

# The Therapeutic Antibody Profiler (TAP): Five Computational Developability Guidelines

Matthew I. J. Raybould<sup>1</sup>, Claire Marks<sup>1</sup>, Konrad Krawczyk<sup>1</sup>, Bruck Taddese<sup>2</sup>, Jaroslaw Nowak<sup>1</sup>, Alan P. Lewis<sup>3</sup>, Alexander Bujotzek<sup>4</sup>, Jiye Shi<sup>5</sup>, and Charlotte M. Deane<sup>1</sup>✉

<sup>1</sup>Department of Statistics, University of Oxford, Oxford, OX1 3LB, UK

<sup>2</sup>Department of Antibody Discovery and Protein Engineering, MedImmune, Milstein Building, Granta Park, Cambridge, CB21 6GH, UK

<sup>3</sup>Computational and Modelling Sciences, GlaxoSmithKline Research and Development, Stevenage, SG1 2NY, UK

<sup>4</sup>Roche Pharma Research and Early Development, Large Molecule Research, Roche Innovation Center Munich, Penzberg, Germany

<sup>5</sup>Chemistry Department, UCB Pharma, Slough, SL1 3WE, UK

Therapeutic monoclonal antibodies (mAbs) must not only bind to their target but must also be free from ‘developability issues’, such as poor stability or high levels of aggregation. While small molecule drug discovery benefits from Lipinski’s rule of five to guide the selection of molecules with appropriate biophysical properties, there is currently no *in silico* analog for antibody design. Here, we model the variable domain structures of a large set of post-Phase I clinical-stage antibody therapeutics (CSTs), and calculate an array of metrics to estimate their typical properties. In each case, we contextualize the CST distribution against a snapshot of the human antibody gene repertoire. We describe guideline values for five metrics thought to be implicated in poor developability: the total length of the Complementarity-Determining Regions (CDRs), the extent and magnitude of surface hydrophobicity, positive charge and negative charge in the CDRs, and asymmetry in the net heavy and light chain surface charges. The guideline cut-offs for each property were derived from the values seen in CSTs, and a flagging system is proposed to identify non-conforming candidates. On two mAb drug discovery sets, we were able to selectively highlight sequences with developability issues. We make available the Therapeutic Antibody Profiler (TAP), an open-source computational tool that builds downloadable homology models of variable domain sequences, tests them against our five developability guidelines, and reports potential sequence liabilities and canonical forms. TAP is freely available at <http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/TAP.php>

therapeutic monoclonal antibodies | developability guidelines | immunoglobulin gene sequencing | surface hydrophobicity | surface charge

Correspondence: [deane@stats.ox.ac.uk](mailto:deane@stats.ox.ac.uk)

## Introduction

Monoclonal antibodies (mAbs) are increasingly used as therapeutics targeting a wide range of membrane-bound or soluble antigens - of the 73 antibody therapies approved by the EU or FDA since 1986 (valid as of June 12<sup>th</sup>, 2018), 10 were first approved in 2017 (1). There are many barriers to therapeutic mAb development, besides achieving the desired affinity to the antigen. These include intrinsic immunogenicity, chemical and conformational instability, self-association, high viscosity, polyspecificity, and poor expression. *In vitro* screening for these negative characteristics is now routine in industrial pipelines (2).

While some cases of poor developability are subtle in origin, others are less ambiguous. High levels of hydropho-

bicity, particularly in the highly variable complementarity-determining regions (CDRs), have repeatedly been implicated in aggregation, viscosity and polyspecificity (2–8). Asymmetry in the net charge of the heavy and light chain variable domains is also correlated with self-association and viscosity at high concentrations (4, 9). Patches of positive (10) and negative (11) charge in the CDRs are linked to high rates of clearance and poor expression levels. Product heterogeneity (e.g. through oxidation, isomerisation, or glycosylation) often results from specific sequence motifs liable to post- or co-translational modification.

An improved understanding of the factors governing these biophysical properties has enabled the development of *in silico* assays, which are faster and cheaper than their experimental equivalents. Computational tools already facilitate the identification of sequence liabilities, such as sites of lysine glycation (12), aspartate isomerisation (13), asparagine deamidation (13), and the presence of cysteines or N-linked glycosylation sites (14). A primary focus in recent years has been in designing software that can better predict aggregation proclivity. Many algorithms designed for this purpose use only the antibody sequence (4, 7, 8), although some suggest an analogous equation to use if a structure is available (4). One purely structure-based method is the structural aggregation propensity (SAP) metric (5), later included in the Developability Index (6). This has been shown to detect aggregation-prone regions, such as surface patches (15), and to be able to rank candidates relative to a known antibody developability profile (11), using a closely related antibody crystal structure. It is likely that SAP’s atomic-resolution analysis would be too sensitive to use in comparing static homology models of diverse antibodies, given the current accuracy of structure prediction (16).

An alternative approach to predict antibodies likely to have poor developability profiles is to highlight those candidates whose characteristics differ greatly from clinically-tested therapeutic mAbs; a similar strategy in the field of pharmacokinetics led to the Lipinski rules for small molecule drug design (17). Here, we build three-dimensional models of a large set of post-Phase I therapeutics and survey their sequence and structural properties, all of which are contextualized against human immunoglobulin gene sequencing (Ig-seq) data.

Using the distributions of these properties, we build the Therapeutic Antibody Profiler (TAP), an open-source computational tool that highlights antibodies with anomalous values compared to therapeutics. TAP builds a downloadable structural model of an antibody variable domain sequence, and tests it against guideline thresholds of five calculated measures likely to be linked to poor developability. It also reports potential sequence liabilities and all non-CDRH3 loop canonical forms.

