Molecular Dynamics Informed Prediction of Antibody Thermostability

Hew Phipps

Department of Statistics, The University of Oxford

Supervised by Prof. Charlotte Deane, Santiago Villalba and Dr Matthew Raybould

December 14, 2025

1 Introduction

Antibodies are large Y-shaped proteins constituting the adaptive immune system's primary effectors thought to bind with high specificity to their targets. Typical monoclonal antibodies (mAbs) are made up of two pairings of heavy-light chain heterodimers where the heavy chains join at the stem (Fc region) via disulfide bridges. The arms (Fab region) are split into conserved and variable sections. The variable Fv region contains the binding interface which is primarily made up of three variable extended loop regions on the end of each chain called the CDR loops. As the Fv region determines binding affinity it is often the primary target in antibody design.

Historically target binding affinity has been the focal design objective in antibody development, mostly because before modern *in silico* methods therapeutic design was best achieved by rational mutation of, and or chimeric recombination of Fv regions onto reliable IgG1 Immunoglobulins, the dominant class of antibodies in humans. Such "CDR grafting" has resulted in a multitude of successful biologic drugs, like adalimumab, as it helps maintain the clinical viability of antibodies by optimising their "humanness" and other antibody attributes colloquially known as developability [1].

Developability traits remain a major bottleneck in drug development pipelines. These traits often have a complex relationship with binding affinity [2]. Computational prediction of developability traits therefore makes for an appealing solution to reducing developmental costs and timelines by providing the opportunity to vet drug candidates before reaching clinical stages where such traits could preclude advancement [3]. One such developability trait is thermostability. Good thermostability is vital for manufacture, storage and shelf life [4], reducing aggregation propensity [1] whilst helping prevent misfolding, and even affecting binding potency [5].

Since AlphaFold2 [6], sequence and structure based deep learning models have proven powerful predictors of protein structure. Antibody specific models such as ImmuneBuilder [7], and generative models conditioned on binding targets [8] perform well *in silico* highlighting the power of machine learning in drug design [9]. Indeed, the first Machine Learning (ML) aided antibody therapeutic to reach clinical use, Bimekizumab, shows superiority over previous state of the art treatments for plaque psoriasis, adalimimab and secukinumab [10], and AU-007 represents perhaps the first antibody therapeutic designed using artificial intelligence in clinical trials [11]. Deep learning language models in particular have been frequently applied to the problem of thermostability prediction [12, 13] and seem to perform well at producing antibody variants with increased stability [14]. Deep learning however benefits significantly from large volumes of high quality data, which is readily available for sequence based learning (almost 2 million unique paired antibody sequences and over 2 billion unpaired sequences [15]) and structure prediction (over 10,000

Table 1: Experimental Thermostability Metrics

Metric	**Description**	**Experimental Method**
Tm1 (Melting temperature)	Temperature at which 50% of the protein is unfolded	Differential Scanning Fluorometry (DSF)
Tm2 (Second Melting temperature)	Second melting curve transition state	DSF melt curve 2nd transition
Ton (Onset temperature)	Temperature at which unfolding begins	DSF inflection point of melt curve
Tagg (Onset of Aggregation)	Temperature at which aggregation begins	Dynamic Light Scattering (DLS)

resolved antibody structures in the PDB [16]) but less so for developability (on the scale of 100s) due to these traits' multiparameter nature and the costs of experimental methods.

Applying sequence or structure based deep learning methods to the task of thermostability prediction has so far resulted in inconsistent performance with particularly poor results on out of training family distribution datasets [17], hinting at poor generalisability. Recent work suggests $\Delta\Delta$G antibody-antigen binding prediction could require data on the order of 10^4 suggesting similar numbers will be required for developability traits [18]. To further convolute the problem of training data, there are several experimental metrics for thermostability, as summarised in Table 1.

Experimental conditions such as the protein:dye ratio, temperature ramp rate, buffer, salts and pH, down to variations within the PCR instruments used, can contribute noise to data [19, 20, 21]. Even the method of calculating Tm from a melt curve can vary from taking the peak of the negative derivative, applying a two state model or sigmoidal fit, or just the median between the baseline and the peak of the curve [22, 23].

This highlights a larger issue in that thermostability is a result of not just the chemical composition of the molecule but also the surrounding solvent [4] and pH [24], as shown by a 96 condition analysis of Trasuzumab Tm [1]. Such external factors are difficult to incorporate into the training regimes of common deep learning models. Ideally, predictive models will be capable of generalising beyond environmental context, which is why many approaches focus on intrinsic chemical traits known to contribute to thermostability such as the number of internal contacts, the overall Lennard Jones (LJ) potential and the

[1] https://www.proteinstable.com/assets/files/pdf/Formulation-screen-of-Trastuzumab-using-the-SUPR-DSF.pdf

presence of salt bridges and disulfide bonds [25, 26]. Aromatic ring stacking at the interface between the constant and variable regions has been shown to be a large contributor to the stability of adalimumab [27] and the introduction of disulfide bonds increased denaturation temperature by 6.5°C [28].

Deep learning methods may therefore benefit from added physics-based information. Indeed, ThermML shows how just providing Rosetta energy features can improve thermostability classification of single chain Fv region variants [29]. Physics informed approaches like PIDiff [30] have shown how incorporating the LJ of the antigen-antibody binding interface can significantly improve the binding affinity of diffusion model-generated antibodies. Such a "physics informed" approach could be of great benefit to stability prediction by reducing the reliance on data whilst suffering less from the likes of evolutionary bias seen in sequence based deep learning models. For example, Chungyoun et al. [17] show how sequence based deep learning models have poor performance predicting thermostability for antibodies out of the training set families and assign higher fitness to the wild type antibody sequence despite achieving lower thermostability than select mutants, whilst a Rosetta energy-based model avoids this pitfall.

Proteins are dynamic and exist across a multitude of conformations responsive to their environment which are not easily captured by classic structure prediction models trained on solved crystal structures [31]. Classical statistical mechanics enables us to estimate molecule stability from ensembles of conformations with states more commonly represented being energetically more stable. Thus, equilibrium ensemble approaches can learn the energy landscape of a protein [32], potentially enabling thermostability prediction with higher accuracy.

MD provides a solution to the above within the limit of computational efficiency. The ergodic hypothesis states that conformational statistics captured from simulations over sufficient time approximate large ensembles [33]. As MD is an atomistic level simulation driven by Newtonian mechanics it also serves as a plausible ground truth for molecular physics. So, training a model on MD trajectories offers an opportunity for physics informed learning that can capture ensemble statistics. Such an MD informed approach for antibody thermostability prediction was recently applied to nanobodies [34] as well as mAb Fv regions by Merck [35], who simulated 25 internal IgG1 Fv structures for 100ns at 300K, 350K and 400K and used a number of descriptor features from the MD data to predict Tm, Tagg and Ton with high accuracy, a method they called AbMelt.

