nature portfolio

Peer Review File

Biophysical cartography of the native and human-engineered antibody landscapes quantifies the plasticity of antibody developability

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

This paper represents a huge contribution and certainly an impressive tour-de-force of Ab characterization, analysis and systems-level model building and rationalization of data.

I think this paper is of great interest for the immunology community, and deserves to be published as is.
It my case, it does not always happen to see such interesting pieces of work.
I have no comments that would add to the paper.

Reviewer #2 (Remarks to the Author):

Bashour & Smorodina et al. have made a computational study comparing several developability parameters predicted from primary sequence and structure and compare human engineered antibodies with natural ones. One interesting outcome of their study is that human engineered antibodies only explore a little of the sequence/structure space of the natural ones. Another strength of this study, is the high number of antibody (scFv) sequences that are part of this study, as also outlined by the authors, most other studies rely on computational analysis of 100-1000 of antibodies. I believe that this study has broad interest and would suit publication in Communication Biology. It is relevant for an active scientific field to start bridging studies with high number of sequence data with the experimental studies available, and this study act as a good contribution to the first part. Though to ensure that the study reach broad interest, i.e. ensuring that also experimentalists would read it, I believe that the authors have to make a more detailed description of their selected DP's. I acknowledge that they have a good descriptive table in SI, but I think they should discuss it further in the manuscript. As an example, why have they decided to included extinction coefficient which is not that interesting, while they omit e.g. TAP parameters? Beside this major comment, then I only have minor comments for the authors to consider, not necessarily to follow.

Minor comments to be addressed/considered:
- Considering the 40 and 46 DP's predicted based on sequence and structure, respectively, some of the parameters seem more interesting and validated than others. It could perhaps be interesting with a version of Fig 2, that only include the parameters which the study suggest are most relevant (on their own and in combination) for developability assessment.
- Chemical degradation hotspots in CDR's, such as DG sites could be very interesting to also have addressed in this dataset. I.e. with the requirements from regulatories, it could be that the human-engineered antibodies has significantly less of such sites in CDR's to avoid CQA's, in comparison to the natural repetoire?
- The authors write: "We also found that the values of DPs measured on the conformational ensembles of antibodies evolve throughout their molecular dynamics (MD)". This is expected, but still very interesting and highlight a challenge for structure based predictions. Would be interesting if the authors could speculate further on, how this could be addressed in future studies.
- The authors write: "Furthermore, many studies have focused on extracting developability guidelines from a limited number of successful mAbs, considering them as a "ground truth" of desired developability (7, 10, 17, 29, 52)". I find the "ground truth" statement as unnecessary, and for some of the cited studies, I also believe it is an unfair statement from the authors.
- Overall figures are very nice, both visually and content wise. Some figures have too small text, which makes it difficult for the reader to really digest it, e.g. figure 2C as example. Perhaps Fig 2C could be an individual figure or part of SI.
- The authors write: "Notably, none of the tested DPs displayed high average sensitivity (median excess kurtosis << 0, Figure 4A, Supp. Figure 12), suggesting that developability is relatively stable to the average single amino acid mutation with few outliers. Nevertheless, since DP values were normalized, small relative shifts induced by a mutation may still have a large effect in practice.". There are many examples of single mutations having radical effects on especially self-association/viscosity and non-specific binding. Perhaps the authors could elute to which "real world" developability issues that they expect can be predicted from the selected DP's?

- Interesting how single mutations had higher effect on charge and hydrophobicity predictors for FR1, FR2, FR3 compared to CDR1 and CDR2. Perhaps the authors could elute to this, and what the reason for this could be?

Reviewer #3 (Remarks to the Author):

In this work, the authors have used sequence and structure-based developability measures to distinguish between human-engineered and natural antibodies. One of the significant contributions of the manuscript is applying 86 in-silico developability measures (higher than previous papers) to over 2 million natural and human engineered antibody sequences. The other contribution is a ML based prediction tool for all developability measures. The curated large dataset and the predicted developability measure can be useful for multifactor-optimized antibody design.

Below, I have listed my comments regarding the manuscript.
Minor:
1. Page 6: paragraph 3, spell check for 'stricture'.
2. I highly recommend increasing the resolution of the figures. The labels are not clear at all.
3. Page 15, paragraph 2, last line: How could the conclusion be drawn that the developability profile is independent of CDR3 sequence similarity only? It seems like the correlation is low across all regions.
4. Page 25: How could Figure 7b help conclude that DP profiling agreed with PLM-based profiling?