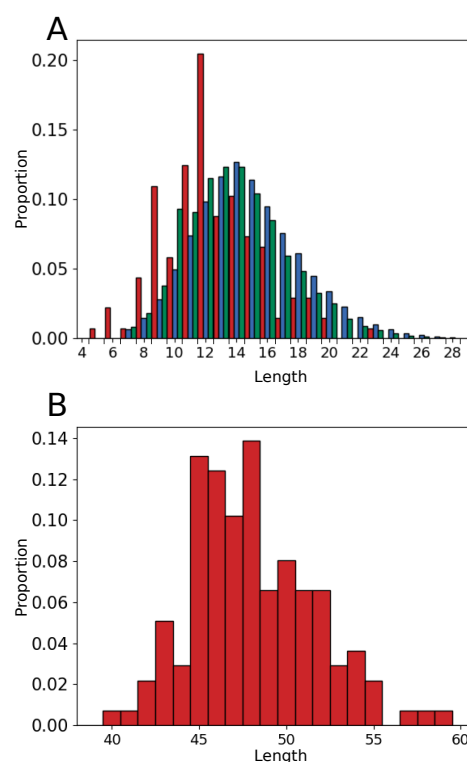
## Results

**Sequence Data.** As a dataset of mAbs unlikely to suffer with developability issues, we used the variable domain heavy and light chain sequences of 137 clinical-stage antibody therapeutics (“137 CSTs”) (18). To contextualize the properties of the CST set, we used a snapshot of the human antibody chain repertoire from a recent Ig-seq study by UCB Pharma Ltd. (“human Ig-seq”). The human Ig-seq dataset was analyzed as a set of non-redundant heavy or light chains (“human Ig-seq non-redundant chains”), and as a set of non-redundant CDR sequences (“human Ig-seq non-redundant CDRs”). We chose this Ig-seq data as it contains simultaneously sequenced heavy and light chains, and so is a promising starting point for realistic *in silico* pairing, required to make complete structural models.

**Model Structures.** High quality structural information is critical to accurately predict the surface properties of antibodies. As crystal structures are often unavailable, or difficult to attain, accurate modeling is a necessary step of an effective antibody profiler. Accordingly, all our comparisons are made between models (even when crystal structures are available) to avoid a bias in terms of structural quality. ABodyBuilder (19) was run on the 56 CSTs with a reference PDB (20) structure (as of May 4<sup>th</sup>, 2018). Sequence identical templates were not included, and each resulting model was aligned to its reference to evaluate the backbone root-mean-square deviation (RMSD) across all IMGT regions (see *SI Methods*). The mean framework and CDR RMSDs (Table S1) were commensurate with the current state of the art (16). For our structural property calculations, we class surface-exposed residues as having a side chain with relative accessible surface area ( $ASA_{rel,X} \geq 7.5\%$ , compared to Alanine-X-Alanine for each residue X (21, 22)). Using this definition, we identified all exposed residues in the models and PDB structures. Of the 7,057 exposed crystal structure residues, only 265 (3.76%) were wrongly assigned as buried in the models.

The results suggest that ABodyBuilder models are accurate enough for our analysis, we used this software to model all 137 CSTs (“137 CST models”) and a diverse subset of paired human Ig-seq chains (19,019 “human Ig-seq models”, see *SI Methods*). We then performed a series of *in silico* assays to determine the TAP guideline metrics.

**IMGT CDR Lengths.** Loop length has a major impact on the nature of antigen binding. For example, if an antibody has a long CDRH3 loop, it tends to form most of the interactions



**Fig. 1.** (A) Comparing the CDRH3 length distributions of the 137 CSTs (red), 1,696,918 human Ig-seq non-redundant CDRH3s (blue), and 4,587,907 human Ig-seq non-redundant heavy chains (green). The CSTs have a lower median CDRH3 length. (B) The distribution of total CDR length for the 137 CST dataset.

with an antigen, whilst shorter CDRH3 loops contribute to concave binding sites where other CDRs more often assist in binding (23).

The 137 CST and human Ig-seq sequences were IMGT-numbered (24), and IMGT CDR definitions were used to split the sequences by region. The 137 CST CDRH3 loops had a median length of 12, compared to 14 for the human Ig-seq datasets (Fig. 1A). In the case of CDRL3 the distributions were closer, with a median length of 9 for the 137 CSTs, compared to 10 for the human Ig-seq datasets (Fig. S1E).

To test whether hybridoma development might account for these findings - as it is known that mouse antibodies tend to have shorter CDRH3 loops than human antibodies (25) - we split the 137 CST dataset by developmental origin (Fig. S2). Fully human therapeutics were disproportionately represented at longer CDRH3s (mean: 13.21, median: 12), compared to chimeric, humanized, or fully murine therapeutics (mean: 11.91, median: 12). However, both therapeutic subsets still have shorter CDRH3s than human-expressed antibodies.

