

PERSPECTIVE

Open Access



Utilizing cross-product prior knowledge to rapidly de-risk chemical liabilities in therapeutic antibody candidates

Alex W. Jacobitz¹, Wilfredo Rodezno¹ and Neeraj J. Agrawal^{2*}

Abstract

There is considerable pressure in the pharmaceutical industry to advance better molecules faster. One pervasive concern for protein-based therapeutics is the presence of potential chemical liabilities. We have developed a simple methodology for rapidly de-risking specific chemical concerns in antibody-based molecules using prior knowledge of each individual liability at a specific position in the molecule's sequence. Our methodology hinges on the development of sequence-aligned chemical liability databases of molecules from different stages of commercialization and on sequence-aligned experimental data from prior molecules that have been developed at Amgen. This approach goes beyond the standard practice of simply flagging all instances of each motif that fall in a CDR. Instead, we de-risk motifs that are common at a specific site in commercial mAb-based molecules (and therefore did not previously pose an insurmountable barrier to commercialization) and motifs at specific sites for which we have prior experimental data indicating acceptably low levels of modification. We have used this approach successfully to identify candidates in a discovery phase program with exclusively very low risk potential chemical liabilities. Identifying these candidates in the discovery phase allowed us to bypass protein engineering and accelerate the program's timeline by 6 months.

Keywords: Antibody, Chemical liabilities, In silico, Cross-product analysis

Introduction

Pharmaceutical candidates must be screened for many different potential liabilities that could adversely affect their manufacturability, storage, and function. Many of these potential liabilities for protein-based therapeutics including monoclonal antibodies (mAb) like viscosity, aggregation, pharmacokinetics, and process yields, can be measured as attributes for the molecule as a whole. Chemical liabilities on the other hand are associated with a specific amino acid residue within the sequence of the molecule. Manifestations of any of these attributes at an inappropriate level could be catastrophic to drug development, so considerable time and money are invested to predict and prevent their appearance.

Predicting these attributes as early as possible can allow molecules that are likely to harbor unfavorable characteristics to be removed from the screening process early on or engineered to remediate any shortcomings. The majority of the in silico predictive models created to date have been based on first principles (Agrawal et al. 2016, 2018; Chennamsetty et al. 2015; Sáenz-Suárez et al. 2016; Sharma et al. 2014), as it is difficult to amass datasets that are both large enough, and diverse enough, to train effective machine learning models for biologics—save for a few notable exceptions (Jia and Sun 2017; Sankar et al. 2018; Yang et al. 2017). These models have seen different levels of success, but none has been widely implemented in the industry. There are a handful of themes that are becoming generally accepted: negative charge patches in the variable region of a mAb predictive of high viscosity (Agrawal et al. 2016; Sharma et al. 2014; Chaudhri et al. 2013; Yadav et al. 2012; Li et al. 2014; Buck et al. 2015),

*Correspondence: agrawaln@amgen.com

² Amgen Inc., 360 Binney St., Cambridge, MA 02142, USA
Full list of author information is available at the end of the article

positive charge patches in the variable region correlating with poor pharmacokinetics (Schoch et al. 2015; Boswell et al. 2010; Igawa et al. 2010; Khawli et al. 2004), solvent accessibility correlating with tryptophan (Sharma et al. 2014; Barnett et al. 2019; Folzer et al. 2015; Ehrenshaft et al. 2015), and methionine oxidation (Agrawal et al. 2018; Chennamsetty et al. 2015; Sankar et al. 2018; Yang et al. 2017; Barnett et al. 2019), but each research group tends to develop their own model based on their own dataset, and there is seldom any agreement on a preferred method. Here, we detail a different take on attribute prediction, focusing specifically on de-risking potential chemical liabilities by directly comparing to cross-product sequence and experiment databases of internal and external molecules. Our approach hinges on the growing availability of data from prior antibody-based molecules across all stages of the pipeline, from discovery to late phase, and is sensitive to the fact that in most cases we still lack the requisite size and breadth of data required to yield functional machine learning models.

The case for utilizing cross-product data

The pharmaceutical industry is growing, with a total global market of USD 1.23 trillion in 2020 projected to increase to \$1.7 trillion by 2025. Growth in the market overall leads to increased competition for specific targets and increased pressure to achieve first to market status and the associated 6% average increase in market share (Pharma's first-to-market advantage | McKinsey 2021). Given that most potential drug targets are simultaneously researched by multiple companies, the pressure to be first to market has led to an increasing focus on speed of development. Still, getting to market first with an inferior product is not acceptable, as a better successor molecule can rapidly render those first-to-market benefits obsolete. Molecules in this intensely competitive environment cannot afford to be burdened with liabilities that will affect manufacturing, distribution, or patient experience, but they still need to get to market faster than the competition.

