Computational Strategies in Therapeutic Antibody Development: Current Techniques and Future Directions

Abu Junaid Siddiqui[1]

[a*]Department of Bioengineering, Faculty of Engineering, Integral University, Lucknow, India

abujunaid@iul.ac.in

Prof. (Dr.) Alvina Farooqui [1*]

Professor & Head of Department of Bioengineering, Faculty of Engineering, Integral University, Lucknow, India

alvina@iul.ac.in

Prof. (Dr.) Alvina Farooqui [1*] are corresponding author.

Abstract

Antibodies are specialized proteins that identify and bind to specific molecular targets, playing a central role in the adaptive immune system. In autoimmune conditions, however, they may mistakenly target the body's own healthy tissues. Owing to their remarkable binding specificity and adaptability, antibodies have become the most prominent category of biotherapeutic agents, with monoclonal antibodies comprising a significant portion of the top-selling drugs globally. Recent developments in computational protein modeling and design are significantly contributing to the advancement of antibody-based therapies. These antibody-focused computational approaches are increasingly benefiting from large-scale datasets generated through next-generation sequencing technologies. Additionally, they are being applied to newer antibody formats, such as nanobodies. This review offers a comprehensive summary of current databases, established tools, and innovative methodologies in computational antibody research, with a focus on their relevance to therapeutic antibody design and engineering.

Keywords: homology modeling, therapeutic antibodies, molecular docking, antibody–antigen interactions, bioinformatics databases Introduction

I. INTRODUCTION

Antibodies, also known as immunoglobulins, are essential components of the adaptive immune system. They identify and bind to specific molecular structures—known as antigens—on potentially harmful entities for elimination [1]. In autoimmune disorders, however, these proteins may mistakenly target endogenous molecules, leading to immune responses against healthy tissues [2]. Antibodies have evolved to recognize a broad spectrum of antigenic surfaces, making them highly adaptable binding agents [3].

Due to their specificity and adaptability, antibodies have become a cornerstone of therapeutic interventions and currently represent the largest category within the biotherapeutic market. Among the top-selling drugs globally are five monoclonal antibodies: adalimumab and infliximab (targeting TNFα), rituximab (anti-CD20), bevacizumab (anti-VEGF), and trastuzumab (anti-HER2/neu), whose clinical impact continues to grow [4]. As the demand for effective antibody therapies increases, more efficient discovery and development strategies are needed. Computational approaches offer promising alternatives to traditional, labor-intensive experimental protocols, enabling rapid design and screening of therapeutic candidates [5].

Several established bioinformatics tools—such as homology modeling [6,7], protein–protein docking [8,9], and interface prediction algorithms [10]—are now routinely used in the rational design of antibodies [11–13]. In addition, computational methods are being developed to evaluate critical features like immunogenic potential [14] and biophysical stability [15]. The availability of extensive datasets, including structural [16], sequence [17], and experimental data [18–21], has significantly advanced data-driven antibody design.

A transformative development in this context has been the application of next-generation sequencing (NGS) to characterize B-cell receptor repertoires [22]. NGS enables high-throughput profiling of antibody sequences, capturing millions of variants from the theoretically vast antibody diversity in humans—estimated at 10^{12} to 10^{15} unique sequences [23,24]. Analysis of these repertoires reveals biases and patterns that reflect the natural variability and evolution of the immune system [25]. Such insights are invaluable for benchmarking therapeutic antibody candidates [13] and designing biologically inspired display libraries [26].

Moreover, the growing arsenal of computational tools is now being extended to emerging antibody formats, such as nanobodies, which exhibit favorable biophysical traits like high solubility, stability, and reduced immunogenicity [27]. As a result, computational antibody modeling has matured into a robust discipline, capable of supporting a wide array of therapeutic development initiatives.

This review provides an organized summary of current databases, algorithms, and tools used in computational antibody research. Emphasis is placed on their application in antibody design, structural prediction, and emerging strategies aimed at therapeutic innovation.

II. Antibody Structure, Function, and Therapeutic Formats

Immunoglobulins (antibodies) are synthesized by B lymphocytes in jawed vertebrates and serve as either membrane-bound B-cell receptors or secreted soluble antibodies. Each of the estimated 5×10^9 B cells in the human body produces a unique antibody variant through somatic recombination of variable (V), diversity (D), joining (J), and constant (C) gene segments [23,28,29].

Heavy chains are assembled using V, D, J, and C gene segments from the heavy chain locus, while light chains are formed from V, J, and C segments located at the κ or λ light chain loci. These chains combine to form five major antibody isotypes: IgG, IgD, IgE (monomeric forms), IgA (dimeric), and IgM (pentameric) [30]. The IgG isotype, which predominates in blood circulation and has the most therapeutic relevance, contains one crystallizable fragment (Fc) and two antigen-binding fragments (Fabs).

Each Fab fragment consists of a heavy (VH) and a light (VL) variable domain, which interact with a specific site on the antigen known as the epitope. These variable domains each contain three highly variable loops—termed complementarity-determining regions (CDRs)—which make up the antigen-binding site, or paratope. During antigen exposure, B cells undergo somatic hypermutation within the CDRs, a process known as affinity maturation, resulting in higher-affinity antibodies [31]. Combined with the sequence variability introduced by V(D)J recombination, this process contributes to a theoretical antibody diversity of up to 10^{15} variants [23,24].

Although the CDRs are hypervariable in sequence, most (excluding CDRH3) adopt a limited range of backbone conformations, known as canonical structures [32]. CDRH3 is particularly diverse both in sequence and structure [33], and it plays a crucial role in antigen recognition [34,35]. Consequently, CDR regions—especially CDRH3—are often the primary targets in antibody engineering for monoclonal antibody (mAb) development [36,37].

Despite their advantages, conventional mAbs (~150 kDa) often exhibit poor tissue penetration. To address this, smaller antibody derivatives and engineered formats have been developed. These include single-chain variable fragments (scFv), composed of linked VH and VL domains, and other modular structures such as diabodies and minibodies [38–40]. Additionally, bispecific and multispecific antibodies—engineered to bind two or more distinct antigens—are gaining attention in cancer therapy [41].

Another innovation in antibody engineering is the development of single-domain antibodies, also known as nanobodies or VHHs, which are naturally found in camelids and certain shark species. Nanobodies are approximately half the size of a standard antibody domain but retain comparable specificity and affinity. Their high solubility, thermal stability, and lower immunogenic potential make them promising therapeutic agents [27]. Notably, caplacizumab, the first nanobody-based drug, received regulatory approval in 2018 [42]

Antibody Databases

The effectiveness of computational antibody research depends on access to well-curated and diverse datasets. Several resources provide detailed information on therapeutic antibodies, including the Therapeutic Antibody Database (TABS) and the SAbDab-Therapeutic Antibodies database [13]. These repositories can be categorized based on their content—sequence, structure, or experimental data—with some integrating all three types (see Table 1).

Most databases include both conventional antibodies and nanobodies. However, there are specialized resources like sdAb-DB that focus exclusively on single-domain antibodies [58].

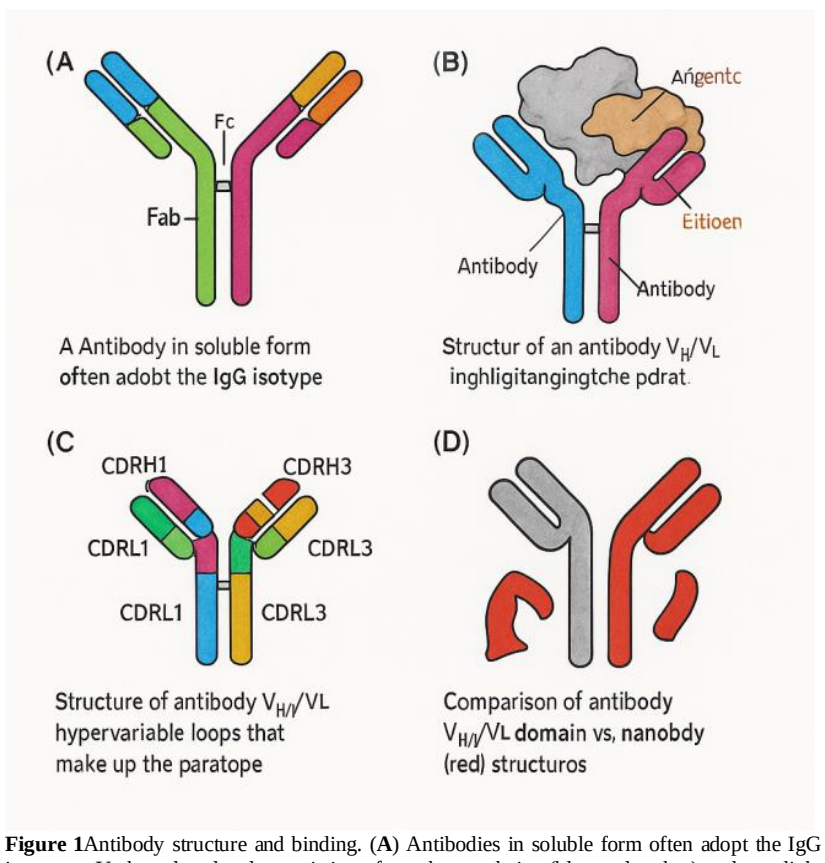

Figure 1 Antibody structure and binding. (**A**) Antibodies in soluble form often adopt the IgG isotype, a Y-shaped molecule consisting of two heavy chains (blue and amber) and two light chains (green and magenta). Each IgG molecule can be subdivided into an Fc and two Fab fragments through papain cleavage of the (hinge) region between these. At each end of a Fab fragment is a variable domain (VH/VL) involved in antigen binding. (**B**) Structure of an antibody VH (blue)/V(magenta) in complex with cognate antigen (grey). The antibody paratope (light green) and antigen epitope (light brown) are highlighted. (**C**) Structure of an antibody VH (blue)/V(magenta) highlighting the six hypervariable loops that make up the paratope; CDRH1 (white), CDRH2 (red), CDRH3 (amber), CDRL1 (green), CDRL2 (light blue), CDRL3 (yellow). (**D**) Comparison of antibody VH/VL domain (grey) and nanobody (red) structures. Nanobodies are devoid of the light chain, thus all the binding is mediated by theVH-homologous portion including its three CDR loops (CDRH1–3).

Sequence Databases

The International Immunogenetics Information System (IMGT) is the primary reference source for germline antibody sequences and is widely used for assigning gene segments in recombined antibodies [46]. Other platforms, such as Abysis [47] and DIGIT [48], typically store recombined variable region sequences (VH and VL), often obtained from repositories like the European Nucleotide Archive (ENA) [60] and the National Center for Biotechnology Information (NCBI) [61].

Databases like DIGIT and Abysis typically contain around 10^5 sequences, including many from artificially engineered antibodies. These sequences are generally high-quality, originating from individual submissions using methods such as Sanger sequencing. In contrast, high-throughput repositories such as iReceptor [49] and Observed Antibody Space [17] aggregate large-scale datasets from NGS experiments, often encompassing more than 10^8 sequences.

NGS-derived sequences can carry inherent error rates due to the scale and speed of data generation [62]. To mitigate this, databases like Observed Antibody Space provide annotations for predicted sequencing errors [62]. Additionally, these sequences often include metadata such as CDR region annotations, standardized numbering schemes, and—when available—details about the immune state of the donor at the time of sampling.

Most NGS repositories currently offer only unpaired heavy and light chains. However, advancements in paired sequencing technologies are expected to make paired chain datasets more readily accessible in the near future [63,64], which will further enhance computational modeling and therapeutic design capabilities. **Structure Databases**

The Protein Data Bank (PDB) serves as the principal global repository for three-dimensional (3D) structural data of proteins [65]. Several specialized tools extract antibody-specific fragments or data from the PDB:

- **PyIgClassify** categorizes CDR loops into canonical classes [51].
- **PCLICK** gathers detailed antibody–antigen interaction data [50].
- Full antibody structures are reachable via **IMGT/3D-Structure-DB** [66], **SAbDab** (Structural Antibody Database) [16], **Abysis** [47], and **AbDb** [52].

As of now, approximately 3,500 structures in the PDB include at least one antibody or nanobody chain, out of a total of ~150,000 entries. SAbDab offers a downloadable weekly-updated dataset, ideal for modelling or docking projects. Meanwhile, Abysis and SAbDab enable structure retrieval by sequence queries or classification of CDR canonical forms. The **Immune Epitope Database (IEDB)** also integrates structural data with experimentally identified epitope information [18].

