A Novel Anti-IL-17A antibody, shows low immunogenicity potential in human in Vitro Assays Comparable to Other Marketed Biotherapeutics with Low Clinical Immunogenicity, *mAbs*, 2016, **8**(3), 536–550.

80 T. J. Eyes, J. I. Austerberry, R. J. Dearman, L. O. Johannissen, I. Kimber, N. Smith, *et al.*, Identification of B cell Epitopes Enhanced by Protein Unfolding and Aggregation, *Mol. Immunol.*, 2019, **105**, 181–189.

81 I. C. Büttel, P. Chamberlain, Y. Chowers, F. Ehmann, A. Greinacher, R. Jefferis, *et al.*, Taking Immunogenicity Assessment of Therapeutic Proteins to the Next Level, *Biologicals*, 2011, **39**(2), 100–109.

82 Z. E. Sauna, D. Lagassé, J. Pedras-Vasconcelos, B. Golding and A. S. Rosenberg, Evaluating and Mitigating the Immunogenicity of Therapeutic Proteins, *Trends Biotechnol.*, 2018, **36**(10), 1068–1084.

83 C. B. Anfinsen, Principles that Govern the Folding of Protein Chains, *Science*, 1973, **181**(4096), 223–230.

84 M. Rahban, S. Zolghadri, N. Salehi, F. Ahmad, T. Haertlé, N. Rezaei-Ghaleh, *et al.*, Thermal Stability Enhancement: Fundamental Concepts of Protein Engineering Strategies to Manipulate the Flexible Structure, *Int. J. Biol. Macromol.*, 2022, **214**, 642–654.

85 N. B. Pham and W. S. Meng, Protein Aggregation and Immunogenicity of Biotherapeutics, *Int. J. Pharm.*, 2020, **585**, 119523.

86 F. Shojaei, N. Tavakolinia, A. Divsalar, T. Haertlé, A. A. Saboury, M. Nemat-Gorgani, *et al.*, Biochemical and Immunological Aspects of Protein Aggregation in Neurodegenerative Diseases, *J. Iran. Chem. Soc.*, 2014, **11**(6), 1503–1512.

87 C. Nick Pace, J. M. Scholtz and G. R. Grimsley, Forces Stabilizing Proteins, *FEBS Lett.*, 2014, **588**(14), 2177–2184.

88 F. Ahmad, Protein Stability Determination Problems, *Front. Mol. Biosci.*, 2022, **9**, 880358.

89 F. Taghavi, M. Habibi-Rezaei, M. Amani, A. A. Saboury and A. A. Moosavi-Movahedi, The Status of Glycation in Protein Aggregation, *Int. J. Biol. Macromol.*, 2017, **100**, 67–74.

90 A. Fardanesh, S. Zibaie, B. Shariati, F. Attar, F. Rouhollah, K. Akhtari, *et al.*, Amorphous Aggregation of Tau in the Presence of Titanium Dioxide Nanoparticles: Biophysical, Computational, and Cellular studies, *Int. J. Nanomed.*, 2019, **14**, 901–911.

91 R. Rajan, S. Ahmed, N. Sharma, N. Kumar, A. Debas and K. Matsumura, Review of the Current State of Protein Aggregation Inhibition From A materials chemistry perspective: special focus on polymeric Materials, *Mater. Adv.*, 2021, **2**(4), 1139–1176.

92 T. L. Pukala, Mass Spectrometric Insights into Protein Aggregation, *Essays Biochem.*, 2023, **67**(2), 243–253.

93 S. Ye, C.-H. Hsiung, Y. Tang and X. Zhang, Visualizing the Multistep Process of Protein Aggregation in Live Cells, *Acc. Chem. Res.*, 2022, **55**(3), 381–390.

94 H. Ramshini, C. Parrini, A. Relini, M. Zampagni, B. Mannini, A. Pesce, *et al.*, Large Proteins Have a Great Tendency to Aggregate but a Low Propensity to Form Amyloid Fibrils, *PLoS One*, 2011, **6**(1), e16075.

95 A. Quigley and D. R. Williams, The Second Virial Coefficient as a Predictor of Protein Aggregation Propensity: A Self-Interaction Chromatography Study, *Eur. J. Pharm. Biopharm.*, 2015, **96**, 282–290.