In order to move molecules through the pipeline more rapidly, one should be able to effectively predict any potential issues a candidate molecule could face and either reject that molecule outright or rapidly engineer out those liabilities. To do this effectively without wasting time and resources, engineering efforts need to be pinpointed to liabilities that are most likely to be problematic to avoid over-engineering. Many of the standard prediction methods for chemical liabilities (e.g., scanning for deamidation by motif) have a tendency to over-predict, leading to extensive engineering efforts to remediate potential issues that would have never posed a problem if left unaltered. Existing predictive modeling efforts seek

to reduce the burden of engineering by using computational models to more effectively predict the specific residues in a molecule's sequence that are most likely to be problematic. In our method, we use a simple statistical approach to effectively utilize the ever-growing array of data that has been collected from previously studied molecules to rapidly and thoroughly de-risk individual sites in new molecules of interest.

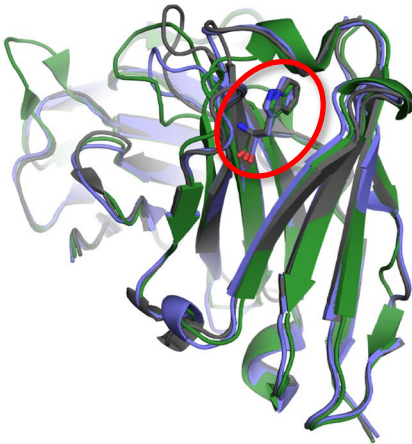
Cross-product sequence analysis to rapidly derisk potential chemical liabilities

Our approach hinges on the fact that antibody-based molecules, which make up the majority of therapeutically relevant proteins, are built onto a highly conserved immunoglobulin scaffold. This conservation makes it possible to rapidly gain structural information about a given site on a molecule simply by aligning its sequence to a reference framework. There are several prominent methodologies in the literature for numbering the residues of an antibody's sequence based on alignment to a reference (Kabat (Te Wu and Kabat 1970), Chothia (Al-Lazikani et al. 1997), AHO (Annemarie and Andreas 2001), IMGT (Lefranc et al. 2005), etc.). For the purpose of aligning residue level information to structure, we prefer the AHO numbering system (Annemarie and Andreas 2001) because it more appropriately handles CDR loops of differing lengths (Fig. 1A, B); however, our approach is general enough to utilize alignment in any other scheme. Having numbered residues correspond to specific positions in the immunoglobulin structure allows you to make inferences about the environment that a given residue (or sequence motif) will experience. We can then use this as the basis for comparing a motif at a given position to both sequence prevalence data and experimental data from prior molecules that also harbored that same motif at the same aligned position.

By pre-aligning tables of predicted hotspots across a database of sequences to a reference numbering system, we can rapidly compare a specific site in a molecule of interest to other molecules that achieved varying levels of success and apply a kind of "survival of the fittest molecule" approach to determining if a particular site in a particular molecule is likely to be problematic (Fig. 2A). Sites that are extremely common in existing commercial mAb-based molecules (commercial mAb-based molecules here means FDA or EU approved molecules for which 94 sequences were publicly available at the time of publication) are unlikely to cause an issue that would be disruptive to commercialization and can be effectively de-risked. As an example, there is a potential methionine oxidation site located at position 41 in H_CDR1 that is present in >50% of commercial mAb-based molecules (Table 1). The high prevalence of methionine residues at

A

Aho	33	34	38	39	40	41	42
4X7T							
Linear	31	-	32	33	34	35	36
IMGT	35	-	36	37	38	39	40
Kabat	31	-	32	33	34	35	35A
Sequence	S	-	G	Y	S	W	N
5SX5							
Linear	30	31	32	33	34	35	36
IMGT	34	35	36	37	38	39	40
Kabat	31	32	33	34	35	35A	35B
Sequence	S	G	D	Y	Y	W	T
6B3S							
Linear	30	31	32	33	34	35	36
IMGT	34	35	36	37	38	39	40
Kabat	31	32	33	34	35	35A	35B
Sequence	S	G	D	Y	Y	W	S



B

Aho	109	110	111	112	113	114	115	116	117	118	128	129	130	131	132	133	134	135	136	137	138
6UMH																					
Linear	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119
IMGT	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
Kabat	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115
Sequence	D	R	L	N	Y	Y	D	S	S	G	Y	Y	H	Y	K	Y	Y	G	M	A	V
3GIZ																					
Linear	99	100	101	102	103	104									105	106	107	108	109	110	111
IMGT	107	108	109	110	111	112									112.1	112.2	113	114	115	116	117
Kabat	95	96	97	98	99	100									100A	100B	100C	100D	100E	100F	100G
Sequence	D	I	Q	Y	G	N	-	-	-	-	-	-	-	-	Y	Y	Y	G	M	D	V
5WUV																					
Linear	99	100	101	102													103	104	105	106	107
IMGT	107	108	109	110													113	114	115	116	117
Kabat	95	96	97	98													99	100	100A	100B	100C
Sequence	G	Y	R	S	-	-	-	-	-	-	-	-	-	-	-	-	Y	A	M	D	Y
6AL4																					
Linear	225	226	227	228	229	230	231								232	233	234	235	236	237	238
IMGT	107	108	109	110	111	112	113								112.2	112.1	112	113	114	115	116
Kabat	95	96	97	98	99	100	101								100B	100C	100D	100E	100F	100G	100H
Sequence	R	E	T	T	T	V	G	-	-	-	-	-	-	-	R	Y	Y	Y	A	M	D

