

Engineering protein-based therapeutics through structural and chemical design

Received: 3 November 2022

Sasha B. Ebrahimi¹✉ & Devleena Samanta²✉

Accepted: 5 April 2023

Published online: 27 April 2023

 Check for updates

Protein-based therapeutics have led to new paradigms in disease treatment. Projected to be half of the top ten selling drugs in 2023, proteins have emerged as rivaling and, in some cases, superior alternatives to historically used small molecule-based medicines. This review chronicles both well-established and emerging design strategies that have enabled this paradigm shift by transforming protein-based structures that are often prone to denaturation, degradation, and aggregation in vitro and in vivo into highly effective therapeutics. In particular, we discuss strategies for creating structures with increased affinity and targetability, enhanced in vivo stability and pharmacokinetics, improved cell permeability, and reduced amounts of undesired immunogenicity.

In 1982, Humulin became the first FDA-approved, recombinant, protein-based therapeutic¹. Today, protein-based drugs constitute a market approaching \$400 billion with hundreds of candidates approved and in clinical trials². Here, we chronicle the chemical design strategies that have transformed the field of medicine from one that has been historically dominated by small molecule pharmaceuticals to one where proteins have emerged as comparable or superior rivals.

Proteins represent the most versatile class of biomolecules, acting as catalysts, signaling molecules, molecular and ion transporters, scaffolds for maintaining cellular and tissue integrity, receptors, and more³. Therefore, as medicines, proteins can be used to serve any of these roles (Box 1) and offer several advantages over small molecule drugs. Proteins are less likely to cause side effects by interfering with normal biological processes because they have evolved to play highly specific roles. Typically, protein therapeutics show high potency and can also execute more complex functions owing to their intricate three-dimensional structures.

Over the last ~150 years, fundamental advances have been made towards harnessing the power of proteins to create new medicines (Fig. 1). The development of an antibody-based treatment for diphtheria in 1891 was a major milestone and was awarded the first Nobel Prize in medicine⁴. The antibody was extracted from the serum of horses that had been challenged with an attenuated form of diphtheria-causing bacteria. Two decades later, the extraction of insulin from porcine pancreas for the treatment of diabetes mellitus marked another significant advance—the use of an exogenous protein

to treat an endogenous deficiency^{5,6}. However, two tons of pig pancreas were required to produce eight ounces of protein⁷. Assuming 72 million patients (a close approximation for the number of diabetics in the world requiring insulin) and the need for 0.02 ounces of insulin per year for the average patient, this would necessitate ~150 million pigs/yr, making it challenging to produce the drug at appropriate scale^{8,9}. Moreover, the use of proteins extracted from animals could elicit an immune response in patients and lead to potential exposure to animal diseases. A breakthrough moment came in 1982 when recombinant DNA technology was used to produce insulin in a bacterial host (*E. Coli*)¹⁰. The successful use of recombinant DNA technology helped to circumvent challenges with both scale-up and immunogenicity of animal-derived proteins. Considering that, in principle, one could express any protein for which its associated gene is known, a viable competitor to the workhorse of medicine at the time – small molecules – had truly emerged.

Still, there were intrinsic challenges that needed to be solved due to the inherent susceptibility of proteins to aggregation, degradation, denaturation, and concomitant loss of activity¹¹. Moreover, untimely clearance from the body, non-specific distribution, immunogenicity, and toxicity also posed relevant concerns¹². Advances in rational design and ability to deliberately introduce chemical and structural modifications have driven a paradigm shift in how these properties can be tuned^{13–15}. In the following sections, we delineate the key considerations that drive the chemical design of protein-based therapeutics and describe the various design strategies that have led to

¹Drug Product Development—Steriles, GlaxoSmithKline, Collegeville, PA 19426, USA. ²Department of Chemistry, The University of Texas at Austin, Austin, TX 78712, USA. ✉ e-mail: sasha.b.ebrahimi@gsk.com; dsamanta@utexas.edu

BOX 1

Common classes of protein-based therapeutics

Antibodies

Antibodies are produced by B lymphocyte cells of the immune system in response to foreign objects, such as invading pathogens. They function by binding to specific molecules on the pathogen's surface (antigens) and inactivating the invader. As therapeutics, antibodies are predominantly used to bind target antigens which can then block specific signaling pathways (e.g. nivolumab) or induce cell death (e.g. binutuzumab). Antibodies can also serve as vehicles for the targeted delivery of potent cytotoxic drugs to diseased cells (e.g. antibody-drug conjugates such as trastuzumab emtansine). Monoclonal antibodies (mAbs), generated using clones of a unique lymphocyte cell, constitute the most prevalent type of antibody-based therapeutics. Structurally, antibodies are Y-shaped, comprising two antigen-binding fragments (Fab) and a crystallizable fragment (Fc).

Enzymes

Enzymes play a key role in the human body by catalyzing a wide variety of biochemical reactions. As therapeutics, they can be delivered to replace a deficient or absent enzyme in so called enzyme replacement therapy. In other cases, they can be used as agents to catalyze the degradation, cleavage, or chemical modification of therapeutically relevant targets. Examples of enzyme-based drugs include PEG-asparaginase, sacrosidase, pegvaliase, and laronidase.

Coagulation factors

Coagulation factors are naturally occurring proteins that play an essential role in the blood clotting process. Consequently, as therapeutics, these proteins can be plasma-derived or produced recombinantly and administered to patients with a deficiency in native levels. Coagulation factor-based medicines are used for various bleeding-related indications, including hemophilia A and B (e.g. eptacog alfa).

Protein hormones

Protein hormones constitute a class of molecules that are secreted by endocrine glands. They act as chemical messengers by binding to receptors which then triggers a signaling cascade and leads to a physiological response. Insulin represents a classic example. After a meal, insulin facilitates glucose removal from the bloodstream by binding to its cell-surface receptor. This initiates an intracellular cascade that results in translocation of glucose transporters to the surface and subsequent uptake of glucose into cells. Other therapeutically-relevant protein hormones include erythropoietin and gonadotropin.

Cytokines

Cytokines comprise a broad group of proteins (encompassing molecules such as interleukins, interferons, and colony-stimulating factors) that mediate cell-to-cell communication during immune responses. This contrasts with protein hormones, which largely regulate the endocrine system (vide supra). Therapeutic cytokines can be used as immunomodulatory agents for a variety of indications, including multiple sclerosis (interferon β -1b) and hairy cell leukemia (interferon α -2b).

-350 approved drug products (Supplementary Tables S1–S3). We end with an outlook on areas that need further development to realize the full potential of these molecules. We particularly focus on deliberate structural and chemical modifications that are made directly to the protein structure as opposed to strategies such as directed evolution and protein encapsulation, or those focusing on tuning protein formulations. We also exclude vaccines.

Key considerations driving the chemical design of protein-based therapeutics

To function as effective therapeutics, proteins must have certain characteristics (Fig. 2). These desirable attributes vary from protein to protein based on the application of interest. A key consideration is the stability of the proteins both under storage conditions and in vivo. Many proteins are susceptible to aggregation, degradation (e.g., via deamidation/oxidation in vitro, through proteases in vivo, etc.), and denaturation which can significantly reduce efficacy^{16,17}. Some proteins are sensitive to moderate changes in temperature, which is an added concern for their transport and storage in different locations¹¹. Moreover, the residues at the surface of certain proteins can display favorable interactions with container surfaces, resulting in adsorption and reducing the concentration of the active ingredient available for therapeutic action¹⁸. Therefore, several structural modifications (e.g., site-specific mutations¹⁹ and PEGylation²⁰) are aimed at improving the solubility and stability of proteins.

Protein-based therapeutics must also exhibit appropriate pharmacokinetics and pharmacodynamics for optimal function. For example, a protein that is rapidly cleared from the body is not suitable for applications that require sustained action¹⁶. On the other hand, a protein that circulates for a very long time may cause side effects.

Similarly, protein drugs must not interfere with the body's natural functions or elicit unwanted immune responses. To address these challenges, various strategies including PEGylation²⁰, glycosylation²¹, lipidation²², and protein fusion³ have been developed.

Another important consideration in the development of protein-based medicines is targetability. The ability to target drugs to specific tissues (e.g., cancer cells) or organs (e.g., brain) is highly desirable both from the standpoint of lowering therapeutic dose as well as reducing side effects. For example, antibody-drug conjugates can be used to direct the drug to cells that express a specific receptor as the antigen. Many protein-based structures are prone to sequestration in the liver, kidney, and spleen²³. Strategies wherein targeting moieties are incorporated to promote distribution outside of these organs are an active area of investigation²⁴. Recently, it has been shown that proteins covalently conjugated to multiple copies of the transferrin aptamer show preferential accumulation in the brain relative to native proteins²⁵.