In this work we take inspiration from AbMelt's approach, applying it to a public antibody thermostability dataset of 137 therapeutic mAbs standardised on human IgG1 isotypes in Hepes-buffered saline [36]. We present HOGROAST, a robust and modular refactor of

AbMelt for antibody Tm prediction from GROMACS MD simulations. With HOGROAST we also implement antibody Fv region angular dynamics analysis using ABangle [37] and provide an additional simulation protocol for coarse graining with Martini which enables a 7x improvement in computation time allowing for simulation at the microsecond level where larger conformational dynamics begin to arise [38]. We show how these new features also contribute to aggregation propensity prediction using experimental results from the Jain et al. [36] set and discuss why our results are significantly different to AbMelt's, alongside the potential for further work.

2 Methods

2.1 Training Test Split

A total of 137 antibody Fv region sequences and associated developability assay data was retrieved from Jain et al. [36] by csv file. Antibodys were split into train and test sets by clustering by Levenshtein distance sequence similarity with the agglomerative clustering algorithm. The train-test split was produced by taking the combination of clusters that came closest to a 70–30% ratio in sequence counts without splitting clusters between sets. The test set was held out from all analyses and training until model testing. For cross validation (CV), the training set was split into 3 folds to give 3 CV sets taking a Leave One Out strategy (LOO). Folds were produced by clustering the train set alone by sequence similarity as before and grouping clusters into folds without splitting clusters across folds to achieve as similar antibody counts for each fold as possible.

2.2 Internal Contacts

As in AbMelt, we define internal contacts as the number of intramolecular residue pairs within a distance of 3.5Å. To calculate this for static pdb structures, we use Biopython's PDB parser tools [39] and iterate through all possible atom pairs, tallying a residue pair once if at least one atom pair is within 3.5Å. We used MDAnalysis [40] to load trajectories and its FastNS neighbour search function to count the number of contacts for each frame.

2.3 Correlation Analysis

All correlations throughout this work refer to Pearson's correlation coefficient (r_p), which denotes the strength and correlation direction (slope of the gradient) of the optimal linear relationship between two variables (Equation 1).

$$r_p = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{1}$$

Where X and Y are sets of observations and σ_X and σ_Y are their standard deviations. Within this text we refer to the 11 Jain et al. assay metrics for each antibody as "fitness features" (Table S2) corresponding with Y whilst the MD descriptor features for each antibody correspond to X.

To compare different datasets of correlations between antibody MD features and fitness features such as Tm we use a number of matrix comparison statistics. For two datasets of Pearson's r_p correlation coefficients as matrices $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{n \times m}$ where each row a_i and b_i corresponds to the vector of correlations for sample i of n across m features, we compute:

1. The mean μ_r of all $|r_p|$ for each dataset is computed as:

$$\mu_r = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} |r_p^{i,j}| \tag{2}$$

2. A paired t-test t on Fisher's z-transformed r_p (z) to determine if the means of the two datasets are significantly different is computed as:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} (z_a^i - z_b^i) \tag{3}$$

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left((z_a^i - z_b^i) - \bar{d}\right)^2} \tag{4}$$

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \tag{5}$$

3. A Wilcoxon signed-rank test W on Fisher's z-transformed r_p, a non-parametric test to determine if the median of differences of the two datasets is significantly different from zero, is computed as:

$$d_i = z_a^i - z_b^i \tag{6}$$

$$R_i = \text{rank}(|d_i|) \tag{7}$$

$$W = \sum_{i=1}^{n} \text{sgn}(d_i) \cdot R_i \tag{8}$$

4. The difference in the Frobenius norms $\|A - B\|_F$ of two matrices of r_p values was computed to gauge the difference in overall magnitude of correlations with:

$$\|A - B\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a^{i,j} - b^{i,j})^2} \tag{9}$$

Where the Fisher's z-transform, z, used to remove skew from the correlation matrices and approximate normality, is defined by:

$$z = \frac{1}{2} \ln \left(\frac{1 + r_p}{1 - r_p} \right) \tag{10}$$

To determine what r_p correlations are to be considered significant we used a one-sample t-test for significance (Equation 12) for each r_p in each dataset to get the number of significantly correlated features and a critical r_p (r_{crit}) at which correlations less than this are rejected under the null hypothesis.

$$t = \frac{r_p\sqrt{n-2}}{\sqrt{1-(r_p)^2}} \tag{11}$$

$$r_{\text{crit}} = \frac{t_{\text{crit}}}{\sqrt{t_{\text{crit}}^2 + df}} \tag{12}$$

Where the degrees of freedom df is equal to $n-2$ and t_{crit} is obtained from a two tailed test equal to the $(1-\alpha/2)$ percentile of the Student's t-distribution where we used an α of 0.05 by default.

2.4 Baselines

The database TheraSAbDab [41] was used to find PDB crystal structures for antibody sequences in the Jain et al. dataset. On 26/07/25 all entries in the database were downloaded from the TheraSAbDab webserver as a csv file. Using Python, the data was scraped for antibodies sharing the same name as any in the Jain et al. dataset. Columns pertaining to antibody name and PDB ID for structures with 100%, 99% or 95-98% sequence similarity were collected.

For some antibodies, several PDB entries were available each listed with their respective heavy and light chain IDs for that PDB structure (often non-canonically labelled). The first PDB code and chain ID per structure was taken for each of the 100%–99% sequence identity PDBs and used to scrape structures from the RSCB PDB site, retrieving a PDB file for each. As the retrieved structures were Fab regions they were renumbered with ANARCI [42] to extract Fv regions before hydrogens were added with PDBfixer. Internal contacts for each structure was determined based on the definition outlined in Section 2.2. The resultant structure contacts per antibody were plotted against their respective Tm values to obtain a baseline correlation.

For benchmarks more comparable to molecular dynamics trajectory data antibody Fv region structure ensembles were generated for each antibody in the Jain set with ImmuneBuilder [7] using the ABodyBuilder2 model. Briefly, for each antibody the heavy and light chain sequences were fed into four separate models to predict 4 separate structures. The closest structure to the average was selected then relaxed with OpenMM [43] to remove structural clashes. This was repeated 100 times for each antibody in the Jain dataset to produce a final ensemble from which the mean and standard deviation of internal contacts across the ensemble was taken and their correlations with Tm determined.

2.5 Molecular Dynamics Simulation with GROMACS

All molecular dynamics simulations were performed with GROMACS-2025.1 [44]. Structures were predicted from VH and VL sequences with ImmuneBuilder's ABodyBuilder2 [7]. Histidine residue protonation states were determined by predicting each structures' pKa using propka 3.1 [45, 46] and deprotonating if it was below the physiological pH of 7.4 at which all simulations were conducted. CDR loops regions were identified with IMGT numbering using ANARCI [42], and for solved PDB structures this was also used to identify heavy and light chains.