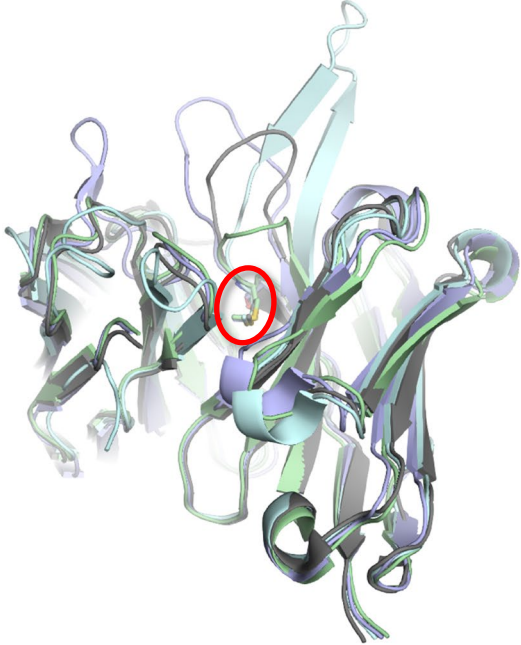


Fig. 1 Reference alignment of sequences. All sequences are aligned to the AHo system with other numbering systems shown above each indicated sequence. **A** Comparison of structures from commercial mAb-based molecules with Trp at AHo position 41 and crystal structures in the PDB. The Trp at position 41 in the Aho numbering system falls at 35 or 35A in the Kabat system (red highlighted column) even though this residue is structurally identical across the different mAb crystal structures (Trp41 shown as sticks and circled in Red; PDB 4X7T (Jensen et al. 2015), green; PDB 5SX5 (Sickmier et al. 2016), grey; PDB 6B3S (Bagchi et al. 2018), blue). **B** Comparison of structures from commercial mAb-based molecules with Met at position 136 in AHo numbering and crystal structures in the PDB. The Met at AHo position 136 is numbered 100A, 100E, 100G, or 100M in the Kabat system across the different molecules (red highlighted column) even though crystal structures indicate the position of this Met is highly homologous across the structures (Met136 shown as sticks and circled in Red; PDB 6AL4 (Teplyakov et al. 2018), light blue; PDB 5WUV (Lee et al. 2017), light green; PDB 6UMH (Garces et al. 2020), light cyan; PDB 3GIZ (Du et al. 2009), light gray)

