frontiers | Frontiers in Molecular Biosciences

TYPE Review
PUBLISHED 07 August 2023
DOI 10.3389/fmolb.2023.1221626

Check for updates

OPEN ACCESS

EDITED BY
Shubhra Ghosh Dastidar,
Bose Institute, India

REVIEWED BY
Moutusi Manna,
Central Salt and Marine Chemicals Research Institute (CSIR), India
Frederic Cadet,
INSERM Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), France

*CORRESPONDENCE
Sandeep Kumar,
Sandeep.Kumar@modernatx.com

†PRESENT ADDRESS
Sandeep Kumar, Computational Protein Design and Modeling, Computational Sciences, Moderna Therapeutics, Cambridge, MA, United States

RECEIVED 12 May 2023
ACCEPTED 10 July 2023
PUBLISHED 07 August 2023

CITATION
Bauer J, Rajagopal N, Gupta P, Gupta P, Nixon AE and Kumar S (2023), How can we discover developable antibody-based biotherapeutics?
Front. Mol. Biosci. 10:1221626.
doi: 10.3389/fmolb.2023.1221626

COPYRIGHT
© 2023 Bauer, Rajagopal, Gupta, Gupta, Nixon and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

How can we discover developable antibody-based biotherapeutics?

Joschka Bauer[1,2], Nandhini Rajagopal[2,3], Priyanka Gupta[2,3], Pankaj Gupta[2,3], Andrew E. Nixon[3] and Sandeep Kumar[2,3]*†

[1]Early Stage Pharmaceutical Development Biologicals, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach/Riss, Germany, [2]In Silico Team, Boehringer Ingelheim, Hannover, Germany, [3]Biotherapeutics Discovery, Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, United States

Antibody-based biotherapeutics have emerged as a successful class of pharmaceuticals despite significant challenges and risks to their discovery and development. This review discusses the most frequently encountered hurdles in the research and development (R&D) of antibody-based biotherapeutics and proposes a conceptual framework called biopharmaceutical informatics. Our vision advocates for the syncretic use of computation and experimentation at every stage of biologic drug discovery, considering developability (manufacturability, safety, efficacy, and pharmacology) of potential drug candidates from the earliest stages of the drug discovery phase. The computational advances in recent years allow for more precise formulation of disease concepts, rapid identification, and validation of targets suitable for therapeutic intervention and discovery of potential biotherapeutics that can agonize or antagonize them. Furthermore, computational methods for *de novo* and epitope-specific antibody design are increasingly being developed, opening novel computationally driven opportunities for biologic drug discovery. Here, we review the opportunities and limitations of emerging computational approaches for optimizing antigens to generate robust immune responses, *in silico* generation of antibody sequences, discovery of potential antibody binders through virtual screening, assessment of hits, identification of lead drug candidates and their affinity maturation, and optimization for developability. The adoption of biopharmaceutical informatics across all aspects of drug discovery and development cycles should help bring affordable and effective biotherapeutics to patients more quickly.

KEYWORDS

biotherapeutics, drug discovery and development, developability, biopharmaceutical informatics, machine learning, computational biophysics

1 Introduction

Since the inception of hybridoma technology, which facilitated large-scale monoclonal antibody (mAb) production, biotherapeutics have experienced significant growth (Koehler and Milstein, 1975). The Food and Drug Administration's (FDA) approval of the pioneering mAb therapeutic, muromonab or Orthoclone OKT3, in 1986 (Smith, 1996), set the stage for numerous groundbreaking developments in biotherapeutics. As of 2022, over 110 approved mAbs and more than 65 mAbs in phase-2/3 and phase-3 clinical trials have emerged (Kaplon et al., 2022). Clinically, mAbs have demonstrated their efficacy in treating serious conditions such as neurodegenerative diseases, autoimmune diseases, and diverse types of cancers (Reichert et al., 2009; Lu et al., 2020).

Despite the promising trajectory of biotherapeutics, the biopharmaceutical industry faces mounting pressure due to decreasing productivity and increasing research and development (R&D) costs. The average R&D cost surged from $1.2 billion in 2007 (adjusted United States dollar value of $1.6 billion in 2020) to $2.8 billion in 2016 (equivalent to $3.1 billion in 2020) (DiMasi and Grabowski, 2007; DiMasi et al., 2016; Farid et al., 2020). Concurrently, the success rate of phase-1 to approval dropped from 30% in 2007 to 12% or lower in 2016 (Farid et al., 2020). These trends suggest the presence of several challenges along various stages of discovery and development of novel biological therapeutics. A lack of detailed understanding of disease biology, the inability of model systems to reliably predict human diseases and outcomes of therapeutic interventions, the lack of efficacy, target-mediated toxicity and other safety issues, and suboptimal developability profiles are among the major reasons that may contribute to drug failures during clinical trials (Mehta et al., 2017; Fogel, 2018). The identification of new targets presents additional challenges toward development of novel therapeutic concepts and discovery of multi-specific biotherapeutics, resulting in low approval rates despite high development costs (Swinney and Anthony, 2011). Next-generation biotherapeutics such as nanobodies, bi- and multi-specific antibodies, and T-cell receptor mimetics are broadening clinical applications (Strohl, 2018); however, these novel formats are often more challenging to develop into marketed biologic drug products (Runcie et al., 2018; Wang et al., 2019; Sawant et al., 2020). Furthermore, as the biopharmaceutical industry shifts its focus toward patient convenience, drug product development processes must be tailored to emerging routes of drug administration such as subcutaneous or intravitreal delivery, necessitating high-concentration protein formulations (HCPFs) (Garidel et al., 2017). These requirements introduce additional challenges to the manufacturability and developability of novel drugs. Integrating developability early in the drug discovery process can help avoid costly delays or failures at later stages and potentially increase the likelihood of success during clinical trials and approvals. Numerous technological advancements have been made since the approval of the first mAb to overcome challenges in the R&D pipelines and accelerate novel drug discovery and development (Martin et al., 2023). However, every new technology comes with associated risks and limitations (Gray A. C. et al., 2020; González-Fernández et al., 2020).

In silico techniques have been well established in small-molecule drug discovery (Shaker et al., 2021). Over the past decade, considerable progress has been made toward developing *in silico* strategies for the discovery and development of biologic drugs as well. In fact, developability has emerged as a key concept for biologic drugs over this time (Jarasch et al., 2015; Kumar and Singh, 2015; Bailly et al., 2020; Garripelli et al., 2020; Khetan et al., 2022; Mieczkowski et al., 2023). A variety of computational tools and procedures are now employed across various stages of drug development, such as hit selection, lead identification, optimization, affinity maturation, and early developability assessment. However, a significant potential of *in silico* technologies toward the discovery of biotherapeutics still remains untapped. As collaborative academic and industrial initiatives continue to demonstrate the viability of *in silico* antibody discovery techniques, it is important to acknowledge that the nascent nature of these methods often results in a lack of historical evidence to support their success and therefore requires a cultural shift toward proactive adoption of innovation to continually improve drug discovery and development processes. To address these challenges and enhance the success rate of novel targets, there is an urgent requirement for an integrated vision to create a platform that streamlines biotherapeutic discovery and development via syncretic use of experimentation and computation. Such a vision would not only accelerate the development of new biotherapeutics and reduce costs but also expand the druggable target space.

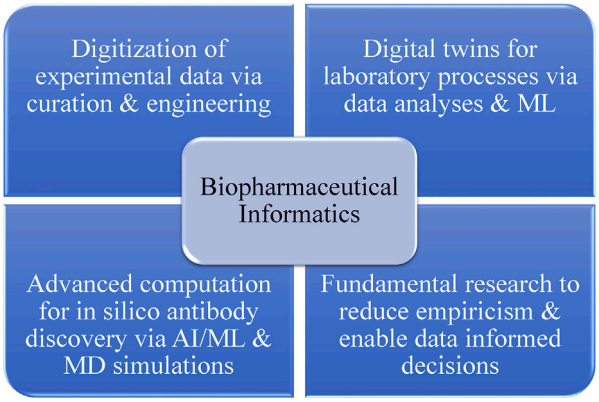

FIGURE 1
Strategic components for the vision of biopharmaceutical informatics. The digital transformation of the biopharmaceutical industry, achieved through capturing and curing experimental data, can enable the development and continuous improvement of digital twins for laboratory processes and prediction of experimental results before their execution. Fundamental research connecting molecular sequences, structures, and dynamics of biologic drug candidates can enhance our understanding of experimental observations, reduce empiricism, and enable more data-informed decision-making at various project stages. Moreover, the integration of computational learning technologies with principles of molecular modeling and simulations can potentially facilitate the *in silico* discovery of biotherapeutics. It is important to note that the key to biopharmaceutical informatics lies in the syncretic use of experimentation and computation, with a shared goal of making the discovery and development of biotherapeutics more efficient.

2 Biopharmaceutical informatics: integrating drug discovery and development

In the realm of biotherapeutics, it is crucial for drug candidates to be both developable and functional. Biotherapeutic drug candidates often encounter developability challenges related to manufacturing, safety, immunogenicity, efficacy, pharmacology, and drug product heterogeneity. Many of these risks can be linked to the inherent physicochemical properties of a biologic drug candidate, as determined by its protein sequence, three-dimensional structure, and molecular dynamics (MD) (Xu et al., 2018). Considering the intrinsic physicochemical properties of a biotherapeutic drug candidate, which are encoded in its amino acid

sequence and structure, early in the discovery and development can help identify and mitigate risks associated with various developability issues, such as chemical, conformational, colloidal, and physical instabilities. Moreover, by employing the innovative approach of biopharmaceutical informatics, these sequence–structural attributes can be modified for improved developability as described previously by Kumar et al. (2018a). Figure 1 outlines the primary components of biopharmaceutical informatics. This interdisciplinary field advocates for the digital transformation of the biopharmaceutical industry by converting experimental data collected during drug discovery and development phases into FAIR (findable, accessible, interoperable, and reusable) information systems. These systems can be leveraged by data scientists to create predictive tools such as digital twins of actual laboratory processes. Additionally, the field promotes the increased use of AI/ML (artificial intelligence/machine learning) and computational biophysics to address fundamental challenges in drug discovery and development through research. Biopharmaceutical informatics seeks to enable data-driven decision-making at every stage of biologic drug discovery and development. Developability is a key aspect of biopharmaceutical informatics, encompassing both *in silico* tools and experimental studies such as developability assessments. Rooted in the energy landscape theory, the concept of developability posits that the conformational ensembles and potential energy landscapes of large macromolecules, like mAbs, change with their environment (e.g., pH, temperature, and physicochemical state) (Onuchic, 1997; Ma et al., 2000; Kumar et al., 2009). As a result, the physicochemical properties of conformational ensembles of biotherapeutics under a given set of environmental conditions dictate their biophysical experiment outcomes. If proteins with the same size and fold are analyzed under identical conditions using standardized experiments, differences in the results should be attributable to sequence–structural variations among the proteins. The ability to predict experimental outcomes by analyzing the sequence–structural characteristics of biotherapeutic drug candidates is a primary goal for biopharmaceutical informatics, as part of the development of computational methods that facilitate discovery of antibodies *in silico* (DAbI).

Optimal synergies and benefits can be achieved by integrating cost-effective, rapid computational methods with standardized biophysical experimental studies, which are characteristic of current developability assessments in biologic drug discovery and early-stage product development (Zurdo, 2013; Jarasch et al., 2015; Xu et al., 2018). Late-stage development approaches typically focus on assessing the changing conditions of a single molecule in the drug manufacturing process using quantitative unit operation models (Smiatek et al., 2020), while early-stage approaches require analyzing a diverse set of molecules under identical conditions. Biopharmaceutical informatics plays a pivotal role in bridging the gap between biologic drug discovery and development by improving the understanding of the relationship between macromolecular sequence–structure–function and developability.

A key challenge in biopharmaceutical informatics is correlating the "macroscopic" experimentally determined properties of a biologic with its "microscopic" sequence–structure features computed *in silico*. Uncovering these correlations can guide molecular sequence optimization strategies, proactively addressing potential obstacles in drug product development by predicting the performance of the final drug candidate in the streamlined platform processes used during development stages. This process necessitates combining data from standardized biophysical experiments with descriptors computed from molecular modeling and simulations in a common database. Various statistical and machine learning approaches can be employed to develop mathematical models that predict the solution behavior of mAbs based solely on their sequence–structure information, depending on the available data (Tomar et al., 2016; Jain et al., 2017b; Chiu et al., 2019; Hebditch and Warwicker, 2019; Lecerf et al., 2019; Raybould et al., 2019; Starr and Tessier, 2019; Kuroda and Tsumoto, 2020; Zhang et al., 2020). As a result, the interdisciplinary field of biopharmaceutical informatics aims to seamlessly integrate techniques from computational and experimental biophysics, information technology, and data science to provide data-driven inputs for the decision-making framework for all stages of biologic drug discovery and development.

3 Opportunities for computation at various stages of biotherapeutic discovery and early development

There are numerous opportunities to collaboratively apply computational and experimental tools to facilitate faster and more efficient drug engineering and development. In this review article, we present a diverse set of use cases at various stages of biotherapeutics discovery and development projects that could benefit with increased use of computation in synchrony with the experiments to demonstrate the practical feasibility of our vision. The major challenges faced at distinct stages of biotherapeutic discovery and early drug development are described in Table 1 along with potential computational opportunities to address them. The pros and cons of these computational opportunities are also presented in Table 1. It is important to note that the field has not matured uniformly across all stages of discovery and development cycles for biotherapeutics. For example, computational approaches to developability assessments and lead optimization (LO) are currently more advanced than *in silico* antibody discovery and *in silico* formulation development. Moreover, there are also opportunities to modify the workflows and transitions between the different discoveries and development stages in view of the rapidly growing capabilities of computation. These opportunities are described in the following sections.

3.1 Antigen optimization

The discovery of antibody-based biotherapeutics adheres to a stepwise approach once a target antigen or multiple antigens for simultaneous targeting in a multi-specific format have been identified. The initial phase entails producing enough target antigens to enable animal immunization, *in vitro* selection of antigen-specific antibodies, and functional activity characterization. However, some antigens exhibit favorable expression *in vivo* but encounter conformational stability and solubility issues *in vitro*, outside the cellular context (Qing et al.,

TABLE 1 Opportunities for the expanded use of computational approaches throughout the discovery and development process of biotherapeutics.

Process stage	Typical problems	Potential applications of computational approaches	Pros	Cons
In vitro synthesis of immunogens/antigens to generate corresponding antibodies	1. Availability of structural models for immunogens and accurate definition of epitope(s) of therapeutic interest 2. Aggregation tendency, protein insolubility, and reduced conformational stability may result in limited material availability for immunization experiments 3. Epitope(s) of therapeutic interest might not be immunodominant	1. Protein structure prediction and precise definition of epitope(s) of therapeutic interest 2. Sequence/structure-based optimization for improved solubility via APR disruption, supercharging; and increased conformational stability via residue scan can help improve quantity as well as quality of material needed for immunization 3. Strategies for disruption or masking of immunodominant but therapeutically irrelevant epitopes to improve chance of antibody binders to the therapeutically relevant epitopes	1. Protein structure is crucial for structure-based approaches to drug discovery defining epitopes of therapeutic interest. The emergence of AI-based protein structure prediction methods has enhanced the structural definition of immunogens in recent years 2. Judiciously selected mutations at single or multiple sites can significantly improve the availability of immunogen material in the laboratory	1. Confidence levels in different regions of the structure should be considered, as flexible regions are typically predicted with lower confidence levels 2. Defining the epitope(s) of therapeutic interest and avoiding mutations in and around them is important 3. Implementing site-directed mutagenesis of immunogens to improve material availability also requires a cultural shift among experimental scientists
Antibody generation	1. Animal immunizations can be time-consuming, expensive, and may yield inconsistent results 2. The lead antibody molecule identified through animal immunization may necessitate humanization and developability enhancements 3. Humanized mice and display technologies do not entirely capture the complete human immunome 4. Phage and yeast display technologies can quickly identify high-affinity binders, but these may require further optimization for developability	1. Generative AI can aid in designing antigen-specific and agnostic libraries with incorporated developability features 2. Virtual screening of antibody libraries against given antigen(s)/epitope(s), followed by docking and structure-based affinity enhancements 3. Utilizing computational methods to design phage and yeast display libraries for enhanced developability and/or affinity 4. Employing computational approaches to redesign antibody CDRs for altered specificities	1. Adopting computational methods can reduce timelines and costs associated with antibody discovery 2. Expanded druggable antigen space 3. Opportunities to explore a broader sequence diversity, thereby maximizing the odds for antibody discovery compared to conventional methods 4. Addressing developability during library design can help reduce time required for lead optimization	1. Emerging technology 2. Necessitates more extensive validation and experimental demonstration of its capabilities before routine project use 3. Requires a cultural shift from experimentally driven antibody discovery to computationally driven approaches
Hit selection and lead identification	1. Sequencing of identified hits 2. Epitope mapping of the hits to ensure the desired therapeutic effect in the absence of structural models for the antigen-antibody complex 3. Experimental evaluation of several hundreds of candidates for functionality and developability can be time and resource-intensive	1. Establishment of suitable sequencing pipelines 2. Computational prediction of epitopes and paratopes for epitope mapping purposes 3. In-silico evaluations of candidates for developability and manufacturability can facilitate the selection of developable hits and identification of lead candidate(s) with favorable developability characteristics 4. Development of digital twins for biophysical processes via computational biophysics and data science	1. Incorporation of computational assessments can aid in guiding hit selection for experimental testing 2. Proactive consideration of developability can help reduce costs and efforts to identify lead molecules 3. Opportunities to enhance our understanding of the connection between molecular sequence-structural properties and experimental outcomes	1. Greater availability of data is needed to connect 'microscopic' sequence-structural features of antibodies with the 'macroscopic' biophysical outcomes 2. Lack of digitization and digital transformation present significant challenges 3. A cultural shift from protecting experimental data to sharing it with computational scientists is required among discovery scientists
Lead optimization	Lead candidates may require humanization, affinity optimization, and elimination of physicochemical liabilities in the CDRs for enhanced developability	1. Structure-based modeling of the lead candidates can assist in their humanization, affinity maturation, and identification of potential sequence/structural motifs that may contribute to their physicochemical degradation. Access to this information can help direct protein engineering strategies for lead optimization 2. Assessment of the optimized lead candidates for their drug likeness	1. Computational guidance for lead optimization efforts can decrease timelines and costs 2. This aspect represents the most developed application of computational protein design in biotherapeutic drug discovery 3. Numerous well-developed computational solutions are available	1. There remains cultural resistance to the adoption of computational protein design for lead optimization among industrial scientists 2. Greater dissemination of successful case studies, where computational protein design makes a difference, is needed to raise awareness

(Continued on following page)

TABLE 1 (*Continued*) Opportunities for the expanded use of computational approaches throughout the discovery and development process of biotherapeutics.

