Methods 131 (2017) 33–65

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

In silico methods for design of biological therapeutics

Ankit Roy [a,1], Sanjana Nair [a,1], Neeladri Sen [a], Neelesh Soni [a], M.S. Madhusudhan [a,b,*]

[a] Indian Institute of Science Education and Research, Pune, Dr. Homi Bhabha Road, Pune 411008, India
[b] Bioinformatics Institute, 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore

ARTICLE INFO

Article history:
Received 15 June 2017
Received in revised form 21 September 2017
Accepted 23 September 2017
Available online 27 September 2017

Keywords:
Biotherapeutics
Protein/peptide design
In silico engineering of antibodies and vaccines

ABSTRACT

It has been twenty years since the first rationally designed small molecule drug was introduced into the market. Since then, we have progressed from designing small molecules to designing biotherapeutics. This class of therapeutics includes designed proteins, peptides and nucleic acids that could more effectively combat drug resistance and even act in cases where the disease is caused because of a molecular deficiency. Computational methods are crucial in this design exercise and this review discusses the various elements of designing biotherapeutic proteins and peptides. Many of the techniques discussed here, such as the deterministic and stochastic design methods, are generally used in protein design. We have devoted special attention to the design of antibodies and vaccines. In addition to the methods for designing these molecules, we have included a comprehensive list of all biotherapeutics approved for clinical use. Also included is an overview of methods that predict the binding affinity, cell penetration ability, half-life, solubility, immunogenicity and toxicity of the designed therapeutics. Biotherapeutics are only going to grow in clinical importance and are set to herald a new generation of disease management and cure.

Contents

* Corresponding author at: Indian Institute of Science Education and Research, Pune, Dr. Homi Bhabha Road, Pune 411008, India.
E-mail address: madhusudhan@iiserpune.ac.in (M.S. Madhusudhan).
[1] Equal contribution.

https://doi.org/10.1016/j.ymeth.2017.09.008
1046-2023/

1. Introduction

Healing illnesses and diseases, caused by malfunctioning organs or by pathogens, with therapeutic agents is as old as recorded history. The ancient cultures practiced what we now categorize as traditional medicine. The therapeutic agents were typically herbal or animal extracts. From there we progressed to synthesizing new therapeutic compounds, such as chloral hydrate [1]. This acted as a great fillip to the development of the pharmaceutical industry. A large number of compounds were screened for the efficacy in combating diseased conditions by a procedure of trial and error. The screening of large libraries of compounds proved expensive and cumbersome. Moreover, these small molecule drugs had many off-target effects leading to adverse drug reactions [2–4].

In all of these therapies, the administered drugs helped cure the symptoms and/or the disease without prior knowledge of the mechanism of drug action. Given the advances in structure determination methods, the first rationally designed drug, dorzolamide, was introduced into the market about 20 years ago [5]. In rational design, treatments are devised by first getting insights into the molecules and molecular pathways involved in the diseased condition. A compound or molecule is then designed/synthesized considering the molecule against which it is targeted. This ensures that the therapeutic agent acts specifically on a target of choice. All drugs that are rationally designed today follow the protocol of first finding an appropriate bio-molecular target and then creating a target-specific inhibitor.

Small molecule inhibitors make for attractive drugs – many are easily synthesized, orally administered and could act on intracellular or extracellular targets [6–8]. But many of these small molecules do not have a high (enough) specificity and still result in side effects [9]. Also, certain diseases caused due to deficiency of a protein or enzyme, such as hemophilia, could not be managed by small molecule drugs [10]. To overcome these challenges, the pharmaceutical industry turned to designing biological therapeutics or biotherapeutics. Biological therapeutics include proteins, peptides and nucleic acids. Since nature has optimized these molecules to demonstrate specificity in target recognition within the crowded cell, they override the shortcomings of small molecules. The interest of the pharmaceutical industry on biotherapeutics was also due to the sharp rise in antibiotic-resistant strains of infectious organisms [11,12]. This rise is a result of indiscriminate antibiotic use. The advantage of using biotherapeutics in this scenario is that they have a large binding site on the target protein of the pathogen. This large surface that is recognized would still allow pathogens to mutate and become resistant to the drug, but the timescale would be much longer than for small molecule drugs [13]. However, biotherapeutics have their own set of challenges such as production costs, invocation of an immune response, reduced half-lives and limited modes of administration [14]. Despite these difficulties, biotherapeutics present effective strategies to surpass the disadvantages of small molecule drugs.

The focus of this review is to discuss the methods involved in designing protein and peptide-based biotherapeutics. Nucleic acid biotherapeutics have been discussed in detail elsewhere [15]. The first step of the design process, as with all rational design procedures, is to identify a suitable target. This selection is dependent on factors such as the similarity of the target with other proteins in humans (or other recipient organisms), its cellular location and its precise role in the progression of the disease *etc.* [16]. In this review, we assume that the target has already been identified and we restrict our discussion to the methods employed in designing and developing biotherapeutic agents, with special emphasis on antibody and vaccine design. The methods involved in the design of proteins and peptides are similar and will be discussed in Section 2. Methods specific to designing antibodies and vaccines are discussed in Sections 3 and 4 respectively (Fig. 1). The *in silico* efficacy of the designed biotherapeutics and the methods for their improvement are described in Section 5, followed by the challenges in the field and the way forward.

2. Computational methods for designing therapeutic proteins and peptides

In this section, we discuss the principles of designing biotherapeutic proteins and peptides. We define peptides as polymers of 40 amino acids or less. Most of the methods are developed for designing proteins, as they are more complex, but these methods are also applicable for designing peptides, unless mentioned otherwise. The aim of these design exercises is to create/engineer a new molecule that would intervene in the diseased condition by binding an appropriate target or by affecting a chemical pathway. The design of proteins/peptides refers to changing the arrangement of amino acids to either create an entirely novel protein/peptide (designing sequence and 3D structure) or to fit a pre-defined structural template (designing the sequence alone). The approach most commonly adopted for rational design is either the local modification of a pre-existing protein or the fitting of an amino acid sequence onto a given protein fold. The use of templates circumvents the difficulty of predicting the fold of an unknown sequence. Since the fold is unchanged, the backbone atoms are directly placed on this framework. The side chains that would best stabilize the structure are then added to the backbone to create a functional protein [17,18].

As mentioned earlier, we assume that the target is already identified and has a known three-dimensional structure. Therapeutic proteins against a known target can be designed by two approaches – (a) By identifying proteins that have a fold complementary to the target and modifying/changing some residues to facilitate binding, or (b) By searching/designing a sequence that would adopt the complementary fold and exhibit binding capabilities. In the following section, we discuss methods relevant to each of these approaches.

2.1. Template based design methods

Here we discuss methods that use existing structural templates for the design of therapeutic proteins. The overall approach can be summarized in two steps; (1) searching for a suitable template, and (2) modifying the template's interface to improve binding with the target. Templates are required to possess a fold geometrically complementary to the target structure. Once a template is chosen, the amino acid arrangement at the interface of the template is modified to have preferential contacts between the target and the template. When the structures of target binding proteins are available, the binding residues could be grafted from the target binding proteins onto the new template. The grafted region needs

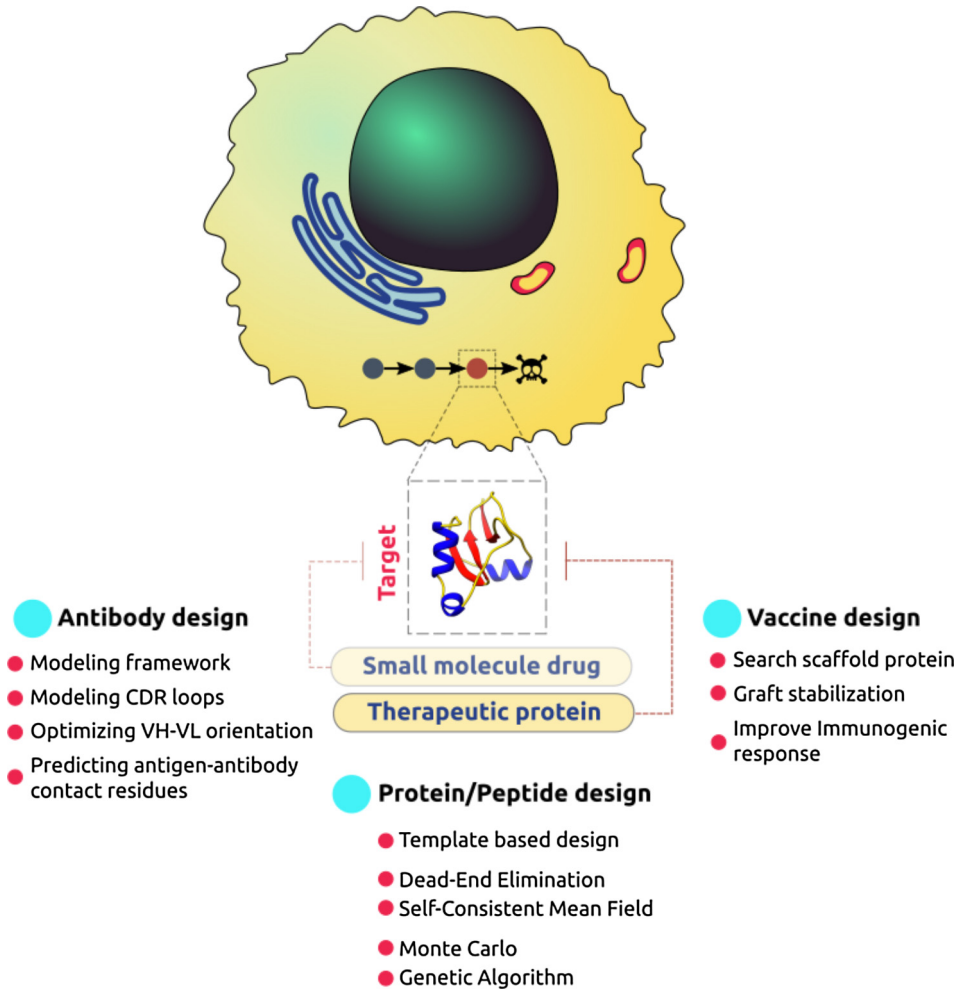

Fig. 1. An overview of the methods discussed in this review. A cellular process that leads to a disease condition is represented schematically as a set of dots and arrows connecting them. One of the molecules in this process/pathway serves as a therapeutic target. In this review, we discuss different methods of the three broad categories of designed biotherapeutics including proteins/peptides, antibodies, and vaccines.

to be stabilized by optimizing side chain torsion angles and by energy minimization using tools such as RosettaDesign [19,20] that improve the hydrogen bonding network and van der Waals interactions. Additional mutations for further stabilization could be introduced around the grafted regions. Care needs to be taken that there is minimum scaffold perturbation while modifying these amino acids. Existing protein structures are used as scaffolds for protein design as they are robust to residue substitutions [21,22].

Most methods such as GRAFTER [23], FITSITE [24], ProdaMatch [25] and ScaffoldSelection [22] require geometric restraints in the form of coordinates where representative atoms from templates could be placed. Representative atoms are backbone atoms, usually C^{α}, or both C^{α} and C^{β}. Once the spatial restraints are set, templates that satisfy these restraints are searched from a library of non-redundant protein structures. CLICK [26–28] is an example of a method that can efficiently perform topology independent structural comparisons between a template library and the geometric descriptor. Tools such as AutoMatch [29] also account for backbone flexibility while searching for templates.

Once a suitable template is found, the amino acid side chains at its interface need to be modified to improve binding affinity. Side chain modifications are made by evaluating various factors that determine binding affinity and structural stability. These factors include geometric and chemical complementarity, avoiding steric hindrance and an increase in bound surface area of the target-template interface [22]. Tools such as ORBIT [30] use a force field based evaluation method to guide side chain modifications that enable target template binding. RosettaMatch [31] uses a geometric hashing technique for optimal placement of side chains. If the binding requires the presence of cofactors, such as metals, methods such as OptGraft [32] and MetalSearch [33] could be used to design metal binding sites on proteins.

Designed protein structures can be refined using geometric algorithms that optimize side chain positioning. These algorithms include Voronoi (and Delaunay tessellations) partitioning based algorithms along with alpha and beta shape approximations [34,35]. They use Voronoi diagrams for a set of N points (seeds) to partition the space into N regions such that each region occupies one seed. Algorithms such as BetaSCP pack the side chains such that they minimize the intersection volume between atoms calculated through Voronoi diagrams [36]. This algorithm attempts to reduce the repulsive forces and increase the attractive forces between atoms to obtain a lower energy structure. Voronoi based algorithms have already been shown to be effective in modeling ligand/peptide conformations at their binding sites [37–40] and could be further explored for efficient side chain positioning at the designed binding interface.

The above-mentioned methods have been shown to be successful in a variety of protein design problems. Mayo and co-workers have used thioredoxin as the scaffold protein and modified its

active site to catalyze the reaction that converts p-nitrophenol acetate to p-nitrophenol [41]. Baker and co-workers have designed enzymes that catalyze the Kemp elimination reaction [42], retro-aldol reaction [43] and Diels-Alder reaction [44]. Liu et al. have designed a non-natural interaction of a PH domain protein with erythropoietin receptor by grafting interface residues from erythropoietin onto the non-homologous PH domain scaffold [45]. Fleishman et al. have designed two proteins that bind to influenza hemagglutinin with low nanomolar affinity [46]. Although not all of these designed proteins are used as therapeutics, the design principles remain the same and can be directly applied to the design of protein therapeutics.

2.2. Designing novel sequences

The previous section dealt with template based protein design where we inherit the sequence of the protein scaffold and make minor changes to improve binding affinity. Here, we discuss methods that predict a completely novel sequence that would be accommodated on a desired fold. Using deterministic or stochastic sequence search methods, a novel sequence is predicted that would fold into a desired conformation obtained from an existing protein scaffold. These methods search the sequence space to obtain a sequence that would represent the global minimum energy conformation. These methods are essentially used for solving the inverse protein folding problem.

The sequence of the protein is chosen such that it satisfies the geometric and energetic constraints imposed by the desired fold. Constraints usually include various parameters involved in intramolecular interactions, such as hydrogen bonding, van der Waals interactions, hydrophobic interactions, polar and electrostatic interactions *etc.* Typically, a scoring function is employed that takes into account various energetic contributions of the above-mentioned parameters. Sequence search methods sample multiple sequences and calculate their energies to find the one with the minimum energy. This problem could be tackled deterministically or stochastically. Deterministic methods search the complete sequence space to arrive at a sequence that folds into the global minimum energy conformation whereas stochastic methods search the sequence space heuristically.

2.2.1. Deterministic search methods

Deterministic search methods attempt to sample the complete sequence space and then converge onto a solution. This solution is the sequence that would accurately adopt the template backbone conformation with the lowest energy. Here, we examine two of the most commonly used deterministic methods namely dead-end elimination and self-consistent mean field.

2.2.1.1. Dead-End elimination. Dead-End Elimination (DEE) is an exhaustive search algorithm that identifies and eliminates sequence-rotameric states that are not part of the global minimum energy conformation [47]. The algorithm uses an energy function of two-body interactions to iteratively eliminate amino acids or rotameric states until no further amino acids or rotamers could be eliminated.

DEE compares two amino acid rotamers and eliminates the one with higher interaction energy. Interaction energies are calculated for every rotamer of the test amino acid with all the other rotamers of every other amino acid. At every iteration, the following condition is tested for any two rotamers i_a and i_b at amino acid position i:

$$E(i_a) + \sum_{j \neq i}^{N} \min_{x} E(i_a, j_x) > E(i_b) + \sum_{j \neq i}^{N} \max_{x} E(i_b, j_x) \qquad \text{(i)}$$

where $E(i_a)$ is the backbone-side chain interaction energy of i_a, and $E(i_a, j_x)$ is the side chain-side chain interaction energy with rotamers at all other amino acid positions j_x. If the condition is true, then i_a is incompatible with the global minimum energy conformation and is eliminated. This condition is iteratively tested for all amino acid positions and their rotamers until it no longer holds true [47,48].

DEE has been successfully implemented in the design of a novel 28 amino acid peptide that folds into a ββα motif based on a zinc finger template [49]. It has also been implemented in a partial design procedure where a Streptococcal protein Gβ1 domain was designed to have enhanced thermostability with a melting temperature above 100 °C [50]. Mayo and co-workers have developed an automated protein design pipeline that uses DEE and Monte Carlo methods to design the sequence of a protein that fits a desired fold [30].

With increasing sequence length, the combinatorial complexity of DEE increases exponentially. This renders DEE practically intractable for designing sequences of 30 amino acids or larger [51]. DEE is being continuously improved to decrease the combinatorial complexity. Improvements include the use of rotamer clusters for comparisons and revisions in the elimination criteria by incorporating Monte Carlo calculations to eliminate high energy rotameric states [52]. Extended DEE has been developed to account for multi-state protein structures to obtain multiple low-energy states from discrete states [53]. Another method for multi-state protein design is the type-dependent DEE [54], where energy comparisons are only made between different rotameric states of the same amino acid. Flexible backbone DEE [55] allows backbone flexibility by providing upper and lower bounds to rotameric interaction energies by specifying a range of backbone dihedral angles where the amino acid rotamer can be placed. These advances in DEE are mostly to address issues related to scaling-up to larger systems [51].

2.2.1.2. Self-consistent mean field. The self-consistent mean field method for side-chain modeling of proteins uses a global conformational matrix (CM) that contains the probabilities for every rotameric state at each amino acid position on the protein. CM is a matrix of dimension $N \times R$ where N is the total number of amino acids in the protein and R is the total number of rotamers. Any element CM_{ij} is defined as the probability of rotamer j occurring at the ith position of the protein [56]. The method populates CM by assigning probabilities to all rotameric states of an amino acid position based on its interaction energy with all other amino acids in the protein. The final sequence is derived by choosing the amino acids with the highest probabilities at each position.

Initially, the CM has equal probabilities for all rotamers. This is updated by calculating the average local energy generated by the interaction of a side-chain rotamer j with all other neighboring side-chains (mean field). This is repeated multiple times for every position i until convergence. The energy function used for calculating the average local energy of a side-chain rotamer at position x is:

$$\varepsilon_x(\alpha_x, r(\alpha_x)) = \varepsilon_x^0(\alpha_x, r(\alpha_x)) + \sum_{y,\alpha_y} \sum_{r(\alpha_y)} \omega_y(\alpha_y, r(\alpha_y)) \gamma_{xy}(\alpha_x, r(\alpha_x); \alpha_y, r(\alpha_y)) \qquad \text{(ii)}$$

where, α_x and $r(\alpha_x)$ are the amino acid identity and its rotameric conformation at position x, $\varepsilon_x^0(\alpha_x, r(\alpha_x))$ is the interaction energy of the side-chain rotamer with the backbone, $\omega_y(\alpha_y, r(\alpha_y))$ is the site specific probability of amino acids and their side-chain conformations at the position y, and $\gamma_{xy}(\alpha_x, r(\alpha_x); \alpha_y, r(\alpha_y))$ is the two-body interaction energy of side-chains α_x and α_y. At the end of the computation the CM is populated by updated probabilities for every side-chain rotamer at each amino acid position of the protein. The

predicted side-chain conformations are the ones with highest probabilities at each amino acid position of the protein [48,56].

Koehl and Delarue modeled the side chains of a 325 amino acid long protein rhizopuspepsin starting from a given backbone conformation. They were able to correctly predict 81% of all χ^1 dihedrals and 73% of χ^1 and χ^2 dihedrals [56]. This method has been reported to have a higher accuracy in predicting side chain conformations at the hydrophobic core as compared to the protein's surface [51,56]. Since the time required for convergence of *CM* increases linearly with increase in sequence size, the self-consistent mean field is faster in comparison to DEE, but is less accurate [51]. Currently, researchers are combining DEE with self-consistent mean field approaches to improve on both speed and accuracy [51].

2.2.2. Stochastic search methods

The deterministic methods, while accurate when designing small proteins/peptides have practical limitations with increasing sequence size. For large proteins, finding an optimal solution relies on the use of heuristic or stochastic approaches. The stochastic methods most commonly used for protein or peptide design include Monte Carlo sampling and Genetic Algorithm.

