AACDB: Antigen-Antibody Complex Database — a Comprehensive Database Unlocking Insights into Interaction Interface

Yuwei Zhou[1], Wenwen Liu[1], Ziru Huang[1], Yushu Gou[1, 2], Siqi Liu[1], Lixu Jiang[1],Yue Yang[1], Jian Huang[1, 2*]

[1] The Clinical Hospital of Chengdu Brain Science Institute, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China.

[2] School of Healthcare Technology, Chengdu Neusoft University, Chengdu, 611844, China.

[*] Corresponding author:

hj@uestc.edu.cn (Jian Huang)

Abstract

Antibodies are critical components of the vertebrate immune system and possess a wide array of biomedical applications. Elucidating the complex interactions between antibodies and antigens is an important step in drug development. However, the complex and vast nature of the data presents significant challenges in accurately identifying and comprehending these interactions. To overcome these challenges and deepen our understanding of the antibody-antigen interface, we developed the Antigen-Antibody Complex Database (AACDB). The current version provides a comprehensive collection of 7,498 manually processed antigen-antibody complexes, ensuring accuracy and detail. This database provides extensive metadata and rectifies annotation errors found in the PDB database. Furthermore, it integrates data on antibody developability and antigen-drug target relationships, making it valuable for assisting new antibody therapies development. Notably, the database includes comprehensive paratope and epitope annotation information, thereby serving as a valuable benchmark for immunoinformatics research. The AACDB interface is designed to be user-friendly, providing researchers with powerful search and visualization tools that enable effortless querying, manipulation, and visualization of complex data. Researchers can access AACDB completely online at http://i.uestc.edu.cn/AACDB. Regular updates are promised to ensure the timely provision of scientific and valuable information.

Key words: Antibody; Antigen; Interaction Interface; Benchmark; Database

1. Introduction

Antibodies are not only an essential part of the vertebrate immune system, but also has wide application in biomedical field [1, 2]. In clinical practice, the advent of monoclonal antibodies (mAbs) has transformed the treatment for various cancers, autoimmune diseases, and numerous other conditions.. Due to their large binding interfaces, high affinity and specificity, antibody drugs have successfully targeted numerous proteins that were previously impervious to small molecule therapies. The expansion of antibody drug formats, such as antibody-drug conjugates (ADCs), domain antibodies, bispecific antibodies and antibody fusion proteins, has broadened their indications to a wider range of diseases including blood disorders, infections, neurological conditions, ocular diseases, and metabolic disorders [3-5]. Antibody drug development typically encompasses three stages: three stages: preclinical research, clinical trial and post-marketing surveillance study [6]. The preclinical research of antibodies mainly refers to a series of in vitro experiments and animal model studies conducted before the application of antibodies in clinical

trials. In this stage, an important task is to determine the binding interfaces and interaction sites between antibodies and their drug targets. Understanding how antibodies specifically interact with their targets or antigens can enable help optimize antibody drug leads further, as well as provide insights into mechanisms of interaction [7].

The field of machine learning and deep learning-based methods for antigen and antibody binding interfaces prediction have gained significant attention in recent years [8-17]. The advancement of this field heavily relies on a reliable foundation of data. Althoughthe Protein Data Bank (PDB) [18, 19] is a well-established repository for protein complexes, identifying antigen-antibody complexes within its extensive collection of general protein structures presents considerable challenges. In 2018, AbDb [20] was introduced as a specialized database for antigen-antibody complexes. However, its adoption has been limited by the absence of advanced visualization tools, interactive functionalities and abundant metadata annotations. We acknowledge that AbDb has made exceptional contributions to the field by aggregating antibody-antigen structures over the past decade. Our work builds upon these foundations, aiming to address complementary challenges in annotation and interoperability. In comparison, the series of SAbDab databases [21-23] indeed provide a comprehensive data source of antibody structure and achieve in time updating. However, this rapid updating process may inadvertently overlook a significant amount of information that requires thorough verification, thereby increasing the likelihood of incorporating erroneous annotations directly from the PDB database.

A further challenge in the field of interface prediction is the lack of a standardized definition for interacting amino acids. Some datasets are based on differences in the solvent accessible surface area (ΔSASA) for each residue upon binding [9, 24]; Others are based on distances between atoms [8, 13-15, 25]. Different studies employ different datasets and definitions, resulting in a lack of comparability. There is an urgent need for a standard benchmark dataset to enable objective comparisons of these methods. None of the databases mentioned above provide details of interaction residues. To address these limitations, we developed the Antigen-Antibody Complex Database (AACDB). It has a user-friendly interface for convenient querying, manipulation, browsing, and visualization of comprehensive information about antibody-antigen complexes. Notably, AACDB also provides detailed information on interacting residues using two methods (ΔSASA and atom distance), facilitating the benchmarking studies of the predictive methods for paratopes and epitopes.

2. Materials and Methods

We search the PDB database using the keywords “antibody”, “antigen”, “immunoglobulin” and “complex”. Over 32000 experimental structures were extracted (up to December 2023). Each entry was processed following the steps in Figure 1.

2.1 Entry screening

We aimed to retrieve all antibody-related complexes from the PDB using the chosen search approach. However, we encountered a notable number of false hits in the results, prompting us to implement a stringent structural filtering process. Antibody complexes were defined as structures containing at least one antibody molecule and another protein exceeding 50 amino acids in length. Structures that lacked antibodies or consisted solely of one type of antibody were excluded from further analysis. It is crucial to emphasize that not all proteins that bind to antibodies qualify as antigens. For instance, immunoglobulin-binding proteins like Protein A and Protein G, expressed

by Staphylococcus aureus and Streptococcal species, are commonly employed for antibody purification procedures [26]. Although protein A may bind to the Fab region of antibodies, the interacting amino acids might be located in the framework region (FR) rather than the complementarity-determining regions (CDRs) [27]. These proteins are often misidentified as antigen. Consequently, structures exhibiting such characteristics require meticulous verification and should be excluded from our dataset.