We also plotted the combined length of all CDRs for each antibody in the 137 CST dataset (Fig. 1B, median value 48). The human Ig-seq data is not paired, and so we could not plot its total CDR length. Our artificially paired human Ig-seq models were built to share a similar total CDR length distribution to the CSTs (Fig. S3), so that CDR length would not bias comparisons in other metrics. The 137 CST total CDR length was highly correlated to CDRH3 length (Pear-

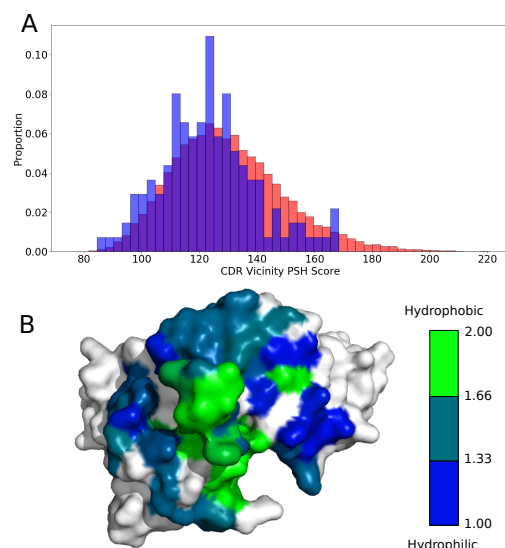
son's correlation coefficient of +0.77, with a two-tailed p-value of  $2.44 \times 10^{-28}$ ). As the total length of the CDRs captures both binding site shape (lower value, more concave) and CDRH3 length (typically shorter in CSTs than our human Ig-seq heavy chains), this metric was selected for inclusion in the final five TAP guidelines.

**Canonical Forms.** In natural antibodies, all CDR loops, apart from CDRH3, are thought to fall into structural classes known as canonical forms (26, 27). We assigned length-independent canonical forms (see *Methods*) to the 137 CST and human Ig-seq models (Fig. S4). All assignable CST model CDRs were labeled with a canonical form also present in the human Ig-seq model dataset. Fewer than 19% of CST CDRs remained unassigned in each loop region, suggesting that, despite engineering, a clear majority of non-CDRH3 therapeutic CDR loops adopt well-characterized canonical forms. TAP reports the canonical form of each modeled loop, highlighting if any cannot be assigned.

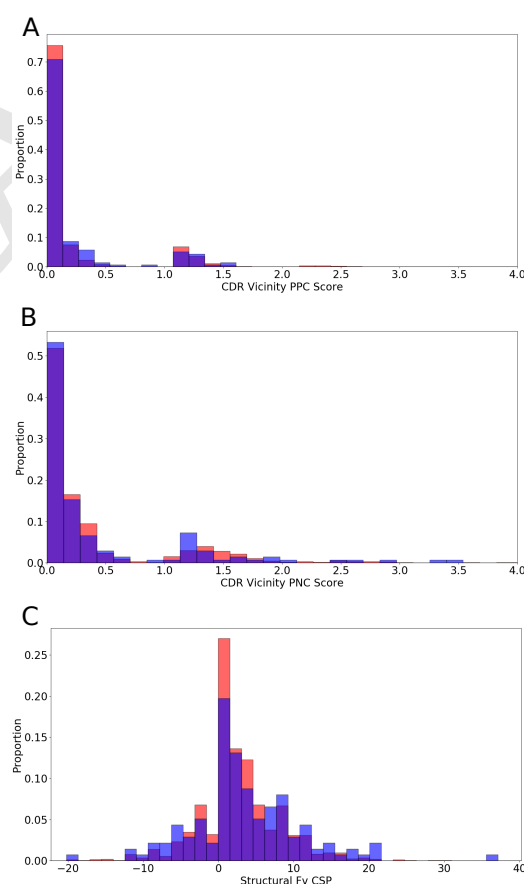
**Hydrophobicity.** Hydrophobicity in the CDR regions has been repeatedly linked to aggregation propensity in mAbs (2, 6–8). Using our homology models, we estimated the effective hydrophobicity of each residue by considering not only its degree of apolarity, but also whether or not it is solvent-exposed (side chain relative  $ASA_{rel} > 7.5\%$  (21, 22)). As the energy of the hydrophobic effect is approximately proportional to the interface area (28), we developed a metric (Patches of Surface Hydrophobicity, PSH, see *Methods*) that yields higher scores if hydrophobic residues tend to neighbor one another in a region, rather than being evenly separated. We evaluated PSH for the 137 CST and human Ig-seq models across two regions (the CDR vicinity (see *Methods*) and the entire variable (Fv) region), and with five different hydrophobicity scales (29–33).

The results of all hydrophobicity scales were highly correlated (e.g.  $R^2 \geq 0.91$  between all scales in the CDR vicinity). The mean CDR vicinity PSH values for the CST and Ig-seq distributions were  $123.30 \pm 16.60$  and  $130.10 \pm 19.53$  respectively (Kyte hydrophobicity scale (29), Fig. 2A). CSTs were noticeably underrepresented at higher CDR PSH values; galiximab is a rare example of a therapeutic antibody with a high value (Fig. 2B). The same divergence occurred to a much lesser extent across the entire Fv region, with mean values of  $357.69 \pm 22.95$  and  $363.13 \pm 20.64$  respectively (Fig. S5). This supports the theory that the high concentration conditions under which therapeutics are stored may render them less tolerant of large patches of hydrophobicity in the highly-exposed CDR vicinity, and also suggests that a subset of natural human antibodies would be unsuitable therapeutic candidates. We therefore included the CDR vicinity PSH score as a TAP guideline metric.

**Charge.** Surface patches of positive or negative charge have also been linked to negative biophysical characteristics (10, 11). We calculated two metrics designed to highlight regions of dense charge: the Patches of Positive Charge (PPC) and Patches of Negative Charge (PNC) measures (see *Methods*).