2.2.2.1. Monte Carlo sampling. The Monte Carlo method samples conformational energy from a Boltzmann distribution [57]. The goal is to sample sequence variations to identify the lowest energy sequence that can adopt a desired fold. The lowest energy sequence is searched by making random moves and accepting or rejecting the move based on the energy of the new state. A move is defined as a change from the previous state of the protein that could either be a change in the amino acid sequence or a change in the rotameric state of an amino acid. The move is always accepted if the energy of the new state is less than that of the previous state, otherwise it is accepted with a probability P_{accept} (Metropolis criterion) that is expressed as:

$$P_{accept} = min(1, e^{-\beta\Delta E}), \quad \beta = 1/kT \tag{iii}$$

Here, ΔE refers to the change in energy as a result of the move, β represents the inverse of temperature, k is the Boltzmann constant and T is the temperature. Various modifications to this method exist such as Simulated Annealing, Simulated Tempering, Biased Monte Carlo, and Replica Exchange Monte Carlo. These modifications are explained below:

1. *Simulated Annealing*: In this method, the system is initially simulated at a very high temperature and is gradually cooled [58]. The moves that reduce the energy of the system are accepted, and those that increase the energy are accepted with the probability of their occurrence (Eq. (iii)). This probability depends on the temperature and the energy difference between the two states. This method, at higher temperatures, allows the system to overcome local minima. As the temperature approaches 0, moves that encourage downhill movement on the energy landscape are preferred. Thus, with adequate sampling, it results in finding the global minima.
2. *Simulated Tempering*: Since the temperature jumps are random, every run of simulated annealing has a probability of ending at a different minimum and does not guarantee the achievement of global minima when run at finite temperature ranges. A modified version of Monte Carlo, called simulated tempering, was introduced to include discrete temperature ranges [59,60]. This method has a modified probability of acceptance of a move. The simulation starts at an initial temperature. Classical Monte Carlo protocol is followed that is dependent on the energy of the resultant protein and the Metropolis criterion for acceptance of moves. Change in temperature to a higher or a lower value is also subject to the acceptance probability and a separate function incorporating both the energy and the temperature contribution is created. This enables exploration of the energy space as well as temperature space, hence, resulting in the convergence of the moves into a global minimum.
3. *Biased Monte Carlo*: This method is biased for the selection of the sequence [61]. The amino acids are arranged in a two-dimensional lattice such that each amino acid occupies one grid. Thus, the total number of grids corresponds to the length of the protein. A fixed number of combinations of these grids are made. The energy of the amino acid at each position depends on its neighbors and whether the amino acid is buried or exposed. The Monte Carlo protocol is followed, with the exception that the choice of amino acid for replacement depends on a probability function that favors replacements having better energies. Thus, the moves have higher chances of getting accepted. The acceptance probability is calculated from the ratio of Rosenbluth weights after $(W_{a'})$ and before (W_a) accepting the move. Details of the formulation of the probability distribution and weights are explained in detail in previous studies [61,62].

$$P_{accept} = min\left(1, \frac{W_{a'}}{W_a}\right) \tag{iv}$$

4. *Replica Exchange Monte Carlo*: Replica Exchange Monte Carlo [63] is another modification of the classic Monte Carlo that is efficient in uneven energy landscapes. The Monte Carlo simulations are simultaneously carried out at different temperatures. Since, the moves are random, these simulations or "replicas" would have different conformations. At specific intervals of time, the conformations from two different systems at different temperatures are swapped. Then the probability of acceptance of the swapping move depends upon the temperature of the two systems and the difference between the energies of the two states represented by:

$$P_{accept} = min(1, e^{(\beta_1\beta_2)(\Delta E_1 - \Delta E_2)}) \tag{v}$$

Here, β_1 and β_2 represent inverse temperatures and ΔE_1 and ΔE_2 are energies at temperatures t_1 and t_2 respectively.

The most commonly implemented Mote Carlo based protein design methods are RosettaDesign [20], PROFASI [64], PHAISTOS [65], CHARMM [66], MCPRO [67], EvoDesign [68] and eVolver [69]. RosettaDesign uses Monte Carlo with simulated annealing to find a sequence that would fit the desired fold. A modification of RosettaDesign that allows change in the backbone conformation is RosettaBackrub [70]. Even though RosettaBackrub was not specifically built for designing proteins, its allowance of backbone flexibility could be incorporated into RosettaDesign. eVolver uses simulated annealing with a structure based sequence profile to search the sequence space. PROFASI is another Monte Carlo based simulation software that allows backbone flexibility, like the Rosetta package, but favors local backbone deformations. It allows users to choose between simulated annealing and simulated tempering. The energy function is a linear combination of the hydrogen bonds formed by backbone atoms, hydrophobic interactions between non-polar groups, repulsion between atoms and electrostatic attraction between sequential neighbors. PHAISTOS is another method that utilizes PROFASI's forcefield in its Monte Carlo program. It differs from PROFASI in having an additional energy function of OPLS-AA/L and a larger move set. CHARMM implements hybrid Monte Carlo that adopts its acceptance criteria from a change in total energy and not a change in potential energy. MCPRO uses simulated annealing like the other methods discussed and is used for analyzing protein-ligand interactions. EvoDesign

uses classical Monte Carlo with a limited sequence search space. It creates an evolutionary profile of sequences based on the selected scaffold. This restricts the sequence search space for Monte Carlo. A study by Dantas et al. [71] has shown concurrence between designed proteins and their experimentally determined structures, illustrating the usefulness of such methods.

2.2.2.2. Genetic algorithm. Genetic algorithm is another stochastic search method, which derives its principle from Darwin's theory of "Survival of the Fittest". Its biological basis is evident in the steps involved in its operation – reproduction, mutation, cross-over and selection [72]. Its application is suggested when the search space is large, replete with local minima or maxima, and there is lack of knowledge about the route that needs to be followed to obtain the final structure. Initially, a set of random seed sequences are chosen either based on *a priori* knowledge or randomly. These sequences are then subjected to mutation simultaneously. A mutation involves changing an amino acid to one of the other 19 amino acids. The initial sequence is the parent and the mutated one is its offspring that forms the next generation. Another change that can occur is a cross-over, which refers to swapping of segments of one sequence with another. The resulting chimera is now a part of the next generation. After each generation, the "fitness" of the sequence is analyzed. The fitness of a sequence is evaluated by different measures depending upon the desired application. For designing stable folds of protein or peptide, the fitness score is obtained from the pairwise-energy potentials as well as solvation effects. When the protein or peptide is being designed to bind to a specific target, the fitness score is determined by the binding affinity of the three-dimensional structure of the sequence with the target. Sequences that form low energy structures are considered to have better fitness. Selection of the sequences that will be carried forward to the next generation can be done using various algorithms. The first algorithm proposed was the proportional or roulette wheel algorithm, where the proportion of a sequence in the population is directly proportional to its relative fitness compared to other sequences. The best percent selection algorithm conserves sequences having the best fitness scores. This is done by choosing a fixed percentage of sequences with the top fitness scores that are passed onto the next generation without any mutation or cross over. Other algorithms include linear rank selection, random selection, binary tournament selection, Q-tournament selection and universal sampling (discussed in detail in [73]). This iterative process of forming new sequences is carried out either for a fixed number of times (usually 100 generations) or until the convergence of different sequences into one. As this method involves parallel processing of many seed sequences, the number of computations and time required to reach the optimum solution is multiplied considerably. But as individual systems are independent of the others, the computation could be parallelized. Two webservers that use genetic algorithms for protein design are GAPSSIF [74] and EGAD [75].

3. Antibody design

The previous sections dealt with protein design in general. However, antibodies form a special case of proteins and are one of the most important classes of biotherapeutics today [76,77]. In recent years, the largest fraction of approved biological therapeutics has been monoclonal antibodies. About 27% of all the approved biological therapeutics between the years 2010 and 2014 were antibodies, with worldwide sales of $75 billion and with a projected rise to $125 billion dollars by 2020. As of January 2017, the FDA has approved 68 therapeutic antibodies with 10 of them being approved in 2016 alone [77–79].

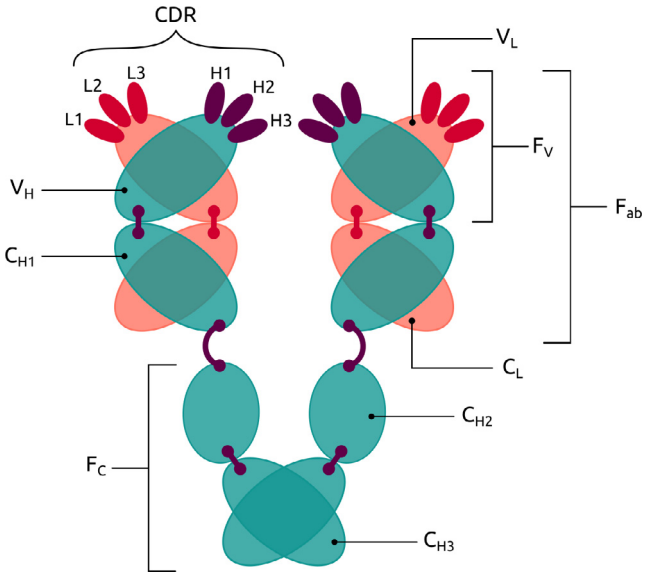

Fig. 2. Schematic representation of the structure of an antibody. Antibodies consist of a shorter light chain (peach) and a longer heavy chain (teal). The antibody structure is divided into its constant region (F_C) and its antigen binding fragment (F_{ab}). F_{ab} harbors a variable region (F_V) that is responsible for antigen binding. F_V is composed of light chain variable region (V_L) and a heavy chain variable region (V_H). The complementarity determining region (CDR) on V_H and V_L dictate antigen binding specificity and affinity. CDR is composed of six hypervariable loops, three from the light chain (L1, L2, L3) and three from the heavy chain (H1, H2, H3). The constant regions of the heavy chains are labelled as C_{H1}, C_{H2} and C_{H3} whereas the constant region of the light chain as C_L.

The current method to develop antibodies against a specific antigen is to screen a large library of antibodies for potential binders. Molecules that bind to the specific antigen are then mutated to generate a new library of antibodies that are in turn screened to check for an increase in binding affinity. This iterative process while efficient at designing antibodies with higher binding affinities is both time consuming and expensive [80,81]. Clearly, better strategies are needed to get high affinity antibodies. To better appreciate these newer methods, we first preview the peculiarities of antibody 3D structure.

Antibodies are made of two types of chains namely the light chain and the heavy chain, each of which has multiple domains. The structure of an antibody can be divided broadly into two domains the constant domain and the antigen binding domain. Antigen binding is mediated by six hypervariable loops that form the Complementarity Determining Region (CDR) present on the antigen binding domain. A description of the structural characteristics of antibodies is shown in Fig. 2.

Antibody design is divided into modeling the six hypervariable loops in the CDR and the rest of the antibody known as the framework region. Since the framework shares substantial similarity with other antibodies, it is easier to model, whereas the CDR has hypervariable loops and require additional constraints to be modeled accurately. The CDR-H3 loop being the most variable, both in length and sequence, requires a different approach for modeling as compared to the other CDR loops. The CDR-H3 is particularly important in antigen binding, and therefore it is crucial to model it with significant accuracy [82].

Prediction of the antibody structure can be briefly summarized into annotating the sequence into structurally equivalent residues of the framework and CDR, modeling the framework, modeling the CDR loops and predicting or optimizing the V_H-V_L orientation. In this section, we briefly summarize the methods and the tools (Table 1) required for each step of rational antibody design. For a comprehensive overview, we recommend the review by Krawczyk et al. [82].

The antibody sequence is initially annotated into its framework and CDR segments using numbering schemes. Various numbering schemes such as Kabat [83], Chothia [84], Enhanced Chothia [85], IMGT [86], and AHo [87] have been used for these annotations. Multiple online web servers are available that utilize one or more of the methods mentioned above to number the antibody sequence; Abnum [85] uses Kabat, Chothia, and Enhanced Chothia schemes, DomainGapAlign [88,89] uses IMGT scheme, PyIgClassify [90] uses AHo scheme, and ANARCI [91] uses all the five numbering schemes. Once the sequence is annotated, the framework and the CDR are modeled using approaches mentioned below.

3.1. Modeling the framework region

Modeling the framework is relatively easier as compared to the CDR since they share a substantial similarity both in terms of sequence and structure. Usually, using templates that are 80% or higher in sequence identity in the framework regions, results in an accurate model [82]. Using such templates, all participants of the Antibody Modeling Assessment-II [92] were able to predict the framework region with an average RMSD below 1 Å (for both V_L and V_H frameworks). Tools such as IgBLAST [93] and SAbDab template search [94] are used for searching templates to model V_H and V_L framework regions. These template search methods differ from conventional methods in being able to delineate the framework from the CDR and being able to utilize the search for homologous framework sequences from antibody sequence databases. Templates obtained using these methods can be utilized by different modeling protocols.

3.2. Modeling CDR loops

Five of the six hypervariable loops (L1, L2, L3, H1 and H2) of the CDR can be clustered into a limited number of structural conformations, known as canonical structures [84,95–97]. The conformation of the loop can be decided by identifying a few key structure determining residues [98]. The other residues (not identified as key) do not impact the conformation of the loop. Tools such as PyIgClassify [90] look for the structure determining residues in a sequence and compare it with the canonical clusters to determine the conformation of the loop. Another method to predict the non-CDR-H3 loops is to treat these loops like any other loop modeling problem, without the canonical classification [99]. Tools akin to FREAD [100] and the MODELLER loop modeling protocol [101] can be used to deal with the non-CDR-H3 loops like any other loop and predict their structure.

Prediction of the CDR-H3 loop is relatively difficult as compared to the other CDR loops owing to their high variability in both length and sequence. The CDR-H3 loop lies at the center of the antigen binding site and has a critical role in antigen recognition due to which modeling of this loop with significant accuracy is crucial for antibody engineering [82]. The structure of CDR-H3 loop can be divided into the base region (proximal to framework) and the β-harpin region (distal to framework). Although CDR-H3 loops do not have canonical structures, they can be classified to some degree based on subtypes of these two regions [95,102,103]. Residue preference at few positions of the antibody sequence can help guide the choice of CDR-H3 loop conformations, like the presence or absence of asparagine at Chothia position 101 [102,103]. Both template-based and *ab initio* methods have been utilized for modeling the conformation of the CDR-H3 loop. Template-based methods such as PIGS [104] and FREAD [100] use sequence derived rules for choosing the correct template. As with all other template based structure determination methods, the accuracy of the CDR-H3 loop depends on the choice of the template and loops cannot be predicted in the absence of a template. *Ab initio* methods like Kotai Antibody Builder [105] and Rosetta Antibody [106,107] sample multiple conformations to build loop decoys and score them using their energy functions. Reliable prediction of the CDR-H3 loop remains a hurdle in the rational design of antibodies.

Table 1
List of all tools used for different stages of antibody design along with their modes of access.

Tool Name	Type	Access
Antibody Structure Prediction		
ABodyBuilder	Web server	Open
Accelrys Discovery Studio	Standalone	Proprietary
Kotai Antibody Builder	Web server	Open
LYRA (LYmphocyte Receptor Automated modeling)	Web server	Open
MOE (Molecular Operating Environment)	Standalone	Proprietary
PIGS (Prediction of ImmunoGlobulin Structure)	Web server	Open
Rosetta Antibody	Web server	Open
Schrodinger Biologics Suite	Standalone	Proprietary
SmrtMolAntibody	Web server	Proprietary
Antibody Sequence Annotation		
Abnum	Web server	Open
ANARCI (Antigen receptor Numbering And Receptor Classification)	Web server and Standalone	Open
IMGT Domain Gap Align	Web server	Open
PyIgClassify (Python-based Immunoglobulin Classification)	Web server	Open
Template Search		
IgBLAST	Web server and Standalone	Open
SAbDab Structural Antibody Database	Database	Open
Loop Prediction		
FREAD	Web server	Open
H3Loopred	Standalone	Open
MODELLER	Standalone	Open
ModLoop	Web server	Open
VH-VL Orientation		
ABangle	Web server and Standalone	Open
Antibody Humanization		
TabHu (Tool for Antibody Humanization)	Web server	Open
Antibody Prediction Suite		
SAbPred (Structural Antibody Prediction)	Web server	Open
Paratope Prediction		
Antibody-i-Patch	Web server and Standalone	Open
Paratome	Web server	Open
ProABC	Web server	Open
Epitope Prediction		
Conformational Epitope Database	Database	Open
EpiPred	Web server	Open
Immune Epitope Database and Analysis Resource	Web server and Database	Open
Antigen – Antibody Docking		
ClusPro (Antibody mode)	Web server	Open
DockSorter	Standalone	Open
SnugDock	Standalone	Open

3.3. Optimization of the V_H-V_L domain orientation

Amino acids on the antibody that are in direct contact with the antigen are collectively termed as the paratope. The corresponding interacting region on the antigen is termed the epitope. The orientation of the V_H-V_L domain determines the extent to which the paratope is accessible to the antigen. Sub-optimal orientation of the V_H-V_L domains decrease the antigen binding affinity and may also lead to a complete loss of antigen recognition; hence it is important to optimize the orientation of these domains [108–114]. There are multiple ways to determine the orientation of the

V_H-V_L domains, the easiest of which is to mimic the orientation from a structure with high sequence similarity. Sequence similarity can be calculated by considering the complete variable region or just the residues that are known to make contacts with the antigen. Other methods include the use of energy functions to choose the best orientation from a set of possible conformations or to optimize the orientation during the prediction procedure iteratively [106,115]. ABangle [116] uses an absolute measure of the domain orientation; it measures two torsion angles, two twist angles, two tilt angles and the distance between the domains. Some methods use information from key residues to predict the orientation of the V_H-V_L domains [117,118].

3.4. Predicting the antigen-antibody contact residues

In a natural immune response, the binding residues on the antibody are often mutated to increase the antigen binding affinity and specificity in a process known as affinity maturation. If antigen-antibody contact residues are known, one can computationally mutate them to screen for residues, which increase the binding affinity and specificity of the antibody. This process can be divided into three steps; paratope prediction, epitope prediction, and antigen-antibody docking.

3.4.1. Paratope prediction

On average, about 80% of the amino acids constituting the paratope are in the CDR, accounting for about a third of the number of residues in this region [119]. Since most of the paratope occupies the CDR, traditional methods for predicting CDRs, such as those by Chothia, Kabat, and IMGT, are often regarded as tools for predicting antigen binding sites [82]. Tools such as Paratome [120] use sequence and structure information to predict paratope residues both in the CDRs and outside of the traditionally defined CDRs.

As only a third of the CDR residues participate in antigen binding, it is of importance to identify these residues precisely. This reduces the number of residues that need to be mutated for affinity maturation. Tools for such high-precision antigen binding residue predictions are available, namely ProABC [121] and Antibody-i-Patch [122]. Antibody-i-Patch provides an antigen-contact-likelihood for all predicted residues. These likelihoods can be used to decide what residues are to be chosen for mutagenesis.

3.4.2. Epitope prediction

In a search for immunogenic motifs on antigens, databases such as the Conformational Epitope Database [123] and the Immune Epitope Database [124] have been created that map the structural/sequential epitope motifs onto antigens. A majority of the methods for epitope prediction rely on these databases to search for immunogenic motifs on antigens [125]. These methods are based on the assumption that antigens have certain motifs that are inherently more immunogenic as compared to the rest of the protein and do not require any information of the antibody. It has been reported that immunogenic motifs are indistinguishable from the rest of the protein which suggests that any part of the protein can be a part of an epitope [126]. Methods that incorporate antibody information for epitope prediction have been shown to outperform methods that do not [127]. EpiPred [128] is an epitope prediction tool that analyzes the geometric complementarity of the antigen-antibody interface along with preferential interface contact residue frequencies using an antigen-antibody specific knowledge-based statistical potential.

3.4.3. Antigen-antibody docking

Predictions of the epitope and paratope help us identify the key residues involved in the formation of the antigen-antibody interface but do not provide the pairwise relationship between them. Docking the antigen onto the antibody reveals the contact residue pairs from the interface. This information can be utilized to rationally mutate the antibody residues to improve the interface complementarity and thereby enhancing the binding affinity.

Protein-protein docking is a method used to predict the best binding mode of two interacting proteins. It involves two steps; generation of decoys using conformational sampling and reordering of decoys using scoring functions which sort the decoys by binding affinity. The best scoring complex (lowest energy structure) should, ideally, be the native complex (or the best binding complex). Although the problem of antigen-antibody docking is a subset of a more general problem of protein-protein docking, tools which utilize antibody-specific decoy generation and scoring methods perform better than the general methods [126,127]. SnugDock [129] and the Antibody Mode of ClusPro [130] are antibody specific docking protocols.

Results of the antigen-antibody docking can be substantially improved by providing information about the antigen epitope and the antibody paratope. This information can be used to apply constraints for the docking procedure. Since not many antigen-antibody specific docking algorithms exist, one can generate a large number of decoys using a generic docking algorithm [82], such as ZDOCK [131] and PatchDock [132], and then reorder the decoys using antibody specific scoring functions such as DockSorter [122].