2.2 PDB splitting

We directly incorporated the PDB entry of antigen-antibody complex with only one antigen and one antibody into the AACDB database. However, quite a few antigen-antibody complexes contain several antigens and antibodies. In such cases, before splitting the structures, we need to determine the correct pairing of light and heavy chains by examining the information on equivalent chain interactions in the PDBsum database [28]. Next, we assign the correct antigen chains to the antibodies by identifying whether the remaining chains in the structure interact with the CDR regions of determined antibodies. For example, in the case of 1AHW, which contains two copies of antigen-antibody complexes, chains BA and DE are identified as two antibody pairs, where chains C and F bind to BA and DE, respectively. Consequently, 1AHW is split into two files, BAC and DEF (Figure 2A). In 6OGE, Pertuzumab (chains C and B) and Trastuzumab (chains E and D) bind to different epitopes of the receptor tyrosine-protein kinase erbB-2 (chain A). This generates two records in AACDB (CBA and EDA) (Figure 2B). Specifically, In the analysis of anti-idiotypic antibody complexes, the partner antibody is treated as a dual-chain antigen, necessitating individual evaluation of heavy chain and light chain interactions with the anti-idiotypic component. When both chains engage in binding, the complex is divided into two distinct entries: "anti-idiotypic antibody-heavy chain" and "anti-idiotypic antibody-light chain." Conversely, if only one chain participates, the non-interacting chain will be excluded from AACDB records. For instance, in 3BQU, the anti-idiotype 3H6 Fab (chains DC) only interact with the heavy chain (chain B) of 2F5 Fab. In AACDB, 2F5 Fab will be split and the light chain (chain A) will be discarded(Figure 2C).

For each PDB entry, we utilized the corresponding split chains information to divide the downloaded .pdb based on the ATOM records. Notably, certain structures are exclusively provided in the “mmCIF” format by the PDB database. While “.pdb” and “.cif” files store atomic coordinates in distinct text formats, the segmentation of these structure files is automatically performed based on manually annotated antibody-antigen chains. To address this, we integrated these considerations into our file processing pipeline enabling a fully automated file segmentation process. Additionally, the Naccess software does not support .cif files as input, we converted the “.cif” files into “.pdb”format files while performing the splitting process (**see Interacting residues definition**). Furthermore, we refined the annotation of the split “.fasta” files to ensure coherence with AACDB records.

2.3 Metadata

AACDB provides detailed metadata for each entry, including chain IDs, antibody name, antigen name, method, resolution, organism, and more. To ensure data accuracy, we have conducted comprehensive verification by consulting original literature sources. AACDB has addressed many annotation errors identified within the corresponding PDB entries. These errors include but not limited to: (1) mislabeling of species (e.g., the entry 7WRL where the organism of BD55-1239H antibody was erroneously labeled as “SARS coronavirus B012”); (2) Resolution

annotation errors (e.g., 1NSN, in which the resolution of 2.9 Å is misannotated as 2.8 Å). ; (3) mislabeling of antibody chains as other proteins (e.g., in 3KS0, the light chain of B2B4 antibody was misnamed as heme domain of flavocytochrome b2); (4) misidentification of heavy chains as light chains (e.g., both two chains of antibody were labeled as light chain in 5EBW); (5) mutation status annotation errors. We have identified cases in which PDB entries indicate "NO" for mutations, while in reality, mutations exist (e.g., bevacizumab (Avastin) in 6BFT was labeled as none mutation. When aligned with the bevacizumab sequence, however, mutation T8D/T30D in heavy chain and S52D/S53D in light chain were observed.); and (6) incomplete annotations. Certain entries only provide the name of the mutant without specifying the precise mutation site (e.g., in 7SU1, antibody was described as Ipilimumab variant Ipi.106. but PDB database do not provide any mutation amino acid or position, which can be identified by blast to Ipilimumab sequence).

We carefully checked each entry manually to find out all possible annotation problems. Given the context-dependent nature of mutations, sequence alignment was not universally performed across all antibody sequences. Instead, therapeutic antibody sequences served as definitive references. For example, for entries in the PDB labeled as "Bevacizumab mutant" or "Ipilimumab variant Ipi.106" where the "Mutation(s)" field was annotated as "NO," we retrieved corresponding therapeutic antibody sequences from Thera-SAbDab and performed sequence alignment with the PDB entries to identify mutated residues. For other antibodies lacking well-defined wild-type references, mutation information was directly extracted from the PDB or original literature. All corrections have been publicly archived in AACDB.

The antibody nomenclature follows the title of the corresponding search entry in the RCSB PDB database, with verification done through the original literature. In cases where the name in the original literature differs from that in the RCSB PDB, we used the name in the published literature as the standard. Furthermore, for antibody fragments lacking names in both the RCSB PDB database and original literature, we adopt a naming convention of "PDBID" + "antibody fragment" (e.g., 4WEB Fab).

Biological and physicochemical properties are critical considerationsin the development pipelines for therapeutic antibodies. These properties include solubility, immunogenicity, aggregation tendencies, expression level, stability, and hydrophobicity. We provide the International Non-proprietary Name (INN) and clinical trial information for each therapeutic antibody entry, linked to the DOTAD database [29]. Numerous antigens have been successfully identified as targets for antibodies or small molecule drugs. We conducted a comparison between antigen sequences and the drug targets listed in the DrugBank database [30]. A threshold of percent identity > 90% was applied to determine the corresponding drug targets.

2.4 Interacting residues definition

We labeled the interacting residues based on SASA and atom distance. Naccess V2.1.1 and Bio.PDB module were employed to calculate SASA value for each residue in antibody and antigen, respectively. The residues with a SASA loss (ΔSASA) in binding of more than 1Å^2 were classified as interacting residues. In addition, we also defined another set of interacting paratope-epitope residues by a distance cutoff of 6Å . Two amino acids are considered as interacting residues if they have at least one pair of non-hydrogen atoms atom within a distance of 6Å.

2.5 Data integration and website implementation

The main data processing algorithm is implemented in Python. The front-end web interface of AACDB was constructed by HTML and enhanced with JavaScript, CSS and Bootstrap technologies. We developed a dynamic 3D structure visualization window based on PV, a WebGL-based protein viewer, inspired by Dunbar *et al.* [21]. All the data were managed within the MySQL database system. For the back-end functionality, PHP is utilized to enable data browsing, searching, and downloading features.