The majority of current protein-based pharmaceuticals have extracellular targets. The relatively large sizes compared to small-molecule drugs and heterogeneous surface charges render most proteins impermeable to the cell-membrane. Intracellular delivery is highly sought after as the ability to intercept deleterious intracellular processes can lead to highly effective medicines. Emerging chemical strategies such as appending cell-penetrating peptides²⁶, supercharging²⁷, and dense DNA grafting²⁸ have shown promising results in this regard.

While chemical modifications to protein structures can impart several advantageous properties, they can also alter the activity or potency of the drugs²⁰. Typically, high activity or potency is a desirable trait as it can lower the therapeutic dose necessary. The central

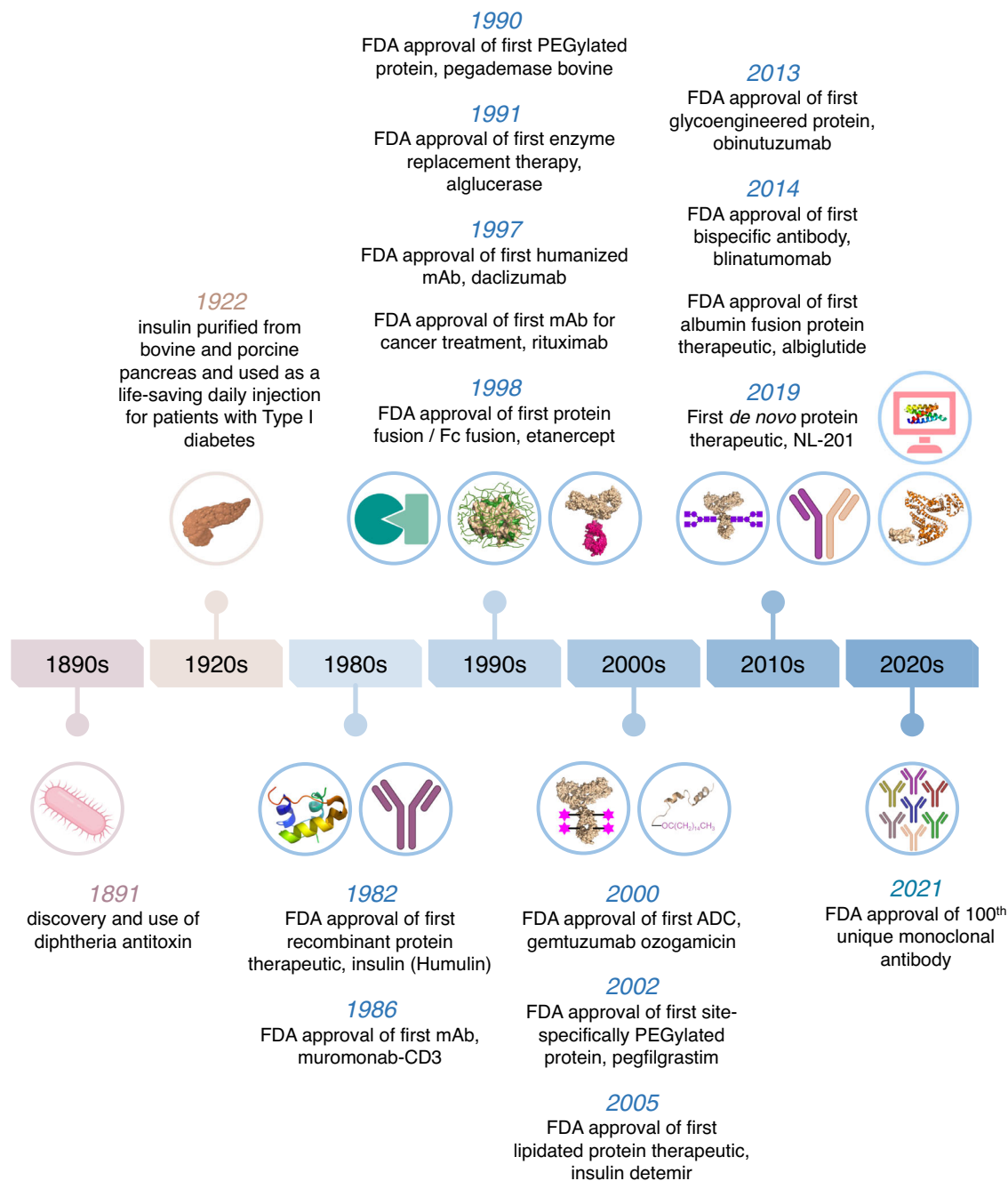


Fig. 1 | A timeline of significant advances in the development of protein-based therapeutics.

challenge in designing protein-based therapeutics is to choose a strategy that leads to the gain of a certain function without causing the loss of another. Table 1 summarizes the ability of established and emerging chemical design strategies (Fig. 3) to achieve this balance.

Established chemical design strategies

Site-specific mutagenesis. The use of site-specific mutagenesis to introduce amino acid point mutations has been a widely employed method to confer enhanced properties to protein-based therapeutics²⁹. A classic example of this has been in the development of insulin variants with different kinetics of action. (Supplementary Table S1). For instance, the substitution of asparagine by glycine at amino acid 21 of the α chain and the addition of 2 arginines to the β chain gives rise to insulin glargine, a long-acting variant with duration of action up to 24 h³⁰. These amino acid modifications

increase the isoelectric point (pI) of the structure towards physiological pH, resulting in precipitation upon injection and therefore a decrease in absorption rate²⁹. In other cases, substitutions can be made that decrease self-association and increase the rate of absorption³⁰. Insulin glulisine, for example, has a modified amino acid sequence wherein β chain asparagine (position 3) and lysine (position 29) are exchanged with lysine and glutamic acid, respectively³¹. The subsequent decrease in pI from 5.5 (native insulin) to 5.1 promotes increased solubility with less propensity for hexamer formation, leading to fast-acting effect.

Targeted mutations can also be used for tuning protein stability in solution, resulting in more robust formulations less prone to unwanted aggregation. A well-established modification involves the substitution of free cysteines with other amino acids such as serine to prevent the formation of non-native disulfide bonds or oxidation of

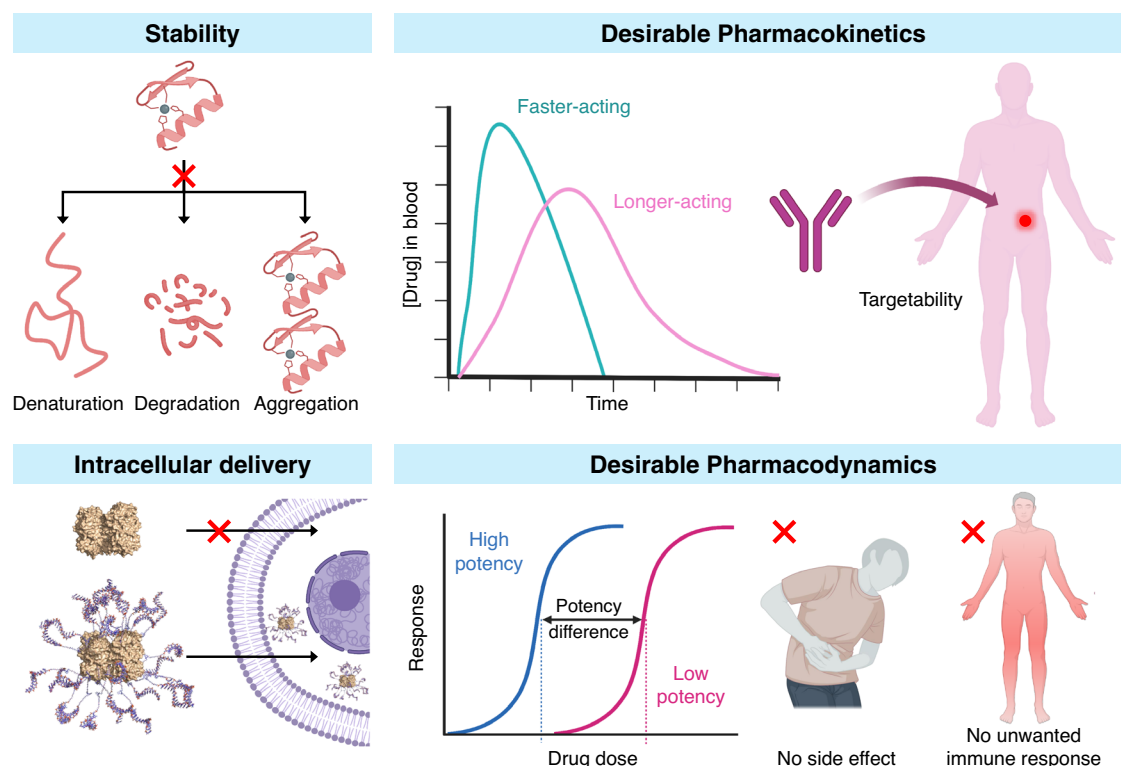


Fig. 2 | Desirable characteristics of protein-based therapeutics. These include high stability (top left), ability to enter cells (bottom left), and appropriate pharmacokinetics (top right) and pharmacodynamics (bottom right).