**Fig. 2.** (A) CDR vicinity PSH scores across the 137 CST (blue) and human Ig-seq (red) models. The CSTs are underrepresented at higher PSH values. (B) Galiximab (Kyte & Doolittle CDR vicinity PSH score of 167.89) has a large surface-exposed patch of hydrophobicity in its CDRH3 loop. Heavy and light chain surfaces outside the CDR vicinity are colored in white.



**Fig. 3.** Histograms of 137 CST (blue) and human Ig-seq model (red) values of (A) patch-weighted positive charge (PPC) in the CDR vicinity, and (B) patch-weighted negative charge (PNC) in the CDR vicinity. In both metrics, the datasets are biased away from higher scores. (C) Histogram of Structural Fv Charge Symmetry Parameter values. Both datasets show a bias away from negative values.

| Metric            | Amber Region     | Red Region  |
|-------------------|------------------|-------------|
| 1. Total CDR Len. | Bottom 5%/Top 5% | Above/Below |
| 2. PSH            | Bottom 5%/Top 5% | Above/Below |
| 3. PPC            | Top 5%           | Above       |
| 4. PNC            | Top 5%           | Above       |
| 5. SFvCSP         | Bottom 5%        | Below       |

**Table 1.** TAP amber and red flag cut-off thresholds, with respect to the clinical-stage therapeutic distributions. Len = Length.

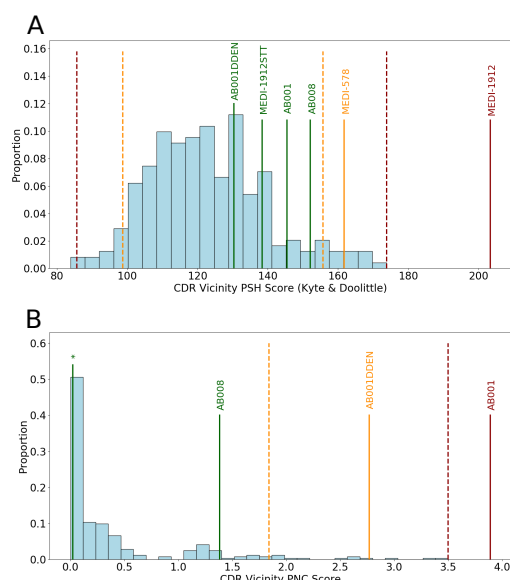
| Metric | Amber Region                        | Red Region               |
|--------|-------------------------------------|--------------------------|
| 1.     | $54 \leq L < 60$                    | $L > 60$                 |
| 2.     | $83.84 \leq \text{PSH} < 100.71$    | $\text{PSH} < 83.84$     |
|        | $156.20 < \text{PSH} \leq 173.85$   | $\text{PSH} > 173.85$    |
| 3.     | $1.25 \leq \text{PPC} < 3.16$       | $\text{PPC} > 3.16$      |
| 4.     | $1.84 \leq \text{PNC} < 3.50$       | $\text{PNC} > 3.50$      |
| 5.     | $-20.40 \leq \text{SFvCSP} < -6.30$ | $\text{SFvCSP} < -20.40$ |

**Table 2.** TAP amber and red flag regions, as defined by the entire set of 242 CSTs. PSH score is calculated with the Kyte & Doolittle hydrophobicity scale (29). Metric numbers refer to Table 1. L = Length.

All surface residues were initially assigned the appropriate charge for their averaged  $pK_a$  values, as neighboring residues appear to have a limited effect at pH 7.4 (4). The charge of residues found to be engaging in salt bridges was then revised to 0.

The 137 CSTs tend to avoid patches of charge in their CDR vicinities, with 88.32% and 80.30% of them having PPC (Fig. 3A) and PNC (Fig. 3B) values below 1, respectively. Human Ig-seq models displayed similar PPC and PNC distributions, with the top 5% of PPC values being noticeably smaller than the top 5% of PNC values. Both PPC and PNC assays were carried forward as TAP guideline metrics.

When mAbs have oppositely charged  $V_H$  and  $V_L$  chains, they typically have higher *in vitro* viscosity values (4). This aggregate-inducing electrostatic attraction is captured at the sequence level by the Fv Charge Symmetry Parameter (FvCSP) metric - the mAb tends to be more viscous if the product of net  $V_H$  and  $V_L$  charges is negative (4). Harnessing our structural models, we calculated a variant (the Structural Fv Charge Symmetry Parameter, SFvCSP), which only includes residues that are surface-exposed, and not locked in salt bridges, in the evaluation of net charge. In galiximab, for example, we ‘correct’ the charge of arginine H108 and aspartic acid L56 to 0, as the model indicates that they form a salt bridge. The charges of the glutamic acid at position H6, the aspartic acids at positions H107, L98, and L108, and the histidine at position L40 are ignored as their side chains are buried. The FvCSP score for this antibody would be 0 (net heavy chain charge of 0, net light chain charge of -2.9), whilst the SFvCSP score is +2.0 (net heavy chain charge of +2, net light chain charge of +1). A similarly low percentage of CST models (21.9%) and human Ig-seq models (19.7%) had negative SFvCSP scores (Fig. 4C), with mean values of  $3.34 \pm 7.44$  and  $2.52 \pm 5.54$  respectively. With such a bias away from negative products, we chose the SFvCSP as our final TAP guideline property.