3. Results

3.1 Statistics

From over 32000 experimental structures in the PDB database, we manually curated 7498 antigen−antibody entries for the current version of AACDB, encompassing to 16 antibody fragment types across 14 species(Figure 3). It is obvious that fab fragments and human antibodies accounted for the largest proportion of the data, accounting for 71.98% and 60.95%, respectively. Our statistical analysis reveals a significant increase in the number of antigen-antibody complex entries within the PDB during the period from 2021 to 2023. These entries accounted for approximately 45% of the total antigen-antibody complex entries. Our search was conducted with an end date of November 2023. Notably, Figure 2B contains a 2024-dated entry resulting from database versioning updates: the original 7SIX record (added November 16, 2022) was superseded by the 8TM1 entry through an official revision on January 12, 2024. Furthermore, the developability properties of antibodies in 325 entries can be queried in the DOTAD database, at the meanwhile, 3,733 antigen records have been identified as drug targets in DrugBank (data not shown).

3.2 Database browse and search

All the data can be browsed directly by clicking the “Datasets” item on the top menu (Figure 4). The summary table includes nine columns as following:

(1) AACDB_ID: The unique id in the AACDB database, linking to the “Detail” page;
(2) PDBID: The identifier of the RCSB PDB database (https://www.rcsb.org/), linking to PDB;
(3) Chains: The chain ids containing in this entry. Antibody and antigen chains were separated by “_”. The heavy chain id precedes the light chain id if antibody is complete;
(4) Antibody: This column is represented by “antibody name + fragment”;
(5) Protein: The corresponding antigen;
(6) Organism: The source organism of antibody and antigen;
(7) Method: The experiment method used to solve the structure;
(8) Resolution: The indicator that measures the resolution of protein structures in experiments is expressed in units of angstroms (Å).
(9) Reference: The DOI linker of original literature that produced this structure.

The table in AACDB can be easily searched through the search panel at the top section of the “Datasets” page. Users can perform a quick search by specifying one or more fields and entering relevant keywords. The search results can be downloaded as files in either txt or csv format. Figure 4 shows the search results using the keyword “Lysozyme” in “Protein” column, returning 117 hits.

An individual structure can be accessed using its AACDB_ID accession code. When click the AACDB_ID hyperlink, the user will be brought to its details page as shown in Figure 5. There, the

complex structure can be visualized with different colors and styles (Figure 5A). Besides the visualization window, we provide more details of this entry, including the mutation, INN and clinical trial of antibody, the ID in DrugBank of antigen (if exist) (Figure 5B). Under the structure information tab, further details about each chain can be found. These include: sequence, mutate amino acid type and position, interacting residues in each chain based on the ΔSASA method, interacting plot of paratope-epitope residues by a distance cutoff of < 6Å (Figure 5D). It should be pointed out that while our manuscript emphasizes widely accepted thresholds for consistency with prior benchmarks, AACDB explicitly provides raw ΔSASA and distance values for all annotated residues. Users can dynamically filter the data from the downloaded file (e.g., selecting 5Å instead of 6Å) by excluding entries exceeding their preferred thresholds.

3.3 Data download

We provide two ways for downloading the data:

(1) Download data of the single entry.

When access the detailed information about an entry using the corresponding AACDB_ID, user can click the hyperlink at the "Download" section of the bottom of the page to download the data for a single certain entry (Figure 5E).

(2) Download all the data of AACDB.

AACDB provide the download page for users. All the sequence and structure files and the interacting data based on different methods were packaged in different .zip file that can be downloaded.

Moreover, the website provides a user-friendly 'Help' page that presents a step-by-step tutorial to assist users in manipulating, querying, browsing and downloading the AACDB database.

4. Discussion and conclusion

Research on antigen-antibody interactions contributes to the advancement of the antibody-related industry. Databases such as PDB and SabDab provide the foundational data for this purpose. However, there are still many unresolved issues. In this study, we developed the AACDB database to provide a curated and reliable dataset of antigen-antibody interactions. During the process of data collection and organization, we identified numerous annotation errors in the PDB database. Some of these errors had been directly introduced into SabDab. For example, the species of the antibody in 7WRL was incorrectly labeled as "SARS coronavirus B012" in both PDB and SabDab. We also invested significant effort and time to manually cross-reference with original literature in order to rectify these errors and exclude antibody binding proteins that were erroneously annotated as antigens by SabDab. Apart from the curation and reannotation of structural data, AACDB offers features not provided by other antigen-antibody complex databases: 1) AACDB's data processing pipeline supports mmCIF files, and 2) we provide amino acids in the interaction interface through two methods, enabling the definition of unified standards for epitopes and paratopes. This provides a more accurate and comprehensive benchmark dataset for developed interaction interface prediction tools, enhancing the comparability of various tools.

However, AACDB still has some limitations. Despite our best efforts, the limitations of our team's resources and knowledge mean that our database may not capture all antigen-antibody complex structures. Since the data is manually curated, eliminating errors completely during the information processing is a challenge. While we strive to fill in any gaps, we also hope that experts and users within the community can provide timely feedback to help us improve these

issues. Additionally, currently AACDB only includes antigen proteins with a length greater than 50 amino acids. The current AACDB release exclusively includes antigen proteins exceeding 50 amino acids in length. Future iterations will expand structural coverage to diverse antigen classes (peptides, nucleic acids, haptens) while incorporating literature-validated affinity data through systematic cross-referencing.

In summary, AACDB is a novel database of antigen-antibody complexes that provides information on antibody-developability, antigen-drug target relationships, and detailed antigen-antibody interaction interfaces. It is fully accessible at http://i.uestc.edu.cn/AACDB. We are committed to a regular data update, ensuring that researchers in immunoinformatics have access to timely and valuable resources.

5. Acknowledgment(s)

This work was supported by the National Natural Science Foundation of China [grant numbers 62071099, 62371112] and Sichuan Province Science and Technology Support Program [2024NSFSC0636]. We are grateful to Steve's Scholar Memory Team for polishing our final manuscript.