cysteine residues leading to thiol adducts or cysteic acid³². This strategy has been employed in various FDA-approved protein therapeutics, including aldesleukin (Proleukin), interferon β 1b (Betaseron), and pegfilgrastim (Neulasta). To explore a more general strategy, Trout et al. developed spatial aggregation propensity (SAP), a molecular dynamics-based simulation technique for identifying key regions in proteins that drive aggregation³³. Based on this information, they were able to make specific mutations that enhanced stability compared to the native protein.

Validated point mutations are also commonly used to control IgG-based antibody behavior in vivo. For instance, circulation half-life can be tuned by introducing substitutions into the Fc region that change the nature of binding interactions with the neonatal Fc receptor (FcRn), a receptor present in the endosomes of a variety of cell types. Specifically, promoting binding at endosomal pH (<6.5) leads to the antibody being trafficked to the cell surface with the FcRn rather than being sent to the lysosome for degradation. At the cell surface, the increase in pH results in loss of binding affinity and release of the antibody into circulation. Taken together, this “recycling” mechanism mediated by FcRn plays a key role in dictating the circulation half-life of antibodies. By tuning the binding strength of Fc to FcRn, circulation half-life can be increased or decreased³⁴. Fc domains with the amino acid substitutions M428L/N434S (LS variant) and M252Y/S254T/T256E (YTE variant) constitute two common examples for such modifications^{34,35}. Notably, the LS variant has been used in the FDA-approved ravulizumab (Ultomiris) to increase circulation half-life in comparison to the parent antibody, eculizumab (Soliris)³⁶. Furthermore, the discovery of new advantageous mutations remains an active area of research. Beyond tuning half-life, Fc mutations can also be made for either increasing or decreasing antibody effector function (e.g. antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), complement-dependent cytotoxicity (CDC)). A notable example is the fusion protein abatacept (Orencia) for arthritis, where C220S/C226S/C229S/P238S

mutations in the linked Fc moiety led to significant decrease of cytotoxicity associated with CDC and ADCC³⁷.

The main challenge with introducing site-specific mutations is that the mutation may negatively impact structure (i.e. prevent proper protein folding, decrease conformational/colloidal stability) or lead to decreased protein function. Furthermore, in some cases, it can be difficult to predict a priori how mutations might affect protein stability and function.

Antibody-drug conjugates (ADCs). ADCs constitute an emerging class of medicines that are generally used in the treatment of various forms of cancer. These structures consist of an antibody conjugated to a cytotoxic drug via a linker module. Attachment of the cytotoxic drug to the antibody is commonly achieved via modification of reactive residues such as cysteines or lysines. ADCs use the antibody component as the targeting moiety (e.g., to bind an antigen present in high abundance on cancer cells) to achieve cell-selective delivery of potent drug payloads that would be too toxic for administration on their own. This method increases the payload’s therapeutic index³⁸. The linker plays the critical role of keeping the cytotoxic payload conjugated to the antibody while in circulation as premature release could result in significant toxicity³⁸. Once at the target site (e.g. within a cancer cell), the linker should release the payload to allow for therapeutic effect to take place. While trials in patients began in the 1980s, challenges such as in vivo immunogenicity, in vitro instability, and lack of sufficient potency historically made ADC translation into the clinic difficult¹⁵. Continuous advances in knowledge in these challenging areas have resulted in the approval of 12 ADCs over the last -11 years (Supplementary Table 2)³⁸. A particularly interesting example is Kadcyla, approved by the FDA in 2013. Kadcyla consists of the FDA-approved trastuzumab (Herceptin) conjugated to the chemotherapeutic emtansine, illustrating the utility of repurposing a clinically validated antibody as the targeting element in the design of ADCs. Moreover, early efforts in ADC development utilized drug payloads that

Table 1 | Structural and chemical design strategies and their impact on the properties of protein-based therapeutics

	Stability			Desirable pharmacokinetics	Targeted delivery			Potency of drug molecule	Biocompatibility	
	Aggregation	Degradation	<i>In vivo</i>		Intracellular	Cell-specific	Tissue / organ-specific		Toxicity	Unwanted immune response
Site-specific mutagenesis (including the incorporation of unnatural amino acids)	●	●	●	●	●	●	●	●	●	●
ADCs	●	●	●	●	●	●	●	●	●	●
PEGylation	●	●	●	●	●	●	●	●	●	●
Glycoengineering	●	●	●	●	●	●	●	●	●	●
Fc fusion	●	●	●	●	●	●	●	●	●	●
Fusion to other proteins	●	●	●	●	●	●	●	●	●	●
Lipidation	●	●	●	●	●	●	●	●	●	●
Attachment of non-PEG polymers	●	●	●	●	●	●	●	●	●	●
Supercharging	●	●	●	●	●	●	●	●	●	●
Computational design	●	●	●	●	●	●	●	●	●	●

● strategy is typically used for improving this property

● strategy is not commonly used for improving this property or has shown promising results in some case and mixed results in others

● strategy adversely affects this property in many cases

substantially lost potency upon conjugation to antibodies. This largely arose from differences in cellular uptake mechanisms between free and attached payloads. Specifically, while a hydrophobic drug molecule on its own can diffuse into cells in large quantities, its conjugation to an antibody limits its uptake while also requiring an additional release step from the antibody in order to initiate a therapeutic effect. The use of even more potent payloads, such as auristatin-based drugs, helped overcome these challenges. Other advances, including the design of linkers that effectively release payloads intracellularly, were also pivotal in the rise of ADC therapeutics^{38,39}.

Several areas of research remain open in the ADC field. Conjugation of payloads to lysine or cysteine residues results in structures with varying drug loadings, meaning that a mixture of species is administered to patients^{40,41}. This is especially important to consider as each drug-loading variant will have a unique profile with regards to toxicity, aggregation propensity, pharmacokinetics, target affinity, and potency⁴⁰. Consequently, the exploration of strategies for site-specific conjugation of antibodies to achieve uniform drug loading remains a continuous area of work. Beyond several ADCs being tested in clinical trials, the development of new ADCs for various indications such as neuroblastoma⁴², colon cancer⁴³, and hepatocellular carcinoma⁴⁴ remains ongoing.

PEGylation. One of the most common strategies for extending the half-life of protein therapeutics is the attachment of polyethylene glycol chains (PEGylation) to the surface. PEGylation increases the overall hydrodynamic diameter. As kidneys filter molecules based on size (e.g., particles with hydrodynamic diameters below 5–6 nm are rapidly cleared), PEGylation decreases the clearance rate of the drug from the body. Moreover, PEGylation can enhance solubility, protect against proteolytic degradation or shield immunogenic epitopes from recognition by the immune system²⁰. The first PEGylated protein, adenosine deaminase (Adagen), was approved in 1990 with several more cleared by the FDA over the course of the next three decades³.

In some cases, previously approved proteins have been PEGylated to create longer acting variants that require less frequent administration. For example, turoctocog alfa (NovoEight, FDA-approved in 2013), used for treating hemophilia A, was transformed into a longer-acting drug turoctocog alfa pegol (Esperoct, FDA approval in 2019)⁴⁵. Early work in the PEGylation field largely modified proteins in a nonspecific

fashion through conjugation of amino handles on lysines and the N-terminus²⁰. Moreover, the PEG starting materials generally consist of a mixture of various length chains. This presents a challenge as the resultant heterogeneous pool of products can have vastly differing therapeutic properties. In this regard, later efforts focused on the site-specific modification of proteins towards overcoming these drawbacks⁴⁶. One example is certolizumab pegol (Cimzia), containing an engineered unpaired cysteine for conjugation to PEG⁴⁷.