Process stage	Typical problems	Potential applications of computational approaches	Pros	Cons
Early stage developability assessments	1. Assessing molecular stability and compatibility of drug candidates, identified during drug discovery, with platform processes utilized in drug development 2. Adapting to multiple product development goals such routes of administration and product presentations	1. Structure prediction of full length antibodies and novel formats 2. In-silico development of formulations 3. Employing multi-scale simulations to anticipate platform compatibility and evaluate molecular responses to stresses encountered during manufacturing, storage, and transportation 4. Utilization of predictive algorithms to determine suitable bioprocess conditions 5. Establishing digital twins for various facets of drug development	1. Developing full-length models of the drug substance can facilitate improved prediction of molecular origins of dominant degradation routes during manufacturing, storage, and shipping 2. Accelerating formulation process development and saving costs of drug development can be achieved through pH and buffer screening of antibody formulations via in-silico characterization of molecular integrity of the drug substance 3. Resource savings can be realized with the development of digital twins	1. Computationally intensive calculations 2. Need for improved correlations between experimental results and molecular simulations 3. Consistent availability of development data across different projects 4. Requirement for greater investments in the digitalization of drug development data

2022). Producing recombinant antigens can be particularly challenging for certain target classes, such as membrane proteins (G protein–coupled receptors and ion channels) (Bill et al., 2011). If antigen binding is impacted by the *in vitro* conformational stability and/or solubility of the antigen, then these issues may hinder the entire antibody discovery strategy and functional validation of the antibody hits.

Computational methods can aid in the redesign of antigens with enhanced conformational stability and solubility when a threedimensional crystal structure or model is available. Bioinformatic tools can enable crystal structure refinement, modeling of breaks and gaps, loop modeling, energy minimization and molecular dynamics simulations to support antigen redesign. When the crystal structure of an antigen is unavailable, protein structure prediction techniques can often estimate it (Nimrod et al., 2018). For example, homology-based structure modeling can be employed using crystal homologs. A sequence identity of at least 30% between the protein of interest and its crystal homologue is typically sufficient for structure generation through homology modeling. However, some novel targets may not have homologs with existing crystal structures. This can be due to the inherent difficulty in obtaining crystal structures of membrane-associated proteins, which often have poor solubility. Membrane proteins represent a significant class of drug targets, and the discovery pipeline frequently proceeds without knowledge of the antigen structure. In such challenging cases, recent groundbreaking advances in *de novo* protein structure prediction techniques have achieved remarkable success and accuracy by leveraging machine learning and deep learning algorithms (AlQuraishi, 2019; Gao et al., 2020; Pereira et al., 2021; Jones and Thornton, 2022). Deep learning–based structure prediction methods, such as AlphaFold2 and RoseTTAFold, combined with physical modeling, have outperformed numerous conventional approaches (Baek et al., 2021; Jumper et al., 2021; Pereira et al., 2021; Jones and Thornton, 2022). Understanding of the antigen's three-dimensional structure can be crucial for accurately assessing its stability and solubility, computationally. This knowledge can also help enhance solubility without sacrificing stability and functional activity, allowing for the extraction of crystal structures, and facilitating experimental assays that measure target binding. Care should be taken, however, to minimize the impact of such mutations on the overall molecular structure of the target antigen and preserve its potential to generate adequate immune response to epitopes of therapeutic interest. Bioinformatics can also support rational strategies to immunize only therapeutically relevant epitopes on the antigen surface. This means epitopes that may be immune-dominant but are of no therapeutic interest or relevance can be either eliminated or masked to facilitate the immunization of the desired epitopes of therapeutic importance.

3.2 Antibody generation

Immunization strategies have long been employed to generate high-affinity antibodies, using previously expressed and purified antigens to establish immune reactions in animals (typically laboratory mice, humanized/transgenic mice, or other animals like chickens, rabbits, or cows). Antibody binding to specific antigens can be obtained through techniques such as hybridoma (Koehler and Milstein, 1975), single B cells (Yu et al., 2008), or screening natural and/or synthetic antibody libraries via display technologies using phage or yeast (Benatuil et al., 2010; Chen and Sidhu, 2014; Alfaleh et al., 2020; Gray A. et al., 2020; Nagano and Tsutsumi, 2021; Ledsgaard et al., 2022; Valldorf et al., 2022). Promising candidates are selected and validated using antigen-binding assays that align with the research target profile. Currently used methods in the biopharmaceutical industry for antibody generation are almost exclusively experimental, and depending on the techniques used, it can take several months before an initial set of antibody-based binders is available for further investigation and lead identification. Fully synthetic human antibody libraries containing Fabs chosen for their biophysically favorable development characteristics have been developed using experimental means (Valldorf et al., 2022). Special emphasis has been placed on selecting molecules with enhanced chemical, conformational, and colloidal stabilities (Tiller et al., 2013). The availability of such libraries can significantly help accelerate the discovery of antibody-based biotherapeutics by pre-paying for developability.

The concept of optimized antibody libraries for generating developable antibodies can be integrated with *de novo* computational databases containing an immense variety of human-like light- and heavy-chain combinations (Pan and Kortemme, 2021; Akbar et al., 2022). Targeted mutations at specific sequence positions [e.g., complementarity-determining regions (CDRs)] in the antibody sequences could further broaden the library, either to recognize different antigens or to optimize binding affinity toward a specific antigen (Ledsgaard et al., 2022). Recently, a generative adversarial network was successfully employed to create a diverse library of novel antibodies that emulate somatically hypermutated human repertoire responses (Amimeur et al., 2020). This *in silico* method further revealed residue diversity throughout the variable region, which could be useful for additional computational tools like CDR redesign. CDR redesign utilizes a highly developable antibody framework and modifies the original CDRs, or paratope, to recognize a new antigen. In recent years, noteworthy progress has been made in designing not only thermodynamically stable but also biologically functional antibodies (Baran et al., 2017).

Computational technologies, initially developed for small-molecule drug discovery, can also be applied to antibody-based drug discovery. Once fully developed and implemented, these computational methods will provide additional means to generate diverse antibody binders against a target antigen. These methods will not only help reduce animal use in biologic drug discovery but also decrease reliance on experimental trial and error for finding initial hits. Initial case studies describing such methods are beginning to emerge in the literature (bioRxiv.org for preprints) (Sever et al., 2019; Wilman et al., 2022). Additionally, it becomes feasible to find potential binders to difficult targets, thereby expanding the druggable target space for antibody-based biotherapeutics.

Figure 2 provides an overall conceptual roadmap for Discovery of antibodies in silico (DAbI). The proposed roadmap encompasses three major parts where each part can have multiple stages depending upon the project in hand. In the first part, the key is to use different computational algorithms to generate medicine-like human antibody sequence libraries *in silico*. These libraries can be either antigen-specific or antigen-agnostic and are of orthogonal utilities. For example, creation of antigen- or epitope-specific antibody libraries via machine learning can help us achieve early success in each antibody discovery project by facilitating a focused path to the discovery of lead candidates toward the antigen and support the therapeutic concept. A biological analog of such libraries shall be the sequence repertoires obtained from immunized animals, hybridomas, or the results obtained by panning the display libraries against a specific antigen. However, such libraries have to be generated repeatedly for each different antigen or epitope. Antigen-/epitope-agnostic libraries on the other hand can be incredibly useful toward supporting multiple drug discovery projects simultaneously. Such libraries can be thought of as naive B-cell repertoires obtained from humanized animals prior to immunization with specific antigens. The computationally generated naive antibody repertoires can potentially capture greater sequence diversities than those feasible from humanized animals, display technologies, or observable B-cell repertoires. Within a discovery organization, such libraries have to be constructed only once and be potentially useful toward pre-computation of binders for all the targets of interest to the organization. These pre-computed antibody binder libraries can potentially accelerate early antibody discovery projects because now the discovery process does not have to wait for availability of target reagent in the laboratory. Therefore, such libraries can be particularly useful toward difficult to express and purify targets such as membrane proteins. Irrespective of the purpose of *in silico* generated antibody libraries, it is important to generate structural models of (at least) the variable regions of the antibodies sampled from these libraries. The generated structures can then be used for assessing their medicine-likeness and developability. Early elimination of non–medicine-like antibodies from such libraries can improve their utility and differentiate them from those generated using the experimental means solely. The structural models can also be used for predicting antibody paratopes. Many computational methods are currently available for the structural prediction of antibodies. The major challenges in this field include prediction of HCDR3 conformation and pairing of the light- and heavy-chain variable regions (Fernández-Quintero et al., 2023).

In addition to the design of the *in silico* antibody libraries, currently available computational methods also provide an opportunity to design single or a few human antibody variable regions against specific antigen epitopes *de novo* (Chowdhury et al., 2018; Nimrod et al., 2018). The design process can also commence with a structural model of an antigen:antibody (Ag:Ab) complex, generated using molecular docking of the antigen and antibody structures (Nimrod et al., 2018). Subsequently, the affinity of the antigen toward the antibody can be either altered by randomly introducing sequence variations or selectively re-designing interfaces using structure-based approaches (Nimrod et al., 2018). For example, interfacial residues in the Ag:Ab complexes that significantly contribute to their stability and instability can be identified through computational alanine (Ala) scanning. In the following step, the identified residue positions can be scanned for mutations that either increase or decrease the stability of the Ag:Ab complex and enhance or reduce the affinity of the antibody toward its cognate antigen (Sheng et al., 2022), depending on the project requirements. Another appealing alternative for rational antibody design involves hotspot grafting with CDR loop swapping, which only requires information about interactions with the antigen (Liu et al., 2017).

The goal of epitope-driven antibody generation is to design an antibody variable region with a paratope that complements the given epitope. Since CDRs make up most of the paratope, initial efforts to design epitope-specific antibodies have focused on *ab initio* CDR redesign and modeling. OptCDR (Pantazes and Maranas, 2010), used in conjunction with Rosetta Antibody Modeler, generates epitope-specific high-affinity CDRs by selecting the most feasible canonical loop conformations followed by iterative model optimization and improvements in binding energy. This method enables the generation of a focused library of antibody binders, quite like hit sequences obtained from experiments. OptCDR was later optimized (OptMAVEn) to consider the entire fragment variable (Fv) region rather than just CDRs as the starting point for generating antibody binders (Li et al., 2014), allowing for the incorporation of humanness at the antibody generation stage through careful selection of human framework region residues. Further advances have incorporated MD simulations for accurate evaluation of

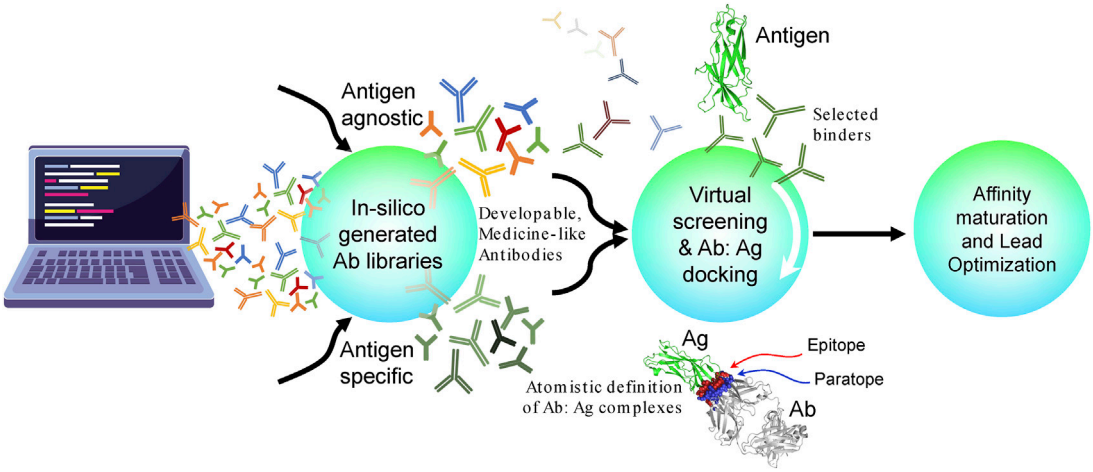

FIGURE 2
Conceptual roadmap for the discovery of antibodies *in silico* (DAbI). This conceptual roadmap can be divided into three major parts that can be developed either independently or in synchrony. The first part focuses on the *in silico* generation of medicine-like, antigen-agnostic, or specific antibody sequence libraries. Several machine learning algorithms are currently being developed to facilitate the *in silico* generation of antibodies. In the second part, these *in silico* generated antibodies and their structural models can be used to screen against a given antigen or an epitope on an antigen via virtual screening, docking, or other computational chemistry-based algorithms. Conversely, a large set of potential antigens can also be pre-screened against the antibody libraries using the same computational technologies. In both cases, the goal is to obtain atomistic definitions of putative antibody–antigen complexes. At this stage, it is preferable to virtually screen a larger number of antibodies (e.g., 1–10 million) and then select a much smaller number (e.g., 10–100) for docking simulations. This will help speed up the calculations and save computational resources. It is also important to quantitatively assess the quality of modeled antibody–antigen complexes by comparing them against crystal structures of other antigen–antibody complexes. A third option is to convert the whole or portions of the *in silico* generated antibody libraries into molecular libraries suitable for phage or yeast display and then pan them against a diverse panel of desired antigens. In the third part, the structural models of the putative antibody–antigen complexes obtained previously can be used to identify potential lead antibody candidates and modify their binding affinities to the desired levels via single- or multi-residue mutations in the paratope regions through computational protein design. These structural models can also be used to impart cross-reactivity to homologous antigens from other non-human species and/or to even create surrogate antibodies. Care should be taken to avoid introducing residues susceptible to physicochemical degradation and therefore reducing the developability of the lead candidates. It is important to note that DAbI will require changing the discovery workflows because it is pre-paying for developability and may therefore require significantly reduced effort during lead optimization (LO).

binding energetics (Chowdhury et al., 2018). A one-to-one residue matching method called epitoping, which starts from antibody structures with basic shape complementarity, was developed to obtain an accurate epitope–paratope binding match (Nimrod et al., 2018). Although this process requires a pre-identified approximate match, it can be considered for lead optimization to improve binding.

Recent advancements in generative deep learning and the availability of approximately 2,000 solved crystal structures of the antibody–antigen complexes have opened possibilities for structure-based *de novo* antibody generation. A proof-of-concept study utilizing a variational autoencoder (VAE)–based generative algorithm demonstrated the capability to directly generate 3D coordinates of antibody backbones that complement a specific epitope (Eguchi et al., 2022). Additionally, another deep learning algorithm was developed to learn the 3D features of antibodies from 1D sequences, enabling the generation of antibody sequences with desired structural characteristics (Akbar et al., 2022). Although the proof-of-concept study primarily aimed to achieve high-affinity binder antibody sequences for a given epitope, the method holds potential for encoding additional features, allowing the model to be tailored to produce highly developable sequences. As stated previously, generation of epitope-specific antibodies or libraries thereof has immediate applications for individual drug discovery projects, since the knowledge of epitopes is often required for defining novel therapeutic concepts.

3.3 Early screening for developability of *in silico* generated antibody libraries

Once the *in silico* antibody sequence libraries have been generated, it is worth assessing the generated antibody sequences for developability and advancing highly developable sequences to further stages of discovery. The developability assessment tools to be employed here can be ported over easily from those used at the hit selection and lead identification, lead optimization, and early development stages in the conventional biotherapeutic discovery and development workflows.