6. CRediT author statement

Yuwei Zhou: Writing - Original Draft, Software, Conceptualization **Wenwen Liu:** Validation, Visualization **Ziru Huang:** Validation **Yushu Gou:** Data Curation **Siqi Liu:** Data Curation **Lixu Jiang:** Data Curation **Yue Yang:** Data Curation **Jian Huang:** Supervision, Funding acquisition

7. References

[1] X. Lyu, Q. Zhao, J. Hui, T. Wang, M. Lin, K. Wang, et al., The global landscape of approved antibody therapies, Antib Ther, 5(4) (2022) 233-257. https://doi.org/10.1093/abt/tbac021

[2] X.D. Su, Y. Shuai, Macromolecules and Antibody-Based Drugs, Adv Exp Med Biol, 1248 (2020) 485-530. https://doi.org/10.1007/978-981-15-3266-5_20

[3] P.J. Carter, A. Rajpal, Designing antibodies as therapeutics, Cell, 185(15) (2022) 2789-2805. https://doi.org/10.1016/j.cell.2022.05.029

[4] K. Tsuchikama, Y. Anami, S.Y.Y. Ha, C.M. Yamazaki, Exploring the next generation of antibody-drug conjugates, Nat Rev Clin Oncol, 21(3) (2024) 203-223. https://doi.org/10.1038/s41571-023-00850-2

[5] X.K. Chi, X.L. Xu, B.Y. Chen, J. Su, Y.Z. Du, Combining nanotechnology with

monoclonal antibody drugs for rheumatoid arthritis treatments, J Nanobiotechnology, 21(1) (2023) 105. https://doi.org/10.1186/s12951-023-01857-8

[6] Y. Zhou, Z. Huang, W. Li, J. Wei, Q. Jiang, W. Yang, et al., Deep learning in preclinical antibody drug discovery and development, Methods, 218 (2023) 57-71. https://doi.org/10.1016/j.ymeth.2023.07.003

[7] Y. Myung, D.E.V. Pires, D.B. Ascher, Understanding the complementarity and plasticity of antibody-antigen interfaces, Bioinformatics, 39(7) (2023). https://doi.org/10.1093/bioinformatics/btad392

[8] S. Pittala, C. Bailey-Kellogg, Learning context-aware structural representations to predict antigen and antibody binding interfaces, Bioinformatics, 36(13) (2020) 3996-4003. https://doi.org/10.1093/bioinformatics/btaa263

[9] T. Qiu, L. Zhang, Z. Chen, Y. Wang, T. Mao, C. Wang, et al., SEPPA-mAb: spatial epitope prediction of protein antigens for mAbs, Nucleic Acids Res, 51(W1) (2023) W528-W534. https://doi.org/10.1093/nar/gkad427

[10] C. Zhou, Z. Chen, L. Zhang, D. Yan, T. Mao, K. Tang, et al., SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens, Nucleic Acids Res, 47(W1) (2019) W388-W394. https://doi.org/10.1093/nar/gkz413

[11] T. Qi, T. Qiu, Q. Zhang, K. Tang, Y. Fan, J. Qiu, et al., SEPPA 2.0--more refined server to predict spatial epitope considering species of immune host and subcellular localization of protein antigen, Nucleic Acids Res, 42(Web Server issue) (2014) W59-63. https://doi.org/10.1093/nar/gku395

[12] A.M. Papadopoulos, A. Axenopoulos, A. Iatrou, K. Stamatopoulos, F. Alvarez, P.

Daras, ParaSurf: a surface-based deep learning approach for paratope-antigen interaction prediction, Bioinformatics, 41(2) (2025). https://doi.org/10.1093/bioinformatics/btaf062

[13] L. Chinery, N. Wahome, I. Moal, C.M. Deane, Paragraph-antibody paratope prediction using graph neural networks with minimal feature vectors, Bioinformatics, 39(1) (2023). https://doi.org/10.1093/bioinformatics/btac732

[14] D. Li, F. Pucci, M. Rooman, Prediction of Paratope-Epitope Pairs Using Convolutional Neural Networks, Int J Mol Sci, 25(10) (2024). https://doi.org/10.3390/ijms25105434

[15] E. Liberis, P. Velickovic, P. Sormanni, M. Vendruscolo, P. Lio, Parapred: antibody paratope prediction using convolutional and recurrent neural networks, Bioinformatics, 34(17) (2018) 2944-2950. https://doi.org/10.1093/bioinformatics/bty305

[16] M. Kalemati, A. Noroozi, A. Shahbakhsh, S. Koohi, ParaAntiProt provides paratope prediction using antibody and protein language models, Sci Rep, 14(1) (2024) 29141. https://doi.org/10.1038/s41598-024-80940-y

[17] T.I. Shashkova, D. Umerenkov, M. Salnikov, P.V. Strashnov, A.V. Konstantinova, I. Lebed, et al., SEMA: Antigen B-cell conformational epitope prediction using deep transfer learning, Front Immunol, 13 (2022) 960985. https://doi.org/10.3389/fimmu.2022.960985

[18] S.K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, H. Chao, L. Chen, et al., RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning, Nucleic Acids Res, 51(D1) (2023) D488-D508.

https://doi.org/10.1093/nar/gkac1077

[19] S.K. Burley, D.W. Piehl, B. Vallat, C. Zardecki, RCSB Protein Data Bank: supporting research and education worldwide through explorations of experimentally determined and computationally predicted atomic level 3D biostructures, IUCrJ, 11(Pt 3) (2024) 279-286. https://doi.org/10.1107/S2052252524002604

[20] S. Ferdous, A.C.R. Martin, AbDb: antibody structure database-a database of PDB-derived antibody structures, Database (Oxford), 2018 (2018). https://doi.org/10.1093/database/bay040

[21] J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, et al., SAbDab: the structural antibody database, Nucleic Acids Res, 42(Database issue) (2014) D1140-6. https://doi.org/10.1093/nar/gkt1043

[22] M.I.J. Raybould, C. Marks, A.P. Lewis, J. Shi, A. Bujotzek, B. Taddese, et al., Thera-SAbDab: the Therapeutic Structural Antibody Database, Nucleic Acids Res, 48(D1) (2020) D383-D388. https://doi.org/10.1093/nar/gkz827

[23] C. Schneider, M.I.J. Raybould, C.M. Deane, SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker, Nucleic Acids Res, 50(D1) (2022) D1368-D1372. https://doi.org/10.1093/nar/gkab1050

[24] J. Sun, D. Wu, T. Xu, X. Wang, X. Xu, L. Tao, et al., SEPPA: a computational server for spatial epitope prediction of protein antigens, Nucleic Acids Res, 37(Web Server issue) (2009) W612-6. https://doi.org/10.1093/nar/gkp417

