Prediction of Antibody Non-Specificity using Protein Language Models and Biophysical Parameters

Laila I. Sakhnini[1,3,*], Ludovica Beltrame[2], Simone Fulle[2], Pietro Sormanni[3], Anette Henriksen[1], Nikolai Lorenzen[1], Michele Vendruscolo[3,*], Daniele Granata[2,*]

[1]Therapeutics Discovery, Novo Nordisk A/S, Copenhagen, Denmark
[2]Digital Chemistry and Design, Novo Nordisk A/S, Copenhagen, Denmark
[3]Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, UK

[*]Corresponding authors:
L.I. Sakhnini (llsh@novonordisk.com), M. Vendruscolo (mv245@cam.ac.uk), D. Granata (dngt@novonordisk.com)

Abstract

The development of therapeutic antibodies requires optimizing target binding affinity and pharmacodynamics, while ensuring high developability potential, including minimizing non-specific binding. In this study, we address this problem by predicting antibody non-specificity by two complementary approaches: (i) antibody sequence embeddings by protein language models (PLMs), and (ii) a comprehensive set of sequence-based biophysical descriptors. These models were trained on human and mouse antibody data from Boughter *et al.* (2020) and tested on three public datasets: Jain *et al.* (2017), Shehata *et al.* (2019) and Harvey *et al*. (2022). We show that non-specificity is best predicted from the heavy variable domain and heavy-chain complementary variable regions (CDRs). The top performing PLM, a heavy variable domain-based ESM 1v LogisticReg model, resulted in 10-fold cross-validation accuracy of up to 71%. Our biophysical descriptor-based analysis identified the isoelectric point as a key driver of non-specificity. Our findings underscore the importance of biophysical properties in predicting antibody non-specificity and highlight the potential of protein language models for the development of antibody-based therapeutics. To illustrate the use of our approach in the development of lead candidates with high developability potential, we show that it can be extended to therapeutic antibodies and nanobodies.

Keywords: Therapeutic antibodies, non-specificity, protein-language models, machine learning, isoelectric point

1. Introduction

Monoclonal antibodies (mAbs) continue to be one of the leading drug modalities in the pharmaceutical industry, with more than 100 unique mAbs approved by the FDA since 2021 [1] and global sales forecasted to 300 billion US dollars by 2025 [2,3]. The success of mAbs for therapeutic application is the result of advances in *in vivo* and *in vitro* discovery platforms, which have enabled fast generation of high-affinity binders towards a highly diverse set of targets [4,5]. Recently, *de novo* design has gained increasing interest in the field as a third-generation discovery approach with the potential of significantly accelerating drug discovery and development timelines [6,7]. Moreover, with the advances in mAb engineering, there has also been an increased interest in the development of mAbs with ultra-high target affinity (pM to fM) [8], context-dependent target binding (e.g. pH-dependent [9,10] or ligand induced target binding [11]), and various multi-specific functionalities [12]. To reach optimal binding affinity and potency, mAb hits identified during discovery are often subjected to comprehensive screening campaigns using display platforms (libraries of 10^3 to 10^{10} variants) of and recombinant well-plate variant generation workflows (libraries of 10^2 to 10^4 variants) [13].

When selecting an antibody lead candidate for development towards clinical testing, optimal target binding affinity and/or pharmacodynamics are key selection parameters. In addition, in the last decade there has been an increased focus on the importance of progressing antibodies to clinical stage which also possess a good developability potential. Antibody developability requires the intersection of multiple disciplines, in which diverse parameters such as expression levels, immunogenicity, processability, and formulation feasibility are addressed, to ensure optimal potential for successful clinical development of lead candidates. Non-specific binding, i.e. weak non-covalent interactions with off-target molecules or interfaces, has emerged as one of the key developability parameters to increase the chance for clinical success [14,15]. Specifically, several studies reported that high tendency for non-specific binding can translate into faster *in vivo* clearance, thereby compromising pharmacokinetics [16,17,18,19,20,21,22]. Furthermore, there is an inherent risk that non-specific interactions can translate into undesirable side-effects [23]. Non-specific binding is not a rare phenomenon, as recent reports suggest the presence of a trade-off between affinity and specificity [24,25,26,27]. Thus, optimization of affinity and potency comes with an inherent risk of compromising target specificity [28,29,30,31].

Given the high level of interest in measuring non-specific interactions, there are several *in vitro* screening assays available for this purpose. A commonly used assay is the Enzyme-Linked Immunosorbent Assay (ELISA) with a panel of common antigens, typically insulin, DNA, albumin, cardiolipin and lipopolysaccharide (LPS) [31,15]. Initially, these biomolecules were studied as model antigens for autoimmune responses and diseases. For example, insulin is a self-antigen for autoantibodies associated with type 1 diabetes [32,33]. In addition, DNA, albumin and cardiolipin are self-antigens for autoantibodies associated with several diseases such as systemic lupus erythematosus [34,35] and anti-phospholipid syndrome [36], and LPS is an antigen for immune responses to bacterial infections [37]. Moreover, ELISA is widely used in immunology, where non-

specific antibodies are often referred to as poly-reactive antibodies. Such antibodies are characterised as having low-affinity binding to multiple distinct antigens, including self-antigens, and they have been widely studied for targets such as HIV and Influenza viruses, as they can be broadly neutralizing [38,39]. While this feature is beneficial for immunity to infectious diseases, to potentially confer broad protection against viruses, it is a highly undesirable feature for therapeutic mAbs. Other common non-specificity assays include baculovirus particle (BVP) ELISA [40], poly-specific reagent (PSR) [41], and cross-interaction chromatography with ligands such as heparin [42] and human IgG from serum [43].

In addition to these *in vitro* assays, *in silico* methods have been gaining interest, and several tools have been reported recently for prediction of non-specific interactions [44,45,46,47]. The development and implementation of predictive computational methods for the prediction and re-design of monoclonal antibody non-specificity at an early stage is of great interest, as it facilitates the generation of safe and efficacious lead candidates with high developability potential. Several studies reported on the identification of non-specificity by *in silico* approaches. Short, linear sequence motifs (e.g. GG, RR, VG, VV, YY, WW and WxW, where x can be any amino acid) enriched in non-specific antibodies, as reported by Kelly *et al.* [48], have been utilized to create synthetic antibody libraries free from such motifs in the CDRs [49]. Moreover, AI/ML models that classifies non-specific antibodies, leveraging experimental data and sequence-based information, have been reported. Boughter *et al.* [44] developed a classifier to identify non-specific antibodies based on experimental data acquired from ELISA with a panel of common antigens. Harvey *et al.* [45] developed a one-hot LogisticReg model based on a naïve Nb library assessed by the PSR assay.

In this study, we developed machine learning (ML) models to estimate the non-specificity of antibodies (**Figure 1A**). Commonly used biophysical properties were tested alongside protein language models (PLMs) to embed antibody sequences. PLMs have emerged as powerful tools for extracting informative features from raw protein sequences by leveraging patterns learned from massive sequence databases [50,51,52,53]. Among these, Evolutionary Scale Modeling (ESM) models have shown particular promise in capturing structural, functional, and physico-chemical properties (including antibody specificity) without requiring explicit structural data [54,55,56,57]. ESM models, such as ESM-1v, encode sequences into high-dimensional embeddings that reflect residue context, conservation, and evolutionary information, which are all factors known to influence antibody behaviour. These features make ESM models well-suited for predicting complex properties like non-specificity, which can arise from subtle sequence-dependent effects not easily captured by traditional descriptors. By applying ESM models to antibody variable regions, we aim to harness its representation power to identify sequence signatures of non-specific binding and improve early-stage developability assessment.

Besides testing which encoding provided the best prediction performance, one aspect of the study was to identify which part of the antibody contributes to non-specificity. Furthermore, to gain biophysical insight, sequence-based biophysical descriptors were analysed to support the predictive models. Our results indicate that the computational models that we considered enable the prediction

of non-specific interactions that can be used to guide the design and selection of antibodies with improved specificity and efficacy.

2. Results & Discussion

2.1 Public antibody data

Four different datasets were retrieved from public sources; (i) a curated dataset of >1000 mouse IgA, influenza-reactive and HIV-reactive antibodies with their respective non-specificity flag from ELISA with a panel of common antigens [44], (ii) 137 clinical-stage IgG1-formatted antibodies with their respective non-specificity flag from ELISA with a panel of common antigens [15], (iii) 398 antibodies, originating from naïve, IgG memory and long-lived plasma cells, with their respective poly-specific reagent score [58], and (iv) 140 000 nanobody (Nb) clones assessed by the PSR assay from a naïve Nb library [45]. These four datasets are referred to as the Boughter, the Jain, the Shehata and the Harvey datasets, respectively.