4. Vaccine design

The previous parts of the review dealt with computational methods for designing proteins, antibodies and peptides to combat disease conditions. However, the prevention of diseases offers a viable if not more effective alternative. Preventive measure often involves administering vaccines that mimic the pathogenic antigen proteins, so that antibodies can be raised against them, which can in turn prevent future infections. With the increase in the incidence of Zika, Dengue *etc*., the computational design of vaccines is a key step in any preventive measure. The *in silico* design of vaccines involves grafting of the epitope residues on a structurally similar template, as mentioned earlier in Section 2.1 followed by mutations to stabilize the graft (Fig. 3). Because of structural and chemical similarity, the epitope grafted onto a non-pathogenic scaffold should elicit an antibody response against it, forming the basis for long term immune protection. The newly engineered protein is then expressed and is injected into a mice/rabbit to generate antibodies. The generated antibodies are checked for binding to the antigenic epitope. After passing through various clinical trials involving toxicity, binding affinity, cross reactivity tests *etc*. the engineered immunogen can be administered as a vaccine.

Various *in silico* vaccine design studies have been carried out with HIV-1 epitopes. Initial studies with insertion of continuous linear epitope of HIV-1 gp41 subunit into a rhinovirus carrier showed that the length, hydrophobic character and precise insertion sites of the epitope are important for such design [133]. Continuous and discontinuous epitopes of gp41 subunit [134], gp120 subunit [135] and 4E10 [136] from HIV have been used for such studies. Since side chain grafting sometimes introduces conformational differences in the backbone of the scaffold and the epitope, up to RMSD ~1 Å, backbone grafting is resorted to solve this problem [137]. This involves incorporation of the epitope onto the scaffold followed by introduction of novel backbone regions in the

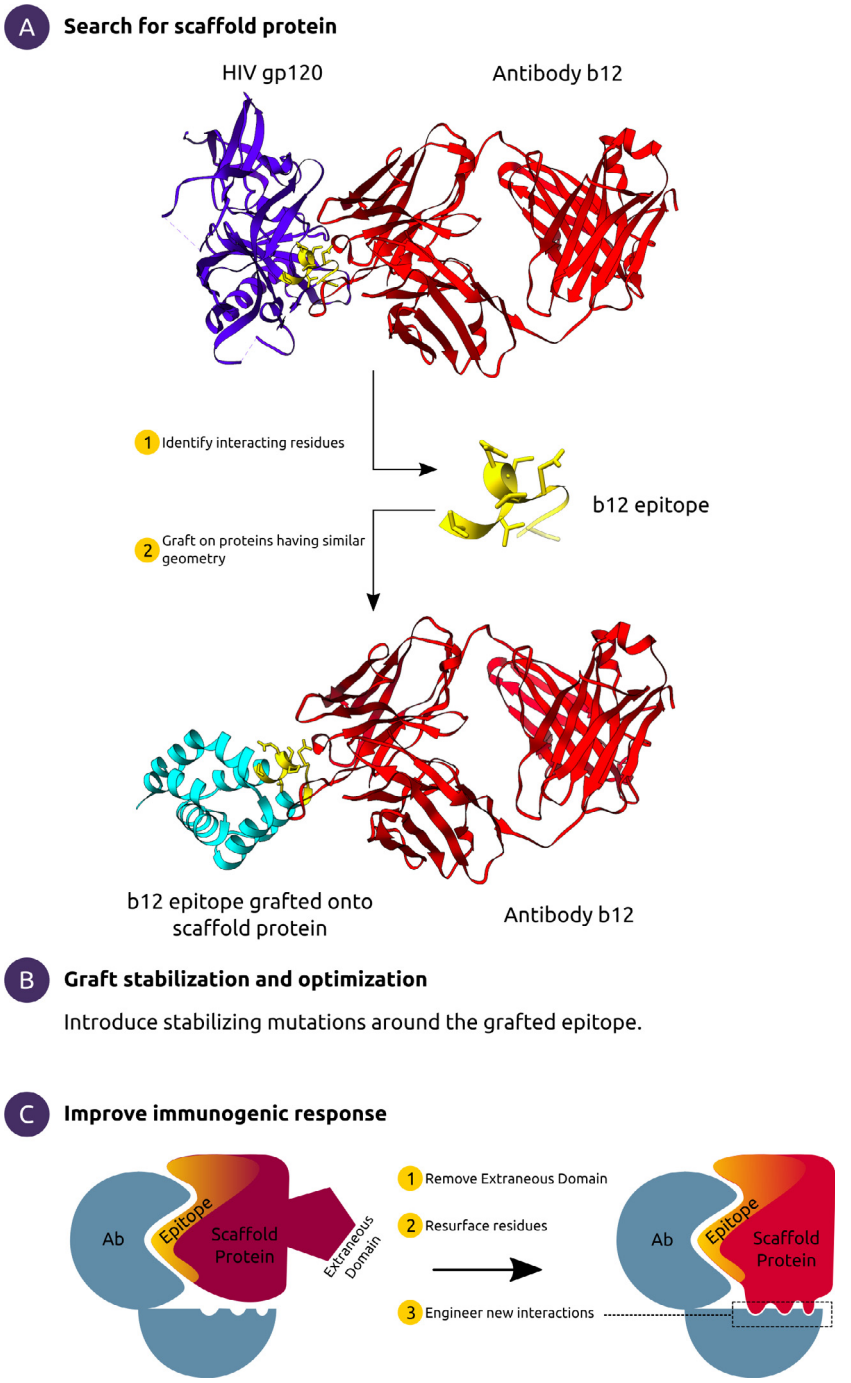

Fig. 3. The steps involved in vaccine design illustrated using the HIV gp120 protein (blue ribbons). A structure of the neutralizing antibody b12 (red ribbons) in complex with the gp120 is known (pdb: 2NY7). (A) The eptiope of gp120 (yellow ribbons with residues in stick representation) that is recognized by the antibody is hence also known. To elicit the same antibody response, a vaccine has to present the same epitope. Protein with similar geometries as the epitope are identified (eg., the protein shown in cyan ribbons). (B) The epitope residues are grafted onto the new protein and additional mutations could be made to stabilize the grafted region. (C) The affinity of the antibody towards the epitope is optimized by increasing interactions between them. Extraneous regions/domains are removed and amino acids are resurfaced to decrease unwanted secondary binding sites and increase binding affinity. All protein representations were rendered using UCSF Chimera 1.11.2 [208]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

insertion junction. The scaffolds that clash the least with the antibody are retained for further investigation. Different sequence modifications are made via backbone and side chain modifications, changing length and secondary structure of the connecting segments to generate structures with progressively lower ROSETTA energies. This procedure reduced the RMSD between the scaffold and epitope to 0.2 Å. The antibodies produced by backbone grafting showed a thirty-fold improvement in its affinity to 2F5 epitope as compared to those produced by side chain grafting. Human-guided design is followed to remove extraneous domains, optimize the solubility and eliminate undesired mutations like unpaired cystines, solvent exposed hydrophobic groups, buried hydrogen bond acceptor or donor groups, extra interactions with the epitope *etc.* Using similar techniques, discontinuous epitopes from two α

helices of RSV (Respiratory Syncytial Virus) F peptide that binds the motavizamav antibody were designed [138]. The design involved searching for scaffold for each of the two epitope segments. On finding a match, the scaffold was then searched for the other epitope while maintaining a rigid body conformation for the first epitope. The epitope specific antibodies produced were precise, but in low titre.

The above-mentioned techniques might produce cross-reactive epitopes, reduced solubility and/or stability of the engineered protein. Resurfacing, is a method used to overcome such problems that involves changing most of the exposed residues other than the epitope of interest (Fig. 3). Antigens were designed by keeping the core, glycan site and the antibody binding region of the HIV-1 gp120 protein constant. The remaining surface amino acids were resurfaced based on evolutionary information, structural and solubility consideration and with similarities and differences with wild type or pre-existing designs [139]. RosettaDesign was then used to select low energy sequences. Similarly resurfaced antigens were created for the 4E10 epitope [140]. During resurfacing sequences with smaller sequence identity to parent sequence were preferred to increase the sampling space. Another method involves removal of extra epitopes on the scaffold to reduce the size of the epitope-scaffold (Fig. 3). The 4E10 epitope scaffold was trimmed to get rid of an extra domain of the scaffold to reduce the size of the antigen. Newer backbone fragments of different lengths were inserted in the trimmed region and optimal sequences were obtained such that the newly modeled region was maximally stabilized. The affinity between the antigen and antibody can be enhanced by increasing interactions of the scaffold with the CDR H3 loop of the antibody [141]. The epitope scaffold produced bound the 2F5 antibodies with sub-nanomolar affinity.

Earlier methods were bound by the scaffold proteins of predetermined structures. Newer methods, such as Fold From Loop [142], were developed to allow full backbone flexibility to improve the tailoring of the epitope. The functional motif and the target topology are taken as an input for the design. *Ab initio* folding is then carried out to produce diverse backbone conformations similar to the target topology. The sequences are then iteratively designed followed by structural relaxation and full atom optimization to select for low energy sequences. The conformation of the functional motif is mostly fixed throughout the computation. This is followed by human guided computational optimization. Immunization of macaques with these engineered scaffolds produced potent neutralizing antibodies against RSV F epitope [142].

5. Prediction and improvement of *in vivo* efficacy

The efficacy of the designed biotherapeutics depend on multiple factors including their affinity to bind specifically to their targets, their ability to be retained in the circulatory system for longer durations, their ability to penetrate cells, their immunogenicity, solubility and toxicity. This section will briefly discuss how each of these properties is predicted in the designed biotherapeutics.

5.1. Binding affinity

Protein-protein interactions are crucial to the functioning of cells and are often regarded as potential drug targets. For instance, monoclonal antibodies could competitively interact with their cognate antigens with high binding affinities. Predicting the binding affinities, usually given in terms of the dissociation constant K_d, of protein complexes is therefore an important step in the rational design of such proteins. The relationship between K_d and the binding free energy is given by:

$$\Delta G = -RTln(K_d) \quad \text{(vi)}$$

where ΔG is the change in the free energy, is the gas constant and T is the absolute temperature. The prediction of binding affinities usually utilizes energy functions that evaluate the binding free energy of a protein complex. These energy functions are either force field-based energy functions or knowledge-based statistical potentials.

Force-field based energy functions extensively calculate energies derived from various parameters like van der Waals interactions, hydrogen bonding, electrostatic interactions, hydrophobic effect, desolvation energies and entropic effects [143–145]. The binding energy of a protein complex is calculated by summing up the individual contribution of these parameters. Most commonly used force field-based methods include free energy perturbation (FEP) [146,147], Molecular Mechanics – Poisson Boltzmann Surface Area (MM-PBSA) [148,149], Molecular Mechanics - Generalized Born Surface Area (MM-GBSA) [148,150], and thermodynamic integration [146,151]. FEP has high accuracy but is computationally expensive and time consuming. MM-PBSA and MM-GBSA use the Poisson Boltzmann equation and the Generalized Born approximation respectively, for calculation of the electrostatic contribution and are computationally more efficient [145,151]. Knowledge-based potentials are energy functions derived from statistically analyzing various descriptors of known protein structures. These functions exploit the regularities in protein structures to calculate potentials, based on the assumption that frequently occurring states correspond to low energy states [152,153].

Multiple tools can identify hot spots on protein interfaces, which when mutated to alanine strongly attenuate binding. These tools calculate the values for the change in binding energy of the protein complex upon mutation. A few of the most commonly used hot spot prediction tools are: HotPOINT [154], which uses a knowledge-based potential; KFC [155], KFC2a and KFC2b [156] use a combination of knowledge-based potential and machine learning; MINERVA [157] uses machine learning; FoldX [158] and Robetta [159] use force field-based energy functions. Alanine Scanning can be used to study the effect of a particular amino acid on the binding affinity of the protein complex. AlaScan [160] is a user-friendly graphical user interface based on FEP calculations and is aimed at providing an easy platform to perform *in silico* Alanine Scanning Mutagenesis. Although multiple methods for prediction of binding affinities exist, as of now their accuracy is limited as can be seen from their mediocre performance in the CAPRI (Critical Assessment of Prediction of Interactions) binding affinity test [161].

5.2. Cell penetration ability

Currently, biotherapeutics are either enzymes/proteins that overcome a deficiency, or drugs that act by interacting with cell surface receptors. However, the sub-cellular localization varies from one target to another. An effective biotherapeutic should be able to cross the lipid bilayer of the cell plasma membrane. Since the cellular receptors and transporters only allow selective entry of molecules, the therapeutics must possess the ability to cross the membrane without the aid of these membrane proteins. Some peptides are capable of crossing the cell membrane and are termed as Cell Penetrating Peptides (CPPs). These CPPs act as the vehicles that carry the therapeutics to their required site of action. CPPs are 5–30 residue long peptides that have a strong hydrophobic character and frequently exhibit a set of terminal cationic charges that allows them to penetrate the cell membrane while still being in the soluble form [162,163]. Some therapeutic peptides can enter the cell membrane by themselves and are called bioactive CPPs. For other therapeutics, a CPP tag can act as a transporter that carries the therapeutic into the cell. Antimicrobial peptides produced

by various organisms have the ability to penetrate the cell membrane and kill the microbe, either by forming a pore in the cell membrane or by acting as a metabolic inhibitor [164]. The ability to design therapeutics with such properties is desirable.

Mechanisms by which CPPs enter the cells have been proposed to be endocytosis, ATP based transport or micropinocytosis [162,163]. But the structural changes that lead to this transport are unclear. Nevertheless, the analysis of various naturally occurring proteins and descriptors for CPP characteristics has led to the discovery and development of novel CPPs. Some of them are tissue-specific and even organelle-specific, thus leading to increased bioavailability and effective targeting. CPPsite [165] is a database that includes around 1800 CPPs and their various characteristics such as nature of sequence, structure, chemical modifications, type of therapeutic cargo delivered and experimental validations. The description of CPPs is collected through data mining. Apart from the already available CPPs, servers such as CPPpred [166] and CellPPD [167] predict whether a query peptide sequence can act as a CPP. CPPpred uses artificial neural networks, and CellPPD uses support vector machines to predict the cell penetration ability of peptides. Since CPPs are not a part of the host system antigen repertoire, they can be immunogenic. Many of the CPPs are not immunogenic by themselves but induce an immune response when delivered with certain drugs [168]. Thus, the major limitation in the use of CPPs is the non-availability of methods to predict immunogenicity of the CPP tagged biotherapeutic.

5.3. Half-life

The half-life of the therapeutic has significant implications for the treatment of diseases. If the rate of clearance of the drug from the system is high, the number of doses needed to maintain its effective concentration would increase. A more rapid clearance rate is undesirable as the mode of administration for a majority of biotherapeutics is through injections. Therefore, once a biotherapeutic is designed, its half-life must be evaluated. *In silico* prediction of the half-life of proteins and peptides allows for rapid estimation of the efficiency of the biotherapeutics. Half Life Prediction [169], SprotP [170] and ProtParam [171] are the web servers that predict the half-lives of proteins and peptides. ProtParam uses the identity of the N-terminal amino acid to estimate the half-life of the protein. There is experimental evidence of a correlation between the N-terminal residue and its half-life. However, the server is organism-specific, and the same N-terminal residue in different biological systems (Yeast, Mammalian, *E. coli*) results in a vastly different estimate of the half-life. Half Life Prediction is based on an SVM technique that considers amino acid composition as a feature. It also suggests mutations that could potentially improve the half-life. SprotP also utilizes SVMs, but the predictions are limited to human cells. The experimentally determined half-lives of peptides are stored in the PEPlife [172] repository, which stores about 1193 unique peptides. In the case of peptide and protein therapeutics, the presence of proteases results in extremely short half-lives, in the order of a few minutes. On account of this, it is essential to modify the peptide or proteins so that they have an enhanced half-life. These modifications include PEGylation [173], glycosylation [174] and co-injection with an unstructured protein XTEN [175]. Apart from this, the fixed chirality of amino acids in biological systems can be exploited. In living systems L-amino acids are incorporated in proteins hence proteases do not recognize D-amino acids. The stability of designed protein and peptide therapeutics can be enhanced by incorporation of D-amino acids instead of L-amino acids [176]. Furthermore, cyclization of peptides ensures protection from proteases. A more detailed account of half-life extension is presented in [177].

5.4. Solubility

Designed proteins and peptides need to be water soluble to prevent aggregation and to increase bioavailability. Most of the biotherapeutics, especially proteins, are synthesized in bacteria as recombinants (commonly in *E. coli*) [178]. Since well-designed proteins already exhibit hydrophilic surface and hydrophobic core, they are water soluble. However, the overexpression of a desired protein or peptide could lead to its aggregation or retention in inclusion bodies [179]. This necessitates biotherapeutics to be expressed in a water-soluble form. The solubility of the biotherapeutics can be assessed by various web servers like ccSOL omics [180] and PROSO II [181]. CcSOL omics uses SVM technique to quantify the solubility of the protein. The descriptors for SVM include, degree of hydrophilicity, hydrophobicity and propensities to exist in secondary structures like α helices, β sheets or coils. PROSO II also uses SVM but the features are the composition of monopeptides and dipeptides. Both these methods have comparable accuracies and are freely available. These tools also suggest mutations that could increase the solubility of the proteins. Better solubility is associated with more bioavailability and for injected therapeutics it ensures maintenance of an effective dose in the serum.

5.5. Immunogenicity

As therapeutic proteins and peptides are foreign entities, their introduction into the body can potentially evoke an immune response. The immune response is highest for subcutaneous injection and progressively reduces with intramuscular, intranasal and intravenous injections. Oral biotherapeutics are least likely to induce an immune response [182]. Not only the mode of administration, but also the dosage and the nature of the biotherapeutic determine its immunogenicity. Post translational modifications such as glycosylation could alter immunogenicity [183]. The immune response is mediated by B cells and T cells. Biotherapeutics on contact with the antigen presenting cells (APCs) get processed into small peptide fragments and are expressed on the membrane of APCs. B cells encounter APCs with these fragments (epitopes) that may or may not be sequentially contiguous and result in an activation of the immune response [184]. The epitopes for T cells are small contiguous peptides that are recognized in conjugation with MHC molecules. On the one hand, high immunogenicity can deter the efficacy of some biotherapeutics by reducing their half-life. On the other hand, this is a desirable characteristic of a vaccine. Hence, it is prudent to predict the immunogenicity of biotherapeutics *in silico*. Various sequence and structure based methods have been proposed to predict the immunogenicity of a therapeutic. Sequence based methods rely on machine learning and quantitative matrices whereas structure based methods use docking, threading and molecular dynamics simulations to infer the binding efficacy of epitopes. Most of these methods have been developed for the prediction of T cell epitopes and as described in previous sections, are extensively used in vaccine design (see reviews [185,186] for more details). However, their scores can also be indicative of the extent of immune activation that they can provide. There have been successful attempts at these predictions. EpiMatrix [187], for example, predicts the immunogenicity with respect to the T cell receptor and has been experimentally shown to be accurate [188]. Prediction of immunogenicity does not help in drug development unless the proteins can be modified to exclude regions responsible for eliciting an immune response. This has been shown to be feasible, by King et al., who have used an SVM based method to predict and eliminate epitopes while retaining protein stability and target specificity [189]. Although their method is capable of predicting T-cell epitopes,

Table 2
A comprehensive list of all clinically approved biotherapeutics compiled from DrugBank [209] and FDA [https://www.accessdata.fda.gov/scripts/cder/daf/]. Biotherapeutics have been listed with their year of approval, type of macromolecule, route of administration, target molecule, location of the target and the disease/disorder they are used to treat.