PEGylation can negatively affect protein activity, in some cases with loss of up to -99% activity⁴⁸. Moreover, some reports have claimed that PEG itself can be immunogenic, which has motivated research into the use of non-PEG polymers as an alternative (*vide infra*)⁴⁹. PEGylation also raises concerns about unwanted PEG accumulation *in vivo*. To alleviate this issue, some researchers have explored the incorporation of biocleavable moieties into PEG chains^{14,50}.

Fc fusion. Fc-fusion proteins constitute a class of engineered therapeutics wherein the Fc region of an antibody is fused to a biologically active protein or peptide. Typically, Fc-fusion is done to enhance circulation half-life. The Fc moiety is able to extend circulation half-life by increasing the hydrodynamic diameter of the fusion-protein to slow kidney filtration and via interaction with the FcRn receptor. Two examples of this approach are efmorocog alfa (Eloctate) and eftrenonacog alfa (Alprolix), both of which are FDA-approved. Compared to conventional coagulation Factor VIII proteins, Eloctate exhibits -1.4–1.8 fold enhanced circulation time, leading to a less frequent requirement for patient injection⁵¹. Alprolix shows a more significant change, exhibiting -3-fold higher circulation time than unmodified Factor IX protein⁵². The addition of the Fc moiety can also help drive effector response (ADCC, ADCP, etc.) towards the creation of more potent protein therapeutics^{53,54}. Another potential benefit of Fc fusions is illustrated by etanercept (Enbrel), an FDA-approved TNF- α blocker for treating inflammatory conditions. Here, the TNF- α receptor is fused to an Fc moiety which then dimerizes due to disulfide bond formation between Fc groups. This dimerization leads to a structure that has significantly higher affinity for TNF- α compared to the receptor alone⁵⁵.

Fc fusions offer the benefit of being able to genetically encode the entire structure. This is in contrast to PEGylation, where PEG moieties are chemically conjugated after protein expression. Drawbacks include

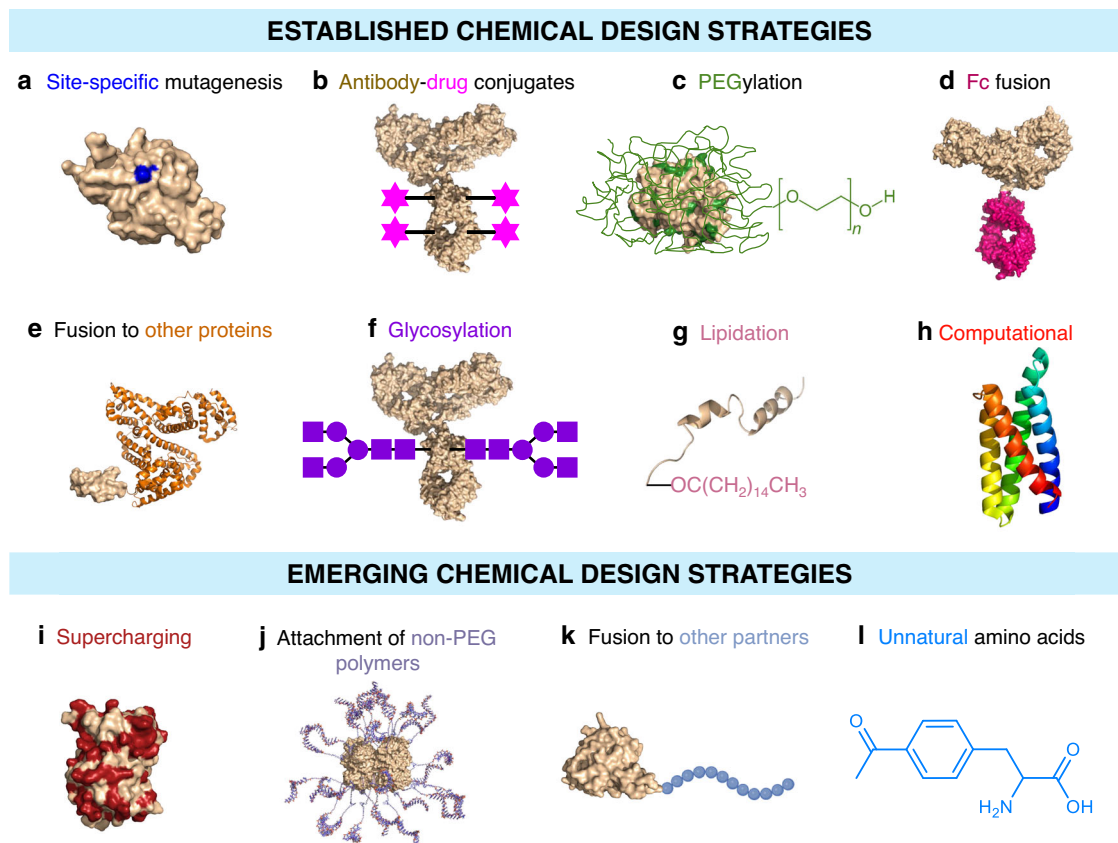


Fig. 3 | Established and emerging chemical design strategies employed to generate protein-based therapeutics. a Site-specific mutagenesis. **b** Antibody-drug conjugates. **c** PEGylation. **d** Fc fusion. **e** Fusion to other proteins. **f** Glycosylation. **g** Lipidation. **h** Computational. **i** Supercharging. **j** Attachment of

non-PEG polymers. **k** Fusion to other partners. **l** Unnatural amino acids. The structure shown in **(h)** represents neoleukin (NL)-2/15 (PDB ID: 6DG6) which is a predecessor to NL-201, the world's first de novo protein therapeutic.

the fact that incorporation of an Fc component can elicit effector response (ADCC, ADCP, etc.) even in situations where this is not desirable⁵⁶. The Fc region can also be prone to degradation by proteases in vivo.

Fusion to other proteins. A common strategy for increasing the half-life of a protein therapeutic involves fusing it to another protein with a long circulation time. Human serum albumin (HSA) has been a widely used fusion partner owing to its relatively long half-life of ~19 days⁵⁷. Moreover, conjugation to HSA increases the size of the structure reducing the rate of kidney filtration and protects its fusion partner from potential in vivo protease degradation. Taken together, HSA-fusions confer a significant improvement in the pharmacokinetic profile of a drug. A notable example of a clinical success involves albutrepenonacog alfa (Idelvion), indicated for the treatment of hemophilia B. In this case, the fusion of HSA extends the half-life from ~22 to ~102 h⁵⁸.

Fusing proteins can also lead to the creation of bispecific agents. The FDA-approved T-cell engager, blinatumomab, constitutes one such example. Here, two different single-chain variable fragments (scFv) of an antibody are linked together by a peptide moiety. One scFv recognizes CD3 on T cells, while the other scFv recognizes CD19 on malignant B cells. Therefore, simultaneous binding to both targets bridges cancer cells with activated immune cells. The T cells can secrete various enzymes, leading to cancer cell death^{59,60}.

In a similar but distinct example, the FDA-approved caplacizumab constitutes a case of genetically fusing two identical nanobodies via a short amino acid chain to create a bivalent therapeutic agent^{61,62}. Nanobodies initially garnered interest after the discovery of antibodies

in camelids that contain only heavy chains. These antibodies have a single variable domain that binds antigens. This domain, termed a nanobody, still recognizes its targets when isolated from the whole structure. With a size of around 15 kDa, these small therapeutic agents can potentially access hard to reach areas in vivo, exhibit high stability, and can be produced with relative ease compared to conventional antibodies. A general risk associated with fusion proteins is that the therapeutic protein may lose activity if the binding/active site is occluded upon conjugation⁵².

Glycoengineering. Glycosylation is a post-translational modification mediated by intracellular enzymatic machinery wherein oligosaccharide groups are covalently conjugated to the protein structure, most often at asparagine (N-linked) or serine/threonine (O-linked) residues. Certain therapeutic proteins, such as insulin, are not glycosylated and can therefore be expressed in prokaryotic hosts⁶³. However, the majority of approved structures are glycosylated. Hosts incapable of glycosylation are not suitable for their expression. In this regard, a key advance in the rise of protein therapeutics was the use of recombinant DNA technology for glycoprotein expression in appropriate host cells. It was found that “appropriate” host cells are typically mammalian – the resultant glycosylated structures were less immunogenic in humans compared to those expressed in other eukaryotic hosts⁶⁴.