As therapeutic antibodies are engineered to be closely related to human antibodies to avoid immunogenic responses, it is important to exploit human antibody data for development of optimal ML models. As the Boughter dataset partly consists of mouse IgA antibodies, the sequence similarity of these mouse antibodies was compared to the human antibodies to ensure that there are not too large sequence differences within the dataset. The mouse IgA antibodies appear to differ mostly in the H/L-CDRs (**Figure S1B**). Another notable difference within the Boughter dataset is that the mouse IgA antibodies have a slightly shorter H-CDR3 loop relative to the human antibodies (**Figure S1C**).

Ideally, training data sets for classification ML-Models should be balanced when it comes to positive and negative data points. The distribution of non-specificity for the three datasets is visualised in **Figure 1B**. The Boughter and the Jain datasets are relatively balanced in terms of specific (zero flags), mildly non-specific (1-3 flags) and non-specific (>4 flags) antibodies, while the Shehata dataset is unbalanced, with 7 out of 398 antibodies characterised as non-specific only. In this study, the most balanced dataset (i.e. Boughter one) was selected for training of ML models, while the remining three (i.e. Jain, Shehata and Harvey, which consists exclusively of VHH sequences) were used for testing.

2.2 Protein language models enable the representation of antibody non-specificity

Following the study original study [44], the Boughter dataset was first parsed into two groups: specific (0 flags) and non-specific group (>3 flags), leaving out the mildly non-specific antibodies (1-3 flags) (**Figure 1A**). The amino acid sequences of the parsed dataset were then annotated in the CDRs, and various fragments of the antibody sequences were embedded into vectors representing their physico-chemical and structural properties (i.e. ESM 1b, ESM 1v, ESM 2, Protbert bfd, AntiBERTy, and AbLang2). This procedure resulted into the training of 12 different antibody fragment-specific binary classification models were trained (see **Table 4**). Overall, all of the protein language models (PLMs) performed well with 66-71% 10-fold CV accuracy, including the antibody-specific ones AntiBERTy and AbLang2 **(Figure 1C**). These deep learning models were trained on

large datasets of protein sequences in the million-to-billion range, encoding protein sequences into vectors for representation of their physiological properties, remote homology, and secondary/tertiary structure. PLMs were originally developed for the prediction of protein contacts and structure [81, 82]. Going forward, the Evolutionary Scale Modelling (ESM) 1v was selected as the embedder of choice for this study.

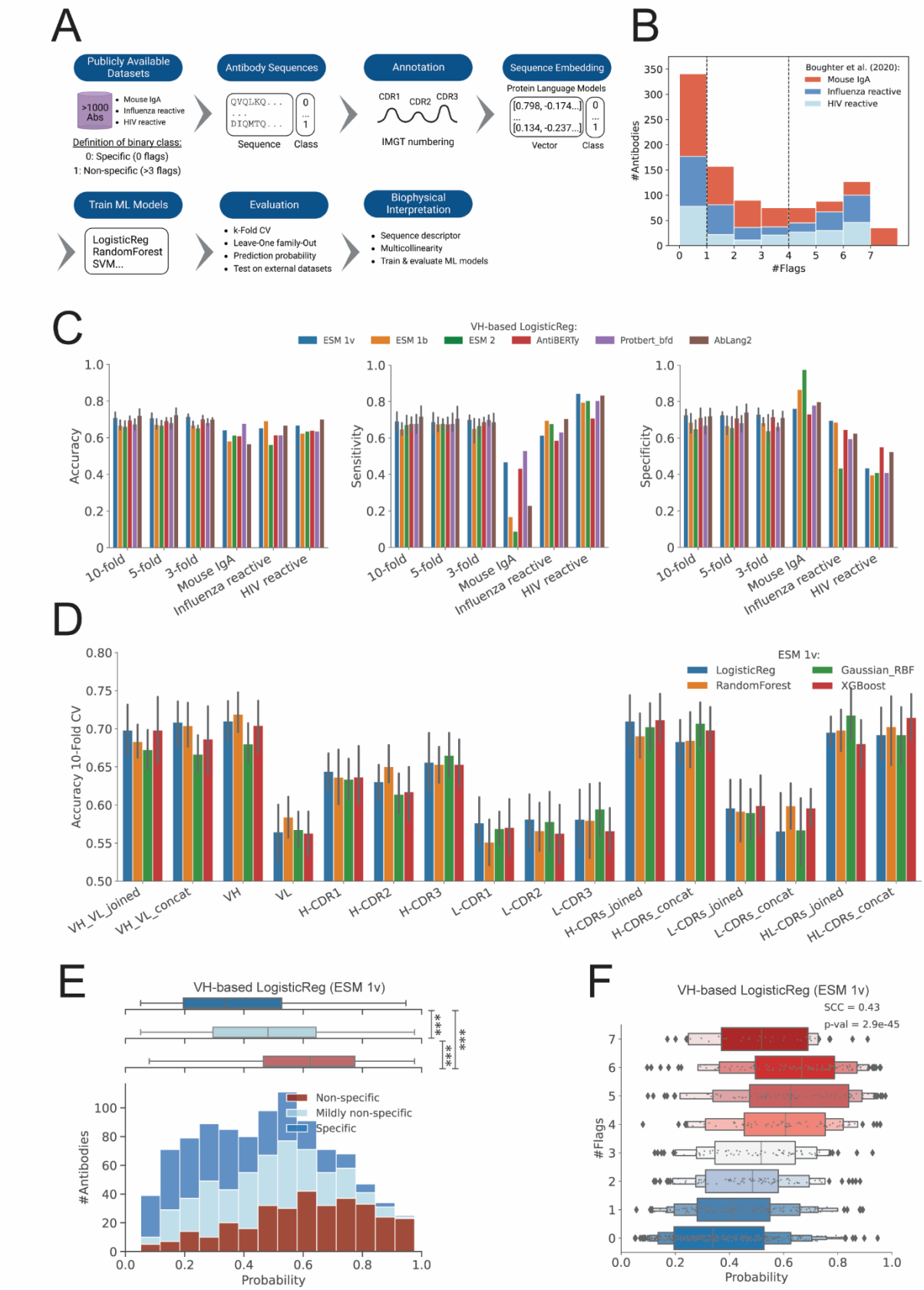

Figure 1. Performance Evaluation of Machine Learning Models for Predicting Antibody Non-Specificity. (A) Schematic workflow of the study. Publicly available datasets containing antibody sequences were used. These sequences were annotated and

embedded using sequence-based biophysical descriptors and protein language models (PLMs). Different ML models were trained and evaluated using k-fold cross-validation, sensitivity-specificity analysis, and external datasets. **(B)** Histogram showing the distribution of antibody sequences based on the number of flags in the Boughter dataset. Sequences are categorized as Influenza reactive (blue), HIV non-reactive (red), and HIV reactive (light blue). **(C)** Bar plots showing the validation performance (k-Fold CV and Leave-One Family-Out) for a top performing ML algorithm (LogisticReg) and PLMs across various validation schemes (k-fold CV and leave-one-family-out). **(D)** Bar plot of 10-fold CV accuracy for different antibody sequences embedded by top performing language model (ESM 1v mean-mode). **(E)** Histogram of predicted probabilities of antibody non-specificity using the VH-based Logistic Regression (ESM 1v) model for the Jain dataset. Antibodies are classified into non-specific (dark blue), mildly non-specific (light blue), and specific (red) categories. **(F)** Boxplot comparing the predicted non-specificity probabilities for antibodies across different datasets using VH-based Logistic Regression (ESM 1v). The boxplot displays the median, interquartile range, and outliers, with significant differences indicated by SCC and p-values (*** indicate p-value <0.001).

2.3 The highest PLM-based predictability is achieved by encoding the VH domain

An overview of the different antibody fragment-specific models based on the ESM 1v embedder is shown in **Figure 1D**. Highest predictability (71% 10-fold CV accuracy of non-specificity) was obtained for the models trained on the VH and H-CDRs sequences. These results suggest that the non-specificity primarily originates from the VH domain, with main contributions from the H-CDR loops. When looking at the models based on the individual H-CDR loops, the order of low-to-high predictability of non-specificity follows H-CDR2, H-CDR1 and H-CDR3 (**Figure 1D**). H-CDR3 has the highest predictability of non-specificity among all the H-CDR loops. The importance of the H-CDR3 loop for non-specificity is in agreement with Guthmiller *et al.* [59], who showed by using MD simulation that the flexibility of the H-CDR3 loop plays an important role in the non-specific behaviour of antibodies.