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
ALBUKED	1942	Protein	Liquid (subcutaneous); solution (intravenous); suspension (intravenous)	Apolipoprotein e; serum amyloid a-1 protein; protein ambp; cytochrome p450 2c9; myeloperoxidase; udp-glucuronosyltransferase1–9	Secreted; endoplasmic reticulum membrane; lysosome; microsome	Severe blood loss, hypervolemia, hypoproteinemia
ACTHAR	1950	Peptide	Gel (intramuscular; subcutaneous); powder	Adrenocorticotropic hormone receptor; corticoliberin; 3 beta-hydroxysteroid dehydrogenase/delta5 –>4-isomerase type 2; 25-hydroxyvitamin d-1 alpha hydroxylase	Cell membrane; secreted; endoplasmic reticulum membrane; mitochondrion membrane; mitochondrion	Infantile spasms, exacerbations of multiple sclerosis, rheumatic; collagen; dermatologic; allergic states; ophthalmic; respiratory; and edematous state
HYPERTET	1957	Protein	Injection (intramuscular)	Immune system	Plasma	Clostridium tetani
FIBRINOLYSIN	1964	Protein	Ointment; topical	Fibrin	Secreted	Minor burns, superficial wounds, ulcers, surgical wounds, and superficial hematomas
SULODEXIDE	1971	Glycan	Oral (capsule)	Heparin cofactor 2; antithrombin-iii	Secreted	Anticoagulant
Hyp Rho-D	1971	Protein	Injection (intramuscular)	(rho) d immunoglobin	Secreted	Rh disease
SOFRADEX	1972	Peptide	Solution/drops (auricular (otic)); solution/drops (ophthalmic); solution (ophthalmic); liquid (ophthalmic); solution/drops (auricular (otic); ophthalmic); liquid (auricular (otic); ophthalmic); cream (topical); solution (auricular (otic); ophthalmic); ointment (auricular (otic); ophthalmic); solution/drops (topical); spray (nasal); ointment (topical)	Multidrug resistance protein 1	Cell membrane	Skin lesions, surface wounds and eye infections
CHORIONIC GONADOTROPIN	1974	Protein	Liquid; powder	Lutropin-choriogonadotropic hormone receptor	Cell membrane	Induction of final follicular maturation, ovulation and early luteinization in infertile women
ALLERGENIC EXTRACT DP	1974	Protein	Solution (percutaneous; intradermal; subcutaneous); liquid (intradermal; subcutaneous; percutaneous); concentrate (intradermal; percutaneous; subcutaneous); concentrate (intradermal; percutaneous; subcutaneous); concentrate (intradermal; subcutaneous); concentrate (intradermal; subcutaneous); injection (cutaneous; intradermal; subcutaneous)	Immune system	Immune cells	Treatment for dust mite allergies.
HYPERRAB	1974	Protein	Liquid (intramuscular); injection (intramuscular); solution (intramuscular);	Immune system	Plasma	Rabies
KINLYTIC	1978	Protein	Injection	Urokinase plasminogen activator surface receptor; plasminogen; urokinase-type plasminogen activator; tissue-type plasminogen activator; plasminogen activator inhibitor 2; plasminogen activator inhibitor 1; plasma serine protease inhibitor; nidogen-1; low-density lipoprotein receptor-related protein 2; suppressor of tumorigenicity 14 protein	Cell membrane; secreted; cytoplasm; membrane	Pulmonary emboli

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
ELSPAR	1978	Protein	Injection (intramuscular; intravenous; subcutaneous)	L-asparagine	Plasma	Acute lymphoblastic leukemia
ILETIN II	1979	Peptide	Injection	Insulin receptor; hla class ii histocompatibility antigen	Cell membrane; nucleus; lysosome; cytoplasm; cytoplasmic vesicle; secreted; membrane; endoplasmic reticulum membrane	Diabetes mellitus
FEIBA VH	1979	Protein	Injection (intravenous)	Blood	Plasma	Hemophilia a or b
PITOCIN	1980	Peptide	Injection (intramuscular; intravenous)	Oxytocin-neurophysin 1; oxytocin receptor; prolyl endopeptidase	Secreted; cell membrane; cytoplasm	Enhance uterine contractions
STREPTASE	1981	Protein	Intracoronary; intravenous	Plasminogen; proteinase-activated receptor 1; cytosolic phospholipase a2	Secreted; cell membrane; cytoplasm	Acute myocardial infarction, deep vein thrombosis, pulmonary embolism, acute or subacute thrombosis of peripheral arteries and chronic occlusive arterial diseases, occlusion of central retinal artery or veins
THYROGLOBULIN	1981	Protein	Tablet (oral)	Thyroid gland	Secreted	Hypothyroidism
ATGAM	1981	Protein	Injection	Thymus lymphocytes	Cell membrane; other cells	Prevention of renal transplant rejection and for the treatment of aplastic anemia
HUMULIN	1982	Peptide	Powder	Insulin receptor; retinoblastoma-associated protein; cathepsin d; insulin-like growth factor 1 receptor; insulin-degrading enzyme; neuroendocrine convertase 2; carboxypeptidase e; neuroendocrine convertase 1; protein nov homolog; low-density lipoprotein receptor-related protein 2; insulin-like growth factor-binding protein 7; synaptotagmin-like protein 4; cytochrome p450 1a2	Cell membrane; nucleus; lysosome; cytoplasm; cytoplasmic vesicle; secreted; membrane; endoplasmic reticulum membrane	Diabetes mellitus
ROFERON-A	1986	Protein	Liquid (intramuscular; subcutaneous); powder	Interferon alpha/beta receptor 1; interferon alpha/beta receptor 2	Membrane	Hairy cell leukemia, aids-related kaposi's sarcoma, condylomata acuminata and chronic myeloid leukemia, chronic hepatitis b, chronic hepatitis c (adults), recurrent or metastatic renal cell carcinoma, non-hodgkin's lymphoma, malignant melanoma
DIGIBIND	1986	Protein	Injection (intravenous)	Digoxin	Secreted	Digitoxin overdose
INTRON	1986	Protein	Injection	Interferon alpha/beta receptor 2; interferon alpha/beta receptor 1; cytochrome p450 1a2	Membrane; endoplasmic reticulum membrane	Hairy cell leukemia, malignant melanoma, follicular lymphoma, condylomata acuminata, aids-related kaposi's sarcoma, chronic hepatitis c, chronic hepatitis b
CIBACALCIN	1986	Peptide	Injection	Thyroid gland	Secreted	Hypercalcemia
CATHFLO	1987	Protein	Injection (intravenous)	Urokinase plasminogen activator surface receptor; fibrinogen alpha chain; plasminogen; plasminogen activator inhibitor 1	Cell membrane; secreted	Acute ischemic stroke (ais), acute myocardial infarction (ami)
PROTAMINE SULFATE	1987	Peptide	Injection	Heparin	Secreted	Reversal of heparin
SUPREFACT	1988	Peptide	Solution (nasal); solution (subcutaneous); liquid (subcutaneous; nasal); spray (nasal);	Lutropin-choriogonadotropic hormone receptor; gonadotropin-releasing hormone receptor; cytochrome p450 19a1	Cell membrane; membrane	Prostate cancer

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
EPOGEN/PROCRIT	1989	Protein	Injection	Erythropoietin receptor	Cell membrane	Anemia
ALFERON	1989	Protein	Injection (subcutaneous)	Interferon alpha/beta receptor 1; interferon alpha/beta receptor 2	Membrane	Condylomata acuminata.
EMINASE	1989	Protein	Injection (intravenous)	Urokinase plasminogen activator surface receptor; fibrinogen alpha chain; plasminogen; plasminogen activator inhibitor 1	Cell membrane; secreted	For lysis of acute pulmonary emboli, intracoronary emboli and management of myocardial infarction
ENGERIX B	1989	Virus	Injection (intramuscular)	Immune system	Plasma	Hepatitis b
GEREF	1990	Peptide	Injection	Growth hormone-releasing hormone receptor	Cell membrane	Idiopathic growth hormone deficiency
ADAGEN	1990	Protein	Injection	Adenosine; growth factor receptor-bound protein 2	Nucleus	Severe combined immunodeficiency disease
ALPHANINE SD	1990	Protein	(intravenous); powder	Blood	Plasma	Hemophilia b
LEUKINE	1991	Protein	Injection	Granulocyte-macrophage colony-stimulating factor receptor subunit alpha; bone marrow proteoglycan; interleukin-3 receptor subunit alpha; cytokine receptor common subunit beta; syndecan-2	Cell membrane; secreted; membrane	Acute myelogenous leukemia
BOTOX	1991	Protein	Injection	Synaptosomal-associated protein 25; rho-related gtp-binding protein rhob	Cytoplasm; late endosome membrane	Chronic migraine, upper limb spasticity, cervical dystonia, axillary hyperhidrosis, blepharospasm, strabismus
CEREDASE	1991	Protein	Injection	Glucocerebroside	Plasma	Type 1 gaucher disease
NEUPOGEN	1991	Protein	Injection	Granulocyte colony-stimulating factor receptor; neutrophil elastase	Secreted; cytoplasmic	Neutropenia
SURVANTA	1991	Protein	Suspension (endotracheal); suspension (endotracheal)	Lung	Secreted	Respiratory distress syndrome (rds)
THROMBATE III	1991	Protein	Powder	Blood	Plasma	Hereditary antithrombin iii deficiency (at-iii) in surgical or obstetrical procedures and thromboembolism
LENOGRASTIM	1991	Protein	Injection	Granulocyte colony-stimulating factor receptor	Secreted	Bone marrow transplantation, cytotoxic-induced neutropenia, mobilisation of peripheral blood progenitor cells, neutropenia, reduction in the duration of neutropenia following bone marrow transplantation.
PROLEUKIN	1992	Protein	Injection (intravenous)	Interleukin-2 receptor subunit beta; interleukin-2 receptor subunit alpha; cytokine receptor common subunit gamma; prostaglandin g/h synthase 2; cytosolic phospholipase a2; cytochrome p450 3a4; xanthine dehydrogenase/oxidase; cytochrome p450 2e1	Membrane; microsome membrane; cytoplasm; endoplasmic reticulum membrane	Metastatic renal cell carcinoma, metastatic melanoma (adults)
ORTHOCLONE	1992	Protein	Injection	T-cell surface glycoprotein cd3 epsilon chain; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc	Cell membrane; cytoplasmic; secreted; cytoplasm; membrane	Renal transplant acute rejection

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
				region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; t-cell surface glycoprotein cd3 delta chain; t-cell surface glycoprotein cd3 amma chain; t-cell surface glycoprotein cd3 zeta chain		
PULMOZYME	1993	Protein	Solution (respiratory (inhalation));	DNA	Nucleus	Cystic fibrosis
BETASERON	1993	Protein	Injection	Interferon alpha/beta receptor 1; interferon alpha/beta receptor 2	Membrane	Multiple sclerosis
TRASYLOL	1993	Protein	Injection	Trypsin-1; chymotrypsinogen b; plasminogen; kallikrein-1	Secreted; cytoplasmic	Reduce perioperative blood loss and the need for blood transfusion in patients undergoing cardiopulmonary bypass
CEREZYME	1994	Protein	Injection	Glucocerebroside	Secreted	Non-neuronopathic (type 1) or chronic neuronopathic (type 3) gaucher disease
REOPRO	1994	Protein	Injection	Integrin beta-3; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; integrin alpha-iib; vitronectin	Cell membrane; cytoplasmic; secreted; cytoplasm; membrane	Unstable angina, prevention of problems in percutaneous coronary intervention
ONCASPAR	1994	Protein	Injection	L-asparagine	Plasma	Acute lymphoblastic leukemia
TARGOCID	1994	Peptide	Injection (intramuscular; intravenous);	D-ala-d-ala moiety of nam/nag peptide subunits of peptidoglycan	Bacterial membrane	Osteomyelitis, septic arthritis, non-cardiac bacteremia, septicaemia
GENOTROPIN	1995	Protein	Injection (subcutaneous)	Growth hormone receptor; prolactin receptor	Cell membrane; membrane	Growth hormone deficiency (ghd), prader-willi syndrome, small for gestational age, turner syndrome, and idiopathic short stature
HAVRIX	1995	Virus	Injection (intramuscular)	Immune system	Immune cells	Hepatitis a
HUMALOG	1996	Protein	Injection (intravenous)	Insulin receptor; insulin-like growth factor 1 receptor; cytochrome p450 1a2	Cell membrane; endoplasmic reticulum membrane	Diabetes mellitus
AVONEX	1996	Protein	Injection	Interferon alpha/beta receptor 1; interferon alpha/beta receptor 2	Membrane	Multiple sclerosis
COTAZYM	1996	Protein	Capsule; oral	Dietary fat; dietary protein; dietary starch	Secreted	Exocrine pancreatic insufficiency in cystic fibrosis, chronic pancreatitis
COPAXONE	1996	Protein	Injection	Hla class ii histocompatibility antigen	Cell membrane	Multiple sclerosis
WELLFERON	1997	Protein	Liquid (intramuscular; subcutaneous);	Interferon alpha/beta receptor 2; interferon alpha/beta receptor 1	Membrane	Hairy cell leukemia

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
NEUMEGA	1997	Protein	Injection (subcutaneous)	Interleukin-11 receptor subunit alpha	Membrane	Reduced platelets
INFERGEN	1997	Protein	Solution (subcutaneous)	Interferon alpha/beta receptor 1; interferon alpha/beta receptor 2	Membrane	Chronic hepatitis c
RITUXAN	1997	Protein	Injection	Low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; b-lymphocyte antigen cd20	Cell membrane; cytoplasmic; secreted; cytoplasm	Non-hodgkin's lymphoma (nhl), chronic lymphocytic leukemia (cll), rheumatoid arthritis (ra)
REGRANEX	1997	Protein	Gel (cutaneous); gel (topical);	Platelet-derived growth factor receptor beta; alpha-2-macroglobulin; platelet-derived growth factor receptor alpha	Cell membrane; secreted	Lower extremity diabetic neuropathic ulcers
GONAL-F	1997	Protein	Injection	Primary sexual organs	Cells	Induction of ovulation and pregnancy in the oligo-anovulatory infertile patient
CARTICEL	1997	Cells	Intra-articular	Cartilage	Cells	Cartilage defects
REFLUDAN	1998	Protein	Injection	Prothrombin	Secreted	Anticoagulation in patients with heparin-associated thrombocytopenia
ENBREL	1998	Protein	Injection	Tumor necrosis factor; lymphotoxin-alpha; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; tumor necrosis factor receptor superfamily member 1b; prostaglandin g/h synthase 2	Cell membrane; secreted; cytoplasmic; cytoplasm; microsome membrane	Rheumatoid arthritis
RETAVASE	1998	Protein	Injection (intravenous)	Urokinase plasminogen activator surface receptor; fibrinogen alpha chain; plasminogen; plasminogen activator inhibitor 1	Cell membrane; secreted	Acute myocardial infarction (ami) the reduction of the incidence of congestive heart failure
GLUCAGEN	1998	Peptide	Injection	Glucagon receptor; glucagon-like peptide 2 receptor; glucagon-like peptide 1 receptor	Cell membrane	Hypoglycemia

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
LYMERIX	1998	Protein	Injection (intramuscular)	Toll-like receptor 2	Membrane	Lyme disease (vaccine)
REMICADE	1998	Protein	Injection	Tumor necrosis factor	Cell membrane	Crohn's disease
HERCEPTIN	1998	Protein	Injection	Low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; receptor tyrosine-protein kinase erbb-2; epidermal growth factor receptor; cytochrome p450 19a1	Cell membrane; cytoplasmic; secreted; cytoplasm; membrane	Her2 overexpressing breast cancer, metastatic gastric or gastroesophageal junction adenocarcinoma
SIMULECT	1998	Protein	Injection	Interleukin-2 receptor subunit alpha; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; interleukin-2 receptor subunit beta; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Membrane; cell membrane; cytoplasmic; secreted; cytoplasm	Acute organ rejection
THYMOGLOBULIN	1998	Protein	Injection (intravenous)	T-cell surface glycoprotein cd1a; major histocompatibility complex class i-related gene protein; integrin alpha-l; t-lymphocyte activation antigen cd86; low affinity immunoglobulin gamma fc region receptor ii-b; t-cell surface glycoprotein cd4; integrin beta-1; integrin alpha-v; integrin beta-3	Cell membrane; membrane	Renal transplant acute rejection
SYNAGIS	1998	Protein	Injection	Fusion glycoprotein f0; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high	Virion membrane; cell membrane; cytoplasmic; secreted; cytoplasm	Respiratory syncytial virus (rsv)

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
				affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c		
SUCRAID	1998	Protein	Solution (oral)	Sucrose	Gastrointestinal tract	Congenital sucrose-isomaltase deficiency (csid)
ONTAK	1999	Protein	Injection	Interleukin-2 receptor subunit alpha; cytokine receptor common subunit gamma; interleukin-2 receptor subunit beta	Membrane	Cutaneous t-cell lymphoma
ACTIMMUNE	1999	Protein	Injection	Interferon gamma receptor 1; interferon gamma receptor 2; cytochrome p450 1a2	Membrane; endoplasmic reticulum membrane	Chronic granulomatous disease (cgd) associated infections, malignant osteopetrosis
NOVOSEVEN	1999	Protein	Injection (intravenous)	Coagulation factor x; tissue factor; serine protease hepsin; tissue factor pathway inhibitor; vitamin k-dependent gamma-carboxylase; coagulation factor vii	Secreted; membrane; endoplasmic reticulum membrane	Glanzmann's thrombasthenia, congenital fvii deficiency, congenital or acquired hemophilia
STEMGEN	1999	Protein	Injection	Hematopoietic progenitors	Cells	Autologous peripheral blood progenitor cell (pbpc) transplantation
CUROSURF	1999	Protein	Suspension (endotracheal);	Lung	Secreted	Respiratory distress syndrome (rds)
TNKASE	2000	Protein	Injection	Urokinase plasminogen activator surface receptor; plasminogen activator inhibitor 1; fibrinogen alpha chain; plasminogen activator inhibitor 2; tetranectin; keratin	Cell membrane; secreted; cytoplasm; endoplasmic reticulum lumen; endoplasmic reticulum membrane	Acute myocardial infarction
MYOBLOC	2000	Protein	Injection	Vesicle-associated membrane protein 1; vesicle-associated membrane protein 2; synaptotagmin-2	Cytoplasmic vesicle	Cervical dystonia
MYLOTARG	2000	Protein	Injection	Myeloid cell surface antigen cd33; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Cell membrane; cytoplasmic; secreted; cytoplasm	Cd33 positive acute myeloid leukemia
OVIDREL	2000	Protein	Solution (subcutaneous);	Follicle-stimulating hormone receptor; lutropin-choriogonadotropic hormone receptor	Cell membrane	Induction of final follicular maturation, ovulation and early luteinization in infertile women
NOVOLOG	2000	Protein	Injection	Insulin receptor; cytochrome p450 1a2	Cell membrane; endoplasmic reticulum membrane	Diabetes mellitus

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
ZADAXIN	2000	Peptide	Injection (subcutaneous)	T-cell	Plasma	Hepatitis b and c
ARANESP	2001	Protein	Injection	Erythropoietin receptor	Cell membrane	Anemia
SYLATRON	2001	Protein	Injection (subcutaneous)	Interferon alpha/beta receptor 1; interferon alpha/beta receptor 2; cytochrome p450 1a2; cytochrome p450 2d6	Membrane; endoplasmic reticulum membrane	Melanoma
KINERET	2001	Protein	Injection	Interleukin-1 receptor type 1	Membrane	Rheumatoid arthritis
XIGRIS	2001	Protein	Injection (intravenous)	Coagulation factor viii; coagulation factor v; plasminogen activator inhibitor 1; thrombomodulin; vitamin k-dependent protein s; ceruloplasmin; prothrombin; platelet factor 4; plasma serine protease inhibitor; serpin b6; vitamin k-dependent gamma-carboxylase; endothelial protein c receptor	Secreted; membrane; cytoplasm; endoplasmic reticulum membrane	Severe sepsis
CAMPATH	2001	Protein	Injection (intravenous)	Campath-1 antigen; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Cell membrane; cytoplasmic; secreted; cytoplasm	B-cell chronic lymphocytic leukemia (b-cll)
NATRECOR	2001	Peptide	Injection	Atrial natriuretic peptide receptor 1; atrial natriuretic peptide receptor 2; atrial natriuretic peptide receptor 3	Membrane; cell membrane	Acutely decompensated congestive heart failure with dyspnea at rest
ELIGARD	2002	Peptide	Injection	Gonadotropin-releasing hormone receptor	Cell membrane	Prostate cancer
PEGASYS	2002	Protein	Injection	Interferon alpha/beta receptor 2; interferon alpha/beta receptor 1; cytochrome p450 1a2	Membrane; endoplasmic reticulum membrane	Hepatitis b and c
NEULASTA	2002	Protein	Injection (subcutaneous); solution (subcutaneous)	Granulocyte colony-stimulating factor receptor; neutrophil elastase	Secreted; cytoplasmic	Mobilization of peripheral blood progenitor cells, neutropenia
ELITEK	2002	Protein	Injection	Uric acid	Secreted	Management of plasma uric acid levels in pediatric patients with leukemia, lymphoma, and solid tumor malignancies
HUMIRA	2002	Protein	Injection	Tumor necrosis factor; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin	Cell membrane; cytoplasmic; secreted; cytoplasm	Rheumatoid arthritis

