# Overview and prospects of physical and computational antibody design

Bas Baartmans

18-12-22

Supervisor: Dr. Maximilian Fürst

Formal supervisor: Dr. Gerrit Poelarends

Computational Enzymology

University of Groningen

# Table of Contents

# Abstract

Monoclonal antibodies are used as therapeutic drugs because of their predisposition for highly specific binding to macromolecular targets. Because of their therapeutic value, multiple physical approaches have been developed in the last decades for the production and isolation of strong binding antibodies. But, with the increasing computational power of the last years, the field of computational protein design has expanded greatly, and therefore questions must be asked whether these physical procedures will soon be replaced. This essay provides an overview of currently used physical and computational methods for the design of antibodies, as well as for the computational design of *de novo* binders. Firstly, it describes the hybridoma and display techniques and explains current methods of DNA library optimization. Secondly, state-of-the-art computational antibody and *de novo* protein design strategies are described. The trends and prospects of both sides will be discussed to answer the question of whether the experimental design of antibodies will become obsolete with the emergence of new computational methods.


# Introduction

Since 1975, monoclonal antibodies (mAbs) have been developed as effective therapeutic drugs (Liu 2014). To put this in perspective, four out of the top ten sold pharmaceuticals were mAbs in 2021 (Urquhart 2022). Additionally, the 100$^{th}$ mAb was recently approved by the FDA (Mullard 2021).

Antibodies are an essential part of various species' immune systems (Vadnais, Criscitiello, and Smider 2017). During an immune response, unique antibodies are produced to neutralize immunogenic particles (antigens). This neutralization is based on the binding between the paratope of the antibody and the epitope of the antigen, which can consequently signal to the immune system to remove the contamination (Vadnais et al. 2017). Most animals' adaptive immune systems contain an enormous repertoire of different antibody isoforms of low to high affinity to these antigens (Sun et al. 2020). By harnessing this endogenous machinery, researchers have been able to generate extremely specific binders to interesting therapeutic targets, as well as biochemical assays (Iha et al. 2019). For instance, mAb trastuzumab has been developed against the extracellular HER-2 receptor that is upregulated in some forms of breast cancer (Lv et al. 2018). Other examples include rheumatic diseases (Norman 2017), and melanoma (Kwok et al. 2016).

Antibodies are proteins that can bind with high specificity to molecular surfaces. An antibody, which is a covalently linked dimer consisting of two heavy chains and two light chains (as depicted in Fig. 1A), has a binding site called the complementary determining region (CDR). The genes encoding for this region are susceptible to recombination and, upon immunization, will undergo somatic hypermutation, resulting in an extremely diverse set of non to high-affinity binders against antigens (Wang et al. 2020). This region is therefore also named the variable fragment ($F_v$). The CDR of each chain is made up of three alpha-helices (as depicted in Fig. 1B). These helices were long considered variable, however, recently canonical parts in CDR1 and CDR2 were found, as well as a small set of canonical structures in CDR3 (Chiu and Gilliland 2016). Opposing the $F_v$ region, the crystallizable (constant) fragment ($F_c$) is needed as a supportive structure and for its' effector functions, such as activating the complement system, a complex cascade of immune reactions ultimately dismantling damaged or non-self objects (Goldberg and Ackerman 2020)(Bordron et al. 2020). The $F_c$-domain is also a contributor to other biophysical and pharmacokinetic properties such as plasma half-life, aggregation, and stability (Chiu and Gilliland 2016) (Alfaleh et al. 2020) (Khan et al. 2022).

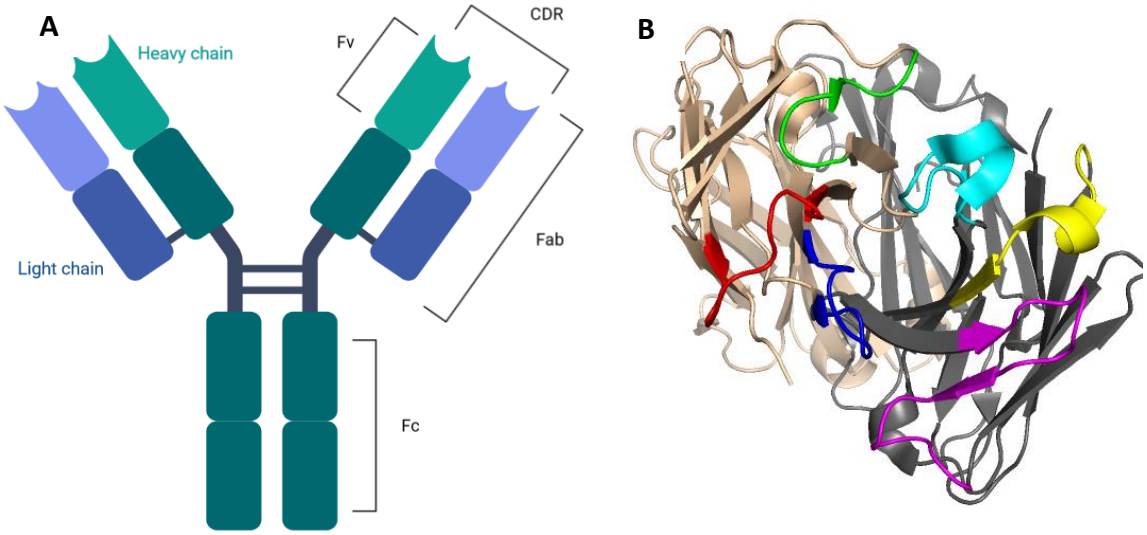


*Figure 1 - A: schematic overview of an IgG1 antibody. The following structures are depicted: heavy chain (green), light chain (blue), $F_c$-region, $F_{ab}$-region, Fv-region, CDR. B: top view of the mAb trastuzumab Fab region depicting the CDR. The following structures are visualized: light chain (light orange), heavy chain (gray), CDR-L1(red), CDR-L2 (green), CDR-L3 (blue), CDR-H1 (yellow), CDR-H2 (magenta), CDR-H3 (cyan). PDB structure is 4HKZ* (Donaldson et al. 2013). *CDRs were colored according to identified CDR sequences as described by Hermanto et al.* (Hermanto et al. 2017).