Accuracy of around 70% was consistently observed across 3, 5 and 10-Fold CV for the top performing models (**Figure 1C**), and similar performance obtained for sensitivity and specificity. Moreover, when looking at the predictability of one antibody family to another, the overall accuracy was consistently above >60% for the Leave-One Family-Out validations. However, when comparing sensitivity and specificity, classifiers trained on human antibodies performed poorly when tested on mouse antibodies. This is not surprising as mouse and human antibodies have notable sequence differences, such as mouse IgA having a shorter H-CDR3 and larger sequence differences in the CDRs relative to human antibodies (**Figure S1B**). Moreover, classifiers trained on mouse IgA and HIV reactive antibodies perform well across all evaluation metrics (accuracy, sensitivity and specificity) when tested on Influenza reactive antibodies, while classifiers trained on mouse IgA and Influenza reactive antibodies seem to be better in predicting non-specific HIV reactive antibodies than specific ones.

2.4 Classification probability of non-specificity against non-specificity ELISA flags mimics regression behaviour

A prediction probability of non-specificity was computed in addition to the binary output from the binary classification models. When comparing the prediction probability of non-specificity for all the antibodies from the Boughter dataset (test antibodies sampled from the 10-Fold CV and the mildly non-specific antibodies), three distinct distributions of prediction probabilities for the specific, mildly non-specific and non-specific antibody groups appeared (**Figure 1E**). The premise that non-specificity is not a binary property is exemplified by the overlaps between the distributions. This

illustrates that the prediction probability can be used beyond the binary output to assess antibodies of varying degree of non-specificity.

When comparing those to non-specificity ELISA flags, a significant regression-like behaviour (SCC 0.43) was observed for one of the top performing classifiers, ESM 1v mean-mode VH-based LogisticReg model (**Figure 1F).** The prediction probability for non-specificity followed an uptrend when compared to the non-specificity ELISA flags. An exception to this trend was the antibodies with seven non-specificity ELISA flags, as those were exclusively mouse IgA antibodies (differences discussed in previous section).

2.5 The isoelectric point: a key biophysical driver of non-specificity

To gain insight into the biophysical origins of antibody non-specificity, a set of 68 sequence descriptors (**Table S1**) was computed for the parsed Boughter dataset. These descriptors encompass a wide range of biophysical properties derived from the antibodies sequence, including theoretical isoelectric point (pI), secondary structure propensity, and hydrophobicity. To assess the presence of redundancy among the descriptors, we constructed a Spearman's correlation matrix, which revealed that several descriptors, such as hydrophobicity descriptors, exhibited strong correlation among each other (SCC > 0.5, **Figure S9**), thus indicating redundancy. All the descriptors were ranked according to the absolute logistic regression coefficients (**Table S2**), whereafter top 25 descriptors were selected, and used for training of a VH-based LogisticReg model. The in-depth analysis of the importance of the 25 descriptors is shown in four different plots in **Figure 2A**:

- The first plot shows the LogisticReg coefficients, indicating the relative importance of each descriptor. Notably, Disorder_Propensity_DisProt, Aggrescan_a4v, and theoretical pI show significant positive coefficients, suggesting that they are strong drivers of non-specificity.
- The second plot displays the permutation importance, highlighting the change in accuracy when each descriptor is permuted. Descriptors like theoretical pI, bulkiness, Hplc_Hfba_retention and Polarity_Zimmerman demonstrate substantial decrease in accuracy upon permutation.
- The third plot illustrates the accuracy of models based on single descriptors to underscore their individual predictive power. Theoretical pI resulted in the highest accuracy compared to the other descriptors, confirming its critical role in the prediction of non-specificity.
- The fourth plot shows the leave-one-feature-out accuracy, revealing how the exclusion of each descriptor affects the overall model performance. Most of the descriptors result in a minimal drop in accuracy, indicating that the model performance remains unaffected when a certain descriptor is left out. This can be explained by that there remains a certain level of redundancy among the 25 descriptors, e.g. theoretical pI appears to be negatively correlated with Polarity_Zimmerman, according to the Spearman's correlation matrix in **Figure 2B**.

Table 1. Top 2, 3, 4 and 5 combined VH-based sequence descriptors.

Top	Descriptors

2	Theoretical pI, Bulkiness
3	Theoretical pI, Bulkiness, Aggrescan_Nr_hotspots
4	Theoretical pI, Average_Flexibility_BP, Beta_Turn_Chou_Fasman, Hplc_Hfba_Retention
5	Theoretical pI, Bulkiness, Disorder_Propensity_DisProt, Percentage_Acessible_Res, Aggrescan_av4

To further narrow down the redundancy, the 25 descriptors were tested in all possible combinations of 2, 3, 4 and 5 descriptors for training of new LogisticReg models. The results of the top descriptors from this analysis are shown in

Table 1. The results indicate that, among the top 5 descriptors, the theoretical pI appear to be the most important driver for non-specificity. This conclusion is also supported by the frequency of this particular descriptor among the top models (**Figure S10**). Additionally, as a parallel check, we performed Principal Component Analysis (PCA) for dimensionality reduction and feature selection. The primary objective of this study was again to address multicollinearity among the descriptors and to identify the most significant features contributing to the variance in the dataset. We thus evaluated the performance of the LogisticReg models trained on the top 3, 5, and 10 principal components identified by PCA. In agreement with our previous findings regarding the theoretical pI, the presence of this descriptor among the selected features significantly influenced the model performance, particularly when it was included in the top 5 and 10 components, but it was absent in the top 3 components (**Figure S19**). The isoelectric point is known to influence PKPD behaviour/clearance of antibodies [60,61].