Experimental Databases

To extend structural and sequence insights, various databases provide experimental measurements relevant to antibody binding:

- The **IEDB** includes epitope-specific antibody sequences linked to structural data [67].
- Binding affinity details can be found in SAbDab and in the broader **PDBBind** database [54].
- Targeted data such as mutation-driven changes in affinity are cataloged in **Ab-Bind** (covering 1,101 mutations in 32 antibody complexes) [19].
- The **SKEMPI** database provides binding energy changes for diverse protein complexes, not limited to antibodies [55].

Database Name	Link	Description	Reference
TABS (commercial use)	TABS	Repository of approved therapeutic antibodies	n/a
SAbDab - therapeutic antibodies	SAbDab	Collection of therapeutic antibodies with structural data	[13]
PCLICK	PCLICK	Database of antibody-antigen binding clusters	[50]
Andrew Martin's Antibody Resources	Link	Compilation of antibody-related bioinformatics tools and resources	[43]
AAAAA	AAAAA	Educational resources on antibody structures and engineering	[43]
AbMiner	AbMiner	Database providing monoclonal antibody data	[44]
Igpdb	Igpdb	Archive of inferred germline immunoglobulin variants	[45]
IMGT®	IMGT	Authoritative database for immunoglobulin gene sequences	[46]
Abysis (commercial license)	Abysis	Combines sequence and structure information of antibodies	[47]
DIGIT	DIGIT	Antibody sequence analysis tool	[48]
IReceptor	IReceptor	Platform for sharing and querying B-cell receptor NGS data	[49]
Observed Antibody Space	OAS	Repository of antibody and BCR sequences obtained via NGS	[17]
SystemsDB (commercial license)	SystemsDB	Repository for antibody and TCR sequence data from high-throughput sequencing	n/a
PyIgClassify	PyIgClassify	Canonical class database for CDR loop conformations	[51]
Structural Antibody Database (SAbDab)	SAbDab	Automatically updated antibody/nanobody structure database	[16]
AbDb	AbDb	Comprehensive database of antibody 3D structures	[52]
Immune Epitope Database	IEDB	Manually curated repository of immune epitope data	[18]
AntigenDB	AntigenDB	Resource for antigenic proteins	[53]
PDBBind	PDBBind	Protein-ligand binding affinity data from PDB	[54]
Ab-Bind	Ab-Bind	Database of mutations affecting antibody binding affinity	[55]
SKEMPI	SKEMPI	Database for mutations influencing non-antibody protein interactions	[55, 56]
Non-redundant Nanobody Database	Article	Curated non-redundant structural database of nanobodies	[57]
SAbDab-Nano	SAbDab-Nano	Nanobody-specific extension of the SAbDab structure database	[58]
Institute of Analysis and Collection of Nanobodies	IACN	Database with nanobody sequences and structural models	[59]

Table 1. Databases containing information on antibody and nanobody structure and sequence. Most of the databases are free for academic use. In cases where the authors made it clear that a commercial version is available, this is indicated next to the database name. In some cases, such as IMGT or SKEMPI, conditions for non-commercial reuse are defined. In such cases, the authors of the respective databases should be contacted for details on commercial re-use of their material. Example contents of the databases are summarized in Supplementary Section 1. An up-to date list of antibody-related database resources is maintained at http://naturalantibody.com/tools

III. Computational Approaches for Antibody Engineering

Bioinformatics tools build upon the wealth of antibody data to support engineering endeavours throughout therapeutic development (see Table 2). Computational methods assist both during Lead Identification—where initial candidates are found—and Lead Optimization—where candidates are refined. These tools help evaluate binding strength, stability, immunogenicity, and other critical attributes before moving forward to clinical testing.

Antibody Numbering

A foundational step in computational antibody characterization is assigning sequence positions using standardized numbering frameworks (Table 2A). Nucleotide sequences of variable domains are first translated and aligned to germline gene references (e.g., via **IgBLAST** [68] or **IMGT VQuest** [69]), identifying V, D, and J gene usage. This alignment facilitates mapping residues into standardized numbering systems—like Kabat [152], Chothia [32], or IMGT [153]. Tools such as **ANARCI** [83], **Abnum** [82], and **AbRSA** [81] automate this process, enabling consistent identification of framework and CDR regions essential for subsequent modeling and prediction.

A. Antibody Annotation/Numbering Tools

Tool Name	Function	Link	Reference
IgBLAST	Processes raw antibody data	https://www.ncbi.nlm.nih.gov/igblast/	[68]
IMGT V-Quest	Raw data sequence processing	http://www.imgt.org/IMGTindex/V-QUEST.php	[69]
MiXCR	Analyzes immune sequencing data	https://mixcr.readthedocs.io/en/master/	[70]
Immcantation	Antibody repertoire data processing	https://immcantation.readthedocs.io	[71, 72]
IgReC	Constructs immune repertoires	https://yana-safonova.github.io/ig_repertoire_constructor/	[73]
ImmuneD	Analyzes	https://bitbucket.org/Imm	[74]

iversity	immune repertoire diversity	unediversity/ImmuneDiversity/	
IMSEQ	Preprocesses immune sequencing data	http://www.imtools.org/	[75]
Partis	V(D)J inference and clonal clustering	https://github.com/psathyrella/partis	[76]
IGOR	Models B cell receptor generation	https://github.com/mikemc/igor	[77]
Vidjil	Immune repertoire visualization	http://www.vidjil.org/	[78, 79]
ImmuneDB	Immune sequencing data analysis	https://immunedb.readthedocs.io/en/latest/	[80]
AbRSA	Numbering system for antibodies	http://cao.labshare.cn/AbRSA/	[81]
Abnum	Antibody numbering resource	http://www.bioinf.org.uk/abs/abnum/	[82]
ANARCI	Numbering scheme classification	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/ANARCI.php	[83]

B. Structural Antibody Modelling Tools

Tool	Function	Link	Reference
AbodyBuilder	Comprehensive modeling of antibody variable regions	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Modelling.php	[84]
LYRA	Modeling of full variable regions	http://www.cbs.dtu.dk/services/LYRA/index.php	[85]
PIGS	Full variable region modeling	https://cassandra.med.unito.it/pigspro/	[86]
Kotai Antibody Builder	Complete variable region modeling	http://kotaiab.org/	[87]
Rosetta Antibody	Rosetta-based full variable region modeling	https://rosie.rosettacommons.org/antibody	[88, 89]
BIOVIA	General modeling tool including antibody support	https://www.3dsbiovia.com/	[90]
MoFvAb	Full variable region modeling	-	[91]
WAM	Antibody variable modeling	-	[92]
BioLuminate	Full variable region modeling via Schrödinger	https://www.schrodinger.com/products/bioluminate	[93]
MOE	Modeling of antibody variable regions	https://www.chemcomp.com/	[94]
ABGEN	Antibody modeling tool	-	[95]
AbPredi	Rosetta-based modeling method	http://abpredict.weizmann.ac.il/bin/steps	[96]
SmrToAntibody	Complete antibody modeling	https://www.macromoltek.com/	[97]
PEARS	Prediction of antibody side chains	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/PEARS.php	[98]
H3LoopPred	Specific prediction of H3 loop	-	[99]
SCWRL	Predicts side chain conformations	http://dunbrack.fccc.edu/scwrl4/	[100]
BetaScpWeb	Predicts side chain placement	http://voronoi.hanyang.ac.kr/betascpweb	[101]
SPHINX	Ab initio loop prediction	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Sphinx.php	[102]
FREAD	Database search-based loop modeling	http://opig.stats.ox.ac.uk/webapps/fread/php	[103]
PLOP	Predicts antibody loop regions	http://www.jacobsonlab.org/plop_manual/plop_overview.htm	[104]
Chothia Canonical	Assigns loop structures based on Chothia rules	http://www.bioinf.org.uk/abs/chothia.html	[105]
SCALOP	CDR classification and structure assignment	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/SCALOP.php	[106]
Roche VH/VL orientation	Determines VH/VL orientation	Part of Rosetta Suite	[107]
Rosetta VH/VL orientation	Models VH/VL orientation	Part of Rosetta Suite	[108]
ABangle	Defines VH/VL orientation angle	http://opig.stats.ox.ac.uk/webapps/abangle/index.html	[109]

Computational Tools for Antibody–Antigen Interaction Prediction and Design

C. Antibody–Antigen Interface Prediction

Tool/Platform	Function	Access Link	Reference
Antibody i-Patch	Predicts paratope regions	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/ABipatch.php	[110]
Paratome	Predicts paratope regions	http://ofranservices.biu.ac.il/site/services/paratome/	[111]
ProABC	Predicts paratope regions	http://circe.med.uniroma1.it/proABC/	[112]
Parapred	Predicts paratope regions	https://github.com/eliberis/parapred	[113]
AntibodyInterfacePrediction	Predicts paratope regions	https://github.com/sebastiananderlaku/AntibodyInterfacePrediction	[114]
AG-FAST-Parapred	Paratope predictor	-	[115]
ISMBL AB-PPI	Predicts protein contacts	http://ismblab.genomics.sinica.edu.tw/predict-	[3]

		ppi?pred=PPI	
Rapberger et al. 2007	Epitope prediction	-	[116]
PEASE	Epitope prediction	http://ofranservices.biu.ac.il/site/services/epitope/index.html	[117, 118]
PpiPred	Epitope prediction	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/PpiPred.php	[119]
Jespersen et al.	Epitope prediction	-	[120]
EpiScore	Epitope prediction	-	[121]
MabTope	Epitope prediction	-	[122]
ASEP	Epitope prediction	-	[123]
BEPAR	Epitope prediction	-	[124]
ABEPAR	Epitope prediction	-	[125]
ClusPro	Antibody docking	https://cluspro.bu.edu/login.php	[8, 126]
Surfit	Antibody docking	https://sysimm.ifrec.osaka-u.ac.jp/docking/main/	[127]
SnugDock	Antibody docking	http://rosie.graylab.jhu.edu/snug_dock	[128]
FRODOCK	Antibody docking	http://frodock.chaconlab.org/	[129]
DockSorter	Docking (not Ab-specific)	http://www.stats.ox.ac.uk/~krawczyk/dockingsupp.html	[110]
Hex	Docking (not Ab-specific)	http://hex.loria.fr/	[–]
ZDOCK	Docking (not Ab-specific)	https://zdock.umassmed.edu/	[130]
HADDOCK	Docking (not Ab-specific)	https://haddock.science.uu.nl/services/HADDOCK2.2/	[131, 132]
ATTRACT	Docking (not Ab-specific)	http://www.attract.ph.tum.de/services/ATTRACT/attract.html	[133]
GRAMM-X	Docking (not Ab-specific)	http://vakser.compbio.ku.edu/resources/gramm/grammx/	[134]
pyDockWeb (pyDock, FTDock)	Docking (not Ab-specific)	https://life.bsc.es/pid/pydockweb	[135]
Swarmdock	Docking (not Ab-specific)	https://bmm.crick.ac.uk/~svr6/swc-bmm-swarmdock	[136]
PatchDock	Docking (not Ab-specific)	https://bioinfo3d.cs.tau.ac.il/PatchDock/	[137, 138]

D. Antibody Design

Tool/Platform	Function	Access Link	Reference
OPTCDR	Design Protocol	http://www.maranasgroup.com/submission/OptCDR_2.htm	[139]
OPTMAVEN	Design Protocol	https://github.com/maranasgroup/OptMAVEn_2.0	[140, 141]
Rosetta AntibodyDesign	Design Protocol	https://www.rosettacommons.org/docs/latest/application_documentation/antibody/RosettaAntibodyDesign	[142]
AbDesign	Design Protocol	https://www.rosettacommons.org/node/9206	[12, 143]

Table E Antibody Design

Tools for Humanization and Developability in Pharmaceutical Applications

Tool/Platform	Primary Use	Access Link	Citation
Humanness Score Evaluator	Humanization	http://www.bioinf.org.uk/abs/shab/	[14]
Humanizer	Humanization	https://drive.google.com/file/d/1seCQYMlMG4_oC1-0EjDhZHMt9D-l8R5/view?usp=sharing	[141]
Tabhu	Humanization	http://circe.med.uniroma1.it/tabhu/	[144]
Human String Content	Humanization	Not available	[145]
Human String Content (Alternate)	Humanization	Not available	[145]
T20 Score	Humanization	https://dm.lakepharma.com/bioinformatics/	[146]
codaH	Humanization	Not available	[147]
Developability Index	Developability	Not available	[148]
Delayed Heavy Chain Retention Predictor	Developability	Not available	[149]
Therapeutic Antibody Profiling Tool	Developability	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/TAP.php	[13]
Lonza Tool	Developability	Not available	[15]

Antibody Modeling

Antibody homology modelling generates 3D structures from amino acid sequences. Conservation across framework regions and canonical CDR loops makes these models highly reliable [7]. Typically, modeling involves:

1. **Template selection** for heavy and light chains.
2. **VH-VL orientation** alignment.
3. **CDR loop modeling**, which is routine for canonical loops but complex for CDRH3 (necessitating ab initio methods or hybrid approaches like Sphinx [102]).
4. **Side-chain placement**, refined by tools like SCWRL [100] or antibody-specific PEARS [98].
5. **Energetic refinement** using platforms such as Rosetta [89].