96 H.-X. Zhou and X. Pang, Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation, *Chem. Rev.*, 2018, **118**(4), 1691–1741.

97 M. R. G. Kopp, F. Grigolato, D. Zürcher, T. K. Das, D. Chou, K. Wuchner, *et al.*, Surface-Induced Protein Aggregation and Particle Formation in Biologics: Current

Understanding of Mechanisms, Detection and Mitigation Strategies, *J. Pharm. Sci.*, 2023, **112**(2), 377–385.

98 W. M. Babinchak and W. K. Surewicz, Liquid–Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation, *J. Mol. Biol.*, 2020, **432**(7), 1910–1925.

99 H. C. Mahler, W. Friess, U. Grauschopf and S. Kiese, Protein Aggregation: Pathways, Induction Factors and Analysis, *J. Pharm. Sci.*, 2009, **98**(9), 2909–2934.

100 R. Wälchli, P. J. Vermeire, J. Massant and P. Arosio, Accelerated Aggregation Studies of Monoclonal Antibodies: Considerations for Storage Stability, *J. Pharm. Sci.*, 2020, **109**(1), 595–602.

101 V. Joshi, T. Shivach, V. Kumar, N. Yadav and A. Rathore, Avoiding Antibody Aggregation During Processing: Establishing Hold Times, *Biotechnol. J.*, 2014, **9**(9), 1195–1205.

102 C. Siska, P. Harber and B. A. Kerwin, Shocking Data on Parcel Shipments of Protein Solutions, *J. Pharm. Sci.*, 2020, **109**(1), 690–695.

103 A. Hawe, J. C. Kasper, W. Friess and W. Jiskoot, Structural Properties of Monoclonal Antibody Aggregates Induced by Freeze-Thawing and Thermal Stress, *Eur. J. Pharm. Sci.*, 2009, **38**(2), 79–87.

104 B. Boll, J. Bessa, E. Folzer, A. Ríos Quiroz, R. Schmidt, P. Bulau, *et al.*, Extensive Chemical Modifications in the Primary Protein Structure of IgG1 Subvisible Particles Are Necessary for Breaking Immune Tolerance, *Mol. Pharm.*, 2017, **14**(4), 1292–1299.

105 E. Lopez, N. E. Scott, B. D. Wines, P. M. Hogarth, A. K. Wheatley, S. J. Kent, *et al.*, Low pH Exposure During Immunoglobulin G Purification Methods Results in Aggregates That Avidly Bind Fcγ Receptors: Implications for Measuring Fc Dependent Antibody Functions, *Front. Immunol.*, 2019, **10**, 2415.

106 B. Liu, H. Guo, J. Xu, T. Qin, L. Xu, J. Zhang, *et al.*, Acid-Induced Aggregation Propensity of Nivolumab is Dependent on the Fc, *mAbs*, 2016, **8**(6), 1107–1117.

107 H. Imamura and S. Honda, Kinetics of Antibody Aggregation at Neutral pH and Ambient Temperatures Triggered by Temporal Exposure to Acid, *J. Phys. Chem. B*, 2016, **120**(36), 9581–9589.

108 H. Wu, Y. Shen, D. Wang, H. Herrmann, R. D. Goldman and D. A. Weitz, Effect of Divalent Cations on the Structure and Mechanics of Vimentin Intermediate Filaments, *Biophys. J.*, 2020, **119**(1), 55–64.

109 G. V. Barnett, V. I. Razinkov, B. A. Kerwin, T. M. Laue, A. H. Woodka, P. D. Butler, *et al.*, Specific-Ion Effects on the Aggregation Mechanisms and Protein-Protein Interactions for Anti-Streptavidin Immunoglobulin Gamma-1, *J. Phys. Chem. B*, 2015, **119**(18), 5793–5804.

110 H. Oyama, H. Koga, T. Tadokoro, K. Maenaka, A. Shiota, M. Yokoyama, *et al.*, Relation of Colloidal and Conformational Stabilities to Aggregate Formation in a Monoclonal Antibody, *J. Pharm. Sci.*, 2020, **109**(1), 308–315.