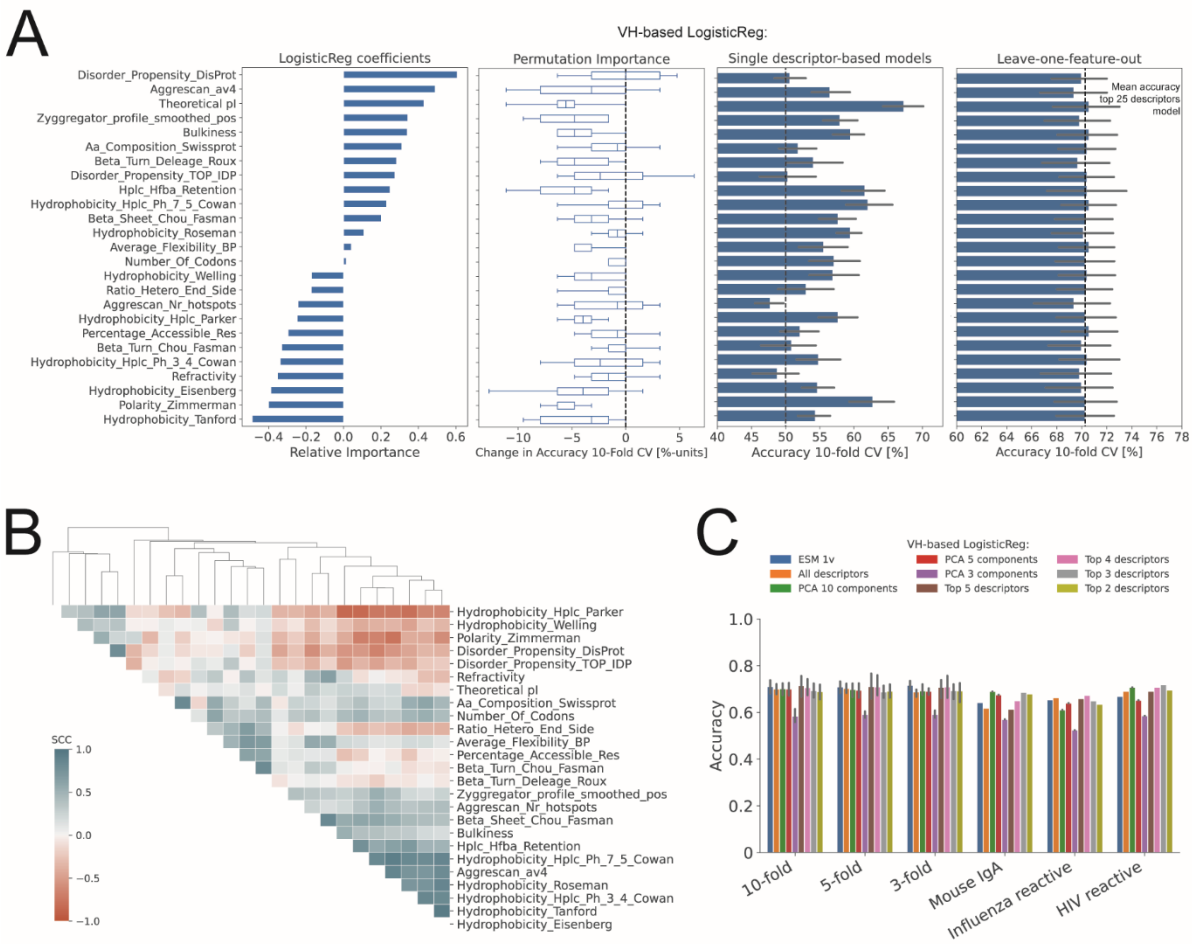

Figure 2. Analysis of Descriptor Importance and Model Performance for VH-Based Logistic Regression. (A) Analysis of descriptor importance using various metrics for the VH-based Logistic Regression model; (first panel) Logistic regression coefficients indicating the relative importance of different features, (second panel) permutation importance showing the change in 10-fold CV accuracy when each descriptor is permuted, (third panel) 10-fold CV accuracy of models based on single descriptors, and (fourth panel) 10-fold CV accuracy of leave-one-feature-out models compared to the mean accuracy of model using the top 25 descriptors. **(B)** Heatmap displaying the Spearman's correlation coefficient (SCC) between the top 25 descriptors selected based on highest Logistic Regression coefficient in model with all descriptors. The dendrogram shows hierarchical clustering of descriptors based on their SCC. **(C)** Bar plot comparing the 10-fold CV accuracy of different models in predicting antibody non-specificity across various validation schemes: k-fold CV and leave-one-family-out validation. Models compared include VH-based sequences embedded by ESM 1v, all descriptors, PCA with 3, 5 and 10 components, and top 2, 3, 4, and 5 descriptors. Sensitivity and specificity bar plots can be found in **Figure S12**.

Altogether, a comparison between the PLM-based and descriptor-based ML models in terms of accuracy of across different validation schemes is shown in **Figure 2C**. The results indicate that the ESM 1v model consistently achieves high accuracy across all validation schemes. The VH-based Logistic Regression model using all descriptors also performs well, though slightly lower than the ESM 1v model. Notably, the PCA-based models show comparable performance, demonstrating the effectiveness of dimensionality reduction in maintaining model accuracy. Interestingly, the VH-based Logistic Regression models using the top descriptors, 2, 3, 4, and 5 combinations (

Table 1), exhibit robust performance. This finding suggests that a smaller subset of key descriptors can achieve similar predictive power as using the full set of descriptors, highlighting the potential for model simplification without compromising accuracy.

2.6 VH-based LogisticReg classification model is applicable to clinical-stage therapeutic antibodies

To show applicability of the non-specificity classification model on therapeutic antibodies, the ESM 1v mean-mode VH-based LogisticReg model was tested on the Jain dataset. As in the Boughter dataset, the Jain dataset was parsed into two groups, specific (0 flags) and non-specific (>3 flags), leaving out the mildly non-specific antibodies (1-3 flags). An accuracy of 69% was obtained for the parsed Jain dataset (see confusion matrix in **Figure S14A**). This value is comparable to the mean accuracy of 71% obtained for the same classifier across 3, 5 and 10-Fold CV for the parsed Boughter dataset. Moreover, as in the case of the Boughter dataset, a similar distribution of prediction probability of non-specificity was obtained for the full Jain dataset, and it appears to mimic regression-like behaviour when compared to the non-specificity ELISA flags, although to a slightly weaker extent (**Figure 3A** and **S13**). The same trend can be observed for the top 5 descriptors model (**Figure 3C**). The overall performance of the classifier on the Jain dataset illustrates that it can be applied to therapeutic antibodies.

2.7 Antibodies characterised by the PSR assay appear to be on a different non-specificity spectrum than that from the non-specificity ELISA assay

During recent years, alternative assays to ELISA have been developed to meet the demand of high-throughput screening during drug discovery, and such one is the poly-specific reagent (PSR) assay [62], where antibodies displayed on the surface of yeast cells are counter-selected when non-specifically bound to soluble membrane protein in a flow cytometry-setup. Several studies have been reported using this assay for assessing antibody non-specificity [15,58,63]. Recently, Harvey and co-authors [45] developed a one-hot LogisticReg model based on >140 000 clones assessed by the PSR assay from a naïve Nb library. They found a significant correlation of the PSR with the gold-standard ELISA based on six Nbs. To find out whether our ESM 1v mean-mode VH-based LogisticReg model can extend its applicability further to the non-specificity scored by the PSR assay, the Shehata dataset and the VH-based Nb dataset by Harvey and co-authors [45], here referred to as the Harvey dataset, were tested. The classifier did not appear to separate the PSR-scored specific and non-specific antibodies well. All the specific PSR-scored antibodies of the Shehata dataset were distributed along the entire prediction probability scale, while the few non-specific ones were on the probability end towards higher non-specificity (**Figure 3C,D**). A similar forecast was observed for the Harvey dataset; all the specific PSR-scored Nbs resulted in a broad probability distribution, while the non-specific PSR-scored ones resulted in a narrower probability distribution towards higher non-specificity (**Figure 3E,F**). Thus, the classifier appears to be better at predicting non-specific PSR-scored antibodies, than specific PSR-scored antibodies. This result suggests that the spectrum of non-specificity from the PSR assay is different than the one from the non-specificity ELISA assay, of which the classifier is trained on. Thus, a specific antibody classified by the PSR assay may necessarily not translate into a specific antibody classified by the non-specificity ELISA assay. The

specific PSR-scored Nbs could partly consist of mildly non-specific clones in addition to specific ones, thus resulting in this broad probability distribution.

An interesting remark can be made about the distributions of predicted probabilities obtained from the two different LogisticReg models tested on the Harvey dataset in **Figure 3E,F**. The ESM 1v VH-based LogisticReg model produces a more uniform distribution of predicted probabilities across the dataset, while the top 5 descriptors VH-based LogisticReg model exhibits a clear biphasic distribution. It is no surprise that this bimodal pattern closely resembles the distribution of pI (**Figures S15-S18**), as this descriptor is the main driver of non-specificity in the top 2, 3, 4 and 5 descriptors LogisticReg models (

Table 1 and **Figure S10**). Nonetheless, the distribution of non-specific antibodies in the Harvey dataset appears to exclusively be of high pI (>8) according to **Figure S18A**. The distributions of the other descriptors do not appear to differ significantly between specific and non-specific antibodies see (**Figures S15-S18**). The distinct separation suggests that the pI plays a crucial role in differentiating between specific and non-specific antibodies.