All all-atom simulations followed the same protocol set out in AbMelt [35]. In brief, all-atom simulations used the CHARMM27 (CHARMM22 + CMAP) force field [47]. The system was solvated with 9000 water molecules with the TIP3P water model [48] in rectangular boxes of $5.8 \times 7.5 \times 7.3\,\mathrm{nm}^3$ and 60 Na+ and Cl- ions were added to reach a salt concentration of 150 mM. The total number of atoms reached 30,000 per simulation. The LeapFrog algorithm was used with 2 fs timesteps. First

Simulations were carried out at each temperature of 300K, 350K, 400K with four steps:

1. The system was relaxed with steepest descent energy minimization.
2. Simulation with constant particle, volume and temperature (NVT) canonical ensemble for 100 ps using the velocity rescale (V-rescale) thermostat with a 0.1 ps time constant.
3. Simulation with constant particle, pressure and temperature (NPT) ensemble for 100 ps using the Berendsen barostat and a 2.0 ps time constant to maintain isotropic pressure at 1.0 bar.
4. Production simulation without restraints for 100 ns with the Nose-Hoover thermostat and Parrinello-Rahman barostat to maintain temperature and pressure with 2.0 ps and 1.0 ps time constants, respectively, whilst saving frame information every 10 ps.

Long range electrostatics used the Particle Mesh Ewald algorithm (PME) [49]. The Verlet cutoff scheme [50] was used to determine short range non-bonded interactions. Bond lengths were constrained with the LINCS algorithm [51] and water with the SHAKE algorithm [52].

Simulation trajectories were fixed to correct for periodic boundary jumps then the final trajectories were stripped of added water molecules and ions and saved as compressed .xtc files for use with GROMACS tools and MDAnalysis.

2.6 Feature Extraction

Features were extracted from .xtc and relevant topology files using built-in GROMACS tools with the IMGT defined group indexes and removing the first 20 ns of simulation as further equilibration time following AbMelt's approach [35]. The following GROMACS tools listed in Table 2 were used.

Additionally, a number of extra features were computed from the above, namely:

1. Partition-SASA was calculated as the SASA per residue using the Shrake-Rupley algorithm [53] with MDAnalysis then selecting core and surface residues by the 20 residues with the least and most SASA respectively to get core and surface partition-wise SASA.

2. N-H backbone vector order parameter S^2 was calculated as in AbMelt with a block size of 10 ns and excluding prolines [35].

3. The Lambda metric is calculated as the temperature dependence of S^2 by the gradient and r_p fit of $\ln(1 - S)$ vs $\ln(T)$, as described in AbMelt [35]

All feature data was collected for each antibody into a single dataset of both means and stdevs (summarised in Table S3) of each feature at each temperature which was used for all subsequent correlation analyses. The data was further processed for model feature selection and training by filtering out erronous results and replacing any missing values present by imputation with means to preserve the underlying statistics.

2.7 ABangle

ABangle [37] is a tool used for the calculation of antibody Fv region inter-chain angles (Figure **??**). Specifically, by setting a centroid vector between the structural mass centroids of the heavy and light chain it calculates 5 angles: HL, HC1, LC1, HC2, LC2 (Table 3).

To complement the angle analysis we also describe 6 distances based on four points, Hp1 and Lp1 which are each set 10Å from the centroid point of each chain in the direction of vectors HC1 and LC1 respectively, and Hp2 and Lp2 similarly for vectors HC2 and LC2, respectively (Table 4).

Analysis was performed with MDAanalysis on downsampled trajectories for computational efficiency where every 20th frame was converted to a temporary pdb file for processing with ABangle. For details on how ABangle defines the centroid vector and angles see [37].

Table 2: GROMACS Tools Used For Feature Extraction

GROMACS Command	**Feature**	**Region**	**Data Type**	**Units**
`gmx sasa`	Solvent Accessible Surface Area	Fv, H, L, CDRS, CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, CDRL3	time series	nm^2
`gmx mindist`	Internal contacts count	Fv, H, L, CDRS, CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, CDRL3	time series	count
`gmx hbond`	Hydrogen bond count	Fv, H, L	time series	count
`gmx rms`	Root Mean Square deviation	Fv, H, L, CDRS, CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, CDRL3	time series	nm
`gmx gyrate`	Radius of gyration	Fv, H, L, CDRS, CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, CDRL3	time series	nm
`gmx dipole`	Dipole moment	Fv	time series	Debye
`gmx energy`	Energies	Fv	time series	various
`gmx rmsf`	Root Mean Square Fluctuation	Fv, H, L, CDRS, CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, CDRL3	per atom	nm
`gmx covar`	Covariance matrix of atomic fluctuations	Fv	NA	NA
`gmx anaeig`	Eigen vector analysis of covariance matrix	Fv	Eigen values	$nm/S^2/N$
`gmx potential`	Electrostatic potentials, charges and field energy across atom groups	Fv, H, L, CDRS, CDRH1, CDRH2, CDRH3, CDRL1, CDRL2, CDRL3	z axis average coordinate	various

Table 3: ABangle Fv Region Angles

Angle	Definition
HL	Angle between heavy and light chain vectors HC1 and LC1 along the centroid plane bissecting the two chains
HC1	Angle between the HC1 and centroid vectors
LC1	Angle between the LC1 and centroid vectors
HC2	Angle between the HC2 and centroid vectors
LC2	Angle between the LC2 and centroid vectors

Table 4: ABangle Fv Region Distances

Distance	Definition
DC	Length of the centroid vector
PP	Distance between points Hp1 and Lp1
H1	Distance between the point Lp1 and the centroid of chain H
L1	Distance between the point Hp1 and the centroid of chain L
H2	Distance between the point Lp2 and the centroid of chain H
L2	Distance between the point Hp2 and the centroid of chain L

2.8 Martini Coarse Graining

Coarse graining improves computational performance by orders of magnitude, allowing for significantly extended simulation times with minimal loss of accuracy [54]. The Martini3 [54] coarse grain force field splits residues into a backbone bead comprising the backbone atoms and a variable number of beads capturing side chains.

To coarse grain an all-atom PDB, the Martinise tool from the package Vermouth [55] was used with elastic network (harmonic spring) parameters of a $700\,\mathrm{kJ/mol/nm^2}$ spring constant, a lower distance cutoff of 0.5 nm and upper distance cutoff of 0.5 nm. Coarse graining performs best on energetically stable structures which can be obtained from a prior molecular dynamics simulation or with an ensemble of structures where the most abundant conformation is deemed most stable. To avoid further MD adding to computation time, we opted for an ensemble strategy using ImmuneBuilder to generate 100 unique structures for each Fv region. A representative structure was obtained by capturing structural variation within 3D space across the ensemble with Principle Component Analysis (PCA) and clustering then taking the medoid structure of the most populated cluster as a loose approximation of the most stable conformation.