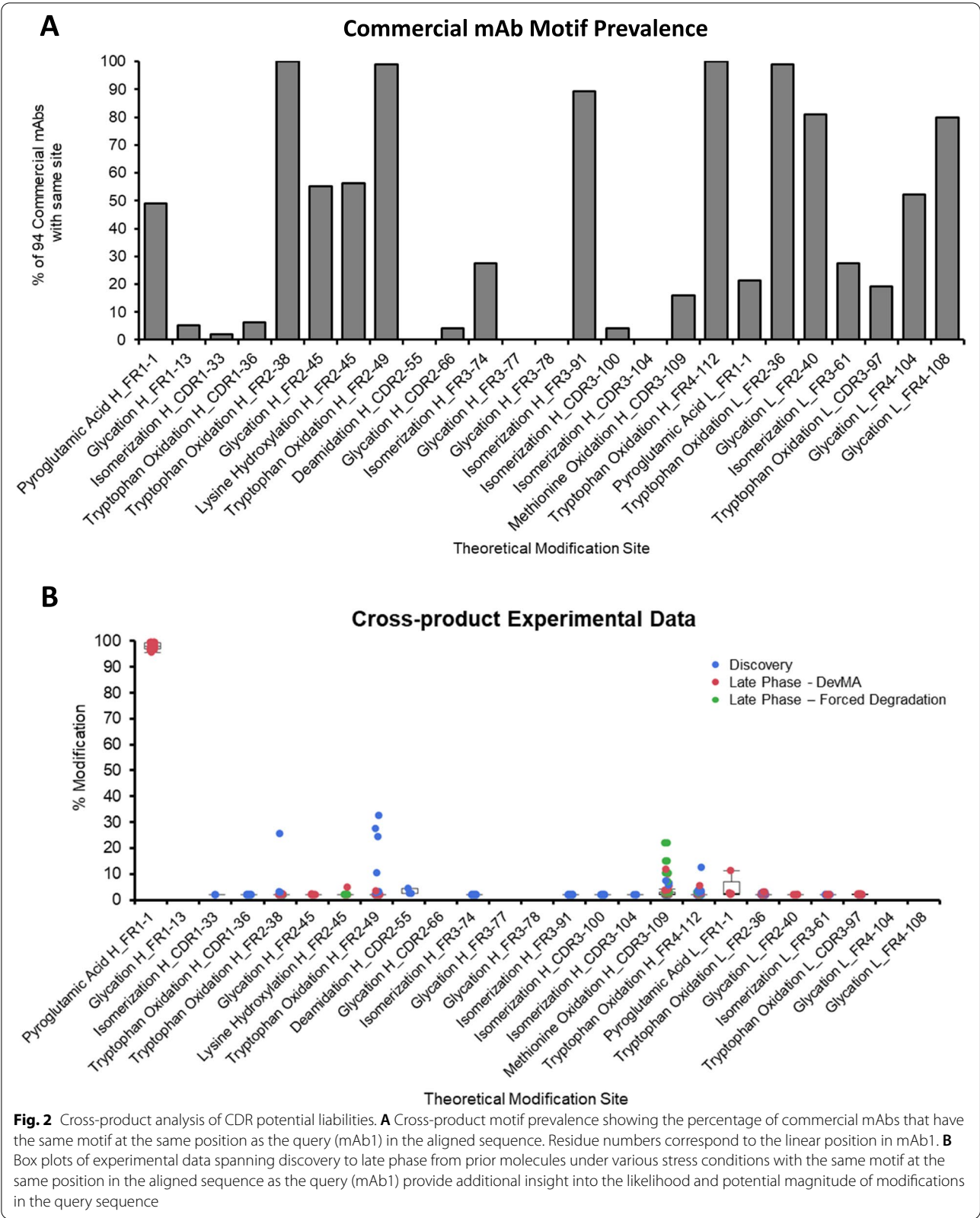


Fig. 2 Cross-product analysis of CDR potential liabilities. **A** Cross-product motif prevalence showing the percentage of commercial mAbs that have the same motif at the same position as the query (mAb1) in the aligned sequence. Residue numbers correspond to the linear position in mAb1. **B** Box plots of experimental data spanning discovery to late phase from prior molecules under various stress conditions with the same motif at the same position in the aligned sequence as the query (mAb1) provide additional insight into the likelihood and potential magnitude of modifications in the query sequence

Table 1 Most common potential chemical liability motifs among variable regions of commercial mAbs

Potential Chemical Liability	AHo#	Position	Motif	Domain	% Commercial mAbs (n = 94)
Tryptophan Oxidation	43	H_FR2	W	HC	100
Tryptophan Oxidation	139	H_FR4	W	HC	100
Tryptophan Oxidation	43	L_FR2	W	LC	98.94
Tryptophan Oxidation	54	H_FR2	W	HC	98.94
Isomerization	100	H_FR3	DT	HC	89.36
O-Xylosylation	81	L_FR3	GSG	LC	89.36
O-Xylosylation	83	L_FR3	GSG	LC	85.11
Glycation	47	L_FR2	KPG	LC	80.85
Glycation	148	L_FR4	KRT	LC	79.79
Methionine Oxidation	41	H_CDR1	M	HC	63.83
Lysine Hydroxylation	75	H_CDR2	KG	HC	58.51
Methionine Oxidation	4	L_FR1	M	LC	57.45
Lysine Hydroxylation	50	H_FR2	KG	HC	56.38
Glycation	50	H_FR2	KGL	HC	55.32
Glycation	144	L_FR4	KVE	LC	52.13
Methionine Oxidation	93	H_FR3	M	HC	51.06
Glycation	75	H_CDR2	KGR	HC	48.94
Pyroglutamic Acid	1	H_FR1	Q	HC	48.94
Deamidation	94	H_FR3	NS	HC	47.87
Pyroglutamic Acid	1	H_FR1	E	HC	47.87
Glycation	53	L_FR2	KLL	LC	46.81
Isomerization	17	L_FR1	DR	LC	46.81
Glycation	50	L_FR2	KAP	LC	43.62
Isomerization	137	H_CDR3	DY	HC	42.55
O-Sulfation	138	H_CDR3	DY	HC	42.55
Glycation	14	H_FR1	KPG	HC	39.36
Isomerization	83	H_FR3	DN	HC	38.3
Methionine Oxidation	91	H_FR3	M	HC	37.23
Deamidation	87	H_FR3	NT	HC	36.17
Non-Consensus N-linked Glycosylation	87	H_FR3	SKN	HC	36.17
Glycation	24	H_FR1	KAS	HC	34.04
Glycation	86	H_FR3	KNT	HC	30.85
Glycation	144	L_FR4	KLE	LC	30.85
Isomerization	72	H_CDR2	DS	HC	28.72
Glycation	13	H_FR1	KKP	HC	27.66
Isomerization	76	L_FR3	DR	LC	27.66
Isomerization	83	H_FR3	DT	HC	27.66
Glycation	20	H_FR1	KVS	HC	25.53
Deamidation	42	L_CDR1	NW	LC	24.47
Deamidation	84	H_FR3	NS	HC	24.47

this site indicates that it is generally not an impediment to commercialization. The exact frequency cutoff below which a site can effectively be de-risked should be defined by the risk tolerance acceptable to a given program.