Historically, mAbs were developed using hybridoma technology (Moraes et al. 2021). This technique is based on the immunization of a mouse with an antigen and the fusing of its effective B lymphocytes with an immortal myeloma cell line ultimately producing specific antibodies. Although still widely used today, there has been pressure to replace this technique due to ethical concerns, time consumption, technical expertise requirements, and low efficiency of the B lymphocyte-hybridoma fusion (Moraes et al. 2021). Now, decades later, high-throughput display methods have been developed to replace the need for hybridomas (Ledsgaard et al. 2018; Newton et al. 2020; Salema and Fernández 2017). In these techniques, the gene encoding an antibody fragment can be fused with a surface protein gene of a microbe or bacteriophage. The translated proteins are then presented on the surface of the respective membrane and can be selected against. Other types are ribosome and mRNA display methods, which are powerful alternatives to phage and cell-display methods and do not rely on phage or microbial systems (Newton et al. 2020). High throughput of this method is possible due to the use of large recombinant or synthetic DNA libraries and a relatively easy affinity selection procedure.

All current FDA-approved mAbs have been developed experimentally, mostly with the use of hybridoma technology (Parray et al. 2020). However, since the increased computing power over the years, *in silico* efforts of protein design have emerged. With the recent developments of software such as RosettaFold and the highly praised AlphaFold2, very accurate protein structures can be predicted, comparable to the structures determined by biophysical approaches (Jumper et al. 2021). This prediction relies partly on multiple sequence alignments, which is problematic for predicting CDR structures as they are hypervariable, particularly CDR-H3 (Abanades et al. 2022). Also considering the advancements in *de novo* protein design (Boyken et al. 2016; Cao et al. 2022; Huang, Boyken, and Baker 2016), the resource-intensive and laborious physical methods might soon come to an end. A milestone in this field is the recent anti-interleukin-2 mAb AU-007, which is the first computationally designed antibody to enter clinical trials in humans (Vasselli et al. 2022).

This essay discusses these state-of-the-art computational methods for *de novo* protein binders and antibodies and will reflect upon the course of both experimental (i.e. physical) and computational methods of design. It will try to answer the question of whether the physical methods will become obsolete with the emergence of computational methods to design protein binders. A subquestion is asked as to whether small *de novo* protein binders will replace future antibodies themselves.

# Experimental approaches to antibody design

The preclinical stage of modern drug discovery initially comprises target selection, lead generation, and lead optimization (Zhou and Zhong 2017). The same is true for antibody development. First, a cellular macromolecular target, the antigen, with potential therapeutic value is selected. Next, a lead generation strategy is employed. These strategies can rely on hybridoma and transgenic mouse models, but high-throughput combinatorial approaches such as display techniques are also used. The lead binders are consequently optimized to increase binding affinity and specificity to the antigen, using affinity maturation-like techniques. Ultimately, biophysical properties such as solubility and thermostability are improved (Chiu and Gilliland 2016). Protein aggregation can especially be detrimental, as it can enhance immunogenicity (Ratanji et al. 2014).

## Hybridoma technology

The hybridoma technique relies on fusing antibody-producing B lymphocytes into myeloma cells (Moraes et al. 2021)(Parray et al. 2020). In principle, this makes the cell immortal, resulting in high antibody production. B lymphocytes were initially harvested from mice, which was problematic since murine antibodies cause an immune response in humans, leading to the invention of murine antibody humanization. In short, this process entails the grafting of mouse CDRs onto a human antibody scaffold, ultimately bypassing the anti-mouse immune response (Kim and Hong 2012). Additionally, transgenic mice have been developed to produce human antibodies. The need to replace this technology is still high, and there have been lots of efforts to do this. The reasons are multifold. For example, the need to immunize and harvest animal cells has great ethical concerns. There is also a low efficiency rate of cell fusion, and the screening process for hybridomas can be time-consuming and challenging (Moraes et al. 2021).

## Display techniques

In search of alternatives, different display methods have been developed. In short, phage and cellular display methods exploit the natural machinery of translocation of surface proteins by fusing antibody fragments into the encoding gene. The protein of interest can be displayed on its respective membrane. Other display methods used include ribosome display and mRNA display. These techniques do not rely on cell-based systems, thus making them an attractive option for protein expression since no gene fusion is necessary. High throughput can be achieved using enormous libraries of antibody DNA, resulting in a diverse collection of protein expression.

### DNA libraries

A DNA library containing antibody sequences must first be generated to perform display methods in high throughput. In principle, there are two classes of libraries that overarch numerous subtypes (Adams and Sidhu 2014). The first class tries to mimic naturally occurring antibodies by recombining light and heavy chain V, D, and J gene segments into a large repertoire of fragments, employing the same paradigm responsible for CDR variety among endogenous antibodies. The gene segments can be isolated from a collection of naive B cells. Cells that were previously exposed to the antigen may improve initial affinity towards the target

of interest. More recent advancements in the development of this type of system include CDR-H3 diversification while maintaining canonical regions in the Hv and Lv. This reduces the diversity of constant structures, therefore increasing the positive folding properties of the protein (Prassler et al. 2011). Minimalistic libraries focus on reducing sequence complexity, but rationally designing CDR structures. For example, it has been shown that naturally occurring CDRs contain amino acid sequences that are inclined to contain tyrosine and serine. Constructing a library with CDRs biased toward these two amino acids can create highly specific antibodies while reducing the sequence possibilities (Fellouse et al. 2005). Furthermore, the CDR-H3 is found to be the most variable CDR (Adams and Sidhu 2014). Knowing this, a set of constraints arises that can be used in designing antibody libraries. More complete explanations of approaches to introducing genetic diversity have been described comprehensively (Beerli and Rader 2010)(Ledsgaard et al. 2018).

## Cell-based display methods

A commonly used bacteriophage for the phage display of antibodies is the filamentous M13 phage. This simple bacteriophage consists of 9 genes, one of which is the pIII gene, transcribing for the major surface protein. Fusing an antibody fragment to the pIII gene of a plasmid containing the M13 phage genome (also called phagemid) and infecting an *E. coli* bacterium will cause the bacteriophage to replicate and present on the surface of the outer virion membrane (Ledsgaard et al. 2018). Often, a helper phage offers utility by providing the genes encoding for proteins used in replication, resulting in libraries that are two orders of magnitude larger than without the use of a helper phage (Rondot et al. 2001).

The need to work with antigens can inherently be a limiting factor, as the antigens must first be purified and immobilized to a surface, such as a microtiter plate or magnetic bead. This can be problematic as the structural modification needed for immobilization can alter antigen conformation and will thus misdirect the antibodies (Alfaleh et al. 2020). After generating the library, the cloned phages are precipitated and concentrated and, in a process called biopanning, the large variety of these clones is affinity captured with the beforementioned antigen (Salema and Fernández 2017). This process is often repeated to capture the highest affinity binders, that are subsequently subjected to affinity maturation processes.