After coarse graining the representative structure the system was solvated with Insane [56]. We used a periodic box size of 7 units, a salt concentration of 0.15 mol/L, and solvated the system with 100 water molecules. As with the all-atom protocol, the system underwent energy minimisation with steepest descent before being equilibrated with NVT and NPT canonical ensembles this time for 200 ps, followed by production in NPT.

Unlike the all-atom configuration which used Nose-Hoover, we use the V-rescale thermostat as we found it more robust for coarse grain systems which don't reproduce exact atomic dynamics. We used a timestep of 10 fs, saving frame information every 10 ps. As with all-atom, the LINCS algorithm and Verlet cutoff schemes were used for bond constraints and neighbour searching respectively. The C-rescale barostat with isotropic pressure coupling was used for pressure scaling.

Since coarse graining removes atomistic resolution a number of analyses are unavailable, such as hydrogen bond analysis which relies on proton acceptor pairs. Thus, to produce the same features analysis the final processed trajectories were converted back to all-atom (finegraining) with the cg2all package [57] and subjected to the same analysis as the all-atom protocol. Due to the nature of energy analysis being performed on the coarse grained trajectory not the cg2all trajectory the following features had to be excluded from martini feature extraction: lj14, ljSR, coloumb14, columbSR, kinetic, potential, and enthalpy (Table S3). Further, due to issues with cg2all sometimes producing trajectory frames with overlapping atoms interfering with MDAnalysis' sasa analysis tool the partition-sasa feature was also excluded.

RMSD analysis of cg2all trajectories was performed after aligning the cg2all trajectory frames and original all-atom frames to the first frame in the all original atom trajectory as reference. RMSD was calculated using MDAnalysis's rmsd function on backbone atoms.

To compare feature values between all-atom and cg2all simulations all missing values were filtered out, and any features not shared between the coarse grain analysis and all-atom were removed. The difference matrix $\mathbf{F}_{diff} \in \mathbb{R}^{n \times m}$ between values for every feature (m) for every antibody (n) was computed with:

$$\mathbf{F}_{diff} = \mathbf{F}_{aa} - \mathbf{F}_{cg2all} \tag{13}$$

Where $\mathbf{F}_{aa} \in \mathbb{R}^{n \times m}$ and $\mathbf{F}_{cg2all} \in \mathbb{R}^{n \times m}$ refer to the matrices of antibody (rows) and feature (cols) values for the all atom and coarse grained (finegrained) simulations respectively.

To gauge the similarity between the two feature datasets the RV coefficient, which captures shared variance and agreement across samples between two datasets, was computed

to give a value between 0 and 1 where 1 indicates the strongest possible overall structural similarity between the two datasets, as follows:

$$\mathrm{RV}(\mathbf{A}, \mathbf{B}) = \frac{\langle \mathbf{A}\mathbf{A}^\top, \mathbf{B}\mathbf{B}^\top \rangle_F}{\|\mathbf{A}\mathbf{A}^\top\|_F \cdot \|\mathbf{B}\mathbf{B}^\top\|_F} \tag{14}$$

Where $\mathbf{A}$ and $\mathbf{B}$ refer to the mean centered $\mathbf{F}_{aa}$ and $\mathbf{F}_{cg2all}$ per antibody feature matrices for the all-atom and coarse grained sets respectively. Mean centering was performed by subtracting the row mean of each row in the matrix from every value in the respective row given rows are samples (antibodies) and columns features.

2.9 Machine Learning on Descriptor Features

Machine learning on MD features was performed as follows. First, highly inter-correlated features (Pearson r_p coefficient > 0.95) were removed. Recursive followed by exhaustive feature selection with CV was performed using the 3-fold train validation splits and the scikit-learn [58] package's RFECV and EFSCV modules with its Random Forest Regressor as the model fit metric whilst optimising hyperparameters with Optuna [59] using 50 trials. This was performed separately for each developability feature from the Jain et al. set.

For each of the developability fitness features, six models were trained from their selected MD descriptor feature subsets (Table 5). During training the same CV split as feature selection was used to optimise hyperparameters and select the optimal performing model before testing.

Each trained model for each developability fitness feature was tested on the holdout test split. Accuracy was computed with three metrics between the predicted value and the ground truth from the Jain et al. set: The coefficient of determination (r^2), Pearson's coefficient (r_p) and Spearman's coefficient (ρ) as below:

$$r^2 = (r_p)^2 \tag{15}$$

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{16}$$

Table 5: Table of models used in training on each developability assay metric (fitness feature) selected features subset

Model	**Description**	**Package**
Elastic Net	Regression with L1 and L2 regression for high dimensional data where the number of predictors is > the number of samples	Scikit-learn
K Nearest Neighbours (KNN)	Non-parametric algorithm that takes the average of K closest neighbours as the regression prediction	Scikit-learn
Linear regression	Models the optimal linear relationship between predictors and an outcome	Scikit-learn
Random forest	Ensemble of decision trees where each tree is trained on a random subset of the data then takes the average of tree predictions as regression output	Scikit-learn
Support Vector Machine (SVM)	Finds a function that approximates outcomes within a margin of tolerance	Scikit-learn
XG Boost	Sequential tree building with L1 and L2 regularisation minimising a loss function that corrects errors from previous trees to find an optimal objective function	Scikit-learn

References

[1] Audrey D McConnell et al. "A General Approach to Antibody Thermostabilization". In: *mAbs* 6.5 (Sept. 2014), pp. 1274–1282. ISSN: 1942-0862. DOI: 10.4161/mabs.29680. (Visited on 07/29/2025).

[2] Shuai Wang et al. "Early Determination of Potential Critical Quality Attributes of Therapeutic Antibodies in Developability Studies through Surface Plasmon Resonance-Based Relative Binding Activity Assessment". In: *mAbs* 16.1 (Dec. 2024), p. 2374607. ISSN: 1942-0862, 1942-0870. DOI: 10.1080/19420862.2024.2374607. (Visited on 08/05/2025).

[3] Tushar Jain, Todd Boland, and Maximiliano Vásquez. "Identifying Developability Risks for Clinical Progression of Antibodies Using High-Throughput in Vitro and in Silico Approaches". In: *mAbs* 15.1 (), p. 2200540. ISSN: 1942-0862. DOI: 10.1080/19420862.2023.2200540. (Visited on 08/13/2025).

[4] Drago Kuzman et al. "Long-Term Stability Predictions of Therapeutic Monoclonal Antibodies in Solution Using Arrhenius-based Kinetics". In: *Scientific Reports* 11 (Oct. 2021), p. 20534. ISSN: 2045-2322. DOI: 10.1038/s41598-021-99875-9. (Visited on 07/29/2025).