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
				gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c		
ZEVALIN	2002	Protein	Injection (intravenous)	B-lymphocyte antigen cd20; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Cell membrane; cytoplasmic; secreted; cytoplasm	Follicular b-cell non-hodgkin's lymphoma (nhl)
BRAVELLE	2002	Protein	(subcutaneous); powder	Follicle-stimulating hormone receptor	Cell membrane	Induction of ovulation
FORTEO	2002	Peptide	Injection	Parathyroid hormone/parathyroid hormone-related peptide receptor	Cell membrane	Osteoporosis
XOLAIR	2003	Protein	Injection	High affinity immunoglobulin epsilon receptor subunit alpha; high affinity immunoglobulin epsilon receptor subunit beta	Cell membrane; membrane	Asthma
ZEMAIRA	2003	Protein	Injection (intravenous)	Neutrophil elastase	Cytoplasmic	Alpha1-proteinase inhibitor deficiency
CUBICIN	2003	Peptide	Injection	Bacterial outer membrane; lipoteichoic acid synthesis	Bacterial membrane	Skin and skin structure infections, staphylococcus aureus bloodstream infections (bacteremia), including those with rightsided infective endocarditis
SOMAVERT	2003	Protein	Injection	Growth hormone receptor; sterol 26-hydroxylase	Cell membrane; mitochondrion membrane; cytoplasm; endoplasmic reticulum membrane; secreted	Acromegaly
ALDURAZYME	2003	Protein	Injection	Iduronic acid	Secreted	Hurler and hurler-scheie forms of mucopolysaccharidosis i (mps i)
AMEVIVE	2003	Protein	Injection (intramuscular; intravenous)	T-cell surface antigen cd2; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low	Membrane; cell membrane; cytoplasmic; secreted; cytoplasm	Plaque psoriasis

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
				affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c		
RAPTIVA	2003	Protein	Injection	Integrin alpha-l; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Membrane; cell membrane; cytoplasmic; secreted; cytoplasm	Plaque psoriasis
FABRAZYME	2003	Protein	Injection	Globotriaosylceramide	Secreted	Fabry disease
PLENAXIS	2003	Peptide	Injection	Gonadotropin-releasing hormone receptor	Cell membrane	Prostate cancer
FUZEON	2003	Peptide	Injection	Envelope glycoprotein; cytochrome p450 2c19; cytochrome p450 2e1	Endoplasmic reticulum membrane	Hiv-1
IPRIVASK	2003	Protein	(subcutaneous)	Carboxypeptidase a1	Secreted	Prophylaxis for deep vein thrombosis
ERBITUX	2004	Protein	Injection	Epidermal growth factor receptor; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; complement c1s subcomponent; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Cell membrane; cytoplasmic; secreted; cytoplasm	Metastatic colorectal cancer, squamous cell cancer of the head and neck
MENOPUR	2004	Protein	Injection (intramuscular; subcutaneous)	Follicle-stimulating hormone receptor; lutropin-choriogonadotropic hormone receptor	Cell membrane	Development of multiple follicles and pregnancy
KEPIVANCE	2004	Protein	Injection	Fibroblast growth factor receptor 2; neuropilin-1; fibroblast growth factor receptor 1; fibroblast growth factor receptor 4; fibroblast growth factor receptor 3; basement membrane-specific heparan sulfate proteoglycan core protein	Cell membrane; secreted	Oral mucositis
LUVERIS	2004	Protein	Injection (subcutaneous)	Lutropin-choriogonadotropic hormone receptor	Cell membrane	Stimulation of follicular development in infertile hypogonadotropic hypogonadal women

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
FOLLISTIM AQ	2004	Protein	Injection	Follicle-stimulating hormone receptor	Cell membrane	Release of multiple folicles, induction of ovulation and pregnancy in anovulatory infertile patients
VITRASE	2004	Protein	Injection (subcutaneous)	Transforming growth factor beta-1; serum albumin; hyaluronan	Secreted	Adjunct, hypodermoclysis
TYSABRI	2004	Protein	Injection (intravenous); solution (intravenous)	Integrin alpha-4; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; intercellular adhesion molecule 1	Membrane; cell membrane; cytoplasmic; secreted; cytoplasm	Multiple sclerosis, crohn's disease
AVASTIN	2004	Protein	Injection	Low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; vascular endothelial growth factor a	Cell membrane; cytoplasmic; secreted; cytoplasm	Metastatic colorectal cancer, non-squamous non-small cell lung cancer, glioblastoma, cervical cancer, metastatic renal cell carcinoma,
APIDRA	2004	Protein	Injection	Insulin receptor; cytochrome p450 1a2	Cell membrane; endoplasmic reticulum membrane	Diabetes mellitus
MACUGEN	2004	Protein	Injection	Neuropilin-1	Cell membrane	Neovascular (wet) age-related macular degeneration
MULTIFERON	2004	Protein	Injection (subcutaneous)	Interferon alpha/beta receptor 1	Membrane	Malignant melanoma
INCRELEX	2005	Protein	Injection	Insulin-like growth factor 1 receptor; insulin-like growth factor-binding protein 3; insulin receptor; cation-independent mannose-6-phosphate receptor	Cell membrane; secreted; lysosome membrane	Primary insulin-like growth factor-1 deficiency (primary igfd)
SYMLIN	2005	Peptide	Injection (subcutaneous)	Receptor activity-modifying protein 1; receptor activity-modifying protein 2; receptor activity-modifying protein 3; calcitonin receptor	Membrane; cell membrane	Type 1 and type 2 diabetes
NAGLAZYME	2005	Protein	Injection	Perilipin-3; dermatan sulfate	Cytoplasm	Mucopolysaccharidosis vi (mps vi; maroteaux-lamy syndrome)
ORENCIA	2005	Protein	Injection	T-lymphocyte activation antigen cd80; t-lymphocyte activation antigen cd86	Membrane; cell membrane	Rheumatoid arthritis, juvenile idiopathic arthritis

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
LEVEMIR	2005	Protein	Injection	Insulin receptor; serum albumin; cytochrome p450 1a2	Cell membrane; secreted; endoplasmic reticulum membrane	Diabetes mellitus
HYLENEX	2005	Protein	Injection	Hyaluronan; transforming growth factor beta-1; serum albumin	Secreted	Adjuvant
VECTIBIX	2006	Protein	Injection	Epidermal growth factor receptor	Cell membrane	Metastatic colorectal carcinoma
LUCENTIS	2006	Protein	Injection	Vascular endothelial growth factor a	Secreted	Neovascular (wet) age-related macular degeneration
MYOZYME	2006	Protein	Injection (intravenous)	Perilipin-3; dermatan sulfate; heparan sulfate	Cytoplasm	Pompe disease (gaa deficiency)
HEPAGAM B	2006	Protein	Injection	Hbsag	Virion membrane	Hepatitis b
SOLIRIS	2007	Protein	Injection	Complement c5	Secreted	Paroxysmal nocturnal hemoglobinuria, atypical hemolytic uremic syndrome
MIRCERA	2007	Protein	Injection	Erythropoietin receptor	Cell membrane	Anaemia associated with chronic kidney disease
NPLATE	2008	Protein	Injection	Thrombopoietin receptor	Cell membrane	Thrombocytopenia
RECOMODULIN	2008	Protein	Injection (intravenous)	Coagulation factor v; prothrombin	Secreted	Disseminated intravascular coagulation
ARCALYST	2008	Protein	Injection	Interleukin-1 beta; interleukin-1 alpha; interleukin-1 receptor antagonist protein	Cytoplasm; secreted	Caps, also known as cryopyrin-associated periodic syndromes, including familial cold auto-inflammatory syndrome (fcas) and muckle-wells syndrome (mws)
CIMZIA	2008	Protein	Injection	Tumor necrosis factor	Cell membrane	Rheumatoid arthritis and crohn's disease
SILAPO	2008	Protein	Injection	Erythropoietin receptor	Cell membrane	Treatment of anaemia associated with chronic renal failure
CALCITONIN-SALMON	2009	Peptide	Liquid (intramuscular; subcutaneous); solution (nasal); solution (intramuscular; subcutaneous); spray	Calcitonin receptor	Cell membrane	Postmenopausal osteoperosis
WILATE	2009	Protein	Injection	Coagulation factor x; phytanoyl-coa dioxygenase	Secreted; peroxisome; membrane; endoplasmic reticulum lumen; endoplasmic reticulum membrane; endoplasmic reticulum-golgi intermediate compartment membrane; cell membrane; endoplasmic reticulum-golgi intermediate compartment	Von willebrand
STELARA	2009	Protein	Injection	Interleukin-12 subunit beta; interleukin-23	Secreted	Plaque psoriasis and psoriatic arthritis, crohn's disease.
ILARIS	2009	Protein	Injection	Interleukin-1 beta	Cytoplasm	Familial cold autoinflammatory syndrome (fcas) and muckle-wells syndrome (mws), which are both part of the cryopyrin-associated periodic syndromes (caps) as well as for patients 2 years of age and older to treat systemic juvenile idiopathic arthritis (sjia).
BERINERT	2009	Peptide	Injection (intravenous)	Complement c1r subcomponent; complement c1s subcomponent; plasma kallikrein; coagulation factor xii; prothrombin; coagulation factor xi; tissue-type plasminogen activator	Cytoplasmic; secreted	Acute abdominal, facial, or laryngeal attacks of hereditary angioedema
ARZERRA	2009	Protein	Injection	B-lymphocyte antigen cd20	Cell membrane	Chronic lymphocytic leukemia (cll)
SIMPONI	2009	Protein	Injection	Tumor necrosis factor	Cell membrane	Active rheumatoid arthritis (ra), active psoriatic arthritis (psa), juvenile idiopathic arthritis, active ankylosing spondylitis (as), ulcerative colitis (uc)

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
RIASTAP	2009	Protein	Powder (topical); injection	Blood	Plasma	Congenital fibrinogen deficiency, afibrinogenemia and hypofibrinogenemia.
ATRYN	2009	Protein	Injection	Blood	Plasma	Hereditary antithrombin deficiency
KALBITOR	2009	Protein	Injection	Kallikrein	Secreted	Hereditary angioedema
XIAFLEX	2010	Protein	Ointment (topical); powder	Collagen alpha-1(i) chain; collagen alpha-1(ii) chain; collagen alpha-1(iii) chain; collagen alpha-2(i) chain	Secreted	Dupuytren's contracture
LUMIZYME	2010	Protein	Injection	Cation-dependent mannose-6-phosphate receptor; glycogen	Lysosome membrane	Pompe disease (gaa deficiency)
ACTEMRA	2010	Protein	Injection	Interleukin-6 receptor subunit alpha	Basolateral cell membrane	Rheumatoid arthritis
XGEVA	2010	Protein	Injection (subcutaneous); solution (subcutaneous); injection (subcutaneous)	Tumor necrosis factor ligand superfamily member 11	Cell membrane	Giant cell tumor of bone
PROVENGE	2010	Cells	Injection (intravenous)	Prostatic acid phosphatase	Secreted	Asymptomatic or minimally symptomatic metastatic castrate resistant (hormone refractory) prostate cancer
VPRIV	2010	Protein	Injection	Glucosylceramidase	Lysosome membrane	Type 1 gaucher disease
KRYSTEXXA	2010	Protein	Injection	Uric acid	Plasma	Chronic gout
PROFILNINE SD	2010	Protein	Injection (intravenous)	Blood	Plasma	Hemophilia b patients, reversal of vitamin k antagonist (vka, e.g., warfarin)
YERVOY	2011	Protein	Injection (intravenous); liquid (intravenous)	Cytotoxic t-lymphocyte protein 4	Cell membrane	Unresectable or metastatic melanoma
NULOJIX	2011	Protein	Injection	T-lymphocyte activation antigen cd86; t-lymphocyte activation antigen cd80	Cell membrane; membrane	Rheumatoid arthritis
ADCETRIS	2011	Protein	Injection	Tumor necrosis factor receptor superfamily member 8; cytochrome p450 3a4; cytochrome p450 3a5; multidrug resistance protein 1	Cell membrane; endoplasmic reticulum membrane	Hodgkin lymphoma and systemic anaplastic large cell lymphoma
BENLYSTA	2011	Protein	Injection	Tumor necrosis factor ligand superfamily member 13b	Cell membrane	Systemic lupus erythematosus (sle)
EYLEA	2011	Protein	Injection	Vascular endothelial growth factor a; placenta growth factor; vascular endothelial growth factor b	Secreted	Neovascular age-related macular degeneration (amd)
ERWINAZE	2011	Protein	Injection	Asparagine	Plasma	Acute lymphoblastic leukemia (all)
BEXXAR THEREPY	2012	Protein	Solution (intravenous)	B-lymphocyte antigen cd20; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Cell membrane; cytoplasmic; secreted; cytoplasm	Cd20 positive, relapsed or refractory, low-grade, follicular, or transformed nonhodgkin's lymphoma
BYDUREON	2012	Peptide	Injection	Glucagon-like peptide 1 receptor	Cell membrane	Type 2 diabetes
LUCINACTANT	2012	Peptide	Intratracheal	Endogenous human surfactant protein b mimic	Secreted	Respiratory distress syndrome (rds)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
PERJETA	2012	Protein	Injection	Receptor tyrosine-protein kinase erbb-2	Cell membrane	Her2-positive metastatic breast cancer
ELELYSO	2012	Protein	Injection	Glucocerebroside	Secreted	Type 1 gaucher's disease.
JETREA	2012	Protein	Injection	Alpha-2-macroglobulin; alpha-2-antiplasmin; fibronectin	Secreted	Symptomatic vitreomacular adhesion
VORAXAZE	2012	Protein	Injection	Methotrexate	Secreted	Toxic plasma methotrexate concentrations
GATTEX	2012	Peptide	Injection	Glucagon-like peptide 2 receptor	Cell membrane	Short bowel syndrome (sbs), malabsorption associated with the removal of the intestine
RAXIBACUMAB	2012	Protein	Injection (intravenous)	Protective antigen	Secreted	Anthrax
VARIZIG	2012	Protein	Liquid (intramuscular); injection (intramuscular; intravenous); solution (intramuscular; intravenous)	Virus	Virus	Varicella zoster virus (anti-vzv).
GRANIX	2012	Protein	Injection	G-csf receptors	Cell membrane	Neutropenia
BIVIGAM	2013	Protein	Injection	Low affinity immunoglobulin gamma fc region receptor iii-b; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c; high affinity immunoglobulin gamma fc receptor ib; complement c3; complement c4-a; complement c4-b; complement c5	Cell membrane; cytoplasm; secreted	Primary humoral immunodeficiency
KADCYLA	2013	Protein	Injection	Cytochrome p450 3a4; cytochrome p450 3a5; receptor tyrosine-protein kinase erbb-2; multidrug resistance protein 1	Endoplasmic reticulum membrane; cell membrane	Metastatic breast cancer
GAZYVA	2013	Protein	Injection	B-lymphocyte antigen cd20	Cell membrane	Chronic lymphocytic leukemia
NOVOEIGHT	2013	Protein	Injection	Blood	Secreted	Hemophilia a
TRETTEN	2013	Protein	Injection	Blood	Plasma	Congenital factor xiii a-subunit deficiency
KCENTRA	2013	Protein	Injection (intravenous)	Blood	Plasma	Reversal of vitamin k agonist
VASOSTRICT	2014	Peptide	Injection (intramuscular; subcutaneous); liquid (intramuscular; subcutaneous);	Vasopressin v1a receptor; vasopressin v2 receptor; vasopressin v1b receptor; canalicular multispecific organic anion transporter 1	Cell membrane; apical cell membrane	Vasodilatory shock
RAGWITEK	2014	Mixed (extract of cell)	Oral	Immune system	Plasma	Ragweed allergy
CYRAMZA	2014	Protein	Solution (intravenous);	Vascular endothelial growth factor receptor 2	Cell junction	Advanced gastric or gastro-esophageal junction adenocarcinoma
SAXENDA	2014	Peptide	Injection	Glucagon-like peptide 1 receptor; dipeptidyl peptidase 4; neprilysin	Cell membrane; secreted	Weight management
ENTYVIO	2014	Protein	Injection	Integrin alpha-4; integrin beta-7	Membrane	Ulcerative colitis and crohn's disease
OPDIVO	2014	Protein	Injection (intravenous);	Programmed cell death protein 1	Membrane	Unresectable (cannot be surgically removed) or metastatic melanoma
SYLVANT	2014	Protein	Injection	Interleukin-6	Secreted	Multicentric castleman's disease (mcd)
KEYTRUDA	2014	Protein	Injection	Programmed cell death protein 1	Membrane	Unresectable or metastatic melanoma
TANZEUM	2014	Peptide	Injection	Glucagon-like peptide 1 receptor	Cell membrane	Type 2 diabetes

(*continued on next page*)

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
TRULICITY	2014	Peptide	Injection	Glucagon-like peptide 1 receptor	Cell membrane	Type 2 diabetes
MYALEPT	2014	Protein	Injection	Leptin receptor	Cell membrane	Congenital or acquired generalized lipodystrophy
VIMIZIM	2014	Protein	Injection	N-acetylgalactosamine-6-sulfatase	Lysosome	Morquio syndrome
BLINCYTO	2014	Protein	(intravenous); powder	B-lymphocyte antigen cd19; t-cell surface glycoprotein cd3 delta chain	Membrane	Philadelphia chromosome-negative relapsed or refractory b-cell precursor acute lymphoblastic leukemia (all)
PLEGRIDY	2014	Protein	Injection	Unknown	Unknown	Multiple sclerosis
RUCONEST	2014	Peptide	Injection	Complement c1r subcomponent; complement c1s subcomponent; plasma kallikrein; coagulation factor xii; prothrombin; coagulation factor xi; tissue-type plasminogen activator	Cytoplasmic; secreted	Hereditary angioedema
OBIZUR	2014	Protein	Injection	Blood	Plasma	Acquired haemophilia a (aha).
ELOCTATE	2014	Protein	Injection (intravenous)	Blood	Plasma	Hemophilia a
ALPROLIX	2014	Protein	Injection (intravenous)	Blood	Plasma	Hemophilia b
BASAGLAR	2015	Protein	Injection	Insulin receptor; insulin-like growth factor 1 receptor; cytochrome p450 1a2	Cell membrane; endoplasmic reticulum membrane	Type 1 and type 2 (adults) diabetes
DEFITELIO	2015	Nucleic acid	Injection (intravenous)	Adenosine receptor a1; adenosine receptor a2a; adenosine receptor a2b	Cell membrane	Hepatic veno-occlusive disease, with renal or pulmonary dysfunction following hematopoietic stem-cell transplantation (hsct)
NATPARA	2015	Protein	Injection	Parathyroid hormone/parathyroid hormone-related peptide receptor; parathyroid hormone 2 receptor	Cell membrane	Hypocalcemia
EMPLICITI	2015	Protein	Injection	Slam family member 7	Membrane	Multiple myeloma
NUCALA	2015	Protein	Injection	Interleukin-5	Secreted	Asthma
COSENTYX	2015	Protein	Injection (subcutaneous); powder	Interleukin-17a	Secreted	Uveitis, rheumatoid arthritis, ankylosing spondylitis, and psoriasis.
ANTHRASIL	2015	Protein	Liquid (intravenous)	Protective antigen	Secreted	Anthrax
UNITUXIN	2015	Protein	Injection (intravenous)	Ganglioside gd2	Cell membrane	High-risk neuroblastoma
STRENSIQ	2015	Protein	Injection	Sphingosine 1-phosphate receptor 1; pyrophosphate	Cell membrane	Hypophosphatasia
NUWIQ	2015	Protein	Injection (intravenous)	Blood	Secreted	Hemophilia a
PRAXBIND	2015	Protein	Injection (intravenous)	Dabigatran etexilate	External drug inhibitor	Reversal of dabigatran
PRALUENT	2015	Protein	Injection	Proprotein convertase subtilisin/kexin type 9	Cytoplasm	High cholesterol
REPATHA	2015	Protein	Injection	Proprotein convertase subtilisin/kexin type 9	Cytoplasm	Heterozygous/homozygous familial hypercholesterolemia or clinical atherosclerotic cardiovascular disease
ADYNOVATE	2015	Protein	Injection (intravenous)	Blood	Plasma	Hemophilia a
DARZALEX	2015	Protein	Injection	Adp-ribosyl cyclase 1	Membrane	Multiple myeloma
PORTRAZZA	2015	Protein	Solution (intravenous)	Epidermal growth factor receptor	Cell membrane	Non-small cell lung cancer (nsclc)
ZARXIO	2015	Protein	Injection	Granulocyte colony-stimulating factor receptor	Secreted	Severe chronic or acute neutropenia
TRESIBA	2015	Peptide	Injection	Insulin receptor	Cell membrane	Diabetes mellitus
KANUMA	2015	Protein	Injection	Cholesteryl esters and triglycerides	Cell membrane	Lysosomal acid lipase deficiency (lal-d)
COAGADEX	2015	Protein	Injection (intravenous)	Blood	Plasma	Heriditary factor x deficiency
EPTIFIBATIDE	2016	Peptide	Injection (intravenous);	Integrin beta-3	Cell membrane	Acute coronary syndrome
IDELVION	2016	Protein	Powder	Coagulation factor x; coagulation factor xi; coagulation factor viii; prothrombin; coagulation factor vii; vitamin k-dependent gamma-carboxylase; prolow-density lipoprotein receptor-related protein 1	Secreted; endoplasmic reticulum membrane; cell membrane	Hemophilia b