Drawbacks include the need to fuse an antibody fragment gene with the G3P gene of the phage. Naturally, the transformation of the phage into *E. coli* is not completely efficient, as not all E. coli take up the phage DNA, resulting in an incomplete display of the library. With the indirect transformation of a phagemid, higher display rates are observed. Additionally, due to difficulties in the expression and folding of complete immunoglobulins (Xiao et al. 2017), often only heavy chain Fv fragments are expressed. More recently, camelid fragments, called nanobodies, consisting of the variable part of both heavy and light chains were expressed. Nanobodies have distinctive qualities that make them attractive pharmaceuticals, such as their small size and high specificity, which simultaneously amend complete expression and biopanning during the display process (Muyldermans 2021). Moreover, complete Fab regions were also expressed (Alfaleh et al. 2020). Furthermore, a comparative study found that phage-derived antibodies are more likely to exhibit developmental risks compared to antibodies derived from immunized mice. Antibodies extracted from mice show favorable biophysical properties since an immune system will inherently produce antibodies tolerable for the mouse (Alfaleh et al. 2020). For example, phage-derived antibodies are more susceptible to self-interaction and polyreactivity likely as an effect of higher aliphatic content in the CDRs (Kaleli, Karadag, and Kalyoncu 2019).

Yeast or mammalian cell display methods are other techniques where a DNA library is directly introduced into a cell, as contrasted to using E. coli as a host for phage proliferation. An

advantage of these types of cell systems is that they can produce proteins with eukaryotic post-translational modifications. For example, antibodies can be glycosylated, giving them biophysical properties similar to mass-produced antibodies. Other advantages include ease of cell culture modification and compatibility with flow cytometric analysis (Teymennet-Ramírez, Martínez-Morales, and Trejo-Hernández 2022). Of note should be that transfection efficiency is lower in these types of systems than phage-display, resulting in smaller and less diverse libraries.

## Cell-free display methods

mRNA display is based on the fusion of the translated protein to the encoding mRNA, with a puromycin linker (Newton et al. 2020). This *in vitro* technique does not rely on the context of a cell system and is therefore not limited by the efficiency of plasmid fusion and incorporation into the cell or phage, resulting in libraries of $10^{14}$ variants. Other advantages include the ability to incorporate unnatural amino acids into the sequence, as well as the robustness of the system, allowing precise control over experimental conditions such as pH and ionic strength (Newton et al. 2020). Even though larger libraries can be constructed, it should be noted that due to the loss of the cell context, it is currently not possible to incorporate PTMs into the proteins. Additionally, mRNA display is usually limited to monomeric proteins, since a link to the mRNA is required. However, there have been successful attempts to display $F_{ab}$ regions with mRNA display (Sumida, Yanagawa, and Doi 2012).

Ribosome display is similar to mRNA display in the sense that it couples the phenotype (protein) directly to a genotype (mRNA)(Li et al. 2019).

In this case, a DNA library is set up for which each sequence often contains a T7 promotor, stem-loops at 5' and 3' ends, the gene of interest, and a spacer. The stop codon is not incorporated into the segment, thereby stalling the ribosome, resulting in an mRNA-ribosome-protein complex. Subsequent selections of protein fragments are performed using the same techniques as other display methods. Ribosome display has been extensively reviewed by Li and colleagues (Li et al. 2019).

The fundamental bottleneck in display technology is the transfection and transformation efficiency (Hoogenboom et al. 1998). To illustrate this, consider that prokaryotic display methods have a typical library size of $10^8$ to $10^{10}$ variants. Phage display libraries are typically made of $10^9$ variants but can be up to $10^{12}$ variants (Newton et al. 2020). The severity of the bottleneck can be demonstrated based on the work of Keefe and Szostak, who screened for an ATP-binder using mRNA display (Keefe and Szostak 2001). With the use of a random synthetic library, they calculated that only 1 in $10^{11}$ variants showed high affinity to ATP, indicating that big and highly diverse libraries are a must that cannot be achieved by cell-surface and phage display. In principle, a second affinity maturation step can be performed to gain better binders from a set of lower-affinity antibody fragments, however, the DNA space might not be completely utilized which could lead to a reduced amount of high-affinity binders.

## Next-generation sequencing in display methods

Normally, screening for high-affinity binders using display techniques involves subsequent biopanning cycles and sequence determination of clones. However, this step is often overshadowed by the high expression or display rate of a small number of proteins, instead of providing sequences for lower abundant proteins (Nannini et al. 2021; Rouet et al. 2018). Traditionally, Sanger sequencing was used to sequence the most abundant clones, and although highly accurate, considering average display libraries contain $10^8$-$10^{10}$ variants, this method has become insufficient (He et al. 2018). There is a need for high-throughput, full-length sequencing methods that can be used during the biopanning process of highly diverse libraries. A current solution of this problem is PacBio® HiFi sequencing, which allows high-

throughput sequencing with read lengths of up to 25000 base pairs and 99.5% accuracy (Hon et al. 2020).

# Computational design of proteins

Affinity reagents, such as antibodies, are needed to bind highly specific to therapeutic targets. Such binders are, however, not limited be structured like antibodies. Interactions between proteins and macromolecules are based on the binding of intricate positioned polar and apolar residues between the two entities. With the great advancements in computation power of the last decades, *in silico* techniques have been developed that simulate these interfaces, making the computational design of protein binders amendable. In practice, two problems must be addressed in the computational design of proteins. The first one is the problem of accurately predicting protein folding, and the second one is the prediction of binding to a molecular surface.

## Protein folding

Protein folding is primarily driven by hydrophobic forces, hydrogen bonding, van der Waals interactions, and Coulombic interactions among others (Huang et al. 2016)(Newberry and Raines 2019), which drive the hydrophobic core away from the solvent, resulting in a thermodynamically stable structure. For example, the polar carbonyl and amide groups in a protein backbone can form hydrogen bonds, resulting in secondary structures such as alpha-helices and beta-sheets (Huang et al. 2016). Accurate protein fold prediction is, in principle, a search through chemical space for the beforementioned energetically stable conformations. However, two problems arise from this, namely the accurate calculation of the free energy and the efficient sampling of the search space. The search space can be envisioned as a multidimensional array containing all theoretical protein conformations and its size depends among other things on polypeptide sequence length, backbone degrees of freedom, and side chain rotation. In the early days of the field of structure prediction, Cyrus Levinthal postulated that a polypeptide chain of 150 amino acids has $10^{300}$ different conformations (Levinthal 1969), therefore, finding low-energy conformational states in this landscape is very challenging as it is computationally impossible to sample all conformations. Accurate energy calculations are crucial to determine these energy states, however, obvious caveats are the trade-off between accurate predictions and time (Huang et al. 2016).