Lipinski's "rule-of-five" revolutionized the discovery and development of small molecules by providing guidelines for improving their solubility and permeability (Lipinski, 2000). However, establishing similar rules for new biological entities (NBEs) has proven challenging due to their complex structures. In response, researchers have turned to biophysical evaluations and computational approaches to better understand these entities and overcome inherent obstacles. Biophysical evaluations of clinical-stage antibodies have contributed to the empirical definition of analogous boundaries, offering valuable insights for NBE development (Jain et al., 2017b; Raybould et al., 2019; Jain et al., 2023). Additionally, marketed antibodies have been profiled using calculated physicochemical descriptors, in an approach known as the DEvelopability Navigator *In Silico* (DENIS) (Ahmed et al., 2021;

Licari et al., 2022). These advances have significantly contributed to our understanding of NBEs and their development processes.

Biotherapeutics can undergo various levels of conformational changes over time, which presents significant challenges regarding conformational stability during manufacturing, shipping, and storage. This is because the environment of a biotherapeutic drug candidate can influence its structure, highlighting the importance of understanding these complex molecules in more detail. To address this, biophysical analysis employs a variety of techniques, such as thermodynamic, spectroscopic, and hydrodynamic methods, for characterizing protein-based drug candidates. These techniques are routinely used during the discovery phase to guide the identification and characterization of the lead drug candidates. Some properties commonly assessed during biophysical analysis include post-translational modifications (e.g., glycosylation, deamidation, isomerization, oxidation, and fragmentation), aggregation, self-association, hydrophobicity, molecule pI, and viscosity for high-concentration liquid formulations. While these techniques are well established, they can be time- and resource-consuming and demand expert knowledge and advanced instrumentation. This has driven researchers to seek more efficient and accessible methods for obtaining critical data. *In silico* tools can predict the intrinsic biophysical properties of drug candidates along with identifying their degradation routes, whose knowledge is important for establishing appropriate formulation strategies. These tools demonstrate significant relationships between the Fv domain sequences and physicochemical properties that define antibody developability. For example, post-translational modification sites, such as deamidation, aspartate isomerization, oxidation, and fragmentation can be identified using computational approaches (Irudayanathan et al., 2022; Vatsa, 2022). Similarly, hydrophobic interaction chromatography (HIC) retention times have been successfully correlated with sequence and structure features through diverse methods such as quantitative structure–property relationship (QSPR) modeling and machine learning (Jain et al., 2017a; Jetha et al., 2018; Karlberg et al., 2020). Although solution and colloidal state properties are challenging to predict due to multiple influencing factors, computational tools like SOLpro and PROSO II have demonstrated their ability to predict solubility upon expression with an accuracy of ~75% (Magnan et al., 2009; Smialowski et al., 2012). The isoelectric point (pI) is a crucial physicochemical property for mAbs. It has been associated with specific developability aspects such as thermostability, viscosity, and resistance to high molecular weight species formation at low pH. Tools like MassLynx, Vector NTI, and EMBOSS (Rice et al., 2000) calculate pI based on sequence data, achieving results within a 15% range of experimentally determined values (Goyon et al., 2017). Tools that predict the pI based on protein structure can provide a more accurate result, since the underlying residue p*K*a values are calculated by considering the residual microenvironments. Viscosity is also a critical factor in the colloidal stability of biologics and is influenced by electrostatics and hydrophobicity, which are in turn determined by the Fv sequence and structure. The *in silico* tool, spatial charge map (SCM), can identify highly viscous antibodies based on the mAb structure (Agrawal et al., 2015). Biomolecule aggregation is related to sequence and structural characteristics, such as the presence of aggregation-prone regions, hydrophobicity (Münch and Bertolotti, 2010), electrostatics (Buell et al., 2013), and dipole moments (Tartaglia et al., 2004), which enable both sequence- and structure-based computational predictions. Various *in silico* tools play a significant role in guiding mAb candidate design with high colloidal stability by predicting the impact of single or multiple amino acid exchanges on aggregation propensity. Alternative tools such as TANGO, PASTA, FoldAmyloid, SALSA, and AggreRATE-Pred can detect aggregation-prone regions based on the physicochemical principles of secondary structure elements, particularly the ability to form intermolecular cross-β-structures (Fernandez-Escamilla et al., 2004; Trovato et al., 2007; Zibaee et al., 2007; Garbuzynskiy et al., 2010; Walsh et al., 2014; Rawat et al., 2019). In summary, these *in silico* tools can effectively predict various biophysical properties of biotherapeutics. Their high-throughput capabilities make them particularly attractive for biophysical assessments during various stages of the drug discovery process.

3.4 Hit selection and lead identification

Following the production of antigen-binding antibodies through immunized animals, hybridoma cells, or phage and yeast display techniques, the variable regions of the antibodies are sequenced, and the binders are validated in the conventional workflows adapted by the biopharmaceutical industry. The immunization methods, strength and diversity of the immune responses, and sequencing technologies used can yield numerous unique hits, particularly via B-cell cloning and repertoire sequencing. Subsequently, these diverse hits must be prioritized to identify the most promising lead candidates, necessitating extensive resources to experimentally test each hit and confirm antigen binding.

Several bioinformatic techniques can aid in prioritizing and selecting hits for *in vitro* confirmation of antigen binding and lead identification (Figure 3). A common strategy involves clustering hits into high-, medium-, and low-binding bins based on the initial estimates, analyzing each bin for heavy- and light-chain germline diversity, and then examining CDR diversity to select multiple representatives from each germline pair in each bin for experimental testing. Alternatively, hits can be binned based on the germline pair and CDR diversity, with selections made according to their estimated antigen binding. At this stage of hit selection, developability aspects can also be considered using computational tools introduced in the previous section. In a basic application, heavy- (HC) and light-chain (LC) sequences of hits can be scored based on the presence of potential chemical degradation motifs, aggregation-prone regions (APRs), and T-cell immune epitopes present in or overlapping with the CDRs of the heavy and light chains. The scoring schemes can be further optimized by assigning different weights based on which CDRs contain these motifs and whether they are in the Vernier zones or middle of the CDRs.

Structure-based approaches require accurate three-dimensional antibody fold information, typically generated via homology modeling. This process includes 1) identifying a high-identity structural template for framework (FW) regions, 2) loop

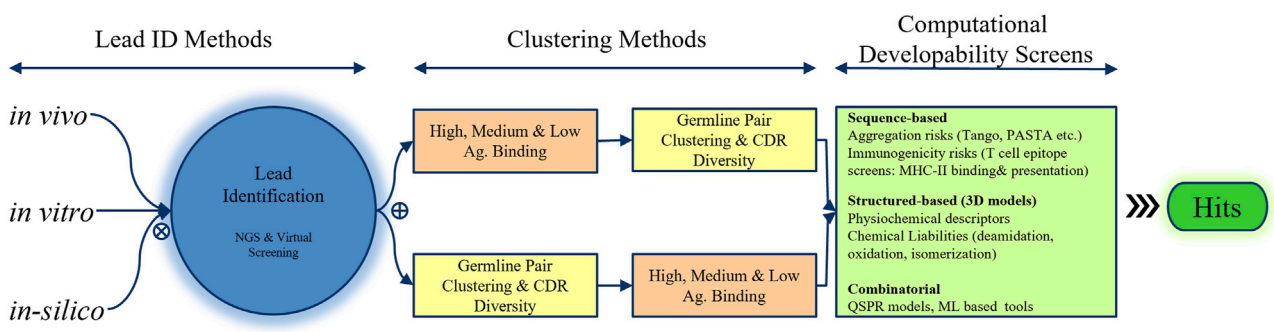

FIGURE 3
Integration of *in vivo*, *in vitro*, and *in silico* approaches for hit selection in the discovery phase of the pharmaceutical industry. Next-generation screening and virtual screening methods are employed to identify promising leads, which are then prioritized using clustering techniques based on 1) antigen binding and 2) a combination of germline pair clustering and CDR diversity. Finally, computational developability screens that analyze the amino acid sequence, structure, and combinatorial methods such as QSPR or machine learning are performed to select the most promising hits.

modeling of LCDR1-3 and HCDR1-2 using canonical loop conformations, 3) HCDR3 loop modeling and optimization of the orientation of heavy-chain variable region (VH) and light-chain variable region (VL), and 4) sidechain packing and refinement. The key challenges involve obtaining high-resolution templates with optimal VH-VL orientations and accurately modeling loops, particularly the HCDR3 loop. Recent progress in Fv structure modeling has led to advanced tools, such as RosettaAntibody (Weitzner et al., 2017; Adolf-Bryfogle et al., 2018; Schoeder et al., 2021), AbPredict2 (Lapidoth et al., 2018), ABodyBuilder (Leem et al., 2016), LYRA (Klausen et al., 2015), MoFvAb (Bujotzek et al., 2015), and Kotai Antibody Builder (Yamashita et al., 2014), which demonstrate high performance in the AMA-II benchmark test. Commercial packages like Molecular Operating Environment (MOE) and BioLuminate are popular for high-throughput full-length Fv structure modeling. A detailed discussion of recent advancements in Fv structure modeling tools can be found in focused reviews (Fernández-Quintero et al., 2023). Additionally, tools like FREAD, H3LoopPred, SPHINX, MODELER, PLOP, SCWRL, BetaSCPWeb, Chothia canonical assignment, and SCALOP have significantly contributed to full-length Fv region three-dimensional structure modeling. Tools such as TopModel efficiently examine the structure for cis-amide bonds, D-amino acids, and steric clashes, allowing for rapid evaluation of model quality and accuracy prior to conducting further analysis (Norman et al., 2019; Wilman et al., 2022; Fernández-Quintero et al., 2023). The generated three-dimensional structural models of all or a subset of hits can then be analyzed regarding their physicochemical descriptors, such as pI, charge, dipole moment, and solvent-exposed hydrophobic and ionic patches. These physicochemical properties have been demonstrated to potentially influence the chemical, conformational, colloidal, and physical stabilities of antibodies, and consequently their developability. In subsequent studies, a few of the best hits are rigorously tested in the laboratory for biological function, cross-reactivity across species, non-specific binding, and pharmacological indicators, such as serum stability. This process results in the identification of one or more lead candidates.

3.5 Virtual screening and docking as potential alternatives to *in vitro* hit selection and lead identification

Identification of potential binders through immunization campaigns can be accomplished using bioinformatics tools for paratope and epitope prediction, followed by rapid virtual screening, as outlined in Part 2 of the *in silico* roadmap, we call DAbI (Figure 2). This approach involves three-dimensional structure modeling of a diverse antibody sequence library and screening it against a given antigen by taking advantage of the shape and charge complementarity between the epitopes and paratopes. The antibody libraries to be screened can be endowed with the biophysical characteristics desired from a developability perspective as described previously.

Small-molecule drug discovery has successfully employed virtual screening to identify binders from a library of drug candidates (Gorgulla et al., 2020; Maia et al., 2020; Yan et al., 2020). Typically, millions of small-molecule drug candidates undergo structural and energetic screening processes through docking, pharmacophore-, or ligand-based approaches. Modern techniques involving computer vision, image-based, and geometric learning–based algorithms have reached advanced stages of validation and are now well established among the marketed small-molecule drugs designed using *in silico* methods (Eguida and Rognan, 2020; Gorgulla et al., 2020; Yan et al., 2020). Similarly, a curated and modeled antibody library may be treated as a potential set of drugs to be screened against a given antigen. However, directly applying these techniques may not be feasible due to the significant structural and functional differences between small-molecule drugs and large antibodies, with size (molecular weights, 500–1,000 Da versus approximately 25,000 Da for the Fv) being a primary concern even when considering only the Fv regions. Additionally, given the estimated theoretical diversity of B-cell repertoire (BCR) based on V(D)J recombination, which is about 10^{13}–10^{20} unique sequences, it is crucial to consider large antibody libraries to allow screening over a highly diverse sample space of paratopes.

Hypothetically speaking, we consider an antibody library of 1 million Fv sequences and assume a screening time of 1 min per Fv for a given antigen, the total runtime would amount to approximately 695 days (close to 23 months) for screening a single antigen against the library, which consists of only a small fraction of BCR diversity. Currently existing docking methods have runtimes of several minutes per complex. On the bright side, rapid virtual screening may not necessarily require rigorous energy-based binding evaluations employed in modern docking programs. Sacrificing the accuracy afforded by pose refinement can allow for greater speed in the screening process. Consequently, novel techniques have to be developed to enable the screening of large antibody libraries by considering the key aspects of the structural and chemical complementarity of the antigen:antibody interfaces and ensuring high-throughput rapid execution. An ideal *in silico* antibody virtual screening process could narrow down the potential binding hits to the order of 10^1–10^2, meaning that virtual screening would enable identifying binders at least as accurately as about one in a thousand to a few thousand sequences from the library, significantly impacting the discovery pipeline.

While *in silico* virtual screening does not replicate the generation of antigen binders via experimental methods in terms of binding affinity or functional efficacy, it can allow for comprehensive screening of the antibody library to identify all possible structural matches of epitopes and paratopes. Iterative refinement of these matches can help discover antibody binders to a given antigen with a diverse set of binding affinities and therefore suitable for antagonist as well as agonist function. Novel techniques, such as image-based and graph-based deep learning algorithms, have been proposed for identifying complementary paratope/epitope interfaces. These approaches can be further accelerated through pre-identified or predicted paratope and epitope information (Gainza et al., 2020; Pittala and Bailey-Kellogg, 2020; Akbar et al., 2021; Ripoll et al., 2021). Schneider et al. (2021) proposed a structure-based virtual screening method using voxel representation of the interfacing surface atom groups in their screening method called Deep Learning for AntiBodies (DLAB), adapted and extended from its small-molecule counterpart (Imrie et al., 2018). Recently proposed image fingerprinting–based approaches, with analogous applications in small molecules, show promising potential for protein interface matching and could be further expanded to predict paratope/epitope binders for hit selection (Gainza et al., 2020; Ripoll et al., 2021). More recently, a geometric deep learning method called ScanNet has been introduced to predict protein–protein and protein–antibody binding interfaces through geometric deep learning of three-dimensional structural features (Tubiana et al., 2022). Moreover, some of the paratope/epitope prediction methods involving deep learning of interfacial interactions may be extrapolated to interface screening and predicting binders.

The *in silico* virtual screening of antibodies against a given antigen can also borrow techniques such as fragment-based drug design (Sormanni et al., 2015; Sormanni et al., 2018) and pharmacophore modeling from the realm of small-molecule drug discovery. By facilitating the identification of binding sites, improving antibody–antigen docking, and enabling more accurate structure-based virtual screening, these methods can accelerate the development of novel therapeutic antibodies and enhance our ability to target a wider range of diseases and conditions.

Recent molecular docking protocols feature highly robust, energy-based scoring functions for evaluating and ranking protein–protein or protein–antibody binding partners. This offers a suitable toolkit for further optimization of hits identified through virtual screening of target antigens against an antibody library. Docking methods have demonstrated accurate prediction of protein-binding interfaces; however, speed has not been a priority for molecular docking programs. Although the current speed of implementation poses a bottleneck, rapid advancements in the field of protein–protein docking have spurred the development of new methods utilizing advanced machine learning algorithms and hybrid physics and learning-based technologies, promising faster docking methods soon. Moreover, such advancements may bridge the gap between virtual screening and docking, further accelerating *in silico* antibody screening, hit selection, and lead identification processes altogether.

Antibody–antigen docking has often been considered with paratope/epitope prediction and improving CDR modeling accuracy. SnugDock combines docking with accurate modeling prediction of the paratope (CDR loop construction), where the Rosetta Antibody Modeler operates alongside the docking protocol, iteratively improving docking and model prediction (Sircar and Gray, 2010; Jeliazkov et al., 2021). Additionally, methods employing more rigorous energy-defined constructs to evaluate multiple docking poses through the MM-GBSA (molecular mechanics—generalized Born solvent accessibility) method have shown promising outcomes (Shimba et al., 2016). Information-driven docking methods depend on a set of data to reduce the number of decoys, thus saving prediction time. Interface prediction-based methods, such as Antibody i-patch and EpiPred, focus on refining docking poses through paratope/epitope interface prediction (Krawczyk et al., 2013) By contrast, proABC adopts a more site-directed approach driven by the interface (paratope) (Olimpieri et al., 2013; Krawczyk et al., 2014). Advances in machine learning and deep learning algorithms have significantly contributed to enhancing docking prediction methods.

Other widely employed programs such as ClusPro, LightDock, ZDOCK, and HADDOCK, coupled with CDR and binding epitope information for directed/biased docking approaches, have shown promising results, with HADDOCK demonstrating notable performance improvement (Ambrosetti et al., 2020a). Pro-ABC-2, another information-driven docking approach and an updated version of Pro-ABC, utilizes deep learning convolutional neural networks (CNNs) for paratope prediction to assist in docking (Ambrosetti et al., 2020b). Such information-driven methods may also be applicable in pipelines using commercial docking techniques offered by MOE from Chemical Computing Group, PIPER from Schrodinger, and others with additional efforts.