[25] S. Daberdaku, C. Ferrari, Antibody interface prediction with 3D Zernike descriptors and SVM, Bioinformatics, 35(11) (2019) 1870-1876.

https://doi.org/10.1093/bioinformatics/bty918

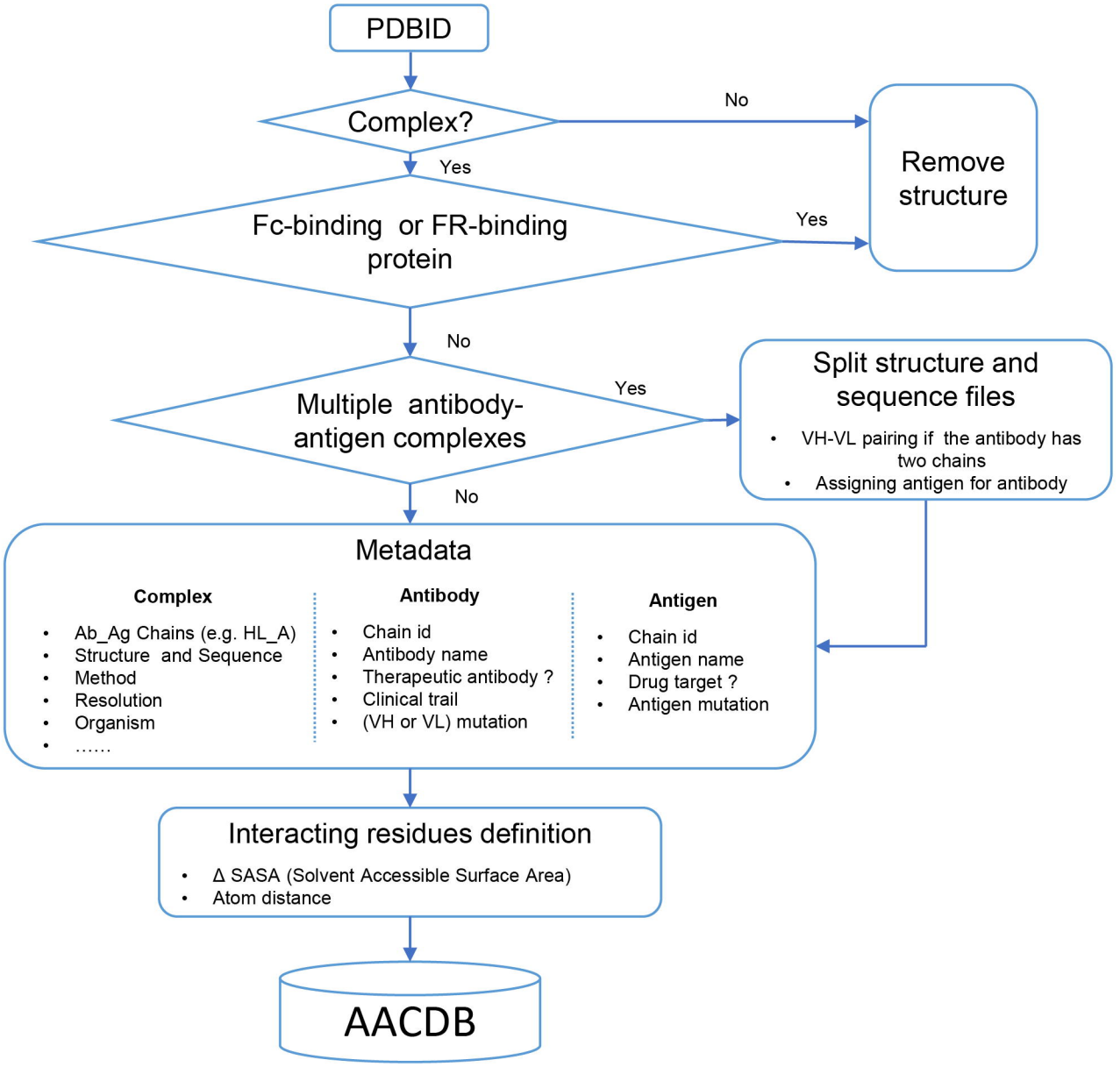
[26] J.B. Fishman, E.A. Berg, Protein A and Protein G Purification of Antibodies, Cold Spring Harb Protoc, 2019(1) (2019). https://doi.org/10.1101/pdb.prot099143