[5] Hui Ma, Ciarán Ó'Fágáin, and Richard O'Kennedy. "Antibody Stability: A Key to Performance - Analysis, Influences and Improvement". In: *Biochimie* 177 (Oct. 2020), pp. 213–225. ISSN: 0300-9084. DOI: 10.1016/j.biochi.2020.08.019. (Visited on 07/29/2025).

[6] John Jumper et al. "Highly Accurate Protein Structure Prediction with AlphaFold". In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. (Visited on 08/12/2025).

[7] Brennan Abanades et al. "ImmuneBuilder: Deep-Learning Models for Predicting the Structures of Immune Proteins". In: *Communications Biology* 6.1 (May 2023), p. 575. ISSN: 2399-3642. DOI: 10.1038/s42003-023-04927-7. (Visited on 07/29/2025).

[8] Joseph L. Watson et al. "De Novo Design of Protein Structure and Function with RFdiffusion". In: *Nature* 620.7976 (Aug. 2023), pp. 1089–1100. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06415-8. (Visited on 07/29/2025).

[9] Luiz Felipe Vecchietti et al. "Artificial Intelligence-Driven Computational Methods for Antibody Design and Optimization". In: *mAbs* 17.1 (Dec. 2025), p. 2528902. ISSN: 1942-0862. DOI: 10.1080/19420862.2025.2528902. (Visited on 08/03/2025).

[10] Ralph Adams et al. "Bimekizumab, a Novel Humanized IgG1 Antibody That Neutralizes Both IL-17A and IL-17F". In: *Frontiers in Immunology* 11 (Aug. 2020), p. 1894. ISSN: 1664-3224. DOI: 10.3389/fimmu.2020.01894. (Visited on 07/29/2025).

[11] Sophia Frentzas et al. “A Phase 1/2 Study of AU-007, a Monoclonal Antibody (mAb) That Binds to IL-2 and Inhibits CD25 Binding, in Patients with Advanced Solid Tumors: Interim Results.” In: *Journal of Clinical Oncology* 41.16_suppl (June 2023), e14507–e14507. ISSN: 0732-183X. DOI: 10.1200/JCO.2023.41.16_suppl.e14507. (Visited on 07/29/2025).

[12] Ieva Pudžiuvelytė et al. “TemStaPro: Protein Thermostability Prediction Using Sequence Representations from Protein Language Models”. In: *Bioinformatics* 40.4 (Apr. 2024), btae157. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btae157. (Visited on 08/03/2025).

[13] Chiara Rodella, Symela Lazaridi, and Thomas Lemmin. “TemBERTure: Advancing Protein Thermostability Prediction with Deep Learning and Attention Mechanisms”. In: *Bioinformatics Advances* 4.1 (Jan. 2024), vbae103. ISSN: 2635-0041. DOI: 10.1093/bioadv/vbae103. (Visited on 08/03/2025).

[14] Mark Hutchinson et al. “Toward Enhancement of Antibody Thermostability and Affinity by Computational Design in the Absence of Antigen”. In: *mAbs* 16.1 (Dec. 2024), p. 2362775. ISSN: 1942-0862. DOI: 10.1080/19420862.2024.2362775. (Visited on 07/29/2025).

[15] Tobias H. Olsen, Fergus Boyles, and Charlotte M. Deane. “Observed Antibody Space: A Diverse Database of Cleaned, Annotated, and Translated Unpaired and Paired Antibody Sequences”. In: *Protein Science: A Publication of the Protein Society* 31.1 (Jan. 2022), pp. 141–146. ISSN: 1469-896X. DOI: 10.1002/pro.4205.

[16] James Dunbar et al. “SAbDab: The Structural Antibody Database”. In: *Nucleic Acids Research* 42.D1 (Jan. 2014), pp. D1140–D1146. ISSN: 0305-1048. DOI: 10.1093/nar/gkt1043. (Visited on 08/05/2025).

[17] Michael Chungyoun, Jeffrey Ruffolo, and Jeffrey Gray. *FLAb: Benchmarking Deep Learning Methods for Antibody Fitness Prediction*. Jan. 2024. DOI: 10.1101/2024.01.13.575504. (Visited on 07/30/2025).

[18] Alissa M. Hummer et al. “Investigating the Volume and Diversity of Data Needed for Generalizable Antibody–Antigen $\Delta\Delta G$ Prediction”. In: *Nature Computational Science* (July 2025), pp. 1–13. ISSN: 2662-8457. DOI: 10.1038/s43588-025-00823-8. (Visited on 08/06/2025).

[19] Stephane Boivin, Sandra Kozak, and Rob Meijers. “Optimization of Protein Purification and Characterization Using Thermofluor Screens”. In: *Protein Expression and Purification* 91.2 (Oct. 2013), pp. 192–206. ISSN: 1046-5928. DOI: 10.1016/j.pep.2013.08.002. (Visited on 08/13/2025).

[20] Negin Gooran and Kari Kopra. “Fluorescence-Based Protein Stability Monitoring—A Review”. In: *International Journal of Molecular Sciences* 25.3 (Feb. 2024), p. 1764. ISSN: 1422-0067. DOI: 10.3390/ijms25031764. (Visited on 08/13/2025).

[21] Srishti Joshi, Chinmoyee Maharana, and Anurag S. Rathore. "An Application of Nano Differential Scanning Fluorimetry for Higher Order Structure Assessment between mAb Originator and Biosimilars: Trastuzumab and Rituximab as Case Studies". In: *Journal of Pharmaceutical and Biomedical Analysis* 186 (July 2020), p. 113270. ISSN: 0731-7085. DOI: 10.1016/j.jpba.2020.113270. (Visited on 07/29/2025).

[22] Nan Bai et al. "Isothermal Analysis of ThermoFluor Data Can Readily Provide Quantitative Binding Affinities". In: *Scientific Reports* 9.1 (Feb. 2019), p. 2650. ISSN: 2045-2322. DOI: 10.1038/s41598-018-37072-x. (Visited on 08/13/2025).

[23] Taiasean Wu, Zachary J. Gale-Day, and Jason E. Gestwicki. "DSFworld: A Flexible and Precise Tool to Analyze Differential Scanning Fluorimetry Data". In: *Protein Science : A Publication of the Protein Society* 33.6 (May 2024), e5022. ISSN: 0961-8368. DOI: 10.1002/pro.5022. (Visited on 08/13/2025).

[24] Harumi Fukada et al. "Long-Term Stability and Reversible Thermal Unfolding of Antibody Structure at Low pH: Case Study". In: *Journal of Pharmaceutical Sciences* 107.11 (Nov. 2018), pp. 2965–2967. ISSN: 0022-3549. DOI: 10.1016/j.xphs.2018.07.001. (Visited on 07/29/2025).

[25] Francesca Peccati, Sara Alunno-Rufini, and Gonzalo Jiménez-Osés. "Accurate Prediction of Enzyme Thermostabilization with Rosetta Using AlphaFold Ensembles". In: *Journal of Chemical Information and Modeling* 63.3 (Feb. 2023), pp. 898–909. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.2c01083. (Visited on 08/05/2025).