Table 2 (*continued*)

Name	Year of Approval	Type of macromolecule	Route of administration	Target	Target location	Disease or disorder
ZINBRYTA	2016	Protein	Injection	Interleukin-2 receptor subunit alpha; low affinity immunoglobulin gamma fc region receptor iii-b; complement c1r subcomponent; complement c1q subcomponent subunit a; complement c1q subcomponent subunit b; complement c1q subcomponent subunit c; low affinity immunoglobulin gamma fc region receptor iii-a; high affinity immunoglobulin gamma fc receptor i; low affinity immunoglobulin gamma fc region receptor ii-a; interleukin-2 receptor subunit beta; low affinity immunoglobulin gamma fc region receptor ii-b; low affinity immunoglobulin gamma fc region receptor ii-c	Membrane; cell membrane; cytoplasmic; secreted; cytoplasm	Multiple sclerosis
ANTHIM	2016	Protein	Solution (intravenous)	Anthrax toxin	Plasma	Anthrax
EXONDYS 51	2016	Nucleic acid	Injection (intravenous)	Dmd-001 gene (exon 51 target site)	Nucleus	Duchenne muscular dystrophy
LARTRUVO	2016	Protein	Injection (intravenous)	Platelet-derived growth factor receptor alpha	Cell membrane	Soft tissue sarcoma
TALTZ	2016	Protein	Injection	Interleukin-17a	Secreted	Plaque psoriasis
TECENTRIQ	2016	Protein	Injection	Programmed cell death 1 ligand 1	Cell membrane	Locally advanced or metastatic urothelial carcinoma
ZINPLAVA	2016	Protein	Injection	Clostridium difficile toxin b	Toxin	Neutralizes clostridium difficile toxin b
SPINRAZA	2016	Protein	Injection	Exonuclease	Nucleus	Spinal muscular atrophy
AMJEVITA	2016	Protein	Injection	Tnf-alpha	Cell membrane	Rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, adult crohn's disease, ulcerative colitis, plaque psoriasis
ERELZI	2016	Protein	Injection	Tnf	Secreted	Asthma
CINQAIR	2016	Protein	Injection	Il-5	Secreted	Rheumatoid arthritis, polyarticular juvenile idiopathic arthritis, ankylosing spondylitis, plaque psoriasis
ODACTRA	2017	Protein	Solution (percutaneous; subcutaneous); liquid (intradermal; percutaneous; subcutaneous); concentrate (intradermal; percutaneous; subcutaneous); injection (cutaneous; intradermal; subcutaneous)	Immune system	Immune cells	Treatment for dust mite allergies.
SILIQ	2017	Protein	Injection (subcutaneous)	Interleukin (il)-17 receptor a	Secreted	Asthma, psoriasis, crohn's disease, psoriatic arthritis, and rheumatoid arthritis.

computational prediction of B cell epitopes is still a challenge. The non-contiguous nature of B cell epitopes make it difficult to predict using the current computational methods [190]. The most accurate B-cell epitope predicting method, SEPIa [191] has an area under the Receiver Operating Characteristic (AUROC) curve of 0.65, where the value of 0.5 is that of a random classifier. This is an area of active research and the prediction of non-contiguous B-cell epitopes is still a challenging problem.

5.6. Toxicity

The toxicity of chemical drugs mainly arises from off-target effects and undesirable products formed as a result of their metabolism. Biotherapeutics have the advantage of avoiding these off-target effects due to the specificity of these molecules towards their targets. The toxicity of therapeutic proteins arises from two factors; immunogenicity, which has been discussed previously, and aggregation. Various diseases such as Alzheimer's and Parkinson's are suspected to be caused by aggregation of proteins. These protein aggregation events are not well understood and are being actively investigated [192]. Even though the precise mechanism has not been completely elucidated, some sequence motifs that promote aggregation have been identified [193]. Based on this information, tools such as PASTA 2.0 [194], TANGO [195], Zyggregator [196] have been developed to detect the aggregation-prone segments in proteins based on their amino acid sequence. A direct consequence of aggregation on administration of the biological therapeutic is hyper-activation of the target molecule. This is quite evident in monoclonal antibodies where the antibodies aggregate on the platelets leading to thrombocytopenia [197]. Some peptide therapeutics have toxicity associated with them whose origins are not clear [198]. The designed peptide must be non-toxic and this can be tested *in silico* by tools such as ToxinPred, which classify peptides as toxic or non-toxic based on an SVM algorithm [199].

6. Conclusions and challenges

Over the past few years, there has been an upsurge in the number of clinically approved biotherapeutics (see Table 2 for a comprehensive list). Computational methods for the rational design of proteins and peptides are vital in growing the repertoire of biotherapeutics.

Generally, the computational methods involved in designing biotherapeutic peptide and proteins are the same as those for designing any other proteins/peptides. Protein design methods could be classified into two broad categories: (a) prediction of the structure of the backbone and (b) prediction of the amino acid sequence.

Most of the methods usually adopt a predetermined backbone conformation or use scaffold libraries to optimize geometric complementarity with the target. This precludes the possibility of adopting a novel fold. Searching for novel folds would increase the range of possible designs and hence increase the versatility of the design methods. Although methods that sample the backbone conformation have been developed and have successfully designed proteins with novel folds [200], it still remains a computational challenge to exhaustively sample all possible backbone folds. Another problem is the accuracy with which the energy functions score the backbone conformations. Not only do we need better backbone sampling methods but also better scoring methods to come up with energetically favorable novel folds.

After the backbone conformation is predicted, sequence search methods are used to predict the amino acid sequence of the protein that would adopt a desired fold. Sequence search methods can be divided into two broad classes, deterministic search methods and stochastic search methods. Deterministic methods are computationally extensive but provide accurate results for small proteins. The computation time for methods such as DEE increases exponentially with increase in size of protein sequence, for such cases stochastic methods are used that trade off accuracy for speed. Stochastic methods do not always converge to the same solution. Hybrid methods need to be developed using deterministic and stochastic algorithms together to reduce the combinatorial complexity while maintaining accuracy.

Although this review deals with the design of proteins with natural amino acids, some of the designs also incorporate non-natural amino acids. Non-natural amino acids can be incorporated into proteins/peptides either cotranslationally [201–203] by an extended codon-anticodon pair system or by site/residue specific chemical modifications [204]. Incorporation of non-natural amino acids to generate mimetics of therapeutic peptides has been shown to lower their susceptibility to proteolysis and improve bioavailability [205].

The largest fraction of approved biological therapeutics in recent years has been antibodies. Antibodies exhibit favorable design properties as their binding preferences can be modulated by making small changes in their complementarity determining regions (CDRs). The CDRs are constituted by six hyper-variable loops, of which five have canonical structures and can be modeled using templates. The conformation of CDR-H3 loops, which do not have canonical structures, are subject to techniques of loop modeling that either involve using various templates and/or *ab initio* methods. The CDR-H3 loop lies at the center of the antigen-binding site and is therefore crucial for determining the binding affinity of the antibody. Reliable prediction of the CDR-H3 loop remains a hurdle in the rational design of antibodies. Affinity maturation of a computationally designed antibody can be performed if the contact residues of the epitope and paratope are known along with their mode of interaction. Paratope residues can be mutated to residues that improve the complementation between the paratope and epitope interfaces. Methods for determination of the epitope and paratope need to be improved for reliable prediction of mutations for improving the affinity of the antibody. Antibody specific docking methods need to be developed that could sample V_H-V_L orientations to predict the best binding mode of the antigen with the designed antibody. In one such recent study Hattori and coworkers designed a pH sensitive antibody against IL-6 receptor. The antibody binds to its antigen in the slightly alkaline blood plasma (pH 7.4) whereas it rapidly dissociates from its antigen in the acidic environment of the endosome (pH 6.0) and gets recycled [206]. The design of such antibodies could potentially decrease the dose size of the administered therapeutic.

Antibodies are used as curative agents that do not generate an immunological memory. Vaccines raise an immune response that generates antibodies and confer immunological memory. Vaccine design involves a search for structurally similar epitope scaffolds in the protein databank. This is followed by transplanting the epitope residues onto the scaffold and introducing mutations to stabilize these transplanted residues. Additional mutations are made to increase solubility, affinity to antibody, removal of cross reactive epitopes *etc*. With improved computational techniques, rational design of vaccines is becoming more promising, but still the technique cannot be developed into an automated pipeline. Stabilization of the grafted epitope is still a challenging step. Moreover, the statistical potentials used for identification of stabilizing/destabilizing mutations need improvement. Mutations predicted as stabilizing can actually be destabilizing and vice versa. Newly engineered proteins with the epitope might have the epitope in a different conformation that may not elicit the required antibody response.

Designed biological therapeutics need to be tested for their *in vivo* efficacies namely binding affinities to their targets, cell penetration abilities, toxicity, half-life, solubility and immunogenicity. Binding affinity prediction is important for designing biotherapeutics, as they involve interactions with other biomolecules. Various tools exist for its prediction but their accuracies are low as seen by their performance on CAPRI tests. Cell penetration ability, ensures that the developed biotherapeutic enters the cell/nucleus via the membranes. Although methods exist for predicting the cell penetration ability of biotherapeutics they are limited by their accuracies. The study of cell penetrating peptides could reveal important features that could be engineered into biotherapeutics to improve their cells penetration abilities. Improved half-life of these therapeutics is necessary to reduce their dosage. In addition, solubility of the designed molecule is important for its bioavailability. Various tools have been developed for the prediction of half-life and solubility using machine learning algorithms that could predict beneficial mutations to improve both half-life and solubility. Designed biotherapeutic also needs to be non-toxic *i.e.* should not have adverse immunogenic response or form aggregates. Aggregation prone segments can be predicted and removed using various tools described earlier. Various tools have been developed for predicting T-cell epitopes but prediction of B-cell epitopes yet remains a major challenge.

Many challenges remain and new ones may present themselves, such as finding the best mode of delivery for designed biotherapeutics, prediction and minimization of off-target effects, prediction and optimization of absorption, distribution, metabolism and excretion (ADME) of the designed biotherapeutics. Efforts are being made to tackle these problems either by modifying the therapeutic proteins by covalent attachment of various compounds or by using different formulations for delivery. Covalent attachment of polyethylene-glycol (PEG), sialic acid, glycolic acid, *etc.* prolong circulation and decrease glomerular filtration rate of therapeutic proteins. Different formulations with colloidal systems such as liposomes or nano/microparticulate materials such as PLGA (a polymer of lactic acid and glycolic acid) microspheres are used for efficient delivery of therapeutic proteins [207]. With improvements in computational capabilities and algorithms, *in silico* design of biological therapeutics is a promising step ahead.

Acknowledgements

MSM would like to acknowledge the Wellcome Trust DBT India Alliance for a senior fellowship. SN and NSen would like to acknowledge CSIR-SPM fellowship.

References

[1] O. Liebreich, Das Chloralhydrat; ein neues Hypnoticum und Anaestheticum und dessen Anwendung in der Medicin: Eine Arzneymittel-Untersuchung (1869) 60.
[2] T.-Y. Wu, M.-H. Jen, A. Bottle, M. Molokhia, P. Aylin, D. Bell, A. Majeed, Ten-year trends in hospital admissions for adverse drug reactions in England 1999–2009, Jrsm. 103 (2010) 239–250, https://doi.org/10.1258/jrsm.2010.100113.
[3] G. Shepherd, P. Mohorn, K. Yacoub, D.W. May, Adverse drug reaction deaths reported in United States vital statistics, 1999–2006, Ann. Pharmacother. 46 (2012) 169–175, https://doi.org/10.1345/aph.1P592.
[4] International drug monitoring: the role of national centres. Report of a WHO meeting, World Heal. Organ. – Tech. Rep. Ser. 498 (1972) 1–25.
[5] J. Greer, J.W. Erickson, J.J. Baldwin, M.D. Varney, Application of the three-dimensional structures of protein target molecules in structure-based drug design, J. Med. Chem. 37 (1994) 1035–1054, https://doi.org/10.1021/jm00034a001.
[6] T. Uhlig, T. Kyprianou, F.G. Martinelli, C.A. Oppici, D. Heiligers, D. Hills, X.R. Calvo, P. Verhaert, The emergence of peptides in the pharmaceutical business: From exploration to exploitation, EuPA Open Proteomics 4 (2014) 58–69, https://doi.org/10.1016/j.euprot.2014.05.003.
[7] K. Imai, A. Takaoka, Comparing antibody and small-molecule therapies for cancer, Nat. Rev. Cancer. 6 (2006) 714–727, https://doi.org/10.1038/nrc1913.
[8] D.J. Craik, D.P. Fairlie, S. Liras, D. Price, The future of peptide-based drugs, Chem. Biol. Drug Des. 81 (2013) 136–147, https://doi.org/10.1111/cbdd.12055.
[9] A. Mócsai, L. Kovács, P. Gergely, What is the future of targeted therapy in rheumatology: biologics or small molecules?, BMC Med 12 (2014) 43, https://doi.org/10.1186/1741-7015-12-43.
[10] B. Leader, Q.J. Baca, D.E. Golan, Protein therapeutics: a summary and pharmacological classification, Nat. Rev. Drug Discov. 7 (2008) 21–39, https://doi.org/10.1038/nrd2399.
[11] K.K. Kumarasamy, M.A. Toleman, T.R. Walsh, J. Bagaria, F. Butt, R. Balakrishnan, U. Chaudhary, M. Doumith, C.G. Giske, S. Irfan, P. Krishnan, A. V. Kumar, S. Maharjan, S. Mushtaq, T. Noorie, D.L. Paterson, A. Pearson, C. Perry, R. Pike, B. Rao, U. Ray, J.B. Sarma, M. Sharma, E. Sheridan, M.A. Thirunarayan, J. Turton, S. Upadhyay, M. Warner, W. Welfare, D.M. Livermore, N. Woodford, Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study, Lancet Infect. Dis. 10 (2010) 597–602, https://doi.org/10.1016/S1473-3099(10)70143-2.
[12] G. Sakoulas, R.C. Moellering Jr., Increasing antibiotic resistance among methicillin-resistant *Staphylococcus aureus* strains, Clin. Infect. Dis. 46 (2008) S360–S367, https://doi.org/10.1086/533592.
[13] R.E.W. Hancock, H.-G. Sahl, Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies, Nat. Biotechnol. 24 (2006) 1551–1557, https://doi.org/10.1038/nbt1267.
[14] A.D.G. Lawson, Antibody-enabled small-molecule drug discovery, Nat. Rev. Drug Discov. 11 (2012) 519–525, https://doi.org/10.1038/nrd3756.
[15] K.E. Lundin, O. Gissberg, C.I.E. Smith, Oligonucleotide therapies: the past and the present, Hum. Gene Ther. 26 (2015) 475–485, https://doi.org/10.1089/hum.2015.070.
[16] J. Knowles, G. Gromo, A guide to drug discovery: target selection in drug discovery, Nat. Rev. Drug Discov. 2 (2003) 63–69, https://doi.org/10.1038/nrd986.
[17] C. Pabo, Molecular technology. Designing proteins and peptides, Nature. 301 (1983) 200, https://doi.org/10.1038/301200a0.
[18] K.E. Drexler, Molecular engineering: an approach to the development of general capabilities for molecular manipulation, Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 5275–5278, https://doi.org/10.1073/pnas.78.9.5275.
[19] B. Kuhlman, D. Baker, Native protein sequences are close to optimal for their structures, Proc. Natl. Acad. Sci. 97 (2000) 10383–10388, https://doi.org/10.1073/pnas.97.19.10383.
[20] Y. Liu, B. Kuhlman, RosettaDesign server for protein design, Nucleic Acids Res. 34 (2006) W235–W238, https://doi.org/10.1093/nar/gkl163.
[21] Y. Lu, S.M. Berry, T.D. Pfister, Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds, Chem. Rev. 101 (2001) 3047–3080, https://doi.org/10.1021/cr0000574.
[22] A.C. Stiel, K. Feldmeier, B. Höcker, Identification of protein scaffolds for enzyme design using scaffold selection, Methods Mol. Biol. (2014) 183–196, https://doi.org/10.1007/978-1-4939-1486-9_9.
[23] D.P. Hearst, F.E. Cohen, Grafter: a computational aid for the design of novel proteins, Protein Eng. Des. Sel. 7 (1994) 1411–1421, https://doi.org/10.1093/protein/7.12.1411.
[24] K. Hornischer, H. Blöcker, Grafting of discontinuous sites: a protein modeling strategy, Protein Eng. 9 (1996) 931–939.
[25] X. Huang, J. Xue, M. Lin, Y. Zhu, Use of an improved matching algorithm to select scaffolds for enzyme design based on a complex active site model, PLoS One 11 (2016) e0156559, https://doi.org/10.1371/journal.pone.0156559.
[26] M.N. Nguyen, K.P. Tan, M.S. Madhusudhan, CLICK – topology-independent comparison of biomolecular 3D structures, Nucleic Acids Res. 39 (2011) W24–W28, https://doi.org/10.1093/nar/gkr393.
[27] M.N. Nguyen, M.S. Madhusudhan, Biological insights from topology independent comparison of protein 3D structures, Nucleic Acids Res. 39 (2011), https://doi.org/10.1093/nar/gkr348. e94–e94.
[28] L. He, J. Zhu, Computational tools for epitope vaccine design and evaluation, Curr. Opin. Virol. 11 (2015) 103–112, https://doi.org/10.1016/j.coviro.2015.03.013.
[29] C. Zhang, L. Lai, Automatch: target-binding protein design and enzyme design by automatic pinpointing potential active sites in available protein scaffolds, Proteins Struct. Funct. Bioinf. 80 (2012) 1078–1094, https://doi.org/10.1002/prot.24009.
[30] B.I. Dahiyat, S.L. Mayo, Protein design automation, Protein Sci. 5 (1996) 895–903, https://doi.org/10.1002/pro.5560050511.
[31] A. Zanghellini, L. Jiang, A.M. Wollacott, G. Cheng, J. Meiler, E.A. Althoff, D. Röthlisberger, D. Baker, New algorithms and an in silico benchmark for computational enzyme design, Protein Sci. 15 (2006) 2785–2794, https://doi.org/10.1110/ps.062353106.
[32] H. Fazelinia, P.C. Cirino, C.D. Maranas, OptGraft: a computational procedure for transferring a binding site onto an existing protein scaffold, Protein Sci. 18 (2009) 180–195, https://doi.org/10.1002/pro.2.
[33] N.D. Clarke, S.M. Yuan, Metal search: a computer program that helps design tetrahedral metal-binding sites, Proteins Struct. Funct. Bioinf. 23 (1995) 256–263, https://doi.org/10.1002/prot.340230214.
[34] H. Edelsbrunner, E.P. Mücke, Three-dimensional alpha shapes, ACM Trans. Graph. 13 (1994) 43–72, https://doi.org/10.1145/174462.156635.
[35] D.S. Kim, J. Seo, D. Kim, J. Ryu, C.H. Cho, Three-dimensional beta shapes, CAD Comput. Aided Des. 38 (2006) 1179–1191, https://doi.org/10.1016/j.cad.2006.07.002.