Beyond influencing immunogenicity, glycosylation can also impact protein stability, in vivo activity, and pharmacokinetics. The use of glycoengineering has therefore been a well-studied strategy for tuning protein properties⁶⁵. In some cases, glycoengineering involves changing or adding glycosylation sites. One of the most well-known

examples is the erythropoietin analog darbepoetin alfa (Aranesp), engineered with two extra locations for N-linked glycosylation⁶⁶. This modification improved the half-life of the drug by ~3-fold owing to an increase in the protein's size, making kidney filtration slower.

Glycoengineering can also involve changing the identity of the conjugated oligosaccharide groups. In darbepoetin alfa, enhanced half-life is also attributed to its increased sialic acid content, making clearance mediated by asialoglycoprotein receptors (found in liver cells) less favorable⁶⁶. Glycoengineering of antibodies has also yielded several examples of structures with enhanced properties. For instance, Chinese Hamster Ovary (CHO) host cells can be programmed to express antibodies that lack fucose sugar groups (i.e. afucosylation). Afucosylation in the Fc region of antibodies can increase binding to FcγRIIIa receptors. This enhances recruitment and activation of immune effector cells, leading to enhanced ADCC or ADCP potency towards cancer cells. Mogamulizumab (Poteligeo) constitutes one example used in the clinic⁶⁶. While a powerful approach, challenges to optimizing proteins through glycoengineering remain, including difficulty in predicting a priori how changing glycopatterns will influence protein properties and loss of activity upon varying glycosylation identity and sites.

Lipidation. The conjugation of lipid groups to proteins has been utilized to enhance therapeutic properties in several approved drugs. Lipidation is often done to enhance pharmacokinetics owing to the ability of lipids to bind to HSA, thereby conferring enhanced characteristics to the structure as a whole. This technique is an alternative to direct fusion of a protein partner to a therapeutic (*vide supra*). A potential advantage offered by this strategy is that the interaction between lipids and HSA is reversible; therefore, the therapeutic protein component of the structure can elicit its effect once no longer bound and sterically blocked by the partner. In certain situations, the presence of the lipid group can also promote reversible multimer formation upon subcutaneous injection that leads to extended release in the body²². Insulin detemir was the first lipidated protein to gain FDA approval. It consists of desB30 (i.e. an insulin analog wherein amino acid 30, threonine, is not present) modified with myristic acid at LysB29⁶⁷. Lipidation promotes the formation of dihexamers upon injection that slows absorption and also increases albumin binding, leading to a half-life of ~4–7 h⁶⁸. A more recent and dramatic example of half-life increase is represented by insulin icodec, a lipidated insulin analog investigated in clinical trials that only requires once-a-week administration⁶⁹. Another benefit of lipidation involves its ability to promote intracellular delivery of proteins, which can potentially allow for targeting molecules inside of cells²². As is the case with several other modifications, one limitation of lipidation is the potential for loss of protein activity or binding strength if conjugation is close to the active/binding site. Depending on their nature, some lipids may also promote aggregation *in vitro* or elicit an immune response *in vivo*^{22,70}.

Computational design. Protein function can be augmented via computational design⁷¹. The central assumption driving computational protein design is Anfinsen's thermodynamic hypothesis—proteins fold into their minimum-energy conformation⁷². Therefore, the effect of varying amino acid sequences on structure can be systematically studied and correlated to function. Computational design can be used to optimize various chemical, physical, and pharmacological properties of therapeutic proteins^{73–76}. Specifically, computational design has been shown to improve stability, binding affinity, antibody effector activity, and immunogenicity, among others.

Recently, *de novo* protein design has emerged as an especially attractive route for designing therapeutic proteins⁷⁷. This allows proteins to be designed from scratch using the fundamental principles of protein biophysics. Whereas the majority of protein engineering focuses on enhancing the functions of existing proteins, often using

targeted mutations, *de novo* protein design focuses on proteins with amino acid sequences not found in nature. The rationale behind this strategy is as follows. A typical protein formed from 200 amino acids can have 20^{200} ($\approx 10^{260}$) different primary sequences. However, the total number of proteins found in existing organisms is on the order of 10^{12} . Therefore, a large portion of sequence space remains unexplored by evolution, and it is reasonable to imagine that sequences within this space may fold to form proteins with novel properties. Efforts in this area have led to the identification of potent mimics of cytokines interleukin-2 (IL-2) and IL-15⁷⁸, programmed cell death protein-1 (PD-1) agonists⁷⁹, and SARS-CoV-2 inhibitors⁸⁰, among others^{81,82}. A notable example is the IL-2 mimic NL-201, the world's first protein therapeutic designed *de novo* which has shown promise as an anti-cancer immunotherapeutic. Native IL-2 has limited clinical application due to its toxicity which stems from binding to the alpha chain of the IL-2 receptor. In contrast, NL-201 binds exclusively to the beta and gamma chains of the IL-2 receptor which induces the proliferation of anti-tumor effector T cells while avoiding toxicity⁷⁸. In addition, due to structural similarities between the IL-2 and IL-15 receptors, NL-201 is also able to bind to the beta and gamma chains of the IL-15 receptor which causes the expansion of anti-tumor natural killer cells^{75–79}.

The main challenge in computational protein design is navigating through the complex conformational energy landscape of proteins (that may contain many local minima) and locating desirable low-energy structures. The method also requires efficient sampling methods, computational power, and continual improvements in machine learning algorithms to accurately predict desired structures.

Emerging chemical design strategies

Supercharged proteins. Supercharged proteins constitute a group of structures containing greater than one net charge for every kilodalton of weight⁸³. These highly charged proteins can either be engineered or found in nature and offer unique properties. For example, they show strong resistance to aggregation arising from thermal or chemical stress²⁷. This means that upon unfolding due to stresses, supercharged variants can refold and exhibit significant activity even in cases where the unmodified protein cannot²⁷. Additionally, positively supercharged proteins can be taken up into cells in large amounts via binding to cell-surface proteoglycans, making them useful for applications where intracellular delivery is desired⁸⁴. Several proteins have been engineered via mutagenesis to create supercharged variants, including glutathione S-transferase and green fluorescent protein (GFP)⁸³. Importantly, supercharged proteins can be used as fusion partners to deliver other proteins intracellularly⁸³. Moreover, supercharged proteins can be used as fusion partners for functional protein delivery *in vivo*, including in retinal and pancreatic tissue^{27,85}.

Importantly, the residue used for attachment should be chosen such that the supercharged protein does not interfere with the binding or activity of the protein. Furthermore, it should be considered that the use of positively supercharged proteins may lead to toxicity or elicit an *in vivo* immune response⁸⁶.

Attachment of non-PEG polymers. Covalent modification of proteins with polymers other than PEG remains an active area of research. DNA represents a promising option in this regard, wherein dense functionalization of proteins with DNA results in structures called protein spherical nucleic acids (ProSNAs) with several advantageous properties²⁸. Firstly, the DNA shell can impart enhanced stability to the structure, both in the form of increased resistance to protease degradation and increased solubility²³. Secondly, ProSNAs show enhanced circulation times and greater distribution to areas outside of the liver in comparison to the unmodified protein⁸⁷. Finally, the dense arrangement of DNA around the protein core results in recognition by cell surface scavenger receptors, leading to robust cellular uptake up to 280-fold greater than the unmodified protein^{23,28,87}. Several reports

have utilized ProSNAs, including for the in vivo delivery of highly active proteins and for the intracellular detection of cancer biomarkers^{87,88}. One drawback of these structures is that their uptake occurs via endocytosis, meaning that a sufficient quantity must escape the endosome if therapeutic targets are in the cytosol⁸⁹.