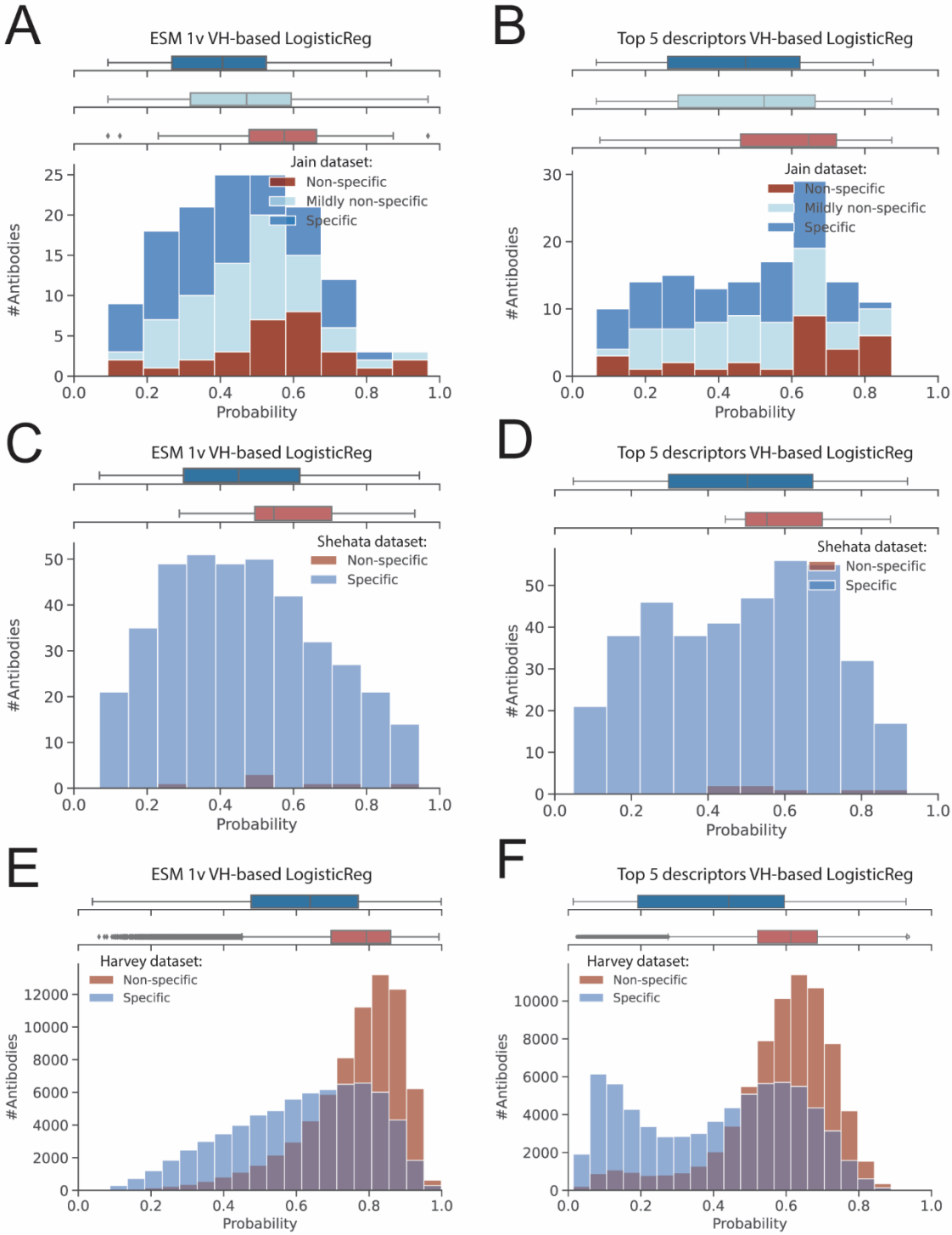

Figure 3. Logistic Regression Models Predicting Antibody Non-Specificity Across Different Datasets. (A-F) Distributions of predicted probabilities of antibody non-specificity for three different datasets using two logistic regression models: predictions for the Jain dataset (A,B), for the Shehata dataset (C,D), and for the Harvey dataset (E,F). (A, C, and E) depict results from the ESM 1v VH-based logistic regression model, while (B, D, and F) depict results from the top 5 descriptors VH-based logistic regression model. For each dataset, antibodies are classified into specific, mildly non-specific (only in the Jain dataset), and non-specific categories, represented by different colours.

2.8 VH-based LogisticReg models performs on par or better than existing predictors

The ESM 1v VH-based LogisticReg model can be compared to two existing predictors in the literature - the predictors reported in the Boughter *et al.* study [44], and the Harvey *et al.* study [45]. Boughter and colleagues stated that while no notable difference could be observed between specific and non-specific antibodies in both gene usage level and amino acid-usage level in CDRs, the positional context of biophysical properties can show the differences. They showed to be successful in developing a binary classifier based on a position-sensitive biophysical matrix with accuracy up to 75%. Their reported performance is on par with the achieved performance of the PLM-based classifier (ESM 1v VH-based LogisticReg), which was trained on the same data (parsed Boughter dataset).

Furthermore, Harvey and colleagues developed a one-hot LogisticReg model based on >140 000 clones assessed by the PSR assay from a naïve Nb library with an accuracy >80%. Using their published web-based predictor [64], we tested its performance on the Boughter, Jain, and Shehata datasets. The results show that the Harvey predictor does not separate well the different antibody groups in the Boughter dataset (**Figure S19A,B**), with overlapping distributions of prediction scores. Similarly, the Jain and the Shehata datasets demonstrate significant overlap between specific and non-specific antibodies (**Figure S19C-F**), indicating some limitations for the Harvey method in predicting non-specificity of antibodies as compared to Nbs.

2.9 Prediction of non-specificity in antibody drug development programs

Consequences of non-specificity in the clinic: Non-specific binding of therapeutic antibodies can lead to significant adverse effects in the clinic. Such antibodies can bind to structurally unrelated off-targets and thereby potentially result in unwanted toxicity [65] or reduced efficacy [66]. They can also interact with tissues like subcutis and thereby result in faster clearance via pinocytosis independently of FcRn [60,61,67]. Ultimately, non-specificity can compromise the safety and efficacy of therapeutic antibodies, potentially resulting in clinical trial failures and increased development costs. Thus, early-stage prediction of non-specificity, such as during selection and optimisation stage, is essential to reduce the risk of failing at late-stage during clinical trials. Otherwise, the further into the development program, the harder it becomes to allow additional protein engineering to mitigate biophysical liabilities, as new *in vitro* and *in vivo* data must be reproduced.

A powerful strategy to address this problem is to combine *in silico* prediction with *in vitro* developability assessment. To identify and flag non-specific antibodies, we propose a combined strategy that integrates *in silico* prediction models with traditional *in vitro* developability assessments during the lead optimization stage in the drug development process. This hybrid approach leverages the strengths of computational predictions and experimental validations, ensuring the selection of lead candidates with high developability potential.

3. Conclusions

In this study, we developed ML models to predict the non-specificity of antibodies, utilizing both PLMs and biophysical properties to embed antibody sequences. In agreement with previous reports

on different datasets [68], our results indicate that the VH domain, particularly the H-CDR loops, as the main contributor to non-specificity, and that the biophysical parameter pI is a key biophysical driver of non-specificity. The resulting computational models enable the prediction of non-specific interactions of antibodies with accuracy of 71% in 10-fold CV, thus providing a valuable tool to guide the design and selection of monoclonal antibodies with improved specificity and efficacy. These findings have important implications for the development of safe and efficacious lead candidates with high developability potential.

4. Methods

4.1 Data sources

All antibody datasets used in this study were retrieved from public sources only. A list of the datasets and their corresponding sources are reported Table 2.

Table 2. List of public antibody datasets with their corresponding size, non-specificity assay and reference.

Dataset	Size	Poly-reactive assay	Reference
Boughter dataset	>1000 antibodies (HIV-1 broadly neutralizing, Influenza reactive, IgA mouse) of varying degree of non-specificity	ELISA with a panel of 7 ligands (DNA, insulin, LPS, albumin, cardiolipin, flagellin, KLH)	[44]
Jain dataset	137 clinical stage IgG1-formatted antibodies	ELISA with a panel of 6 ligands (ssDNA, dsDNA, insulin, LPS, cardiolipin, KLH)	[15]
Shehata dataset	398 antibodies from naïve, IgG memory, IgM memory and long-lived plasma cells	Poly-specific reagent (PSR) assay	[58]
Harvey dataset	>140 000 naïve nanobodies	Poly-specific reagent (PSR) assay	[45]

4.2 Python programming

All coding was performed in Python using Spyder IDE and Jupyter Notebook (Anaconda software distribution) [69], and a list of used Python modules is reported in **Table 3**.

Table 3. List of software and Python modules.

Module	Additional information	Reference
NumPy	https://numpy.org	[70]
SciPy	https://www.scipy.org	[71]
Statsmodels	https://www.statsmodels.org	[72]
Pandas	https://pandas.pydata.org	[73]
Matplotlib	https://matplotlib.org	[74]
Seaborn	https://seaborn.pydata.org	[75]
Scikit-Learn	https://scikit-learn.org	[76]
Json	https://docs.python.org/3/library/json	[77]
Collections	https://docs.python.org/3/library/collections	[78]
Itertools	https://docs.python.org/3/library/itertools	[79]
ANARCI	https://github.com/oxpig/ANARCI	[80]

4.3 Training and validation of binary classification models

First, the Boughter dataset was parsed into three groups as previously done in [44]: specific group (0 flags), mildly poly-reactive group (1-3 flags) and poly-reactive group (>3 flags). The primary

sequences were annotated in the CDRs using ANARCI following the IMGT numbering scheme. Following this, 16 different antibody fragment sequences were assembled and embedded by three state-of-the-art protein language models (PLMs), ESM 1v [81], Protbert bfd [82], and AbLang2 [83], for representation of the physico-chemical properties and secondary/tertiary structure, and a physico-chemical descriptor of amino acids, the Z-scale [84] **Table 4**). For the embeddings from the PLMs, *mean* (average of all token vectors) was used. The vectorised embeddings were served as features for training of binary classification models (e.g. LogisticReg, RandomForest, GaussianProcess, GradeintBoosting and SVM algorithms) for non-specificity (class 0: specific group, and class 1: poly-reactive group). The mildly poly-reactive group was left out from the training of the models.

Table 4. Type and description of sequence input for the binary classification models.

Type of input	Description
VL, VH, L-CDR 1-3, H-CDR 1-3,	A vectorised embedding of VL, VH or an individual CDR sequence
VH/VL joined, L-CDRs joined, H-CDRs joined, H/L-CDRs joined	A vectorised embedding of a joined sequence

The trained classification models were validated by (i) 3, 5 and 10-Fold cross-validation (CV), (ii) Leave-One Family-Out validation, e.g. training on HIV and Influenza reactive antibodies, while testing on mouse IgA antibodies, (iii) comparing probability of predicted poly-reactive class to true class, and (iv) testing on the Jain dataset. The evaluation metrics included accuracy, sensitivity and specificity (Eqs. 1-3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sensitivity = \frac{TP}{P} \tag{2}$$

$$Specificity = \frac{TN}{N} \tag{3}$$

Where TP is true positive (true poly-reactive), TN is true negative (true specific), FP is false positive (false poly-reactive), FN is false negative (false specific), P is all positives (all poly-reactive), and N is all negatives (all specific).