Major comments:
1. Authors found that structure-based developability parameters correlate poorly. Do the authors have any thoughts about the observation? Is it because the context of the antigen is not taken.
2. One of the developability parameters was steric clash. Were the structures predicted by ABB2 used for calculating the developability parameters relaxed?
3. Having supplementary Table 1 in the main manuscript would help readability.
4. Figure 6b. Why is the error bar so high? Are the values reliable, given that the error bars exceed their mean values?

Point-by-point response for the manuscript ***Biophysical cartography of the native and human-engineered antibody landscapes quantifies the plasticity of antibody developability.***

In the point-by-point response document, the reviewers' comments are in black and our responses are in blue. Revision-based changes to the manuscript are indicated by page and line numbers in this document. In the revised manuscript, the modified manuscript text has been yellow-highlighted.

The comments by Reviewer 2 (R2) and Reviewer 3 (R3) are referenced as RX.Y, where Y is the comment number. For example, R2.3 refers to Reviewer 2's 3rd comment. Reviewer 1 had no comments to address.

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

This paper represents a huge contribution and certainly an impressive tour-de-force of Ab characterization, analysis and systems-level model building and rationalization of data. I think this paper is of great interest for the immunology community, and deserves to be published as is. In my case, it does not always happen to see such interesting pieces of work. I have no comments that would add to the paper.

We are very humbled and grateful for these remarks made by Reviewer 1. We thank them for taking the time to review our work.

Reviewer #2 (Remarks to the Author):

Bashour & Smorodina et al. have made a computational study comparing several developability parameters predicted from primary sequence and structure and compare human engineered antibodies with natural ones. One interesting outcome of their study is that human-engineered antibodies only explore a little of the sequence/structure space of the natural ones. Another strength of this study, is the high number of antibody (scFv) sequences that are part of this study, as also outlined by the authors, most other studies rely on computational analysis of 100-1000 of antibodies. I believe that this study has broad interest and would suit publication in Communication Biology. It is relevant for an active scientific field to start bridging studies with high number of sequence data with the experimental studies available, and this study act as a good contribution to the first part.

We thank Reviewer 2 and appreciate the acknowledgment of the relevance and importance of our study.

R2.1. Though to ensure that the study reach broad interest, i.e. ensuring that also experimentalists would read it, I believe that the authors have to make a more detailed description of their selected DP's. I acknowledge that they have a good descriptive table in SI, but I think they should discuss it further in the manuscript.

We thank the reviewer for this constructive comment. We agree that including more descriptions to the DP table would improve the readability and reach of this paper to a wider audience. We have now added further descriptive text to the table and moved the table from SI to the main body of the manuscript (Table 1). Furthermore, we have added a new paragraph within the result section where we clarify the rationale for the selection and the categorization of the DPs in the main text ("Developability parameters ", page 6).

R2.2. As an example, why have they decided to included extinction coefficient which is not that interesting, while they omit e.g. TAP parameters?

We thank R2 for making this important point. The inclusion of the extinction coefficient was following the work described by [1] whose study suggested that including the extinction coefficient parameter in an ML-based model assisted the prediction of antibody developability.

We agree with R2 that the TAP paper and parameters are an important concept and directly relevant to our study. Regarding the use of TAP and/or its parameters:

- TAP parameters were not fully omitted in principle. For instance, we included structure-based positive and negative charge heterogeneity and sequence-based hydrophobicity measures measured on the full Fv region in our study.
- Several observations made throughout this manuscript highlighted a certain level of alignment with observations made previously and recently by TAP, including (i) the localization of Erenumab in what might be a "developability risk zone" in the upper cluster of Figure 7B, and (ii) the limited dissimilarity between IgK and IgL light chain antibodies in regards to developability (Supp Figure 11A).
- Technically, we were unable to fully integrate the TAP software within this work as it requires paired antibodies as input. In this study, most analyses were performed on unpaired antibody data, being the most abundant and accessible native repertoire Fv sequences.

