Journal Title Here, 2022, pp. 1–2

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Application Note

APPLICATION NOTE

Developability Assessment for Monoclonal Antibodies Through Motif-based Sequence Search

William Tran[1,*]

[1]Department of Chemical Engineering, Columbia University, New York, 10027, NY

Abstract

Summary: A meta-analysis on the relationship of sequence motifs to antibody degradation was conducted, which resulted in a panel of 63 sequence motifs that were categorized into their respective degradation risk, and subsequently searched for in a database of 742 commercialized monoclonal antibody sequences. The resulting distribution of motifs can then serve as a baseline against which new therapeutics can be compared against to understand the developability risk of candidates relative to successful antibodies.
Availability and Implementation: Source code is freely available at . All programming was done in Python. Demo data was extracted from open source databases and formatted reference copies are included at the prior link.
Contact: wt2293@columbia.edu

Introduction

Therapeutic monoclonal antibodies (Mabs) are one of the fastest growing sectors of the pharmaceutical market, particularly in the field of oncology, even with 99 percent of preclinical drug candidates failing to make it through the regulatory process. The high risk associated with discovering new therapeutics thus discourages investment from big pharmaceutical companies, and instead, small companies backed by veture capital increasingly take on discovery and proof of concepts for new drug candidates. Consequently, this race-to-acquisition business model tends to emphasize speed over manufacturability for early drug discovery efforts, which in turn causes problems further down the line.

Due to a lengthy development process and high cost of clinical trials, chemistry manufacturing and controls (CMC) experience is needed to effectively screen for drug candidates more likely to succeed. Reasons for the high failure rate of drugs can include insufficient efficacy, suboptimal bioavailability safety and toxicology concerns, or stability and quality issues with the drug product(1). Currently, many early drug candidates are screened primarily for affinity and functionality, and less attention is cast on other important biophysical properties such as aggregation propensity, chemical stability, and possible degradation. The idea of developability is a comprehensive understanding of these properties that make a suitable drug candidate and evaluate its feasibility to progress through the development process. Developability knowledge must be made accessible so that new business entities in the drug discovery space can create therapeutic candidates with the highest probability of success.[1]

A number of in silico predictive tools are being developed as indications of developability, including aggregation prediction by SAP score and the Developability Index (DI), which has been commercialized in software by Biovia. However, due to the complexity of predicting protein function in the human body, a mechanistic understanding of the protein is needed, and many of these methods involve 3D structure modelling and significant computational cost, which are difficult to apply in a high throughput manner as needed for early discovery. Better mechanistic understanding of drugs can inform methods that reduce computational burdern, but much of the existing data of biophysical assays on sucessful therapeutics remain proprietary, or are stored in a way that is not accessible for computational modeling, which makes transferring the expertise required for developability assessment challenging.

To that end, this work leverages recently published knowledge around protein motifs in the antibody heavy and light chains which can increase post translational modifications that can affect developability risk. It is proposed that with thousands of molecules failing to make the clinic for every successful drug, the sequence dataset of commercialized antibodies has a significant survivor bias compared to the natural diversity of antibodies. Thus, characterizing the presence of such motifs in the sequences of successful antibodies will provide insight into an acceptable amount of risk for developability. To our knowledge, there has yet to be creation of an open-source tool that utilizes these novel motifs to inform naive drug development.

[1] *To whom correspondence should be addressed.

Implementation

We provide an implementation in Python that takes open-source data and generates distributions of protein sequence motifs implicated in developability risk. This tool runs on standard Python distributions and does not depend on the installation of other sequence analysis packages. Users can input a protein sequence for the heavy chain or light chain for a novel antibody therapeutic and check for the number of liabilities for antibody degradation through post-translational modifications. The modifications were assigned categories of either: oxidation, isomerization, glycosylation, deamidation, disulfide bonding, or aggregation based on the sequence motif and exposed amino acid residues. Liability flags occur when the input sequence has a number of motifs implicated in antibody degradation risk that are outside of the range of 5 to 95 percentile based on the distribution of existing commercialized Mabs. An example query is included in the source code, and the sequence can be replaced by one of interest.

