OXFORD

Antibody Therapeutics, 2024, 7(3), 256–265
https://doi.org/10.1093/abt/tbae018
Advance access publication date 8 July 2024
Research Article

rAbDesFlow: a novel workflow for computational recombinant antibody design for healthcare engineering

Sowmya Ramaswamy Krishnan[1,‡], Divya Sharma[1,‡], Yasin Nazeer[1], Mayilvahanan Bose[2], Thangarajan Rajkumar[3,4,5], Guhan Jayaraman[1], Narayanan Madaboosi[1,*], M. Michael Gromiha[1,6,7,*]

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India
[2]Department of Molecular Oncology, Cancer Institute (WIA), Adyar, Chennai 600020, India
[3]Department of Applied Mechanics and Biomedical Engineering, Indian Institute of Technology Madras, Chennai 600036, India
[4]MedGenome, Bengaluru 560099, Karnataka, India
[5]Department of Nanosciences and Molecular Medicine, Amrita Institute of Medical Sciences, Kochi 682041, Kerala, India
[6]International Research Frontiers Initiative, School of Computing, Tokyo Institute of Technology, Yokohama 226-8501, Japan
[7]School of Computing, National University of Singapore (NUS), Singapore 119077, Singapore

[*]Corresponding authors. Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India. E-mail: narayananms@smail.iitm.ac.in; gromiha@iitm.ac.in
[‡]Sowmya Ramaswamy Krishnan and Divya Sharma with equal contribution.

Abstract

Recombinant antibodies (rAbs) have emerged as a promising solution to tackle antigen specificity, enhancement of immunogenic potential and versatile functionalization to treat human diseases. The development of single chain variable fragments has helped accelerate treatment in cancers and viral infections, due to their favorable pharmacokinetics and human compatibility. However, designing rAbs is traditionally viewed as a genetic engineering problem, with phage display and cell free systems playing a major role in sequence selection for gene synthesis. The process of antibody engineering involves complex and time-consuming laboratory techniques, which demand substantial resources and expertise. The success rate of obtaining desired antibody candidates through experimental approaches can be modest, necessitating iterative cycles of selection and optimization. With ongoing advancements in technology, *in silico* design of diverse antibody libraries, screening and identification of potential candidates for *in vitro* validation can be accelerated. To meet this need, we have developed rAbDesFlow, a unified computational workflow for recombinant antibody engineering with open-source programs and tools for ease of implementation. The workflow encompasses five computational modules to perform antigen selection, antibody library generation, antigen and antibody structure modeling, antigen–antibody interaction modeling, structure analysis, and consensus ranking of potential antibody sequences for synthesis and experimental validation. The proposed workflow has been demonstrated through design of rAbs for the ovarian cancer antigen Mucin-16 (CA-125). This approach can serve as a blueprint for designing similar engineered molecules targeting other biomarkers, allowing for a simplified adaptation to different cancer types or disease-specific antigens.

Statement of Significance: A computational workflow, rAbDesFlow, was developed for template-based humanized recombinant antibody design. The workflow utilizes open-sourced databases and tools with existing antibody sequences as templates. This approach can serve as a blueprint for designing similar engineered molecules targeting other biomarkers, allowing for a straightforward and simplified adaptation.

Keywords: rAbDesFlow; recombinant antibody engineering; computational workflow; CA-125; ovarian cancer

Introduction

Recombinant antibodies (rAbs) have indeed revolutionized the field of medicine by overcoming the challenges posed by hybridoma-derived monoclonal antibodies (mAbs), providing higher affinity and antigen specificity along with capabilities to withstand a wide range of functionalization elements [1]. The groundbreaking advent of hybridoma technology in 1975 led to the development of the first monoclonal antibody, which was subsequently licensed in 1986 [2, 3]. Hybridoma technology fuses mouse B cells producing specific antibodies with immortalized myeloma cells, resulting in the generation of hybrid cells that produce mAbs with specificity to the mouse-derived antigen. Initially, the production of antibodies heavily relied on animal immunization, employing experimental mice, rabbits, and other laboratory animals until the late 1980s [4]. However, a significant challenge encountered in the production and application of mAbs lies in the inadequate immune response elicited against highly toxic or conserved antigens. In addition, antibody-producing immune response in hybridoma is initiated by proteolysis of the antigen, leading to inadequate affinity of the derived mAbs to the native antigen [1]. Consequently, most clinical antibodies are either derived from human sources or undergo humanization to mitigate potential immunogenicity concerns [5]. While the development of transgenic mice and rabbits carrying human

Received: February 5, 2024. **Revised:** May 11, 2024

antibody genes has addressed the issue of immunogenicity, it does not address the necessity of an effective immune response following immunization. To transcend these challenges in improving affinity, selectivity and scalability in antibody production, a shift from monoclonal to recombinant antibody engineering was initiated [6].

rAbs are obtained through an *in vitro* generation processes using various antibody engineering technologies such as phage display, construction of antibody fragments, immunomodulatory antibodies, and cell-free systems [7], and through computational approaches. Phage display technology utilizes bacteriophages to present a vast library of antibody fragments, enabling the selection and production of mAbs with high specificity and affinity against a target antigen [8]. Approaches have also emerged to genetically engineer host cells to produce and secrete antibodies, enabling scalable and customizable production of highly specific therapeutic molecules [9]. The emergence of bispecific antibodies with multiple engineered subdomains capable of recognizing multiple epitopes simultaneously showcases the capability of genetic engineering to revolutionalize antibody-based therapy [10]. Single chain variable fragment (scFv), nanobodies, suprabodies, and affibodies have also emerged as versatile antibody formats. scFv offer smaller size and improved tissue penetration, nanobodies possess high stability and target accessibility, while affibodies exhibit small size, stability, and broad target recognition, collectively expanding the possibilities for targeted therapies and diagnostics [11]. Collectively, recombinant antibody production has also decreased the usage of animal models, while maintaining high affinity and scalability to a wide range of antigenic fragments.

In past few years, there has been a notable surge in the approval and investigation of engineered antibody drugs in phase II or III clinical trials, demonstrating the accelerated progress and growing interest in utilizing these therapeutics for diverse diseases [12]. FDA-approved antibodies, including Necitumumab, Nivolumab, and ramucirumab, have shown efficacy in treating various cancers by targeting specific antigens [3, 13]. Additionally, rAbs such as Bamlanivimab [14] and mAbs such as Bebtelovimab, REGEN-CoV, and others have received emergency use authorization for the treatment of COVID-19 [15], further highlighting the broad applications of engineered antibodies in addressing critical medical needs. Two novel bispecific rAbs namely, Epcoritamab and Glofitamab were approved in 2023 by FDA, to treat relapsed or refractory diffuse large B-cell lymphoma [16]. In addition to engineered antibodies, plasma therapy has garnered attention and has been investigated for various infectious diseases, including COVID-19 [17]. It involves using plasma derived from individuals who have recovered from a particular disease to treat those currently battling the same illness. This therapeutic approach harnesses the presence of antibodies present in the plasma of recovered individuals to provide passive immunity and aid in the neutralization of the pathogen. Apart from the bispecific rAbs, direct targeted engineered molecules for therapeutics include antibody-drug conjugates and CAR-T cell therapies [18–20]. These molecules are specifically designed to target and bind to specific antigens or receptors on diseased cells, delivering therapeutic agents or activating immune responses for targeted treatment.