Other fusion partners. The use of polypeptide chains as fusion partners for protein therapeutics is an active area of research^{90–92}. Compared to PEGylated structures, these constructs offer several potential advantages, including the ability to genetically encode their production to obviate the need for chemical conjugation, achieve more homogenous end products, and offer a large design space wherein peptide length and identity can be precisely tuned to produce desirable overall properties. Stemmer et al. reported an early example where they used *E. coli* to express and screen a large library of different polypeptide sequences each containing 864 amino acids⁹⁰. Candidates were assessed for factors such as genetic stability, aggregation propensity, heat sensitivity, and solubility. Based on these criteria, a candidate polypeptide termed XTEN was used as a fusion partner and its potential to impart advantageous properties was assessed. The XTEN moiety was shown to be non-immunogenic even in cases where PEGylation elicited an immune response, yielded a 12-fold increase in circulation time when fused to GFP, and had robust solubility that conferred enhanced stability to its payload partner. Notably, the half-life was highly dependent on the length of the XTEN sequence, offering a tunable way to change pharmacokinetic properties. Efanesoctocog Alfa, a protein-based therapeutic employing the XTEN technology, was recently given breakthrough therapy designation by the FDA⁹³. Several other examples of polypeptide fusion-inspired strategies exist, including the use of superhydrophilic zwitterionic peptides for enhancing circulation time and stability^{94,95}. The use of polypeptides also offers an additional advantage over PEG owing to their inherent biodegradability. In other work, especially where intracellular delivery is desired, fusion to cell penetrating peptides has been explored as a viable strategy^{26,96}. It should however be noted that immunogenicity can be a concern with certain peptide partners²³.

The use of other fusion partners can endow structures with interesting properties related to targeting. For instance, Lee et al. showed that proteins modified with tannic acid exhibit prolonged circulation time and appreciable accumulation in the heart²⁴. Similarly, Zuchero et al. engineered Fc fragments as fusion partners that can bind the transferrin receptor and consequently transport proteins across the blood brain barrier⁹⁷.

Unnatural amino acids (UAAs). Incorporation of UAAs into protein structures is a particularly interesting area of work as they can be placed site-specifically, and their biorthogonal functional groups can be leveraged for chemical conjugation. For instance, this has allowed for controlled site-specific conjugation of PEG molecules^{98,99}. In one case, Cho et al. incorporated an UAA with an acetyl functional group for reaction with oxime-functionalized PEG¹⁰⁰. In their study, they screened a set of structures, each with a single UAA at one of 6 unique positions. A 3-fold range in activity of the conjugated protein was observed depending on where the PEG was attached, showing the advantage of being able to control the position of PEG modification. In similarly inspired examples, the use of unnatural amino acids has enabled the site-specific installation of ADC payloads^{40,101}, lipid groups¹⁰², and fusion proteins¹⁰³. In another case, Sullivan et al. incorporated UAAs into proteins in a clustered fashion promoting enhanced and targeted cellular uptake of a therapeutic enzyme (yeast cytosine deaminase)¹⁰⁴. Specifically, the group used the UAAs as chemical handles to conjugate multiple copies of epidermal growth factor receptor (EGFR)-targeting peptides to the protein surface in relatively close proximity to one another. This allowed for multivalent binding to the EGFR receptor, overexpressed on inflammatory breast cancer (IBC)

cells, and ultimately enabled the delivery of the enzyme intracellularly. In a recent example, Wang et al. developed a strategy called proximity-enabled reactive therapeutics (PERx) for generating covalent protein drugs¹⁰⁵. The authors incorporated a bioreactive amino acid fluorosulfate-L-tyrosine (FSY) into human programmed cell death protein-1 (PD-1) which could covalently attach to a proximal histidine on PD-L1 only upon PD-1-PD-L1 interaction. The resulting structures showed a dramatic increase in potency over the noncovalent wild-type PD-1¹⁰⁵.

Challenges with UAAs are similar to those observed with site-specific mutagenesis, including the potential for mutations to decrease protein stability or activity. Moreover, the susceptibility of UAAs for inefficient incorporation often results in low protein yield¹⁰⁶.

Conclusions and outlook

Protein-based therapeutics have already revolutionized medicine and are increasingly growing in scope and impact. Each therapeutic has a set of desirable properties, including conformational and colloidal stability, sufficient circulation time, high potency, and lack of toxicity. The advent of chemical design strategies that enhance these characteristics has been indispensable to the field and has catalyzed its progress. In the future, we envision the continued use of established modifications, including PEGylation, ADCs, and fusions. We also envision the sustained growth of emerging strategies that improve certain properties without negatively impacting others. In this regard, further advances in computational capabilities that enable de novo therapeutic protein design will be instrumental in creating structures that exhibit desirable properties across the board. Considering that many modification strategies lead to heterogeneous end products, further research into site-specific conjugation remains highly necessary. These efforts are particularly important as each heterogeneous end product in a complex mixture can differ in its pharmacological property including safety and efficacy. We anticipate that intracellular therapeutic targets will also be an active area of research. This will put the spotlight on strategies that can mediate the delivery of proteins into cells, such as supercharging or modifying the surface with DNA. Once in the cell, the next hurdle to overcome will be attaining organelle-level specificity for therapeutic action. The further development of structures that can reach hard to target locations in the body, like the brain, will also unlock new capabilities in the field. The chemical design of orally bioavailable proteins will be a continued area of interest, helping to develop a more convenient delivery alternative to injections¹⁰⁷. Taken together, these advances will fuel the sustained evolution of proteins in their role as essential tools in medicine.

References

1. Mathieu, C., Martens, P.-J. & Vangoitsenhoven, R. One hundred years of insulin therapy. *Nat. Rev. Endocrinol.* **17**, 715–725 (2021).
2. Chung, Y. H. et al. Integrating plant molecular farming and materials research for next-generation vaccines. *Nat. Rev. Mater.* **7**, 372–388 (2022).
3. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7**, 21–39 (2008). **This review paper established classifications for protein therapeutics based on their pharmacological action.**
4. Kaufmann, S. H. E. Emil von Behring: translational medicine at the dawn of immunology. *Nat. Rev. Immunol.* **17**, 341–343 (2017).
5. Wang, M. Y. et al. Leptin therapy in insulin-deficient type I diabetes. *Proc. Natl. Acad. Sci. USA* **107**, 4813–4819 (2010).
6. Banting, F. G. & Best, C. The internal secretion of the pancreas. *J. Lab. Clin. Med.* **7**, 251–266 (1922).
7. Gebel, E. Making Insulin: A behind-the-scenes look at producing a lifesaving medication. *Diabetes Forecast* **66**, 54–57 (2013).
8. Buzzetti, R. et al. Adult-onset autoimmune diabetes. *Nat. Rev. Dis. Prim.* **8**, 63 (2022).