To supplement this purely sequence-based approach, we have applied a similar principle to the experimental chemical liability data from prior molecules. By

templating our prior molecule experimental chemical liabilities data and aligning it to the AHo numbering system, we can rapidly query the available data across prior molecules that share any given hotspot with a molecule of interest. In this way, we can instantly examine both the average and the spread of the available data for a given motif at a given site across prior molecules and use this

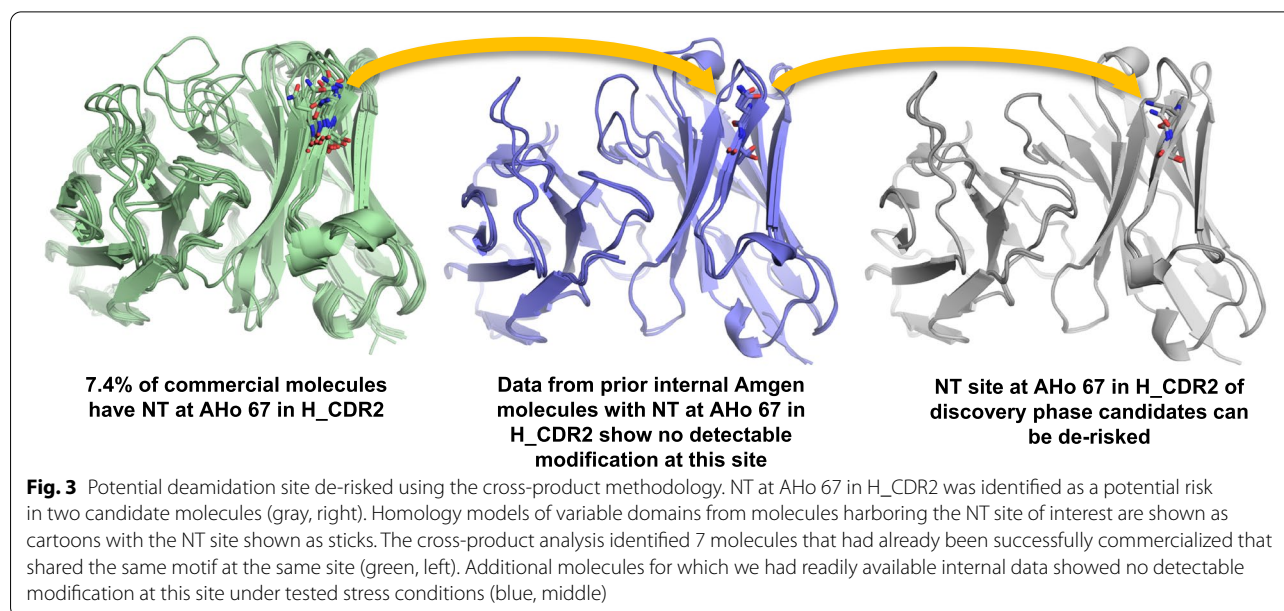
as an indicator of risk likelihood for the same motif at the same site in a molecule of interest (Fig. 2B). Specific sites that show minimal or no modification across all prior programs are unlikely to suddenly become problematic in a new program, and our confidence in this assertion increases as the number of data points increases or as variance in the data decreases.

Currently, we leverage this cross-product sequence analysis method on applicable discovery and development stage programs to accelerate molecules with acceptably low risk profiles or to target protein engineering resources towards the specific sites that we cannot effectively de-risk. In a noteworthy example of this method's success, we utilized an early version of this methodology to de-risk key theoretical sites in a discovery stage mAb program (Fig. 3). By using prior knowledge from cross-product analysis, we were able to bypass the protein engineering stage and move molecules from a discovery campaign directly to development without engineering to remediate potential chemical liability sites. In this case, an Asn residue in H_CDR2 of multiple lead candidates was identified as a potential deamidation risk. This NT motif beginning at AHo position 67 is present in 7.4% of commercial molecules, indicating that there were multiple examples of mAb-based molecules that were able to successfully reach commercialization with this site at this position. Prior internal molecules with the same motif at the same structural position were also identified, and these molecules had previously been shown to have extremely low levels of modification at this site during development (available prior molecule experimental data in our dataset showed deamidation levels

were below the limit of detection for available molecules under all recorded stress conditions). By leveraging this prior knowledge in the decision to move these discovery phase molecules forward without protein engineering, the program's timeline was accelerated by approximately 6 months. Different companies will likely have different existing internal metrics to define what an acceptable level of modification is, and this will also likely be variable for different programs with diverse target candidate profiles. We make no recommendations for what will be necessary to meet the needs of any specific program, but as a general rule, we often look for candidates with all possible chemical liability sites having modification levels below 2–5% under mild stress conditions in prior molecule data (temperatures between 4 and 40 °C, time points out to a max of 4 weeks, pH from 4.5–8.0, cool white light stress up to 200 klux*hr). In order to meet the realization that different programs will have different requirements, additional data under harsher stress conditions (longer time points, added chemical oxidants, etc.) is available in the dataset on demand.