Although antibody engineering has led to significant advancements in the development of therapeutic antibodies, it is important to recognize the constraints related to factors such as time, antigen immunogenicity, production scale, and other experimental approaches. The process of antibody engineering involves complex and time-consuming laboratory techniques such as protein expression, purification, and characterization, which demand substantial resources and expertise. Additionally, traditional methods for generating antibody libraries, such as immunization or hybridoma technology, have constraints in terms of library diversity compared to computationally designed libraries. Experimental screening methods, relying on assays like Enzyme-Linked Immunosorbent Assay (ELISA) or flow cytometry, can be laborious and slow, hindering high-throughput screening for identifying antibodies with high affinity. Furthermore, the success rate of obtaining desired antibody candidates through experimental approaches can be modest, necessitating iterative cycles of selection and optimization. However, ongoing advancements in technology and the integration of computational methods could offer promising solutions to overcome these limitations. Computational tools encompass antibody design software [21], which allows for sequence modifications and optimization of properties. Epitope prediction tools [22, 23] aid in identifying potential binding sites on antigens, while antibody–antigen docking tools [24, 25] simulate binding interactions. Antibody library design tools [26] help to create diverse libraries for screening, and sequence analysis tools [27] enable the examination of antibody sequences. Additionally, protein engineering and optimization tools [28, 29], although not antibody-specific, assist in improving stability and designing variants. Collectively, these tools streamline multiplex antibody development, enabling the design of antibodies with enhanced properties such as specificity, affinity, stability, and therapeutic efficacy [30]. Therefore, there is a need for a pipeline that integrates various computational tools and algorithms to handle different aspects of antibody engineering, such as structure prediction, epitope identification, and property optimization.

In this study, we focus on using the epithelial ovarian cancer biomarker CA-125 as a model antigen to design and map an engineered antibody for both onco-diagnostic and onco-therapeutic purposes. In this regard, we developed a computational workflow, rAbDesFlow, for template-based humanized recombinant antibody design. The workflow utilizes completely open-sourced databases and tools and *in silico* site saturation mutagenesis using existing antibody sequences as templates. The resulting antibodies were extensively analyzed based on various profiling metrics and humanness scores. The top 10 antibodies were selected and used to model the antigen–antibody interaction through structure prediction and docking methods. Through detailed analysis, we identified four predominant epitopes on the antigen and prioritized the top 3 antibodies for further experimental validation based on predicted binding energy. This approach can serve as a blueprint for designing similar engineered molecules targeting other biomarkers, allowing for a straightforward and simplified adaptation to different cancer types or disease-specific antigens. By leveraging this strategy, we can potentially enhance the accuracy and effectiveness of cancer diagnosis and treatment by specifically targeting relevant biomarkers associated with different malignancies.

Materials and methods

The workflow of rAbDesFlow involves five distinct modules (Fig. 1), which provide an end-to-end solution for the computational antibody design problem from sequence engineering to antigen–antibody binding mode prediction. Since the workflow is based only on open-source programs and tools, it can be readily implemented in laboratories for accelerating antibody engineering studies. The functionality of each module in rAbDesFlow is explained in detail below. The workflow was tested on a case study to design antibodies against the ovarian cancer

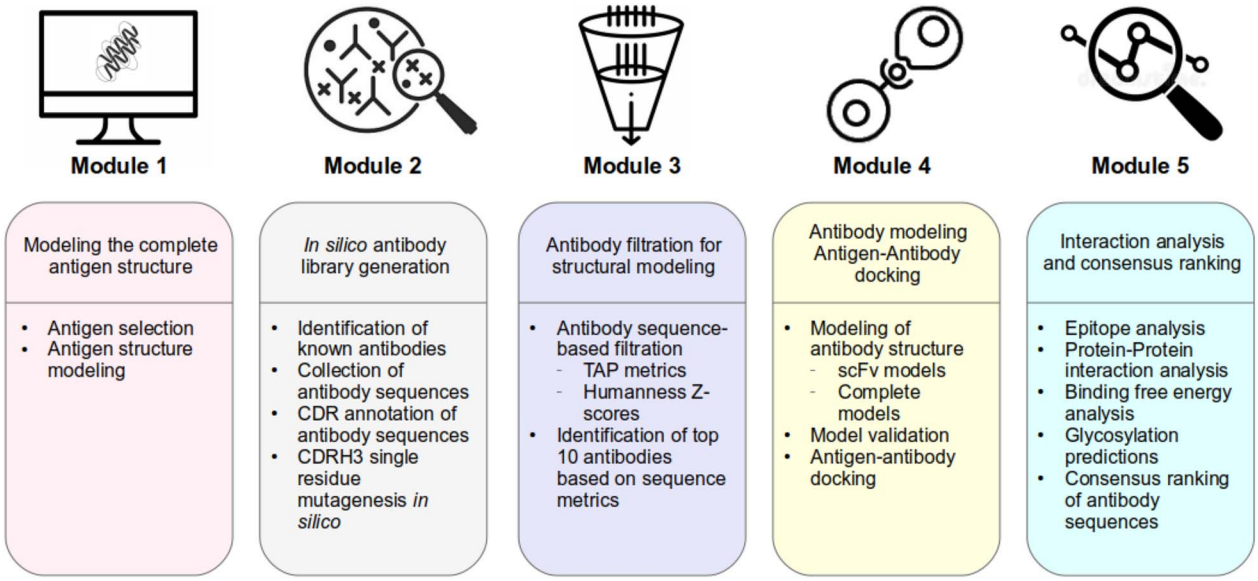

Figure 1. A schematic of the rAbDesFlow method showing the five computational modules (M1 to M5) developed for computational antibody design. The different databases and tools used in each step of the workflow are provided as in supplementary information (Table S1).

antigen, Mucin-16 or CA-125 [31]. It is notable that the workflow has been designed to use open-source standalone tools and web servers only, for ease of accessibility. Also, wherever possible, the same tool was repeatedly used for similar tasks to ensure minimal resource overheads. The complete list of software used in this workflow is provided in the Supplementary Information (Table S1).

Module 1: Modeling the complete antigen structure

The protein sequence of the antigen of interest is obtained from the UniProt database [32] and it is compared with the protein sequence obtained from experimentally determined structures, if available. If there are missing residues in comparison with the UniProt-derived sequence, the complete antigen structure is modeled using multiple computational 3D structure prediction methods such as ITASSER [33], AlphaFold2 [34], and RoseTTAFold [35]. The best model of the antigen is selected based on metrics specific to each structure prediction program such as C-Score (ITASSER), confidence score (AlphaFold2), and energy value (RoseTTAFold). The agreement between the three selected models from each structure prediction programs is quantified using root-mean-square deviation (RMSD) and the model with better terminal and loop region folding is chosen for further analysis. Since AlphaFold2 has been shown to have poor performance in modeling long loop regions [36, 37], only the RMSD of structured domains of the antigen are considered during the comparison [37].

Module 2: Library preparation for antibody sequences

The second module utilizes information from experimentally validated antibodies known against the antigen of interest to enumerate a set of new antibody sequences, which can potentially elicit better binding affinity and activity against the antigen *in vitro*. The light and heavy chain sequences of the template antibodies are collected from the TheraSAbDab database [38] or can be supplied by the user. The complementarity determining regions (CDRs) of the light and heavy chain sequences are mapped using the Kabat CDR definitions [39] available via the ANARCI python package [40]. Based on existing knowledge of the variability of CDR regions among antibodies, the CDRH3 region is chosen for single residue substitution, as this region is known to be highly variable [41] and hence the predominant determinant of antibody specificity. Each residue in CDRH3 region is substituted with the other 19 residues, resulting in a library of antibody sequences per template antibody utilized. The residue substitutions are automated using custom Python scripts.