Comparing to existing methods

This tool was compared against TAP, the Therapeutic Antibody Profiler, one of the tools of SAbPred from the Oxford Protein Informatics Group (OPIG)(2). OPIG models the 3D structure of the antibody and uses 5 aggregated metrics, which include: total CDR length, patches of surface hydrophobicity, patches of positive charge, patches of negative charge, and structural symmetry of charge in the variable fragment. Compared to the proposed tool, TAP focuses heavily on attractive and repulsive forces, which is important for aggregation propensity, but does not inform users of antibody stability and degradation as the motif-based approach does. Nevertheless, TAP is one of the few open access tools that are still maintained and attempt to inform developability risk. Of note, OPIG also utilizes sequence liability information provided by their tool ABodyBuilder, but do not benchmark against successful commercial antibodies, nor do they include motif understanding more recent than 2015. Thus, while the methodology would be based on the same sequence understanding, it would not serve as a comparable benchmark.

To assess the proposed tool, the commercial antibody dataset was randomly split 80/20 percent to serve as a model building set and a verification set. 5 randomly chosen antibodies were run through TAP and the motif-search, and all five yielded no flags in either tool. To look for a negative control, randomly chosen antibodies were selected from a database of naturally occuring non-therapeutic antibodies. Unfortunately, all the sequences tested did not trigger a red flag in the TAP, and manual modifications to induce poor developability traits either stayed within the range of experience from commercial Mabs or returned an error from the tool. More understanding is needed to inform the confidence of prediction and the conclusions drawn.

Validity of assumptions

To further understanding of this tool's validity, we sought to test the hypothesis that survivor bias in the commercialized antibody dataset was significant by comparing to a natural diversity antibody dataset, consisting of almost 90,000 protein sequences. Distributions of motifs were calculated and it was found that diifferences in motifs were not significantly different between the two datasets by chi-squared and the Mann-Whitney U tests, as shown in figure 1 below. Such a result suggests that motifs alone do not have sufficient resolution to differentiate qualities of successful Mabs from random natuirally occuring ones.

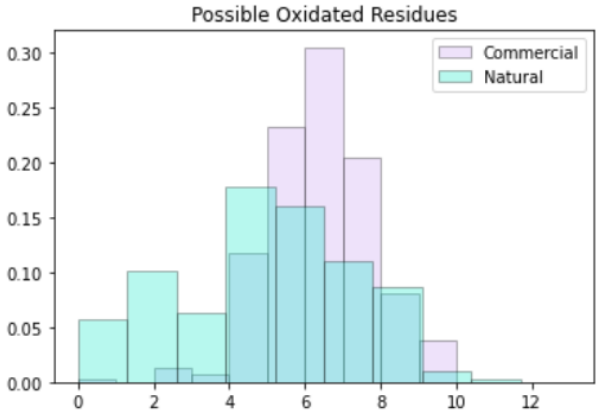

Fig. 1. Comparing the distribution of motifs in commercialized Mabs versus naturally occuring antibodies.

Acknowledgments

The authors thank Professor Itsik Pe'er of Columbia University, the CBMF4761 class, and its TA's for the valuable feedback.

References

1. Bailly M, Mieczkowski C, Juan V, et al. Predicting Antibody Developability Profiles Through Early Stage Discovery Screening. *MAbs.* , 2020;12(1):1743053. doi:10.1080/19420862.2020.1743053.
2. TAP: Raybould, M. I. J. et al. Five computational developability guidelines for therapeutic antibody profiling *PNAS*, 116, 4025-4030, 2019.
3. Xiaojun Lu, R. Paul Nobrega, Heather Lynaugh, Tushar Jain, Kyle Barlow, Todd Boland, Arvind Sivasubramanian, Maximiliano Vásquez, and Yingda Xu Deamidation and isomerization liability analysis of 131 clinical-stage antibodies. *mAbs*, 11:1, 45-57, DOI:10.1080/19420862.2018.1548233, 2019.
4. Yan Q, Huang M, Lewis MJ, Hu P.; Structure Based Prediction of Asparagine Deamidation Propensity in Monoclonal Antibodies *MAbs*,10(6):901-912. doi:10.1080/19420862.2018.1478646, 2018.
5. Wang, X., Das, T. K., Singh, S. K., Kumar, S Potential aggregation prone regions in biotherapeutics: a survey of commercial monoclonal antibodies. In *MAbs*, (Vol. 1, No. 3, pp. 254-267).
6. Yan, Qingrong, et al. Yan, Qingrong, et al. "Structure based prediction of asparagine deamidation propensity in monoclonal antibodies. *MAbs.* , Vol. 10. No. 6. Taylor Francis, 2018.