111 R. Yousefi, S. K. Ardestani, A. A. Saboury, A. Kariminia, M. Zeinali and M. Amani, Investigation on the Surface Hydrophobicity and Aggregation Kinetics of Human Calprotectin in the Presence of Calcium, *J. Biochem. Mol. Biol.*, 2005, **38**(4), 407–413.

112 M. Amani, R. Yousefi, A. A. Moosavi-Movahedi, F. Pintus, A. Mura, G. Floris, *et al.*, Structural Changes and Aggregation Process of Cu/Containing Amine Oxidase in the Presence of 2,2,2'-Trifluoroethanol, *Protein Pept. Lett.*, 2008, **15**(5), 521–527.

113 R. John, J. Mathew, A. Mathew, C. T. Aravindakumar and U. K. Aravind, Probing the Role of Cu(II) Ions on Protein Aggregation Using Two Model Proteins, *ACS Omega*, 2021, **6**(51), 35559–35571.

114 V. Ramirez-Bello, J. Martinez-Seoane, A. Fernández-Silva and C. Amero, Zinc and Copper Ions Induce Aggregation of Human β-Crystallins, *Molecules*, 2022, **27**(9), 2970.

115 M. E. Alam, G. V. Barnett, T. R. Slaney, C. G. Starr, T. K. Das and P. M. Tessier, Deamidation Can Compromise Antibody Colloidal Stability and Enhance Aggregation in a pH-Dependent Manner, *Mol. Pharm.*, 2019, **16**(5), 1939–1949.

116 M. E. Alam, T. R. Slaney, L. Wu, T. K. Das, S. Kar, G. V. Barnett, *et al.*, Unique Impacts of Methionine Oxidation, Tryptophan Oxidation, and Asparagine Deamidation on Antibody Stability and Aggregation, *J. Pharm. Sci.*, 2020, **109**(1), 656–669.

117 K. P. Kent, C. E. Schroeder and C. Sharma, Solution pH Jump During Antibody and Fc-Fusion Protein Thaw Leads to Increased Aggregation, *J. Pharm. Anal.*, 2018, **8**(5), 302–306.

118 J. A. Ji, B. Zhang, W. Cheng and Y. J. Wang, Methionine, Tryptophan, and Histidine Oxidation in a Model Protein, PTH: Mechanisms and Stabilization, *J. Pharm. Sci.*, 2009, **98**(12), 4485–4500.

119 A. Delobel, Glycosylation of Therapeutic Proteins: A Critical Quality Attribute, *Methods Mol. Biol.*, 2021, **2271**, 1–21.

120 P. B. Lawrence, Y. Gavrilov, S. S. Matthews, M. I. Langlois, D. Shental-Bechor, H. M. Greenblatt, *et al.*, Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic and Proteolytic Stability, *J. Am. Chem. Soc.*, 2014, **136**(50), 17547–17560.

121 R. Menacho-Melgar, J. S. Decker, J. N. Hennigan and M. D. Lynch, A Review of Lipidation in the Development of Advanced Protein and Peptide therapeutics, *J. Controlled Release*, 2019, **295**, 1–12.

122 X. Chen, J. L. Zaro and W. C. Shen, Pharmacokinetics of Recombinant Bifunctional Fusion Proteins, *Expert Opin. Drug Metab. Toxicol.*, 2012, **8**(5), 581–595.

123 B. L. Duivelshof, A. Murisier, J. Camperi, S. Fekete, A. Beck, D. Guillarme, *et al.*, Therapeutic Fc-Fusion Proteins: Current Analytical Strategies, *J. Sep. Sci.*, 2021, **44**(1), 35–62.

124 V. Gupta, S. Bhavanasi, M. Quadir, K. Singh, G. Ghosh, K. Vasamreddy, *et al.*, Protein PEGylation for Cancer Therapy: Bench to Bedside, *J. Cell Commun. Signaling*, 2019, **13**(3), 319–330.

125 J. M. Harris and R. B. Chess, Effect of Pegylation on Pharmaceuticals, *Nat. Rev. Drug Discovery*, 2003, **2**(3), 214–221.

126 A. D. AlQahtani, D. O'Connor, A. Domling and S. K. Goda, Strategies for the Production of Long-Acting Therapeutics and Efficient Drug Delivery for Cancer Treatment, *Biomed. Pharmacother.*, 2019, **113**, 108750.