The enumerated antibody sequences within each library are subject to five filtering metrics available in the Therapeutic Antibody Profiler (TAP) server [42]. The TAP metrics were developed based on a database of therapeutic humanized mAbs and their optimal ranges were taken from the TAP study [42]. Antibody sequences which pass all five metrics from TAP were only considered for further analysis. Post-filtration with TAP metrics, the humanness score of the antibody sequences is computed using the HScore server [43]. The returned z-scores are used to prioritize the antibody sequences for modeling their three-dimensional structures. The top 10 antibodies in terms of their humanness z-score are considered for detailed structure modeling studies and antigen–antibody interaction analysis.

Module 3: Modeling the three-dimensional antibody structure

The antibody structure can be modeled using two approaches: scFv modeling and complete antibody modeling. scFv is similar to camelid nanobodies in structure and includes only the Fab regions of the light and heavy chains of the antibody, which is sufficient for eliciting an immune response as proven by experiments [44]. The antibody grafting mode of the Rosetta ROSIE server [45] is used for *ab initio* modeling of the scFv antibody structure. Rosetta returns the top 1000 antibody models for a given sequence, by performing extensive CDRH3 loop modeling and optimization. The binding free energies of the top 1000 models are used to prioritize the final antibody scFv model for antigen–antibody docking. Using this procedure, the antibody scFv models for the top 10 antibody sequences prioritized in module 2 are obtained, along with the structures for the template antibodies considered. The models obtained from Rosetta were also compared with the scFv models from the OPIG antibody modeling server [40]. The

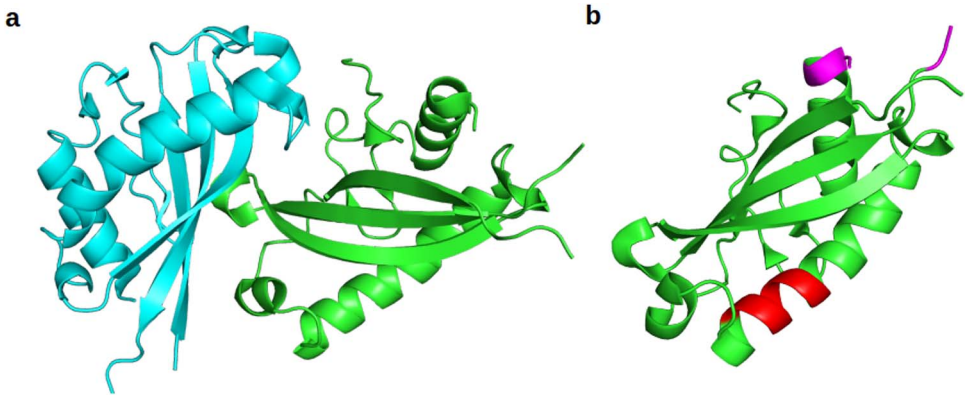

Figure 2. (a) Crystal structure of the partial human CA-125 SEA domain homodimer (PDB ID: 7SA9); (b) the epitopes of known CA-125 antibodies (5E11, OC125 and M11) mapped to the SEA domain monomer.

RMSD between the two models is analyzed for a good agreement in the overall structure obtained.

ITASSER server is used to model the complete antibody structure containing both the Fab and Fc regions. The sequence of the Fc region of one of the four immunoglobulins (IgG1, IgG2, IgG3, or IgG4) can be considered for predicting the model, depending on the predominance of the IgG subtype in related humanized monoclonal antibodies. The Fab and Fc region sequences are combined into the complete light chain and heavy chain sequence for each antibody considered, and are provided as input to ITASSER. ITASSER was chosen as the primary method of choice for complete antibody design, as AlphaFold2 multimer [46] models did not yield good results for the same set of sequences. Latest antibody-specific deep learning models for structure prediction such as IgFold [47] can also be used instead of ITASSER in this module.

Module 4: Antigen–antibody docking using ClusPro advanced docking protocol

The antigen–antibody complex for the wild-type (template antibodies) and top 10 prioritized antibodies are obtained using ClusPro antibody docking mode [48]. During the docking process, the antibody is used as the receptor and antigen is used as the ligand, as suggested by the ClusPro documentation. Thirty models are built for each complex and the final model is chosen based on an analysis of the orientation of the antigen among all antibodies, to ensure that the epitope region is nearly identical between them. This will also ensure that the results of the residue substitution will be comparable between the antibodies thereby, allowing the identification of the substitution with maximal contribution to the overall binding affinity of the complex. Based on the epitope analysis, multiple competing orientations of the antigen with respect to the antibody are identified. Further analysis is performed on all orientations of the antigen–antibody complex to understand the orientation, which provides a better binding affinity compared with the selected template antibody.

Module 5: Computational prediction of the antigen–antibody interaction

The predominant residue-residue interactions present at the interface of the antigen–antibody complexes obtained from ClusPro were extracted using the RING-3.0 stand-alone program [49]. In terms of the interactions observed, the predominant interaction types among the complexes are identified along with their frequency of occurrence. The paratope and epitope residues for each antigen–antibody complex are extracted from the observed interactions and are compared to the known epitopes characterized for the template antibodies utilized in module 1 and from literature. The top 10 antigen–antibody complexes are further ranked on the basis of their predicted binding free energies and dissociation constants (K_d) from the PRODIGY server [50]. A consensus ranking of the antibodies is derived based on the predicted binding free energy of different orientations of the antibody with respect to the antigen. The consensus ranking can be used to prioritize antibodies based on their different predicted binding modes *in silico*.

Further, the prioritized antibodies are also filtered based on the observed overlap in the epitope regions, to avoid steric hindrance upon binding to antibody during *in vitro* evaluation. The final set of antibodies obtained are also substantiated with information on potential glycosylation hotspots and solvent accessibility to optimize the solubility of the antibody for experimental validation.

Results and discussion

The proposed workflow was validated through the design of scFv antibodies against the human ovarian cancer antigen, Mucin-16, also known as CA-125. The ovarian cancer antigen CA-125 has been extensively studied as a therapeutic target and biomarker for humanized monoclonal and rAbs, due to its essential role in the carcinogenesis process [31]. Although most clinicians recommends CA-125 test for individuals with symptoms of ovarian cancer, it is currently not recommended as a diagnostic marker for screening women in general population [31]. The clinical utility of this test is more effective as a follow up marker in monitoring the treatment for ovarian cancer and for detecting disease recurrence [51–53]. Unfortunately till date there is no single specific biomarker available for early detection of ovarian cancer. It is notable that, recently our group has reported a panel of five plasma proteins (including CA-125) which has potential as a diagnostic assay for epithelial ovarian cancer [54]. Globally the market for CA-125 test is expected to grow at 5.80% annually for the next decade and is projected to hit USD 1091.7 million by 2030 according to a recent market survey.