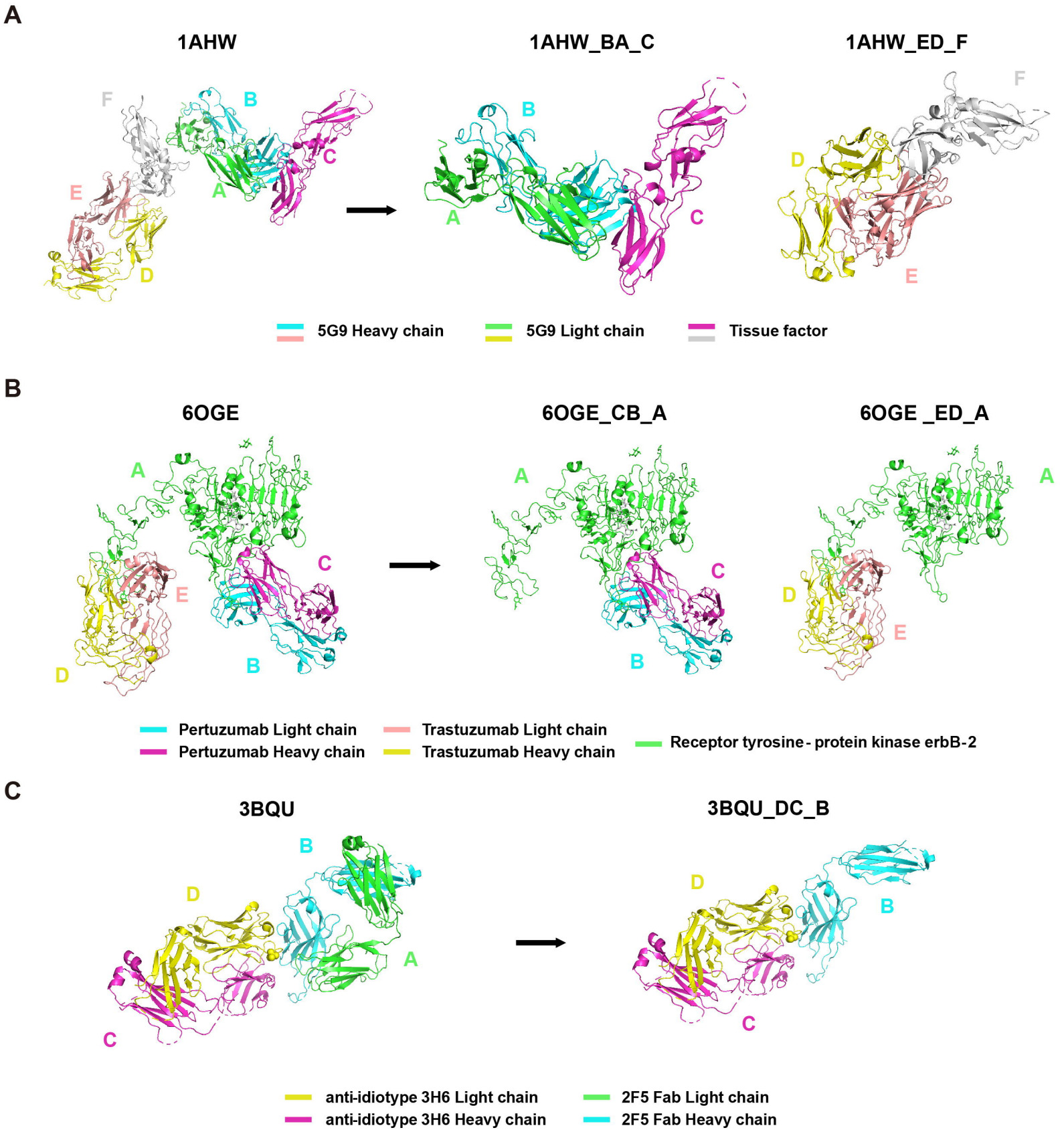
[27] M. Graille, E.A. Stura, A.L. Corper, B.J. Sutton, M.J. Taussig, J.B. Charbonnier, et al., Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity, Proc Natl Acad Sci U S A, 97(10) (2000) 5399-404. https://doi.org/10.1073/pnas.97.10.5399

[28] R.A. Laskowski, J. Jablonska, L. Pravda, R.S. Varekova, J.M. Thornton, PDBsum: Structural summaries of PDB entries, Protein Sci, 27(1) (2018) 129-134. https://doi.org/10.1002/pro.3289

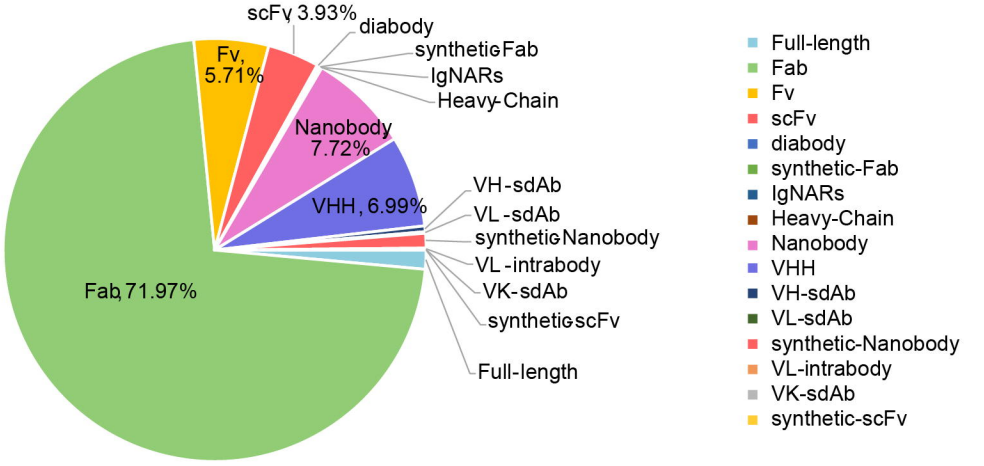
[29] W. Li, H. Lin, Z. Huang, S. Xie, Y. Zhou, R. Gong, et al., DOTAD: A Database of Therapeutic Antibody Developability, Interdiscip Sci, (2024). https://doi.org/10.1007/s12539-024-00613-2

[30] C. Knox, M. Wilson, C.M. Klinger, M. Franklin, E. Oler, A. Wilson, et al., DrugBank 6.0: the DrugBank Knowledgebase for 2024, Nucleic Acids Res, 52(D1) (2024) D1265-D1275. https://doi.org/10.1093/nar/gkad976

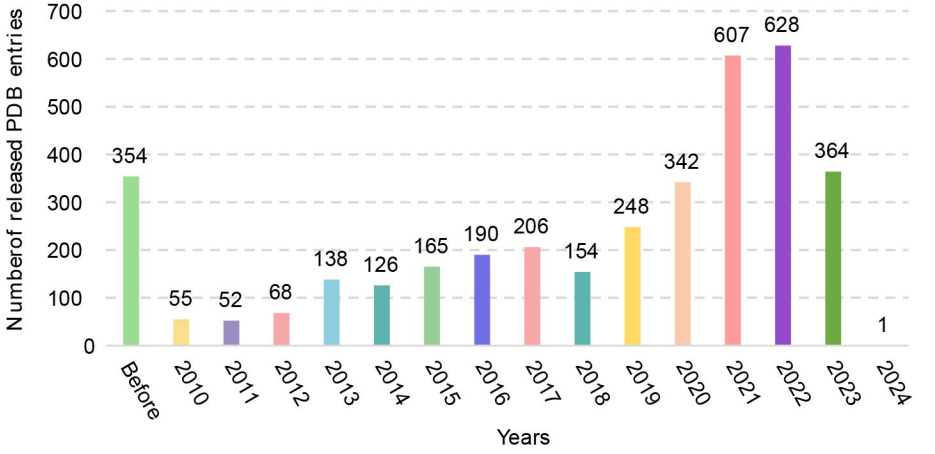

PDBID
Complex?
No
Yes
Remove structure
Fc-binding or FR-binding protein
Yes
No
Multiple antibody-antigen complexes
Yes
Split structure and sequence files
VH-VL pairing if the antibody has two chains
Assigning antigen for antibody
No
Metadata
Complex
Ab_Ag Chains (e.g. HL_A)
Structure and Sequence
Method
Resolution
Organism
……
Antibody
Chain id
Antibody name
Therapeutic antibody ?
Clinical trail
(VH or VL) mutation
Antigen
Chain id
Antigen name
Drug target ?
Antigen mutation
Interacting residues definition
Δ SASA (Solvent Accessible Surface Area)
Atom distance
AACDB