[36] J. Ryu, M. Lee, J. Cha, R.A. Laskowski, S.E. Ryu, D.-S. Kim, BetaSCPWeb: side-chain prediction for protein structures using Voronoi diagrams and geometry prioritization, Nucleic Acids Res. 44 (2016) W416–W423, https://doi.org/10.1093/nar/gkw368.
[37] L.G. Boulu, G.M. Crippen, H.A. Barton, H. Kwon, M.A. Marletta, Voronoi binding site model of a polycyclic aromatic hydrocarbon binding protein, J. Med. Chem. 33 (1990) 771–775.
[38] S. Srivastava, G.M. Crippen, Analysis of cocaine receptor site ligand binding by three-dimensional Voronoi site modeling approach, J. Med. Chem. 36 (1993) 3572–3579.
[39] M.P. Bradley, G.M. Crippen, Voronoi modeling: the binding of triazines and pyrimidines to L. casei dihydrofolate reductase, J Med. Chem. 36 (1993) 3171–3177.
[40] G.M. Crippen, Voronoi binding site models, J. Comput. Chem. 8 (1987) 943–955, https://doi.org/10.1002/jcc.540080703.
[41] D.N. Bolon, S.L. Mayo, Enzyme-like proteins by computational design, Proc. Natl. Acad. Sci. 98 (2001) 14274–14279, https://doi.org/10.1073/pnas.251555398.
[42] D. Röthlisberger, O. Khersonsky, A.M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J.L. Gallaher, E.A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K.N. Houk, D.S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design, Nature 453 (2008) 190–195, https://doi.org/10.1038/nature06879.
[43] L. Jiang, E.A. Althoff, F.R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J.L. Betker, F. Tanaka, C.F. Barbas, D. Hilvert, K.N. Houk, B.L. Stoddard, D. Baker, De Novo computational design of retro-aldol enzymes, Science (80-.) 319 (2008) 1387–1391. 10.1126/science.1152692.
[44] J.B. Siegel, A. Zanghellini, H.M. Lovick, G. Kiss, A.R. Lambert, J.L. St.Clair, J.L. Gallaher, D. Hilvert, M.H. Gelb, B.L. Stoddard, K.N. Houk, F.E. Michael, D. Baker, Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction, Science (80-.) 329 (2010) 309–313, 10.1126/science.1190239.
[45] S. Liu, S. Liu, X. Zhu, H. Liang, A. Cao, Z. Chang, L. Lai, Nonnatural protein-protein interaction-pair design by key residues grafting, Proc. Natl. Acad. Sci. 104 (2007) 5330–5335, https://doi.org/10.1073/pnas.0606198104.
[46] S.J. Fleishman, T.A. Whitehead, D.C. Ekiert, C. Dreyfus, J.E. Corn, E.-M. Strauch, I.A. Wilson, D. Baker, Computational design of proteins targeting the conserved stem region of influenza hemagglutinin, Science (80-.) 332 (2011) 816–821, 10.1126/science.1202617.
[47] J. Desmet, M. De Maeyer, B. Hazes, I. Lasters, The dead-end elimination theorem and its use in protein side-chain positioning, Nature 356 (1992) 539–542, https://doi.org/10.1038/356539a0.
[48] I. Samish, C.M. MacDermaid, J.M. Perez-Aguilar, J.G. Saven, Theoretical and computational protein design, Annu. Rev. Phys. Chem. 62 (2011) 129–149, https://doi.org/10.1146/annurev-physchem-032210-103509.
[49] B.I. Dahiyat, S.L. Mayo, De Novo protein design: fully automated sequence selection, Science (80-.) 278 (1997) 82–87, 10.1126/science.278.5335.82.
[50] S.M. Malakauskas, S.L. Mayo, Design, structure and stability of a hyperthermophilic protein variant, Nat. Struct. Biol. 5 (1998) 470–475, https://doi.org/10.1038/nsb0698-470.
[51] C.A. Voigt, D.B. Gordon, S.L. Mayo, Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design, J. Mol. Biol. 299 (2000) 789–803, https://doi.org/10.1006/jmbi.2000.3758.
[52] D.B. Gordon, G.K. Hom, S.L. Mayo, N.A. Pierce, Exact rotamer optimization for protein design, J. Comput. Chem. 24 (2003) 232–243, https://doi.org/10.1002/jcc.10121.
[53] E. Kloppmann, G.M. Ullmann, T. Becker, An extended dead-end elimination algorithm to determine gap-free lists of low energy states, J. Comput. Chem. 28 (2007) 2325–2335, https://doi.org/10.1002/jcc.20749.
[54] C. Yanover, M. Fromer, J.M. Shifman, Dead-end elimination for multistate protein design, J. Comput. Chem. 28 (2007) 2122–2129, https://doi.org/10.1002/jcc.20661.
[55] I. Georgiev, B.R. Donald, Dead-end elimination with backbone flexibility, in: Bioinformatics, 2007, 10.1093/bioinformatics/btm197.
[56] P. Koehl, M. Delarue, Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy, J. Mol. Biol. 239 (1994) 249–275, https://doi.org/10.1006/jmbi.1994.1366.
[57] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092, https://doi.org/10.1063/1.1699114.
[58] D. Vanderbilt, S.G. Louie, A Monte Carlo simulated annealing approach to optimization over continuous variables, J. Comput. Phys. 56 (1984) 259–271, https://doi.org/10.1016/0021-9991(84)90095-0.
[59] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, New approach to Monte Carlo calculation of the free energy: method of expanded ensembles, J. Chem. Phys. 96 (1992) 1776–1783, https://doi.org/10.1063/1.462133.
[60] E. Marinari, G. Parisi, Simulated tempering: a new Monte Carlo scheme (1992), 10.1209/0295-5075/19/6/002.
[61] A.P. Cootes, P.M.G. Curmi, A.E. Torda, Biased Monte Carlo optimization of protein sequences, J. Chem. Phys. 113 (2000) 2489–2496, https://doi.org/10.1063/1.482067.
[62] X. Yang, J.G. Saven, Computational methods for protein design and protein sequence variability: biased Monte Carlo and replica exchange, Chem. Phys. Lett. 401 (2005) 205–210, https://doi.org/10.1016/j.cplett.2004.10.153.
[63] R.H. Swendsen, J.S. Wang, Replica Monte Carlo simulation of spin-glasses, Phys. Rev. Lett. 57 (1986) 2607–2609, https://doi.org/10.1103/PhysRevLett.57.2607.
[64] A. Irbäck, S. Mohanty, PROFASI: a Monte Carlo simulation package for protein folding and aggregation, J. Comput. Chem. 27 (2006) 1548–1555, https://doi.org/10.1002/jcc.20452.
[65] W. Boomsma, J. Frellsen, T. Harder, S. Bottaro, K.E. Johansson, P. Tian, K. Stovgaard, C. Andreetta, S. Olsson, J.B. Valentin, L.D. Antonov, A.S. Christensen, M. Borg, J.H. Jensen, K. Lindorff-Larsen, J. Ferkinghoff-Borg, T. Hamelryck, PHAISTOS: a framework for Markov chain Monte Carlo simulation and inference of protein structure, J. Comput. Chem. 34 (2013) 1697–1705, https://doi.org/10.1002/jcc.23292.
[66] J. Hu, A. Ma, A.R. Dinner, Monte Carlo simulations of biomolecules: the MC module in CHARMM, J. Comput. Chem. 27 (2006) 203–216, https://doi.org/10.1002/jcc.20327.
[67] W.L. Jorgensen, J. Tirado-Rives, Molecular modeling of organic and biomolecular systems using BOSS and MCPRO, J. Comput. Chem. 26 (2005) 1689–1700, https://doi.org/10.1002/jcc.20297.
[68] P. Mitra, D. Shultis, Y. Zhang, EvoDesign: De novo protein design based on structural and evolutionary profiles, Nucleic Acids Res. 41 (2013) W273–W280, https://doi.org/10.1093/nar/gkt384.
[69] M. Brylinski, eVolver: an optimization engine for evolving protein sequences to stabilize the respective structures, BMC Res. Notes 6 (2013) 303, https://doi.org/10.1186/1756-0500-6-303.
[70] F. Lauck, C.A. Smith, G.F. Friedland, E.L. Humphris, T. Kortemme, RosettaBackrub-a web server for flexible backbone protein structure modeling and design, Nucleic Acids Res. 38 (2010) W569–W575, https://doi.org/10.1093/nar/gkq369.
[71] G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins, J. Mol. Biol. 332 (2003) 449–460, https://doi.org/10.1016/S0022-2836(03)00888-X.
[72] R. Ficek, Genetic Algorithms, Pearson Education India (1997), https://doi.org/10.1007/978-1-60761-842-3_19.
[73] T. Blickle, L. Thiele, A comparison of selection schemes used in evolutionary algorithms, Evol. Comput. 4 (1996) 361–394, https://doi.org/10.1162/evco.1996.4.4.361.
[74] M. Movahedi, F. Zare-Mirakabad, S.S. Arab, Evaluating the accuracy of protein design using native secondary sub-structures, BMC Bioinf. 17 (2016) 353, https://doi.org/10.1186/s12859-016-1199-y.
[75] N. Pokala, T.M. Handel, Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity, J. Mol. Biol. 347 (2005) 203–227, https://doi.org/10.1016/j.jmb.2004.12.019.
[76] J.M. Reichert, C.J. Rosensweig, L.B. Faden, M.C. Dewitz, Monoclonal antibody successes in the clinic, Nat. Biotechnol. 23 (2005) 1073–1078, https://doi.org/10.1038/nbt0905-1073.
[77] G. Walsh, Biopharmaceutical benchmarks, Nat. Biotechnol. 32 (2014) (2014) 992–1000, https://doi.org/10.1038/nbt.3040.
[78] H.H. Cai, Therapeutic monoclonal antibodies approved by FDA in 2016, MOJ Immunol. 5 (2017), https://doi.org/10.15406/moji.2017.05.00145.
[79] D.M. Ecker, S.D. Jones, H.L. Levine, The therapeutic monoclonal antibody market, MAbs. 7 (2015) 9–14, https://doi.org/10.4161/19420862.2015.989042.
[80] R. Kontermann, S. Dubel, Antibody Engineering, Springer, Berlin Heidelberg, 2010. doi:10.1007/978-3-642-01147-4.
[81] J. Maynard, G. Georgiou, Antibody engineering, Annu. Rev. Biomed. Eng. 2 (2000) 339–376, https://doi.org/10.1146/annurev.bioeng.2.1.339.
[82] K. Krawczyk, J. Dunbar, C.M. Deane, Computational tools for aiding rational antibody design, Methods Mol. Biol. (2017) 399–416, https://doi.org/10.1007/978-1-4939-6637-0_21.
[83] T.T. Wu, An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity, J. Exp. Med. 132 (1970) 211–250, https://doi.org/10.1084/jem.132.2.211.
[84] B. Al-Lazikani, A.M. Lesk, C. Chothia, Standard conformations for the canonical structures of immunoglobulins, J. Mol. Biol. 273 (1997) 927–948, https://doi.org/10.1006/jmbi.1997.1354.
[85] K.R. Abhinandan, A.C.R. Martin, Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains, Mol. Immunol. 45 (2008) 3832–3839, https://doi.org/10.1016/j.molimm.2008.05.022.
[86] M.P. Lefranc, IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF, Cold Spring Harbor Protoc. 6 (2011) 643–651, https://doi.org/10.1101/pdb.ip86.
[87] A. Honegger, A. Plückthun, Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool, J. Mol. Biol. 309 (2001) 657–670, https://doi.org/10.1006/jmbi.2001.4662.
[88] F. Ehrenmann, M.P. Lefranc, IMGT/DomainGapAlign: IMGT standardized analysis of amino acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF), Cold Spring Harbor Protoc. 6 (2011) 737–749, https://doi.org/10.1101/pdb.prot5636.
[89] F. Ehrenmann, Q. Kaas, M.P. Lefranc, IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodiesT cell receptors, MHC, IgSF and MHcSF, Nucleic Acids Res. 38 (2009) D301–D307, https://doi.org/10.1093/nar/gkp946.

[90] J. Adolf-Bryfogle, Q. Xu, B. North, A. Lehmann, R.L. Dunbrack, PyigClassify: a database of antibody CDR structural classifications, Nucleic Acids Res. 43 (2015) D432–D438, https://doi.org/10.1093/nar/gku1106.
[91] J. Dunbar, C.M. Deane, ANARCI: antigen receptor numbering and receptor classification, Bioinformatics 32 (2015) 298–300, https://doi.org/10.1093/bioinformatics/btv552.
[92] J.C. Almagro, A. Teplyakov, J. Luo, R.W. Sweet, S. Kodangattil, F. Hernandez-Guzman, G.L. Gilliland, Second antibody modeling assessment (AMA-II), Proteins Struct. Funct. Bioinf. 82 (2014) 1553–1562, https://doi.org/10.1002/prot.24567.
[93] J. Ye, N. Ma, T.L. Madden, J.M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool, Nucleic Acids Res. 41 (2013), https://doi.org/10.1093/nar/gkt382.
[94] J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, C.M. Deane, SAbDab: the structural antibody database, Nucleic Acids Res. 42 (2014), https://doi.org/10.1093/nar/gkt1043.
[95] B. North, A. Lehmann, R.L. Dunbrack, A new clustering of antibody CDR loop conformations, J. Mol. Biol. 406 (2011) 228–256, https://doi.org/10.1016/j.jmb.2010.10.030.
[96] C. Chothia, A.M. Lesk, Canonical structures for the hypervariable regions of immunoglobulins, J. Mol. Biol. 196 (1987) 901–917, https://doi.org/10.1016/0022-2836(87)90412-8.
[97] C. Chothia, A.M. Lesk, A. Tramontano, M. Levitt, S.J. Smith-Gill, G. Air, S. Sheriff, E.A. Padlan, D. Davies, W.R. Tulip, P.M. Colman, S. Spinelli, P.M. Alzari, R.J. Poljak, Conformations of immunoglobulin hypervariable regions, Nature 342 (1989) 877–883, https://doi.org/10.1038/342877a0.
[98] A. Tramontano, C. Chothia, A.M. Lesk, Structural determinants of the conformations of medium-sized loops in proteins, Proteins Struct. Funct. Bioinf. 6 (1989) 382–394, https://doi.org/10.1002/prot.340060405.
[99] Y. Choi, C.M. Deane, Predicting antibody complementarity determining region structures without classification, Mol. Biosyst. 7 (2011) 3327, https://doi.org/10.1039/c1mb05223c.
[100] Y. Choi, C.M. Deane, FREAD revisited: accurate loop structure prediction using a database search algorithm, Proteins Struct. Funct. Bioinf. 78 (2010) 1431–1440, https://doi.org/10.1002/prot.22658.
[101] A. Fiser, A. Sali, ModLoop: automated modeling of loops in protein structures, Bioinf. 19 (2003) 2500–2501, https://doi.org/10.1093/bioinformatics/btg362.
[102] V. Morea, A. Tramontano, M. Rustici, C. Chothia, A.M. Lesk, Conformations of the third hypervariable region in the VH domain of immunoglobulins, J. Mol. Biol. 275 (1998) 269–294, https://doi.org/10.1006/jmbi.1997.1442.
[103] D. Kuroda, H. Shirai, M. Kobori, H. Nakamura, Structural classification of CDR-H3 revisited: a lesson in antibody modeling, Proteins Struct. Funct. Genet. 73 (2008) 608–620, https://doi.org/10.1002/prot.22087.
[104] P. Marcatili, A. Rosi, A. Tramontano, PIGS: automatic prediction of antibody structures, Bioinformatics 24 (2008) 1953–1954, https://doi.org/10.1093/bioinformatics/btn341.
[105] H. Shirai, K. Ikeda, K. Yamashita, Y. Tsuchiya, J. Sarmiento, S. Liang, T. Morokata, K. Mizuguchi, J. Higo, D.M. Standley, H. Nakamura, High-resolution modeling of antibody structures by a combination of bioinformatics, expert knowledge, and molecular simulations, Proteins Struct. Funct. Bioinf. 82 (2014) 1624–1635, https://doi.org/10.1002/prot.24591.
[106] A. Sivasubramanian, A. Sircar, S. Chaudhury, J.J. Gray, Toward high-resolution homology modeling of antibody F v regions and application to antibody-antigen docking, Proteins Struct. Funct. Bioinf. 74 (2009) 497–514, https://doi.org/10.1002/prot.22309.
[107] A. Sircar, E.T. Kim, J.J. Gray, RosettaAntibody: antibody variable region homology modeling server, Nucleic Acids Res. 37 (2009) W474–W479, https://doi.org/10.1093/nar/gkp387.
[108] L. Riechmann, M. Clark, H. Waldmann, G. Winter, Reshaping human antibodies for therapy, Nature 332 (1988) 323–327, https://doi.org/10.1038/332323a0.
[109] J. Foote, G. Winter, Antibody framework residues affecting the conformation of the hypervariable loops, J. Mol. Biol. 224 (1992) 487–499, https://doi.org/10.1016/0022-2836(92)91010-M.
[110] J. Chatellier, M.H. Van Regenmortel, T. Vernet, D. Altschuh, Functional mapping of conserved residues located at the VL and VH domain interface of a Fab, J. Mol. Biol. 264 (1996) 1–6, https://doi.org/10.1006/jmbi.1996.0618.
[111] M.J. Banfield, D.J. King, A. Mountain, R.L. Brady, V(L):V(H) domain rotations in engineered antibodies: crystal structures of the fab fragments from two murine antitumor antibodies and their engineered human constructs, Proteins Struct. Funct. Genet. 29 (1997) 161–171, https://doi.org/10.1002/(SICI)1097-0134(199710)29:2<161::AID-PROT4>3.0.CO;2-G.
[112] M. Ben Khalifa, M. Weidenhaupt, L. Choulier, J. Chatellier, N. Rauffer-Bruyère, D. Altschuh, T. Vernet, Effects on interaction kinetics of mutations at the VH-VL interface of Fabs depend on the structural context, J. Mol. Recogn. 13 (2000) 127–139, https://doi.org/10.1002/1099-1352(200005/06)13:3<127::AID-JMR495>3.0.CO;2-9.
[113] T. Nakanishi, K. Tsumoto, A. Yokota, H. Kondo, I. Kumagai, Critical contribution of VH-VL interaction to reshaping of an antibody: the case of humanization of anti-lysozyme antibody, HyHEL-10, Protein Sci. 17 (2008) 261–270, https://doi.org/10.1110/ps.073156708.
[114] D. Fera, A.G. Schmidt, B.F. Haynes, F. Gao, H.-X. Liao, T.B. Kepler, S.C. Harrison, Affinity maturation in an HIV broadly neutralizing B-cell lineage through reorientation of variable domains, Proc. Natl. Acad. Sci. 111 (2014) 10275–10280, https://doi.org/10.1073/pnas.1409954111.
[115] A. Narayanan, B.D. Sellers, M.P. Jacobson, Energy-based analysis and prediction of the orientation between light- and heavy-chain antibody variable domains, J. Mol. Biol. 388 (2009) 941–953, https://doi.org/10.1016/j.jmb.2009.03.043.
[116] J. Dunbar, A. Fuchs, J. Shi, C.M. Deane, ABangle: characterising the VH-VL orientation in antibodies, Protein Eng. Des. Sel. 26 (2013) 611–620, https://doi.org/10.1093/protein/gzt020.
[117] K.R. Abhinandan, A.C.R. Martin, Analysis and prediction of VH/VL packing in antibodies, Protein Eng. Des. Sel. 23 (2010) 689–697, https://doi.org/10.1093/protein/gzq043.
[118] A. Bujotzek, J. Dunbar, F. Lipsmeier, W. Schäfer, I. Antes, C.M. Deane, G. Georges, Prediction of VH-VL domain orientation for antibody variable domain modeling, Proteins Struct. Funct. Bioinf. 83 (2015) 681–695, https://doi.org/10.1002/prot.24756.
[119] V. Kunik, B. Peters, Y. Ofran, Structural consensus among antibodies defines the antigen binding site, PLoS Comput. Biol. 8 (2012) e1002388, https://doi.org/10.1371/journal.pcbi.1002388.
[120] V. Kunik, S. Ashkenazi, Y. Ofran, Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure, Nucleic Acids Res. 40 (2012) W521–W524, https://doi.org/10.1093/nar/gks480.
[121] P.P. Olimpieri, A. Chailyan, A. Tramontano, P. Marcatili, Prediction of site-specific interactions in antibody-antigen complexes: the proABC method and server, Bioinformatics 29 (2013) 2285–2291, https://doi.org/10.1093/bioinformatics/btt369.
[122] K. Krawczyk, T. Baker, J. Shi, C.M. Deane, Antibody i-patch prediction of the antibody binding site improves rigid local antibody-antigen docking, Protein Eng. Des. Sel. 26 (2013) 621–629, https://doi.org/10.1093/protein/gzt043.
[123] J. Huang, W. Honda, CED: a conformational epitope database, BMC Immunol. 7 (2006) 7, https://doi.org/10.1186/1471-2172-7-7.
[124] Y. Kim, J. Ponomarenko, Z. Zhu, D. Tamang, P. Wang, J. Greenbaum, C. Lundegaard, A. Sette, O. Lund, P.E. Bourne, M. Nielsen, B. Peters, Immune epitope database analysis resource, Nucleic Acids Res. 40 (2012) W525–W530, https://doi.org/10.1093/nar/gks438.
[125] H. Shirai, C. Prades, R. Vita, P. Marcatili, B. Popovic, J. Xu, J.P. Overington, K. Hirayama, S. Soga, K. Tsunoyama, D. Clark, M.P. Lefranc, K. Ikeda, Antibody informatics for drug discovery, Biochim. Biophys. Acta – Proteins Proteomics 2014 (1844) 2002–2015, https://doi.org/10.1016/j.bbapap.2014.07.006.
[126] V. Kunik, Y. Ofran, The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops, Protein Eng. Des. Sel. 26 (2013) 599–609, https://doi.org/10.1093/protein/gzt027.
[127] I. Sela-Culang, M.R.E.I. Benhnia, M.H. Matho, T. Kaever, M. Maybeno, A. Schlossman, G. Nimrod, S. Li, Y. Xiang, D. Zajonc, S. Crotty, Y. Ofran, B. Peters, Using a combined computational-experimental approach to predict antibody-specific B cell epitopes, Structure 22 (2014) 646–657, https://doi.org/10.1016/j.str.2014.02.003.
[128] K. Krawczyk, X. Liu, T. Baker, J. Shi, C.M. Deane, Improving B-cell epitope prediction and its application to global antibody-antigen docking, Bioinformatics 30 (2014) 2288–2294, https://doi.org/10.1093/bioinformatics/btu190.
[129] A. Sircar, J.J. Gray, SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models, PLoS Comput. Biol. 6 (2010) e1000644, https://doi.org/10.1371/journal.pcbi.1000644.
[130] R. Brenke, D.R. Hall, G.Y. Chuang, S.R. Comeau, T. Bohnuud, D. Beglov, O. Schueler-Furman, S. Vajda, D. Kozakov, Application of asymmetric statistical potentials to antibody-protein docking, Bioinformatics 28 (2012) 2608–2614, https://doi.org/10.1093/bioinformatics/bts493.
[131] R. Chen, L. Li, Z. Weng, ZDOCK: an initial-stage protein-docking algorithm, Proteins Struct. Funct. Genet. 52 (2003) 80–87, https://doi.org/10.1002/prot.10389.
[132] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, PatchDock and SymmDock: servers for rigid and symmetric docking, Nucleic Acids Res. 33 (2005) W363–W367, https://doi.org/10.1093/nar/gki481.
[133] M. Lapelosa, E. Gallicchio, G.F. Arnold, E. Arnold, R.M. Levy, In Silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 Epitopes, J. Mol. Biol. 385 (2009) 675–691, https://doi.org/10.1016/j.jmb.2008.10.089.
[134] G. Ofek, F.J. Guenaga, W.R. Schief, J. Skinner, D. Baker, R. Wyatt, P.D. Kwong, Elicitation of structure-specific antibodies by epitope scaffolds, Proc. Natl. Acad. Sci. 107 (2010) 17880–17887, https://doi.org/10.1073/pnas.1004728107.
[135] M.L. Azoitei, B.E. Correia, Y.-E.A. Ban, C. Carrico, O. Kalyuzhniy, L. Chen, A. Schroeter, P.-S. Huang, J.S. McLellan, P.D. Kwong, D. Baker, R.K. Strong, W.R. Schief, Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold, Science (80-.) 334 (2011) 373–376, 10.1126/science.1209368.
[136] B.E. Correia, Y.E.A. Ban, M.A. Holmes, H. Xu, K. Ellingson, Z. Kraft, C. Carrico, E. Boni, D.N. Sather, C. Zenobia, K.Y. Burke, T. Bradley-Hewitt, J.F. Bruhn-Johannsen, O. Kalyuzhniy, D. Baker, R.K. Strong, L. Stamatatos, W.R. Schief, Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope, Structure 18 (2010) 1116–1126, https://doi.org/10.1016/j.str.2010.06.010.
[137] M.L. Azoitei, Y.E.A. Ban, J.P. Julien, S. Bryson, A. Schroeter, O. Kalyuzhniy, J.R. Porter, Y. Adachi, D. Baker, E.F. Pai, W.R. Schief, Computational design of high-