Several other methods that employ deep learning through CNNs, recurrent neural networks (RNNs), or graph-based learning have demonstrated promise in predicting binding interfaces, consequently improving docking accuracy (Liberis et al., 2018; Deac et al., 2019; Pittala and Bailey-Kellogg, 2020; Lu et al., 2021; Myung et al., 2021; Vecchio et al., 2021; Davila et al., 2022). Additionally, research groups have been exploring the exceptional modeling performance of AlphaFold2 in docking

prediction. The accelerated advancements in AI related to AlphaFold and other docking methods offer significant potential for the development of faster and more accurate docking programs in the future.

3.6 *In silico* affinity maturation of lead candidates

In a conventional discovery workflow, the lead candidates identified may have to be optimized for affinity, cross-reactivity, and developability. Among these, the focus is often on developability of the lead candidates. By contrast, DAbI may yield developable lead candidates already since the sequence and structural features that support good developability are already included in the library design (Part 1 of DAbI, see Figure 2). Depending on the library choice (antigen-agnostic or antigen-specific), the *in silico* generated lead candidate may have to be optimized for binding affinity and any residual physicochemical developability issues, particularly from the CDRs. For these reasons, the third part of our conceptual roadmap, DAbI (Figure 2), envisages an ability to adjust the binding affinity as per the project requirements. Depending upon the novel therapeutic concept (NTC), both enhancement (affinity maturation) and decrease (affinity de-maturation) in binding affinities may be required. However, affinity maturation may be required more often than de-maturation, particularly when the lead antibody binders have been derived from antigen-agnostic libraries. In our conceptual roadmap, both affinity maturation and de-maturation begin with a structural representation of the atomic interaction between two proteins, namely, the antigen and the antibody. The methodology's reliability depends on accurately analyzing the interacting sites. Therefore, co-crystallized antibody–antigen complexes are typically preferred over structure-based homology models or AI predictions, which may lead to less reliable results if CDRs are not precisely modeled. The *in silico* affinity maturation relies on accurate molecular interactions for free energy or MM-GBSA–based calculations (Comeau et al., 2023; Thorsteinson et al., 2023), highlighting the importance of improving antibody–antigen complex predictions and the implicit incorporation of multiple conformational ensembles to enhance the effectiveness of *in silico* calculations and optimize library design. Despite this limitation, these methods have been already applied to predicted antibody–antigen complexes (Rangel et al., 2022), facilitating the generation of *in silico* affinity maturation libraries (Conti et al., 2022; Thorsteinson et al., 2023).

The *in silico* scanning of the individual paratope residues yields potential mutations and estimates of the corresponding free energy changes in binding to the target. The subsequent challenge involves designing a combinatorial assembly of these mutations into a library suitable for phage/yeast display. This is because the *in silico* affinity maturation often involves computationally expensive calculations that tend to be more accurate at identifying the single point mutations rather than combinations thereof (Comeau et al., 2023; Thorsteinson et al., 2023). The physical display libraries built using computational guidance can be used to pan combinatorial mutations. Therefore, this part of DAbI requires an understanding of the limitations associated with the library size and panning methodology (Tsumoto and Kuroda, 2022). It is also in consonance with the spirit of biopharmaceutical informatics which calls for taking advantage of the strengths of computation and experiments in a synergistic manner. When combined with library technologies like phage display, computational tools have proven particularly powerful in guiding the design of affinity maturation libraries (Tiller et al., 2017; Nelson et al., 2018; Wang et al., 2018; Thorsteinson et al., 2023). Incorporation of additional considerations along with the binding affinity can help narrow down the mutations for experimental testing and therefore the size of the display libraries. At this stage, the mutations that enhance specificity, humanness, and CDR germlining along with developability can be considered by incorporating relevant physicochemical properties and stability criteria (Khan et al., 2023; Svilenov et al., 2023). Consequently, the selection of lead antibody candidates with high binding affinity and favorable biophysical properties can be achieved simultaneously. In-house, we successfully improved binding affinities of the antibody drug candidates 10- to 1,000-fold in multiple proprietary projects using this strategy.

Several studies have demonstrated the computational design of functional antibodies using multiple structural models supported by statistical or machine learning models (Nimrod et al., 2018; Liu et al., 2019; Amimeur et al., 2020). Upon selecting an initial antibody scaffold, mutations to enhance complementarity with a given epitope can be designed to obtain specific antibody binders to an antigen. For example, the generative adversarial network (GAN) model was trained on over 400,000 light- and heavy-chain human antibody sequences to learn the rules of human antibody formation (Amimeur et al., 2020). The resulting model outperforms common *in silico* techniques, generating diverse libraries of novel antibodies mimicking somatically hypermutated human repertoire responses. Through transfer learning, the GAN can generate molecules with improved stability, developability, lower predicted major histocompatibility complex class II binding, and specific CDR characteristics. In-house, we could independently train the GAN on a much smaller set of approximately 31,500 paired antibody sequences belonging to the VH3-VK1 germline pair and format them as single chain variable regions (ScFvs). These sequences were selected based on their high percent humanness, low incidence of chemical liabilities in the CDRs, and high medicine-likeness. The in-house developed GAN model was then used to generate 100,000 unique antibody ScFv sequences and a small yet highly diverse subset of them was produced in the laboratory as immunoglobulin G1K (IgG1K) antibodies. The initial experimental characterization showed that most of the generated antibodies showed desirable attributes for expression, purification, thermal stability, and colloidal stabilities that compare favorably with those of trastuzumab, a biotherapeutic well known for its good developability profile (unpublished results). In summary, these *in silico* approaches enable the control of pharmaceutical properties for antibodies, potentially offering a more rapid and cost-effective screening, docking, and binding affinity maturation against a given target antigen.

3.7 Humanization and optimization of lead candidates

During the conventional discovery workflow, lead optimization (LO) is carried out as soon as one or more lead candidates have been identified and revalidated for function. The Fv regions may require humanization if the lead molecule is from a non-human source, the removal of post-translational modification (PTM) sites, optimization of affinity, and ideally, improvement of developability (Figure 4). When all parts of DAbI are fully enabled, time and efforts required for LO may be significantly reduced, if not eliminated completely as stated earlier. However, for now, humanization and optimization of the functional lead candidates remain an integral part of biotherapeutic drug discovery. The following describes how computation can support every aspect of the LO process for therapeutic antibodies.

Humanization optimizes the amino acid sequence of non-human Fv regions, decreasing immunogenicity and anti-drug antibodies (ADAs) (Roguska et al., 1994; Townsend et al., 2015). Computational protein design methods can efficiently increase antibody humanness while maintaining structural stability (Choi et al., 2015). State-of-the-art software like MOE (ULC, 2021) enables CDR grafting and humanness optimization through *in silico* calculations (Abhinandan and Martin, 2007; Lazar et al., 2007; Gao et al., 2013; Seeliger, 2013; Olimpieri et al., 2015; Choi et al., 2017; Kuroda and Tsumoto, 2020). Bioinformatic studies have also revealed structural differences between the lambda (VL) and kappa (VK) isotypes, which must be considered during re-engineering (van der Kant et al., 2019). Structure-guided approaches can aid in enhancing the biophysical properties of a therapeutic mAb by transitioning from a problematic lambda framework (FWR) region to a more stable kappa FWR (Lehmann et al., 2015).

The humanized sequences progress to liability engineering campaigns. Pre-formulation assessments, forced degradation studies, and *in silico* evaluations are incorporated into the engineering design plan. Phage display or other screening technologies can be employed to screen a large panel of variants. *In silico* tools monitor and guide the redesign of candidates' individual liabilities (see Figure 4), and medicine-likeness can be estimated by comparing molecular characteristics with marketed antibodies (Ahmed et al., 2021).

Computational tools have successfully guided antibody optimization campaigns, improving solubility, viscosity, self-association, colloidal stability, and binding specificity (Yadav et al., 2011, 2012; Nichols et al., 2015; Kumar et al., 2018b; Shan et al., 2018; Zhang et al., 2018; Navarro and Ventura, 2019; Sakhnini et al., 2019; Bauer et al., 2020). *In silico*–guided LO campaigns have demonstrated single amino acid residue exchanges that can improve multiple chemistry, manufacturing, and control (CMC) properties, such as expression titer, yield, purity, and colloidal stability (Bauer et al., 2020). A case study enhanced antibody developability using a multi-stage approach, starting with *in silico* screening for mutations addressing liabilities while preserving thermodynamic stability, followed by production and characterization of stable candidates (Sakhnini et al., 2019). An alternative hybrid method combined computational and experimental alanine scans to identify CDR positions for mutagenesis, maintaining antigen binding and creating antibody libraries (Tiller et al., 2017). Structure-based computational designs have been effectively employed to improve the affinity and specificity of therapeutic antibodies by pinpointing the key residues in the paratope for site-directed single, double, or even triple mutations (Kiyoshi et al., 2014; Grossman et al., 2016; Kumar et al., 2018b; Chiba et al., 2020). Computational methods offer conformational stability predictions for humanization or LO (Dehouck et al., 2011; Baets et al., 2015; Folkman et al., 2016; Quan et al., 2016; Pandurangan et al., 2017; Cao et al., 2019; Leman et al., 2020), with some tools using ML on experimental data (Pandurangan et al., 2017; Cao et al., 2019). Furthermore, glycoengineering reduces aggregation propensity and enhances conformational stability of biotherapeutics (Hristodorov et al., 2013; Courtois et al., 2015).

Recommendations for amino acid substitutions help design a customized humanization and optimization strategy for the lead mAb candidate. The top lead optimized candidates (3–6) are selected for large-scale production and biophysical characterizations. These processes can be extended to multi-specific antibodies, with additional engineering for optimizing Fv or ScFv domains and identifying optimal multi-specific formats.

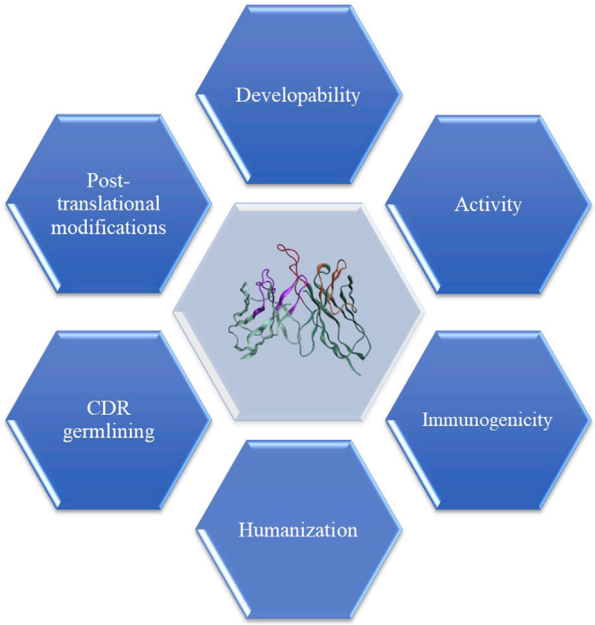

FIGURE 4
Lead humanization and optimization involve converting non-human sequences to human-like sequences while maintaining critical key attributes. *In vitro* binding affinity, which acts as surrogate for function, is the paramount criteria for accepting the mutations. Furthermore, *in silico* tools can be used to identify potential T-cell reactive epitopes, resulting in leads with lowest potential for immunogenicity and high percentage human content by germlining of the CDRs. Another aspect of optimization includes developability, which involves identifying leads with desirable biophysical properties and avoiding incidence of the post-translational modification sites such as N-linked glycosylation, unpaired cysteines, oxidation, deamidation, or aspartate isomerization, particularly in the CDRs.

3.8 Formatting of conventional and next-generation antibodies

After optimizing Fv regions, biotherapeutic engineering proceeds with formatting Fvs into the desired antibody format, combining Fv with the chosen IgG Fc isotype. Fc engineering may be required to adjust receptor-mediated functions like antibody-dependent cell-mediated cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), complement-dependent cytotoxicity (CDC), and endosomal recycling (Mimoto et al., 2016). For next-generation biotherapeutics like bi- and multi-specific antibodies, an intermediate formatting step assesses compatibility and developability properties. Structure-based engineering supports antibody formatting, as demonstrated in a study where TGFβ1 (transforming growth factor β1) binder affinity was restored after converting from ScFv to IgG (Lord et al., 2018). Similar approaches can support formatting complex next-generation antibodies.

In the discovery process's final step, top-performing lead variants undergo pre-formulation studies before transferring to development for cell line generation and early developability assessments (Bailly et al., 2020). The research phase concludes with the final candidate selection, after which conventional and DAbI-enabled workflows for antibody discovery are identical.

3.9 *In silico* assessments in early development

The initial stages of drug substance and drug product development are resource intensive, with full development programs justified only for the final candidate. At the time of selecting the final lead candidate, experimental data are often scarce due to material limitations. The sequence of the final lead candidate becomes locked at the start of development. This decision puts product development at a disadvantage, as real-time stability data are typically unavailable but crucial for meeting regulatory requirements concerning shelf-life, Critical Quality Attributes (CQA), and product heterogeneity. There is significant demand for early, rapid, and reliable stability predictions addressed through hybrid approaches combining *in vitro* and *in silico* techniques. Computational approaches can help estimate a molecule's fit to specific platform processes and tailor subsequent development programs to the biologic candidate's inherent liabilities and characteristics (Figure 5). Conversely, platform processes continuously gather data for new molecules, improving existing and developing novel bioinformatic predictions.

One platform step is the ultrafiltration/diafiltration (UF/DF), typically employed to process the antibody into the desired formulation. Recently, *in silico* models have demonstrated that protein charge can predict common UF/DF effects, such as Gibbs–Donnan and volume-exclusion phenomena (Kannan et al., 2023). After antibody formulation, certain stability aspects become most relevant for evaluating the developability of the final lead candidates using hybridized assessments.

Conformational stability is generally not an issue for conventional mAbs but can pose a significant challenge for next-generation biologics like ScFvs and multi-specific antibodies (Bailly et al., 2020). Numerous bioinformatics tools have been developed to calculate conformational stability, mostly applicable during LO for analyzing stability changes upon point mutations (Koenig et al., 2017; Pandurangan et al., 2017; Steinbrecher et al., 2017; Cao et al., 2019; Kuroda and Tsumoto, 2020; Leman et al., 2020; Harmalkar et al., 2023). Prediction accuracy heavily relies on the quality of the underlying structure or homology model, allowing comparisons between similar sequence variants.

Recent advancements in homology modeling and MD-based free energy calculations offer potential for enhancing thermal stability prediction (Kuhlman and Bradley, 2019; Berner et al., 2021; Tomar et al., 2021; Ko et al., 2022; Licari et al., 2022). Soon, these simulation approaches will extend from antibody fragments to full-length structures (Tomar et al., 2021). MD-derived predictions will improve by considering formulation aspects influencing conformational stability (Somani et al., 2021; Blanco, 2022; Saurabh et al., 2022; Shmool et al., 2022). High-throughput (HTP) screening of biologics' thermal stabilities in platform formulations enables AI, ML, and neural networks to train computational tools to predict the thermal stabilities of diverse candidates (Gentiluomo et al., 2019a; Cao et al., 2019; Wei, 2019; Bailly et al., 2020; Harmalkar et al., 2023). The pharmaceutical industry will benefit from bioinformatic tools predicting optimal formulation composition for specific candidates or identifying the best-suited candidate for a given formulation.

Predicting colloidal stability and aggregation propensity of drug products is critical, with bioinformatics offering significant advantages in development efforts. First, real-time stability studies may take years, allowing bioinformatics to reduce development time and risk of late-stage failure. Second, stability studies require large material amounts, particularly for HCPF, increasing the cost of failures. Third, extrapolations from accelerated stability studies often inaccurately reflect molecular behavior under storage conditions. Simplified approaches using conformational stability to estimate aggregation propensity only account for non-native aggregation (Brader et al., 2015), neglecting self-association and aggregation of natively folded mAbs. Fourth, analytical techniques like HIC, dynamic light scattering (DLS), self-interaction nanoparticle spectroscopy (SINS), size exclusion chromatography (SEC), and micro-flow imaging (MFI) partially characterize colloidal instability and aggregation, often necessitating a comprehensive analytical panel (Kopp et al., 2020). Last, colloidal instability and aggregation can be triggered by various intrinsic (molecule-related) (Alam et al., 2019; Gentiluomo et al., 2019b; Lai et al., 2022) and extrinsic (process-related) factors, following complex mechanisms. Conventional methods struggle to accurately predict shelf-life, leading to resource-intensive development studies and troubleshooting efforts when the development success is at risk.