[26] Feixiang Zhou et al. "ProCeSa: Contrast-Enhanced Structure-Aware Network for Thermostability Prediction with Protein Language Models". In: *Journal of Chemical Information and Modeling* 65.5 (Mar. 2025), pp. 2304–2313. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.4c01752. (Visited on 08/05/2025).

[27] Moeka Yoshikawa et al. "Analysis of Thermostability for Seven Phe to Ala and Six Pro to Gly Mutants in the Fab Constant Region of Adalimumab". In: *The Journal of Biochemistry* 174.4 (Sept. 2023), pp. 345–353. ISSN: 0021-924X. DOI: 10.1093/jb/mvad047. (Visited on 07/29/2025).

[28] Moeka Yoshikawa et al. "Stabilization of Adalimumab Fab through the Introduction of Disulfide Bonds between the Variable and Constant Domains". In: *Biochemical and Biophysical Research Communications* 700 (Mar. 2024), p. 149592. ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2024.149592. (Visited on 07/29/2025).

[29] Ameya Harmalkar et al. "Toward Generalizable Prediction of Antibody Thermostability Using Machine Learning on Sequence and Structure Features". In: *mAbs* 15.1 (), p. 2163584. ISSN: 1942-0862. DOI: 10.1080/19420862.2022.2163584. (Visited on 08/03/2025).

[30] Seungyeon Choi et al. "PIDiff: Physics Informed Diffusion Model for Protein Pocket-Specific 3D Molecular Generation". In: *Computers in Biology and Medicine* 180 (Sept. 2024), p. 108865. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2024.108865. (Visited on 07/29/2025).

[31] Xinyue Cui et al. "Beyond Static Structures: Protein Dynamic Conformations Modeling in the Post-AlphaFold Era". In: *Briefings in Bioinformatics* 26.4 (July 2025), bbaf340. ISSN: 1477-4054. DOI: 10.1093/bib/bbaf340. (Visited on 08/12/2025).

[32] Sarah Lewis et al. *Scalable Emulation of Protein Equilibrium Ensembles with Generative Deep Learning*. Dec. 2024. DOI: 10.1101/2024.12.05.626885. (Visited on 08/03/2025).

[33] Alan Grossfield and Daniel M. Zuckerman. "Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations". In: *Annual reports in computational chemistry* 5 (Jan. 2009), pp. 23–48. ISSN: 1574-1400. DOI: 10.1016/S1574-1400(09)00502-7. (Visited on 08/13/2025).

[34] Gert-Jan Bekker, Benson Ma, and Narutoshi Kamiya. "Thermal Stability of Single-domain Antibodies Estimated by Molecular Dynamics Simulations". In: *Protein Science : A Publication of the Protein Society* 28.2 (Feb. 2019), pp. 429–438. ISSN: 0961-8368. DOI: 10.1002/pro.3546. (Visited on 08/05/2025).

[35] Zachary A. Rollins et al. "AbMelt: Learning Antibody Thermostability from Molecular Dynamics". In: *Biophysical Journal* 123.17 (Sept. 2024), pp. 2921–2933. ISSN: 0006-3495, 1542-0086. DOI: 10.1016/j.bpj.2024.06.003. (Visited on 07/29/2025).

[36] Tushar Jain et al. "Biophysical Properties of the Clinical-Stage Antibody Landscape". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.5 (Jan. 2017), pp. 944–949. ISSN: 0027-8424. DOI: 10.1073/pnas.1616408114. (Visited on 07/27/2025).

[37] J. Dunbar et al. "ABangle: Characterising the VH–VL Orientation in Antibodies". In: *Protein Engineering, Design and Selection* 26.10 (Oct. 2013), pp. 611–620. ISSN: 1741-0126. DOI: 10.1093/protein/gzt020. (Visited on 07/29/2025).

[38] Kresten Lindorff-Larsen et al. "Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation". In: *Journal of the American Chemical Society* 134.8 (Feb. 2012), pp. 3787–3791. ISSN: 1520-5126. DOI: 10.1021/ja209931w.

[39] Peter J. A. Cock et al. "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics". In: *Bioinformatics* 25.11 (June 2009), pp. 1422–1423. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp163. (Visited on 08/13/2025).

[40] Naveen Michaud-Agrawal et al. "MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations". In: *Journal of Computational Chemistry* 32.10 (2011), pp. 2319–2327. ISSN: 1096-987X. DOI: 10.1002/jcc.21787. (Visited on 08/13/2025).

[41] Matthew I J Raybould et al. "Thera-SAbDab: The Therapeutic Structural Antibody Database". In: *Nucleic Acids Research* 48.D1 (Jan. 2020), pp. D383–D388. ISSN: 0305-1048. DOI: 10.1093/nar/gkz827. (Visited on 07/29/2025).

[42] James Dunbar and Charlotte M. Deane. "ANARCI: Antigen Receptor Numbering and Receptor Classification". In: *Bioinformatics* 32.2 (Jan. 2016), pp. 298–300. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv552. (Visited on 07/29/2025).

[43] Peter Eastman et al. "OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials". In: *The Journal of Physical Chemistry B* 128.1 (Jan. 2024), pp. 109–116. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.3c06662. (Visited on 07/29/2025).

[44] Mark Abraham et al. *GROMACS 2025.1 Source Code*. Zenodo. Mar. 2025. DOI: 10.5281/zenodo.15006630. (Visited on 08/13/2025).

[45] Mats H. M. Olsson et al. "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions". In: *Journal of Chemical Theory and Computation* 7.2 (Feb. 2011), pp. 525–537. ISSN: 1549-9618. DOI: 10.1021/ct100578z. (Visited on 07/29/2025).

[46] Chresten R. Søndergaard et al. "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values". In: *Journal of Chemical Theory and Computation* 7.7 (July 2011), pp. 2284–2295. ISSN: 1549-9618. DOI: 10.1021/ct200133y. (Visited on 07/29/2025).

[47] Alexander D. Mackerell, Michael Feig, and Charles L. Brooks. "Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations". In: *Journal of Computational Chemistry* 25.11 (Aug. 2004), pp. 1400–1415. ISSN: 0192-8651. DOI: 10.1002/jcc.20065.

[48] William L. Jorgensen et al. "Comparison of Simple Potential Functions for Simulating Liquid Water". In: *The Journal of Chemical Physics* 79.2 (July 1983), pp. 926–935. ISSN: 0021-9606. DOI: 10.1063/1.445869. (Visited on 08/05/2025).

[49] Tom Darden, Darrin York, and Lee Pedersen. "Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems". In: *The Journal of Chemical Physics* 98.12 (June 1993), pp. 10089–10092. ISSN: 0021-9606. DOI: 10.1063/1.464397. (Visited on 08/03/2025).