In terms of the protein structure, CA-125 is the second largest protein in human proteome, made of ~14 000 residues. It is a transmembrane protein with a long extracellular region consisting of 60–70 tandem repeats of a Sperm protein, Enterokinase and Agrin (SEA) domain made of 196 residues [55]. Different isoforms of CA-125 exist, which vary primarily based on the number of repeats of the SEA domain. Several popular antibodies such as 5E11 [56], OC125 [57] and M11 have been designed to target the SEA domain of CA-125. Recently, the partial crystal structure of

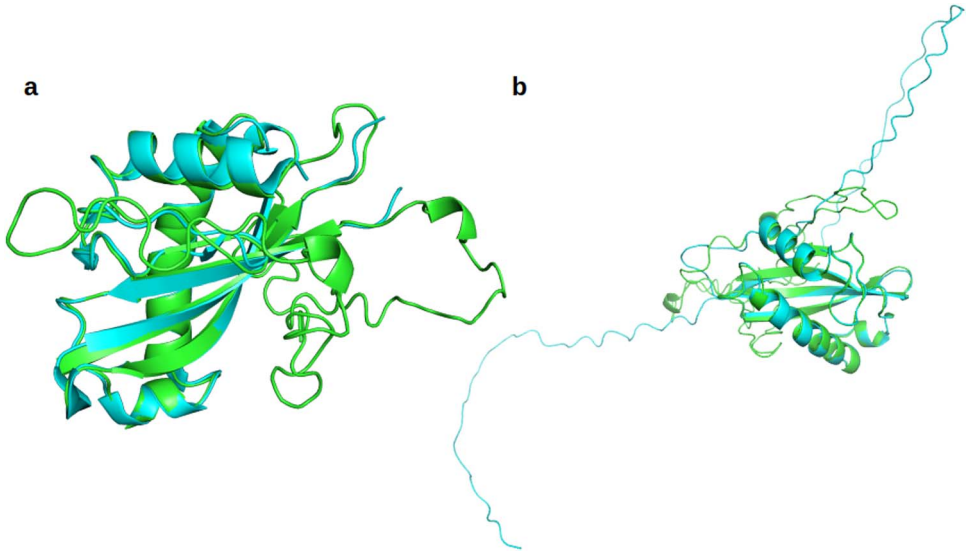

Figure 3. Structural overlap of the ITASSER model of human CA-125 SEA domain with (a) the partial experimental structure from PDB; (b) the model obtained using AlphaFold2 program.

Table 1. Results from the comparison of the five threading models of human CA-125 SEA domain from ITASSER with the experimental structure of partial human SEA domain (PDB ID: 7SA9). The RMSD values were calculated using superposition of the structures in PyMOL.

Model	C-Score[a]	RMSD (Å)
Model 1	−2.45	0.826
Model 2	−2.61	0.740
Model 3	−2.83	0.864
Model 4	−3.13	0.990
Model 5	−3.08	1.074

[a]C-score stands for the ITASSER confidence score used to quantify the model quality. The values of this score range from [−5, 2], with higher values indicating better confidence in the model.

Human SEA domain (residues 35 to 160) was solved to 1.69 Å resolution (PDB ID: 7SA9) [55] (Fig. 2). However, it is notable that none of the antigen–antibody complexes involving CA-125 have been experimentally solved earlier. Hence, with this partial structure of SEA domain from CA-125 as the basis for antibody design, the application of each module of rAbDesFlow is detailed in the sections below.

Module 1: Modeling the complete antigen structure

Since only a partial structure of the human CA-125 SEA domain was available experimentally, the ITASSER server was used to model the complete antigen structure. The complete SEA domains sequence of Mucin-16 protein was obtained from UniProt [32] (UniProt ID: Q8WXI7). The C-scores [33] obtained for the top five models from threading are tabulated below (Table 1). Based on the results, Model 2 provided the best trade-off between the C-Score and RMSD with respect to the crystal structure (Fig. 3a), and it was chosen as the final model for further steps of antibody design. The antigen structure was further confirmed using the ColabFold model [58] from AlphaFold2 team (Fig. 3b). It is notable that the long loops at the N- and C-terminal regions of the protein from the AlphaFold2 program was not favorable for further analysis with docking programs, and hence was discarded despite a good overall structural agreement with an RMSD of 0.742 Å between the two models obtained.

Table 2. Basic statistics of the antibody sequences designed for CA-125 antigen.

Antibody features	Data
Final set of mutations	D96G, D96P, D97G, D97P, D97R, D97T, D97Y, Y98G, D99G, M102F
Range of heavy chain H-score[a]	−1.098 to −1.133 (WT = −1.142)
Heavy chain length	118 residues
Light chain length	112 residues
Fab[a] + Fc[a] region length (1 monomer)	668 residues
Total antibody length	2 monomers = 668∗2 = 1336 residues (~150 kDa)

[a]H-score—Humanness Z-score quantifies the ability of animal-derived or synthetic antibodies to mimic endogenous human antibodies for better immune response; Fab—Fragment-antigen binding region of the antibody; Fc- Fragment crystallizable region of the antibody.

Module 2: Enumeration of CA-125 antibody sequences

CA-125 being an extensively studied biomarker in ovarian cancer has multiple humanized mAbs designed against it, including 5E11, OC125, M11, Sofituzumab, and Abagovomab. Among them the binding epitopes of 5E11, OC125, and M11 have been extensively characterized through experimental mutagenesis studies [59]. Therefore, the sequences of these three antibodies (specifically the CDRH3 region) were considered as templates for enumeration of novel antibody sequences with better binding affinity to CA-125. Using the ANARCI package, the CDRH3 region sequence of the template antibody was found to be SDDYDYGMDY. Some more basic statistics of the antibody sequence are provided in Table 2. The antibody sequences obtained were subject to TAP profiling and humanness score filtering. The top 10 antibodies in terms of their humanness z-score were considered for further analysis.

Module 3: Modeling the three-dimensional antibody structure

The antibody structure was modeled using two approaches: scFv modeling and complete antibody modeling. scFv stands for single chain variable fragment, which is similar to camelid nanobodies

Table 3. Rosetta scores of the scFv antibody structures obtained from the Rosetta ROSIE server.

Antibody	Mutation	Rosetta score[a]
WT	–	−597.699
ab1_26	D96G	−565.436
ab1_33	D96P	−608.073
ab1_46	D97G	−618.948
ab1_53	D97P	−595.807
ab1_55	D97R	−589.909
ab1_57	D97T	−634.401
ab1_60	D97Y	−572.697
ab1_66	Y98G	−612.439
ab1_86	D99G	−646.430
ab1_145	M102F	−597.086

[a]Rosetta score corresponds to the energy of the antibody model. It is obtained using a scoring function with energy terms and probabilities tailored to achieve the allowed geometrical parameters of a protein structure.

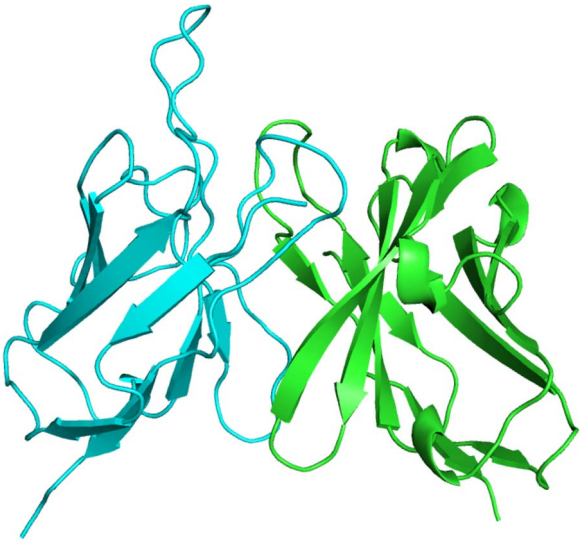

Figure 4. An example of the scFv antibody structure model against CA-125 antigen, obtained from the Rosetta ROSIE server with light chain in right and heavy chain in left. The scFv model obtained from the OPIG server is overlapped on the Rosetta model using PyMOL. The scFv models from ROSIE server for all 10 antibodies are provided in supplementary information (Fig. S4).