A thorough understanding of molecular behavior is essential for addressing self-association, aggregation, or particulate formation issues. Computational approaches have been developed to estimate mechanistic and kinetic characteristics for better comprehension and prediction of colloidal instability and aggregation. Mechanistic tools aid in screening and minimizing APRs during the discovery phase (Kuhn et al., 2017; Prabakaran et al., 2017, 2020; van der Kant et al., 2017; Gil-Garcia et al., 2018; Rawat et al., 2018; Bauer et al., 2020; Ebo et al., 2020; Shahfar et al., 2022), while kinetic predictors

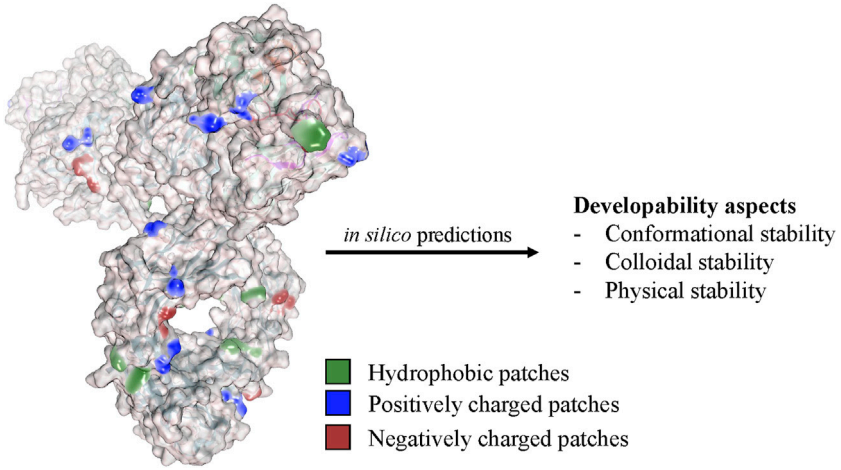

FIGURE 5
Computational approaches analyze the physicochemical properties of the antibody structure to predict various developability aspects and stability factors. These *in silico* methods evaluate factors such as aggregation propensity, conformational stability, colloidal stability, and post-translational modifications and help to select candidates with improved developability and reduced risk of immunogenicity or manufacturing challenges.

estimate aggregation rates, crucial for liquid formulation development meeting regulatory requirements for shelf life (Rawat et al., 2019; Yang et al., 2019; Santos et al., 2020). Machine Learning (ML) can train kinetic models using extensive data sets with experimental and sequence/structure information (Rawat et al., 2019; Yang et al., 2019), facilitating prediction of optimal formulation compositions (pH, salt, excipients) for minimal kinetics.

In the final development stage, creating liquid drug products with stable physical properties is vital. Manufacturing, processing, and administration of highly concentrated antibody formulations often face viscosity challenges. Viscosity is linked to surface charge and hydrophobicity of the mAb (Tomar et al., 2018; Apgar et al., 2020; Lai et al., 2021; Blanco, 2022; Han et al., 2022; Lai, 2022). Studies have shown computational ability to predict viscosity profiles at platform conditions using mAb sequence and structure (Tilegenova et al., 2019; Bauer et al., 2020; Thorsteinson et al., 2021; Han et al., 2022; Lai et al., 2022; Rosace et al., 2022). A recent deep learning approach utilized a 3D convolutional neural network to predict high-concentration viscosity of therapeutic antibodies (Rai et al., 2023). Feature attribution analysis identified key biophysical drivers of viscosity, such as the electrostatic potential surface. The predictor was successfully trained despite limited data. Early integration of viscosity predictors enables addressing viscosity issues and adjusting platform formulations and technologies before finalizing the development strategy.

4 Discussion and conclusion

In this review, we have presented numerous opportunities for computation to play a greater role in biotherapeutics discovery and development. However, the excitement around computation's enhanced role should be tempered with pragmatism. Machine learning experts often lack practical experience in biotherapeutics discovery and development and *vice versa*. Thus, a strong collaboration between bench scientists and data scientists is recommended. Computational biophysics and antibody structure–function–developability relationship experts should work with machine learning and artificial intelligence experts, as well as experimentalists, to fully enable biopharmaceutical informatics. Additionally, technical limitations exist in emerging technologies like machine learning and artificial intelligence. For instance, deep learning model performance often depends on size and diversity within training data sets (Wittmund et al., 2022), posing challenges in sparse or less diverse data settings. Moreover, the lack of insights into the latent space and interpretability of AI models in terms of the underlying physicochemical rules hinders our ability to better understand the models and extend their applicability beyond the tasks they have been trained for. For example, AI-based methods have transformed protein structure prediction, but contrary to popular belief, they have not solved the protein folding problem (Chen et al., 2023), as they do not provide insights into protein folding processes, such as initial building blocks, intermediate states, energy landscapes, and pathways.

In the specific context of protein engineering, the complexity of prediction tasks is escalated by non-additive mutational interactions or epistatic effects, which can significantly alter the impact of single or multiple mutational outcomes (Reetz, 2013; Miton and Tokuriki, 2016; Cadet et al., 2022). A further layer of complication is presented by the dynamic interplay between mutated amino acids and the subsequent establishment of intramolecular interaction networks, which can alter the protein function (Acevedo-Rocha et al., 2021). The situation is exacerbated by the limitations of tools such as AlphaFold2 or ProteinMPNN, which may struggle to predict how individual amino acid changes affect protein structure due to their heavy reliance on evolutionary perspectives and variant sequences (Eisenstein, 2021; Dauparas et al., 2022). Deep learning methods offer a way to investigate protein attributes, such as stability, solubility, aggregation, and binding affinity. However, these methods operate within the confines of the training data.

Although this does not eliminate the possibility of identifying beneficial protein variations within these parameters, it may fail to recognize or accurately predict variants exhibiting fitness values outside the learned range. This means that while beneficial variants can be identified, the optimal variant, particularly if it is an epistatic variant, might be overlooked. Against this backdrop, the use of deep learning models in conjunction with conventional neural network architectures is being explored as a solution for these challenges. By representing numerical quantities as individual neurons without non-linearity, these models can learn to perform systematic numerical computation, enabling them to handle data that lie outside the range used during training (Trask et al., 2018). The adaptability of these models across various task domains augments their potential to tackle challenges encountered in antibody therapeutics. Importantly, the ability to harness epistatic effects and predict mutational outcomes could significantly enhance the design of therapeutic antibodies. Moreover, other studies have indicated the potency of a Machine learning (ML) approach focused exclusively on sequences in accurately predicting epistatic phenomena (Cadet et al., 2018). Unlike most ML and deep learning methodologies that predominantly capture low-order non-linear interactions and predict the additive effects of mutations, this innovative strategy comprehensively encapsulates both low- and high-order non-linear interactions. By utilizing ML in tandem with digital signal processing such as Fourier transform, case studies have demonstrated a significant improvement in the resistance of proteins to unfavorable unfolding and aggregation. Crucially, this method unveils the correlation between epistatic mutational interactions and protein resilience, offering unique, predictive insights beyond those provided by conventional machine learning or deep learning approaches (Li et al., 2021). This approach has considerably enhanced precision, reduced overfitting, and surpassed conventional methods without increasing complexity (Medina-Ortiz et al., 2022). Understanding the rules underlying these interactions could contribute to a more efficient model design and a more predictive performance, thereby bolstering the success of deep learning in the realm of biopharmaceutical informatics.

In conclusion, this review article aims to broaden our strategic perspective on biopharmaceutical informatics. Initially, we emphasized the syncretic use of computation and experimentation for the drug product development of antibody-based biotherapeutics (Kumar et al., 2015; Kumar et al., 2018a). Subsequently, Khetan et al. (2022) demonstrated its feasibility by spelling out different methods and published studies already available in the public domain to support our vision. Here, we propose a more generalized vision of biopharmaceutical informatics by including DAbI and digital transformation. It is widely agreed that digital transformation is essential for modernizing the biopharmaceutical industry's work processes, leading to more judicious use of resources and reduced costs in biotherapeutics discovery and development. Recent advancements in AI and ML, along with the availability of large-scale antibody sequencing data in the public domain, have fueled excitement for DAbI. When fully embraced by the biopharmaceutical industry, DAbI will revolutionize the way biotherapeutic drugs are discovered and developed. Current drug discovery processes and workflows are dominated by experimental trials and errors, with computation playing an assistive role at the best. DAbI can support the start of projects even before the availability of antigen material for *in vitro* experimental studies. This is particularly attractive when the antigens involved are difficult to express and purify. DAbI can also accelerate discovery projects by pre-paying for developability and therefore save on resources and time required to fix these issues at the later stages. These two features may eventually lead to situations where computation plays an equal, if not greater, role alongside experimentation in supporting biotherapeutics discovery and development projects. Therefore, our vision of biopharmaceutical informatics points to an exciting future where we can better serve patients by addressing unmet medical needs through more successful, faster, and affordable discovery and development of biotherapeutics. Additionally, the discovery and development of antibody-based biotherapeutics are rapidly becoming industrialized, with several aspects becoming more uniform (e.g., discovery processes and drug formulations), while multiple options are being explored for others, such as molecular formats, routes of administration, and dosing options (Martin et al., 2023). Biopharmaceutical informatics contributes toward accelerating this industrialization and helping to improve human health.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

All authors thank numerous colleagues in Biotherapeutics Discovery and Development Biologicals for excellent discussions and support.

Conflict of interest

JB was employed by the company Boehringer Ingelheim Pharma GmbH & Co. KG; NR, PrG, PaG, AN, and SK were employed by the company Boehringer Ingelheim Pharmaceuticals Inc.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, editors, and reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abhinandan, K. R., and Martin, A. C. R. (2007). Analyzing the "degree of humanness" of antibody sequences. *J. Mol. Biol.* 369, 852–862. doi:10.1016/j.jmb.2007.02.100

Acevedo-Rocha, C. G., Li, A., D'Amore, L., Hoebenreich, S., Sanchis, J., Lubrano, P., et al. (2021). Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics. *Nat. Commun.* 12, 1621. doi:10.1038/s41467-021-21833-w

Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B. D., Hu, X., Adachi, Y., et al. (2018). RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *Plos Comput. Biol.* 14, e1006112. doi:10.1371/journal.pcbi.1006112

Agrawal, N. J., Helk, B., Kumar, S., Mody, N., Sathish, H. A., Samra, H. S., et al. (2015). Computational tool for the early screening of monoclonal antibodies for their viscosities. *Mabs* 8, 43–48. doi:10.1080/19420862.2015.1099773

Ahmed, L., Guptaa, P., Martin, K. P., Scheer, J. M., Nixon, A. E., and Kumar, S. (2021). Intrinsic physicochemical profile of marketed antibody-based biotherapeutics. *PNAS* 118, e2020577118. doi:10.1073/pnas.2020577118

Akbar, R., Robert, P. A., Pavlović, M., Jeliazkov, J. R., Snapkov, I., Slabodkin, A., et al. (2021). A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. *Cell Rep.* 34, 108856. doi:10.1016/j.celrep.2021.108856

Akbar, R., Robert, P. A., Weber, C. R., Widrich, M., Frank, R., Pavlović, M., et al. (2022). *In silico* proof of principle of machine learning-based antibody design at unconstrained scale. *Mabs* 14, 2031482. doi:10.1080/19420862.2022.2031482

Alam, M. E., Barnett, G. V., Slaney, T. R., Starr, C. G., Das, T. K., and Tessier, P. M. (2019). Deamidation can compromise antibody colloidal stability and enhance aggregation in a pH-dependent manner. *Mol. Pharm.* 16, 1939–1949. doi:10.1021/acs.molpharmaceut.8b01311

Alfaleh, M. A., Alsaab, H. O., Mahmoud, A. B., Alkayyal, A. A., Jones, M. L., Mahler, S. M., et al. (2020). Phage display derived monoclonal antibodies: From bench to bedside. *Front. Immunol.* 11, 1986. doi:10.3389/fimmu.2020.01986

AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics* 35, 4862–4865. doi:10.1093/bioinformatics/btz422

Ambrosetti, F., Jiménez-García, B., Roel-Touris, J., and Bonvin, A. M. J. J. (2020a). Modeling antibody-antigen complexes by information-driven docking. *Structure* 28, 119–129.e2. doi:10.1016/j.str.2019.10.011

Ambrosetti, F., Olsen, T. H., Olimpieri, P. P., Jiménez-García, B., Milanetti, E., Marcatilli, P., et al. (2020b). proABC-2: PRediction Of AntiBody Contacts v2 and its application to information-driven docking. *Bioinformatics* 36, 5107–5108. doi:10.1093/bioinformatics/btaa644

Amimeur, T., Shaver, J. M., Ketchem, R. R., Taylor, J. A., Clark, R. H., Smith, J., et al. (2020). Designing feature-controlled humanoid antibody discovery libraries using generative adversarial networks. *Biorxiv*. doi:10.1101/2020.04.12.024844

Apgar, J. R., Tam, A. S. P., Sorm, R., Moesta, S., King, A. C., Yang, H., et al. (2020). Modeling and mitigation of high-concentration antibody viscosity through structure-based computer-aided protein design. *Plos One* 15, e0232713. doi:10.1371/journal.pone.0232713

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi:10.1126/science.abj8754

Baets, G. D., Durme, J. V., van der Kant, R., Schymkowitz, J., and Rousseau, F. (2015). Solubis: Optimize your protein. *Bioinformatics* 31, 2580–2582. doi:10.1093/bioinformatics/btv162

Bailly, M., Mieczkowski, C., Juan, V., Metwally, E., Tomazela, D., Baker, J., et al. (2020). Predicting antibody developability profiles through early stage discovery screening. *Mabs* 12, 1743053. doi:10.1080/19420862.2020.1743053

Baran, D., Pszolla, M. G., Lapidoth, G. D., Norn, C., Dym, O., Unger, T., et al. (2017). Principles for computational design of binding antibodies. *Proc. Natl. Acad. Sci.* 114, 10900–10905. doi:10.1073/pnas.1707171114

Bauer, J., Mathias, S., Kube, S., Otte, K., Garidel, P., Gamer, M., et al. (2020). Rational optimization of a monoclonal antibody improves the aggregation propensity and enhances the CMC properties along the entire pharmaceutical process chain. *Mabs* 12, 1787121. doi:10.1080/19420862.2020.1787121

Benatuil, L., Perez, J. M., Belk, J., and Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* 23, 155–159. doi:10.1093/protein/gzq002

Berner, C., Menzen, T., Winter, G., and Svilenov, H. L. (2021). Combining unfolding reversibility studies and molecular dynamics simulations to select aggregation-resistant antibodies. *Mol. Pharm.* 18, 2242–2253. doi:10.1021/acs.molpharmaceut.1c00017

Bill, R. M., Henderson, P. J. F., Iwata, S., Kunji, E. R. S., Michel, H., Neutze, R., et al. (2011). Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* 29, 335–340. doi:10.1038/nbt.1833

Blanco, M. A. (2022). Computational models for studying physical instabilities in high concentration biotherapeutic formulations. *Mabs* 14, 2044744. doi:10.1080/19420862.2022.2044744

Brader, M. L., Estey, T., Bai, S., Alston, R. W., Lucas, K. K., Lantz, S., et al. (2015). Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies. *Mol. Pharm.* 12, 1005–1017. doi:10.1021/mp400666b

Buell, A. K., Hung, P., Salvatella, X., Welland, M. E., Dobson, C. M., and Knowles, T. P. J. (2013). Electrostatic effects in filamentous protein aggregation. *Biophys. J.* 104, 1116–1126. doi:10.1016/j.bpj.2013.01.031

Bujotzek, A., Fuchs, A., Qu, C., Benz, J., Klostermann, S., Antes, I., et al. (2015). MoFvAb: Modeling the Fv region of antibodies. *Mabs* 7, 838–852. doi:10.1080/19420862.2015.1068492

Cadet, F., Fontaine, N., Li, G., Sanchis, J., Chong, M. N. F., Pandjaitan, R., et al. (2018). A machine learning approach for reliable prediction of amino acid interactions and its application in the directed evolution of enantioselective enzymes. *Sci. Rep.* 8, 16757. doi:10.1038/s41598-018-35033-y

Cadet, X. F., Gelly, J. C., van Noord, A., Cadet, F., and Acevedo-Rocha, C. G. (2022). "Learning strategies in protein directed EvolutionDirected evolution (DE)," in *Methods in molecular biology* (New York, NY: Springer US), 225–275. doi:10.1007/978-1-0716-2152-3_15

Cao, H., Wang, J., He, L., Qi, Y., and Zhang, J. Z. (2019). DeepDDG: Predicting the stability change of protein point mutations using neural networks. *J. Chem. Inf. Model* 59, 1508–1514. doi:10.1021/acs.jcim.8b00697

Chen, G., and Sidhu, S. S. (2014). Design and generation of synthetic antibody libraries for phage display. *Methods Mol. Biol.* 1131, 113–131. doi:10.1007/978-1-62703-992-5_8

Chen, S.-J., Hassan, M., Jernigan, R. L., Jia, K., Kihara, D., Kloczkowski, A., et al. (2023). Opinion: Protein folds vs. protein folding: Differing questions, different challenges. *Proc. Natl. Acad. Sci.* 120, e2214423119. doi:10.1073/pnas.2214423119

Chiba, S., Tanabe, A., Nakakido, M., Okuno, Y., Tsumoto, K., and Ohta, M. (2020). Structure-based design and discovery of novel anti-tissue factor antibodies with cooperative double-point mutations, using interaction analysis. *Sci. Rep-uk* 10, 17590. doi:10.1038/s41598-020-74545-4