Actions taken:

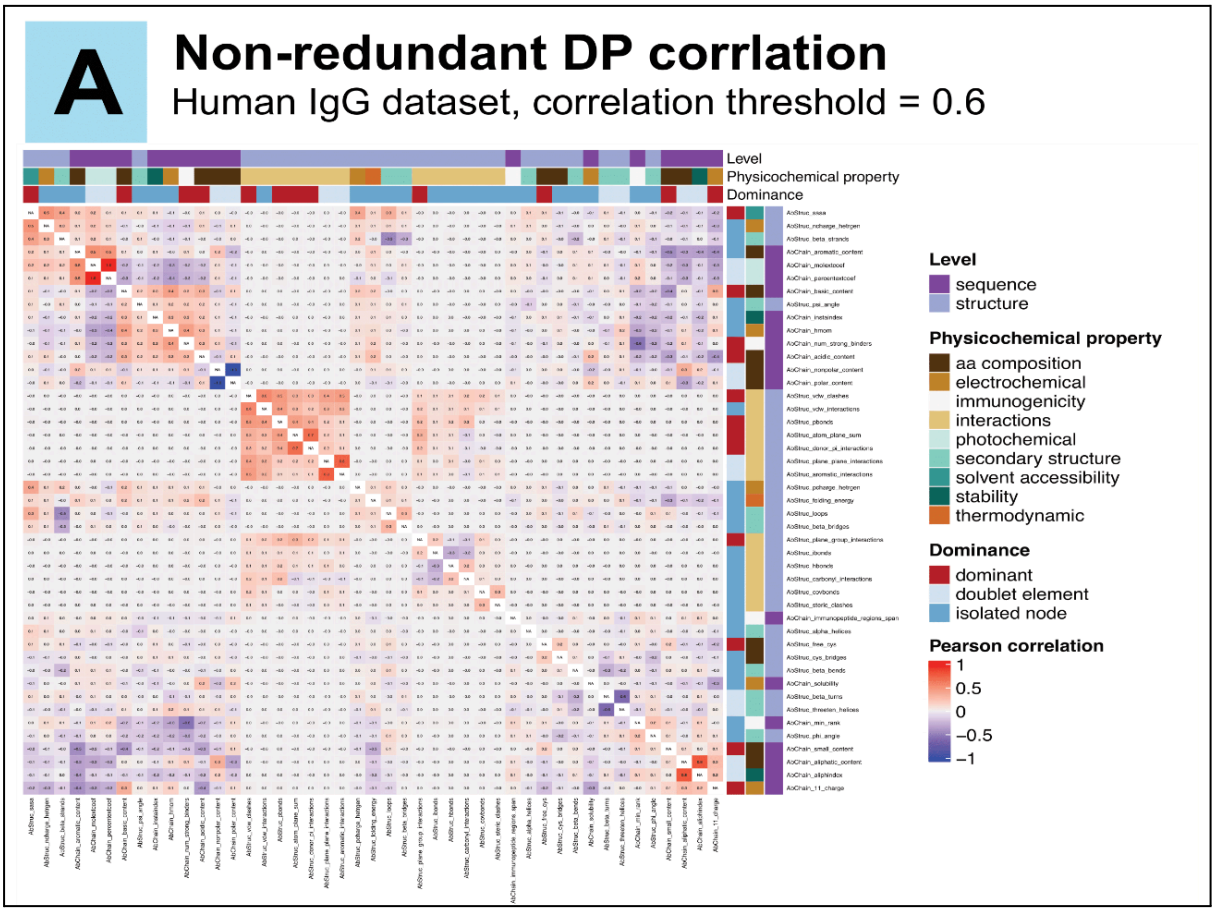
- We added a part of text highlighting the inclusion of some TAP developability parameters within our study in page 6 under the section "Developability parameters".
- We made the points of alignment mentioned above between our findings and the TAP studies more prominent within our result section by adding a few pieces of text in page 27.

Beside this major comment, then I only have minor comments for the authors to consider, not necessarily to follow.

Minor comments to be addressed/considered:

R2.3. Considering the 40 and 46 DP's predicted based on sequence and structure, respectively, some of the parameters seem more interesting and validated than others. It could perhaps be interesting with a version of Fig 2, that only includes the parameters, which the study suggests are most relevant (on their own and in combination) for developability assessment.

We thank R2 for this comment. We have now made this figure and added it to Supp. Figure 2 as Supp. Supp. Figure 2A.