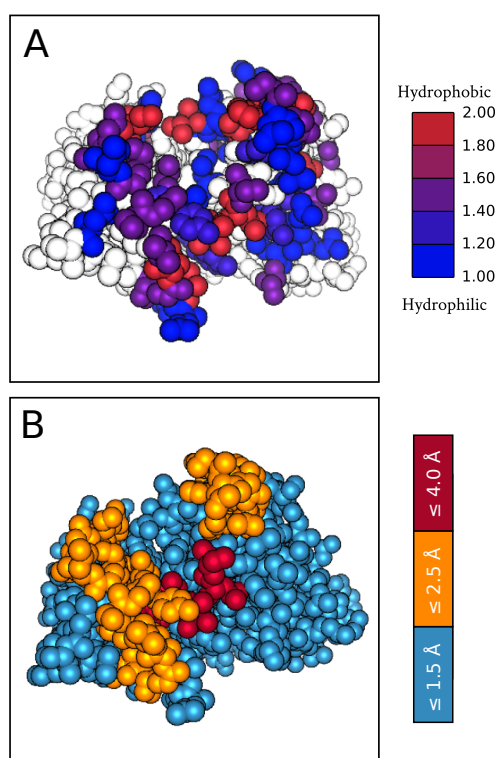


**Fig. 4.** The (A) CDR vicinity PSH and (B) CDR vicinity PNC metrics for the combined set of 242 CSTs (light blue), and MedImmune case studies (colored by assigned flag). MEDI-578, MEDI-1912, and MEDI-1912SST all have the CDR vicinity PNC value labeled by an asterisk. Amber and red dashed lines delineate the 242 CST guideline thresholds. Case studies with prohibitive developability issues (MEDI-1912, AB001) are red-flagged for the PSH and PNC metrics respectively. Engineered versions without developability issues (MEDI-1912SST, AB001DDEN) return to the range of values previously seen in CSTs for all metrics.

**Developability Guidelines.** While CSTs predictably share many features in common with human antibodies, our CDR length and hydrophobicity distributions imply that not every human antibody would make a good therapeutic. Consequently, our developability guidelines were set solely by CST values across the five selected metrics (Table 1). Amber flags indicate that the antibody lies within the extremes of the distributions, and should be interpreted as ‘be aware’. Red flags indicate a previously-unobserved value for that property, and should be interpreted as ‘cause for concern’.

To confirm that these threshold definitions do not typically flag mAbs without developability issues, we identified a further 105 whole antibody therapies (“105 CSTs”, listed in Dataset S1), not included in the 137 CST dataset, that had advanced to at least Phase II in clinical development.

Only eight of this set (7.69%) were assigned a red developability flag according to the boundaries set by the 137 CSTs, an average of 0.08 red flags per newly-tested therapeutic (Table S3). Erenumab received the most red flags - for total CDR length (60), CDR vicinity PSH (173.85), and CDR vicinity PPC (1.53). All other red-flagged therapeutics received only one: ravidurumab for total CDR length (60), intetumumab for CDR PSH (83.84), adacanumab, derlotumumab, lanadelumab and teprotumumab for CDR PPC (2.67, 2.66, 2.48, and 3.16 respectively), and quilzumab for Fv charge asymmetry (-20.40). The low red-flagging rate confirms that these guideline characteristics are highly conserved across therapeutic-like antibodies. Incorporating both sets of CSTs into a larger dataset (“242 CSTs”) led to the new guideline values shown in Table 2. While this had a large effect on the PPC thresholds, all other metrics were only slightly adjusted.



**Fig. 5.** (A) An example TAP web application output showing the heavy atoms of an antibody as spheres colored by the hydrophobicity (Kyte & Doolittle scale, normalized between 1 and 2) of each residue in the CDR vicinity. (B) The ABodyBuilder predicted model accuracy assignments (19) for each IMGT region, with heavy atoms shown as spheres. These are colored according to three backbone RMSD thresholds at a 75% confidence interval (both thresholds and confidence intervals can be modified in the web application). Better quality models will yield more reliable TAP metric values.

**Case Studies.** We tested whether these updated guideline values could highlight candidates with developability problems by running TAP on two datasets supplied by MedImmune (Fig. 4). A lead anti-NGF antibody, MEDI-578, showed minor aggregation issues during *in vitro* testing, of a level usually rectifiable in development, whereas the affinity matured version, MEDI-1912, exhibited unrectifiably high levels of aggregation (34). This observation was rationalized through SAP score (6) values, indicating that a large hydrophobic patch on the surface was responsible. TAP assigns MEDI-578 an amber flag, and MEDI-1912 a red flag - by a large margin - in the CDR vicinity PSH metric (Fig. 4A). Back-mutation of three hydrophobic residues in MEDI-1912 to those of MEDI-578 led to MEDI-1912STT, fixing the aggregation issue while maintaining potency. TAP assigns MEDI-1912STT no developability flags (Fig. 4A).

A lead anti-IL13 candidate, AB008, had no developability issues, but the affinity-matured version, AB001, had very poor levels of expression (seven times lower than AB008) (11). The authors highlighted the role of four consecutive negatively charged residues in the L2 loop - mutation of the fourth negatively charged residue to neutral asparagine (AB001DDEN) was able to stabilize the loop backbone, mitigating the ionic repulsion of the DDE motif, and returning acceptable levels of expression. TAP assigns no developability flags to AB008, but a red flag to AB001, and an amber flag

to AB001DDEN for its CDR vicinity PNC metric (Fig. 4B), again red-flagging the candidate with prohibitive developability issues. Both AB001 and AB008, confirmed monomers in solution (11), did not flag for CDR vicinity PSH score (Fig. 4A).