[44]. It includes only the Fab regions of the light and heavy chains of the antibody, which is sufficient for eliciting an immune response as proven by experiments. The antibody grafting mode of the Rosetta ROSIE server was used for *ab initio* modeling of the scFv antibody structure. Rosetta returns the top 1000 antibody models for a given sequence, by performing extensive CDRH3 loop modeling and optimization. The binding free energies of the top 1000 models were used to prioritize the final antibody scFv model for antigen–antibody docking (Table 3). Using this procedure, the antibody scFv models for the top 10 antibody sequences prioritized earlier were obtained, along with the structure for the template antibody considered (Supplementary Information—Fig. S5). The models obtained from Rosetta were also compared with the scFv models from the OPIG antibody modeling server (Fig. 4). It was observed that all models showed an RMSD <0.5 Å with the OPIG models, indicating a good agreement in the overall structure obtained. The complete antibody model, which contains both the Fc and Fab regions, was also obtained using ITASSER. However, since the overlap RMSD of scFv models with the Fab region of the complete antibody was less than 0.4 Å (Fig. S2), it was decided to use only the scFv models of the antibody for further analysis. This was also done to facilitate antigen–antibody docking, as all the available docking programs can consider only the scFv models of antibody as input, rather than the complete antibody structure. The workflow for designing complete antibody models is discussed in the Supplementary Information (Section S1 and Fig. S1).

Module 4: Antigen–antibody docking using ClusPro advanced docking mode

The antigen–antibody complex for the wild-type and top 10 mutants were obtained using ClusPro antibody docking mode. During the docking process, the antibody was used as the receptor and antigen was used as the ligand, as suggested by the ClusPro server. Thirty models were built for each complex and the final model was chosen based on an analysis of the orientation of the antigen among all antibodies, to ensure that the epitope region is nearly identical between them. This will also ensure that the results of the residue substitution will be comparable between the antibodies thereby, allowing the identification of the substitution with maximal contribution to the overall binding affinity of the complex. Based on the epitope analysis, four possible orientations of the antigen with respect to the antibody were identified. However, due to spatial overlap between the epitopes in the predicted antigen structure, two orientations where the epitopes were positioned without overlaps (Fig. S3) were only considered as valid. Further analysis was performed on both the orientations finalized, to understand which orientation provided a better binding affinity compared to the template antibody considered.

Module 5: Analysis of the antigen–antibody interaction

The predominant residue-residue interactions present at the interface of the antigen–antibody complexes (Fig. 5) obtained from ClusPro were extracted using the RING-3.0 stand-alone program. In terms of the interactions observed, hydrogen bonds and van der Waals interactions were found to be predominant among the complexes, followed by ionic, cation-pi and pi-pi stacking interactions. No disulfide bridges (covalent interactions) were observed between the antigen and antibody in the complex. The paratope and epitope residues for each antigen–antibody complex were identified based on the interacting residues and are provided in Table 4 for epitope 1 and Table S2 for epitope 2. To understand the influence of potential glycosylation of the residues part of the epitope, both sequence-based and structure-based glycosylation prediction methods available in literature were utilized. From the results it was observed that, no potential N- and O-linked glycosylation sites were present within the predicted epitope regions, indicating their potential to interact with the antibody without glycan interference (Supplementary Information Section S2). The frequency of various inter-residue interactions at the antigen–antibody interface as identified with the RING3.0 program were also studied in detail and are summarized in Supplementary Tables S4 and S5.

The 10 antigen–antibody complexes designed for CA-125 were further ranked on the basis of their predicted binding free energy and dissociation constant (K_d) value from the PRODIGY server (Table 5). A consensus binding free energy prediction was also performed based on multiple sequence and structure-based methods [50, 60–64], to establish a consistency in the antibody ranking process (Supplementary Table S7). Based on the results observed, it is conclusive that the designed antibodies provide ∼1 kcal/-mol improvement in binding affinity compared to the wild-type.

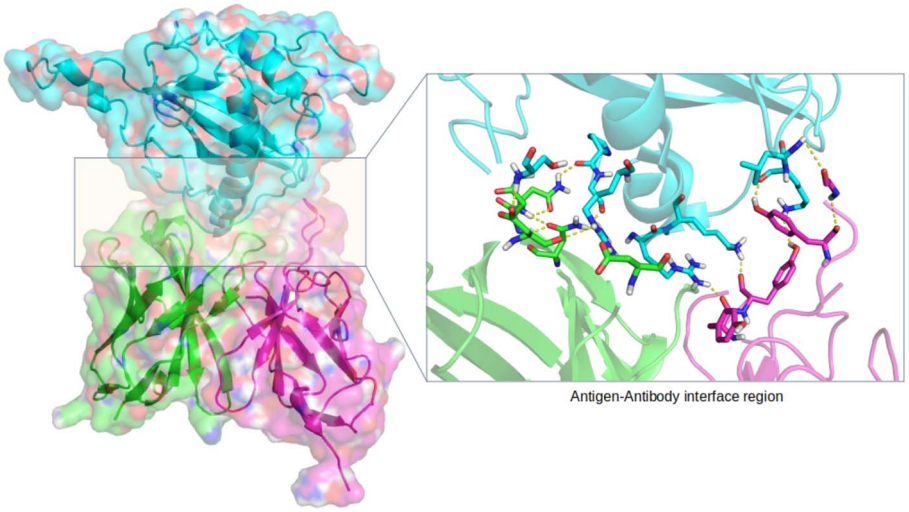

Figure 5. An example of the interactions observed at the human CA125 antigen–antibody interface region where the hydrogen bonds are highlighted in the inset with dotted lines. The antigen is shown on top; antibody light chain is shown in bottom left and heavy chain is shown in bottom right. The visualization was created using PyMOL.

Table 4. Paratope and epitope residues for each antigen–antibody ScFv region model with orientation 1. The residues were extracted based on the results from PPI analysis using the RING3.0 stand-alone program.