Supp. Figure 2A. Pairwise correlation among non-redundant DPs as selected by the ABC-EDA algorithm for an absolute Pearson correlation threshold of 0.6, starting from the correlation heatmap presented in Figure 2B. Of note, doublet DPs were all retained in this analysis as no meaningful selection based only on correlation can be made by the algorithm.

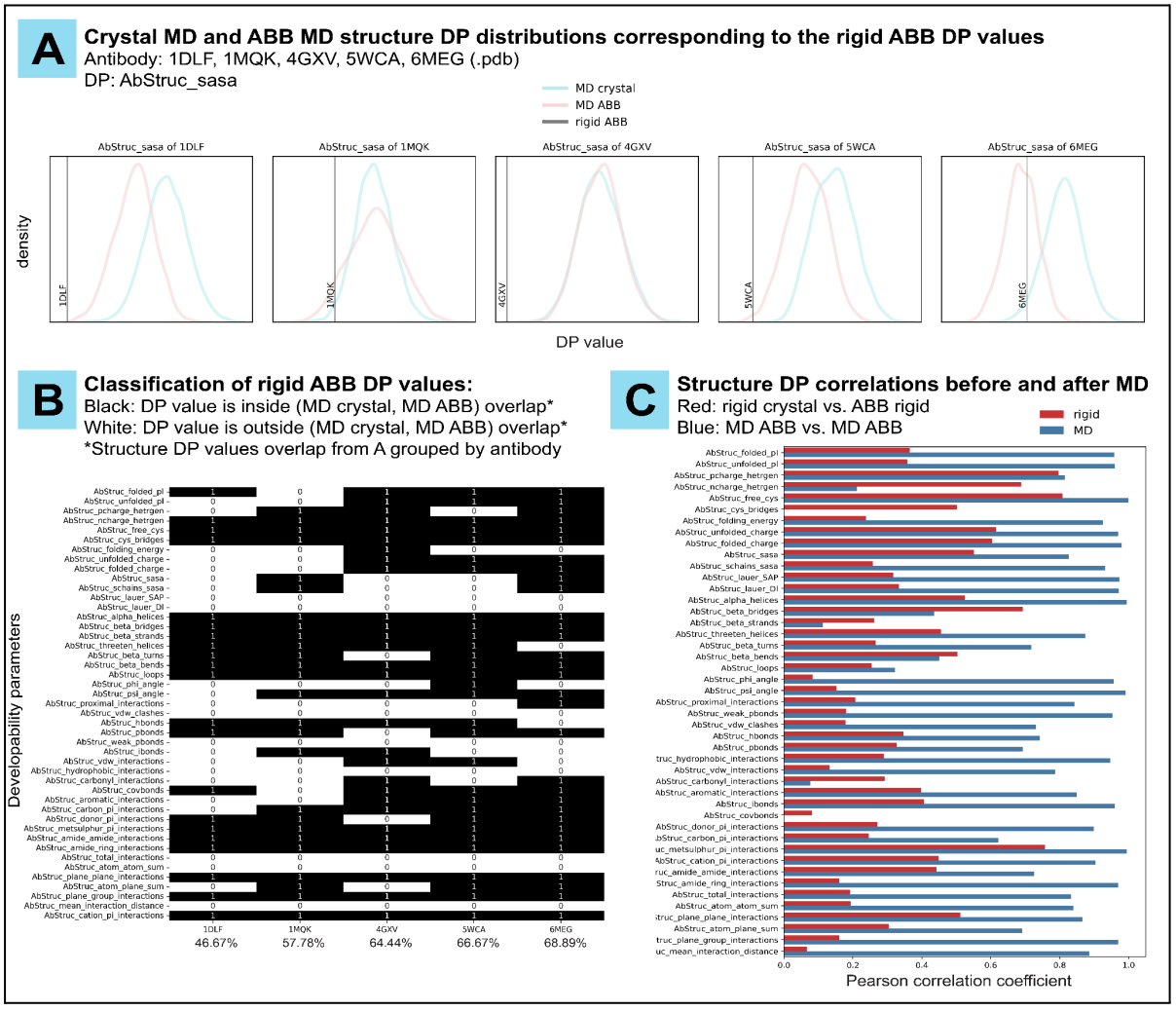
R2.4. Chemical degradation hotspots in CDR's, such as DG sites could be very interesting to also have addressed in this dataset. I.e. with the requirements from regulatories, it could be that the human-engineered antibodies has significantly less of such sites in CDR's to avoid CQA's, in comparison to the natural repertoire?