Antibody Modeling: Five-Step Process and Available Tools

Antibody structure prediction typically involves a five-step pipeline (Figure 2A). The process begins with the selection of a suitable framework template, which serves as the structural base for grafting complementarity-determining regions (CDRs). This step usually involves identifying high sequence similarity in known antibody structures for both heavy (VH) and light (VL) chains using structural databases (16).

The next critical step is determining the correct relative orientation between the VH and VL domains. This spatial relationship significantly influences the overall geometry of the paratope. Dedicated tools like *AbAngle* have been developed to calculate these orientations accurately (107, 109).

Following orientation, the CDR loops—especially the five canonical ones—are modeled. Knowledge-based algorithms can predict these loop structures with high accuracy if suitable templates are available (103, 156). However, modeling the CDRH3 loop remains a significant challenge due to its high structural diversity (156). When no suitable template is available, ab initio approaches generate loop conformations from scratch. Although powerful, they are computationally intensive and often require further steps to select the best loop among numerous candidates (102). Hybrid strategies, such as *Sphinx*, integrate both knowledge-based and ab initio techniques, improving reliability in template-limited scenarios (102).

Once the loop conformations are modeled, the fourth phase focuses on predicting and refining side-chain orientations. General-purpose side-chain modeling tools like *SCWRL* (100) are frequently used, but specialized methods like *PEARS*—designed specifically for antibodies—can produce more accurate side-chain conformations (98).

The final step involves energy minimization to refine the full antibody structure and improve atomic packing. This can be performed using tools like *Rosetta* (89), which optimize the model's energetic landscape to yield a physically plausible conformation.

Multiple software tools and platforms are available to implement these modeling strategies. Some of the freely accessible web servers include *PIGS* (86) and *AbodyBuilder* (84). Commercial packages offering antibody modeling functionalities include *Biovia* from Accelrys (3dsbiovia.com), *SmrtMolAntibody* from Macromoltek (macromoltek.com), *MOE* from Chemical Computing Group (chemcomp.com), and *BioLuminate* by Schrödinger Inc. (schrodinger.com). Tools like *AbPredict* (96) and *Rosetta* (89) are also available for local installation.

These platforms differ significantly in computational efficiency. For instance, *AbodyBuilder* can generate a structural model in about one minute, whereas Rosetta-based frameworks may require several hours to complete a run. Nevertheless, the predictive accuracy across different tools tends to be comparable. In the Antibody Modeling Assessment II (AMA II) study (7), multiple software packages underwent blind benchmarking. The results revealed an average root mean square deviation (RMSD) of 1.1 Å for the predicted Fv regions, although modeling accuracy for the CDRH3 loop remained limited, sometimes exceeding 5 Å RMSD.

While these computational models cannot fully match the resolution of experimental structural data, an RMSD of ~1.0 Å is sufficient to infer meaningful structural and functional insights. These models can be instrumental during the **Lead Identification** phase, such as in identifying paratope residues for mutagenesis (110), or during **Lead Optimization**, for evaluating developability features like paratope surface hydrophobicity (13), which require detailed structural information about the antibody–antigen interface (119).

Popular modeling tools include **AbodyBuilder** [84], which quickly delivers models (~1 min), and more computationally intensive platforms such as Rosetta. Benchmarking in Antibody Modelling Assessment II shows that these tools achieve ~1.1 Å RMSD accuracy overall, although CDRH3 can deviate by over 5 Å [7].

IV. Paratope & Epitope Prediction and Antibody Docking

Paratope prediction

Identifying antigen-contacting residues (paratopes) is critical; about half the residues in CDR regions directly bind antigen surfaces [157–159]. Computational tools—ranging from statistical predictors like **Antibody i-Patch** [110] and **Paratome** [111]—to machine learning models like proABC [112], **A**ntibodyInterfacePrediction [114], and deep learning systems **Parapred** [113] and **AG-Fast-Parapred** [115]—help highlight binding residues. These guides are key during optimization to pinpoint mutation targets.

Epitope prediction

Understanding the antigen-binding site (epitope) informs therapeutic targeting and patent strategy. While experimental mapping is definitive, computational alternatives exist. Linear epitope predictors rely on sequence patterns, but conformational predictors—particularly those accounting for antibody–antigen context—offer more accurate results. Antibody-specific tools (e.g., ASEP [123], EpiPred [119], MabTope [122], Jespersen et al. [120]) prioritize paratope-epitope pairs for improved precision (Table 2C).

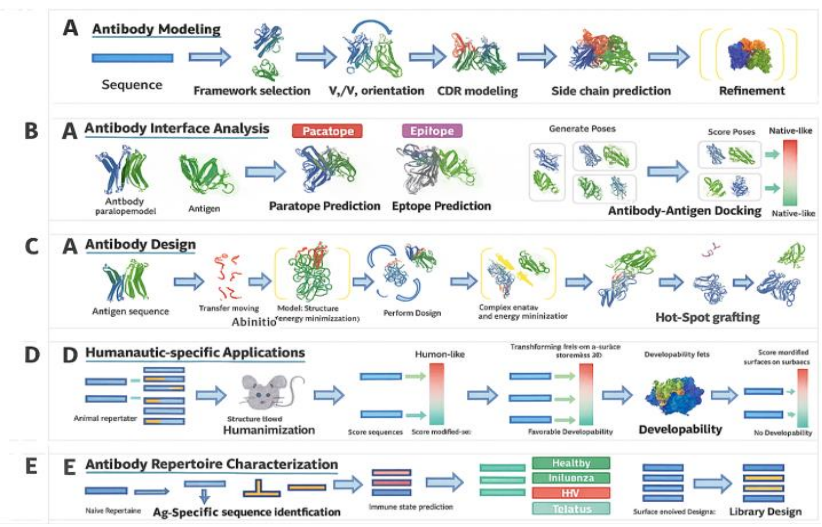

Figure 2.Computational antibody methods schematic. (A) Antibody modelling produces three dimensional coordinates from the sequence of an antibody. Framework templates are identified and the VH/VL domains can be oriented with respect to each other if the two regions originate from different molecules. CDRs are modeled into the framework followed by side-chain prediction and refinement of the entire structure by energy minimization. (B) Antibody interface prediction identifies the residues on the antibody (paratope) that are in contact with the antigen (epitope). This is a special case of molecular docking in which the antibody–antigen docking aims to recapitulate the complex between the antibody and the antigen. (C) Antibody design optimizes the binding of an antibody against an epitope of choice through a series of modelling, docking and energy minimization steps. In *ab initio* design, novel paratopes are generated computationally and their structural stability and binding propensity against the cognate epitope assessed by energy functions. Hotspot grafting involves transferring known interaction motifs from the antigen partner protein to an antibody template. (D) Antibodies need to be immunologically safe and have favorable biophysical properties in order to be administered to humans. Humanization involves modifying an animal-derived sequence to resemble one with a higher degree of human amino acid content without affecting its affinity and specificity. Develop ability-specific applications annotate regions on the surface that might lead to poor solubility or aggregation altogether. (E) Entire antibody repertoires can be used to draw information on the mechanics of the adaptive immune system. Identification of antigen-specific sequences post-vaccination can identify antibodies that could bestow passive immunity. The dynamic state of the repertoire can be analyzed to identify diseases in the organism. The diversity of antibodies can be harnessed to create surface display libraries recapitulating naturally evolved preferences and advantages.

Docking

Predicting antibody–antigen complexes uses protein docking techniques:

- **Global "ab initio" docking**, as employed by **ClusPro** [8,126] and **ZDOCK** [130].
- **Information-driven docking** (e.g., **SnugDock** [9,89], **HADDOCK** [131,132]), which incorporates CDR positions or experimental constraints.

Docking involves sampling potential complex structures followed by ranking (e.g., **ZRANK**, **FireDock**, **Dock-Sorter**). Flexibility-aware tools like **SwarmDock**, **HADDOCK**, and **SnugDock** can model conformational changes, improving accuracy.

HADDOCK supports the integration of experimental restraints—NMR, HDX, mutagenesis—to refine docking predictions, even with minimal epitope guidance [178]. Performance continues to be evaluated in benchmarks like CAPRI [179].

V. Ultimately, combining paratope/epitope predictions with docking offers a cost-effective

route to understanding antigen recognition, guiding experimental design. Computational Approaches for Therapeutic Antibody Discovery

Antibody Design and Modeling

Antibody modelling and antigen-binding interface prediction tools play crucial roles in both the early and advanced phases of therapeutic antibody development. During lead identification, these tools can be employed to design new antibodies from scratch (ab initio), while in lead optimization, they help refine candidates for improved binding and efficacy (as illustrated in Figure 2C and summarized in Table 2D). If the structure of the target antigen is known, it opens opportunities to computationally develop novel antibodies against it [180]. Pioneering work by Lippow and colleagues demonstrated that an existing antibody–antigen complex structure can be computationally modified to enhance binding affinity [181]. Their method involved comprehensive in silico mutagenesis of complementarity-determining regions (CDRs), followed by binding affinity evaluation using the CHARMM force field [182]. Some of these engineered variants showed increased target affinity, proving that computational tools alone can support affinity maturation.

Since then, several ab initio antibody design protocols have emerged, notably OptCDR [139], OptMAVEn [140], AbDesign [143], and RosettaAntibodyDesign [142]. These tools typically follow a four-step pipeline: CDR creation, structural modelling, docking with the antigen, and interaction energy evaluation. OptCDR and RosettaAntibodyDesign derive CDR conformations using databases of canonical structures and model the CDRH3 loop specifically. On the other hand, OptMAVEn and AbDesign adopt a modular approach, assembling antibodies through recombination-like processes akin to V(D)J rearrangement. The resulting constructs are optimized using established energy functions such as RosettaEnergy [183] or CHARMM [182]. These designs are then tested by docking simulations and scored based on binding energy between the antibody and antigen. Although still relatively new, these approaches have shown experimental validation in some cases. Their broader utility in industrial drug development settings, however, awaits further confirmation.

These methods also enable refinement of CDRs to improve stability and affinity through targeted mutagenesis and energy optimization. A distinct strategy termed "hot-spot grafting," proposed by Liu and colleagues, involves transplanting key binding motifs from known protein complexes onto antibodies [11]. Another innovative method, "re-epitoping," developed by Ofran's team, uses existing antibodies to probe epitope complementarity and guides the design of focused display libraries [184], speeding up the identification of therapeutic leads.

These computational methodologies not only streamline early-stage antibody discovery but also support lead optimization by evaluating properties like immunogenicity and developability.

Immunogenicity Assessment

A significant portion of therapeutic antibodies originate from animal immunization, particularly in mice. These non-human-derived antibodies often provoke immune responses in patients, leading to anti-drug antibodies (ADAs). To mitigate this, a process known as humanization is used, in which mouse-derived CDRs are inserted into human antibody frameworks or the frameworks themselves are engineered to resemble human sequences [185, 186]. Traditional humanization involves aligning the sequence with human germline genes to choose a suitable template. However, germline comparisons may not reflect the full diversity of human antibody repertoires.

Computational humanization methods have evolved to address this limitation by comparing the candidate sequence to thousands of recombined human antibody sequences (see Figure 2D and Table 2E). One of the earliest tools, Tabhu, matches a query antibody sequence against a vast repertoire of human antibodies from databases like DIGIT [144]. While this approach considers antibody diversity, simple sequence alignment is often inadequate. More sophisticated, statistically driven methods have since been developed. For instance, the Humanness Score by Andrew Martin's group evaluates how closely a sequence resembles the human amino acid distribution [14]. This score serves as a global metric for humanness.

Further refinement came with the Human String Content (HSC) score, developed by Lazar and colleagues. HSC assesses short peptide segments (e.g., 9-mers) to flag potentially immunogenic regions that diverge from human norms [145]. Both Humanness Score and HSC are based on sequence similarity but newer methods now consider positional residue dependencies, improving predictive accuracy [187, 188]. Though still sequence-based, some like HSC incorporate structural contact data to enhance predictions.

Structural models can also aid in a process called "resurfacing," where exposed immunogenic residues are replaced to reduce immune recognition. Choi and colleagues effectively combined structure-based design with HSC scoring to create de-immunized antibodies [147].