**Web Application.** We have packaged the Therapeutic Antibody Profiler (TAP) into a web application, available at <http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/TAP.php>. TAP only requires the heavy and light chain variable domain sequences as an input, returning a detailed profile of your antibody with a typical runtime of less than 30 seconds. Flags (green, amber, or red) are assigned to each of the TAP metrics, with accompanying histograms. An interactive molecular viewer allows the user to visualize hydrophobicity (Fig. 5A), charge, and probable sequence liabilities on the antibody model surface. Estimated model quality can be easily accessed to help guide interpretation of the results (Fig. 5B). Finally, canonical forms are assigned to each non-CDRH3 loop. A full sample output is shown in Fig. S6.

## Discussion

By analyzing characteristics linked to poor developability, we have found evidence that suggests that not every human antibody would make a good therapeutic. This would be somewhat intuitive, as therapeutics suffer a range of stresses during development (including variation in pH and temperature, sheer forces, and high concentration storage conditions) that human-expressed antibodies are not exposed to. The TAP metrics therefore depend on the values seen across CSTs alone.

Our simple TAP guidelines will not capture the whole spectrum of developability issues. For example, they will not detect sources of immunogenicity, nor more subtle mechanisms that lead to poor stability. Nevertheless, we have shown that the TAP guidelines can selectively highlight antibodies with expression or aggregation issues (11, 34). The choice of 5<sup>th</sup>/95<sup>th</sup> percentile amber flag cut-offs proved useful in distinguishing weak aggregators from non-aggregators (Fig. 4A). We intend to recalculate the threshold values on a monthly basis. This will take into account new mAbs that enter Phase II of clinical trials. It will also allow for the inevitable fluctuation in PSH, PPC, PNC, and SFvCSP values returned by every CST, as ABodyBuilder models will improve as the number of antibodies in the PDB increases (35). When sufficient numbers of mAbs have reached the market, we could increase the reliability of the TAP thresholds by only considering FDA/EU approved therapies, rather than all mAbs that have reached Phase II.

The thresholds themselves should not be interpreted as hard-and-fast rules, and the distance of red-flagged candidates outside the previously-observed bounds should be taken into consideration. Advances in process development and formulation may soon redefine the limits of permissible values (18).

## Methods

All modeling procedures are described in *SI Methods*.

**Human Ig-seq Data.** Procured by UCB Pharma, this dataset contains 4,587,907 non-redundant heavy and 7,120,100 non-redundant light chains. The number of non-redundant CDR sequences is as follows: 174,490 CDRH1s, 279,873 CDRH2s, 1,696,918 CDRH3s, 455,125 CDRL1s, 8,708 CDRL2s, and 980,158 CDRL3s. For further information, including the chain pairing protocol, see *SI Methods*.

**Clinical-Stage Therapeutics.** The initial set of 137 clinical-stage therapeutic (CST) antibody sequences were sourced from the Supporting Information of Jain *et al.* (18). The test set of 105 CST sequences was found through an extensive search of online resources, including the IMGT mAb [https://www.imgt.org/mAb-DB/] and Antibody Society [https://www.antibodysociety.org/late-stage-clinical-pipeline/] databases. The names and sequences of each CST are supplied in Dataset S1, with PDB structures (where available) listed at <http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Therapeutic.html>. All 242 CST ABody-Builder models are available on our group website, at <http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/TAP.php>.

**Canonical Forms.** A length-independent canonical form clustering protocol (36) was run on the North-defined (37) CDR loops of a SABDab (35) snapshot from 26th September 2017. Model loops were inferred to have identical canonical forms to the template used by ABodyBuilder (19).

**Surface-Exposed Residues.** Residues are classed as surface-exposed if they have > 7.5% relative exposure (22) across side chain atoms, compared to the open-chain form Alanine-R-Alanine, as calculated with the Shrake & Rupley algorithm (21).

**CDR Vicinity.** The “CDR vicinity” comprises every surface-exposed IMGT-defined CDR and anchor residue, and all other surface-exposed residues with a heavy atom within a 4 Å radius.

**Salt Bridges.** Salt bridges were defined as pairs of lysines/arginines and aspartic acids/glutamic acids with a N<sup>+</sup>-O<sup>-</sup> distance ≤ 3.2 Å.

**Hydrophobicity.** Where R<sub>1</sub> and R<sub>2</sub> are two surface-exposed residues with a closest heavy atom distance, r<sub>12</sub>, < 7.5 Å and H(R,S) is the normalized hydrophobicity score (between 1 and 2) for residue R in scheme S, the Patches of Surface Hydrophobicity (PSH) metric can be calculated as:

$$\sum_{R_1 R_2} \frac{H(R_1, S) H(R_2, S)}{r_{12}^2}$$

The hydrophobicity scales tested were: Kyte & Doolittle (29), Wimley & White (30), Hessa *et al.* (31), Eisenberg & McLachlan (32) and Black & Mould (33). Salt bridge residues were assigned the same value as glycine in each hydrophobicity scale.

**Charge.** The following charges were assigned by sequence: Aspartic acid: -1, Glutamic acid: -1, Lysine: +1, Arginine: +1, Histidine: +0.1 (Henderson-Hasselbalch equation applied: pK<sub>a</sub> 6, pH 7.4, and rounded-up to one decimal place). Tyrosine hydroxyl deprotonation was not considered. Salt bridge residues were assigned a charge of 0. The Patches of Positive Charge (PPC) and Patches of Negative Charge (PNC) metrics are analogous in form to PSH, with H(R,S) substituted for |Q(R)|, the absolute value of the charge assigned to residue R. Structural Fv Charge Symmetry Parameter (SFvCSP) values were calculated as:

$$\left[ \sum_{R_H} Q(R_H) \right] \left[ \sum_{R_L} Q(R_L) \right]$$

where R<sub>H</sub>, R<sub>L</sub> are surface-exposed V<sub>H</sub>, V<sub>L</sub> residues respectively.

## ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) and the Medical Research Council (MRC) [grant number EP/L016044/1], GlaxoSmithKline plc, MedImmune Limited, F. Hoffmann-La Roche AG, and UCB Celltech. We thank Sebastian Kelm and James Heads for their helpful comments concerning our metrics. MIJR is grateful to Jinwoo Leem for his assistance in designing and implementing the web application.

## Bibliography

1. Approved antibodies, 2018. Published by The Antibody Society. Available to members at: <https://www.antibodysociety.org/news/approved-antibodies/> [Accessed June 12, 2018].
2. Alexander Jarasch, Hans Koll, Joerg T. Regula, Martin Bader, Apollon Papadimitriou, and Hubert Kettenberger. Developability assessment during the selection of novel therapeutic antibodies. *J Pharm Sci*, 104(6):1885–1898, 2015.
3. Yingda Xu, William Roach, Tingwan Sun, Tushar Jain, Bianka Prinz, Ta-Yi Yu, Joshua Torrey, Jerry Thomas, Piotr Bobrowicz, Maximiliano Vásquez, K. Dane Wittrup, and Eric Krauland. Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: A FACS-based, high-throughput selection and analytical tool. *Protein Eng Des and Sel*, 26(10):663–670, 2013.
4. Vikas K. Sharma, Thomas W. Patapoff, Bruce Kabakoff, Eric Hilario, Boyan Zhang, Charlene Li, Oleg Borisov, Robert F. Kelley, Ilya Chorny, Joe Z. Zhou, and Ken A. Dill. *In silico* selection of therapeutic antibodies for development: Viscosity, clearance, and chemical stability. *Proc Natl Acad Sci USA*, 111(52):18601–18606, 2014.
5. Nareesh Chennamsetty, Vladimir Voynov, Veyssel Kayser, Bernhard Helk, and Bernhardt L. Trout. Design of therapeutic proteins with enhanced stability. *Proc Natl Acad Sci USA*, 106(29):11937–11942, 2009.
6. Timothy M. Lauer, Neeraj J. Agrawal, Nareesh Chennamsetty, Kamal Egodage, Bernhard Helk, and Bernhardt L. Trout. Developability index: A rapid in silico tool for the screening of antibody aggregation propensity. *J Pharm Sci*, 101(1):102–115, 2012.
7. Tushar Jain, Todd Boland, Asparouh Lilov, Irina Burnina, Michael Brown, Yingda Xu, and Maximiliano Vásquez. Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning. *Bioinformatics*, 33:3758–3766, 2017.
8. Olga Obrezanova, Andreas Arnell, Ramón Gómez De La Cuesta, Maud E. Berthelot, Thomas R. A. Gallagher, Jesús Zurdo, and Yvette Stallwood. Aggregation risk prediction for antibodies and its application to biotherapeutic development. *mAbs*, 7(2):352–363, 2015.
9. Sandeep Yadav, Thomas M. Laue, Devendra S. Kalonia, Shubhadra N. Singh, and Steven J. Shire. The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. *Mol Pharm*, 9(4):791–802, 2012.
10. Amita Datta-Mannan, Arunkumar Thangaraju, Donniemne Leung, Ying Tang, Derrick R. Witcher, Jirong Lu, and Victor J. Wroblewski. Balancing charge in the complementarity-determining regions of humanized mAbs without affecting pI reduces non-specific binding and improves the pharmacokinetics. *mAbs*, 7(3):483–493, 2015.
11. Bojana Popovic, Suzanne Gibson, Tarik Senussi, Sara Carmen, Sara Kidd, Tim Slidel, Ian Strickland, Jianqing Xu, Jennifer Spooner, Amanda Lewis, Nathan Hudson, Lorna Mackenzie, Jennifer Keen, Ben Kemp, Colin Hardman, Diane Hatton, Trevor Wilkinson, Tristan Vaughan, and David Lowe. Engineering the expression of an anti-interleukin-13 antibody through rational design and mutagenesis. *Protein Eng Des Sel*, 30(4):303–311, 2017.
12. Markus Habberger, Katrin Bomans, Katharina Diepold, Michaela Hook, Jana Gassner, Tilman Schlothauer, Adrian Zwick, Christian Spick, Jochen Felix Kepert, Brigitte Hienz, Michael Wiedmann, Hermann Beck, Philipp Metzger, Michael Mølhoj, Constanze Knoblich, Ulla Grauschopf, Dietmar Reusch, and Patrick Bulau. Assessment of chemical modifications of sites in the CDRs of recombinant antibodies. *mAbs*, 6(2):327–339, 2014.
13. Jasmin F. Sydow, Florian Lipsmeier, Vincent Larraillet, Maximiliane Hilger, Bjoern Mautz, Michael Mølhoj, Jan Kuentzer, Stefan Klostermann, Juergen Schoch, Hans R. Voelger, Joerg T. Regula, Patrick Cramer, Apollon Papadimitriou, and Hubert Kettenberger. Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions. *PLoS ONE*, 9(6):e100736, 2014.