Key requirements, learnings, and future perspectives

This data-driven cross-product approach to de-risking attributes provides incredible value for rapidly assessing the likelihood that a potential chemical liability will impact a candidate molecule. In order to implement this approach, an institution will first need to establish aligned sequence databases of commercial molecules and clinical candidates and pre-screen those sequences for hotspots. For our sequence sets, we focused on three different levels of success to get some granularity into



the frequency of hotspots at specific sites in molecules moving through the drug development pipeline: molecules that progressed to process development, molecules selected for first-in-human (FIH) development, and molecules that achieved marketing authorization approval. Second, the institution will need to develop a suitable data template to maintain sequence-aligned experimental chemical liabilities data and, ideally, populate this database both proactively and retroactively through all stages of development from discovery to late stage. We are actively templating data from our historical programs, with data from hundreds of unique experiments available in our dataset at the time of this publication. While we have certainly seen increasing value from our growing dataset, we also found considerable value in early iterations of this endeavor when data on only a limited number of molecules was available, making this approach easily within reach even for many smaller organizations in the industry. Finally, to make relevant subsets of the data rapidly available for de-risking the specific hotspots in a molecule of interest, software needs to be developed to rapidly access the relevant data from the datasets and present it in a concise manner. The most challenging part of this process for us has been developing a template for our reference aligned chemical liabilities data and populating it with data from past programs as most of this data has historically been stored in PowerPoint and PDF files which has required significant manual effort to find, interpret, structure, and clean so that it can be compared across various programs. Forward-thinking companies will do well to proactively implement strategies to guarantee that future data is automatically stored in a data-science friendly template to make it easily accessible for this and any future use-cases.

Limitations and caveats

The approach described herein is an extremely rapid data-driven method for de-risking potential chemical liability sites in candidate molecules. While there are certainly many benefits to directly querying and assessing sequence and analytical data from prior molecules, it is also important to keep in mind the limitations of this approach. Perhaps the most important of which is that in its present form, these insights are principally of value to antibody-based molecules. While antibody-based molecules represent one of the largest and fastest-growing classes of pharmaceuticals, they are certainly not the only class of biologic therapeutics, and in its present form, this methodology would not be translatable to other molecule types. Additionally, this approach centers on the assumption that the modification propensity of a specific potential chemical liability site can be predicted from the combination of its sequence motif and the position

of that motif within a sequence alignment. This assumes that chemical liability motifs which align together via one of the major antibody sequence alignment/numbering systems (again, our preference is for AHO) will end up in a similar structural position with a similar local environment. While the fact that all antibody domains reliably adopt immunoglobulin folds and the preponderance of available structures would indicate that there is a tremendous amount of structural similarity in these domains, certainly individual chemical liability motifs on individual molecules could fall outside of these assumptions and render the prior molecule data less relevant. Changes in solvent exposure, adjacent charged or hydrophobic patches, or regional flexibility could all potentially affect the modification propensity of a specific residue. While some of this variability in local structural environments will be reflected in the spread of prior analytical data, it is of course still possible for a new molecule to be a structural outlier and subsequently behave outside of what would be expected based on prior data. It is thus important to use this sort of data-driven approach to guide the de-risking process in concert with expert oversight in a phase-appropriate manner. As the number of candidates is reduced towards the selection of a final first-in-human trial candidate, additional scrutiny of candidates for structural relatedness to molecules in the prior knowledge dataset is advisable.

Conclusions and future directions

We have developed a straightforward approach to utilizing prior knowledge to de-risk potential chemical liabilities. The method does not require any type of sophisticated modeling and is therefore not burdened by the need for extremely large datasets or continual retraining of the model. Instead, our method relies on the fact that all immunoglobulin-based molecules will adopt the same fold and that residues and motifs can be mapped to specific positions on the 3-dimensional structure simply by aligning the sequence to a reference numbering system. We can then assume that information we gather about a motif at a particular reference position will be more relevant to that motif at the same reference position in future molecules because of the high degree of conservation inherent in the immunoglobulin fold. This approach will of course lose some predictive value in regions like H-CDR3 where loop lengths and conformations can be quite variable, but it does not completely abrogate all predictive value here. By using a numbering system like that proposed by Honegger et al. (Annemarie and Andreas 2001), it is still possible to draw parallels using this relative spatial positioning even in the more variable CDR loops. This information will be more relevant among CDR loops of similar length and

sequence, and data from motifs falling at the same site in CDR loops that are more closely related to the sequence of interest should be viewed with additional weight. Even without narrowing the dataset to the most similar CDRs, potential predictive power can still be rapidly assessed by simply looking at the variance in the cross-product experimental data at a particular site, presuming it is common enough to have yielded multiple data-points.

This proposed methodology for de-risking attributes is intended to bridge the gap between the current paradigm of gross over-prediction of potential liabilities and a future state where there is enough data available to support high accuracy predictive modeling. This main requirement for implementing this methodology is the creation of templated cross-product datasets which will become more valuable not only within the predictive framework described here, but also for training machine learning models in the future, as the size of the datasets grows.

Acknowledgements

The authors would like to thank Camille Ybanez and Arun Raghu for assisting with data collection and Andrew Dykstra for helpful discussions.