Acknowledgements

We would like to thank Mauricio Augilar Rangel, Dillon Rinauro and Ross Taylor from the University of Cambridge for their valuable discussions to this work.

Funding

Financial support from Novo Nordisk A/S is acknowledged.

P.S. is a Royal Society University Research Fellow (grant no. URF\R1\201461) and acknowledges funding from UK Research and Innovation (UKRI), and Engineering and Physical Sciences Research Council (grant no. EP/X024733/1).

The authors declare no competing financial interest.

References

[1] A. Mullard, FDA approves 100th monoclonal antibody product. *Nat Rev* (2021) 491-495.

[2] H. Kaplon, A. Chenoweth, S. Crescioli, J.M. Reichert, Antibodies to watch in 2022. mAbs (2022) e2014296.

[3] R.M. Lu, Y.C. Hwang, I.J. Liu, C.C. Lee, H.Z. Tsai, H.J. Li, H.C. Wu. Development of therapeutic antibodies for the treatment of diseases. *J Biomed Sci* (2020) 1-30.

[4] A. Pedrioli, A. Oxenius, Single B cell technologies for monoclonal antibody discovery. *Trends Immunol* (2021) 1143-1158.

[5] M. Tabasinezhad, Y. Talebkhan, W. Wenzel, H. Rahimi, E. Omidinia, F. Mahboudi, Trends in therapeutic antibody affinity maturation: From in-vitro towards next-generation sequencing approaches. *Immunol Lett* (2019) 106-113.

[6] T.M. Chidyausiku, S.R. Mendes, J.C. Klima, M. Nadal , U. Eckhard , J. Roel-Touris, S. Houliston, T. Guevara, H.K. Haddox, A. Moyer, C.H. Arrowsmith, F.X. Gomis-Rüth, D. Baker, E. Marcos, De novo design of immunoglobulin-like domains. *Nat Commun* (2022) 5661.

[7] J.L. Watson, D. Juergens, N.R. Bennett, B.L. Trippe, J. Yim, H.E. Eisenach, W. Ahern, A.J. Borst, R.J. Ragotte, L.F. Milles, B.I.M. Wicky, N. Hanikel, S.J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S.V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T.S. Jaakkola, F. DiMaio, M. Baek, D. Baker, De novo design of protein structure and function with RFdiffusion. *Nature* (2023) 10.1038/s41586-023-06415-8.

[8] E.T. Boder, K.S. Midelfort, K.D. Wittrup, Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences USA* (2000) 10701-10705.

[9] T. Igawa, S. Ishii, T. Tachibana, A. Maeda, Y. Higuchi, S. Shimaoka, C. Moriyama, T. Watanabe, R. Takubo, Y. Doi, T. Wakabayashi, A. Hayasaka, S. Kadono, T. Miyazaki, K. Haraya, Y. Sekimori, T. Kojima, Y. Nabuchi, Y. Aso, Y. Kawabe, K. Hattori, Antibody recycling by engineered pH-dependent antigen binding improves the duration of antigen neutralization. *Nat Biotechnol* (2010) 1203-1207.

[10] T. Klaus, S. Deshmukh, pH-responsive antibodies for therapeutic applications. *J. Biomed. Sci.* (2021) 10.1186/s12929-021-00709-7.

[11] M. Kamata-Sakurai, Y. Narita, Y. Hori, T. Nemoto, R. Uchikawa, M. Honda, N. Hironiwa, K. Taniguchi, M. Shida-Kawazoe, S. Metsugi, T. Miyazaki, N.A. Wada, Y. Ohte, S. Shimizu, H. Mikami, T. Tachibana, N. Ono, K. Adachi, T. Sakiyama, T. Matsushita, S. Kadono, S.I. Komatsu, Akihisa Sakamoto, S. Horikawa, A. Hirako, K. Hamada, S. Naoi, N. Savory, Y. Satoh, M. Sato, Y.

Noguchi, J. Shinozuka, H. Kuroi, A. Ito, T. Wakabayashi, M. Kamimura, F. Isomura, Y. Tomii, N. Sawada, A. Kato, O. Ueda, Y. Nakanishi, M. Endo, K.I. Jishage, Y. Kawabe, T. Kitazawa, T. Igawa, Antibody to CD137 Activated by Extracellular Adenosine Triphosphate Is Tumor Selective and Broadly Effective In Vivo without Systemic Immune Activation. *Cancer Discov*. (2021) 158-175.

[12] M. Elshiaty, H. Schindler, P. Christopoulos, Principles and Current Clinical Landscape of Multispecific Antibodies against Cancer. *Int J Mol Sci* (2021) 5632.

[13] H. Østergaard, J. Lund, P.J. Greisen, S. Kjellev, A. Henriksen, N. Lorenzen, E. Johansson, G. Røder, M.G. Rasch, L.B. Johnsen, T. Egebjerg, S. Lund, H. Rahbek-Nielsen, P.S. Gandhi, K. Lamberth, M. Loftager, L.M. Andersen, A.C. Bonde, F. Stavenuiter, D.E. Madsen, X. Li, T.L. Holm, C.D. Ley, P. Thygesen, H. Zhu, R. Zhou, K. Thorn, Z. Yang, M.B. Hermit, J.R. Bjelke, B.G. Hansen, I. Hilden, A factor VIIIa-mimetic bispecific antibody, Mim8, ameliorates bleeding upon severe vascular challenge in hemophilia A mice. *Blood* (2021) 1258-1268.

[14] C.G. Starr, P.M. Tessier, Selecting and engineering monoclonal antibodies with drug-like specificity. *Curr Opin Biotechnol* (2019) 119-127.

[15] T. Jain, T. Sun, S. Durand, A. Hall, N.R. Houston, J.H. Nett, B. Sharkey, B. Bobrowicz, I. Caffry, Y. Yu, Y. Cao, H. Lynaugh, M. Brown, H. Baruah, L.T. Gray, E.M. Krauland, Y. Xu, M. Vásquez, K.D. Wittrup, Biophysical properties of the clinical-stage antibody landscape. Proceedings of the National Academy of Sciences USA (2017) 944-949.

[16] L.B. Avery, J. Wade, M. Wang, A. Tam, A. King, N. Piche-Nicholas, M.S. Kavosi, S. Penn, D. Cirelli, J.C. Kurz, M. Zhang, O. Cunningham, R. Jones, B.J. Fennell, B. McDonnell, P. Sakorafas, J. Apgar, W.J. Finlay, L. Lin, L. Bloom, D.M. O'Hara. Establishing in vitro in vivo correlations to screen monoclonal antibodies for physicochemical properties related to favorable human pharmacokinetics. *MAbs* (2018) 244-255.

[17] R.L. Kelly, T. Sun, T. Jain, I. Caffry, Y. Yu, Y. Cao, H. Lynaugh, M. Brown, M. Vásquez, K.D. Wittrup, Y. Xu, High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. *MAbs* (2015) 770-777.

[18] T.E. Kraft, W.F. Richter, T. Emrich, A. Knaupp, M. Schuster, A. Wolfert, H. Kettenberger, Heparin chromatography as an in vitro predictor for antibody clearance rate through pinocytosis. *MAbs* (2020) 1683432.

[19] C.L. Dobson, P.W. Devine, J.J. Phillips, D.R. Higazi, C. Lloyd, B. Popovic, J. Arnold, A. Buchanan, A. Lewis, J. Goodman, C.F. van der Walle, P. Thornton, L. Vinall, D. Lowne, A. Aagaard, L.L. Olsson, A. Ridderstad Wollberg, F. Welsh, T.K. Karamanos, C.L. Pashley, M.G. Iadanza, N.A. Ranson, A.E. Ashcroft, A.D. Kippen, T.J. Vaughan, S.E. Radford, D.C. Lowe, Engineering the

surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Sci Rep* (2016) 38644.

[20] A. Datta-Mannan, J. Lu, D.R. Witcher, D. Leung, Y. Tang, V.J. Wroblewski, The interplay of non-specific binding, target-mediated clearance and FcRn interactions on the pharmacokinetics of humanized antibodies. *MAbs* (2015) 1084-1093.

[21] A. Datta-Mannan, A. Thangaraju, D. Leung, Y. Tang, D.R. Witcher, J. Lu, V.J. Wroblewski, Balancing charge in the complementarity-determining regions of humanized mAbs without affecting pI reduces non-specific binding and improves the pharmacokinetics. *MAbs* (2015) 483-493.