127 B.-M. Chen, T.-L. Cheng and S. R. Roffler, Polyethylene Glycol Immunogenicity: Theoretical, Clinical, and Practical Aspects of Anti-Polyethylene Glycol Antibodies, *ACS Nano*, 2021, **15**(9), 14022–14048.

128 N. Kommineni, A. Butreddy, V. G. S. Sainaga Jyothi and P. Angsantikul, Freeze-Drying for the Preservation of Immunoengineering Products, *iScience*, 2022, **25**(10), 105127.

129 A. Butreddy, N. Dudhipala, K. Y. Janga and R. P. Gaddam, Lyophilization of Small-Molecule Injectables: an Industry Perspective on Formulation Development, Process Optimization, Scale-Up Challenges, and Drug Product Quality Attributes, *AAPS PharmSciTech*, 2020, **21**(7), 252.

130 K. R. Ward and P. Matejtschuk, The Principles of Freeze-Drying and Application of Analytical Technologies, *Methods Mol. Biol.*, 2021, **2180**, 99–127.

131 B. S. Bhatnagar, R. H. Bogner and M. J. Pikal, Protein Stability During Freezing: Separation of Stresses and Mechanisms of Protein Stabilization, *Pharm. Dev. Technol.*, 2007, **12**(5), 505–523.

132 S. Thakral, J. Sonje, B. Munjal and R. Suryanarayanan, Stabilizers and their Interaction with Formulation Components in Frozen and Freeze-Dried Protein Formulations, *Adv. Drug Delivery Rev.*, 2021, **173**, 1–19.

133 S. Ohtake, Y. Kita and T. Arakawa, Interactions of formulation Excipients With Proteins in Solution and in the Dried state, *Adv. Drug Delivery Rev.*, 2011, **63**(13), 1053–1073.

134 M. A. Mensink, H. W. Frijlink, K. van der Voort Maarschalk and W. L. J. Hinrichs, How Sugars Protect Proteins In The Solid State and Suring Srying (review): Mechanisms of Stabilization in Relation to Stress Conditions, *Eur. J. Pharm. Biopharm.*, 2017, **114**, 288–295.

135 H. J. Lee, A. McAuley, K. F. Schilke and J. McGuire, Molecular Origins of Surfactant-Mediated Stabilization of Protein Drugs, *Adv. Drug Delivery Rev.*, 2011, **63**(13), 1160–1171.

136 M. Adler and G. Lee, Stability and Surface Activity of Lactate Dehydrogenase in Spray-Dried Trehalose, *J. Pharm. Sci.*, 1999, **88**(2), 199–208.

137 P. J. Carter, Introduction to Current and Future Protein Therapeutics: A Protein Engineering Perspective, *Exp. Cell Res.*, 2011, **317**(9), 1261–1269.

138 A. R. S. Iyengar, S. Gupta, S. Jawalekar and A. H. Pande, Protein Chimerization: A New Frontier for Engineering Protein Therapeutics with Improved Pharmacokinetics, *J. Pharmacol. Exp. Ther.*, 2019, **370**(3), 703–714.

139 R. van Vught, R. J. Pieters and E. Breukink, Site-Specific Functionalization of Proteins and their Applications to Therapeutic Antibodies, *Comput. Struct. Biotechnol. J.*, 2014, **9**(14), e201402001.

140 P. H. Tobin, D. H. Richards, R. A. Callender and C. J. Wilson, Protein Engineering: A New Frontier For Biological Therapeutics, *Curr. Drug Metab.*, 2014, **15**(7), 743–756.

141 L. Balabanova, V. Golotin, A. Podvolotskaya and V. Rasskazov, Genetically Modified Proteins: Functional Improvement And Chimeragenesis, *Bioengineered*, 2015, **6**(5), 262–274.

142 M. L. Kucera and J. P. Graham, Insulin Lispro, A New Insulin Analog, *Pharmacotherapy*, 1998, **18**(3), 526–538.

143 R. Hilgenfeld, G. Seipke, H. Berchtold and D. R. Owens, The Evolution of Insulin Glargine and Its Continuing Contribution to Diabetes Care, *Drugs*, 2014, **74**(8), 911–927.