14. Andrei J. Petrescu, Adina L. Milac, Stefana M. Petrescu, Raymond A. Dwek, and Mark R. Wormald. Statistical analysis of the protein environment of N-glycosylation sites: Implications for occupancy, structure, and folding. *Glycobiology*, 14(2):103–114, 2004.
15. Fabienne Courtois, Neeraj J. Agrawal, Timothy M. Lauer, and Bernhardt L. Trout. Rational design of therapeutic mAbs against aggregation through protein engineering and incorporation of glycosylation motifs applied to bevacizumab. *mAbs*, 8(1):99–112, 2016.
16. Juan C. Almagro, Alexey Teplyakov, Jinqian Luo, Raymond W. Sweet, Sreekumar Kodan-gattil, Francisco Hernandez-Guzman, and Gary L. Gilliland. Second antibody modeling assessment (AMA-II). *Proteins*, 82(8):1553–1562, 2014.
17. Christopher A. Lipinski, Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev*, 23(1-3):3–25, 1997.
18. Tushar Jain, Tingwan Sun, Stéphanie Durand, Amy Hall, Nga Rewa Houston, Juergen H. Nett, Beth Sharkey, Beata Bobrowicz, Isabelle Caffry, Yao Yu, Yuan Cao, Heather Lynaugh, Michael Brown, Hemanta Baruah, Laura T. Gray, Eric M. Krauland, Yingda Xu, Maximiliano Vásquez, and K. Dane Wittrup. Biophysical properties of the clinical-stage antibody landscape. *Proc Natl Acad Sci USA*, 114(5):944–949, 2017.
19. Jinwoo Leem, James Dunbar, Guy Georges, Jiye Shi, and Charlotte M. Deane. ABODY-Builder: Automated antibody structure prediction with data-driven accuracy estimation. *mAbs*, 8(7):1259–1268, 2016.
20. Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Phillip E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–242, 2000.
21. A. Shrake and J. A. Rupley. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol*, 79(2):361–371, 1973.
22. Zhan Yang Zhu and Tom L. Blundell. The use of amino acid patterns of classified helices and strands in secondary structure prediction. *J Mol Biol*, 260(2):261–276, 1996.
23. Yoko Tsuchiya and Kenji Mizuguchi. The diversity of H3 loops determines the antigen-binding tendencies of antibody CDR loops. *Prot Sci*, 25(4):815–825, 2016.
24. Marie-Paule Lefranc, Christelle Pommé, Manuel Ruiz, Véronique Giuducelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, and Gérard Lefranc. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev and Comp Immunol*, 27(1):55–77, 2003.
25. Bin Shi, Long Ma, Xiaoyan He, Xiaomei Wang, Peng Wang, Li Zhou, and Xinsheng Yao. Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT / LIGM-DB with IMGT / HighV-QUEST. *Theor Biol Med Model*, 11:30, 2014.
26. Cyrus Chothia and Arthur M. Lesk. Canonical Structures for the Hypervariable Regions of Immunoglobulins. *J Mol Biol*, 196(4):901–917, 1987.
27. Cyrus Chothia, Arthur M. Lesk, Anna Tramontano, Michael Levitt, Sandra J. Smith-Gill, Gillian Air, Steven Sherriff, Eduardo A. Padlan, David Davies, William R. Tulip, Peter M. Colman, Silvia Spinelli, Pedro M. Alzari, and Roberto J. Poljak. Conformations of immunoglobulin hypervariable regions. *Nature*, 342(6252):877–883, 1989.
28. Jacqueline A. Reynolds, David B. Gilbert, and Charles Tanford. Empirical Correlation Between Hydrophobic Free Energy and Aqueous Cavity Surface Area. *Proc Natl Acad Sci USA*, 71(8):2925–2927, 1974.
29. Jack Kyte and Russell F. Doolittle. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157(1):105–132, 1982.
30. William C. Wimley and Stephen H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol*, 3(10):842–848, 1996.
31. Tara Hessa, Hyun Kim, Karl Bihlmaier, Carolina Lundin, Jorrit Boekel, Helena Andersson, IngMarie Nilsson, Stephen H. White, and Gunnar von Heijne. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433(7024):377–381, 2005.
32. David Eisenberg and Andrew D. McLachlan. Solvation energy in protein folding and binding. *Nature*, 319(6050):199–203, 1986.
33. Shaun D. Black and Diane R. Mould. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal Biochem*, 193(1):72–82, 1991.
34. Claire L. Dobson, Paul W.A. Devine, Jonathan J. Phillips, Daniel R. Higazi, Christopher Lloyd, Bojana Popovic, Joanne Arnold, Andrew Buchanan, Arthur Lewis, Joanne Goodman, Christopher F. Van Der Walle, Peter Thornton, Lisa Vinall, David Lowne, Anna Aagaard, Lise Lotte Olsson, Anna Ridderstad Wollberg, Fraser Welsh, Theodoros K. Karamanos, Clare L. Pashley, Matthew G. Iadanza, Neil A. Ranson, Alison E. Ashcroft, Alistair D. Kippen, Tristan J. Vaughan, Sheena E. Radford, and David C. Lowe. Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Sci Rep*, 6(38644):1–14, 2016.
35. James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M. Deane. SABDab: The structural antibody database. *Nucleic Acids Res*, 42(D1):1140–1146, 2014.
36. Jaroslaw Nowak, Terry Baker, Guy Georges, Sebastian Kelm, Stefan Klostermann, Jiye Shi, Sudharsan Sridharan, and Charlotte M. Deane. Length-independent structural similarities enrich the antibody CDR canonical class model. *mAbs*, 8(4):751–760, 2016.
37. Benjamin North, Andreas Lehmann, and Roland L. Dunbrack Jr. A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology*, 406(2):228–256, 2011.