Antibody	Mutation	Antigen residues (Epitope)[a]	Antibody residues (Paratope)[a]
WT	–	ASP122A, THR176A, GLU129A, VAL121A, GLN132A, SER177A, GLY120A, THR179A, GLU124A, ASP173A, PRO167A, PHE40A, LEU114A, SER118A, HIS157A, LEU133A, PRO119A, TRP128A, ARG123A, GLN125A, LYS117A	ASN30L, TYR92L, ASP97H, PHE58H, SER30L, THR93L, TYR100H, TYR94L, ASN54H, ASP56H, TYR30L, TYR32L, LYS30L, ASN52H, TRP50L, LYS64H
ab1_26	D96G	PRO119A, GLU129A, THR176A, GLU124A, ASP122A, LEU133A, PRO167A, THR179A, LYS117A, ASP173A, TRP128A, SER177A, GLY120A, ARG123A, PHE196A, HIS157A, GLN125A	GLN27L, TYR33H, GLN61H, TYR32L, ASN54H, LYS64H, TYR30L, SER30L, PHE58H, ASP99H, THR93L, TYR94L, TYR98H, TYR92L, ASP97H
ab1_33	D96P	ASN80A, ASN153A, GLN125A, LYS117A, SER118A, ARG123A, GLU124A, GLU129A, ASP122A, THR176A, VAL121A, LEU133A, SER82A, TRP128A, GLN132A, PRO119A	TYR94L, TYR33H, PHE58H, THR93L, TYR92L, ASP97H, ASP56H, TYR98H, GLN27L, SER30L, LYS64H, ASP99H, TYR32L, TYR30L, LYS30L
ab1_46	D97G	VAL121A, SER177A, VAL172A, PRO119A, GLU124A, HIS157A, ARG123A, TRP128A, GLY120A, GLN132A, GLN125A, ASP173A, GLU129A, THR176A, PRO167A, ASP122A	ASP56H, THR93L, TYR32L, GLN27L, GLN61H, ASP99H, TYR30L, LYS30L, TYR94L, TYR98H, THR57H, SER30L, PHE58H, ASN30L, LYS64H, TYR33H, TYR92L
ab1_53	D97P	LEU114A, GLY120A, GLN132A, GLU124A, TRP128A, LYS117A, PHE40A, SER177A, GLU129A, ARG123A, VAL121A, THR179A, ASP173A, ASN153A, PRO119A, LEU133A, THR176A, ASP122A, PHE196A	TYR30L, TYR92L, ASN54H, ASN30L, SER30L, THR93L, TYR94L, TYR98H, PHE58H, TYR33H, TYR100H, ASP99H, TYR32L, TRP50L, GLN27L, ASP56H
ab1_55	D97R	GLU129A, THR179A, LYS117A, ASP173A, SER177A, LEU133A, SER82A, THR176A, GLN132A, GLN125A, PHE196A, PHE40A, LEU114A, SER118A, GLY120A, ASP122A, ARG123A, TRP128A, PRO119A	ASP56H, PHE58H, TYR33H, SER30L, TYR32L, TYR30L, TYR92L, ASP99H, GLN27L, TYR100H, ASN54H, LYS30L, TYR94L, TYR98H, ASN30L
ab1_57	D97T	ASN153A, GLN125A, GLU124A, SER177A, THR176A, THR179A, ARG123A, GLU129A, GLN132A, PRO119A, PHE196A, GLY120A, HIS157A, LYS117A, LEU133A, TRP128A, ASP173A	TYR98H, TYR30L, ASP99H, TYR94L, PHE58H, ASN54H, ASN52H, THR97H, TYR92L, THR93L, ASP56H, SER30L, LYS64H
ab1_60	D97Y	ASN80A, LYS79A, LYS117A, GLN125A, GLU124A, SER82A, LEU133A, THR176A, GLU129A, SER177A, ASP122A	TYR32H, TYR94L, ASN54H, SER30L, TYR30L, TYR92L, PHE58H, ASP56H, TYR33H, THR93L, ASP99H
ab1_66	Y98G	SER177A, ARG123A, SER118A, PRO119A, LEU133A, THR81A, ASP122A, THR176A, GLU124A, SER82A, ASN80A, GLN125A, GLU129A, LEU114A, LYS117A, TRP128A	TYR33H, ASN52H, ASN53H, PHE58H, TYR100H, TYR92L, THR93L, ASN30L, ASP97H, TYR32L, ASN54H, TYR30L
ab1_86	D99G	ASP173A, SER177A, LEU133A, PRO119A, LEU114A, THR179A, HIS157A, TRP128A, PHE40A, PRO167A, ASP122A, LYS117A, ARG123A, GLN132A, GLU124A, GLN125A, PHE196A, GLY120A, VAL121A, SER118A, ASN153A, THR176A, GLU129A	LYS64H, TYR100H, LYS30L, ASN54H, GLN61H, TYR92L, TYR98H, TYR30L, ASN30L, GLN27L, TYR32L, SER30L, TYR94L, TYR33H, PHE58H, THR93L, ASP56H
ab1_145	M102F	ASP122A, GLU129A, GLN132A, TRP128A, THR176A, ARG123A, SER177A, PHE40A, VAL121A, GLY120A, SER118A, PRO119A, ASP173A, LEU133A, LYS117A, GLN125A, PHE196A, LEU114A	ASP97H, TYR98H, SER30L, TYR92L, ASN54H, LYS30L, TYR30L, TYR33H, ASP56H, TYR94L, TYR100H, PHE58H, GLN27L, TYR32L, LEU30L, ASN30L

[a]The chain to which each residue belongs has the following nomenclature: A—antigen; L—antibody light chain; H—antibody heavy chain.

Table 5. Overall rank of the 10 proposed antibodies (orientation 1) with respect to the wild-type antibody for CA125. The ranks are assigned based on the interaction energy values predicted by PRODIGY program. The three energy values in bold indicate the top three antibodies predicted by rAbDesFlow against CA-125.

Antibody	Mutation in CDRH3 region	PRODIGY predicted K_d (M) at 25 °C	PRODIGY energy (kcal/mol)
WT	–	2.2E-09	−11.8
ab1_26	D96G	2.0E-09	**−11.9**
ab1_33	D96P	3.7E-09	−11.5
ab1_46	D97G	3.9E-10	**−12.8**
ab1_53	D97P	1.4E-08	−10.7
ab1_55	D97R	1.9E-08	−10.5
ab1_57	D97T	1.1E-08	−10.8
ab1_60	D97Y	5.8E-08	−9.9
ab1_66	Y98G	3.4E-08	−10.2
ab1_86	D99G	1.1E-09	**−12.2**
ab1_145	M102F	1.3E-08	−10.7

Further, replacement of aspartic acid residue (D) seems to be the most favorable substitution, with D97G being consistently ranked as the topmost substitution among the selected antibodies. Since the lower energy of orientation 2 indicates more stable complex being formed between the antigen and antibody, the consensus rank was used to prioritize complexes with higher rank (see Supplementary Information Table S3) in orientation 2. ab1_46, ab1_33 and ab1_145 are proposed to be the best antibodies from the analysis for further experimental validation. Based on consensus glycosylation prediction, no potential N- and O-linked glycosylation sites were found within the predicted epitope regions, indicating their potential to interact with the antibody without glycan interference (Supplementary Information Section S2 and Table S6). To investigate potential off-target binding effects of the top three antibodies, a set of secretory cancer-associated proteins were curated (Supplementary Table S8) and their interaction with the antibodies were predicted using ClusPro program. The results from this analysis indicate that, for all the tested antigens, the binding affinity values of the best antigen–antibody docking pose obtained from ClusPro is not as high as that of the Mucin-16 (Supplementary Table S9). This justifies that, the chances of off-target binding for the top three antibodies identified in this study are very low. The complete nucleotide and protein sequences of all 10 antibodies analyzed in this work are provided in the Supplementary Information (Section S3).

Conclusions

In this study, a computational workflow for template-based humanized recombinant antibody design named, rAbDesFlow, was created. The workflow is based on completely open-sourced databases and tools enabling researchers worldwide to implement the workflow flexibly upon their needs. The proposed method was validated through design of antibodies specific to the human ovarian cancer antigen, CA-125 or Mucin-16, using existing antibody sequences as templates for *in silico* site saturation mutagenesis. From an extensive analysis of various antibody profiling metrics and humanness scores, the top 10 antibodies were identified and used to model the antigen–antibody interaction through structure prediction and docking methods. A detailed analysis of the interfacial interactions and orientations of the antigen and antibody enabled identification of two predominant epitopes in the antigen. Finally, their predicted binding energy and dissociation constant were used to prioritize the top 3 antibodies for further experimental validation. rAbDesFlow will enable significant acceleration in prioritization of potential antibody sequences for experimental validation and hence, is the first of many steps toward fully automated computational antibody design.

Acknowledgements

We thank the Bioinformatics Infrastructure facility, Department of Biotechnology and Indian Institute of Technology Madras for computational facilities. N.M. acknowledges support from the Department of Biotechnology, Govt. of India for Ramalingaswami Re-Entry Fellowship 2020-2021.

