Article

Extracting the Interfacial Electrostatic Features from Experimentally Determined Antigen and/or Antibody-Related Structures inside Protein Data Bank for Machine Learning-Based Antibody Design

Wei Li *

[1] Institute of Special Environment Medicine, Nantong University, No. 9, Seyuan Road, Nantong City, Jiangsu Province, P. R. China

* Correspondence: wli148@aucklanduni.ac.nz

Received: date; Accepted: date; Published: date

Abstract: The importance of antibodies in health care and the biotechnology research and development demands not only knowledge of their experimental structures at high resolution, but also practical implementation of this knowledge for both effective and efficient design and production of antibody for its use in both medical and research applications. While the experimental wet-lab approach is usually costly, laborious and time-consuming, computational (dry-lab) approaches, in spite of their intrinsic limitations in comparison with its experimental (wet-lab) counterpart, provide a cheaper and faster alternative option. For the first time, this article reports a comprehensive set of structural electrostatic features extracted from experimentally determined antigen-antibody-related structures, including especially those structural electrostatic features at the interfaces of all experimentally determined antigen-antibody complex structures as of February 29, 2020, to facilitate effective and efficient machine learning-based computational antibody design using currently available experimental structures inside Protein Data Bank.

Keywords: Antigen-antibody complex structure; Interfacial electrostatic feature; Machine Learning-Based Antibody Design; Protein Data Bank

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 March 2020 doi:10.20944/preprints202003.0011.v1

1. Introduction

Usually known as an immunoglobulin (Ig) [1–3], an antibody (Ab) is a large, Y-shaped protein that is used by the immune system to neutralize pathogens such as pathogenic bacteria and viruses [4–8]. The antibody recognizes a unique molecule of the pathogen, called an antigen(Ag), via the fragment antigen-binding (Fab) variable region [9]. As is widely known, antibodies are important both in health care [10] and biotechnology industry, demanding not only high resolution experimental structural determination of antibodies and also effective and efficient implementation of this hard-earned structural knowledge inside Protein Data Bank (PDB, [11]), either experimentally [12–14] or computationally [15–17], or both approaches in hybrid [18,19]. However, the process of obtaining therapeutic antibodies remains time consuming and empirical [3,10,20–22].