We thank the reviewer for raising this point, and we would like to refer them to Supp Figure 13A. In this analysis, we studied the representation of numerous liability motifs, including DG isomerisation, as described by [2] among native and human engineered antibodies. We reported no consistent increased or decreased representation (trend) of these motifs among the different datasets.

R2.5. The authors write: "We also found that the values of DPs measured on the conformational ensembles of antibodies evolve throughout their molecular dynamics (MD)". This is expected, but still very interesting and highlight a challenge for structure based predictions. Would be interesting if the authors could speculate further on how this could be addressed in future studies.

We thank the reviewer for this important note. We would like to refer them to the Supplemental File; subheadings "*(2) Molecular dynamics analysis*" (presented with the original submission) and "(3)

Rigid model vs. Molecular dynamics analysis" (presented with the revised submission) where this point is fully expanded on (pages 29–33 in Supplementary Information). We also added the following text and Supp. Figure 20 (see below) to further investigate how challenges in structure-based developability predictions can be addressed.



Supp. Figure 20 | ABodyBuilder (ABB) models reflect the real-world structures in terms of DP values in 60% of the case (on average), and molecular dynamics (MD) improves poor correlations between structure DPs caused by the used structure prediction tool. (A) The positioning of DP values measured on the rigid ABB structures for each antibody (PDB IDs: 1DLF, 1MQK, 4GXV, 5WCA, 6MEG) in relation to the overlap between DP value distribution when measured on crystal MD and ABB MD ensembles for the structure-based solvent accessible surface area DP (AbStruc_sasa). The corresponding rigid DP values are shown as a black vertical line. Each subplot represents a separate antibody system. (B) Binary classification: 1 (black) indicates rigid ABB structure DP value is within the overlap of crystal MD and ABB MD structure DP distributions; 0 (white) – otherwise. Numbers shown at the bottom of each column represent the proportion of black cells for each antibody (PDB IDs: 1DLF, 1MQK, 4GXV, 5WCA, 6MEG). (C) Improvement in correlations of DP values across ABB and crystal data between the values from the rigid (red) and dynamic (blue) structures.

"To improve the correlation between structure DP across different sources of antibody structures (ABB models and crystal data), we computed the DP values for each converged conformation from MD simulations of five selected antibodies (1DLF, 1MQK, 4GXV, 5WCA, 6MEG). Next, we calculated the average of these values (one value per system) for each structure DP parameter. We

correlated averaged ABB MD with averaged crystal MD (blue in Supp. Figure 19C), and compared that with the correlation of rigid ABB with rigid crustal (red bars in Supp. Figure 19C). We observed a substantial improvement in correlations of DP across ABB and crystal data between the values from the rigid (red) and dynamic (blue) structures (rigid DP value range: 0.07–0.80, mean: 0.36; MD DP value range: 0.00–1.00, mean: 0.76) (Supplementary File, sections "(3) Rigid model vs. Molecular dynamics analysis:", Supp. Figure 20C) [...] Our findings are in line with recent MD-driven findings in regards to structure-based DPs [3–5] where a correlation between DP values measured on experimental and predicted structures was reported only after averaging the DPs values of the two conformational ensembles."

R2.6. The authors write: "Furthermore, many studies have focused on extracting developability guidelines from a limited number of successful mAbs, considering them as a "ground truth" of desired developability (7, 10, 17, 29, 52)". I find the "ground truth" statement as unnecessary, and for some of the cited studies, I also believe it is an unfair statement from the authors.

We agree with R2. We have now replaced the term "ground truth" with the term "gold standard" in the text (page 3).

R2.7. Overall figures are very nice, both visually and content wise. Some figures have too small text, which makes it difficult for the reader to really digest it, e.g. figure 2C as example. Perhaps Fig 2C could be an individual figure or part of SI.

We thank the reviewer for their comment. We have now moved this figure to Supp figure XYZB to make it easier for the reader to digest. Additionally, high quality versions of all figures are deposited on the project GitHub repository (link: https://github.com/csi-greifflab/developability_profiling/tree/main/figures).