9. World Health Organization. *Keeping the 100-year-old promise: making insulin access universal*. <https://www.who.int/publications/i/item/9789240039100> (2021).
10. Johnson, I. S. Human insulin from recombinant DNA technology. *Science* **219**, 632–637 (1983).
11. Frokjaer, S. & Otzen, D. E. Protein drug stability: a formulation challenge. *Nat. Rev. Drug Discov.* **4**, 298–306 (2005).
12. Zhang, P. et al. Zwitterionic gel encapsulation promotes protein stability, enhances pharmacokinetics, and reduces immunogenicity. *Proc. Natl Acad. Sci.* **112**, 12046–12051 (2015).
13. Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L. & Bernardes, G. J. L. Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **3**, 147–171 (2019).
14. Ekladios, I., Colson, Y. L. & Grinstaff, M. W. Polymer–drug conjugate therapeutics: advances, insights and prospects. *Nat. Rev. Drug Discov.* **18**, 273–294 (2019).
15. Drago, J. Z., Modi, S. & Chandrapaty, S. Unlocking the potential of antibody–drug conjugates for cancer therapy. *Nat. Rev. Clin. Oncol.* **18**, 327–344 (2021). **This is a comprehensive overview delineating the advantages and limitations of ADCs with a particular emphasis on their use for cancer treatment.**
16. Kontos, S. & Hubbell, J. A. Drug development: longer-lived proteins. *Chem. Soc. Rev.* **41**, 2686–2695 (2012).
17. Berkowitz, S. A., Engen, J. R., Mazzeo, J. R. & Jones, G. B. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat. Rev. Drug Discov.* **11**, 527–540 (2012).
18. Eu, B., Cairns, A., Ding, G., Cao, X. & Wen, Z.-Q. Direct visualization of protein adsorption to primary containers by gold nanoparticles. *J. Pharm. Sci.* **100**, 1663–1670 (2011).
19. Fryszkowska, A. et al. A chemoenzymatic strategy for site-selective functionalization of native peptides and proteins. *Science* **376**, 1321–1327 (2022).
20. Harris, J. M. & Chess, R. B. Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* **2**, 214–221 (2003). **This paper reviews how PEGylation can lead to protein-based drugs with enhanced properties for therapeutic applications.**
21. Solá, R. J. & Griebenow, K. Glycosylation of therapeutic proteins. *BioDrugs* **24**, 9–21 (2010).
22. Menacho-Melgar, R., Decker, J. S., Hennigan, J. N. & Lynch, M. D. A review of lipidation in the development of advanced protein and peptide therapeutics. *J. Control. Release* **295**, 1–12 (2019).
23. Ebrahimi, S. B., Samanta, D., Kusmierz, C. D. & Mirkin, C. A. Protein transfection via spherical nucleic acids. *Nat. Protoc.* **17**, 327–357 (2022). **This work describes the methodology to make proteins cell permeable by densely functionalizing their surface with nucleic acids.**
24. Shin, M. et al. Targeting protein and peptide therapeutics to the heart via tannic acid modification. *Nat. Biomed. Eng.* **2**, 304–317 (2018).
25. Kusmierz, C. D., Callmann, C. E., Kudruk, S., Distler, M. E. & Mirkin, C. A. Transferrin aptamers increase the in vivo blood–brain barrier targeting of protein spherical nucleic acids. *Bioconjug. Chem.* **33**, 1803–1810 (2022).
26. Teo, S. L. Y. et al. Unravelling cytosolic delivery of cell penetrating peptides with a quantitative endosomal escape assay. *Nat. Commun.* **12**, 3721 (2021).
27. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–10112 (2007). **This report shows that supercharging proteins can result in structures with extraordinary resistance to aggregation. In later work, the authors showed that these structures can also effectively transfect cells.**
28. Brodin, J. D., Sprangers, A. J., McMillan, J. R. & Mirkin, C. A. DNA-mediated cellular delivery of functional enzymes. *J. Am. Chem. Soc.* **137**, 14838–14841 (2015).
29. Marshall, S. A., Lazar, G. A., Chirino, A. J. & Desjarlais, J. R. Rational design and engineering of therapeutic proteins. *Drug Discov. Today* **8**, 212–221 (2003).
30. Vwarewijk, A. J. & Janssen, J. A. M. J. L. Insulin and its analogues and their affinities for the IGF1 receptor. *Endocr. Relat. Cancer* **19**, F63–F75 (2012).
31. Gillis, R. B. et al. Analysis of insulin glulisine at the molecular level by X-ray crystallography and biophysical techniques. *Sci. Rep.* **11**, 1737 (2021).
32. Xia, X., Longo, L. M. & Blaber, M. Mutation choice to eliminate buried free cysteines in protein therapeutics. *J. Pharm. Sci.* **104**, 566–576 (2015).
33. Chennamsetty, N., Voynov, V., Kayser, V., Helk, B. & Trout, B. L. Design of therapeutic proteins with enhanced stability. *Proc. Natl Acad. Sci. USA* **106**, 11937–11942 (2009).
34. Saunders, K. O. Conceptual approaches to modulating antibody effector functions and circulation half-life. *Front. Immunol.* **10**, 1296 (2019).
35. Loo, Y.-M. et al. The SARS-CoV-2 monoclonal antibody combination, AZD7442, is protective in nonhuman primates and has an extended half-life in humans. *Sci. Transl. Med.* **14**, eabl8124 (2022). **This is a recent example of the utility of point mutations to create drugs with extended half-lives, in this case for the creation of therapeutic or prophylactic agents for the SARS-CoV-2 virus.**
36. Lee, C. H. et al. An engineered human Fc domain that behaves like a pH-toggle switch for ultra-long circulation persistence. *Nat. Commun.* **10**, 5031 (2019).
37. Strohl, W. R. Optimization of Fc-mediated effector functions of monoclonal antibodies. *Curr. Opin. Biotechnol.* **20**, 685–691 (2009).
38. Chari, R. V. J., Miller, M. L. & Widdison, W. C. Antibody–drug conjugates: An emerging concept in cancer therapy. *Angew. Chem. Int. Ed.* **53**, 3796–3827 (2014).
39. Agarwal, P. & Bertozzi, C. R. Site-specific antibody–drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug. Chem.* **26**, 176–192 (2015).
40. Axup, J. Y. et al. Synthesis of site-specific antibody–drug conjugates using unnatural amino acids. *Proc. Natl. Acad. Sci.* **109**, 16101–16106 (2012).
41. Guo, J. et al. Characterization and higher-order structure assessment of an interchain cysteine-based ADC: Impact of drug loading and distribution on the mechanism of aggregation. *Bioconjug. Chem.* **27**, 604–615 (2016).
42. Sano, R. et al. An antibody–drug conjugate directed to the ALK receptor demonstrates efficacy in preclinical models of neuroblastoma. *Sci. Transl. Med.* **11**, eaau9732 (2019).
43. Junttila, M. R. et al. Targeting LGR5+ cells with an antibody–drug conjugate for the treatment of colon cancer. *Sci. Transl. Med.* **7**, e314ra186 (2015).
44. Kong, F.-E. et al. Targeting tumor lineage plasticity in hepatocellular carcinoma using an anti-CLDN6 antibody–drug conjugate. *Sci. Transl. Med.* **13**, eabb6282 (2021).
45. Rondon, A., Mahri, S., Morales-Yanez, F., Dumoulin, M. & Vanbever, R. Protein engineering strategies for improved pharmacokinetics. *Adv. Funct. Mater.* **31**, 2101633 (2021).
46. Ramirez-Paz, J. et al. Thiol–maleimide poly(ethylene glycol) crosslinking of L-asparaginase subunits at recombinant cysteine residues introduced by mutagenesis. *PLoS One* **13**, e0197643 (2018).
47. Park, A. et al. Carbohydrate-mediated polyethylene glycol conjugation of TSH improves its pharmacological properties. *Endocrinology* **154**, 1373–1383 (2013).
48. Podobnik, B. et al. Conjugation of PolyPEG to interferon alpha extends serum half-life while maintaining low viscosity of the conjugate. *Bioconjug. Chem.* **26**, 452–459 (2015).