A
1AHW
1AHW_BA_C
1AHW_ED_F
F
B
C
E
A
D
5G9 Heavy chain
5G9 Light chain
Tissue factor
B
6OGE
6OGE_CB_A
6OGE _ED_A
Pertuzumab Light chain
Pertuzumab Heavy chain
Trastuzumab Light chain
Trastuzumab Heavy chain
Receptor tyrosine - protein kinase erbB-2
C
3BQU
3BQU_DC_B
anti-idiotype 3H6 Light chain
anti-idiotype 3H6 Heavy chain
2F5 Fab Light chain
2F5 Fab Heavy chain

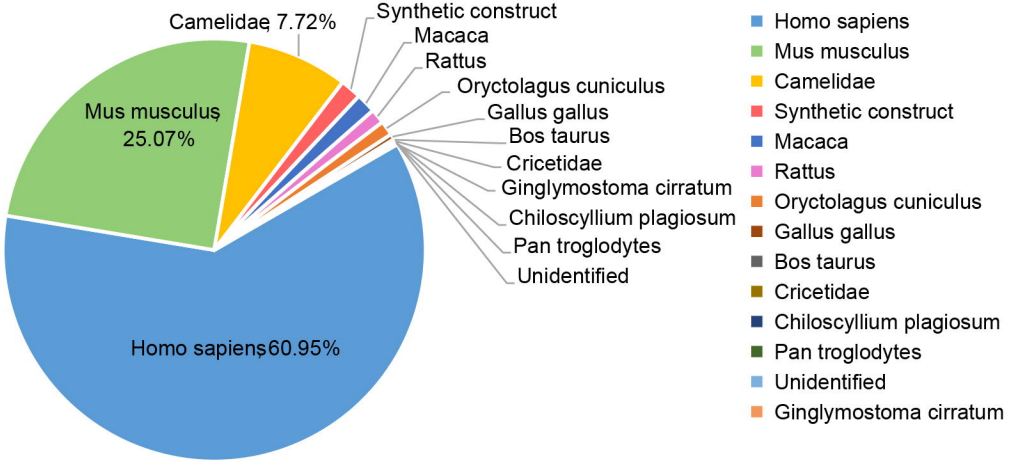
A


scFv, 3.93%
diabody
synthetic-Fab
IgNARs
Heavy-Chain
Fv, 5.71%
Nanobody 7.72%
VHH, 6.99%
VH-sdAb
VL-sdAb
synthetic-Nanobody
VL-intrabody
VK-sdAb
synthetic-scFv
Full-length
Fab, 71.97%
Full-length
Fab
Fv
scFv
diabody
synthetic-Fab
IgNARs
Heavy-Chain
Nanobody
VHH
VH-sdAb
VL-sdAb
synthetic-Nanobody
VL-intrabody
VK-sdAb
synthetic-scFv

B


Number of released PDB entries
700
600
500
400
300
200
100
0
354
55
52
68
138
126
165
190
206
154
248
342
607
628
364
1
Before
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
Years

C


Camelidae, 7.72%
Synthetic construct
Macaca
Rattus
Oryctolagus cuniculus
Gallus gallus
Bos taurus
Cricetidae
Ginglymostoma cirratum
Chiloscyllium plagiosum
Pan troglodytes
Unidentified
Mus musculus, 25.07%
Homo sapiens, 60.95%
Homo sapiens
Mus musculus
Camelidae
Synthetic construct
Macaca
Rattus
Oryctolagus cuniculus
Gallus gallus
Bos taurus
Cricetidae
Chiloscyllium plagiosum
Pan troglodytes
Unidentified
Ginglymostoma cirratum

Datasets

Search

PDBID: | Antibody: | Protein: Lysozyme

Organism: | Resolution(<): | All fields:

Submit search → Search

Click to download the search results

Results list: **117** hits (Download the results: or)