R2.8. The authors write: "Notably, none of the tested DPs displayed high average sensitivity (median excess kurtosis << 0, Figure 4A, Supp. Figure 12), suggesting that developability is relatively stable to the average single amino acid mutation with few outliers. Nevertheless, since DP values were normalized, small relative shifts induced by a mutation may still have a large effect in practice.". There are many examples of single mutations having radical effects on especially self-association/viscosity and non-specific binding. Perhaps the authors could elute to which "real world" developability issues that they expect can be predicted from the selected DP's?

We thank the reviewer for correctly pointing out that single mutations can have radical effects on antibody properties (e.g., [6]). Although we use single amino acid substituted mutants in this experiment, our sensitivity metrics are based aggregate scores across all sampled antibodies and their possible variants. Here, the sensitivity of normalized DP values can be used for guidance to identify parameters where single amino acid substitutions are expected to result in largely similar or more diverging parameter values. We have added a clarification of the uses and limits to the discussion section on page 32.

R2.9. Interesting how single mutations had higher effect on charge and hydrophobicity predictors for FR1, FR2, FR3 compared to CDR1 and CDR2. Perhaps the authors could elute to this, and what the reason for this could be?

We thank the reviewer for perceptively pointing out the stark difference in potential sensitivity between framework regions 1–3 (heavy chain) and CDRs 1–2 (heavy chain) when observing charge and hydrophobicity parameters. The potential sensitivities or ranges depend on the most charged/polar residue in a given region. A potential explanation could be, since FRs are longer than CDR1 and 2, there is a higher probability that FRs contain at least one maximally charged/polar residue while CDRs might be missing them resulting in a lower possible overall range of either charge or hydrophobicity. We have added a hypothesis to the results section on page 18.

Reviewer #3 (Remarks to the Author):

In this work, the authors have used sequence and structure-based developability measures to distinguish between human-engineered and natural antibodies. One of the significant contributions of the manuscript is applying 86 in-silico developability measures (higher than previous papers) to over 2 million natural and human engineered antibody sequences. The other contribution is a ML based prediction tool for all developability measures. The curated large dataset and the predicted developability measure can be useful for multifactor-optimized antibody design. Below, I have listed my comments regarding the manuscript.

Minor:

R3.1. Page 6: paragraph 3, spell check for ‘stricture’.

We thank the reviewer for pointing out this spelling mistake, this has now been rectified.

R3.2. I highly recommend increasing the resolution of the figures. The labels are not clear at all.

We thank the reviewer for their comment. High-resolution figures are deposited on the project GitHub (link: https://github.com/csi-greifflab/developability_profiling/tree/main/figures) and were uploaded separately to the Communications Biology platform. In this revised version, the reviewer should be able to inspect the high-resolution figures appended to the revised manuscript version.

R3.3. Page 15, paragraph 2, last line: How could the conclusion be drawn that the developability profile is independent of CDR3 sequence similarity only? It seems like the correlation is low across all regions.

We thank the reviewer for highlighting the lack of clarity in this conclusion. In this analysis, we sampled 5000 Fv sequences that belong to the same sequence similarity group (Group 1 – Supp. Figure 10B) and share the same IGHV gene family annotation. As Fv sequences belonging to the same IGHV gene family share high (up to 80%) sequence similarity [7,8], most of which is attributed to conserved sequences in the frameworks, CDRH1 and CDRH2 regions [9], the majority of sequence variation is credited to the CDRH3. This means that within our sampled antibodies, the higher the sequence similarity (or lower Normalized Levenshtein distance) it is, the more similar their CDRH3 are in sequence. As such, the lack of correlation between the pairwise Normalised Levenshtein distance and the pairwise Euclidean distance in the developability space for these antibodies (shown in Supp. Figure 10C) indicates an independence of developability profile and CDRH3 sequence similarity among antibodies.

We have now clarified this statement in the manuscript and updated the text on page 20 accordingly.

R3.4. Page 25: How could Figure 7b help conclude that DP profiling agreed with PLM-based profiling?

Thank you for your comment. We acknowledge that we should clarify that we base our conclusion regarding the agreement between DP profiling and PLM-based profiling on the comparison of the PC projections depicted in Figure 7B alongside Supplementary Figures 10B (for VH) and 11B (for VL).

Our analysis reveals that the spatial distributions of human-engineered and naturally occurring antibodies in these figures are notably similar, indicating congruent profiling by both DP and PLM methods. Both figures illustrate in fact that while human-engineered antibodies predominantly occupy the same regions as natural antibodies, they tend to cluster in specific areas rather than dispersing uniformly throughout the space.