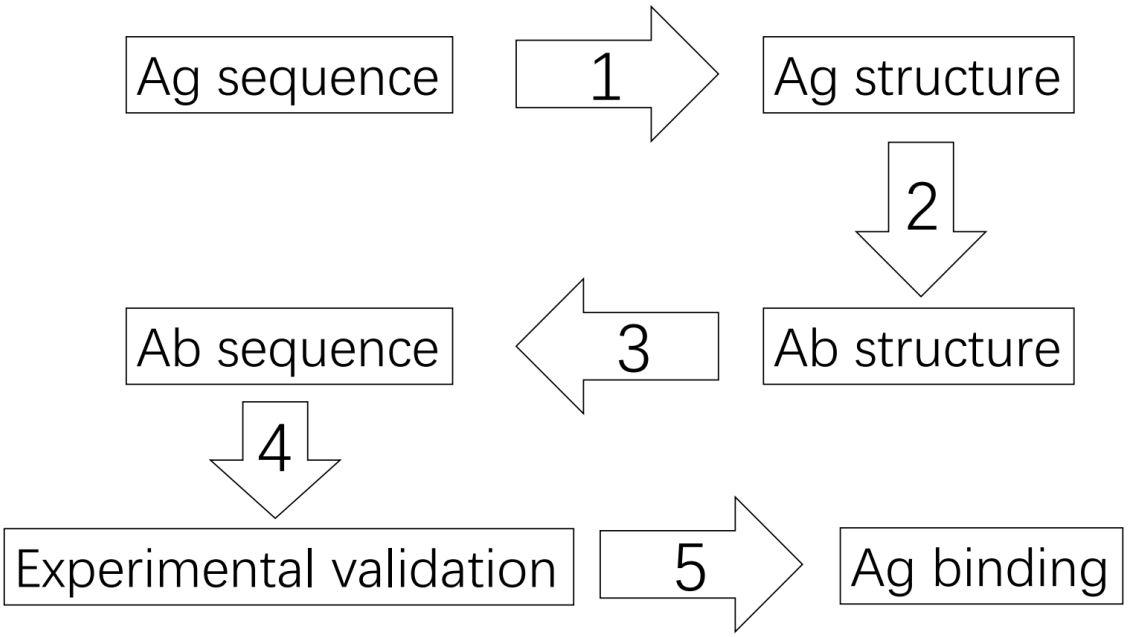

Figure 1. Flowchart of the machine learning-based computational antibody design [23]. In this figure, **Ag** and **Ab** represent antigen and antibody, respectively. **Step 1** represents the process of protein sequence-based protein structure prediction, **Step 2** represents the process of Ag structure-base Ab structure prediction, **Step 3** represents the reverse process of protein sequence-based protein structure prediction, **Step 4** represents the experimental synthesis (either biological or chemical) of the computationally predicted Ab sequence for subsequent experimental validation of its binding activity to Ag (**Step 5**), i.e., the starting point of this flow chart.

To date, as the ability to describe the antibody through binding affinity to the antigen is supplemented by information on antibody structure and amino acid sequences, several computational tools have been developed for the design of antibodies [17,24–27]. For example, Rosetta Antibody is a novel antibody FV (the variable fragment critical for Ag binding) region structure prediction server [24]. To this point, further computational methods with higher accuracy and efficiency should be developed and used in drug development based on experimentally determined structures or homology models, including antigen-antibody dockings [15,28,29] and molecular energy calculations [30–32] with approximate potential functions, providing a potential guide to experimental studies to improve the affinities and physicochemical properties of antibodies [20]. Thus, this article goes through currently available (as of February 29, 2020) antigen-antibody-related structures inside Protein Data Bank [11], aiming at digging all electrostatic features embedded inside out, to facilitate machine learning-based computational antibody design [23] (Figure 1) with reasonable accuracy and efficiency compared with its counterparts currently available [24,25].

 doi:10.20944/preprints202003.0011.v1

2. Materials and Methods

As of February 29, 2020, with a Text Search for: antigen antibody, it turns out that the PDB [33] hosts a total of 888 experimentally determined antigen-antibody-related structures, with their PDB IDs included in the supplementary file **template.pdf**. Among the 888 antigen-antibody-related structures (the titles are included in the supplementary file **template.pdf**), 823 were determined using X-ray diffraction, 50 by electron microscopy, 14 by solution NMR spectrospecopy and 1 by solution scattering, respectively [33].

After the 888 structures (i.e., PDB files with a total size of 753 megabytes as of February 29, 2020) were accessed and downloaded directly from the PDB website [33], a comprehensive set of electrostatic analysis was carried out as described in [34] previously, including both salt bridging and hydrogen bonding analysis for all 888 structures. Specifically, for the 14 NMR structures (with PDB IDs 1CS9, 1CT6, 1CVQ, 1CW8, 1CWZ, 1F3R, 1R21, 1S4H, 1S4J, 1TOR, 1TOS, 2MKL, 2MTW, 2RLL), an in-house python script was used to split the NMR ensemble into single NMR structural models for subsequent electrostatic analysis as described previously in [34].

3. Result

With the electrostatic analysis as described previously in [34], this article puts forward a comprehensive set of electrostatic interaction features for all antigen-antibody-related structures as of February 29, 2020, all included as a large set of separate tables in the supplementary file **supplementary.pdf**, including specifically:

1. all salt bridges formed within all antigen-antibody-related structures as of February 29, 2020.
2. all main chain and side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020.
3. all side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020.
4. all interfacial salt bridges formed within all antigen-antibody-related structures as of February 29, 2020.
5. all interfacial main chain and side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020.
6. all interfacial side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020.
7. all salt bridges formed within all antigen-antibody-related structures as of February 29, 2020, PDB ID-specifically for all 888 antigen-antibody-related structures.
8. all main chain and side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020, PDB ID-specifically for all 888 antigen-antibody-related structures.
9. all side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020, PDB ID-specifically for all 888 antigen-antibody-related structures.
10. all interfacial salt bridges formed within all antigen-antibody-related structures as of February 29, 2020, PDB ID-specifically for all 888 antigen-antibody-related structures.
11. all interfacial main chain and side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020, PDB ID-specifically for all 888 antigen-antibody-related structures.
12. all interfacial side chain hydrogen bonds formed within all antigen-antibody-related structures as of February 29, 2020, PDB ID-specifically for all 888 antigen-antibody-related structures.