affinity epitope scaffolds by backbone grafting of a linear epitope, J. Mol. Biol. 415 (2012) 175–192, https://doi.org/10.1016/j.jmb.2011.10.003.
[138] J.S. McLellan, B.E. Correia, M. Chen, Y. Yang, B.S. Graham, W.R. Schief, P.D. Kwong, Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus, J. Mol. Biol. 409 (2011) 853–866, https://doi.org/10.1016/j.jmb.2011.04.044.
[139] X. Wu, Z.-Y. Yang, Y. Li, C.-M. Hogerkorp, W.R. Schief, M.S. Seaman, T. Zhou, S. D. Schmidt, L. Wu, L. Xu, N.S. Longo, K. McKee, S. O'Dell, M.K. Louder, D.L. Wycuff, Y. Feng, M. Nason, N. Doria-Rose, M. Connors, P.D. Kwong, M. Roederer, R.T. Wyatt, G.J. Nabel, J.R. Mascola, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1, Science (80-.) 329 (2010) 856–861, 10.1126/science.1187659.
[140] B.E. Correia, Y.E.A. Ban, D.J. Friend, K. Ellingson, H. Xu, E. Boni, T. Bradley-Hewitt, J.F. Bruhn-Johannsen, L. Stamatatos, R.K. Strong, W.R. Schief, Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design, J. Mol. Biol. 405 (2011) 284–297, https://doi.org/10.1016/j.jmb.2010.09.061.
[141] M.L. Azoitei, Y.A. Ban, O. Kalyuzhny, J. Guenaga, A. Schroeter, J. Porter, R. Wyatt, W.R. Schief, Computational design of protein antigens that interact with the CDR H3 loop of HIV broadly neutralizing antibody 2F5, Proteins Struct. Funct. Bioinf. 82 (2014) 2770–2782, https://doi.org/10.1002/prot.24641.
[142] B.E. Correia, J.T. Bates, R.J. Loomis, G. Baneyx, C. Carrico, J.G. Jardine, P. Rupert, C. Correnti, O. Kalyuzhniy, V. Vittal, M.J. Connell, E. Stevens, A. Schroeter, M. Chen, S. MacPherson, A.M. Serra, Y. Adachi, M.A. Holmes, Y. Li, R.E. Klevit, B.S. Graham, R.T. Wyatt, D. Baker, R.K. Strong, J.E. Crowe, P.R. Johnson, W.R. Schief, Proof of principle for epitope-focused vaccine design, Nature 507 (2014) 201–206, https://doi.org/10.1038/nature12966.
[143] S.Y. Huang, Search strategies and evaluation in protein-protein docking: principles, advances and challenges, Drug Discov. Today 19 (2014) 1081–1096, https://doi.org/10.1016/j.drudis.2014.02.005.
[144] S.-Y. Huang, S.Z. Grinter, X. Zou, Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions, Phys. Chem. Chem. Phys. 12 (2010) 12899, https://doi.org/10.1039/c0cp00151a.
[145] F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li, T. Hou, Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking, Phys. Chem. Chem. Phys. 18 (2016) 22129–22139, https://doi.org/10.1039/C6CP03670H.
[146] D.L. Beveridge, F.M. DiCapua, Free energy via molecular simulation: applications to chemical and biomolecular systems, Annu. Rev. Biophys. Biophys. Chem. 18 (1989) 431–492, https://doi.org/10.1146/annurev.bb.18.060189.002243.
[147] B.O. Brandsdal, A.O. Smalås, Evaluation of protein-protein association energies by free energy perturbation calculations, Protein Eng. 13 (2000) 239–245, https://doi.org/10.1093/protein/13.4.239.
[148] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc. Chem. Res. 33 (2000) 889–897, https://doi.org/10.1021/ar000033j.
[149] N. Homeyer, H. Gohlke, Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method, Mol. Inform. 31 (2012) 114–122, https://doi.org/10.1002/minf.201100135.
[150] R. Chowdhury, M. Rasheed, D. Keidel, M. Moussalem, A. Olson, M. Sanner, C. Bajaj, Protein-Protein docking with F2Dock 2.0 and GB-Rerank, PLoS One 8 (2013) e51307, https://doi.org/10.1371/journal.pone.0051307.
[151] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, J. Chem. Inf. Model. 51 (2011) 69–82, https://doi.org/10.1021/ci100275a.
[152] M.J. Sippl, Knowledge-based potentials for proteins, Curr. Opin. Struct. Biol. 5 (1995) 229–235, https://doi.org/10.1016/0959-440X(95)80081-6.
[153] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring function to predict protein-ligand interactions, J. Mol. Biol. 295 (2000) 337–356, https://doi.org/10.1006/jmbi.1999.3371.
[154] N. Tuncbag, O. Keskin, A. Gursoy, HotPoint: hot spot prediction server for protein interfaces, Nucleic Acids Res. 38 (2010) W402–W406, https://doi.org/10.1093/nar/gkq323.
[155] S.J. Darnell, L. LeGault, J.C. Mitchell, K.F.C. Server, interactive forecasting of protein interaction hot spots, Nucleic Acids Res. 36 (2008) W265–W269, https://doi.org/10.1093/nar/gkn346.
[156] X. Zhu, J.C. Mitchell, KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features, Proteins Struct. Funct. Bioinf. 79 (2011) 2671–2683, https://doi.org/10.1002/prot.23094.
[157] K. Ii, D. Cho, D. Lee Kim, A feature-based approach to modeling protein-protein interaction hot spots, Nucleic Acids Res. 37 (2009) 2672–2687, https://doi.org/10.1093/nar/gkp132.
[158] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web server: an online force field, Nucleic Acids Res. 33 (2005) W382–W388, https://doi.org/10.1093/nar/gki387.
[159] T. Kortemme, D. Baker, A simple physical model for binding energy hot spots in protein-protein complexes, Proc. Natl. Acad. Sci. 99 (2002) 14116–14121, https://doi.org/10.1073/pnas.202485799.
[160] V. Ramadoss, F. Dehez, C. Chipot, AlaScan: a graphical user interface for alanine scanning free-energy calculations, J. Chem. Inf. Model. 56 (2016) 1122–1126, https://doi.org/10.1021/acs.jcim.6b00162.
[161] M.F. Lensink, S.J. Wodak, Docking, scoring, and affinity prediction in CAPRI, Proteins Struct. Funct. Bioinf. 81 (2013) 2082–2095, https://doi.org/10.1002/prot.24428.
[162] D.M. Copolovici, K. Langel, E. Eriste, Ülo Langel, Cell-penetrating peptides: design, synthesis, and applications, ACS Nano. 8 (2014) 1972–1994. doi:10.1021/nn4057269.
[163] C. Bechara, S. Sagan, Cell-penetrating peptides: 20 years later, where do we stand?, FEBS Lett 587 (2013) 1693–1702, https://doi.org/10.1016/j.febslet.2013.04.031.
[164] K.A. Brogden, Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?, Nat Rev. Microbiol. 3 (2005) 238–250, https://doi.org/10.1038/nrmicro1098.
[165] P. Agrawal, S. Bhalla, S.S. Usmani, S. Singh, K. Chaudhary, G.P.S. Raghava, A. Gautam, CPPsite 2.0: repository of experimentally validated cell-penetrating peptides, Nucleic Acids Res. 44 (2016) D1098–D1103, https://doi.org/10.1093/nar/gkv1266.
[166] T.A. Holton, G. Pollastri, D.C. Shields, C. Mooney, CPPpred: Prediction of cell penetrating peptides, Bioinformatics 29 (2013) 3094–3096, https://doi.org/10.1093/bioinformatics/btt518.
[167] E. Eiríksdóttir, K. Konate, Ü. Langel, G. Divita, S. Deshayes, Secondary structure of cell-penetrating peptides controls membrane interaction and insertion, Biochim. Biophys. Acta – Biomembr. 1798 (2010) 1119–1128, https://doi.org/10.1016/j.bbamem.2010.03.005.
[168] A. Dinca, W.M. Chien, M.T. Chin, Intracellular delivery of proteins with cell-penetrating peptides for therapeutic uses in human disease, Int. J. Mol. Sci. 17 (2016) 263, https://doi.org/10.3390/ijms17020263.
[169] A. Sharma, D. Singla, M. Rashid, G.P.S. Raghava, Designing of peptides with desired half-life in intestine-like environment, BMC Bioinformatics 15 (2014) 282, https://doi.org/10.1186/1471-2105-15-282.
[170] X. Song, T. Zhou, H. Jia, X. Guo, X. Zhang, P. Han, J. Sha, Sprotp: a web server to recognize those short-lived proteins based on sequence-derived features in human cells, PLoS One 6 (2011) e27836, https://doi.org/10.1371/journal.pone.0027836.
[171] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: Proteomics Protoc. Handb, 2005, pp. 571–607, 10.1385/1592598900.
[172] D. Mathur, S. Prakash, P. Anand, H. Kaur, P. Agrawal, A. Mehta, R. Kumar, S. Singh, G.P.S. Raghava, PEPlife: a repository of the half-life of peptides, Sci. Rep. 6 (2016) 36617, https://doi.org/10.1038/srep36617.
[173] S. Jevševar, M. Kusterle, M. Kenig, PEGylation of antibody fragments for half-life extension, Methods Mol. Biol. (2012) 233–246, https://doi.org/10.1007/978-1-61779-931-0_15.
[174] J.W. Baynes, F. Wold, Effect of glycosylation on the in vivo circulating half life of ribonuclease, J. Biol. Chem. 251 (1976) 6016–6024.
[175] V.N. Podust, S. Balan, B.C. Sim, M.P. Coyle, U. Ernst, R.T. Peters, V. Schellenberger, Extension of in vivo half-life of biologically active molecules by XTEN protein polymers, J. Control. Release. 240 (2016) 52–66, https://doi.org/10.1016/j.jconrel.2015.10.038.
[176] M.T. Weinstock, J.N. Francis, J.S. Redman, M.S. Kay, Protease-resistant peptide design-empowering nature's fragile warriors against HIV, Biopolymers 98 (2012) 431–442, https://doi.org/10.1002/bip.22073.
[177] R.E. Kontermann, Strategies to extend plasma half-lives of recombinant antibodies, BioDrugs 23 (2009) 93–109, https://doi.org/10.2165/00063030-200923020-00003.
[178] N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, A. Villaverde, Microbial factories for recombinant pharmaceuticals, Microb. Cell Fact. 8 (2009) 17, https://doi.org/10.1186/1475-2859-8-17.
[179] G. Hannig, S.C. Makrides, Strategies for optimizing heterologous protein expression in Escherichia coli, Trends Biotechnol. 16 (1998) 54–60, https://doi.org/10.1016/S0167-7799(97)01155-4.
[180] F. Agostini, D. Cirillo, C.M. Livi, R. Delli Ponti, G.G. Tartaglia, ccSOL omics: a webserver for large-scale prediction of endogenous and heterologous solubility in E. coli, Bioinformatics 30 (2014) btu420, https://doi.org/10.1093/bioinformatics/btu420.
[181] P. Smialowski, G. Doose, P. Torkler, S. Kaufmann, D. Frishman, PROSO II – A new method for protein solubility prediction, FEBS J. 279 (2012) 2192–2200, https://doi.org/10.1111/j.1742-4658.2012.08603.x.
[182] A. Kuriakose, N. Chirmule, P. Nair, Immunogenicity of biotherapeutics: causes and association with posttranslational modifications, J. Immunol. Res. 2016 (2016) 1298473, https://doi.org/10.1155/2016/1298473.
[183] M. Butler, M. Spearman, The choice of mammalian cell host and possibilities for glycosylation engineering, Curr. Opin. Biotechnol. 30 (2014) 107–112, https://doi.org/10.1016/j.copbio.2014.06.010.
[184] M. Subramanyam, Immunogenicity of biotherapeutics—An overview, J. Immunotoxicol. 3 (2006) 151–156, https://doi.org/10.1080/15476910600845740.
[185] L. Backert, O. Kohlbacher, Immunoinformatics and epitope prediction in the age of genomic medicine, Genome Med. 7 (2015) 119, https://doi.org/10.1186/s13073-015-0245-0.
[186] A. Patronov, I. Doytchinova, T-cell epitope vaccine design by immunoinformatics, Open Biol., 3, 2013. 120139 120139doi:10.1098/rsob.120139.

[187] A.S. De Groot, B.M. Jesdale, E. Szu, J.R. Schafer, R.M. Chicz, G. Deocampo, An interactive web site providing major histocompatibility ligand predictions: application to HIV research, AIDS Res. Hum. Retroviruses 13 (1997) 529–531, https://doi.org/10.1089/aid.1997.13.529.
[188] M.E.E. Elfaki, E.A.G. Khalil, A.S. De Groot, A.M. Musa, A. Gutierrez, B.M. Younis, K.A.M. Salih, A.M. El-Hassan, Immunogenicity and immune modulatory effects of in silico predicted L. donovani candidate peptide vaccines, Hum. Vaccines Immunother. 8 (2012) 1769–1774, https://doi.org/10.4161/hv.21881.
[189] C. King, E.N. Garza, R. Mazor, J.L. Linehan, I. Pastan, M. Pepper, D. Baker, Removing T-cell epitopes with computational protein design, Proc. Natl. Acad. Sci. 111 (2014) 8577–8582, https://doi.org/10.1073/pnas.1321126111.
[190] Y. El-Manzalawy, V. Honavar, Recent advances in B-cell epitope prediction methods, Immunome Res. 6 (2010) S2, https://doi.org/10.1186/1745-7580-6-S2-S2.
[191] G.A. Dalkas, M. Rooman, SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence, BMC Bioinformatics. 18 (2017) 95, https://doi.org/10.1186/s12859-017-1528-9.
[192] E.H. Koo, P.T. Lansbury, J.W. Kelly, Amyloid diseases: abnormal protein aggregation in neurodegeneration, Proc. Natl. Acad. Sci. 96 (1999) 9989–9990, https://doi.org/10.1073/pnas.96.18.9989.
[193] X. Wang, T.K. Das, S.K. Singh, S. Kumar, Potential aggregation prone regions in biotherapeutics: a survey of commercial monoclonal antibodies, MAbs. 1 (2009) 254–267, https://doi.org/10.4161/mabs.1.3.8035.
[194] I. Walsh, F. Seno, S.C.E. Tosatto, A. Trovato, PASTA 2.0: an improved server for protein aggregation prediction, Nucleic Acids Res. 42 (2014) W301–W307, https://doi.org/10.1093/nar/gku399.
[195] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, Nat. Biotechnol. 22 (2004) 1302–1306, https://doi.org/10.1038/nbt1012.
[196] G.G. Tartaglia, M. Vendruscolo, The Zyggregator method for predicting protein aggregation propensities, Chem. Soc. Rev. 37 (2008) 1395, https://doi.org/10.1039/b706784b.
[197] J. Li, C. Yang, Y. Xia, A. Bertino, J. Glaspy, M. Roberts, D.J. Kuter, Thrombocytopenia caused by the development of antibodies to thrombopoietin, Blood 98 (2001) 3241–3248, https://doi.org/10.1182/blood.V98.12.3241.
[198] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G.P.S. Raghava, Peptide toxicity prediction, in: Comput. Pept, 2015, pp. 143–157, 10.1007/978-1-4939-2285-7_7.
[199] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G.P.S. Raghava, In silico approach for predicting toxicity of peptides and proteins, PLoS One 8 (2013) e73957, https://doi.org/10.1371/journal.pone.0073957.
[200] B. Kuhlman, Design of a novel globular protein fold with atomic-level accuracy, Science (80-.) 302 (2003) 1364–1368, 10.1126/science.1089427.
[201] T. Hohsaka, M. Sisido, Incorporation of non-natural amino acids into proteins, Curr. Opin. Chem. Biol. 6 (2002) 809–815, https://doi.org/10.1016/S1367-5931(02)00376-9.
[202] W. Liu, A. Brock, S. Chen, S. Chen, P.G. Schultz, Genetic incorporation of unnatural amino acids into proteins in mammalian cells, Nat. Methods. 4 (2007) 239–244, https://doi.org/10.1038/nmeth1016.
[203] H. Xiao, A. Chatterjee, S.-H. Choi, K.M. Bajjuri, S.C. Sinha, P.G. Schultz, Genetic incorporation of multiple unnatural amino acids into proteins in Mammalian cells, Angew. Chem. Int. Ed. Engl. 52 (2013) 14080–14083, https://doi.org/10.1002/anie.201308137.
[204] K. Lang, J.W. Chin, Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins, Chem. Rev. 114 (2014) 4764–4806, https://doi.org/10.1021/cr400355w.
[205] C. Adessi, C. Soto, Converting a peptide into a drug: strategies to improve stability and bioavailability, Curr. Med. Chem. 9 (2002) 963–978, https://doi.org/10.2174/0929867024606731.
[206] T. Igawa, S. Ishii, T. Tachibana, A. Maeda, Y. Higuchi, S. Shimaoka, C. Moriyama, T. Watanabe, R. Takubo, Y. Doi, T. Wakabayashi, A. Hayasaka, S. Kadono, T. Miyazaki, K. Haraya, Y. Sekimori, T. Kojima, Y. Nabuchi, Y. Aso, Y. Kawabe, K. Hattori, Antibody recycling by engineered pH-dependent antigen binding improves the duration of antigen neutralization, Nat. Biotechnol. 28 (2010) 1203–1207, https://doi.org/10.1038/nbt.1691.
[207] D.S. Pisal, M.P. Kosloski, S.V. Balu-Iyer, Delivery of therapeutic proteins, J. Pharm. Sci. 99 (2010) 2557–2575, https://doi.org/10.1002/jps.22054.
[208] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera – A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.
[209] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maclejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z.T. Dame, B. Han, Y. Zhou, D.S. Wishart, DrugBank 4.0: Shedding new light on drug metabolism, Nucleic Acids Res. 42 (2014) D1091–D1097, https://doi.org/10.1093/nar/gkt1068.