Chiu, M. L., Goulet, D. R., Teplyakov, A., and Gilliland, G. L. (2019). Antibody structure and function: The basis for engineering therapeutics. *Antibodies* 8, 55. doi:10.3390/antib8040055

Choi, Y., Hua, C., Sentman, C. L., Ackerman, M. E., and Bailey-Kellogg, C. (2015). Antibody humanization by structure-based computational protein design. *Mabs* 7, 1045–1057. doi:10.1080/19420862.2015.1076600

Choi, Y., Verma, D., Griswold, K. E., and Bailey-Kellogg, C. (2017). "EpiSweep: Computationally driven reengineering of therapeutic proteins to reduce immunogenicity while maintaining function computational protein design," in *Methods in molecular biology*. Editor I. Samish doi:10.1007/978-1-4939-6637-0_20

Chowdhury, R., Allan, M. F., and Maranas, C. D. (2018). OptMAVEn-2.0: De novo design of variable antibody regions against targeted antigen epitopes. *Antibodies* 7, 23. doi:10.3390/antib7030023

Comeau, S. R., Thorsteinson, N., and Kumar, S. (2023). "Structural considerations in affinity maturation of antibody-based biotherapeutic candidates," in *Methods in molecular biology* (New York, NY: Springer US), 309–321. doi:10.1007/978-1-0716-2609-2_17

Conti, S., Lau, E. Y., and Ovchinnikov, V. (2022). On the rapid calculation of binding affinities for antigen and antibody design and affinity maturation simulations. *Antibodies* 11, 51. doi:10.3390/antib11030051

Courtois, F., Agrawal, N. J., Lauer, T. M., and Trout, B. L. (2015). Rational design of therapeutic mAbs against aggregation through protein engineering and incorporation of glycosylation motifs applied to bevacizumab. *Mabs* 8, 99–112. doi:10.1080/19420862.2015.1112477

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., et al. (2022). Robust deep learning–based protein sequence design using ProteinMPNN. *Science* 378, 49–56. doi:10.1126/science.add2187

Davila, A., Xu, Z., Li, S., Rozewicki, J., Wilamowski, J., Kotelnikov, S., et al. (2022). AbAdapt: An adaptive approach to predicting antibody–antigen complex structures from sequence. *Bioinform Adv.* 2, vbac015. doi:10.1093/bioadv/vbac015

Deac, A., VeliČković, P., and Sormanni, P. (2019). Attentive cross-modal paratope prediction. *J. Comput. Biol.* 26, 536–545. doi:10.1089/cmb.2018.0175

Dehouck, Y., Kwasigroch, J. M., Gilis, D., and Rooman, M. (2011). PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality. *Bmc Bioinforma.* 12, 151. doi:10.1186/1471-2105-12-151

DiMasi, J. A., and Grabowski, H. G. (2007). The cost of biopharmaceutical R&D: Is biotech different? *Manag. Decis. Econ.* 28, 469–479. doi:10.1002/mde.1360

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* 47, 20–33. doi:10.1016/j.jhealeco.2016.01.012

Ebo, J. S., Guthertz, N., Radford, S. E., and Brockwell, D. J. (2020). Using protein engineering to understand and modulate aggregation. *Curr. Opin. Struc Biol.* 60, 157–166. doi:10.1016/j.sbi.2020.01.005

Eguchi, R. R., Choe, C. A., and Huang, P.-S. (2022). Ig-VAE: Generative modeling of protein structure by direct 3D coordinate generation. *Plos Comput. Biol.* 18, e1010271. doi:10.1371/journal.pcbi.1010271

Eguida, M., and Rognan, D. (2020). A computer vision approach to align and compare protein cavities: Application to fragment-based drug design. *J. Med. Chem.* 63, 7127–7142. doi:10.1021/acs.jmedchem.0c00422

Eisenstein, M. (2021). Artificial intelligence powers protein-folding predictions. *Nature* 599, 706–708. doi:10.1038/d41586-021-03499-y

Farid, S. S., Baron, M., Stamatis, C., Nie, W., and Coffman, J. (2020). Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D. *Mabs* 12, 1754999. doi:10.1080/19420862.2020.1754999

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22, 1302–1306. doi:10.1038/nbt1012

Fernández-Quintero, M. L., Kokot, J., Waibl, F., Fischer, A.-L. M., Quoika, P. K., Deane, C. M., et al. (2023). Challenges in antibody structure prediction. *mAbs* 15, 2175319. doi:10.1080/19420862.2023.2175319

Fogel, D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp. Clin. Trials Commun.* 11, 156–164. doi:10.1016/j.conctc.2018.08.001

Folkman, L., Stantic, B., Sattar, A., and Zhou, Y. (2016). EASE-MM: Sequence-Based prediction of mutation-induced stability changes with feature-based multiple models. *J. Mol. Biol.* 428, 1394–1405. doi:10.1016/j.jmb.2016.01.012

Gainza, P., Sverrisson, F., Monti, F., Rodolà, E., Boscaini, D., Bronstein, M. M., et al. (2020). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* 17, 184–192. doi:10.1038/s41592-019-0666-6

Gao, S. H., Huang, K., Tu, H., and Adler, A. S. (2013). Monoclonal antibody humanness score and its applications. *BMC Biotechnol.* 55, 55. doi:10.1186/1472-6750-13-55

Gao, W., Mahajan, S. P., Sulam, J., and Gray, J. J. (2020). Deep learning in protein structural modeling and design. *Patterns* 1, 100142. doi:10.1016/j.patter.2020.100142

Garbuzynskiy, S. O., Lobanov, M. Y., and Galzitskaya, O. V. (2010). FoldAmyloid: A method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* 26, 326–332. doi:10.1093/bioinformatics/btp691

Garidel, P., Kuhn, A. B., Schäfer, L. V., Karow-Zwick, A. R., and Blech, M. (2017). High-concentration protein formulations: How high is high? *Eur. J. Pharm. Biopharm.* 119, 353–360. doi:10.1016/j.ejpb.2017.06.029

Garripelli, V. K., Wu, Z., and Gupta, S. (2020). Developability assessment for monoclonal antibody drug candidates: A case study. *Pharm. Dev. Technol.* 1, 11–20. doi:10.1080/10837450.2020.1829641

Gentiluomo, L., Roessner, D., Augustijn, D., Svilenov, H., Kulakova, A., Mahapatra, S., et al. (2019a). Application of interpretable artificial neural networks to early monoclonal antibodies development. *Eur. J. Pharm. Biopharm.* 141, 81–89. doi:10.1016/j.ejpb.2019.05.017

Gentiluomo, L., Roessner, D., Streicher, W., Mahapatra, S., Harris, P., and Frieß, W. (2019b). Characterization of native reversible self-association of a monoclonal antibody mediated by Fab-Fab interaction. *J. Pharm. Sci.* 109, 443–451. doi:10.1016/j.xphs.2019.09.021

Gil-Garcia, M., Bañó-Polo, M., Varejão, N., Jamroz, M., Kuriata, A., Díaz-Caballero, M., et al. (2018). Combining structural aggregation propensity and stability predictions to redesign protein solubility. *Mol. Pharm.* 15, 3846–3859. doi:10.1021/acs.molpharmaceut.8b00341

González-Fernández, Á., Silva, F. J. B., López-Hoyos, M., Cobaleda, C., Montoliu, L., Val, M. D., et al. (2020). Non-animal-derived monoclonal antibodies are not ready to substitute current hybridoma technology. *Nat. Methods* 17, 1069–1070. doi:10.1038/s41592-020-00977-5

Gorgulla, C., Boeszoermenyi, A., Wang, Z.-F., Fischer, P. D., Coote, P. W., Das, K. M. P., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 580, 663–668. doi:10.1038/s41586-020-2117-z

Goyon, A., Excoffier, M., Janin-Bussat, M.-C., Bobaly, B., Fekete, S., Guillarme, D., et al. (2017). Determination of isoelectric points and relative charge variants of 23 therapeutic monoclonal antibodies. *J. Chromatogr. B* 1065, 119–128. doi:10.1016/j.jchromb.2017.09.033

Gray, A., Bradbury, A. R. M., Knappik, A., Plückthun, A., Borrebaeck, C. A. K., and Dübel, S. (2020). Animal-free alternatives and the antibody iceberg. *Nat. Biotechnol.* 38, 1234–1239. doi:10.1038/s41587-020-0687-9

Gray, A. C., Bradbury, A. R. M., Knappik, A., Plückthun, A., Borrebaeck, C. A. K., and Dübel, S. (2020). Animal-derived-antibody generation faces strict reform in accordance with European Union policy on animal use. *Nat. Methods* 17, 755–756. doi:10.1038/s41592-020-0906-9

Grossman, I., Ilani, T., Fleishman, S. J., and Fass, D. (2016). Overcoming a species-specificity barrier in development of an inhibitory antibody targeting a modulator of tumor stroma. *Protein Eng. Des. Sel.* 29, 135–147. doi:10.1093/protein/gzv067

Han, X., Shih, J., Lin, Y., Chai, Q., and Cramer, S. M. (2022). Development of QSAR models for *in silico* screening of antibody solubility. *Mabs* 14, 2062807. doi:10.1080/19420862.2022.2062807

Harmalkar, A., Rao, R., Xie, Y. R., Honer, J., Deisting, W., Anlahr, J., et al. (2023). Toward generalizable prediction of antibody thermostability using machine learning on sequence and structure features. *Mabs* 15, 2163584. doi:10.1080/19420862.2022.2163584

Hebditch, M., and Warwicker, J. (2019). Charge and hydrophobicity are key features in sequence-trained machine learning models for predicting the biophysical properties of clinical-stage antibodies. *Peerj* 7, e8199. doi:10.7717/peerj.8199

Hristodorov, D., Fischer, R., Joerissen, H., Müller-Tiemann, B., Apeler, H., and Linden, L. (2013). Generation and comparative characterization of glycosylated and aglycosylated human IgG1 antibodies. *Mol. Biotechnol.* 53, 326–335. doi:10.1007/s12033-012-9531-x

Imrie, F., Bradley, A. R., van der Schaar, M., and Deane, C. M. (2018). Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data. *J. Chem. Inf. Model* 58, 2319–2330. doi:10.1021/acs.jcim.8b00350

Irudayanathan, F. J., Zarzar, J., Lin, J., and Izadi, S. (2022). Deciphering deamidation and isomerization in therapeutic proteins: Effect of neighboring residue. *Mabs* 14, 2143006. doi:10.1080/19420862.2022.2143006

Jain, T., Boland, T., Lilov, A., Burnina, I., Brown, M., Xu, Y., et al. (2017a). Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning. *Bioinformatics* 33, 3758–3766. doi:10.1093/bioinformatics/btx519

Jain, T., Sun, T., Durand, S., Hall, A., Houston, N. R., Nett, J. H., et al. (2017b). Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci.* 114, 944–949. doi:10.1073/pnas.1616408114

Jain, T., Boland, T., and Vásquez, M. (2023). Identifying developability risks for clinical progression of antibodies using high-throughput *in vitro* and *in silico* approaches. *Mabs* 15, 2200540. doi:10.1080/19420862.2023.2200540

Jarasch, A., Koll, H., Regula, J. T., Bader, M., Papadimitriou, A., and Kettenberger, H. (2015). Developability assessment during the selection of novel therapeutic antibodies. *J. Pharm. Sci.* 104, 1885–1898. doi:10.1002/jps.24430

Jeliazkov, J. R., Frick, R., Zhou, J., and Gray, J. J. (2021). Robustification of RosettaAntibody and Rosetta SnugDock. *Plos One* 16, e0234282. doi:10.1371/journal.pone.0234282

Jetha, A., Thorsteinson, N., Jmeian, Y., Jeganathan, A., Giblin, P., and Fransson, J. (2018). Homology modeling and structure-based design improve hydrophobic interaction chromatography behavior of integrin binding antibodies. *Mabs* 10, 890–900. doi:10.1080/19420862.2018.1475871

Jones, D. T., and Thornton, J. M. (2022). The impact of AlphaFold2 one year on. *Nat. Methods* 19, 15–20. doi:10.1038/s41592-021-01365-3

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Kannan, A., Chinn, M., Izadi, S., Maier, A., Dvornicky, J., Fedesco, M., et al. (2023). Predicting formulation conditions during ultrafiltration and dilution to drug substance using a donnan model with homology-model based protein charge. *J. Pharm. Sci.* 112, 820–829. doi:10.1016/j.xphs.2022.10.028

Kaplon, H., Chenoweth, A., Crescioli, S., and Reichert, J. M. (2022). Antibodies to watch in 2022. *Mabs* 14, 2014296. doi:10.1080/19420862.2021.2014296

Karlberg, M., de Souza, J. V., Fan, L., Kizhedath, A., Bronowska, A. K., and Glassey, J. (2020). QSAR implementation for HIC retention time prediction of mAbs using fab structure: A comparison between structural representations. *Int. J. Mol. Sci.* 21, 8037. doi:10.3390/ijms21218037

Kemmish, H., Fasnacht, M., and Yan, L. (2017). Fully automated antibody structure prediction using BIOVIA tools: Validation study. *Plos One* 12, e0177923. doi:10.1371/journal.pone.0177923

Khan, A., Cowen-Rivers, A. I., Grosnit, A., Deik, D.-G.-X., Robert, P. A., Greiff, V., et al. (2023). Toward real-world automated antibody design with combinatorial Bayesian optimization. *Cell Rep. Methods* 3, 100374. doi:10.1016/j.crmeth.2022.100374

Khetan, R., Curtis, R., Deane, C. M., Hadsund, J. T., Kar, U., Krawczyk, K., et al. (2022). Current advances in biopharmaceutical informatics: Guidelines, impact and challenges in the computational developability assessment of antibody therapeutics. *Mabs* 14, 2020082. doi:10.1080/19420862.2021.2020082

Kiyoshi, M., Caaveiro, J. M. M., Miura, E., Nagatoishi, S., Nakakido, M., Soga, S., et al. (2014). Affinity improvement of a therapeutic antibody by structure-based computational design: Generation of electrostatic interactions in the transition state stabilizes the antibody-antigen complex. *Plos One* 9, e87099. doi:10.1371/journal.pone.0087099

Klausen, M. S., Anderson, M. V., Jespersen, M. C., Nielsen, M., and Marcatili, P. (2015). LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Res.* 43, W349–W355. doi:10.1093/nar/gkv535

Ko, S. K., Berner, C., Kulakova, A., Schneider, M., Antes, I., Winter, G., et al. (2022). Investigation of the pH-dependent aggregation mechanisms of GCSF using low resolution protein characterization techniques and advanced molecular dynamics simulations. *Comput. Struct. Biotechnol. J.* 20, 1439–1455. doi:10.1016/j.csbj.2022.03.012

Koehler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495–497. doi:10.1038/256495a0

Koenig, P., Lee, C. V., Walters, B. T., Janakiraman, V., Stinson, J., Patapoff, T. W., et al. (2017). Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proc. Natl. Acad. Sci.* 114, E486–E495. doi:10.1073/pnas.1613231114

Kopp, M. R. G., Pérez, A.-M. W., Zucca, M. V., Palmiero, U. C., Friedrichsen, B., Lorenzen, N., et al. (2020). An accelerated surface-mediated stress assay of antibody instability for developability studies. *Mabs* 12, 1815995. doi:10.1080/19420862.2020.1815995

Krawczyk, K., Baker, T., Shi, J., and Deane, C. M. (2013). Antibody i-Patch prediction of the antibody binding site improves rigid local antibody–antigen docking. *Protein Eng. Des. Sel.* 26, 621–629. doi:10.1093/protein/gzt043

Krawczyk, K., Liu, X., Baker, T., Shi, J., and Deane, C. M. (2014). Improving B-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics* 30, 2288–2294. doi:10.1093/bioinformatics/btu190

Kuhlman, B., and Bradley, P. (2019). Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Bio* 20, 681–697. doi:10.1038/s41580-019-0163-x

Kuhn, A. B., Kube, S., Karow-Zwick, A. R., Seeliger, D., Garidel, P., Blech, M., et al. (2017). Improved solution-state properties of monoclonal antibodies by targeted mutations. *J. Phys. Chem. B* 121, 10818–10827. doi:10.1021/acs.jpcb.7b09126

Kumar, S., and Singh, S. K. (2015). in *Developability of biotherapeutics: Computational approaches*. Editors S. Kumar and S. K. Singh

Kumar, S., Singh, S. K., and Gromiha, M. M. (2009). "Temperature-dependent molecular adaptations, microbial proteins," in *Encyclopedia of industrial biotechnology*, 1–22. doi:10.1002/9780470054581.eib516

Kumar, S., Robins, R. H., Buck, P. M., Hickling, T. P., Thangakani, A. M., Li, L., et al. (2015). *Biopharmaceutical informatics: Applications of computation in biologic drug development*. New York, NY: CRC Press, 3–34.