We have updated the manuscript accordingly (page 29).

Major comments:

R3.5. Authors found that structure-based developability parameters correlate poorly.

(1) Do the authors have any thoughts about the observation?
(2) Is it because the context of the antigen is not taken?

We thank the reviewer for these important questions.

(1) We reported that the values of structure-based DPs calculated on different antibody structure prediction tools correlate poorly in Supp Figure 17A (Supplementary File section).

    Although proteins (including antibodies) are flexible and dynamic, high-throughput structure prediction tools only return a single rigid structure with only one set of many possible side-chain conformations. We believe that this is the main factor behind this poor correlation. Indeed, as we show in Supp. Figure 20C, our findings suggest that the correlation of structure-based DPs increases when using more conformations sampled from structural ensembles simulated by molecular dynamics (MD).

(2) Using antigen-complexed structures will stabilize the antibody fluctuations, resulting in a conformational ensemble distribution that is narrower compared to the uncomplexed antibody structure. However, it would mostly affect the antigen-binding loop domains, keeping the dynamics of other domains of the structure similar to the free (uncomplexed) antibody. We, therefore, only expect a negligible increase in DP correlation when using antigen-complexed over uncomplexed antibody structures.

R3.6. One of the developability parameters was steric clash. Were the structures predicted by ABB2 used for calculating the developability parameters relaxed?

The ABB2 pipeline includes a relaxation step after the initial antibody structure prediction. We clarified it in the revised version of the manuscript.

R3.7. Having supplementary Table 1 in the main manuscript would help readability.

We thank the reviewer for this suggestion, which echoes that of R2.1. We have now moved Supplementary Table 1 to page 6 of the main manuscript and labeled it as Table 1.

R3.8. Figure 6b. Why is the error bar so high? Are the values reliable, given that the error bars exceed their mean values?

We thank the reviewer for this comment. As stated in the last paragraph of page 23, the aim of this experiment was not to directly optimize DP prediction but to estimate the amount of relevant information contained in PLM- and DP-based antibody representations. Since the regression models predict each individual DP from all other DPs, the performance varies considerably among models for a given training set size. Although the uncertainty in prediction accuracy is indeed very high for structure-based DPs, our comparison highlights that PLM-based representations almost always outperform pure DP-based representations when predicting sequence-based DPs by linear regression and that neither performs especially well on structure-based DPs.

References

1. Chen, X. *et al.* Predicting Antibody Developability from Sequence using Machine Learning. *biorxiv* 2–8 (2020).
2. Teixeira, A. A. R. *et al.* Simultaneous affinity maturation and developability enhancement using natural liability-free CDRs. *MAbs* **14**, 2115200 (2022).
3. Park, E. & Izadi, S. Molecular Surface Descriptors to Predict Antibody Developability. *bioRxiv* 2023.07.18.549448 (2023) doi:10.1101/2023.07.18.549448.
4. Licari, G. *et al.* Embedding Dynamics in Intrinsic Physicochemical Profiles of Market-Stage Antibody-Based Biotherapeutics. *Mol. Pharm.* (2022) doi:10.1021/acs.molpharmaceut.2c00838.
5. Raybould, M. I. J., Turnbull, O. M., Suter, A., Guloglu, B. & Deane, C. M. Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling. *bioRxiv* 2023.06.28.546839 (2023) doi:10.1101/2023.06.28.546839.
6. Mason, D. M. *et al.* Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat Biomed Eng* **5**, 600–612 (2021).
7. Giudicelli, V. & Lefranc, M. P. Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* **15**, 1047–1054 (1999).
8. Peres, A. *et al.* IGHV allele similarity clustering improves genotype inference from adaptive

immune receptor repertoire sequencing data. *Nucleic Acids Res.* **51**, e86 (2023).

9. Barroso, R., Morrison, W. I. & Morrison, L. J. Molecular Dissection of the Antibody Response: Opportunities and Needs for Application in Cattle. *Front. Immunol.* **11**, 1175 (2020).

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The reviewed version is an improvement over the already excellent first one. The paper is worth publishing

Reviewer #2 (Remarks to the Author):

Thank you for good and interesting responses to the points I raised. All my points were addressed.

Reviewer #3 (Remarks to the Author):

I do not have any further questions. The authors have answered all my questions.