144 F. Wang, J. M. Carabino and C. M. Vergara, Insulin Glargine: A Systematic Review of A Long-acting Insulin Analogue, *Clin. Ther.*, 2003, **25**(6), 1541–1577, discussion 39–40.

145 Z. Chen, X. Wang, X. Chen, J. Huang, C. Wang, J. Wang, *et al.*, Accelerating Therapeutic Protein Design With Computational Approaches Toward The Clinical Stage, *Comput. Struct. Biotechnol. J.*, 2023, **21**, 2909–2926.

146 M. Chungyoun and J. J. Gray, AI Models for Protein Design are Driving Antibody Engineering, *Curr. Opin. Biomed. Eng.*, 2023, 100473.

147 S. S. Jafaripour, S. Gharaghani, E. Nazarshodeh, S. Haider and A. A. Saboury, In Silico Drug Repositioning of FDA-Approved Drugs to Predict New Inhibitors For Alpha-Synuclein Aggregation, *Comput. Biol. Chem.*, 2020, **88**, 107308.

148 M. C. Childers and V. Daggett, Insights From Molecular Dynamics Simulations for Computational Protein Design, *Mol. Syst. Des. Eng.*, 2017, **2**(1), 9–33.

149 M. Rahban, N. Salehi, A. A. Saboury, S. Hosseinkhani, M. H. Karimi-Jafari, R. Firouzi, *et al.*, Histidine Substitution In The Most Flexible Fragments of Firefly Luciferase Modifies Its Thermal Stability, *Arch. Biochem. Biophys.*, 2017, **629**, 8–18.

150 S. A. Hollingsworth and R. O. Dror, Molecular Dynamics Simulation for All, *Neuron*, 2018, **99**(6), 1129–1143.

151 O. M. H. Salo-Ahen, I. Alanko, R. Bhadane, A. M. J. J. Bonvin, R. V. Honorato, S. Hossain, *et al.*, Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development, *Processes*, 2021, **9**(1), 71.

152 D. Macuglia, B. Roux and G. Ciccotti, The Emergence of Protein Dynamics Simulations: How Computational Statistical Mechanics Met Biochemistry, *Eur. Phys. J. H*, 2022, **47**(1), 13.

153 A. Ganesan, M. L. Coote and K. Barakat, Molecular Dynamics-Driven Drug Discovery: Leaping Forward With Confidence, *Drug Discovery Today*, 2017, **22**(2), 249–269.

154 C. Malbranke, D. Bikard, S. Cocco, R. Monasson and J. Tubiana, Machine Learning For Evolutionary-Based And Physics-Inspired Protein Design: Current And Future Synergies, *Curr. Opin. Struct. Biol.*, 2023, **80**, 102571.

155 A. Derry, K. A. Carpenter and R. B. Altman, Training Data Composition Affects Performance of Protein Structure Analysis Algorithms, *Pac. Symp. Biocomput.*, 2022, **27**, 10–21.

156 Z. Hou, Y. Yang, Z. Ma, K.-c Wong and X. Li, Learning The Protein Language of Proteome-Wide Protein-Protein Binding Sites Via Explainable Ensemble Deep Learning, *Commun. Biol.*, 2023, **6**(1), 73.

157 Y. Kwon, S. Park, J. Lee, J. Kang, H. J. Lee and W. Kim, BEAR: A Novel Virtual Screening Method Based on Large-Scale Bioactivity Data, *J. Chem. Inf. Model.*, 2023, **63**(5), 1429–1437.

158 A. Dhasmana, S. Dhasmana, S. Agarwal, S. Khan, S. Haque, M. Jaggi, *et al.*, Integrative Big Transcriptomics Data Analysis Implicates Crucial Role of MUC13 in Pancreatic Cancer, *Comput. Struct. Biotechnol. J.*, 2023, **21**, 2845–2857.

159 I. V. Korendovych, Rational and Semirational Protein Design, *Methods Mol. Biol.*, 2018, **1685**, 15–23.

160 A. Phintha and P. Chaiyen, Rational and Mechanistic Approaches for Improving Biocatalyst Performance, *Chem Catal.*, 2022, **2**(10), 2614–2643.

161 S. Bansal and B. Kundu, Chapter 39 – Protein engineering: Methods and applications, in *Advances in Protein Molecular and Structural Biology Methods*, ed. T. Tripathi and V. K. Dubey, Academic Press, 2022, pp. 641–668.