Kumar, S., Plotnikov, N. V., Rouse, J. C., and Singh, S. K. (2018a). Biopharmaceutical informatics: Supporting biologic drug development via molecular modelling and informatics. *J. Pharm. Pharmacol.* 70, 595–608. doi:10.1111/jphp.12700

Kumar, S., Roffi, K., Tomar, D. S., Cirelli, D., Luksha, N., Meyer, D., et al. (2018b). Rational optimization of a monoclonal antibody for simultaneous improvements in its solution properties and biological activity. *Protein Eng. Des. Sel.* 31, 313–325. doi:10.1093/protein/gzy020

Kuroda, D., and Tsumoto, K. (2020). Engineering stability, viscosity, and immunogenicity of antibodies by computational design. *J. Pharm. Sci.* 109, 1631–1651. doi:10.1016/j.xphs.2020.01.011

Lai, P.-K., Fernando, A., Cloutier, T. K., Gokarn, Y., Zhang, J., Schwenger, W., et al. (2021). Machine learning applied to determine the molecular descriptors responsible for the viscosity behavior of concentrated therapeutic antibodies. *Mol. Pharm.* 18, 1167–1175. doi:10.1021/acs.molpharmaceut.0c01073

Lai, P.-K., Gallegos, A., Mody, N., Sathish, H. A., and Trout, B. L. (2022). Machine learning prediction of antibody aggregation and viscosity for high concentration formulation development of protein therapeutics. *Mabs* 14, 2026208. doi:10.1080/19420862.2022.2026208

Lai, P.-K. (2022). DeepSCM: An efficient convolutional neural network surrogate model for the screening of therapeutic antibody viscosity. *Comput. Struct. Biotechnol. J.* 20, 2143–2152. doi:10.1016/j.csbj.2022.04.035

Lapidoth, G., Parker, J., Prilusky, J., and Fleishman, S. J. (2018). AbPredict 2: A server for accurate and unstrained structure prediction of antibody variable domains. *Bioinformatics* 35, 1591–1593. doi:10.1093/bioinformatics/bty822

Lazar, G. A., Desjarlais, J. R., Jacinto, J., Karki, S., and Hammond, P. W. (2007). A molecular immunology approach to antibody humanization and functional optimization. *Mol. Immunol.* 44, 1986–1998. doi:10.1016/j.molimm.2006.09.029

Lecerf, M., Kanyavuz, A., Lacroix-Desmazes, S., and Dimitrov, J. D. (2019). Sequence features of variable region determining physicochemical properties and polyreactivity of therapeutic antibodies. *Mol. Immunol.* 112, 338–346. doi:10.1016/j.molimm.2019.06.012

Ledsgaard, L., Ljungars, A., Rimbault, C., Sørensen, C. V., Tulika, T., Wade, J., et al. (2022). Advances in antibody phage display technology. *Drug Discov. Today* 27, 2151–2169. doi:10.1016/j.drudis.2022.05.002

Leem, J., Dunbar, J., Georges, G., Shi, J., and Deane, C. M. (2016). ABodyBuilder: Automated antibody structure prediction with data–driven accuracy estimation. *Mabs* 8, 1259–1268. doi:10.1080/19420862.2016.1205773

Lehmann, A., Wixted, J. H. F., Shapovalov, M. V., Roder, H., Dunbrack, R. L., and Robinson, M. K. (2015). Stability engineering of anti-EGFR scFv antibodies by rational design of a lambda-to-kappa swap of the VL framework using a structure-guided approach. *Mabs* 7, 1058–1071. doi:10.1080/19420862.2015.1088618

Leman, J. K., Weitzner, B. D., Lewis, S. M., Adolf-Bryfogle, J., Alam, N., Alford, R. F., et al. (2020). Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nat. Methods* 17, 665–680. doi:10.1038/s41592-020-0848-2

Li, T., Pantazes, R. J., and Maranas, C. D. (2014). OptMAVEn – a New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes. *Plos One* 9, e105954. doi:10.1371/journal.pone.0105954

Li, G., Qin, Y., Fontaine, N. T., Chong, M. N. F., Maria-Solano, M. A., Feixas, F., et al. (2021). Machine learning enables selection of epistatic enzyme mutants for stability against unfolding and detrimental aggregation. *ChemBioChem* 22, 904–914. doi:10.1002/cbic.202000612

Liberis, E., Velickovic, P., Sormanni, P., Vendruscolo, M., and Liò, P. (2018). Parapred: Antibody paratope prediction using convolutional and recurrent neural networks. *Bioinform Oxf Engl.* 34, 2944–2950. doi:10.1093/bioinformatics/bty305

Licari, G., Martin, K. P., Crames, M., Mozdzierz, J., Marlow, M. S., Karow-Zwick, A. R., et al. (2022). Embedding dynamics in intrinsic physicochemical profiles of market-stage antibody-based biotherapeutics. *Mol. Pharm.* 20, 1096–1111. doi:10.1021/acs.molpharmaceut.2c00838

Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol.* 44, 235–249. doi:10.1016/s1056-8719(00)00107-6

Liu, X., Taylor, R. D., Griffin, L., Coker, S.-F., Adams, R., Ceska, T., et al. (2017). Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping. *Sci. Rep-uk* 7, 41306. doi:10.1038/srep41306

Liu, G., Zeng, H., Mueller, J., Carter, B., Wang, Z., Schilz, J., et al. (2019). Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics* 36, 2126–2133. doi:10.1093/bioinformatics/btz895

Lord, D. M., Bird, J. J., Honey, D. M., Best, A., Park, A., Wei, R. R., et al. (2018). Structure-based engineering to restore high affinity binding of an isoform-selective anti-TGFβ1 antibody. *Mabs* 10, 444–452. doi:10.1080/19420862.2018.1426421

Lu, R. M., Hwang, Y. C., Liu, I. J., Lee, C. C., Tsai, H. Z., Li, H. J., et al. (2020). Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* 27, 1. doi:10.1186/s12929-019-0592-z

Lu, S., Li, Y., Wang, F., Nan, X., and Zhang, S. (2021). Leveraging sequential and spatial neighbors information by using CNNs linked with GCNs for paratope prediction. *Ieee Acm Trans. Comput. Biol. Bioinform* 19, 68–74. doi:10.1109/tcbb.2021.3083001

Ma, B., Kumar, S., Tsai, C. J., Hu, Z., and Nussinov, R. (2000). Transition-state ensemble in enzyme catalysis: Possibility, reality, or necessity? *J. Theor. Biol.* 203, 383–397. doi:10.1006/jtbi.2000.1097

Magnan, C. N., Randall, A., and Baldi, P. (2009). SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* 25, 2200–2207. doi:10.1093/bioinformatics/btp386

Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., and Taranto, A. G. (2020). Structure-based virtual screening: From classical to artificial intelligence. *Front. Chem.* 8, 343. doi:10.3389/fchem.2020.00343

Martin, K. P., Grimaldi, C., Grempler, R., Hansel, S., and Kumar, S. (2023). Trends in industrialization of biotherapeutics: A survey of product characteristics of 89 antibody-based biotherapeutics. *Mabs* 15, 2191301. doi:10.1080/19420862.2023.2191301

Medina-Ortiz, D., Contreras, S., Amado-Hinojosa, J., Torres-Almonacid, J., Asenjo, J. A., Navarrete, M., et al. (2022). Generalized property-based encoders and digital signal processing facilitate predictive tasks in protein engineering. *Front. Mol. Biosci.* 9, 898627. doi:10.3389/fmolb.2022.898627

Mehta, D., Jackson, R., Paul, G., Shi, J., and Sabbagh, M. (2017). Why do trials for alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015. *Expert Opin. Inv Drug* 26, 735–739. doi:10.1080/13543784.2017.1323868

Mieczkowski, C., Zhang, X., Lee, D., Nguyen, K., Lv, W., Wang, Y., et al. (2023). Blueprint for antibody biologics developability. *Mabs* 15, 2185924. doi:10.1080/19420862.2023.2185924

Mimoto, F., Kuramochi, T., Katada, H., Igawa, T., and Hattori, K. (2016). Fc engineering to improve the function of therapeutic antibodies. *Curr. Pharm. Biotechnol.* 17, 1298–1314. doi:10.2174/1389201017666160824161854

Miton, C. M., and Tokuriki, N. (2016). How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.* 25, 1260–1272. doi:10.1002/pro.2876

Münch, C., and Bertolotti, A. (2010). Exposure of hydrophobic surfaces initiates aggregation of diverse ALS-causing superoxide dismutase-1 mutants. *J. Mol. Biol.* 399, 512–525. doi:10.1016/j.jmb.2010.04.019

Myung, Y., Pires, D. E. V., and Ascher, D. B. (2021). CSM-AB: Graph-based antibody–antigen binding affinity prediction and docking scoring function. *Bioinformatics* 38, 1141–1143. doi:10.1093/bioinformatics/btab762

Nagano, K., and Tsutsumi, Y. (2021). Phage display technology as a powerful platform for antibody drug discovery. *Viruses* 13, 178. doi:10.3390/v13020178

Navarro, S., and Ventura, S. (2019). Computational re-design of protein structures to improve solubility. *Expert Opin. Drug Dis.* 14, 1077–1088. doi:10.1080/17460441.2019.1637413

Nelson, B., Adams, J., Kuglstatter, A., Li, Z., Harris, S. F., Liu, Y., et al. (2018). Structure-guided combinatorial engineering facilitates affinity and specificity optimization of anti-CD81 antibodies. *J. Mol. Biol.* 430, 2139–2152. doi:10.1016/j.jmb.2018.05.018

Nichols, P., Li, L., Kumar, S., Buck, P. M., Singh, S. K., Goswami, S., et al. (2015). Rational design of viscosity reducing mutants of a monoclonal antibody: Hydrophobic versus electrostatic inter-molecular interactions. *Mabs* 7, 212–230. doi:10.4161/19420862.2014.985504

Nimrod, G., Fischman, S., Austin, M., Herman, A., Keyes, F., Leiderman, O., et al. (2018). Computational design of epitope-specific functional antibodies. *Cell Rep.* 25, 2121–2131.e5. doi:10.1016/j.celrep.2018.10.081

Norman, R. A., Ambrosetti, F., Bonvin, A. M. J. J., Colwell, L. J., Kelm, S., Kumar, S., et al. (2019). Computational approaches to therapeutic antibody design: Established methods and emerging trends. *Brief. Bioinform* 21, 1549–1567. doi:10.1093/bib/bbz095

Olimpieri, P. P., Chailyan, A., Tramontano, A., and Marcatili, P. (2013). Prediction of site-specific interactions in antibody-antigen complexes: The proABC method and server. *Bioinformatics* 29, 2285–2291. doi:10.1093/bioinformatics/btt369

Olimpieri, P. P., Marcatili, P., and Tramontano, A. (2015). Tabhu: Tools for antibody humanization. *Bioinformatics* 31, 434–435. doi:10.1093/bioinformatics/btu667

Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997). Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* 48, 545–600. doi:10.1146/annurev.physchem.48.1.545

Pan, X., and Kortemme, T. (2021). Recent advances in de novo protein design: Principles, methods, and applications. *J. Biol. Chem.* 296, 100558. doi:10.1016/j.jbc.2021.100558

Pandurangan, A. P., Ochoa-Montaño, B., Ascher, D. B., and Blundell, T. L. (2017). SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 45, W229–W235. doi:10.1093/nar/gkx439

Pantazes, R. J., and Maranas, C. D. (2010). OptCDR: A general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Eng. Des. Sel.* 23, 849–858. doi:10.1093/protein/gzq061

Pereira, J., Simpkin, A. J., Hartmann, M. D., Rigden, D. J., Keegan, R. M., and Lupas, A. N. (2021). High-accuracy protein structure prediction in CASP14. *Proteins Struct. Funct. Bioinform* 89, 1687–1699. doi:10.1002/prot.26171

Pittala, S., and Bailey-Kellogg, C. (2020). Learning context-aware structural representations to predict antigen and antibody binding interfaces. *Bioinform Oxf Engl.* 36, 3996–4003. doi:10.1093/bioinformatics/btaa263

Prabakaran, R., Goel, D., Kumar, S., and Gromiha, M. M. (2017). Aggregation prone regions in human proteome: Insights from large-scale data analyses. *Proteins Struct. Funct. Bioinform* 85, 1099–1118. doi:10.1002/prot.25276

Prabakaran, R., Rawat, P., Kumar, S., and Gromiha, M. M. (2020). ANuPP: A versatile tool to predict aggregation nucleating regions in peptides and proteins. *J. Mol. Biol.* 433, 166707. doi:10.1016/j.jmb.2020.11.006

Qing, R., Hao, S., Smorodina, E., Jin, D., Zalevsky, A., and Zhang, S. (2022). Protein design: From the aspect of water solubility and stability. *Chem. Rev.* 122, 14085–14179. doi:10.1021/acs.chemrev.1c00757

Quan, L., Lv, Q., and Zhang, Y. (2016). STRUM: Structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* 32, 2936–2946. doi:10.1093/bioinformatics/btw361

Rai, B. K., Apgar, J. R., and Bennett, E. M. (2023). Low-data interpretable deep learning prediction of antibody viscosity using a biophysically meaningful representation. *Sci. Rep-uk* 13, 2917. doi:10.1038/s41598-023-28841-4

Rangel, M. A., Bedwell, A., Costanzi, E., Taylor, R. J., Russo, R., Bernardes, G. J. L., et al. (2022). Fragment-based computational design of antibodies targeting structured epitopes. *Sci. Adv.* 8, eabp9540. doi:10.1126/sciadv.abp9540

Rawat, P., Kumar, S., and Gromiha, M. M. (2018). An *in-silico* method for identifying aggregation rate enhancer and mitigator mutations in proteins. *Int. J. Biol. Macromol.* 118, 1157–1167. doi:10.1016/j.ijbiomac.2018.06.102

Rawat, P., Prabakaran, R., Kumar, S., and Gromiha, M. M. (2019). AggreRATE-pred: A mathematical model for the prediction of change in aggregation rate upon point mutation. *Bioinformatics* 31, 1439–1444. doi:10.1093/bioinformatics/btz764

Raybould, M. I. J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A. P., et al. (2019). Five computational developability guidelines for therapeutic antibody profiling. *Proc. Natl. Acad. Sci.* 116, 4025–4030. doi:10.1073/pnas.1810576116

Reetz, M. T. (2013). The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed.* 52, 2658–2666. doi:10.1002/anie.201207842

Reichert, J. M., Rosensweig, C. J., Faden, L. B., and Dewitz, M. C. (2009). Monoclonal antibody successes in the clinic. *Nat. Biotechnol.* 23, 1073–1078. doi:10.1038/nbt0905-1073

Rice, P., Longden, I., and Bleasby, A. (2000). Emboss: The European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi:10.1016/s0168-9525(00)02024-2

Ripoll, D. R., Chaudhury, S., and Wallqvist, A. (2021). Using the antibody-antigen binding interface to train image-based deep neural networks for antibody-epitope classification. *Plos Comput. Biol.* 17, e1008864. doi:10.1371/journal.pcbi.1008864

Roguska, M. A., Pedersen, J. T., Keddy, C. A., Henry, A. H., Searle, S. J., Lambert, J. M., et al. (1994). Humanization of murine monoclonal antibodies through variable domain resurfacing. *Proc. Natl. Acad. Sci. U. S. A.* 91, 969–973. doi:10.1073/pnas.91.3.969

Rosace, A., Bennett, A., Oeller, M., Mortensen, M. M., Sakhnini, L., Lorenzen, N., et al. (2022). Automated optimisation of solubility and conformational stability of antibodies and proteins. *Nat. Commun.* 14, 1937. doi:10.1038/s41467-023-37668-6

Runcie, K., Budman, D. R., John, V., and Seetharamu, N. (2018). Bi-specific and tri-specific antibodies-the next big thing in solid tumor therapeutics. *Mol. Med.* 24, 50. doi:10.1186/s10020-018-0051-4

Sakhnini, L. I., Greisen, P. J., Wiberg, C., Bozoky, Z., Lund, S., Perez, A.-M. W., et al. (2019). Improving the developability of an antigen binding fragment by aspartate substitutions. *Biochemistry-us* 58, 2750–2759. doi:10.1021/acs.biochem.9b00251

Santos, J., Pujols, J., Pallarès, I., Iglesias, V., and Ventura, S. (2020). Computational prediction of protein aggregation: Advances in proteomics, conformation-specific algorithms and biotechnological applications. *Comput. Struct. Biotechnol. J.* 18, 1403–1413. doi:10.1016/j.csbj.2020.05.026

Saurabh, S., Kalonia, C., Li, Z., Hollowell, P., Waigh, T., Li, P., et al. (2022). Understanding the stabilizing effect of histidine on mAb aggregation: A molecular dynamics study. *Mol. Pharm.* 19, 3288–3303. doi:10.1021/acs.molpharmaceut.2c00453

Sawant, M. S., Streu, C. N., Wu, L., and Tessier, P. M. (2020). Toward drug-like multispecific antibodies by design. *Int. J. Mol. Sci.* 21, 7496. doi:10.3390/ijms21207496

Schneider, C., Buchanan, A., Taddese, B., and Deane, C. M. (2021). DLAB: Deep learning methods for structure-based virtual screening of antibodies. *Bioinformatics* 38, 377–383. doi:10.1093/bioinformatics/btab660