49. Ko, J. H. & Maynard, H. D. A guide to maximizing the therapeutic potential of protein–polymer conjugates by rational design. *Chem. Soc. Rev.* **47**, 8998–9014 (2018).
50. Lundberg, P. et al. Poly[(ethyleneoxide)-co-(methylene ethylene oxide)]: A hydrolytically degradable poly(ethylene oxide) platform. *ACS Macro Lett.* **1**, 1240–1243 (2012).
51. Frampton, J. E. Efmoroctocog Alfa: A review in haemophilia A. *Drugs* **81**, 2035–2046 (2021).
52. Strohl, W. R. Fusion proteins for half-life extension of biologics as a strategy to make biobetters. *BioDrugs* **29**, 215–239 (2015). **This is a comprehensive review of various fusion partners that can be used to design protein therapeutics with enhanced therapeutic properties.**
53. Sioud, M., Westby, P., Olsen, J. K. E. & Mobergslie, A. Generation of new peptide-Fc fusion proteins that mediate antibody-dependent cellular cytotoxicity against different types of cancer cells. *Mol. Ther. Methods Clin. Dev.* **2**, 15043 (2015).
54. Shoji-Hosaka, E. et al. Enhanced Fc-dependent cellular cytotoxicity of Fc fusion proteins derived from TNF receptor II and LFA-3 by fucose removal from asn-linked oligosaccharides. *J. Biochem.* **140**, 777–783 (2006).
55. Mohler, K. M. et al. Soluble tumor necrosis factor (TNF) receptors are effective therapeutic agents in lethal endotoxemia and function simultaneously as both TNF carriers and TNF antagonists. *J. Immunol.* **151**, 1548–1561 (1993).
56. Shiga, Y. et al. Hinge-deficient IgG1 Fc fusion: Application to human lactoferrin. *Mol. Pharm.* **14**, 3025–3035 (2017).
57. Liu, Z. & Chen, X. Simple bioconjugate chemistry serves great clinical advances: albumin as a versatile platform for diagnosis and precision therapy. *Chem. Soc. Rev.* **45**, 1432–1456 (2016).
58. Nilsen, J. et al. An intact C-terminal end of albumin is required for its long half-life in humans. *Commun. Biol.* **3**, 181 (2020).
59. Labrijn, A. F., Janmaat, M. L., Reichert, J. M. & Parren, P. W. H. I. Bispecific antibodies: a mechanistic review of the pipeline. *Nat. Rev. Drug Discov.* **18**, 585–608 (2019).
60. Einsele, H. et al. The BiTE (bispecific T-cell engager) platform: Development and future potential of a targeted immuno-oncology therapy across tumor types. *Cancer* **126**, 3192–3201 (2020).
61. Yang, E. Y. & Shah, K. Nanobodies: Next generation of cancer diagnostics and therapeutics. *Front. Oncol.* **10**, 1182 (2020).
62. Morrison, C. Nanobody approval gives domain antibodies a boost. *Nat. Rev. Drug Discov.* **18**, 485–487 (2019).
63. Ballew, N. & Gerngross, T. Production of therapeutic proteins in fungal hosts. *Expert Opin. Biol. Ther.* **4**, 623–626 (2004).
64. Ma, B. et al. Protein glycoengineering: An approach for improving protein properties. *Front. Chem.* **8**, 622 (2020).
65. Elliott, S. et al. Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat. Biotechnol.* **21**, 414–421 (2003). **This paper is an early example of glycoengineering for realizing protein-based therapeutics with enhanced in vivo activity.**
66. Dammen-Brower, K. et al. Strategies for glycoengineering therapeutic proteins. *Front. Chem.* **10**, 863118 (2022).
67. Wu, J. et al. Expression, purification, and characterization of the Degludec precursor DesB30. *Protein Expr. Purif.* **161**, 28–39 (2019).
68. Poon, K. & King, A. B. Glargine and detemir: Safety and efficacy profiles of the long-acting basal insulin analogs. *Drug Healthc. Patient Saf.* **2**, 213–223 (2010).
69. Rosenstock, J. et al. Once-weekly insulin for type 2 diabetes without previous insulin treatment. *N. Engl. J. Med.* **383**, 2107–2116 (2020). **This report discusses the unprecedented ability of insulin icodec to treat diabetes with only once a week administration.**
70. Venanzi, M. et al. A spectroscopic and molecular dynamics study on the aggregation process of a long-acting lipidated therapeutic peptide: The case of semaglutide. *Soft Matter* **16**, 10122–10131 (2020).
71. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 (2019).
72. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
73. Lippow, S. M., Witttrup, K. D. & Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **25**, 1171–1176 (2007).
74. Barderas, R., Desmet, J., Timmerman, P., Meloen, R. & Casal, J. I. Affinity maturation of antibodies assisted by in silico modeling. *Proc. Natl Acad. Sci.* **105**, 9029–9034 (2008).
75. Steed, P. M. et al. Inactivation of TNF signaling by rationally designed dominant-negative TNF variants. *Science* **301**, 1895–1898 (2003).
76. Vivona, S. et al. Computer-aided biotechnology: from immunoinformatics to reverse vaccinology. *Trends Biotechnol.* **26**, 190–200 (2008).
77. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
78. Silva, D.-A. et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019). **This paper reports the first protein therapeutic designed de novo (IL-2 mimic, NL-201).**
79. Bryan, C. M. et al. Computational design of a synthetic PD-1 agonist. *Proc. Natl. Acad. Sci.* **118**, e2102164118 (2021).
80. Cao, L. et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
81. Cao, L. et al. Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
82. Chevalier, A. et al. Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
83. Ma, C., Malessa, A., Boersma, A. J., Liu, K. & Herrmann, A. Supercharged proteins and polypeptides. *Adv. Mater.* **32**, 1905309 (2020).
84. Cronican, J. J. et al. Potent delivery of functional proteins into mammalian cells in vitro and in vivo using a supercharged protein. *ACS Chem. Biol.* **5**, 747–752 (2010).
85. Cronican, J. J. et al. A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo. *Chem. Biol.* **18**, 833–838 (2011).
86. Thompson, D. B., Cronican, J. J. & Liu, D. R. Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. in: *Methods Enzymol.* **503**, 293–319 (2012).
87. Kusmierz, C. D., Bujold, K. E., Callmann, C. E. & Mirkin, C. A. Defining the design parameters for in vivo enzyme delivery through protein spherical nucleic acids. *ACS Cent. Sci.* **6**, 815–822 (2020).
88. Samanta, D., Ebrahimi, S., Kusmierz, C., Cheng, H. F. & Mirkin, C. Protein spherical nucleic acids for live-cell chemical analysis. *J. Am. Chem. Soc.* **142**, 13350–13355.
89. Ebrahimi, S. B., Samanta, D. & Mirkin, C. A. DNA-based nanostructures for live-cell analysis. *J. Am. Chem. Soc.* **142**, 11343–11356 (2020).
90. Schellenberger, V. et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27**, 1186–1190 (2009). **This work reports the development of the XTEN polypeptide fusion partner for imparting advantageous properties into protein therapeutics, such as enhanced circulation time.**
91. Hu, J., Wang, G., Liu, X. & Gao, W. Enhancing pharmacokinetics, tumor accumulation, and antitumor efficacy by elastin-like polypeptide fusion of interferon alpha. *Adv. Mater.* **27**, 7320–7324 (2015).
92. Luginbuhl, K. M. et al. One-week glucose control via zero-order release kinetics from an injectable depot of glucagon-like peptide-

- 1 fused to a thermosensitive biopolymer. *Nat. Biomed. Eng.* **1**, 0078 (2017).
93. FDA grants efanesoctocog alfa Breakthrough Therapy designation for hemophilia A. *Sanofi* <https://www.sanofi.com/en/media-room/press-releases/2022/2022-06-01-05-00-00-2453803> (2022).
94. Yuan, Z. et al. Zwitterionic peptide cloak mimics protein surfaces for protein protection. *Angew. Chem. Int. Ed.* **59**, 22378–22381 (2020).
95. McMullen, P., Fang, L., Qiao, Q., Shao, Q. & Jiang, S. Impacts of a zwitterionic peptide on its fusion protein. *Bioconjug. Chem.* **33**, 1485–1493 (2022).
96. Dixon, J. E. et al. Highly efficient delivery of functional cargoes by the synergistic effect of GAG binding motifs and cell-penetrating peptides. *Proc. Natl. Acad. Sci.* **113**, E291–E299 (2016).
97. Kariolis, M. S. et al. Brain delivery of therapeutic proteins using an Fc fragment blood-brain barrier transport vehicle in mice and monkeys. *Sci. Transl. Med.* **12**, eaay1359 (2020).
98. Ptacin, J. L. et al. An engineered IL-2 reprogrammed for anti-tumor therapy using a semi-synthetic organism. *Nat. Commun.* **12**, 4785 (2021).
99. Zhang, B. et al. Site-specific PEGylation of interleukin-2 enhances immunosuppression via the sustained activation of regulatory T cells. *Nat. Biomed. Eng.* **5**, 1288–1305 (2021).
100. Cho, H. et al. Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl. Acad. Sci.* **108**, 9060–9065 (2011).
101. Lyu, Z. et al. A switchable site-specific antibody conjugate. *ACS Chem. Biol.* **13**, 958–964 (2018).
102. Vanderschuren, K. et al. Tuning protein half-life in mouse using sequence-defined biopolymers functionalized with lipids. *Proc. Natl. Acad. Sci.* **119**, e2103099119 (2022).
103. Lim, S. I., Hahn, Y. S. & Kwon, I. Site-specific albumination of a therapeutic protein with multi-subunit to prolong activity in vivo. *J. Control. Release* **207**, 93–100 (2015).
104. Lieser, R. M., Chen, W. & Sullivan, M. O. Controlled epidermal growth factor receptor ligand display on cancer suicide enzymes via unnatural amino acid engineering for enhanced intracellular delivery in breast cancer cells. *Bioconjug. Chem.* **30**, 432–442 (2019).
105. Li, Q. et al. Developing covalent protein drugs via proximity-enabled reactive therapeutics. *Cell* **182**, 85–97.e16 (2020).
106. Wals, K. & Ova, H. Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2** (2014).
107. Jester, B. W. et al. Development of spirulina for the manufacture and oral delivery of protein therapeutics. *Nat. Biotechnol.* **40**, 956–964 (2022).

Acknowledgements

Images are created using BioRender.com, the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, and Avogadro: an open-source molecular builder and visualization tool (Version 1.2.0, <http://avogadro.cc/>).

Author contributions

S.B.E. and D.S. wrote and edited the manuscript, prepared the figures, and compiled the tables.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38039-x>.

Correspondence and requests for materials should be addressed to Sasha B. Ebrahimi or Devleena Samanta.

Peer review information *Nature Communications* thanks Kai H. Griebenow, Rita Vanbever, Daniel-Adriano Silva and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023