To tackle the problem of chemical space, efficient sampling algorithms have been developed, as well as ways to reduce the chemical space. For example, structural data of evolutionary similar homologs can be used as a basis for a prediction model. Homology modeling can effectively be applied when a sequence shares at least 25% similarity (Jisna and Jayaraj 2021). However, if this data is not available, for example in *de novo* design, it becomes very challenging to accurately predict protein structure. Other approaches to guide search space sampling are model accuracy scoring functions. These functions try to estimate the probability of error of protein folds, e.g., based on residue-residue distance, and can give information about how the model should be altered to improve the structure (Hiranuma et al. 2021).

As the physical interaction models require extensive computing power for energy calculations and search space sampling, it has been proven to be very challenging to predict protein structures with these methods. Therefore, machine learning (ML) solutions have been intensively investigated, having resulted in highly accurate protein fold predictions with the development of AlphaFold2 (Jumper et al. 2021). AlphaFold2 uses a neural network approach that incorporates both biophysical structural knowledge and multiple sequence alignment algorithms. The median (n = 87) backbone Cα RMSD of AlphaFold2 is 0.96Å, compared to 2.8Å of the next best computational approach in the CASP14 competition, making AlphaFold2 significantly better than the other prediction methods. Other accurate neural network

approaches are RosseTTAFold, which employs a “three-track” neural network, and the very recent ESM Metagenomic Atlas (Lin et al. 2021) which is based on a language model.

## Computational protein binder design

To design a protein solely using computational methods, the protein structure prediction must be accurate, as well as the prediction of binding. Recently, developments in the field of *de novo* protein design have increased our understanding of this process (Huang et al. 2016; Pan and Kortemme 2021), as well as developments in de field of computational antibody design (Norman et al. 2020).

### *De novo* protein design

Design without any prior information about protein shape is a huge challenge, because, firstly, a binding site on the target must be identified, and secondly, there is no protein to base the design on, as would be with the design using homologs. For example, generating random protein sequences and testing all proteins for function is not feasible as the sequence space is enormous. To illustrate this, consider that there are 20 naturally occurring amino acids, meaning for a randomly designed *de novo* protein of 100 amino acids there would be $20^{100}$ different sequence possibilities (Huang et al. 2016). Between the 1980s and early 2000s, *de novo* design mostly made use of mathematical equations to define backbone conformations, and sidechain repacking algorithms design the final sequence. (Korendovych and DeGrado 2020). With the emergence of online protein structure repositories, a milestone in the field occurred in 2003. Here, Top7 was designed based on residue fragments from the protein databank, having a backbone RMSD of a mere 1.2Å compared to the experimentally determined structure (Kuhlman et al. 2003). Interestingly, the Top7 protein was observed to exhibit topology otherwise not found in nature. This finding paved the way to explore other unnatural protein folds, with new ways of binding and function.

There are two paradigms in computational design of proteins. The first is a design that uses known structures such as homologous structures, PDB fragments, or predefined topology. For example, helical bundles can be designed with atomic accuracy, and allow easy backbone sampling, thus providing good starting models for protein design (Hill et al. 2000). The second paradigm is the design without the knowledge of similar structures. Below, several examples will be given demonstrating the different approaches and processes of the *de novo* design of proteins.

To exemplify the first approach, recently, a *de novo* a-Helical protein was developed that stabilizes the adenosine $A_{2A}$ receptor (Mitsumoto et al. 2021). The authors selected a set of 389 backbone structures from a collection of 1688 globular a-backbone structures and designed amino acid sequences that stabilize each backbone. Designs were selected based on core packing, compatibility between structure and sequence, and the ability to be superimposed onto two helical loops of the alpha subunit of the target protein. Energy landscape limitations were applied, and molecular dynamics simulations were used to identify proteins containing low fluctuating N- and C-terminal helices. This resulted in two highly stable *de novo* fusion proteins, one of which showed nanomolar affinity towards the adenosine $A_{2A}$ receptor in its inactive state. It should be noted that the binding mode between these two proteins does not proceed via the classical orthosteric binding route but rather proceeds allosterically. An advantage of this is that the binder does not compete with endogenous ligands for the orthosteric binding site, therefore being likely to inhibit the receptor disregarding the endogenous effect (Nussinov and Tsai 2012). Proteins like this can be more favorable than therapeutic antibodies, as they contain better stability and size properties.

Another approach to *de novo* design has been described by Cao and colleagues (Cao et al. 2022). Here, the authors attempted to create small protein binders to 12 protein targets that

do not have clearly defined sites that can mediate strong interactions. In short, the process was started by identifying weak interacting residues on the target surface. Using the rotamer interaction field (RIF), disembodied amino acids were first docked against the target surface, thereby identifying polar rotamer binding possibilities and binding energy. RIF docking is highly optimized for speed (algorithmic complexity of O(1)) and generates favorable rotamer placements out of a large collection of possibilities. It has been extensively explained by Dou and colleagues (Dou et al. 2018). By the 'backward growing' of multiple weak binding residues, a library of sequences with correct backbone conformation and positioning of binding residues was created. This miniprotein library was again docked against the target using PatchDock to establish crude positioning of the binder. Using RIF docking, a higher resolution of rigid body orientation was acquired for each protein. Recurrent backbone motifs were then identified and used as guiding structures for further design optimizations of the scaffold library. Further experimental validation resulted in 10 to 100 binders for each target. Even though this process resulted in high-affinity binders, obvious limitations are its resource intensiveness and complex workflow. Besides multiple high throughput docking methods and highly specific optimization conditions, experimental validation is also necessary. For example, only a small percentage out of millions of computed proteins binds to the target, but even after various library size limiting steps, it is still paramount to experimentally validate numerous proteins. In this case, DNA synthesis, multiple rounds of yeast display screening, solubility- and binding assays, as well as the need for generating a mutagenesis library are considerable bottlenecks.

Other examples include the use of ML algorithms, such as the use of a Rossetta-based denoising diffusion model called RF*diffusion* (Watson et al. n.d.). Here, the authors designed high-affinity binders (picomolar range) to apoptosis-related peptide Bim and parathyroid hormone. Using this approach, no topology or binding mode needs to be specified, which greatly enhances the de novo protein design process as it is not restricted to using sequence homologs or predefined scaffolds. Others used Bayesian optimization to make accurate predictions of possible peptide inhibitors (Yang, Milas, and White n.d.). A recurring problem is the lack of accuracy in the output of *in silico* high-throughput screens. By modifying existing pre-trained models, such as AlphaFold2, Bayesian optimization can be used to optimize sequences as error rates are also defined. The authors found likely hemolytic peptides as well as a binder of Ras GTPase. Although new machine-learning approaches like this can be useful in later research, the findings were not experimentally validated.