Authors' contributions

AWJ developed the methodology, aggregated sequence and experimental data, designed the data structure, prototyped software for rapidly analyzing query sequences and wrote the paper. WR designed and implemented the final software tool which allows multiple users to access the data and analyze query molecules. NJA assisted in conceptualizing the initial process, revising the paper, and provided support, guidance and oversight. The author(s) read and approved the final manuscript.

Funding

All research was funded by Amgen Inc.

Availability of data and materials

Sequence data for most commercial molecules are available from <https://www.who.int/teams/health-product-and-policy-standards/inn/inn-lists>. Experimental data from Amgen molecules is proprietary.

Declarations

Competing interests

The authors declare they have no competing interests.

Author details

¹Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA.

²Amgen Inc., 360 Binney St., Cambridge, MA 02142, USA.

Received: 3 December 2021 Accepted: 27 April 2022

Published online: 23 May 2022

References

- Agrawal NJ, Dykstra A, Yang J, Yue H, Nguyen X, Kolvenbach C et al (2018) Prediction of the hydrogen peroxide-induced methionine oxidation propensity in monoclonal antibodies. *J Pharm Sci* 107(5):1282–1289
- Agrawal NJ, Helk B, Kumar S, Mody N, Sathish HA, Samra HS et al (2016) Computational tool for the early screening of monoclonal antibodies for their viscosities. *MAbs* 8(1):43–48 Available from: <https://pubmed.ncbi.nlm.nih.gov/26399600/>. Cited 2021 Jul 20
- Al-Lazikani B, Lesk AM, Chothia C (1997) Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 273(4):927–948 Available from: <https://pubmed.ncbi.nlm.nih.gov/9367782/>. Cited 2021 Jul 22
- Annemarie H, Andreas P (2001) Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *J Mol Biol* 8:657–670 Available from: <https://pubmed.ncbi.nlm.nih.gov/11397087/>
- Bagchi A, Haidar JN, Eastman SW, Vieth M, Topper M, Iacolina MD et al (2018) Molecular basis for necitumab inhibition of EGFR variants associated with acquired cetuximab resistance. *Mol Cancer Ther* 17(2):521–531 Available from: <https://mct.aacrjournals.org/content/17/2/521>. Cited 2021 Sep 2
- Barnett GV, Balakrishnan G, Chennamsetty N, Hoffman L, Bongers J, Tao L et al (2019) Probing the tryptophan environment in therapeutic proteins: implications for higher order structure on tryptophan oxidation. *J Pharm Sci* 108(6):1944–1952
- Boswell CA, Tesar DB, Mukhyala K, Theil FP, Fielder PJ, Khawli LA (2010) Effects of charge on antibody tissue distribution and pharmacokinetics. *Bioconjug Chem* 21:2153–2163 Available from: <https://pubmed.ncbi.nlm.nih.gov/21053952/>. Cited 2021 Jul 20
- Buck PM, Chaudhri A, Kumar S, Singh SK (2015) Highly viscous antibody solutions are a consequence of network formation caused by domain-domain electrostatic complementarities: insights from coarse-grained simulations. *Mol Pharm* 12(1):127–139 Available from: <https://pubmed.ncbi.nlm.nih.gov/25383990/>. Cited 2021 Jul 20
- Chaudhri A, Zarraga IE, Yadav S, Patapoff TW, Shire SJ, Voth GA (2013) The role of amino acid sequence in the self-association of therapeutic monoclonal antibodies: insights from coarse-grained modeling. *J Phys Chem B* 117(5):1269–1279 Available from: <https://pubs.acs.org/doi/abs/10.1021/jp3108396>. Cited 2021 Jul 20
- Chennamsetty N, Quan Y, Nashine V, Sadinani V, Lyngberg O, Krystek S (2015) Modeling the oxidation of methionine residues by peroxides in proteins. *J Pharm Sci* 104(4):1246–1255 Available from: <https://pubmed.ncbi.nlm.nih.gov/25641333/>. Cited 2021 Jul 21
- Du J, Yang H, Guo Y, Ding J (2009) Structure of the Fab fragment of therapeutic antibody Ofatumumab provides insights into the recognition mechanism with CD20. *Mol Immunol* 46(11–12):2419–2423
- Ehrenschaft M, Deterding LJ, Mason RP (2015) Tripping up Trp: modification of protein tryptophan residues by reactive oxygen species, modes of detection, and biological consequences. *Free Radic Biol Med* 89:220–228 Elsevier Inc
- Folzer E, Diepold K, Bomans K, Finkler C, Schmidt R, Bulau P et al (2015) Selective oxidation of methionine and tryptophan residues in a therapeutic IgG1 molecule. *J Pharm Sci* 104(9):2824–2831
- Garces F, Mohr C, Zhang L, Huang CS, Chen Q, King C et al (2020) Molecular insight into recognition of the cgrpr complex by migraine prevention therapy Aimovig (Erenumab). *Cell Rep* 30(6):1714–1723.e6
- Igawa T, Tsunoda H, Tachibana T, Maeda A, Mimoto F, Moriyama C et al (2010) Reduced elimination of IgG antibodies by engineering the variable region. *Protein Eng Des Sel* 23(5):385–392 Available from: <https://pubmed.ncbi.nlm.nih.gov/20159773/>. Cited 2021 Jul 20
- Jensen RK, Plum M, Tjerrild L, Jakob T, Spillner E, Andersen GR et al (2015) Structure of the omalizumab Fab. *urn:issn:2053-230X*. 71(4):419–426 Available from: <http://scripts.iucr.org/cgi-bin/paper?r15093>. Cited 2021 Sep 2
- Jia L, Sun Y (2017) Protein asparagine deamidation prediction based on structures with machine learning methods. *PLoS One* 12(7):e0181347 Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181347>. Cited 2021 Jul 22
- Khawli LA, Mizokami MM, Sharifi J, Hu P, Epstein AL (2004) Pharmacokinetic characteristics and biodistribution of radioiodinated chimeric TNT-1, -2, and -3 monoclonal antibodies after chemical modification with biotin. *Cancer Biother Radiopharm* 17(4):359–370 <https://home.liebertpub.com/cbr>. Available from: <https://www.liebertpub.com/doi/abs/10.1089/108497802760363150>. Cited 2021 Jul 20
- Lee JU, Shin W, Son JY, Yoo K-Y, Heo Y-S (2017) Molecular basis for the neutralization of tumor necrosis factor α by certolizumab pegol in the treatment of inflammatory autoimmune diseases. *Int J Mol Sci* 18(1):228 Available from: <https://www.mdpi.com/1422-0067/18/1/228/htm>. Cited 2021 Sep 2