However, immune reactions to biologics can still occur even with fully humanized antibodies. Immunogenicity is multifactorial—shaped not only by sequence but also by individual patient profiles and protein product quality (e.g., presence of aggregates or degradation products) [190, 191]. A key initial step in immunogenicity is the presentation of therapeutic peptide fragments by MHC class II molecules to T-cells.

Several computational platforms have been designed to predict binding between peptide sequences and MHC class I or II molecules. These tools often use machine learning, including neural networks, to estimate binding affinities of peptides to MHC [192, 193]. Public resources like the IEDB provide validated data and epitope prediction tools, making them essential for immunogenicity assessments [18].

Predicted MHC-II binding peptides in therapeutic antibodies can serve as indicators of immunogenic potential and guide modifications early in development. Epivax Inc.'s immunogenicity scale is one such predictive metric used to evaluate and prioritize antibody candidates [194].

Kumar and co-workers observed that immune epitopes often overlap with aggregation-prone regions (APRs), particularly near the CDRs [195, 196]. This connection implies a shared mechanism between aggregation and immune activation and opens the door for simultaneous optimization of antibody efficacy, solubility, and safety using structure-guided engineering.

VI. Antibody Modelling, Immunogenicity, and Biophysical Properties

Antibody modelling and antigen-binding interface prediction tools play crucial roles in both the early and advanced phases

of therapeutic antibody development. During lead identification, these tools can be employed to design new antibodies from scratch (ab initio), while in lead optimization, they help refine candidates for improved binding and efficacy (as illustrated in Figure 2C and summarized in Table 2D). If the structure of the target antigen is known, it opens opportunities to computationally develop novel antibodies against it [180]. Pioneering work by Lippow and colleagues demonstrated that an existing antibody–antigen complex structure can be computationally modified to enhance binding affinity [181]. Their method involved comprehensive in silico mutagenesis of complementarity-determining regions (CDRs), followed by binding affinity evaluation using the CHARMM force field [182]. Some of these engineered variants showed increased target affinity, proving that computational tools alone can support affinity maturation.

Since then, several ab initio antibody design protocols have emerged, notably OptCDR [139], OptMAVEn [140], AbDesign [143], and RosettaAntibodyDesign [142]. These tools typically follow a four-step pipeline: CDR creation, structural modelling, docking with the antigen, and interaction energy evaluation. OptCDR and RosettaAntibodyDesign derive CDR conformations using databases of canonical structures and model the CDRH3 loop specifically. On the other hand, OptMAVEn and AbDesign adopt a modular approach, assembling antibodies through recombination-like processes akin to V(D)J rearrangement. The resulting constructs are optimized using established energy functions such as RosettaEnergy [183] or CHARMM [182]. These designs are then tested by docking simulations and scored based on binding energy between the antibody and antigen. Although still relatively new, these approaches have shown experimental validation in some cases. Their broader utility in industrial drug development settings, however, awaits further confirmation.

These methods also enable refinement of CDRs to improve stability and affinity through targeted mutagenesis and energy optimization. A distinct strategy termed "hot-spot grafting," proposed by Liu and colleagues, involves transplanting key binding motifs from known protein complexes onto antibodies [11]. Another innovative method, "re-epitoping," developed by Ofran's team, uses existing antibodies to probe epitope complementarity and guides the design of focused display libraries [184], speeding up the identification of therapeutic leads.

These computational methodologies not only streamline early-stage antibody discovery but also support lead optimization by evaluating properties like immunogenicity and developability.

Immunogenicity Assessment

A significant portion of therapeutic antibodies originate from animal immunization, particularly in mice. These non-human-derived antibodies often provoke immune responses in patients, leading to anti-drug antibodies (ADAs). To mitigate this, a process known as humanization is used, in which mouse-derived CDRs are inserted into human antibody frameworks or the frameworks themselves are engineered to resemble human sequences [185, 186]. Traditional humanization involves aligning the sequence with human germline genes to choose a suitable template. However, germline comparisons may not reflect the full diversity of human antibody repertoires.

Computational humanization methods have evolved to address this limitation by comparing the candidate sequence to thousands of recombined human antibody sequences (see Figure 2D and Table 2E). One of the earliest tools, Tabhu, matches a query antibody sequence against a vast repertoire of human antibodies from databases like DIGIT [144]. While this approach considers antibody diversity, simple sequence alignment is often inadequate. More sophisticated, statistically driven methods have since been developed. For instance, the Humanness Score by Andrew Martin's group evaluates how closely a sequence resembles the human amino acid distribution [14]. This score serves as a global metric for humanness.

Further refinement came with the Human String Content (HSC) score, developed by Lazar and colleagues. HSC assesses short peptide segments (e.g., 9-mers) to flag potentially immunogenic regions that diverge from human norms [145]. Both Humanness Score and HSC are based on sequence similarity but newer methods now consider positional residue dependencies, improving predictive accuracy [187, 188]. Though still sequence-based, some like HSC incorporate structural contact data to enhance predictions.

Structural models can also aid in a process called "resurfacing," where exposed immunogenic residues are replaced to reduce immune recognition. Choi and colleagues effectively combined structure-based design with HSC scoring to create de-immunized antibodies [147].

However, immune reactions to biologics can still occur even with fully humanized antibodies. Immunogenicity is multifactorial—shaped not only by sequence but also by individual patient profiles and protein product quality (e.g., presence of aggregates or degradation products) [190, 191]. A key initial step in immunogenicity is the presentation of therapeutic peptide fragments by MHC class II molecules to T-cells.

Several computational platforms have been designed to predict binding between peptide sequences and MHC class I or II molecules. These tools often use machine learning, including neural networks, to estimate binding affinities of peptides to MHC [192, 193]. Public resources like the IEDB provide validated data and epitope prediction tools, making them essential for immunogenicity assessments [18].

Predicted MHC-II binding peptides in therapeutic antibodies can serve as indicators of immunogenic potential and guide modifications early in development. Epivax Inc.'s immunogenicity scale is one such predictive metric used to evaluate and prioritize antibody candidates [194].

Kumar and co-workers observed that immune epitopes often overlap with aggregation-prone regions (APRs), particularly near the CDRs [195, 196]. This connection implies a shared mechanism between aggregation and immune activation and opens the door for simultaneous optimization of antibody efficacy, solubility, and safety using structure-guided engineering.

Despite these advancements, the relationship between computational epitope predictions and real-world ADA generation remains under investigation. Consequently, while computational de-immunization holds promise for more efficient therapeutic development, its clinical impact is yet to be fully validated.

Biophysical Properties

In addition to immunogenicity, the successful development of antibody therapeutics also depends on their biophysical characteristics. Key attributes include colloidal stability, viscosity at high concentrations, and chemical or physical degradation profiles [197–201]. Maintaining good solubility is especially critical [202, 203] to prevent aggregation, which can lead to decreased efficacy, antibody breakdown, or unwanted immune responses.

Protein aggregation, a persistent issue in biopharmaceuticals, has both mechanistic and kinetic dimensions. Mechanistically, it involves identifying unstable regions in proteins, particularly aggregation-prone regions (APRs) characterized by surface-exposed hydrophobic patches. Various algorithms have been evaluated for their ability to predict APRs (Figure 2D, Table 2E) [204, 205]. Wang and colleagues demonstrated that marketed monoclonal antibodies (mAbs) often harbor multiple APR motifs in their CDRs [206]. These motifs not only contribute to antigen binding [160] but also explain how aggregation might reduce antibody potency and suggest targets for selective disruption to maintain activity.

Recently, Rawat et al. collected experimental kinetic data on aggregation and applied machine learning to identify mutations that either promote or reduce aggregation rates in proteins [207]. While several general-purpose tools exist for predicting solubility and APRs [208, 209], specialized antibody-focused tools have also been developed [204, 210]. For example, Lauer and collaborators assessed biophysical parameters of 12 antibodies over two years [148], deriving a Developability Index (DI). This score integrates calculated hydrophobicity, surface aggregation propensity (SAP) [211], and net molecular charge to assess aggregation risk.

Identifying hydrophobic surfaces—a key factor in aggregation risk—ideally requires crystal structures or accurate homology models. Jain and team addressed this by developing a surface accessibility predictor that generates a risk score based on sequence data [149]. Metrics like DI and aggregation propensity leverage hydrophobicity and charge annotations, indicating these alone can provide useful developability insights. Obrezanova et al. expanded this by creating an Adaptive Boosting model using a wide range of physicochemical features to predict aggregation tendencies [15], trained on a dataset of 500 antibody sequences.

These approaches, often relying on proprietary clinical-stage data, enable early candidate filtering for favorable developability. Alternatively, naturally occurring antibody sequences can serve as a proxy for desirable properties [13]. Raybould and colleagues proposed five guidelines based on such sequences. One of these involves comparing structure-based hydrophobicity scores against a large dataset of natural antibodies derived from next-generation sequencing (NGS). Deviations from the natural distribution indicate potential developability risks. This method exemplifies the innovative use of large-scale NGS data in guiding therapeutic antibody design and optimization.

VII. Emerging Trends: Leveraging NGS Data for Antibody Engineering

The advancement of computational strategies for antibody design is increasingly dependent on the integration of next-generation sequencing (NGS) data. This data, particularly from B-cell receptor (BCR) sequencing, is being used as a proxy for antibody repertoire analysis [212, 213]. Numerous online repositories now offer access to NGS datasets [17], which have proven valuable in evaluating therapeutic antibodies [13]. Current bioinformatic efforts primarily focus on interpreting immune repertoire diversity, with several potential applications in therapeutic development [22, 25, 214].

A major use of computational analysis of NGS data involves identifying antigen-specific BCR sequences post-immunization (Figure 2E). When an antigen is introduced, it stimulates the production of specific antibodies, leading to a skewed immune profile. By sequencing the immune repertoire and clustering similar sequences—particularly those sharing V and J genes and CDRH3 regions—researchers can identify candidate antigen-specific antibodies. This technique has been applied to human vaccination studies, such as with Hepatitis B [215], and in mouse models [216]. However, these basic clustering methods can sometimes fail to detect low-abundance antigen-specific sequences or may mistakenly identify unrelated sequences as relevant [215]. More sophisticated statistical models, as demonstrated by Fowler et al., improve accuracy by reducing false positives [217]. Identifying such antibodies is particularly useful in vaccine development, as they can serve as candidates for passive immunization [218].

Beyond antigen recognition, NGS analysis can also help infer an individual's immune status (Figure 2E). Since the immune repertoire reflects overall health, certain antibody signatures may correlate with disease states [219]. For example, classifiers have been trained to differentiate immune profiles linked to chronic lymphocytic leukemia [220], multiple sclerosis [221], and influenza [222]. Expanding these models could eventually lead to diagnostic tools capable of detecting numerous conditions solely through BCR sequencing.

Improving detection of antigen-specific sequences may require a deeper understanding of the sequence and structural principles that govern immune responses. Despite the immense diversity in antibody sequences, recent research has shown that certain sequence motifs are frequently shared across individuals [23, 223]. Even after discarding a majority of sequences (50–90%), key diversity features persist in the human antibody repertoire [214]. Moreover, structural constraints—particularly in the variable CDRH3 region—appear consistent among individuals [224]. Notably, many therapeutic CDRH3 loops are also found in natural repertoires from NGS studies, suggesting convergence between natural immunity and therapeutic design [225].

Recognizing these evolutionary patterns can inform the creation of more effective antibody libraries. For instance, an analysis of antibodies from over 600 donors [24] was used to guide the development of libraries based on naturally preferred sequence positions [26]. Libraries built on this foundation may yield antibodies with superior biophysical and immunological profiles.

Continued progress in NGS-based antibody engineering will depend not only on algorithmic innovation but also on data quality. Most current NGS datasets lack paired heavy and light chain information. Advancements in single-cell sequencing technology are crucial for generating such paired datasets [64, 227], which will significantly enhance computational exploration of the immune system and improve the development of next-generation antibody therapeutics.

Alternative Antibody Formats: Nanobodies

Recent advancements in antibody therapeutics have extended beyond traditional IgG molecules to include alternative molecular formats, particularly nanobodies. These are heavy-chain-only antibodies that naturally occur in camelids such as camels, alpacas, and llamas [27], as well as in certain species of sharks [228, 229]. Interest in nanobodies has grown substantially, especially following the approval of caplacizumab in 2018—the first therapeutic nanobody. This increasing attention has also led to the creation of specialized databases and analytical tools dedicated to nanobody sequences and structures [57, 58, 231, 232].

Nanobodies consist of a single variable domain containing three highly diverse loops: CDRH1, CDRH2, and CDRH3. These loops form a compact and elongated paratope on one side of the folded domain. The absence of a light chain results in significant differences between nanobodies and conventional antibodies in terms of both sequence composition and structural conformation. This allows nanobodies to recognize epitopes that are inaccessible to full-length antibodies, such as those buried within enzyme active sites, viral structures, or G protein-coupled receptors [233, 234].