## Computational antibody design

Besides the *de novo* design of small protein binders, there have been many efforts to computationally design antibodies and CDRs directly. Previous ML approaches for antibody design were often limited by the low amount of training data (~10000), such as PyIgClassify (Robert et al. n.d.)(Adolf-Bryfogle et al. 2015), but recently Absolut! was developed, which is a massive (~$10^8$) dataset containing 6.9 million antibodies docked to 159 different antigens. These antibody-antigen binding pairs were scored by calculated binding energy. The dataset conforms to eight physical constraints, thereby preserving biological complexity, such as antigen topology, making it relevant for real-world applications (Khan et al. 2022). A dataset like this is crucial for benchmarking following antibody design ML initiatives and has been readily used in AntBO (Khan et al. 2022). AntBO was developed to design CDR-H3 sequences based on an antigen of interest and can suggest high-affinity antibodies while maintaining developability, based on a Bayesian optimization algorithm. As an example, the authors computed 200 antibodies against the S protein of SARS-CoV, each with multiple developability parameters. While sounding promising, no experimental validation has yet been performed. A myriad of other computational design approaches has been developed, some of which are OptCDR and its successor OptMAVEn (Li, Pantazes, and Maranas 2014; Pantazes and Maranas 2010), which are based on energy-function models. The recent review article by

Norman et al. provides an excellent overview of current trends in computational antibody design (Norman et al. 2020).

# Closing remarks

This essay has thus far provided an overview of different trends in experimental and computational antibody and binder design. The question of whether experimentally developed antibodies will become obsolete due to the emergence of computational strategies is speculative in nature, therefore the answer should be regarded only as an opinion about possible future prospects. It should be noted that this piece is by no means an exhaustive outline of all trends within the field and will therefore be biased by its limited scope.

Indeed, computational approaches have long remained dormant in the protein design field, but now, with computing power higher than ever and the possibility to produce synthetic DNA has made it possible to completely create proteins from scratch. However, to state that experimental methods will become completely outdated is currently unlikely, as all computational methods still need experimental validation. Computational methods are very promising, but often propose a large number of possible binders, out of which only a fraction will fold correctly, let alone bind to the target. It is also important to realize that *in silico* methods cannot provide definitive accurate binders, but only suggestions at best. It is therefore still crucial to screen most binders for the correct folding and binding, which can be resource and time-consuming. With the emergence of AlphaFold2, the ultimate fold of small and thermostable proteins is mostly solved, which are favorable developability qualities for therapeutics. The influence of the computational design field is and will be significant, based on the coming of age of the field. There has been a myriad of solutions to solve the problem of sequence space sampling, using machine learning algorithms among others, as well as different approaches to identify binding modes to undruggable targets, and we are now paving the way to very accurate structural predictions. With regards to experimental approaches, the recent advancements in next-generation sequencing make it possible to elucidate *in vitro* library diversity in higher amounts than ever before and to sequence possible high-affinity binders with low expression, improving the quality of output of display methods substantially. Indeed, it is still impossible to bring complete antibodies to expression with display methods, but rather fragments of the proteins. Because the context of the entire antibody is lost, the output of these high-throughput methods does not necessarily translate to complete antibodies. To conclude, both approaches are not perfect by any means, are continuing to improve, and will likely keep complementing one another in the following years.

To argue whether *de novo* binders will replace antibodies in the future, let us first look at the qualities of both binders. *De novo* binders are often small, very thermostable, and have very efficient binding to their target, which makes them attractive candidates as therapeutics. And however similar in specificity, antibodies are large macromolecules susceptible to aggregation which can influence immunogenicity. These improved biophysical properties alone are a great incentive for the development of small protein binders and a similar rationale led to the development of nanobodies. Regardless, antibodies can activate the immune system with its $F_c$-region, which can be the leading mode of action against a certain target. This is something that small protein binders cannot readily do, and it will likely be challenging to incorporate similar properties into small protein binders while retaining the aforementioned qualities. That being said, they show great potential as therapeutics and likely will be a substantial group of future pharmaceuticals.

# Literature


Abanades, Brennan, Guy Georges, Alexander Bujotzek, and Charlotte M. Deane. 2022. 'ABlooper: Fast Accurate Antibody CDR Loop Structure Prediction with Accuracy Estimation'. *Bioinformatics* 38(7):1877–80. doi: 10.1093/bioinformatics/btac016.

Adams, Jarrett J., and Sachdev S. Sidhu. 2014. 'Synthetic Antibody Technologies'. *Current Opinion in Structural Biology* 24:1–9. doi: 10.1016/j.sbi.2013.11.003.

Adolf-Bryfogle, Jared, Qifang Xu, Benjamin North, Andreas Lehmann, and Roland L. Dunbrack. 2015. 'PyIgClassify: A Database of Antibody CDR Structural Classifications'. *Nucleic Acids Research* 43(D1):D432–38. doi: 10.1093/nar/gku1106.

Alfaleh, Mohamed A., Hashem O. Alsaab, Ahmad Bakur Mahmoud, Almohanad A. Alkayyal, Martina L. Jones, Stephen M. Mahler, and Anwar M. Hashem. 2020. 'Phage Display Derived Monoclonal Antibodies: From Bench to Bedside'. *Frontiers in Immunology* 11. doi: 10.3389/fimmu.2020.01986.

Beerli, Roger R., and Christoph Rader. 2010. 'Mining Human Antibody Repertoires'. *MAbs* 2(4):365–78. doi: 10.4161/mabs.12187.

Bordron, Anne, Cristina Bagacean, Adrian Tempescul, Christian Berthou, Eléonore Bettacchioli, Sophie Hillion, and Yves Renaudineau. 2020. 'Complement System: A Neglected Pathway in Immunotherapy'. *Clinical Reviews in Allergy & Immunology* 58(2):155–71. doi: 10.1007/s12016-019-08741-0.

Boyken, Scott E., Zibo Chen, Benjamin Groves, Robert A. Langan, Gustav Oberdorfer, Alex Ford, Jason M. Gilmore, Chunfu Xu, Frank DiMaio, Jose Henrique Pereira, Banumathi Sankaran, Georg Seelig, Peter H. Zwart, and David Baker. 2016. 'De Novo Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network–Mediated Specificity'. *Science* 352(6286):680–87. doi: 10.1126/science.aad8865.