- Lefranc MP, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D et al (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Developmental and Comparative Immunology. Dev Comp Immunol:185–203 Available from: <https://pubmed.ncbi.nlm.nih.gov/15572068/>
- Li L, Kumar S, Buck PM, Burns C, Lavoie J, Singh SK et al (2014) Concentration dependent viscosity of monoclonal antibody solutions: explaining experimental behavior in terms of molecular properties. Pharm Res 31(11):3161–3178 Available from: <https://pubmed.ncbi.nlm.nih.gov/24906598/>. Cited 2021 Jul 20
- Pharma's first-to-market advantage | McKinsey. Available from: <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/pharmas-first-to-market-advantage#>. Cited 2021 Jul 21.
- Sáenz-Suárez H, Poutou-Piñales RA, González-Santos J, Barreto GE, Rieto-Navarera LP, Saenz-Moreno JA et al (2016) Prediction of glycation sites: new insights from protein structural analysis. Artic TURKISH J Biol Available from: <https://www.researchgate.net/publication/274780099>
- Sankar K, Hoi KH, Yin Y, Ramachandran P, Andersen N, Hilderbrand A et al (2018) Prediction of methionine oxidation risk in monoclonal antibodies using a machine learning method. MAbs 10(8):1281–1290 Available from: <https://pubmed.ncbi.nlm.nih.gov/30252602/>. Cited 2021 Jul 21
- Schoch A, Kettenberger H, Mundigl O, Winter G, Engert J, Heinrich J et al (2015) Charge-mediated influence of the antibody variable domain on FcRn-dependent pharmacokinetics. Proc Natl Acad Sci 112(19):5997–6002 Available from: <https://www.pnas.org/content/112/19/5997>. Cited 2021 Jul 20
- Sharma VK, Patapoff TW, Kabakoff B, Pai S, Hilario E, Zhang B et al (2014) In silico selection of therapeutic antibodies for development: viscosity, clearance, and chemical stability. Proc Natl Acad Sci U S A 111(52):18601–18606
- Sickmier EA, Kurzeja RJM, Michelsen K, Vazir M, Yang E, Tasker AS (2016) The panitumumab EGFR complex reveals a binding mechanism that overcomes cetuximab induced resistance. PLoS One 11(9):e0163366 Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0163366>. Cited 2021 Sep 2
- Te Wu T, Kabat EA (1970) An analysis of the sequences of the variable regions of bence jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med 132(2):211–250 Available from: <https://pubmed.ncbi.nlm.nih.gov/5508247/>. Cited 2021 Jul 16
- Tepljakov A, Obmolova G, Luo J, Gilliland GL (2018) Crystal structure of B-cell co-receptor CD19 in complex with antibody B43 reveals an unexpected fold. Proteins Struct Funct Bioinforma 86(5):495–500 Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/prot.25485>. Cited 2021 Sep 2
- Yadav S, Laue TM, Kalonia DS, Singh SN, Shire SJ (2012) The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. Mol Pharm 9(4):791–802 Available from: <https://pubs.acs.org/doi/abs/10.1021/mp200566k>. Cited 2021 Jul 20
- Yang R, Jain T, Lynaugh H, Nobrega RP, Lu X, Boland T et al (2017) Rapid assessment of oxidation via middle-down LCMS correlates with methionine side-chain solvent-accessible surface area for 121 clinical stage monoclonal antibodies. MAbs 9(4):646–653 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5419077/?report=abstract>. Cited 2020 Nov 19

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)