Author contributions

Sowmya Ramaswamy Krishnan (Data Curation, Validation, Formal analysis, Software, Writing—Original Draft, Writing—Review & Editing, Visualization), Divya Sharma (Data Curation, Validation, Formal analysis, Software, Writing—Original Draft, Writing—Review & Editing),

Yasin Nazeer (Conceptualization, Methodology, Validation, Writing—Review & Editing, Supervision), Mayilvahanan Bose (Writing—Review & Editing), Thangarajan Rajkumar (Writing—Review & Editing), Guhan Jayaraman (Writing—Review & Editing), Narayanan Madaboosi (Conceptualization, Methodology, Validation, Resources, Writing—Review & Editing, Supervision), and M. Michael Gromiha (Conceptualization, Methodology, Validation, Investigation, Resources, Writing—Review & Editing, Supervision).

Supplementary data

Supplementary data is available at *ABT* online.

Conflict of interest statement

Prof. Thangarajan Rajkumar is affiliated to MedGenome and Mrs. Sowmya Ramaswamy Krishnan is employed at Tata Consultancy Services at the time of submission.

Funding

The authors acknowledge funding from the Science and Engineering Research Board (SERB), Government of India, for the project ID-PPP-OTN on Antibody Engineering (IPA/2021/000147).

Data availability

All data are incorporated into the article and its online supplementary material.

Ethics and consent statement

Not applicable.

Animal research statement

Not applicable.

References

1. Basu K, Green EM, Cheng Y. *et al.* Why recombinant antibodies - benefits and applications. *Curr Opin Biotechnol* 2020;**60**:153–8. https://doi.org/10.1016/j.copbio.2019.01.012.
2. Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975;**256**:495–7. https://doi.org/10.1038/256495a0.
3. Nelson AL, Dhimolea E, Reichert JM. Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* 2010;**9**:767–74. https://doi.org/10.1038/nrd3229.
4. Wang W, Xu R, Li J. Production of native bispecific antibodies in rabbits. *PloS One* 2010;**5**:e10879. https://doi.org/10.1371/journal.pone.0010879.
5. Reichert JM. Which are the antibodies to watch in 2013? *MAbs* 2013;**5**:1–4. https://doi.org/10.4161/mabs.22976.
6. Kuhn P, Fühner V, Unkauf T. *et al.* Recombinant antibodies for diagnostics and therapy against pathogens and toxins generated by phage display. *Proteomics Clin Appl* 2016;**10**:922–48. https://doi.org/10.1002/prca.201600002.
7. Edwards BM, He M. Evolution of antibodies in vitro by ribosome display. *Methods Mol Biol* 2012;**907**:281–92. https://doi.org/10.1007/978-1-61779-974-7_16.
8. Alfaleh MA, Alsaab HO, Mahmoud AB. *et al.* Phage display derived monoclonal antibodies: from bench to bedside. *Front Immunol* 2020;**11**:1986. https://doi.org/10.3389/fimmu.2020.01986.
9. Seigel DL. Recombinant monoclonal antibody technology. *Transfus Clin Biol* 2002;**9**:15–22. https://doi.org/10.1016/s1246-7820(01)00210-5.
10. Labrijn AF, Janmaat ML, Reichert JM. *et al.* Bispecific antibodies: a mechanistic review of the pipeline. *Nat Rev Drug Discov* 2019;**18**:585–608. https://doi.org/10.1038/s41573-019-0028-1.
11. Altunay B, Morgenroth A, Beheshti M. *et al.* HER2-directed antibodies, affibodies and nanobodies as drug-delivery vehicles in breast cancer with a specific focus on radioimmunotherapy and radioimmunoimaging. *Eur J Nucl Med Mol Imaging* 2021;**48**:1371–89. https://doi.org/10.1007/s00259-020-05094-1.
12. Saeed AF, Wang R, Ling S. *et al.* Antibody engineering for pursuing a healthier future. *Front Microbiol* 2017;**8**:495. https://doi.org/10.3389/fmicb.2017.00495.
13. Dienstmann R, Felip E. Necitumumab in the treatment of advanced non-small cell lung cancer: translation from preclinical to clinical development. *Expert Opin Biol Ther* 2011;**11**:1223–31. https://doi.org/10.1517/14712598.2011.595709.
14. Falcone M, Tiseo G, Valoriani B. *et al.* Efficacy of Bamlanivimab/Etesevimab and Casirivimab/Imdevimab in preventing progression to severe COVID-19 and role of variants of concern. *Infect Dis Ther* 2021;**10**:2479–88. https://doi.org/10.1007/s40121-021-00525-4.
15. Nichols RM, Deveau C, Upadhyaya H. Bebtelovimab: considerations for global access to treatments during a rapidly evolving pandemic. *Lancet Infect Dis* 2022;**22**:1531. https://doi.org/10.1016/S1473-3099(22)00592-8.
16. Thieblemont C, Phillips T, Ghesquieres H. *et al.* Epcoritamab, a novel, subcutaneous CD3xCD20 bispecific T-cell-engaging antibody, in relapsed or refractory large B-cell lymphoma: dose expansion in a phase I/II trial. *J Clin Oncol* 2023;**41**:2238–47. https://doi.org/10.1200/JCO.22.01725.
17. Klassen SA, Senefeld JW, Johnson PW. *et al.* The effect of convalescent plasma therapy on mortality among patients with COVID-19: systematic review and meta-analysis. *Mayo Clin Proc* 2021;**96**:1262–75. https://doi.org/10.1016/j.mayocp.2021.02.008.
18. Fu Z, Li S, Han S. *et al.* Antibody drug conjugate: the "biological missile" for targeted cancer therapy. *Signal Transduct Target Ther* 2022;**7**:93. https://doi.org/10.1038/s41392-022-00947-7.
19. Ma J, Mo Y, Tang M. *et al.* Bispecific antibodies: from research to clinical application. *Front Immunol* 2021;**12**:626616. https://doi.org/10.3389/fimmu.2021.626616.
20. Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J* 2021;**11**:69. https://doi.org/10.1038/s41408-021-00459-7.
21. Adolf-Bryfogle J, Kalyuzhniy O, Kubitz M. *et al.* RosettaAntibodyDesign (RAbD): a general framework for computational antibody design. *PLoS Comput Biol* 2018;**14**:e1006112. https://doi.org/10.1371/journal.pcbi.1006112.
22. Fleri W, Paul S, Dhanda SK. *et al.* The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* 2017;**8**:278. https://doi.org/10.3389/fimmu.2017.00278.
23. Jespersen MC, Peters B, Nielsen M. *et al.* BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017;**45**:W24–9. https://doi.org/10.1093/nar/gkx346.
24. Pierce BG, Wiehe K, Hwang H. *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 2014;**30**:1771–3. https://doi.org/10.1093/bioinformatics/btu097.
25. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;**125**:1731–7. https://doi.org/10.1021/ja026939x.
26. Vlachakis D, Feidakis C, Megalooikonomou V. *et al.* IMGT/collier-de-Perles: a two-dimensional visualization tool for amino acid domain sequences. *Theor Biol Med Model* 2013;**10**:14. https://doi.org/10.1186/1742-4682-10-14.
27. Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* 2016;**32**:298–300. https://doi.org/10.1093/bioinformatics/btv552.
28. Buß O, Rudat J, Ochsenreither K. FoldX as protein engineering tool: better than random based approaches? *Comput Struct Biotechnol J* 2018;**16**:25–33. https://doi.org/10.1016/j.csbj.2018.01.002.
29. Yuan S, Chan HCS, Hu Z. Using PyMOL as a platform for computational drug design. *Wiley Interdiscip Rev Comput Mol Sci* 2017;**7**:e1298. https://doi.org/10.1002/wcms.1298.
30. Baran D, Pszolla MG, Lapidoth GD. *et al.* Principles for computational design of binding antibodies. *Proc Natl Acad Sci U S A* 2017;**114**:10900–5. https://doi.org/10.1073/pnas.1707171114.
31. Charkhchi P, Cybulski C, Gronwald J. *et al.* CA125 and ovarian cancer: a comprehensive review. *Cancers (Basel)* 2020;**12**:3730. https://doi.org/10.3390/cancers12123730.
32. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2021;**49**:D480–9. https://doi.org/10.1093/nar/gkaa1100.