Cao, Longxing, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M. Jude, Iva Marković, Rameshwar U. Kadam, Koen H. G. Verschueren, Kenneth Verstraete, Scott Thomas Russell Walsh, Nathaniel Bennett, Ashish Phal, Aerin Yang, Lisa Kozodoy, Michelle DeWitt, Lora Picton, Lauren Miller, Eva-Maria Strauch, Nicholas D. DeBouver, Allison Pires, Asim K. Bera, Samer Halabiya, Bradley Hammerson, Wei Yang, Steffen Bernard, Lance Stewart, Ian A. Wilson, Hannele Ruohola-Baker, Joseph Schlessinger, Sangwon Lee, Savvas N. Savvides, K. Christopher Garcia, and David Baker. 2022. 'Design of Protein-Binding Proteins from the Target Structure Alone'. *Nature* 605(7910):551–60. doi: 10.1038/s41586-022-04654-9.

Chiu, Mark L., and Gary L. Gilliland. 2016. 'Engineering Antibody Therapeutics'. *Current Opinion in Structural Biology* 38:163–73. doi: 10.1016/j.sbi.2016.07.012.

Donaldson, Joshua M., Cindy Zer, Kendra N. Avery, Krzysztof P. Bzymek, David A. Horne, and John C. Williams. 2013. 'Identification and Grafting of a Unique Peptide-Binding Site in the Fab Framework of Monoclonal Antibodies'. *Proceedings of the National Academy of Sciences* 110(43):17456–61. doi: 10.1073/pnas.1307309110.

Dou, Jiayi, Anastassia A. Vorobieva, William Sheffler, Lindsey A. Doyle, Hahnbeom Park, Matthew J. Bick, Binchen Mao, Glenna W. Foight, Min Yen Lee, Lauren A. Gagnon, Lauren Carter, Banumathi Sankaran, Sergey Ovchinnikov, Enrique Marcos, Po-Ssu Huang, Joshua C. Vaughan,

Barry L. Stoddard, and David Baker. 2018. 'De Novo Design of a Fluorescence-Activating β-Barrel'. *Nature* 561(7724):485–91. doi: 10.1038/s41586-018-0509-0.

Fellouse, Frederic A., Bing Li, Deanne M. Compaan, Andrew A. Peden, Sarah G. Hymowitz, and Sachdev S. Sidhu. 2005. 'Molecular Recognition by a Binary Code'. *Journal of Molecular Biology* 348(5):1153–62. doi: 10.1016/j.jmb.2005.03.041.

Goldberg, Benjamin S., and Margaret E. Ackerman. 2020. 'Antibody-mediated Complement Activation in Pathology and Protection'. *Immunology & Cell Biology* 98(4):305–17. doi: 10.1111/imcb.12324.

He, Bifang, Lixu Jiang, Yaocong Duan, Guoshi Chai, Yewei Fang, Juanjuan Kang, Min Yu, Ning Li, Zhongjie Tang, Pengcheng Yao, Pengcheng Wu, Ratmir Derda, and Jian Huang. 2018. 'Biopanning Data Bank 2018: Hugging next Generation Phage Display'. *Database* 2018. doi: 10.1093/database/bay032.

Hermanto, S., M. Yusuf, A. Mutalib, and S. Hudiyono. 2017. 'Molecular Dynamic Simulation of Trastuzumab $F(Ab')_2$ Structure in Corporation with HER2 as a Theranostic Agent of Breast Cancer'. *Journal of Physics: Conference Series* 835:012005. doi: 10.1088/1742-6596/835/1/012005.

Hill, R. Blake, Daniel P. Raleigh, Angela Lombardi, and William F. DeGrado. 2000. '*De Novo* Design of Helical Bundles as Models for Understanding Protein Folding and Function'. *Accounts of Chemical Research* 33(11):745–54. doi: 10.1021/ar970004h.

Hiranuma, Naozumi, Hahnbeom Park, Minkyung Baek, Ivan Anishchenko, Justas Dauparas, and David Baker. 2021. 'Improved Protein Structure Refinement Guided by Deep Learning Based Accuracy Estimation'. *Nature Communications* 12(1):1340. doi: 10.1038/s41467-021-21511-x.

Hon, Ting, Kristin Mars, Greg Young, Yu-Chih Tsai, Joseph W. Karalius, Jane M. Landolin, Nicholas Maurer, David Kudrna, Michael A. Hardigan, Cynthia C. Steiner, Steven J. Knapp, Doreen Ware, Beth Shapiro, Paul Peluso, and David R. Rank. 2020. 'Highly Accurate Long-Read HiFi Sequencing Data for Five Complex Genomes'. *Scientific Data* 7(1):399. doi: 10.1038/s41597-020-00743-4.

Hoogenboom, Hennie R., Adriaan P. de Bruïne, Simon E. Hufton, René M. Hoet, Jan-Willem Arends, and Rob C. Roovers. 1998. 'Antibody Phage Display Technology and Its Applications'. *Immunotechnology* 4(1):1–20. doi: 10.1016/S1380-2933(98)00007-4.

Huang, Po Ssu, Scott E. Boyken, and David Baker. 2016. 'The Coming of Age of de Novo Protein Design'. *Nature* 537(7620):320–27.

Iha, Kanako, Mikio Inada, Naoki Kawada, Kazunari Nakaishi, Satoshi Watabe, Yong Hong Tan, Chieh Shen, Liang-Yin Ke, Teruki Yoshimura, and Etsuro Ito. 2019. 'Ultrasensitive ELISA Developed for Diagnosis'. *Diagnostics* 9(3):78. doi: 10.3390/diagnostics9030078.

Jisna, V. A., and P. B. Jayaraj. 2021. 'Protein Structure Prediction: Conventional and Deep Learning Perspectives'. *The Protein Journal* 40(4):522–44. doi: 10.1007/s10930-021-10003-y.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer,

Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. 2021. 'Highly Accurate Protein Structure Prediction with AlphaFold'. *Nature* 596(7873):583–89. doi: 10.1038/s41586-021-03819-2.

Kaleli, Nazlı Eda, Murat Karadag, and Sibel Kalyoncu. 2019. 'Phage Display Derived Therapeutic Antibodies Have Enriched Aliphatic Content: Insights for Developability Issues'. *Proteins: Structure, Function, and Bioinformatics* 87(7):607–18. doi: 10.1002/prot.25685.

Keefe, Anthony D., and Jack W. Szostak. 2001. 'Functional Proteins from a Random-Sequence Library'. *Nature* 410(6829):715–18. doi: 10.1038/35070613.

Khan, Asif, Alexander I. Cowen-Rivers, Antoine Grosnit, Derrick-Goh-Xin Deik, Philippe A. Robert, Victor Greiff, Eva Smorodina, Puneet Rawat, Kamil Dreczkowski, Rahmad Akbar, Rasul Tutunov, Dany Bou-Ammar, Jun Wang, Amos Storkey, and Haitham Bou-Ammar. 2022. 'AntBO: Towards Real-World Automated Antibody Design with Combinatorial Bayesian Optimisation'.