4. Conclusion and Discussion

For the first time, this article reports a comprehensive set of electrostatic features sucked out of the currently (as of February 29, 2020) available 888 antigen-antibody-related structure inside PDB [11] in both PDF format (supplementary file **supplementary.pdf**) and also LaTeX format (a series of

machine-readable -importable and -analyzable .tex files zipped in the supplementary file **scan.zip**). Combined with the features structurally extracted with other molecular energy terms (hydrophobic interaction for instance) [30–32], the electrostatic features reported here are expected to be potentially useful to facilitate the second step of the machine learning-based antibody design [35], as illustrated in Figure 1.

Author Contributions: Conceptualization, W.L.; methodology, W.L.; software, W.L.; validation, W.L.; formal analysis, W.L.; investigation, W.L.; resources, W.L.; data duration, W.L.; writing–original draft preparation, W.L.; writing–review and editing, W.L.; visualization, W.L.; supervision, W.L.; project administration, W.L.; funding acquisition, not applicable.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

1. Choong, Y.S.; Lee, Y.V.; Soong, J.X.; Law, C.T.; Lim, Y.Y. Computer-Aided Antibody Design: An Overview. In *Recombinant Antibodies for Infectious Diseases*; Springer International Publishing, 2017; pp. 221–243.
2. Yamashita, T. Toward rational antibody design: recent advancements in molecular dynamics simulations. *International Immunology* **2018**, *30*, 133–140.
3. Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nature Reviews Drug Discovery* **2017**, *16*, 315–337.
4. Li, W. Structural Identification of the Electrostatic Hot Spots for Severe Acute Respiratory Syndrome Coronavirus Spike Protein to Be Complexed with Its Receptor ACE2 and Its Neutralizing Antibodies **2020**.
5. Chan, J.F.W.; Kok, K.H.; Zhu, Z.; Chu, H.; To, K.K.W.; Yuan, S.; Yuen, K.Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* **2020**, *9*, 221–236.
6. Chan, J.F.W.; Yuan, S.; Kok, K.H.; To, K.K.W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.Y.; Poon, R.W.S.; Tsoi, H.W.; Lo, S.K.F.; Chan, K.H.; Poon, V.K.M.; Chan, W.M.; Ip, J.D.; Cai, J.P.; Cheng, V.C.C.; Chen, H.; Hui, C.K.M.; Yuen, K.Y. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* **2020**, *395*, 514–523.
7. Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; Xia, J.; Yu, T.; Zhang, X.; Zhang, L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* **2020**, *395*, 507–513.
8. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; Yuan, M.L.; Zhang, Y.L.; Dai, F.H.; Liu, Y.; Wang, Q.M.; Zheng, J.J.; Xu, L.; Holmes, E.C.; Zhang, Y.Z. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**.
9. Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Molecular Biology and Evolution* **1993**.
10. Elgundi, Z.; Reslan, M.; Cruz, E.; Sifniotis, V.; Kayser, V. The state-of-play and future of antibody therapeutics. *Advanced Drug Delivery Reviews* **2017**, *122*, 2–19.
11. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980.
12. Wagner, J.; Schaffer, M.; Fernández-Busnadiego, R. Cryo-electron tomography-the cell biology that came in from the cold. *FEBS Letters* **2017**, *591*, 2520–2533.
13. Selenko, P. Quo Vadis Biomolecular NMR Spectroscopy? *International Journal of Molecular Sciences* **2019**, *20*, 1278.
14. López, C.A.; Swift, M.F.; Xu, X.P.; Hanein, D.; Volkmann, N.; Gnanakaran, S. Biophysical Characterization of a Nanodisc with and without BAX: An Integrative Study Using Molecular Dynamics Simulations and Cryo-EM. *Structure* **2019**.
15. Li, W.; Shi, G. How $Ca_V1.2$-bound verapamil blocks Ca^{2+} influx into cardiomyocyte: Atomic level views. *Pharmacological Research* **2019**, *139*, 153–157.
16. Carter, P.J. Potent antibody therapeutics by design. *Nature Reviews Immunology* **2006**, *6*, 343–357.
17. Leem, J.; Deane, C.M. High-Throughput Antibody Structure Modeling and Design Using ABodyBuilder. In *Methods in Molecular Biology*; Springer New York, 2018; pp. 367–380.