Large-scale computational comparisons between classical antibodies and nanobodies reveal substantial systematic distinctions [231, 232]. Nanobodies show less variation in their framework regions but exhibit similar sequence diversity in the CDRH1 and CDRH2 loops when compared to traditional antibodies [232]. Notably, even with similar sequence diversity, nanobodies display greater structural variation in these regions. Unlike classical antibodies, the CDRH1 and CDRH2 loops in nanobodies do not conform to established canonical structural rules, presenting significant challenges for computational modelling [232, 235].

Additionally, nanobody CDRH3 loops tend to be three to four residues longer than those in conventional antibodies and exhibit greater diversity in both sequence and 3D structure [230, 232, 235, 236]. This variability contributes to unique loop conformations, such as extended finger-like projections, which enable deep insertion into antigen binding pockets.

From a modelling perspective, nanobody paratopes present even more complexity. On average, they include nearly three additional amino acid residues compared to those found in

standard antibody VH domains. Moreover, nanobody paratopes draw from a broader array of sequence positions, roughly equivalent to the combined VH-VL interface seen in conventional antibodies [230, 232]. Since the VL domain in classical antibodies contributes relatively little structural variability, this expanded footprint in nanobodies implies a greater need for refined modelling tools.

Further complicating matters, structural analyses of nanobody–antigen complexes reveal that nanobody paratopes consist of a more diverse array of structural motifs compared to classical antibodies [231]. The highly variable CDRH3 loop is often the primary contributor to antigen interaction, reinforcing the notion that nanobody–antigen interfaces cannot be easily modeled using traditional tools developed for IgG antibodies.

Given these fundamental differences, it is currently uncertain whether existing computational approaches for antibody modelling, docking, and affinity prediction are directly applicable to nanobodies. To clarify this, a comprehensive benchmarking of current antibody modelling tools against nanobody datasets is essential. Such an evaluation would highlight limitations and guide the development of new methods tailored specifically to the unique structural and functional characteristics of nanobodies.

Acknowledgement:

This work was supported by the Department of Bioengineering, Integral University, Lucknow, thankfully acknowledges the support provided by the Head of Department Prof. Alvina Farooqui, Faculty of engineering.

References

1. T. J. Kindt, R. A. Goldsby, B. A. Osborne, et al., *Kuby Immunology*. New York, USA: W. H. Freeman and Co., 2007, ISBN 9780716785903.
2. K. M. Kelly-Scumpia, P. O. Scumpia, J. S. Weinstein, et al., "B cells enhance early innate immune responses during bacterial sepsis," *J. Exp. Med.*, vol. 208, no. 8, pp. 1673–1682, 2011.
3. H.-P. Peng, K. H. Lee, J.-W. Jian, et al., "Origins of specificity and affinity in antibody-protein interactions," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, pp. E2656–E2665, 2014.
4. H. Kaplon, J. M. Reichert, "Antibodies to watch in," *MAbs*, vol. 2018, pp. 1–21, 2018.
5. K. Krawczyk, J. Dunbar, C. M. Deane, "Computational tools for aiding rational antibody design," *Methods Mol. Biol.*, vol. 1529, pp. 399–416, 2017.
6. A. Fiser, A. Šali, "Modeller: generation and refinement of homology-based protein structure models," *Methods Enzymol.*, vol. 374, pp. 461–491, 2003.
7. J. C. Almagro, A. Teplyakov, J. Luo, et al., "Second antibody modeling assessment (AMA-II)," *Proteins Struct. Funct. Bioinform.*, vol. 82, pp. 1553–1562, 2014.
8. R. Brenke, D. R. Hall, G.-Y. Chuang, et al., "Application of asymmetric statistical potentials to antibody-protein docking," *Bioinformatics*, vol. 28, pp. 2608–2614, 2012.
9. A. Sircar, J. J. Gray, "SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models," *PLoS Comput. Biol.*, vol. 6, p. e1000644, 2010.
10. R. Esmaielbeiki, K. Krawczyk, B. Knapp, et al., "Progress and challenges in predicting protein interfaces," *Brief. Bioinform.*, vol. 17, pp. 1–15, 2015.
11. X. Liu, R. D. Taylor, L. Griffin, et al., "Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping," *Sci. Rep.*, vol. 7, p. 41306, 2017.
12. D. Baran, M. G. Pszolla, G. D. Lapidoth, et al., "Principles for computational design of binding antibodies," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 41, pp. 10900–10905, 2017.
13. M. I. J. Raybould, C. Marks, K. Krawczyk, et al., "Five computational developability guidelines for therapeutic antibody profiling," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 10, pp. 4025–4030, 2019.
14. K. R. Abhinandan, A. C. R. Martin, "Analyzing the 'Degree of Humanness' of antibody sequences," *J. Mol. Biol.*, vol. 369, pp. 852–862, 2007.
15. O. Obrezanova, A. Arnell, R. G. De La Cuesta, et al., "Aggregation risk prediction for antibodies and its application to biotherapeutic development," *MAbs*, vol. 7, pp. 352–363, 2015.
16. J. Dunbar, K. Krawczyk, J. Leem, et al., "SAbDab: the structural antibody database," *Nucleic Acids Res.*, vol. 42, pp. 1140–1146, 2013.
17. A. Kovaltsuk, J. Leem, S. Kelm, et al., "Observed antibody space: a resource for data mining next-generation sequencing of antibody repertoires," *J. Immunol.*, vol. 201, no. 8, pp. 2502–2509, 2018.
18. R. Vita, L. Zarebski, J. A. Greenbaum, et al., "The immune epitope database 2.0," *Nucleic Acids Res.*, vol. 38, pp. D854–D862, 2010.
19. S. Sirin, J. R. Apgar, E. M. Bennett, et al., "AB-Bind: antibody binding mutational database for computational affinity predictions," *Protein Sci.*, vol. 25, no. 2, pp. 393–409, 2016.
20. P. Koenig, C. V. Lee, B. T. Walters, et al., "Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 4, pp. E486–E495, 2017.
21. T. Jain, T. Sun, S. Durand, et al., "Biophysical properties of the clinical-stage antibody landscape," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 5, pp. 944–949, 2017.
22. E. Miho, A. Yermanos, C. R. Weber, et al., "Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires," *Front. Immunol.*, vol. 9, p. 224, 2018.
23. B. Briney, A. Inderbitzin, C. Joyce, et al., "Commonality despite exceptional diversity in the baseline human antibody repertoire," *Nature*, vol. 566, pp. 393–397, 2019.
24. J. Glanville, W. Zhai, J. Berka, et al., "Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, pp. 20216–20221, 2009.
25. A. J. Brown, I. Snapkov, R. Akbar, et al., "Augmenting adaptive immunity: progress and challenges in the quantitative engineering and analysis of adaptive immune receptor repertoires," *Mol. Syst. Des. Eng.*, 2019. doi: 10.1039/C9ME00071B, arXiv:1904.04105.
26. W. Zhai, J. Glanville, M. Fuhrmann, et al., "Synthetic antibodies designed on natural sequence landscapes," *J. Mol. Biol.*, vol. 412, no. 1, pp. 55–71, 2011.
27. P. Bannas, J. Hambach, F. Koch-Nolte, "Nanobodies and nanobody-based human heavy chain antibodies as antitumor therapeutics," *Front. Immunol.*, vol. 8, p. 1603, 2017.
28. S. Tonegawa, "Somatic generation of antibody diversity," *Nature*, vol. 302, pp. 575–581, 1983.
29. D. G. T. Hesslein, D. G. Schatz, "Factors and forces controlling V(D)J recombination," *Adv. Immunol.*, vol. 78, pp. 169–232, 2001.
30. U. Storb, "Somatic hypermutation and class switch recombination," *Encycl. Immunobiol.*, vol. 3, pp. 186–194, 2016.
31. A. Peters, U. Storb, "Somatic hypermutation of immunoglobulin genes is linked to transcription initiation," *Immunity*, vol. 4, no. 1, pp. 57–65, 1996.
32. C. Chothia, A. M. Lesk, "Canonical structures for the hypervariable regions of immunoglobulins," *J. Mol. Biol.*, vol. 196, pp. 901–917, 1987.
33. C. Regep, G. Georges, J. Shi, et al., "The H3 loop of antibodies shows unique structural characteristics," *Proteins Struct. Funct. Bioinform.*, vol. 85, pp. 1311–1318, 2017.

34. Y. Tsuchiya, K. Mizuguchi, "The diversity of H3 loops determines the antigen-binding tendencies of antibody CDR loops," *Protein Sci.*, 2016.
35. J. L. Xu, M. M. Davis, "Diversity in the CDR3 region of VH is sufficient for most antibody specificities," *Immunity*, vol. 13, pp. 37–45, 2000.
36. A. Knappik, L. Ge, A. Honegger, et al., "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides," *J. Mol. Biol.*, 2000.
37. J. De Kruif, E. Boel, T. Logtenberg, "Selection and application of human single chain Fv antibody fragments from a semisynthetic phage antibody display library with designed CDR3 regions," *J. Mol. Biol.*, vol. 248, no. 1, pp. 97–105, 1995.
38. P. Holliger, P. J. Hudson, "Engineered antibody fragments and the rise of single domains," *Nat. Biotechnol.*, vol. 23, no. 9, pp. 1126–1136, 2005.
39. S. Farajnia, V. Ahmadzadeh, A. Tanomand, et al., "Development trends for generation of single-chain antibody fragments," *Immunopharmacol. Immunotoxicol.*, vol. 36, no. 5, pp. 297–308, 2014.
40. N.-Y. Kwon, Y. Kim, J.-O. Lee, "Structural diversity and flexibility of diabodies," *Methods*, vol. 154, pp. 136–142, 2019.
41. K. Runcie, D. R. Budman, V. John, et al., "Bi-specific and trispecific antibodies—the next big thing in solid tumor therapeutics," *Mol. Med.*, vol. 24, no. 1, p. 50, 2018.
42. S. Duggan, "Caplacizumab: first global approval," *Drugs*, vol. 78, pp. 1639–1642, 2018.
43. A. Honegger, A. Plückthun, "Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool," *J. Mol. Biol.*, vol. 309, pp. 657–670, 2001.
44. S. M. Major, S. Nishizuka, D. Morita, et al., "AbMiner: a bioinformatic resource on available monoclonal antibodies and corresponding gene identifiers for genomic, proteomic, and immunologic studies," *BMC Bioinformatics*, vol. 7, p. 192, 2006.
45. M. Ohlin, C. Scheepers, M. Corcoran, et al., "Inferred allelic variants of immunoglobulin receptor genes: a system for their evaluation, documentation, and naming," *Front. Immunol.*, vol. 10, p. 435, 2019.
46. M.-P. Lefranc, V. Giudicelli, P. Duroux, et al., "IMGT®, the international ImMunoGeneTics information system® 25 years on," *Nucleic Acids Res.*, vol. 43, pp. D413–D422, 2015.
47. M. B. Swindells, C. T. Porter, M. Couch, et al., "abYsis: integrated antibody sequence and structure—management, analysis, and prediction," *J. Mol. Biol.*, vol. 429, pp. 356–364, 2017.
48. A. Chailyan, A. Tramontano, P. Marcatili, "A database of immunoglobulins with integrated tools: DIGIT," *Nucleic Acids Res.*, vol. 40, pp. D1230–D1234, 2012.
49. B. D. Corrie, N. Marthandan, B. Zimonja, et al., "iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories," *Immunol. Rev.*, vol. 284, no. 1, pp. 24–41, 2018.
50. M. N. Nguyen, M. R. Pradhan, C. Verma, et al., "The interfacial character of antibody paratopes: analysis of antibody-antigen structures," *Bioinformatics*, vol. 33, pp. 2971–2976, 2017.
51. J. Adolf-Bryfogle, Q. Xu, B. North, et al., "PyIgClassify: a database of antibody CDR structural classifications," *Nucleic Acids Res.*, vol. 43, pp. D432–D438, 2015.
52. S. Ferdous, A. C. R. Martin, "AbDb: antibody structure database—a database of PDB-derived antibody structures," *Database*, vol. 2018, p. bay040, 2018.
53. H. R. Ansari, D. R. Flower, G. P. S. Raghava, "AntigenDB: an immunoinformatics database of pathogen antigens," *Nucleic Acids Res.*, vol. 38, pp. D847–D853, 2010.
54. R. Wang, X. Fang, Y. Lu, et al., "The PDBbind database: methodologies and updates," *J. Med. Chem.*, vol. 48, pp. 4111–4119, 2005.
55. I. H. Moal, J. Fernández-Recio, "SKEMPI: a structural kinetic and energetic database of mutant protein interactions and its use in empirical models," *Bioinformatics*, vol. 28, no. 20, pp. 2600–2607, 2012.
56. J. Jankauskaite, B. Jiménez-García, J. Dapkunas, et al., "SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation," *Bioinformatics*, vol. 35, no. 3, pp. 462–469, 2019.
57. U. Zavrtanik, S. Hadži, "A non-redundant data set of nanobody-antigen crystal structures," *Data Br.*, vol. 2019, p. 103754, 2019.
58. E. E. Wilton, M. P. Opyr, S. Kailasam, et al., "sdAb-DB: the single domain antibody database," *ACS Synth. Biol.*, vol. 7, no. 11, pp. 2480–2484, 2018.
59. J. Zuo, J. Li, R. Zhang, et al., "Institute collection and analysis of Nanobodies (iCAN): a comprehensive database and analysis platform for nanobodies," *BMC Genomics*, vol. 18, p. 797, 2017.
60. R. Leinonen, R. Akhtar, E. Birney, et al., "The European nucleotide archive," *Nucleic Acids Res.*, vol. 39, 2011.
61. NCBI Resource Coordinators, "Database Resources of the National Center for Biotechnology Information," *Nucleic Acids Res.*, vol. 45, pp. D12–D17, 2017.
62. A. Kovaltsuk, K. Krawczyk, S. Kelm, et al., "Filtering next-generation sequencing of the Ig gene repertoire data using antibody structural information," *J. Immunol.*, vol. 201, no. 12, pp. 3694–3704, 2018.
63. B. J. DeKosky, O. I. Lungu, D. Park, et al., "Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 113, no. 19, pp. E2636–E2645, 2016.
64. B. J. DeKosky, G. C. Ippolito, R. P. Deschner, et al., "High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire," *Nat. Biotechnol.*, vol. 31, pp. 166–169, 2013.
65. H. M. Berman, J. Westbrook, Z. Feng, et al., "The protein data bank," *Nucleic Acids Res.*, vol. 28, pp. 235–242, 2000.
66. F. Ehrenmann, Q. Kaas, M. Lefranc, "IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF," *Nucleic Acids Res.*, vol. 38, pp. D301–D307, 2010.
67. S. Mahajan, R. Vita, D. Shackelford, et al., "Epitope specific antibodies and T cell receptors in the immune epitope database," *Front. Immunol.*, vol. 9, p. 2688, 2018.
68. J. Ye, N. Ma, T. L. Madden, et al., "IgBLAST: an immunoglobulin variable domain sequence analysis tool," *Nucleic Acids Res.*, vol. 41, pp. W34–W40, 2013.
69. X. Brochet, M.-P. Lefranc, V. Giudicelli, "IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis," *Nucleic Acids Res.*, vol. 36, pp. W503–W508, 2008.
70. D. A. Bolotin, S. Poslavsky, I. Mitrophanov, et al., "MiXCR: software for comprehensive adaptive immunity profiling," *Nat. Methods*, vol. 12, 2015.
71. N. T. Gupta, J. A. Vander Heiden, M. Uduman, et al., "Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data," *Bioinformatics*, vol. 31, no. 20, pp. 3356–3358, 2015.
72. J. A. Vander Heiden, G. Yaari, M. Uduman, et al., "PRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires," *Bioinformatics*, vol. 30, no. 13, pp. 1930–1932, 2014.