Schoeder, C. T., Schmitz, S., Adolf-Bryfogle, J., Sevy, A. M., Finn, J. A., Sauer, M. F., et al. (2021). Modeling immunity with Rosetta: Methods for antibody and antigen design. *Biochemistry-us* 60, 825–846. doi:10.1021/acs.biochem.0c00912

Seeliger, D. (2013). Development of scoring functions for antibody sequence assessment and optimization. *Plos One* 8, e76909. doi:10.1371/journal.pone.0076909

Sever, R., Roeder, T., Hindle, S., Sussman, L., Black, K.-J., Argentine, J., et al. (2019). bioRxiv: the preprint server for biology. *Biorxiv*, 833400. doi:10.1101/833400

Shahfar, H., Du, Q., Parupudi, A., Shan, L., Esfandiary, R., and Roberts, C. J. (2022). Electrostatically driven protein–protein interactions: Quantitative prediction of second osmotic virial coefficients to aid antibody design. *J. Phys. Chem. Lett.* 13, 1366–1372. doi:10.1021/acs.jpclett.1c03669

Shaker, B., Ahmad, S., Lee, J., Jung, C., and Na, D. (2021). *In silico* methods and tools for drug discovery. *Comput. Biol. Med.* 137, 104851. doi:10.1016/j.compbiomed.2021.104851

Shan, L., Mody, N., Sormani, P., Rosenthal, K. L., Damschroder, M. M., and Esfandiary, R. (2018). Developability assessment of engineered monoclonal antibody variants with a complex self-association behavior using complementary analytical and *in silico* tools. *Mol. Pharm.* 15, 5697–5710. doi:10.1021/acs.molpharmaceut.8b00867

Sheng, Z., Bimela, J. S., Katsamba, P. S., Patel, S. D., Guo, Y., Zhao, H., et al. (2022). Structural basis of antibody conformation and stability modulation by framework somatic hypermutation. *Front. Immunol.* 12, 811632. doi:10.3389/fimmu.2021.811632

Shimba, N., Kamiya, N., and Nakamura, H. (2016). Model building of antibody–antigen complex structures using GBSA scores. *J. Chem. Inf. Model* 56, 2005–2012. doi:10.1021/acs.jcim.6b00066

Shmool, T. A., Martin, L. K., Matthews, R. P., and Hallett, J. P. (2022). Ionic liquid-based strategy for predicting protein aggregation propensity and thermodynamic stability. *Jacs Au* 2, 2068–2080. doi:10.1021/jacsau.2c00356

Sircar, A., and Gray, J. J. (2010). SnugDock: Paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *Plos Comput. Biol.* 6, e1000644. doi:10.1371/journal.pcbi.1000644

Smialowski, P., Doose, G., Torkler, P., Kaufmann, S., and Frishman, D. (2012). PROSO II – A new method for protein solubility prediction. *Febs J.* 279, 2192–2200. doi:10.1111/j.1742-4658.2012.08603.x

Smiatek, J., Jung, A., and Bluhmki, E. (2020). Towards a digital bioprocess replica: Computational approaches in biopharmaceutical development and manufacturing. *Trends Biotechnol.* 38, 1141–1153. doi:10.1016/j.tibtech.2020.05.008

Smith, S. (1996). Ten years of Orthoclone OKT3 (muromonab-CD3): A review. *J. Transpl. Coord.* 6, 109–119; quiz 120-121. doi:10.7182/prtr.1.6.3.8145l3u185493182

Somani, S., Jo, S., Thirumangalathu, R., Rodrigues, D., Tanenbaum, L. M., Amin, K., et al. (2021). Toward biotherapeutics formulation composition engineering using site-identification by ligand competitive saturation (SILCS). *J. Pharm. Sci.* 110, 1103–1110. doi:10.1016/j.xphs.2020.10.051

Sormanni, P., Aprile, F. A., and Vendruscolo, M. (2015). Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. *Proc. Natl. Acad. Sci.* 112, 9902–9907. doi:10.1073/pnas.1422401112

Sormanni, P., Aprile, F. A., and Vendruscolo, M. (2018). Third generation antibody discovery methods: In silico rational design. *Chem. Soc. Rev.* 47, 9137–9157. doi:10.1039/c8cs00523k

Starr, C. G., and Tessier, P. M. (2019). Selecting and engineering monoclonal antibodies with drug-like specificity. *Curr. Opin. Biotech.* 60, 119–127. doi:10.1016/j.copbio.2019.01.008

Steinbrecher, T., Zhu, C., Wang, L., Abel, R., Negron, C., Pearlman, D., et al. (2017). Predicting the effect of amino acid single-point mutations on protein stability—large-scale validation of MD-based relative free energy calculations. *J. Mol. Biol.* 429, 948–963. doi:10.1016/j.jmb.2016.12.007

Strohl, W. R. (2018). Current progress in innovative engineered antibodies. *Protein Cell* 9, 86–120. doi:10.1007/s13238-017-0457-8

Svilenov, H. L., Arosio, P., Menzen, T., Tessier, P., and Sormanni, P. (2023). Approaches to expand the conventional toolbox for discovery and selection of antibodies with drug-like physicochemical properties. *Mabs* 15, 2164459. doi:10.1080/19420862.2022.2164459

Swinney, D. C., and Anthony, J. (2011). How were new medicines discovered? *Nat. Rev. Drug Discov.* 10, 507–519. doi:10.1038/nrd3480

Tartaglia, G. G., Cavalli, A., Pellarin, R., and Caflisch, A. (2004). The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* 7, 1939–1941. doi:10.1110/ps.04663504

Thorsteinson, N., Gunn, J. R., Kelly, K., Long, W., and Labute, P. (2021). Structure-based charge calculations for predicting isoelectric point, viscosity, clearance, and profiling antibody therapeutics. *Mabs* 13, 1981805. doi:10.1080/19420862.2021.1981805

Thorsteinson, N., Comeau, S. R., and Kumar, S. (2023). Structure-based optimization of antibody-based biotherapeutics for improved developability: A practical guide for molecular modelers. *Methods Mol. Biol.* doi:10.1007/978-1-0716-2609-2

Tilegenova, C., Izadi, S., Yin, J., Huang, C. S., Wu, J., Ellerman, D., et al. (2019). Dissecting the molecular basis of high viscosity of monospecific and bispecific IgG antibodies. *Mabs* 12, 1692764. doi:10.1080/19420862.2019.1692764

Tiller, T., Schuster, I., Deppe, D., Siegers, K., Strohner, R., Herrmann, T., et al. (2013). A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. *Mabs* 5, 445–470. doi:10.4161/mabs.24218

Tiller, K. E., Chowdhury, R., Li, T., Ludwig, S. D., Sen, S., Maranas, C. D., et al. (2017). Facile affinity maturation of antibody variable domains using natural diversity mutagenesis. *Front. Immunol.* 8, 986. doi:10.3389/fimmu.2017.00986

Tomar, D. S., Kumar, S., Singh, S. K., Goswami, S., and Li, L. (2016). Molecular basis of high viscosity in concentrated antibody solutions: Strategies for high concentration drug product development. *Mabs* 8, 216–228. doi:10.1080/19420862.2015.1128606

Tomar, D. S., Singh, S. K., Li, L., Broulidakis, M. P., and Kumar, S. (2018). *In silico* prediction of diffusion interaction parameter (kD), a key indicator of antibody solution behaviors. *Pharm. Res.* 35, 193. doi:10.1007/s11095-018-2466-6

Tomar, D. S., Licari, G., Bauer, J., Singh, S. K., Li, L., and Kumar, S. (2021). Stress-dependent flexibility of a full-length human monoclonal antibody: Insights from molecular dynamics to support biopharmaceutical development. *J. Pharm. Sci.* 111, 628–637. doi:10.1016/j.xphs.2021.10.039

Townsend, S., Fennell, B. J., Apgar, J. R., Lambert, M., McDonnell, B., Grant, J., et al. (2015). Augmented Binary Substitution: Single-pass CDR germ-lining and stabilization of therapeutic antibodies. *Proc. Natl. Acad. Sci.* 112, 15354–15359. doi:10.1073/pnas.1510944112

Trask, A., Hill, F., Reed, S., Rae, J., Dyer, C., and Blunsom, P. (2018). Neural arithmetic logic units. *arXiv*. doi:10.48550/arxiv.1808.00508

Trovato, A., Seno, F., and Tosatto, S. C. E. (2007). The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* 20, 521–523. doi:10.1093/protein/gzm042

Tsumoto, K., and Kuroda, D. (2022). in *Computer-aided antibody design*. Editors K. Tsumoto and D. Kuroda (Springer-Verlag New York Inc).

Tubiana, J., Schneidman-Duhovny, D., and Wolfson, H. J. (2022). ScanNet: An interpretable geometric deep learning model for structure-based protein binding site prediction. *Nat. Methods* 19, 730–739. doi:10.1038/s41592-022-01490-7

ULC, C. C. G. (2021). *Molecular operating environment (MOE)*. Montreal, QC, Canada. 2019.01. 1010 Sherbooke St. West, Suite #910, H3A 2R7

Valldorf, B., Hinz, S. C., Russo, G., Pekar, L., Mohr, L., Klemm, J., et al. (2022). Antibody display technologies: Selecting the cream of the crop. *Biol. Chem.* 403, 455–477. doi:10.1515/hsz-2020-0377

van der Kant, R., Karow-Zwick, A. R., Durme, J. V., Blech, M., Gallardo, R., Seeliger, D., et al. (2017). Prediction and reduction of the aggregation of monoclonal antibodies. *J. Mol. Biol.* 429, 1244–1261. doi:10.1016/j.jmb.2017.03.014

van der Kant, R., Bauer, J., Karow-Zwick, A. R., Kube, S., Garidel, P., Blech, M., et al. (2019). Adaption of human antibody λ and κ light chain architectures to CDR repertoires. *Protein Eng. Des. Sel.* 32, 109–127. doi:10.1093/protein/gzz012

Vatsa, S. (2022). *In silico* prediction of post-translational modifications in therapeutic antibodies. *Mabs* 14, 2023938. doi:10.1080/19420862.2021.2023938

Vecchio, A. D., Deac, A., Liò, P., and Veličković, P. (2021). Neural message passing for joint paratope-epitope prediction. *Arxiv*.

Walsh, I., Seno, F., Tosatto, S. C. E., and Trovato, A. (2014). PASTA 2.0: An improved server for protein aggregation prediction. *Nucleic Acids Res.* 42, W301–W307. doi:10.1093/nar/gku399

Wang, Q., Chung, C., Chough, S., and Betenbaugh, M. J. (2018). Antibody glycoengineering strategies in mammalian cells. *Biotechnol. Bioeng.* 115, 1378–1393. doi:10.1002/bit.26567

Wang, Q., Chen, Y., Park, J., Liu, X., Hu, Y., Wang, T., et al. (2019). Design and production of bispecific antibodies. *Antibodies* 8, 43. doi:10.3390/antib8030043

Wei, G.-W. (2019). Protein structure prediction beyond AlphaFold. *Nat. Mach. Intell.* 1, 336–337. doi:10.1038/s42256-019-0086-4

Weitzner, B. D., Jeliazkov, J. R., Lyskov, S., Marze, N., Kuroda, D., Frick, R., et al. (2017). Modeling and docking of antibody structures with Rosetta. *Nat. Protoc.* 12, 401–416. doi:10.1038/nprot.2016.180

Wilman, W., Wróbel, S., Bielska, W., Deszynski, P., Dudzic, P., Jaszczyszyn, I., et al. (2022). Machine-designed biotherapeutics: Opportunities, feasibility and advantages of deep learning in computational antibody discovery. *Brief. Bioinform* 23, bbac267. doi:10.1093/bib/bbac267

Wittmund, M., Cadet, F., and Davari, M. D. (2022). Learning epistasis and residue coevolution patterns: Current trends and future perspectives for advancing enzyme engineering. *ACS Catal.* 12, 14243–14263. doi:10.1021/acscatal.2c01426

Xu, Y., Wang, D., Mason, B., Rossomando, T., Li, N., Liu, D., et al. (2018). Structure, heterogeneity and developability assessment of therapeutic antibodies. *Mabs* 11, 239–264. doi:10.1080/19420862.2018.1553476

Yadav, S., Sreedhara, A., Kanai, S., Liu, J., Lien, S., Lowman, H., et al. (2011). Establishing a link between amino acid sequences and self-associating and viscoelastic behavior of two closely related monoclonal antibodies. *Pharm. Res.* 28, 1750–1764. doi:10.1007/s11095-011-0410-0

Yadav, S., Laue, T. M., Kalonia, D. S., Singh, S. N., and Shire, S. J. (2012). The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. *Mol. Pharm.* 9, 791–802. doi:10.1021/mp200566k

Yamashita, K., Ikeda, K., Amada, K., Liang, S., Tsuchiya, Y., Nakamura, H., et al. (2014). Kotai antibody builder: Automated high-resolution structural modeling of antibodies. *Bioinformatics* 30, 3279–3280. doi:10.1093/bioinformatics/btu510

Yan, X. C., Sanders, J. M., Gao, Y.-D., Tudor, M., Haidle, A. M., Klein, D. J., et al. (2020). Augmenting hit identification by virtual screening techniques in small molecule drug discovery. *J. Chem. Inf. Model* 60, 4144–4152. doi:10.1021/acs.jcim.0c00113

Yang, W., Tan, P., Fu, X., and Hong, L. (2019). Prediction of amyloid aggregation rates by machine learning and feature selection. *J. Chem. Phys.* 151, 084106. doi:10.1063/1.5113848

Yu, X., Tsibane, T., McGraw, P. A., House, F. S., Keefer, C. J., Hicar, M. D., et al. (2008). Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* 455, 532–536. doi:10.1038/nature07231

Zhang, C., Samad, M., Yu, H., Chakroun, N., Hilton, D., and Dalby, P. A. (2018). Computational design to reduce conformational flexibility and aggregation rates of an antibody fab fragment. *Mol. Pharm.* 15, 3079–3092. doi:10.1021/acs.molpharmaceut.8b00186

Zhang, Y., Wu, L., Gupta, P., Desai, A. A., Smith, M. D., Rabia, L. A., et al. (2020). Physicochemical rules for identifying monoclonal antibodies with drug-like specificity. *Mol. Pharm.* 17, 2555–2569. doi:10.1021/acs.molpharmaceut.0c00257

Zibaee, S., Makin, O. S., Goedert, M., and Serpell, L. C. (2007). A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, Abeta, and tau using the amino acid sequence alone. *Protein Sci.* 16, 906–918. doi:10.1110/ps.062624507

Zurdo, J. (2013). Developability assessment as an early de-risking tool for biopharmaceutical development. *Pharm. Bioprocess* 1, 29–50. doi:10.4155/pbp.13.3

Glossary

Ab	Antibody
ADA	Anti-drug antibodies
ADCC	Antibody-dependent cell-mediated cytotoxicity
ADCP	Antibody-dependent cellular phagocytosis
Ag	Antigen
AI	Artificial intelligence
Ala	Alanine
APR	Aggregation-prone region
BCR	B-cell repertoire
CDC	Complement-dependent cytotoxicity
CDR	Complementarity determining region
CMC	Chemistry, manufacturing, and control
CNN	Convolutional neural networks
CQA	Clinical Quality Attributes
DAbI	Discovery of antibodies *in silico*
DENIS	DEvelopability Navigator *In Silico*
DLAB	Deep Learning for AntiBodies
DLS	Dynamic light scattering
Fab	Fragment antigen binding
FAIR	Findable, accessible, interoperable, and reusable
FDA	Food and Drug Administration
Fv	Fragment variable
FW	Framework
GAN	Generalized adversarial network
HC	Heavy chain
HCDR1-3	Heavy-chain complementarity determining regions 1–3
HCPF	High-concentration protein formulations
HIC	Hydrophobic interaction chromatography
HTTP	High-throughput
IgG	Immunoglobulin G
LC	Light chain
LCDR1-3	Light-chain complementarity determining regions 1–3
LO	Lead optimization
mAb	Monoclonal antibody
MaSIF	Molecular surface interaction fingerprints
MD	Molecular dynamics
MFI	Micro-flow imaging
MHC	Major histocompatibility complex
ML	Machine learning
MM-GBSA	Molecular mechanics—generalized Born solvent accessibility
MOE	Molecular Operating Environment
NBE	New biologic entity
NTC	Novel therapeutic concept
pI	Isoelectric point
PTM	Post-translational modification
QSAR	Quantitative structure–activity relationship
QSPR	Quantitative structure–property relationship
R&D	Research and development
RNN	Recurrent neural networks
RTP	Research target profile
ScFv	Single-chain fragment variable
SCM	Spatial charge map
SEC	Size exclusion chromatography
SINS	Self-interaction nanoparticle spectroscopy
TGFβ1	Transforming growth factor β1
UF/DF	Ultrafiltration/diafiltration
VH	Heavy-chain variable region
VK	Light-chain variable region (kappa isotype)
VL	Light-chain variable region (lambda isotype)