33. Yang J, Yan R, Roy A. *et al.* The I-TASSER suite: protein structure and function prediction. *Nat Methods* 2015;**12**:7–8. https://doi.org/10.1038/nmeth.3213.
34. Jumper J, Evans R, Pritzel A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2.
35. Baek M, DiMaio F, Anishchenko I. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;**373**:871–6. https://doi.org/10.1126/science.abj8754.
36. Stevens AO, He Y. Benchmarking the accuracy of AlphaFold 2 in loop structure prediction. *Biomolecules* 2022;**12**:985. https://doi.org/10.3390/biom12070985.
37. Akdel M, Pires DEV, Pardo EP. *et al.* A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* 2022;**29**:1056–67. https://doi.org/10.1038/s41594-022-00849-w.
38. Raybould MIJ, Marks C, Lewis AP. *et al.* Thera-SAbDab: the therapeutic structural antibody database. *Nucleic Acids Res* 2020;**48**:D383–8. https://doi.org/10.1093/nar/gkz827.
39. Kabat EA, Wu TT, Bilofsky H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J Biol Chem* 1977;**252**:6609–16.
40. Dunbar J, Krawczyk K, Leem J. *et al.* SAbPred: a structure-based antibody prediction server. *Nucleic Acids Res* 2016;**44**:W474–8. https://doi.org/10.1093/nar/gkw361.
41. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 2000;**13**:37–45. https://doi.org/10.1016/s1074-7613(00)00006-6.
42. Raybould MIJ, Marks C, Krawczyk K. *et al.* Five computational developability guidelines for therapeutic antibody profiling. *Proc Natl Acad Sci U S A* 2019;**116**:4025–30. https://doi.org/10.1073/pnas.1810576116.
43. Thullier P, Huish O, Pelat T. *et al.* The humanness of macaque antibody sequences. *J Mol Biol* 2010;**396**:1439–50. https://doi.org/10.1016/j.jmb.2009.12.041.
44. Ahmad ZA, Yeap SK, Ali AM. *et al.* scFv antibody: principles and clinical application. *Clin Dev Immunol* 2012;**2012**:980250. https://doi.org/10.1155/2012/980250.
45. Moretti R, Lyskov S, Das R. *et al.* Web-accessible molecular modeling with Rosetta: the Rosetta online server that includes everyone (ROSIE). *Protein Sci* 2018;**27**:259–68. https://doi.org/10.1002/pro.3313.
46. Evans R, O'Neill M, Pritzel A. *et al.* Protein complex prediction with AlphaFold-multimer. *bioRxiv* 2021. https://doi.org/10.1101/2021.10.04.463034 October 04, 2021, preprint: not peer reviewed.
47. Ruffolo JA, Chu L, Mahajan SP. *et al.* Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nat Commun* 2023;**14**:2389. https://doi.org/10.1038/s41467-023-38063-x.
48. Brenke R, Hall DR, Chuang G. *et al.* Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics* 2012;**28**:2608–14. https://doi.org/10.1093/bioinformatics/bts493.
49. Clementel D, Conte AD, Monzon AM. *et al.* RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Res* 2022;**50**:W651–6. https://doi.org/10.1093/nar/gkac365.
50. Xue LC, Rodrigues JP, Kastritis PL. *et al.* PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* 2016;**32**:3676–8. https://doi.org/10.1093/bioinformatics/btw514.
51. Pignata S, Cannella L, Leopardo D. *et al.* Follow-up with CA125 after primary therapy of advanced ovarian cancer: in favor of continuing to prescribe CA125 during follow-up. *Ann Oncol* 2011;**22 Suppl 8**:viii40-viii44. https://doi.org/10.1093/annonc/mdr470.
52. Li Z, Yin H, Ren M. *et al.* Prognostic significance of CA125 dynamic change for progression free survival in patients with epithelial ovarian carcinoma. *Med Sci Monit* 2020;**26**:e925051. https://doi.org/10.12659/MSM.925051.
53. Piatek S, Panek G, Lewandowski Z. *et al.* Rising serum CA-125 levels within the normal range is strongly associated recurrence risk and survival of ovarian cancer. *J Ovarian Res* 2020;**13**:102. https://doi.org/10.1186/s13048-020-00681-0.
54. Periyasamy A, Gopisetty G, Subramanium MJ. *et al.* Identification and validation of differential plasma proteins levels in epithelial ovarian cancer. *J Proteomics* 2020;**226**:103893. https://doi.org/10.1016/j.jprot.2020.103893.
55. White B, Patterson M, Karnwal S. *et al.* Crystal structure of a human MUC16 SEA domain reveals insight into the nature of the CA125 tumor marker. *Proteins* 2022;**90**:1210–8. https://doi.org/10.1002/prot.26303.
56. Marcos-Silva L, Narimatsu Y, Halim A. *et al.* Characterization of binding epitopes of CA125 monoclonal antibodies. *J Proteome Res* 2014;**13**:3349–59. https://doi.org/10.1021/pr500215g.
57. Skov RV, Bichel P. OC-125 as a diagnostic aid in the cytological evaluation of ascitic cells in patients with ovarian carcinoma. *Virchows Arch A Pathol Anat Histopathol* 1991;**419**:59–62. https://doi.org/10.1007/BF01600153.
58. Mirdita M, Schütze K, Moriwaki Y. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* 2022;**19**:679–82. https://doi.org/10.1038/s41592-022-01488-1.
59. Marcos-Silva L, Ricardo S, Chen K. *et al.* A novel monoclonal antibody to a defined peptide epitope in MUC16. *Glycobiology* 2015;**25**:1172–82. https://doi.org/10.1093/glycob/cwv056.
60. Myung Y, Pires DEV, Ascher DB. CSM-AB: graph-based antibody-antigen binding affinity prediction and docking scoring function. *Bioinformatics* 2022;**38**:1141–3. https://doi.org/10.1093/bioinformatics/btab762.
61. Nikam R, Yugandhar K, Gromiha MM. Deep learning-based method for predicting and classifying the binding affinity of protein-protein complexes. *Biochim Biophys Acta Proteins Proteom* 2023;**1871**:140948. https://doi.org/10.1016/j.bbapap.2023.140948.
62. Romero-Molina S, Ruiz-Blanco YB, Mieres-Perez J. *et al.* PPI-affinity: a web tool for the prediction and optimization of protein-peptide and protein-protein binding affinity. *J Proteome Res* 2022;**21**:1829–41. https://doi.org/10.1021/acs.jproteome.2c00020.
63. Yang YX, Huang JY, Wang P. *et al.* AREA-AFFINITY: a web server for machine learning-based prediction of protein-protein and antibody-protein antigen binding affinities. *J Chem Inf Model* 2023;**63**:3230–7. https://doi.org/10.1021/acs.jcim.2c01499.
64. Yuan Y, Chen Q, Mao J. *et al.* DG-affinity: predicting antigen-antibody affinity with language models from sequences. *BMC Bioinformatics* 2023;**24**:430. https://doi.org/10.1186/s12859-023-05562-z.