Kim, Jin Hong, and Hyo Jeong Hong. 2012. 'Humanization by CDR Grafting and Specificity-Determining Residue Grafting'. Pp. 237–45 in.

Korendovych, Ivan v., and William F. DeGrado. 2020. '*De Novo* Protein Design, a Retrospective'. *Quarterly Reviews of Biophysics* 53:e3. doi: 10.1017/S0033583519000131.

Kuhlman, Brian, Gautam Dantas, Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, and David Baker. 2003. 'Design of a Novel Globular Protein Fold with Atomic-Level Accuracy'. *Science* 302(5649):1364–68. doi: 10.1126/science.1089427.

Kwok, Gerry, Thomas C. C. Yau, Joanne W. Chiu, Eric Tse, and Yok-Lam Kwong. 2016. 'Pembrolizumab (Keytruda)'. *Human Vaccines & Immunotherapeutics* 12(11):2777–89. doi: 10.1080/21645515.2016.1199310.

Ledsgaard, Line, Mogens Kilstrup, Aneesh Karatt-Vellatt, John McCafferty, and Andreas Laustsen. 2018. 'Basics of Antibody Phage Display Technology'. *Toxins* 10(6):236. doi: 10.3390/toxins10060236.

Levinthal, Cyrus. 1969. 'Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois.' *University of Illinois Press* 22–24.

Li, Ruowei, Guangbo Kang, Min Hu, and He Huang. 2019. 'Ribosome Display: A Potent Display Technology Used for Selecting and Evolving Specific Binders with Desired Properties'. *Molecular Biotechnology* 61(1):60–71. doi: 10.1007/s12033-018-0133-0.

Li, Tong, Robert J. Pantazes, and Costas D. Maranas. 2014. 'OptMAVEn – A New Framework for the de Novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes'. *PLoS ONE* 9(8):e105954. doi: 10.1371/journal.pone.0105954.

Lin, Zeming, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and [2] Alexander Rives. 2021. 'Evolutionary-Scale Prediction of Atomic Level Protein Structure with a Language Model'. doi: 10.1101/2022.07.20.500902.

Liu, Justin K. H. 2014. 'The History of Monoclonal Antibody Development – Progress, Remaining Challenges and Future Innovations'. *Annals of Medicine and Surgery* 3(4):113–16. doi: 10.1016/j.amsu.2014.09.001.

Lv, Shuzhen, Yunchao Wang, Tao Sun, Donggui Wan, Lijun Sheng, Wei Li, Huayun Zhu, Yanping Li, and Janice Lu. 2018. 'Overall Survival Benefit from Trastuzumab-Based Treatment in HER2-Positive Metastatic Breast Cancer: A Retrospective Analysis'. *Oncology Research and Treatment* 41(7–8):450–55. doi: 10.1159/000488202.

Mitsumoto, Masaya, Kanna Sugaya, Kazuki Kazama, Ryosuke Nakano, Takahiro Kosugi, Takeshi Murata, and Nobuyasu Koga. 2021. 'State-Targeting Stabilization of Adenosine A2A Receptor by Fusing a Custom-Made De Novo Designed α-Helical Protein'. *International Journal of Molecular Sciences* 22(23):12906. doi: 10.3390/ijms222312906.

Moraes, Jane Zveiter, Bárbara Hamaguchi, Camila Braggion, Enzo Reina Speciale, Fernanda Beatriz Viana Cesar, Gabriela de Fátima da Silva Soares, Juliana Harumi Osaki, Tauane Mathias Pereira, and Rodrigo Barbosa Aguiar. 2021. 'Hybridoma Technology: Is It Still Useful?' *Current Research in Immunology* 2:32–40. doi: 10.1016/j.crimmu.2021.03.002.

Mullard, Asher. 2021. 'FDA Approves 100th Monoclonal Antibody Product'. *Nature Reviews Drug Discovery* 20(7):491–95. doi: 10.1038/d41573-021-00079-7.

Muyldermans, Serge. 2021. 'Applications of Nanobodies'. *Annual Review of Animal Biosciences* 9(1):401–21. doi: 10.1146/annurev-animal-021419-083831.

Nannini, Francesco, Lenart Senicar, Farhaan Parekh, Khai J. Kong, Alexander Kinna, Reyisa Bughda, James Sillibourne, Xihao Hu, Biao Ma, Yuchen Bai, Mathieu Ferrari, Martin A. Pule, and Shimobi C. Onuoha. 2021. 'Combining Phage Display with SMRTbell Next-Generation Sequencing for the Rapid Discovery of Functional ScFv Fragments'. *MAbs* 13(1). doi: 10.1080/19420862.2020.1864084.

Newberry, Robert W., and Ronald T. Raines. 2019. 'Secondary Forces in Protein Folding'. *ACS Chemical Biology* 14(8):1677–86. doi: 10.1021/acschembio.9b00339.

Newton, Matilda S., Yari Cabezas-Perusse, Cher Ling Tong, and Burckhard Seelig. 2020. '*In Vitro* Selection of Peptides and Proteins—Advantages of MRNA Display'. *ACS Synthetic Biology* 9(2):181–90. doi: 10.1021/acssynbio.9b00419.

Norman, Peter. 2017. 'Humira : Recent Developments'. *Pharmaceutical Patent Analyst* 6(3):89–90. doi: 10.4155/ppa-2017-0012.

Norman, Richard A., Francesco Ambrosetti, Alexandre M. J. J. Bonvin, Lucy J. Colwell, Sebastian Kelm, Sandeep Kumar, and Konrad Krawczyk. 2020. 'Computational Approaches to Therapeutic Antibody Design: Established Methods and Emerging Trends'. *Briefings in Bioinformatics* 21(5):1549–67. doi: 10.1093/bib/bbz095.

Nussinov, Ruth, and Chung-Jung Tsai. 2012. 'The Different Ways through Which Specificity Works in Orthosteric and Allosteric Drugs'. *Current Pharmaceutical Design* 18(9):1311–16. doi: 10.2174/138161212799436377.

Pan, Xingjie, and Tanja Kortemme. 2021. 'Recent Advances in de Novo Protein Design: Principles, Methods, and Applications'. *Journal of Biological Chemistry* 296:100558. doi: 10.1016/j.jbc.2021.100558.

Pantazes, R. J., and C. D. Maranas. 2010. 'OptCDR: A General Computational Method for the Design of Antibody Complementarity Determining Regions for Targeted Epitope Binding'. *Protein Engineering Design and Selection* 23(11):849–58. doi: 10.1093/protein/gzq061.