162 E. C. Alley, G. Khimulya, S. Biswas, M. AlQuraishi and G. M. Church, Unified Rational Protein Engineering With Sequence-Based Deep Representation Learning, *Nat. Methods*, 2019, **16**(12), 1315–1322.

163 S. Wang, H. Tang, Y. Zhao and L. Zuo, BayeStab: Predicting Effects of Mutations on Protein Stability With Uncertainty Quantification, *Protein Sci.*, 2022, **31**(11), e4467.

164 H. Cao, J. Wang, L. He, Y. Qi and J. Z. Zhang, DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks, *J. Chem. Inf. Model.*, 2019, **59**(4), 1508–1514.

165 N. Chennamsetty, V. Voynov, V. Kayser, B. Helk and B. L. Trout, Design of Therapeutic Proteins With Enhanced Stability, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**(29), 11937–11942.

166 G. De Baets, J. Reumers, J. Delgado Blanco, J. Dopazo, J. Schymkowitz and F. Rousseau, An Evolutionary Trade-Off Between Protein Turnover Rate and Protein Aggregation Favors a Higher Aggregation Propensity in Fast Degrading Proteins, *PLoS Comput. Biol.*, 2011, **7**(6), e1002090.

167 O. Conchillo-Solé, N. S. de Groot, F. X. Avilés, J. Vendrell, X. Daura and S. Ventura, Aggrescan: a Server For The Prediction and Evaluation of "Hot Spots" of Aggregation in Polypeptides, *BMC Bioinf.*, 2007, **8**(1), 65.

168 G. G. Tartaglia and M. Vendruscolo, The Zyggregator Method for Predicting Protein Aggregation Propensities, *Chem. Soc. Rev.*, 2008, **37**(7), 1395–1401.

169 J. Santos, J. Pujols, I. Pallarès, V. Iglesias and S. Ventura, Computational Prediction of Protein Aggregation: Advances in Proteomics, Conformation-Specific Algorithms And Biotechnological Applications, *Comput. Struct. Biotechnol. J.*, 2020, **18**, 1403–1413.

170 I. Walsh, F. Seno, S. C. E. Tosatto and A. Trovato, PASTA 2.0: An Improved Server for Protein Aggregation Prediction, *Nucleic Acids Res.*, 2014, **42**(W1), W301.

171 C. W. O'Donnell, J. Waldispühl, M. Lis, R. Halfmann, S. Devadas, S. Lindquist, *et al.*, A Method For Probing The Mutational Landscape Of Amyloid Structure, *Bioinformatics*, 2011, **27**(13), i34–i42.

172 S. O. Garbuzynskiy, M. Y. Lobanov and O. V. Galzitskaya, FoldAmyloid: A Method of Prediction of Amyloidogenic Regions From Protein Sequence, *Bioinformatics*, 2010, **26**(3), 326–332.

173 S. Maurer-Stroh, M. Debulpaep, N. Kuemmerer, M. L. de la Paz, I. C. Martins, J. Reumers, *et al.*, Exploring The Sequence Determinants of Amyloid Structure Using Position-Specific Scoring Matrices, *Nat. Methods*, 2010, **7**(3), 237–242.

174 J. Stanislawski, M. Kotulska and O. Unold, Machine Learning Methods Can Replace 3D Profile Method in Classification of Amyloidogenic Hexapeptides, *BMC Bioinf.*, 2013, **14**(1), 21.

175 C. Família, S. R. Dennison, A. Quintas and D. A. Phoenix, Prediction of Peptide and Protein Propensity for Amyloid Formation, *PLoS One*, 2015, **10**(8), e0134679.

176 P. Gasior and M. Kotulska, Fish Amyloid – A New Method For Finding Amyloidogenic Segments in Proteins Based on Site Specific Cooccurence of Aminoacids, *BMC Bioinf.*, 2014, **15**(1), 54.

177 C. Kim, J. Choi, S. J. Lee, W. J. Welsh and S. Yoon, NetCSSP: Web Application For Predicting Chameleon Sequences And Amyloid Fibril Formation, *Nucleic Acids Res.*, 2009, **37**, W469.