73. A. Shlemov, S. Bankevich, A. Bzikadze, et al., "Reconstructing antibody repertoires from error-prone immunosequencing reads," *J. Immunol.*, vol. 199, no. 9, pp. 3369–3380, 2017.
74. B. Cortina-Ceballos, E. E. Godoy-Lozano, H. Sámano-Sánchez, et al., "Reconstructing and mining the B cell repertoire with ImmunediveRsity," *MAbs*, vol. 7, no. 3, pp. 516–524, 2015.
75. L. Kuchenbecker, M. Nienen, J. Hecht, et al., "IMSEQ—a fast and error aware approach to immunogenetic sequence analysis," *Bioinformatics*, vol. 31, no. 18, pp. 2963–2971, 2015.
76. D. K. Ralph, F. A. Matsen, "Consistency of VDJ rearrangement and substitution parameters enables accurate B cell receptor sequence annotation," *PLoS Comput. Biol.*, vol. 12, no. 1, p. e1004409, 2016.
77. Q. Marcou, T. Mora, A. M. Walczak, "High-throughput immune repertoire analysis with IGoR," *Nat. Commun.*, vol. 9, no. 1, p. 561, 2018.
78. M. Giraud, M. Salson, M. Duez, et al., "Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing," *BMC Genomics*, vol. 15, p. 409, 2014.
79. M. Duez, M. Giraud, R. Herbert, et al., "Vidjil: a web platform for analysis of high-throughput repertoire sequencing," *PLoS One*, vol. 11, no. 11, p. e0166126, 2016.
80. A. M. Rosenfeld, W. Meng, E. T. Luning Prak, et al., "ImmuneDB, a novel tool for the analysis, storage, and dissemination of immune repertoire sequencing data," *Front. Immunol.*, vol. 9, p. 2107, 2018.
81. L. Li, S. Chen, Z. Miao, et al., "AbRSA: a robust tool for antibody numbering," *Protein Sci.*, vol. 28, no. 1, pp. 152–160, 2019.
82. K. R. Abhinandan, A. C. Martin, "Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains," *Mol. Immunol.*, vol. 45, no. 14, pp. 3832–3839, 2008.
83. J. Dunbar, C. M. Deane, "ANARCI: Antigen receptor numbering and receptor classification," *Bioinformatics*, vol. 32, no. 2, pp. 298–300, 2016.
84. J. Leem, J. Dunbar, G. Georges, et al., "ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation," *MAbs*, vol. 8, no. 7, pp. 1259–1268, 2016.
85. M. S. Klausen, M. V. Anderson, M. C. Jespersen, et al., "LYRA, a webserver for lymphocyte receptor structural modeling," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W349–W355, 2015.
86. P. Marcatili, A. Rosi, A. Tramontano, "PIGS: automatic prediction of antibody structures," *Bioinformatics*, vol. 24, no. 17, pp. 1953–1954, 2008.
87. K. Yamashita, K. Ikeda, K. Amada, et al., "Kotai antibody builder: automated high-resolution structural modeling of antibodies," *Bioinformatics*, vol. 30, no. 22, pp. 3279–3280, 2014.
88. A. Sivasubramanian, A. Sircar, S. Chaudhury, et al., "Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking," *Proteins*, vol. 74, no. 2, pp. 497–514, 2009.
89. B. D. Weitzner, J. R. Jeliazkov, S. Lyskov, et al., "Modeling and docking of antibody structures with Rosetta," *Nat. Protoc.*, vol. 12, no. 3, pp. 401–416, 2017.
90. H. Kemmish, M. Fasnacht, L. Yan, "Fully automated antibody structure prediction using BIOVIA tools: validation study," *PLoS One*, vol. 12, no. 5, p. e0177923, 2017.
91. A. Bujotzek, A. Fuchs, C. Qu, et al., "MoFvAb: modeling the Fv region of antibodies," *MAbs*, vol. 7, no. 5, pp. 838–852, 2015.
92. N. R. Whitelegg, A. R. Rees, "WAM: an improved algorithm for modelling antibodies on the WEB," *Protein Eng.*, vol. 13, no. 12, pp. 819–824, 2000.
93. K. Zhu, T. Day, D. Warshaviak, et al., "Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction," *Proteins Struct. Funct. Bioinform.*, vol. 82, no. 8, pp. 1646–1655, 2014.
94. CCG Inc., *Molecular Operating Environment (MOE), 2016.08.* Montreal, QC, Canada: CCG Inc., 2016.
95. C. Mandal, B. D. Kingery, J. M. Anchin, et al., "ABGEN: a knowledge-based automated approach for antibody structure modeling," *Nat. Biotechnol.*, vol. 14, no. 3, pp. 323–328, 1996.
96. G. Lapidoth, J. Parker, J. Prilusky, et al., "AbPredict 2: a server for accurate and unstrained structure prediction of antibody variable domains," *Bioinformatics*, vol. 35, no. 9, pp. 1591–1593, 2018.
97. M. Berrondo, S. Kaufmann, M. Berrondo, "Automated Aufbau of antibody structures from given sequences using Macromoltek's SmrtMolAntibody," *Proteins Struct. Funct. Bioinform.*, vol. 82, no. 8, pp. 1636–1645, 2014.
98. J. Leem, G. Georges, J. Shi, et al., "Antibody side chain conformations are position-dependent," *Proteins Struct. Funct. Bioinform.*, vol. 86, no. 4, pp. 383–392, 2018.
99. M. A. Messih, R. Lepore, P. Marcatili, et al., "Improving the accuracy of the structure prediction of the third hypervariable loop of the heavy chains of antibodies," Bioinformatics, vol. 30, no. 19, pp. 2733–2740, 2014.
100. G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, "Improved prediction of protein side-chain conformations with SCWRL4," Proteins Struct. Funct. Bioinform., vol. 77, no. 4, pp. 778–795, 2009.
101. J. Ryu, M. Lee, J. Cha, et al., "BetaSCPWeb: side-chain prediction for protein structures using Voronoi diagrams and geometry prioritization," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W416–W423, 2016.
102. C. Marks, J. Nowak, S. Klostermann, et al., "Sphinx: merging knowledge-based and ab initio approaches to improve protein loop prediction," *Bioinformatics*, vol. 33, no. 9, pp. 1346–1353, 2017.
103. Y. Choi, C. M. Deane, "FREAD revisited: accurate loop structure prediction using a database search algorithm," *Proteins*, vol. 78, no. 6, pp. 1431–1440, 2010.
104. M. P. Jacobson, D. L. Pincus, C. S. Rapp, et al., "A hierarchical approach to all-atom protein loop prediction," *Proteins Struct. Funct. Genet.*, vol. 55, no. 2, pp. 351–367, 2004.
105. A. C. R. Martin, J. M. Thornton, "Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies," *J. Mol. Biol.*, vol. 263, no. 5, pp. 800–815, 1996.
106. W. K. Wong, G. Georges, F. Ros, et al., "SCALOP: sequence-based antibody canonical loop structure annotation," *Bioinformatics*, vol. 35, no. 10, pp. 1774–1776, 2018.
107. A. Bujotzek, J. Dunbar, F. Lipsmeier, et al., "Prediction of VH-VL domain orientation for antibody variable domain modeling," *Proteins Struct. Funct. Bioinform.*, vol. 83, no. 4, pp. 681–695, 2015.
108. N. A. Marze, S. Lyskov, J. J. Gray, "Improved prediction of antibody VL-VH orientation," *Protein Eng. Des. Sel.*, vol. 29, no. 10, pp. 409–418, 2016.
109. J. Dunbar, A. Fuchs, J. Shi, et al., "ABangle: characterizing the VH-VL orientation in antibodies," *Protein Eng. Des. Sel.*, vol. 26, no. 10, pp. 611–620, 2013.
110. K. Krawczyk, T. Baker, J. Shi, et al., "Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking," *Protein Eng. Des. Sel.*, vol. 26, no. 10, pp. 621–629, 2013.
111. V. Kunik, S. Ashkenazi, Y. Ofran, "Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure," *Nucleic Acids Res.*, vol. 40, no. W1, pp. W521–W524, 2012.
112. P. P. Olimpieri, A. Chailyan, A. Tramontano, et al., "Prediction of site-specific interactions in antibody-antigen complexes: the proABC method and