Parray, Hilal Ahmed, Shivangi Shukla, Sweety Samal, Tripti Shrivastava, Shubbir Ahmed, Chandresh Sharma, and Rajesh Kumar. 2020. 'Hybridoma Technology a Versatile Method for Isolation of Monoclonal Antibodies, Its Applicability across Species, Limitations, Advancement and Future Perspectives'. *International Immunopharmacology* 85:106639. doi: 10.1016/j.intimp.2020.106639.

Prassler, Josef, Stefanie Thiel, Catrin Pracht, Andrea Polzer, Solveig Peters, Marion Bauer, Stephanie Nörenberg, Yvonne Stark, Johanna Kölln, Andreas Popp, Stefanie Urlinger, and Markus Enzelberger. 2011. 'HuCAL PLATINUM, a Synthetic Fab Library Optimized for Sequence Diversity and Superior Performance in Mammalian Expression Systems'. *Journal of Molecular Biology* 413(1):261–78. doi: 10.1016/j.jmb.2011.08.012.

Ratanji, Kirsty D., Jeremy P. Derrick, Rebecca J. Dearman, and Ian Kimber. 2014. 'Immunogenicity of Therapeutic Proteins: Influence of Aggregation'. *Journal of Immunotoxicology* 11(2):99–109. doi: 10.3109/1547691X.2013.821564.

Robert, Philippe A., Rahmad Akbar, Robert Frank, Milena Pavlović, Michael Widrich, Igor Snapkov, Andrei Slabodkin, Maria Chernigovskaya, Lonneke Scheffer, Eva Smorodina, Puneet Rawat, Brij Bhushan Mehta, Mai Ha Vu, Ingvild Frøberg Mathisen, Aurél Prósz, Krzysztof Abram, Alex Olar, Enkelejda Miho, Dag Trygve, Tryslew Haug, Fridtjof Lund-Johansen, Sepp Hochreiter, Ingrid Hobaek Haff, Günter Klambauer, Geir Kjetil Sandve, and Victor Greiff. n.d. 'Unconstrained Generation of Synthetic Antibody-Antigen Structures to Guide Machine Learning Methodology for Real-World Antibody Specificity Prediction'. doi: 10.1101/2021.07.06.451258.

Rondot, Susanne, Joachim Koch, Frank Breitling, and Stefan Dübel. 2001. 'A Helper Phage to Improve Single-Chain Antibody Presentation in Phage Display'. *Nature Biotechnology* 19(1):75–78. doi: 10.1038/83567.

Rouet, Romain, Katherine J. L. Jackson, David B. Langley, and Daniel Christ. 2018. 'Next-Generation Sequencing of Antibody Display Repertoires'. *Frontiers in Immunology* 9. doi: 10.3389/fimmu.2018.00118.

Salema, Valencio, and Luis Ángel Fernández. 2017. '*Escherichia* Coli Surface Display for the Selection of Nanobodies'. *Microbial Biotechnology* 10(6):1468–84. doi: 10.1111/1751-7915.12819.

Sumida, Takeshi, Hiroshi Yanagawa, and Nobuhide Doi. 2012. '*In Vitro* Selection of Fab Fragments by MRNA Display and Gene-Linking Emulsion PCR'. *Journal of Nucleic Acids* 2012:1–9. doi: 10.1155/2012/371379.

Sun, Yi, Tian Huang, Lennart Hammarström, and Yaofeng Zhao. 2020. 'The Immunoglobulins: New Insights, Implications, and Applications'. *Annual Review of Animal Biosciences* 8(1):145–69. doi: 10.1146/annurev-animal-021419-083720.

Teymennet-Ramírez, Karla v., Fernando Martínez-Morales, and María R. Trejo-Hernández. 2022. 'Yeast Surface Display System: Strategies for Improvement and Biotechnological Applications'. *Frontiers in Bioengineering and Biotechnology* 9. doi: 10.3389/fbioe.2021.794742.

Urquhart, Lisa. 2022. 'Top Companies and Drugs by Sales in 2021'. *Nature Reviews Drug Discovery* 21(4):251.

Vadnais, Melissa L., Michael F. Criscitiello, and Vaughn v. Smider. 2017. 'Antibodies from Other Species'. Pp. 85–112 in.

Vasselli, James Robert, Sophia Frentzas, Andrew James Weickhardt, Paul L. de Souza, Jenny Tang, Tim Wyant, Inbar Amit, Yanay Ofran, and Aron Knickerbocker. 2022. 'Trial in Progress: A Phase 1-2, First-in-Human, Open Label, Dose Escalation and Expansion Study of AU-007, a Monoclonal Antibody That Binds to IL-2 and Inhibits IL-2Rα Binding, in Patients with Advanced Solid Tumors.' *Journal of Clinical Oncology* 40(16_suppl):TPS2671–TPS2671. doi: 10.1200/JCO.2022.40.16_suppl.TPS2671.

Wang, Ying, Jun Liu, Peter D. Burrows, and Ji-Yang Wang. 2020. 'B Cell Development and Maturation'. Pp. 1–22 in.

Watson, Joseph L., David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Preetham Venkatesh, Susana Vázquez Torres, Andrew J. Borst, Basile I. M. Wicky, Robert J. Ragotte, Lukas F. Milles, Alexis Courbet, William Sheffler, Jue Wang, Isaac Sappington, Samuel J. Pellock, Nikita Hanikel, Anna Lauko, Regina Barzilay, Tommi S. Jaakkola, Frank Dimaio, Minkyung Baek, and David Baker. n.d. *Broadly Applicable and Accurate Protein Design by Integrating Structure Prediction Networks and Diffusion Generative Models*.

Xiao, Xiaodong, Julie A. Douthwaite, Yan Chen, Ben Kemp, Sara Kidd, Jennifer Percival-Alwyn, Alison Smith, Kate Goode, Bonnie Swerdlow, David Lowe, Herren Wu, William F. Dall'Acqua, and Partha S. Chowdhury. 2017. 'A High-Throughput Platform for Population Reformatting and Mammalian Expression of Phage Display Libraries to Enable Functional Screening as Full-Length IgG'. *MAbs* 9(6):996–1006. doi: 10.1080/19420862.2017.1337617.

Yang, Ziyue, Katarina A. Milas, and Andrew D. White. n.d. 'Now What Sequence? Pre-Trained Ensembles for Bayesian Optimization of Protein Sequences'. doi: 10.1101/2022.08.05.502972.

Zhou, Shu-Feng, and Wei-Zhu Zhong. 2017. 'Drug Design and Discovery: Principles and Applications'. *Molecules* 22(2):279. doi: 10.3390/molecules22020279.