18. Steven, A.C.; Baumeister, W. The future is hybrid. *Journal of Structural Biology* **2008**, *163*, 186–195.
19. Akhouri, R.R.; Öfverstedt, L.G.; Wilken, G.; Skoglund, U. Antibody Complexes. In *Subcellular Biochemistry*; Springer International Publishing, 2019; pp. 23–51.
20. Kuroda, D.; Shirai, H.; Jacobson, M.P.; Nakamura, H. Computer-aided antibody design. *Protein Engineering Design and Selection* **2012**, *25*, 507–522.
21. Tsuchikama, K.; An, Z. Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein & Cell* **2016**, *9*, 33–46.
22. Birrer, M.J.; Moore, K.N.; Betella, I.; Bates, R.C. Antibody-Drug Conjugate-Based Therapeutics: State of the Science. *JNCI: Journal of the National Cancer Institute* **2019**, *111*, 538–549.
23. Wang, H.W.; Pai, T.W. Machine Learning-Based Methods for Prediction of Linear B-Cell Epitopes. In *Methods in Molecular Biology*; Springer New York, 2014; pp. 217–236.
24. Adolf-Bryfogle, J.; Kalyuzhniy, O.; Kubitz, M.; Weitzner, B.D.; Hu, X.; Adachi, Y.; Schief, W.R.; Dunbrack, R.L. RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLOS Computational Biology* **2018**, *14*, e1006112.
25. Li, T.; Pantazes, R.J.; Maranas, C.D. OptMAVEn – A New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes. *PLoS ONE* **2014**, *9*, e105954.
26. Nimrod, G.; Fischman, S.; Austin, M.; Herman, A.; Keyes, F.; Leiderman, O.; Hargreaves, D.; Strajbl, M.; Breed, J.; Klompus, S.; Minton, K.; Spooner, J.; Buchanan, A.; Vaughan, T.J.; Ofran, Y. Computational Design of Epitope-Specific Functional Antibodies. *Cell Reports* **2018**, *25*, 2121–2131.e5.
27. Fischman, S.; Ofran, Y. Computational design of antibodies. *Current Opinion in Structural Biology* **2018**, *51*, 156–162.
28. Duhovny, D.; Nussinov, R.; Wolfson, H.J. Efficient Unbound Docking of Rigid Molecules. In *Lecture Notes in Computer Science*; Springer Berlin Heidelberg, 2002; pp. 185–200.
29. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research* **2005**, *33*, W363–W367.
30. Orioli, S.; Ianeselli, A.; Spagnolli, G.; Faccioli, P. All-atom calculation of protein free-energy profiles. *The Journal of Chemical Physics* **2017**, *147*, 152724.
31. Song, Y.; Tyka, M.; Leaver-Fay, A.; Thompson, J.; Baker, D. Structure-guided forcefield optimization. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 1898–1909.
32. Bejagam, K.K.; Singh, S.; An, Y.; Deshmukh, S.A. Machine-Learned Coarse-Grained Models. *The Journal of Physical Chemistry Letters* **2018**, *9*, 4667–4672.
33. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980.
34. Li, W. How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? *PLOS ONE* **2017**, *12*, e0178519.
35. Frigerio, M.; Kyle, A.F. The Chemical Design and Synthesis of Linkers Used in Antibody Drug Conjugates. *Current Topics in Medicinal Chemistry* **2018**, *17*, 3393–3424.