[22] T. Igawa, H. Tsunoda, T. Tachibana, A. Maeda, F. Mimoto, C. Moriyama, M. Nanami, Y. Sekimori, Y. Nabuchi, Y. Aso, K. Hattori, Reduced elimination of IgG antibodies by engineering the variable region. *Protein Eng Des Sel* (2010) 385-392.

[23] D. Bumbaca, A. Wong, E. Drake, A.E. Reyes II, B.C. Lin, J.P. Stephan, L. Desnoyers, B.Q. Shen, M.S. Dennis, Highly specific off-target binding identified and eliminated during the humanization of an antibody against FGF receptor 4. *mAbs* (2011) 376-386.

[24] K.E. Tiller, L. Li, S. Kumar, M.S. Julian, S. Garde, P.M. Tessier, Arginine mutations in antibody complementarity-determining regions display context-dependent affinity/specificity trade-offs. *J Biol Chem*, (2017) 16638-16652.

[25] R.L. Kelly, D. Le, J. Zhao, K.D. Wittrup, Reduction of Nonspecificity Motifs in Synthetic Antibody Libraries. *J Mol Biol* (2018) 119-130.

[26] H. Ausserwöger, M.M. Schneider, T.W. Herling, P. Arosio, G. Invernizzi, T.P.J. Knowles, N. Lorenzen, Non-specificity as the sticky problem in therapeutic antibody development. *Nat Rev Chem* (2022) 844-861.

[27] O. Cunningham, M. Scott, Z.S. Zhou, W.J.J. Finlay, Polyreactivity and polyspecificity in therapeutic antibody development: risk factors for failure in preclinical and clinical development campaigns. *MAbs* (2021) 1999195.

[28] K.E. Tiller, L. Li, S. Kumar, M.S. Julian, S. Garde, P.M. Tessier, Arginine mutations in antibody complementarity-determining regions display context-dependent affinity/specificity trade-offs. *J Biol Chem*, (2017) 16638-16652.

[29] R.L. Kelly, D. Le, J. Zhao, K.D. Wittrup, Reduction of Nonspecificity Motifs in Synthetic Antibody Libraries. *J Mol Biol* (2018) 119-130.

[30] H. Ausserwöger, M.M. Schneider, T.W. Herling, P. Arosio, G. Invernizzi, T.P.J. Knowles, N. Lorenzen, Non-specificity as the sticky problem in therapeutic antibody development. *Nat Rev Chem* (2022) 844-861.

[31] O. Cunningham, M. Scott, Z.S. Zhou, W.J.J. Finlay, Polyreactivity and polyspecificity in therapeutic antibody development: risk factors for failure in preclinical and clinical development campaigns. *MAbs* (2021) 1999195.

[32] J.P. Palmer, C.M. Asplin, P. Clemons, K. Lyen, O. Tatpati, P.K. Raghu, T.L. Paquette, Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science* (1983) 1337-1339.

[33] C.E. Taplin, J.M. Barker, Autoantibodies in type 1 diabetes. *Autoimmunity* (2008) 11-18.

[34] P.W. Noble, S. Bernatsky, A.E. Clarke, D.A. Isenberg, R. Ramsey-Goldman, J.E. Hansen, DNA-damaging autoantibodies and cancer: the lupus butterfly theory. *Nat Rev Rheumatol* (2016) 429-434.

[35] J. Nehring, L.A. Schirmbeck, J. Friebus-Kardash, D. Dubler, U. Huynh-Do, C. Chizzolini, C. Ribi, M. Trendelenburg, Autoantibodies Against Albumin in Patients With Systemic Lupus Erythematosus. *Front Immunol* (2018) 10.1128/cdli.6.6.775-782.1999.

[36] S. Miyakis, M.D. Lockshin, T. Atsumi, D.W. Branch, R.L. Brey, R. Cervera, R.H. Derksen, P.G. DE Groot, T. Koike, P.L. Meroni, G. Reber, Y. Shoenfeld, A. Tincani, P.G. Vlachoyiannopoulos, S.A. Krilis, International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). *J Thromb Haemost* (2006) 295-306.

[37] I.R. Poxton, Antibodies to lipopolysaccharide. J Immunol Methods (1995) 1-15.

[38] H. Mouquet, J.F. Scheid, M.J. Zoller, M. Krogsgaard, R.G. Ott, S. Shukair, M.N. Artyomov, J. Pietzsch, M. Connors, F. Pereyra, B.D. Walker, D.D. Ho, P.C. Wilson, M.S. Seaman, H.N. Eisen, A.K. Chakraborty, T.J. Hope, J.V. Ravetch, H. Wardemann, M.C. Nussenzweig, Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. Nature (2010) 591-595.

[39] J.J. Guthmiller, L.Y. Lan, M.L. Fernández-Quintero, J. Han, H.A. Utset, D.J. Bitar, N.J. Hamel, O. Stovicek, L. Li, M. Tepora, C. Henry, K.E. Neu, H.L. Dugan, M.T. Borowska, Y.Q. Chen, S.T.H. Liu, C.T. Stamper, N.Y. Zheng, M. Huang, A.E. Palm, A. García-Sastre, R. Nachbagauer, P. Palese, L. Coughlan, F. Krammer, A.B. Ward, K.R. Liedl, P.C. Wilson, Polyreactive Broadly Neutralizing B cells Are Selected to Provide Defense against Pandemic Threat Influenza Viruses. *Immunity* (2020) 1230-1244.

[40] I. Hötzel, F.P. Theil, L.J. Bernstein, S. Prabhu, R. Deng, L. Quintana, J. Lutman, R. Sibia, P. Chan, D. Bumbaca, P. Fielder, P.J. Carter, R.F. Kelley, A strategy for risk mitigation of antibodies with fast clearance. *MAbs* (2012) 753-760.

[41] Y. Xu, W. Roach, T. Sun, T. Jain, B. Prinz, T.Y. Yu, J. Torrey, J. Thomas, P. Bobrowicz, M. Vásquez, K.D. Wittrup, E. 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 Sel* (2013) 663-670.

[42] T.E. Kraft, W.F. Richter, T. Emrich, A. Knaupp, M. Schuster, A. Wolfert, H. Kettenberger, Heparin chromatography as an in vitro predictor for antibody clearance rate through pinocytosis. *MAbs* (2020) 1683432.

[43] R.L. Kelly, T. Sun, T. Jain, I. Caffry, Y. Yu, Y. Cao, H. Lynaugh, M. Brown, M. Vásquez, K.D. Wittrup, Y. Xu, High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. *MAbs* (2015) 770-777.

[44] C.T. Boughter, M.T. Borowska, J.J. Guthmiller, A. Bendelac, P.C. Wilson, B. Roux, E.J. Adams, Biochemical patterna of antibody polyreactivity revealed through a bioinformatics-based analysis of CDR loops. *eLife* (2020) 9:e61393.

[45] E.P. Harvey, J.E. Shin, M.A. Skiba, G.R. Nemeth, J.D. Hurley, A. Wellner, A.Y. Shaw, V.G. Miranda, J.K. Min, C.C. Liu, D.S. Marks, A.C. Kruse, An in silico method to assess antibody fragment polyreactivity. *Nat Commun* (2022) 13, 7554.

[46] P.A. Robert, R. Akbar, R. Frank, M. Pavlović, M. Widrich, I. Snapkov, A. Slabodkin, M. Chernigovskaya, L. Scheffer, E. Smorodina, P. Rawat, B. Bhushan Mehta, M. Ha Vu, I. Frøberg Mathisen, A. Prósz, K. Abram, A. Olar, E. Miho, D. Trygve Tryslew Haug, F. Lund-Johansen, S. Hochreiter, I. Hobæk Haff, G. Klambauer, G. Kjetil Sandve, V. Greiff, Unconstrained generation of synthetic antibody–antigen structures to guide machine learning methodology for antibody specificity prediction. *Nat Comput Sci* (2022) 10.1038/s43588-022-00372-4.

[47] E.K. Makowski, P.C. Kinnunen, J. Huang, L. Wu, M.D. Smith, T. Wang, A.A. Desai, C.N. Streu, Y. Zhang, J.M. Zupancic, J.S. Schardt, J.J. Linderman, P.M. Tessier, Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nat Commun* (2022) 3788.

[48] R.L. Kelly, D. Le, J. Zhao, K.D. Wittrup, Reduction of nonspecificity motifs in synthetic antibody libraries. *J Mol Biol* (2018) 13, 119-130.