server," *Bioinformatics*, vol. 29, no. 18, pp. 2285–2291, 2013.
113. E. Liberis, P. Velickovic, P. Sormanni, et al., "Parapred: antibody paratope prediction using convolutional and recurrent neural networks," *Bioinformatics*, vol. 34, no. 17, pp. 2944–2950, 2018.
114. S. Daberdaku, C. Ferrari, "Antibody interface prediction with 3D Zernike descriptors and SVM," *Bioinformatics*, vol. 35, no. 11, pp. 1870–1876, 2018.
115. A. Deac, P. Velickovic, P. Sormanni, "Attentive cross-modal paratope prediction," *J. Comput. Biol.*, vol. 26, no. 6, pp. 536–545, 2019.
116. R. Rapberger, A. Lukas, B. Mayer, "Identification of discontinuous antigenic determinants on proteins based on shape complementarities," *J. Mol. Recognit.*, vol. 20, no. 2, pp. 113–121, 2007.
117. I. Sela-Culang, S. Ashkenazi, B. Peters, et al., "PEASE: predicting B-cell epitopes utilizing antibody sequence," *Bioinformatics*, vol. 31, no. 8, pp. 1313–1315, 2015.
118. I. Sela-Culang, M. R. E. I. Benhnia, M. H. Matho, et al., "Using a combined computational-experimental approach to predict antibody-specific B cell epitopes," *Structure*, vol. 22, no. 4, pp. 646–657, 2014.
119. K. Krawczyk, X. Liu, T. Baker, et al., "Improving B-cell epitope prediction and its application to global antibody-antigen docking," *Bioinformatics*, vol. 30, no. 16, pp. 2288–2294, 2014.
120. M. C. Jespersen, S. Mahajan, B. Peters, et al., "Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes," *Front. Immunol.*, vol. 10, p. 298, 2019.
121. C. K. Hua, A. T. Gacerez, C. L. Sentman, et al., "Computationally-driven identification of antibody epitopes," *Elife*, vol. 6, p. e29023, 2017.
122. T. Bourquard, A. Musnier, V. Puard, et al., "MAbTope: a method for improved epitope mapping," *J. Immunol.*, vol. 201, no. 10, pp. 3096–3105, 2018.
123. S. Soga, D. Kuroda, H. Shirai, et al., "Use of amino acid composition to predict epitope residues of individual antibodies," *Protein Eng. Des. Sel.*, vol. 23, no. 6, pp. 441–448, 2010.
124. L. Zhao, J. Li, "Mining for the antibody-antigen interacting associations that predict the B cell epitopes," *BMC Struct. Biol.*, vol. 10, no. Suppl 1, p. S6, 2010.
125. L. Zhao, L. Wong, J. Li, "Antibody-specified B-cell epitope prediction in line with the principle of context-awareness," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 8, no. 6, pp. 1483–1494, 2011.
126. D. Kozakov, D. R. Hall, B. Xia, et al., "The ClusPro web server for protein-protein docking," *Nat. Protoc.*, vol. 12, no. 2, pp. 255–278, 2017.
127. N. Shimba, N. Kamiya, H. Nakamura, "Model building of antibody-antigen complex structures using GBSA scores," *J. Chem. Inf. Model.*, vol. 56, no. 10, pp. 2005–2012, 2016.
128. E. Ramírez-Aportela, J. R. López-Blanco, P. Chacón, "FRODOCK 2.0: fast protein-protein docking server," *Bioinformatics*, vol. 32, no. 15, pp. 2386–2388, 2016.
129. G. Macindoe, L. Mavridis, V. Venkatraman, et al., "HexServer: an FFT-based protein docking server powered by graphics processors," *Nucleic Acids Res.*, vol. 38, no. Suppl 2, pp. W445–W449, 2010.
130. R. Chen, L. Li, Z. Weng, "ZDOCK: an initial-stage protein docking algorithm," *Proteins*, vol. 52, no. 1, pp. 80–87, 2003.
131. C. Dominguez, R. Boelens, A. M. J. J. Bonvin, "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information," *J. Am. Chem. Soc.*, vol. 125, no. 7, pp. 1731–1737, 2003.
132. S. J. De Vries, A. D. J. Van Dijk, M. Krzeminski, et al., "HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets," *Proteins Struct. Funct. Genet.*, vol. 69, no. 4, pp. 726–733, 2007.
133. S. J. De Vries, C. E. M. Schindler, I. Chauvot De Beauchêne, et al., "A web interface for easy flexible protein-protein docking with ATTRACT," *Biophys. J.*, vol. 108, no. 3, pp. 462–465, 2015.
134. A. Tovchigrechko, I. A. Vakser, "GRAMM-X public web server for protein-protein docking," *Nucleic Acids Res.*, vol. 34, no. Web Server issue, pp. W310–W314, 2006.
135. B. Jiménez-García, C. Pons, J. Fernández-Recio, "pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring," *Bioinformatics*, vol. 29, no. 13, pp. 1698–1699, 2013.
136. M. Torchala, I. H. Moal, R. A. G. Chaleil, et al., "SwarmDock: a server for flexible protein-protein docking," *Bioinformatics*, vol. 29, no. 6, pp. 807–809, 2013.
137. D. Duhovny, R. Nussinov, H. J. Wolfson, "Efficient Unbound Docking of Rigid Molecules," in *Proc. 2nd Workshop Algorithms Bioinformatics (WABI)*, Rome, Italy, 2002, pp. 185–200.
138. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, et al., "PatchDock and SymmDock: servers for rigid and symmetric docking," *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W363–W367, 2005.
139. R. J. Pantazes, C. D. Maranas, "OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding," *Protein Eng. Des. Sel.*, vol. 23, no. 11, pp. 849–858, 2010.
140. T. Li, R. J. Pantazes, C. D. Maranas, "OptMAVEn—a new framework for the de novo design of antibody variable region models targeting specific antigen epitopes," *PLoS One*, vol. 9, no. 8, p. e105954, 2014.
141. R. Chowdhury, M. F. Allan, C. D. Maranas, "OptMAVEn-2.0: de novo design of variable antibody regions against targeted antigen epitopes," *Antibodies*, vol. 7, no. 3, p. 23, 2018.
142. J. Adolf-Bryfogle, O. Kalyuzhniy, M. Kubitz, et al., "RosettaAntibodyDesign (RAbD): a general framework for computational antibody design," *PLoS Comput. Biol.*, vol. 14, no. 4, p. e1006112, 2018.
143. G. D. Lapidoth, D. Baran, G. M. Pszolla, et al., "AbDesign: an algorithm for combinatorial backbone design guided by natural conformations and sequences," *Proteins Struct. Funct. Bioinform.*, vol. 83, no. 8, pp. 1385–1406, 2015.
144. P. P. Olimpieri, P. Marcatili, A. Tramontano, "Tabhu: tools for antibody humanization," *Bioinformatics*, vol. 31, no. 3, pp. 434–435, 2014.
145. G. A. Lazar, J. R. Desjarlais, J. Jacinto, et al., "A molecular immunology approach to antibody humanization and functional optimization," *Mol. Immunol.*, vol. 44, no. 8, pp. 1986–1998, 2007.
146. S.-H. Gao, K. Huang, H. Tu, et al., "Monoclonal antibody humanness score and its applications," *BMC Biotechnol.*, vol. 13, p. 55, 2013.
147. Y. Choi, C. Hua, C. L. Sentman, et al., "Antibody humanization by structure-based computational protein design," *MAbs*, vol. 7, no. 6, pp. 1045–1057, 2015.
148. T. M. Lauer, N. J. Agrawal, N. Chennamsetty, et al., "Developability index: a rapid in silico tool for the screening of antibody aggregation propensity," *J. Pharm. Sci.*, vol. 101, no. 1, pp. 102–115, 2012.
149. T. Jain, T. Boland, A. Lilov, et al., "Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning," *Bioinformatics*, vol. 33, no. 23, pp. 3758–3766, 2017.
150. L. López-Santibáñez-Jácome, S. E. Avendaño-Vázquez, C. F. Flores-Jasso, "The pipeline repertoire

for Ig-Seq analysis," *Front. Immunol.*, vol. 10, p. 899, 2019.
151. M. Dondelinger, P. Filée, E. Sauvage, et al., "Understanding the significance and implications of antibody numbering and antigen-binding surface/residue definition," *Front. Immunol.*, vol. 9, p. 2278, 2018.
152. T. T. Wu, E. A. Kabat, "An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity," *J. Exp. Med.*, vol. 132, no. 2, pp. 211–250, 1970.
153. M.-P. Lefranc, "IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF," *Cold Spring Harb. Protoc.*, vol. 2011, no. 6, pp. 633–642, 2011.
154. R. M. MacCallum, A. C. R. Martin, J. M. Thornton, "Antibody-antigen interactions: contact analysis and binding site topography," *J. Mol. Biol.*, vol. 262, no. 5, pp. 732–745, 1996.
155. B. North, A. Lehmann, R. L. Dunbrack, "A new clustering of antibody CDR loop conformations," *J. Mol. Biol.*, vol. 406, no. 2, pp. 228–256, 2011.
156. C. Marks, C. M. Deane, "Antibody H3 structure prediction," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 222–231, 2017.
157. J. W. Stave, K. Lindpaintner, "Antibody and antigen contact residues define epitope and paratope size and structure," *J. Immunol.*, vol. 191, no. 3, pp. 1428–1435, 2013.
158. I. Sela-Culang, V. Kunik, Y. Ofran, "The structural basis of antibody-antigen recognition," *Front. Immunol.*, vol. 4, p. 302, 2013.
159. V. Kunik, B. Peters, Y. Ofran, "Structural consensus among antibodies defines the antigen binding site," *PLoS Comput. Biol.*, vol. 8, no. 2, p. e1002388, 2012.
160. X. Wang, S. K. Singh, S. Kumar, "Potential aggregation-prone regions in complementarity-determining regions of antibodies and their contribution towards antigen recognition: a computational analysis," *Pharm. Res.*, vol. 27, no. 8, pp. 1512–1529, 2010.
161. J. V. Kringelum, C. Lundegaard, O. Lund, et al., "Reliable B cell epitope predictions: impacts of method development and improved benchmarking," *PLoS Comput. Biol.*, vol. 8, no. 12, p. e1002829, 2012.
162. A. Kazi, C. Chuah, A. B. A. Majeed, et al., "Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design," *Pathog. Glob. Health*, vol. 112, no. 3, pp. 123–131, 2018.
163. X. Deng, U. Storz, B. J. Doranz, "Enhancing antibody patent protection using epitope mapping information," *MAbs*, vol. 10, no. 2, pp. 204–209, 2017.
164. L. Potocnakova, M. Bhide, L. B. Pulzova, "An introduction to B-cell epitope mapping and in silico epitope prediction," *J. Immunol. Res.*, vol. 2016, p. 6760830, 2016.
165. P. Haste Andersen, M. Nielsen, O. Lund, "Prediction of residues in discontinuous B-cell epitopes using protein 3D structures," *Protein Sci.*, vol. 15, no. 11, pp. 2558–2567, 2006.
166. J. Gao, L. Kurgan, "Computational prediction of B cell epitopes from antigen sequences," *Methods Mol. Biol.*, vol. 1184, pp. 197–215, 2014.
167. V. Kunik, Y. Ofran, "The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops," *Protein Eng. Des. Sel.*, vol. 26, no. 10, pp. 599–609, 2013.
168. J. V. Kringelum, M. Nielsen, S. B. Padkjær, et al., "Structural analysis of B-cell epitopes in antibody: protein complexes," *Mol. Immunol.*, vol. 53, no. 1–2, pp. 24–34, 2013.
169. J. A. Greenbaum, P. H. Andersen, M. Blythe, et al., "Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools," *J. Mol. Recognit.*, vol. 20, no. 2, pp. 75–82, 2007.
170. N. S. Pagadala, K. Syed, J. Tuszynski, "Software for molecular docking: a review," *Biophys. Rev.*, vol. 9, no. 2, pp. 91–102, 2017.
171. B. Pierce, Z. Weng, "ZRANK: reranking protein docking predictions with an optimized energy function," *Proteins Struct. Funct. Genet.*, vol. 67, no. 4, pp. 1078–1086, 2007.
172. N. Andrusier, R. Nussinov, H. J. Wolfson, "FireDock: fast interaction refinement in molecular docking," *Proteins Struct. Funct. Genet.*, vol. 69, no. 1, pp. 139–159, 2007.
173. C. Pons, D. Talavera, X. De La Cruz, et al., "Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): a new efficient potential for protein-protein docking," *J. Chem. Inf. Model.*, vol. 51, no. 2, pp. 370–377, 2011.
174. J. P. G. L. M. Rodrigues, A. M. J. J. Bonvin, "Integrative computational modeling of protein interactions," *FEBS J.*, vol. 281, no. 8, pp. 1988–2003, 2014.
175. A. M. Sevy, J. F. Healey, W. Deng, et al., "Epitope mapping of inhibitory antibodies targeting the C2 domain of coagulation factor VIII by hydrogen-deuterium exchange mass spectrometry," *J. Thromb. Haemost.*, vol. 11, no. 12, pp. 2128–2136, 2013.
176. S. J. Coales, S. J. Tuske, J. C. Tomasso, et al., "Epitope mapping by amide hydrogen/deuterium exchange coupled with immobilization of antibody, on-line proteolysis, liquid chromatography and mass spectrometry," *Rapid Commun. Mass Spectrom.*, vol. 23, no. 5, pp. 639–647, 2009.
177. M. Kotev, R. Soliva, M. Orozco, "Challenges of docking in large, flexible and promiscuous binding sites," *Bioorg. Med. Chem.*, vol. 24, no. 20, pp. 4961–4969, 2016.
178. F. Ambrosetti, B. Jiménez-García, J. Roel-Touris, et al., "Information-driven modelling of antibody-antigen complexes," *SSRN Electron. J.*, 2019. doi: 10.2139/ssrn.3362436.
179. M. F. Lensink, S. Velankar, S. J. Wodak, "Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition," *Proteins Struct. Funct. Bioinform.*, vol. 85, no. 3, pp. 359–377, 2017.
180. D. Kuroda, K. Tsumoto, "Antibody affinity maturation by computational design," *Methods Mol. Biol.*, vol. 1827, pp. 15–34, 2018.
181. S. M. Lippow, K. D. Wittrup, B. Tidor, "Computational design of antibody-affinity improvement beyond in vivo maturation," *Nat. Biotechnol.*, vol. 25, no. 10, pp. 1171–1176, 2007.
182. A. D. MacKerell, Jr, C. L. Brooks, III, L. Nilsson, et al., "CHARMM: the energy function and its parameterization with an overview of the program," *Encycl. Comput. Chem.*, vol. 1, pp. 271–277, 1998.
183. A. Leaver-Fay, M. Tyka, S. M. Lewis, et al., "Rosetta3: an object-oriented software suite for the simulation and design of macromolecules," *Methods Enzymol.*, vol. 487, pp. 545–574, 2011.
184. G. Nimrod, S. Fischman, M. Austin, et al., "Computational design of epitope-specific functional antibodies," *Cell Rep.*, vol. 25, no. 7, pp. 1561–1572, 2018.
185. P. T. Jones, P. H. Dear, J. Foote, et al., "Replacing the complementarity-determining regions in a human antibody with those from a mouse," *Nature*, vol. 321, no. 6069, pp. 522–525, 1986.
186. J. C. Almagro, J. Fransson, "Humanization of antibodies," *Front. Biosci.*, vol. 13, pp. 1619–1633, 2008.
187. A. Clavero-Álvarez, T. Di Mambro, S. Perez-Gaviro, et al., "Humanization of antibodies using a statistical inference approach," *Sci. Rep.*, vol. 8, no. 1, p. 14820, 2018.
188. D. Seeliger, "Development of scoring functions for antibody sequence assessment and