[50] Szilárd Páll and Berk Hess. "A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures". In: *Computer Physics Communications* 184.12 (Dec. 2013), pp. 2641–2650. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2013.06.003. (Visited on 08/03/2025).

[51] Berk Hess et al. “LINCS: A Linear Constraint Solver for Molecular Simulations”. In: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472. ISSN: 1096-987X. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H. (Visited on 08/03/2025).

[52] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C Berendsen. “Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes”. In: *Journal of Computational Physics* 23.3 (Mar. 1977), pp. 327–341. ISSN: 0021-9991. DOI: 10.1016/0021-9991(77)90098-5. (Visited on 08/03/2025).

[53] A. Shrake and J. A. Rupley. “Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin”. In: *Journal of Molecular Biology* 79.2 (Sept. 1973), pp. 351–371. ISSN: 0022-2836. DOI: 10.1016/0022-2836(73)90011-9.

[54] Paulo C. T. Souza et al. “Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics”. In: *Nature Methods* 18.4 (Apr. 2021), pp. 382–388. ISSN: 1548-7105. DOI: 10.1038/s41592-021-01098-3. (Visited on 07/29/2025).

[55] Peter C. Kroon et al. *Martinize2 and Vermouth: Unified Framework for Topology Generation*. May 2025. DOI: 10.48550/arXiv.2212.01191. arXiv: 2212.01191 [q-bio]. (Visited on 07/29/2025).

[56] Tsjerk A. Wassenaar et al. “Computational Lipidomics with Insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations”. In: *Journal of Chemical Theory and Computation* 11.5 (May 2015), pp. 2144–2155. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.5b00209. (Visited on 08/06/2025).

[57] Lim Heo and Michael Feig. *One Particle per Residue Is Sufficient to Describe All-Atom Protein Structures*. May 2023. DOI: 10.1101/2023.05.22.541652. (Visited on 07/29/2025).

[58] Fabian Pedregosa et al. “Scikit-Learn: Machine Learning in Python”. In: *J. Mach. Learn. Res.* 12.null (Nov. 2011), pp. 2825–2830. ISSN: 1532-4435.

[59] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD ’19. New York, NY, USA: Association for Computing Machinery, July 2019, pp. 2623–2631. ISBN: 978-1-4503-6201-6. DOI: 10.1145/3292500.3330701. (Visited on 08/13/2025).

Supplementary Information

Table S1: PDB structures for antibodies from the Jain et al. set retrieved from RCSB PDB

Name	**PDB ID**	**Chain IDs**
Adalimumab	6cr1	HL
Alemtuzumab	1bey	HL
Anifrolumab	4qxg	HL
Atezolizumab	5x8l	FK
Basiliximab	1mim	HL
Belimumab	5y9k	HL
Bevacizumab	1bj1	HL
Bimagrumab	5nhr	HL
Bococizumab	3sqo	HL
Briakinumab	5n2k	LK
Certolizumab	5wuv	HL
Cetuximab	6au5	DC
Crenezumab	5vzx	HL
Dacetuzumab	8yx9	HL
Daclizumab	3nfs	HL
Daratumumab	7dun	HL
Drozitumab	4od2	BA
Dupilumab	6wgb	AB
Eculizumab	5i5k	HL
Efalizumab	3eo9	HL
Gantenerumab	5csz	HL
Gemtuzumab	8w70	BA
Gevokizumab	4g6k	HL
Golimumab	5yoy	RO
Guselkumab	4m6m	HL
Ibalizumab	3o2d	HL
Infliximab	5vh3	HL
Ipilimumab	5tru	HL
Ixekizumab	6nov	AB
Lampalizumab	4d9q	HL
Lebrikizumab	4i77	HL
Lirilumab	8tui	HL

Continued on next page

Name	PDB ID	Chain IDs
Matuzumab	3c08	HL
Motavizumab	3qwo	AB
Muromonab	1sy6	HL
Necitumumab	6b3s	JK
Nivolumab	5ggq	HL
Obinutuzumab	3pp3	HL
Ofatumumab	3giz	HL
Olokizumab	4cni	HL
Omalizumab	4x7s	HL
Onartuzumab	4k3j	HL
Panitumumab	5sx5	HL
Pembrolizumab	5dk3	GF
Pertuzumab	1l7i	HL
Ponezumab	3u0t	BA
Radretumab	7ah1	AA
Ramucirumab	3s34	HL
Ranibizumab	1cz8	HL
Rituximab	2osl	HL
Sarilumab	8iow	HL
Secukinumab	6wio	AB
Sifalimumab	4ypg	HL
Tanezumab	4edw	HL
Tocilizumab	8j6f	HL
Tralokinumab	5l6y	HL
Trastuzumab	6bi0	HL
Tremelimumab	5ggu	HL
Urelumab	6mhr	AB
Ustekinumab	3hmw	HL
Canakinumab	4g5z	HL
Carlumab	4dn3	HL
Epratuzumab	5vkk	AB
Fresolimumab	3eo0	DC
Fulranumab	4m6o	HL
Lumiliximab	3fzu	HL
Natalizumab	4irz	HL
Nimotuzumab	3gkw	HL
Palivizumab	2hwz	HL
Pinatuzumab	6and	HL

Table S2: Jain et al. Antibody developability metrics

Metric	**Description**
Cross-Interaction Chromatography (CIC)	Measures non-specificity and propensity for aggregation against polyclonal human serum antibodies
PSR Binding Assay	Measures polyspecificity by flow cytometry with biotinylated human membrane proteins
CSI-BLI	Measures antibody aggregation propensity with itself
AC-SINS	Another method for antibody self-aggregation propensity
SGAC-SINS	AC-SINS under high salt conditions
Tm (DSF)	Melting temperature measurement using differential scanning fluorometry (DSF)
SMAC	Aggregation, precipitation and solubility assessment
AS SEC Slope	Aggregation and self-association
HIC	Measures surface hydrophobicity of antibodies for aggregation and solubility assessment
BVP Assay	Polyspecificity, immunogenicity and pharmocokinetics by assessing binding to Baculovirus Particles
ELISA	Measures antibody expression levels