[49] A.A.R. Teixeira, S. D'Angelo, M.F. Erasmus, C. Leal-Lopes, F. Ferrara, L.P. Spector, L. Naranjo, E. Molina, T. Max, A. DeAguero, K. Perea, S. Stewart, R.A. Buonpane, H.G. Nastri, A.R.M. Bradbury, Simultaneous affinity maturation and developability enhancement using natural liability-free CDRs. *MAbs* (2022) 14, 2115200.

[50] Hie, Brian L., et al. "Efficient evolution of human antibodies from general protein language models." Nature biotechnology 42.2 (2024): 275-283.

B.L. Hie, V.R. Shanker, D. Xu, T.U.J. Bruun, P.A. Weidenbacher, S. Tang, W. Wu, J.E. Pak, P.S. Kim, Efficient evolution of human antibodies from general protein language models. *Nat Biotechnol* (2024) 42, 275-283.

[51] T.H. Olsen, I.H. Moal, C.M. Deane, AbLang: an antibody language model for completing antibody sequences. *Bioinform Adv* (2022) 10.1093/bioadv/vbac046.

[52] V.R. Shanker, T.U.J. Bruun, B.L. Hie, P.S. Kim, Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science* (2024) 385, 46-53.

[53] R. Singh, C. Im, Y. Qiu, B. Mackness, A. Gupta, T. Joren, S. Sledzieski, L. Erlach, M. Wendt, Y.F. Nanfack, B. Bryson, B. Berger, Learning the language of antibody hypervariability. Proceedings of the National Academy of Sciences USA (2025) 122, e2418918121.

[54] D.M. Mason, S. Friedensohn, C.R. Weber, C. Jordi, B. Wagner, S.M. Meng, R.A. Ehling, L. Bonati, J. Dahinden, P. Gainza, B.E. Correia, S.T. Reddy, Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat Biomed Eng* (2021) 5, 600-612.

[55] Y. Wang, H. Lv, Q.W. Teo, R. Lei, A.B. Gopal, W.O. Ouyang, Y.H. Yeung, T.J.C. Tan, D. Choi, I.R. Shen, X. Chen, C.S. Graham, N.C. Wu, An explainable language model for antibody specificity prediction using curated influenza hemagglutinin antibodies. *Immunity* (2024) 8, 2453-2465.

[56] M. Wang, J. Patsenker, H. Li, Y. Kluger, S.H. Kleinstein, Supervised fine-tuning of pre-trained antibody language models improves antigen specificity prediction. *PLOS Comput Biol* 21.3 (2025) e1012153.

[57] X. Yu, K. Vangjeli, A. Prakash, M. Chhaya, S.J. Stanley, N. Cohen, L. Huang, Protein language models enable prediction of polyreactivity of monospecific, bispecific, and heavy-chain-only antibodies. *Antib Ther* (2024) 30, 199-208.

[58] L. Shehata, D.P. Maurer, A.Z. Wec, A. Lilov, E. Champney, T. Sun, K. Archambault, I. Burnina, H. Lynaugh, X. Zhi, Y. Xu, L.M. Walker, Affinity maturation enhances antibody specificity but compromises conformational stability. *Cell Reports* (2019) 3300-3308.

[59] J.J. Guthmiller, L. Yu-Ling Lan, M.L. Fernández-Quintero, J. Han, H.A. Utset, D.J. Bitar, N. J. Hamel, O. Stovicek, L. Li, M. Tepora, C. Henry, K.E. Neu, H.L. Dugan, M.T. Borowska, Y.Q. Chen, S.T.H. Liu, C.T. Stamper, N.Y. Zheng, M. Huang, A.K.E. Palm, P.C. Wilson, Polyreactive Broadly

Neutralizing B cells Are Selected to Provide Defense against Pandemic Threat Influenza Viruses. *Immunity* (2020) 53, 1230-1244.

[60] R.L. Kelly, Y. Yu, T. Sun, I. Caffry, H. Lynaugh, M. Brown, T. Jain, Y. Xu, K.D. Wittrup, Target-independent variable region mediated effects on antibody clearance can be FcRn independent. *Mab*s (2016) 8, 1269-1275.

[61] B. Li, D. Tesar, C.A. Boswell. H.S. Cahaya, A. Wong, J. Zhang, Y.G. Meng, C. Eigenbrot, H. Pantua, J. Diao, S.B Kapadia, R. Deng, R.F. Kelley. Framework selection can influence pharmacokinetics of a humanized therapeutic antibody through differences in molecule charge. *MAbs* (2014) 6, 1255-1264.

[62] Y. Xu, W. Roach, T. Sun, T. Jain, B. Prinz, T.Y. Yu, J. Torrey, J. Thomas, P. Bobrowicz, M. Vásquez, K.D. Wittrup, E. Krauland, Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. *Protein Engineering, Design and Selection* (2013) 26, 663-670.

[63] R.L. Kelly, T. Sun, T. Jain, I. Caffry, Y. Yu, Y. Cao, H. Lynaugh, M. Brown, M. Vásquez, K.D. Wittrup, Y. Xu, High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. *mAbs* (2015) 7, 770-777.

[64] Web-based predictor of nanobody non-specificity available at Nanobody Polyreactivity Prediction Server.

[65] W.J.J. Finlay, J.E. Coleman, J.S. Edwards, K.S. Johnson, Anti-PD1 'SHR-1210' aberrantly targets pro-angiogenic receptors and this polyspecificity can be ablated by paratope refinement. *MAbs* (2019) 11, 26-44.

[66] D. Bumbaca, A. Wong, E. Drake, A.E. Reyes 2nd, B.C. Lin, J.P. Stephan, L. Desnoyers, B.Q. Shen, M.S. Dennis, Highly specific off-target binding identified and eliminated during the humanization of an antibody against FGF receptor 4. *MAbs* (2011) 3, 376-386.

[67] C.A. Boswell, D.B. Tesar, K. Mukhyala, F.P. Theil, P.J. Fielder, L.A. Khawli, Effects of charge on antibody tissue distribution and pharmacokinetics. *Bioconjug Chem* (2010) 21, 2153-2163.

[68] H.T. Chen, Y. Zhang, J. Huang, M. Sawant, M.D. Smith, N. Rajagopal, A.A. Desai, E. Makowski, G. Licari, Y. Xie, M.S. Marlow, S. Kumar, P.M. Tessier, Human antibody polyreactivity is governed primarily by the heavy-chain complementarity-determining regions. *Cell Rep* (2024) 22, 10.1016/j.celrep.2024.114801.

[69] Anaconda Software Distribution, Anaconda Documentation. Anaconda Inc (2020) Retrieved from https://docs.anaconda.com/.

[70] C.R. Harris, K.J. Millman, S.J. van der Walt *et al.*, Array programming with NumPy. *Nature* 585 (2020) 357–362.

[71] P. Virtanen, R. Gommers, T.E. Oliphant *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17 (2020) 261–272.

[72] Seabold, Skipper, Josef Perktold, statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference* (2010).

[73] W. McKinney, Data Structures for Statistical Computing in Python, *Proceedings of the 9th Python in Science Conference* (2010).

[74] J.D. Hunter, Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9 (2007) 90-95.

[75] M. Waskom, O. Botvinnik, D. O'Kane *et al.*, mwaskom/seaborn: v0.8.1 (September 2017). *Zenodo* (2017) Retrieved from https://doi.org/10.5281/zenodo.883859.

[76] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011) 2825-2830.

[77] Json, The Python Standard Standard, release 3.9.2. Python Software Foundation. Retrieved from https://docs.python.org/3/library/json.html#rfc-errata.

[78] Collections, The Python Standard Reference, release 3.9.2. Python Software Foundation. Retrieved from https://docs.python.org/3/library/collections.html.

[79] Itertools, The Python Standard Reference, release 3.9.2. Python Software Foundation. Retrieved from https://docs.python.org/3/library/itertools.html?highlight=itertools.

[80] J. Dunbar, C.M. Deane, ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* 32 (2016) 298-300.

[81] A. Rives, J., Meier, T., Sercu, S., Goyal, Z., Lin, J., Liu, D., Guo, M., Ott, C.L., Zitnick, J., Ma, R., Fergus, Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences, *Proceedings of the National Academy of Sciences USA* (2021) 118, e2016239118.

[82] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, Y. Wang, L. Jones, T. Gibbs, T. Feher, C. Angerer, M. Steinegger, D. Bhowmik, B. Rost, ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022) 44, 7112-7127.

[83] T.H. Olsen, I.H. Moal, C.M. Deane, Addressing the antibody germline bias and its effect on language models for improved antibody design. *bioRXiv* (2024) 10.1101/2024.02.02.578678v1.

[84] M. Sandberg, L. Eriksson, J. Jonsson, M. Sjöström, S. Wold, New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A Multivariate Characterization of 87 Amino Acids. *J Med Chem* (1998) 41, 2481-2491.