optimization," *PLoS One*, vol. 8, no. 10, p. e76909, 2013.
189. M. A. Roguska, J. T. Pedersen, C. A. Keddy, et al., "Humanization of murine monoclonal antibodies through variable domain resurfacing," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, no. 3, pp. 969–973, 1994.
190. W. Jiskoot, G. Kijanka, T. W. Randolph, et al., "Mouse models for assessing protein immunogenicity: lessons and challenges," *J. Pharm. Sci.*, vol. 105, no. 5, pp. 1567–1575, 2016.
191. S. K. Singh, "Impact of product-related factors on immunogenicity of biotherapeutics," *J. Pharm. Sci.*, vol. 100, no. 2, pp. 354–387, 2011.
192. R. E. Soria-Guerra, R. Nieto-Gomez, D. O. Govea-Alonso, et al., "An overview of bioinformatics tools for epitope prediction: implications on vaccine development," *J. Biomed. Inform.*, vol. 53, pp. 405–414, 2015.
193. J.-W. Sidhom, D. Pardoll, A. Baras, "AI-MHC: an allele-integrated deep learning framework for improving Class I & Class II HLA-binding predictions," *bioRxiv*, 2018. doi: 10.1101/318881.
194. V. Jawa, L. P. Cousens, M. Awwad, et al., "T-cell dependent immunogenicity of protein therapeutics: preclinical assessment and mitigation," *Clin. Immunol.*, vol. 149, no. 3, pp. 534–555, 2013.
195. S. Kumar, S. K. Singh, X. Wang, et al., "Coupling of aggregation and immunogenicity in biotherapeutics: T- and B-cell immune epitopes may contain aggregation-prone regions," *Pharm. Res.*, vol. 28, no. 5, pp. 949–961, 2011.
196. S. Kumar, M. A. Mitchell, B. Rup, et al., "Relationship between potential aggregation-prone regions and HLA-DR-binding T-cell immune epitopes: implications for rational design of novel and follow-on therapeutic antibodies," *J. Pharm. Sci.*, vol. 101, no. 8, pp. 2686–2701, 2012.
197. S. Kumar, N. V. Plotnikov, J. C. Rouse, et al., "Biopharmaceutical informatics: supporting biologic drug development via molecular modelling and informatics," *J. Pharm. Pharmacol.*, vol. 70, no. 5, pp. 595–608, 2017.
198. D. S. Tomar, S. K. Singh, L. Li, et al., "In silico prediction of diffusion interaction parameter (k D), a key indicator of antibody solution behaviors," *Pharm. Res.*, vol. 35, no. 10, p. 193, 2018.
199. D. S. Tomar, L. Li, M. P. Broulidakis, et al., "In-silico prediction of concentration-dependent viscosity curves for monoclonal antibody solutions," *MAbs*, vol. 9, no. 3, pp. 476–489, 2017.
200. N. V. Plotnikov, S. K. Singh, J. C. Rouse, et al., "Quantifying the risks of asparagine deamidation and aspartate isomerization in biopharmaceuticals by computing reaction free-energy surfaces," *J. Phys. Chem. B*, vol. 121, no. 4, pp. 719–730, 2017.
201. D. S. Tomar, S. Kumar, S. K. Singh, et al., "Molecular basis of high viscosity in concentrated antibody solutions: strategies for high concentration drug product development," *MAbs*, vol. 8, no. 2, pp. 216–228, 2016.
202. P. Sormanni, F. A. Aprile, M. Vendruscolo, "The CamSol method of rational design of protein mutants with enhanced solubility," *J. Mol. Biol.*, vol. 427, no. 2, pp. 478–490, 2015.
203. A.-M. Wolf Pérez, P. Sormanni, J. S. Andersen, et al., "In vitro and in silico assessment of the developability of a designed monoclonal antibody library," *MAbs*, vol. 11, no. 2, pp. 388–400, 2019.
204. N. J. Agrawal, S. Kumar, X. Wang, et al., "Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions," *J. Pharm. Sci.*, vol. 100, no. 12, pp. 5081–5095, 2011.
205. P. M. Buck, S. Kumar, X. Wang, et al., "Computational methods to predict therapeutic protein aggregation," *Methods Mol. Biol.*, vol. 899, pp. 425–451, 2012.
206. X. Wang, T. K. Das, S. K. Singh, et al., "Potential aggregation prone regions in biotherapeutics: a survey of commercial monoclonal antibodies," *MAbs*, vol. 1, no. 3, pp. 254–267, 2009.
207. P. Rawat, S. Kumar, G. M. Michael, "An in-silico method for identifying aggregation rate enhancer and mitigator mutations in proteins," *Int. J. Biol. Macromol.*, vol. 118, pp. 1157–1167, 2018.
208. M. Hebditch, M. A. Carballo-Amador, S. Charonis, et al., "Protein-Sol: a web tool for predicting protein solubility from sequence," *Bioinformatics*, vol. 33, no. 19, pp. 3098–3100, 2017.
209. R. Zambrano, M. Jamroz, A. Szczasiuk, et al., "AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W306–W313, 2015.
210. W. Wang, S. Nema, D. Teagarden, "Protein aggregation—pathways and influencing factors," *Int. J. Pharm.*, vol. 390, no. 2, pp. 89–99, 2010.
211. N. Chennamsetty, V. Voynov, V. Kayser, et al., "Prediction of aggregation prone regions of therapeutic proteins," *J. Phys. Chem. B*, vol. 114, no. 19, pp. 6614–6624, 2010.
212. G. Georgiou, G. C. Ippolito, J. Beausang, et al., "The promise and challenge of high-throughput sequencing of the antibody repertoire," *Nat. Biotechnol.*, vol. 32, no. 2, pp. 158–168, 2014.
213. F. Rubelt, C. E. Busse, S. A. C. Bukhari, et al., "Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data," *Nat. Immunol.*, vol. 18, no. 12, pp. 1274–1278, 2017.
214. E. Miho, R. Roškar, V. Greiff, et al., "Large-scale network analysis reveals the sequence space architecture of antibody repertoires," *Nat. Commun.*, vol. 10, p. 1321, 2019.
215. J. Galson, J. Trück, A. Fowler, et al., "Analysis of B cell repertoire dynamics following hepatitis B vaccination in humans, and enrichment of vaccine-specific antibody sequences," *EBioMedicine*, vol. 2, no. 12, pp. 2070–2079, 2015.
216. S. T. Reddy, X. Ge, A. E. Miklos, et al., "Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells," *Nat. Biotechnol.*, vol. 28, no. 9, pp. 965–969, 2010.
217. A. Fowler, J. D. Galson, J. Trück, et al., "Inferring B cell specificity for vaccines using a mixture model," *bioRxiv*, 2018. doi: 10.1101/464792.
218. M. A. Keller, E. R. Stiehm, "Passive immunity in prevention and treatment of infectious diseases," *Clin. Microbiol. Rev.*, vol. 13, no. 4, pp. 602–614, 2000.
219. D. Chaussabel, "Assessment of immune status using blood transcriptomics and potential implications for global health," *Semin. Immunol.*, vol. 27, no. 1, pp. 58–66, 2015.
220. V. Greiff, P. Bhat, S. C. Cook, et al., "A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status," *Genome Med.*, vol. 7, no. 1, p. 49, 2015.
221. J. Ostmeyer, S. Christley, W. H. Rounds, et al., "Statistical classifiers for diagnosing disease from immune repertoires: a case study using multiple sclerosis," *BMC Bioinformatics*, vol. 18, no. 1, p. 401, 2017.
222. R. Arora, J. Kapllinsky, A. Li, et al., "Repertoire-based diagnostics using statistical biophysics," *bioRxiv*, 2019. doi: 10.1101/519108.
223. C. Soto, R. G. Bombardi, A. Branchizio, et al., "High frequency of shared clonotypes in human B cell receptor repertoires," *Nature*, vol. 566, no. 7744, pp. 398–402, 2019.
224. K. Krawczyk, S. Kelm, A. Kovaltsuk, et al., "Structurally mapping antibody repertoires," *Front. Immunol.*, vol. 9, p. 1698, 2018.
225. K. Krawczyk, M. Raybould, A. Kovaltsuk, et al., "Looking for therapeutic antibodies in next

generation sequencing repositories," *bioRxiv*, 2019. doi: 10.1101/572958.

226. A. S. Perelson, G. F. Oster, "Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination," *J. Theor. Biol.*, vol. 81, no. 4, pp. 645–670, 1979.

227. B. J. Dekosky, T. Kojima, A. Rodin, et al., "In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire," *Nat. Med.*, vol. 21, no. 1, pp. 86–91, 2015.

228. M. J. Feige, M. A. Grawert, M. Marcinowski, et al., "The structural analysis of shark IgNAR antibodies reveals evolutionary principles of immunoglobulins," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 22, pp. 8155–8160, 2014.

229. K. Griffiths, O. Dolezal, K. Parisi, et al., "Shark variable new antigen receptor (VNAR) single domain antibody fragments: stability and diagnostic applications," *Antibodies*, vol. 2, no. 4, pp. 66–81, 2013.

230. S. Muyldermans, "Nanobodies: natural single-domain antibodies," *Annu. Rev. Biochem.*, vol. 82, pp. 775–797, 2013.

231. L. S. Mitchell, L. J. Colwell, "Analysis of nanobody paratopes reveals greater diversity than classical antibodies," *Protein Eng. Des. Sel.*, vol. 31, no. 7–8, pp. 267–275, 2018.

232. L. S. Mitchell, L. J. Colwell, "Comparative analysis of nanobody sequence and structure data," *Proteins Struct. Funct. Bioinform.*, vol. 86, no. 7, pp. 697–706, 2018.

233. D. P. Staus, R. T. Strachan, A. Manglik, et al., "Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation," *Nature*, vol. 535, no. 7612, pp. 448–452, 2016.

234. J. Steyaert, B. K. Kobilka, "Nanobody stabilization of G protein-coupled receptor conformational states," *Curr. Opin. Struct. Biol.*, vol. 21, no. 4, pp. 567–572, 2011.

235. A. Sircar, K. A. Sanni, J. Shi, et al., "Analysis and modeling of the variable region of camelid single-domain antibodies," *J. Immunol.*, vol. 186, no. 11, pp. 6357–6367, 2011.

236. S. G. F. Rasmussen, H.-J. Choi, J. J. Fung, et al., "Structure of a nanobody-stabilized active state of the β2 adrenoceptor," *Nature*, vol. 469, no. 7329, pp. 175–180, 2011.

237. J. Haas, A. Barbato, D. Behringer, et al., "Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12," *Proteins Struct. Funct. Bioinform.*, vol. 86, no. S1, pp. 387–398, 2018.

238. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, p. 160018, 2016.