Click for the details

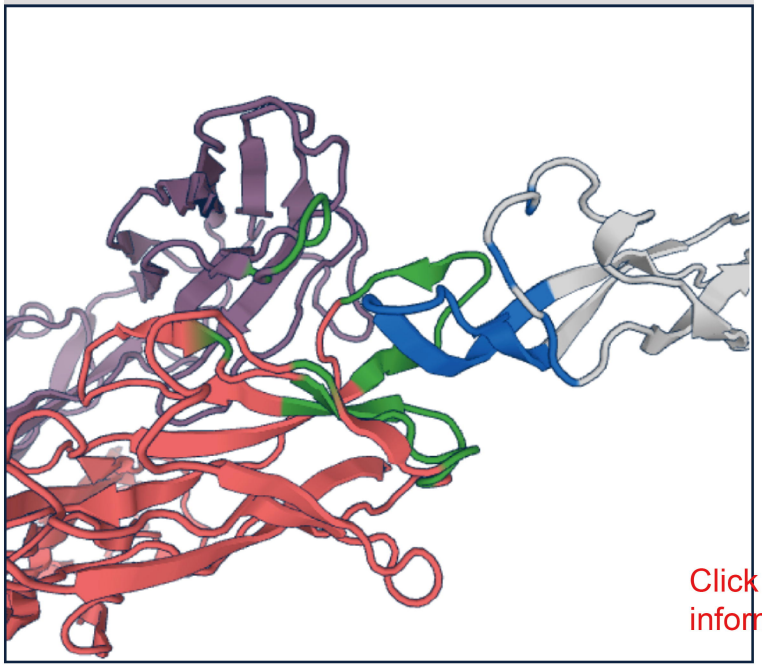
Click to get original literature

Click to view the entry in PDB database

AACDB_ID	PDBID	Chains	Antibody	Protein	Organism	Method	Resolution	Reference
1	1A2Y	BA_C	D1.3 Fv	Lysozyme C	Mus musculus, Gallus gallus	XRD	1.50	10.1021/bi980148j
2	1BQL	HL_Y	HYHEL-5 Fab	BOBWHITE QUAIL Lysozyme.	Mus musculus, Colinus virginianus	XRD	2.60	10.1002/(SICI)1097-0134(199609)26:1<55::AID-PROT5>3.0.CO;2-F
3	1BVK	BA_C	HULYS11 Fv	Lysozyme	Homo sapiens, Gallus gallus	XRD	2.70	10.1084/jem.187.4.479
4	1BVK	ED_F	HULYS11 Fv	Lysozyme	Homo sapiens, Gallus gallus	XRD	2.70	10.1084/jem.187.4.479
5	1C08	BA_C	HYHEL-10 Fv	Hen Egg White Lysozyme	Mus musculus, Gallus gallus	XRD	2.30	10.1074/jbc.274.39.27623
6	1DQJ	BA_C	HyHEL-63 Fab	Lysozyme C	Mus musculus, Gallus gallus	XRD	2.00	10.1021/bi000054l
7	1DZB	A_X	1F9 scFv	Turkey Egg White Lysozyme	Mus musculus, Meleagris gallopavo	XRD	2.00	10.1006/jmbi.2000.3971
8	1DZB	B_Y	1F9 scFv	Turkey Egg White Lysozyme	Mus musculus, Meleagris gallopavo	XRD	2.00	10.1006/jmbi.2000.3971
9	1FBI	HL_X	F9.13.7 Fab	GUINEA FOWL Lysozyme	Mus musculus, Numida meleagris	XRD	3.00	10.1074/jbc.270.30.18067
10	1FBI	QP_Y	F9.13.7 Fab	GUINEA FOWL	Mus musculus, Numida	XRD	3.00	10.1074/jbc.270.30.18067

A

Structure visualisation



Heavy Chain
Light Chain
Antigen Chain
Epitope
Paratope
☐ Spin on/off

Display styles :
○ Spacefill
○ Wire
○ Ball&stick
◉ Cartoon

Display color :
◉ Default
○ Color by B-factor
○ Color by sec.structure
○ Color by element

B

Entry information

Complex	
PDBID:	1BJ1
Chains:	HL_W
Organism:	Mus musculus, Homo sapiens
Method:	XRD
Resolution (Å):	2.40
Reference:	10.1016/s0969-2126(98)00116-6
Antibody	
Antibody:	12 Fab
Antibody mutation:	No
INN (Clinical Trial):	Bevacizumab(Approved); Dilpacimab(Phase-II); Ivonescimab(Preregistration);
Antigen	
Antigen:	Vascular endothelial growth factor A
Antigen mutation:	No
Durg Target:	P15692

Click to access the developability information in DOTAD database

Click to access the target information in DrugBank database

C

Sequence information

Antibody

Heavy Chain: H
Mutation: NULL

>1BJ1_H|Chain B[auth H], F[auth K]|Fab fragment, heavy chain|Mus musculus (10090)
EVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYTGEPTYAADFKRRFTFSLDTSKST
AYLQMNSLRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFP
EPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHT

Light Chain: L
Mutation: NULL

>1BJ1_L|Chain A[auth J], E[auth L]|Fab fragment, light chain|Mus musculus (10090)
DIQMTQSPSSLSASVGDRVTITCSASQDISNYLNWYQQKPGKAPKVLIYFTSSLHSGVPSRFSGSGSGTDFTLTISSLQPE
DFATYYCQQYSTVPWTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQ
ESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Antigen

Chain: W
Mutation: NULL

>1BJ1_W|Chain C[auth V], D[auth W]|Vascular endothelial growth factor A|Homo sapiens (9606)
GQNHHEVVKFMDVYQRSYCHPIETLVDIFQEYPDEIEYIFKPSCVPLMRCGGCCNDEGLECVPTEESNITMQIMRIKPHQ
GQHIGEMSFLQHNKCECRPKKD

D

Interaction

1、**Solvent accessible surface areas (SASA)** were calculated (Naccess V2.1.1) for each residue in antibody and antigen, respectively. The residues with SASA loss in binding of more than 1Å^2 were classified as interacting residues.

Interacting residues (ΔSASA based)

Chain residues position delta_SASA : residuesposition

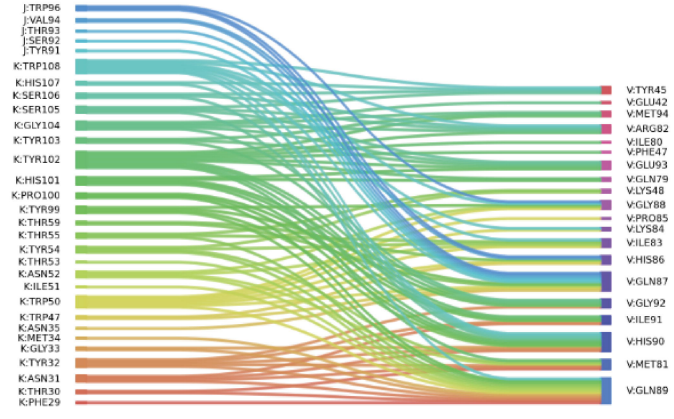
J: TYR91 SER92 THR93 VAL94 TRP96

K: THR30 ASN31 TYR32 GLY33 TRP50 ASN52 THR53 TYR54 THR55 THR59 TYR99 PRO100 HIS101 TYR102 TYR103 GLY104 SER105 SER106 HIS107 TRP108

V: TYR45 LYS48 GLN79 ILE80 MET81 ARG82 ILE83 LYS84 HIS86 GLN87 GLY88 GLN89 HIS90 ILE91 GLY92 GLU93 MET94

2、We defined interacting paratope-epitope residues by a **distance cutoff of < 6 Å**. Two amino acids are considered as interacting residues if they have at least one atom within a distance of 6 Å from any atom.

Interacting residues (Atom distance based)



E

Download

Download sequences
Download structure
Download interacting residues (ΔSASA based)
Download interacting residues (Distance based)