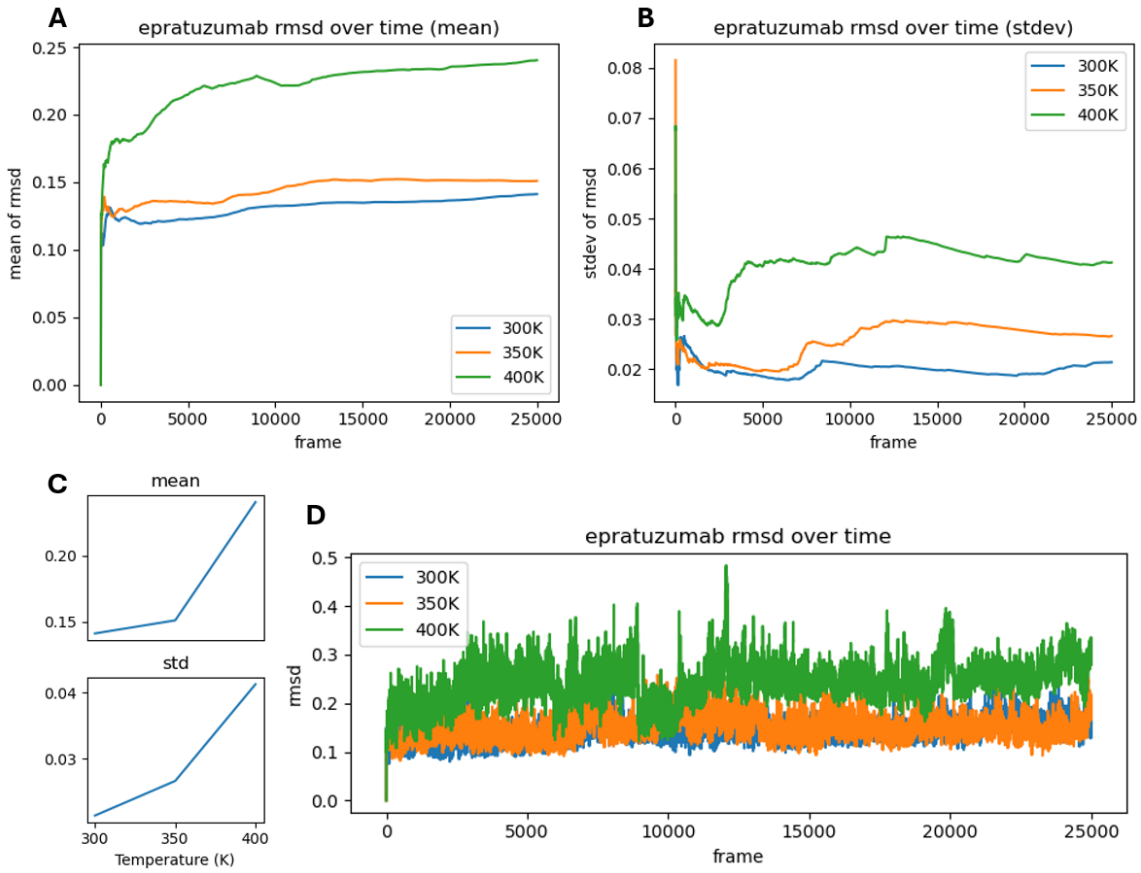

Figure S1: RMSD of epratuzumab all-atom simulation at 300K, 350K, and 400K for 250 ns. Each frame in the x axis represents a 10 ps*timestep*. Dashed red vertical lines mark the 100 ns timepoint. (A) Cumulative mean of the RMSD over the trajectory. (B) Cumulative Stdev of the RMSD over the trajectory. (C) RMSD final frame mean (top) and stdev (bottom) values for each temperature. (D) RMSD over the trajectory.

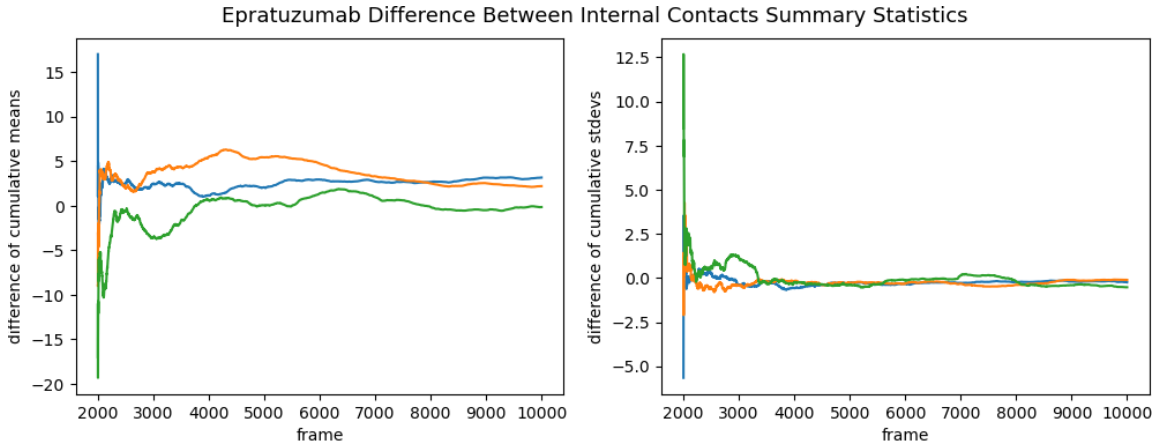

Figure S2: Difference between cumulative summary statistics of the internal contacts of the 100 ns (post-20 ns equilibration cuttoff) and first 80 ns (post-20 ns equilibration cuttoff) of the 250 ns epratuzumab simulations. (Left) difference between cumulative means. (Right) difference between cumulative stdevs.

Table S3: Table listing all descriptor features extracted from MD simulations in this work. The "Group Name" column groups the individual features that each examine a respective part of the group feature.

Group Name	**Feature**
contacts	contacts mean
contacts	contacts std
sasa	sasa mean
sasa	sasa std
partition-sasa	partition-sasa mean
partition-sasa	partition-sasa std
per-res-s2	per-res-s2 mean
per-res-s2	per-res-s2 std
per-block-s2	per-block-s2 mean
per-block-s2	per-block-s2 std
angles	HLangle mean
angles	HLangle std
angles	HC1angle mean
angles	HC1angle std
angles	LC1angle mean
angles	LC1angle std
angles	HC2angle mean
angles	HC2angle std
angles	LC2angle mean
angles	LC2angle std
distances	DCdistance mean
distances	DCdistance std
distances	PPdistance mean
distances	PPdistance std
distances	H1fromCLdistance mean
distances	H1fromCLdistance std
distances	L1fromCHdistance mean
distances	L1fromCHdistance std
distances	H2fromCLdistance mean
distances	H2fromCLdistance std
distances	L2fromCHdistance mean
distances	L2fromCHdistance std
bonds	hbonds mean
bonds	hbonds std

Continued on next page

Group Name	Feature
covar	covar mean
covar	covar std
rmsd	rmsd mean
rmsd	rmsd std
gyr	Rgyr mean
gyr	Rgyr std
gyr	Rgyr Xaxis mean
gyr	Rgyr Xaxis std
gyr	Rgyr Yaxis mean
gyr	Rgyr Yaxis std
gyr	Rgyr Zaxis mean
gyr	Rgyr Zaxis std
dipole	dipole X mean
dipole	dipole X std
dipole	dipole Y mean
dipole	dipole Y std
dipole	dipole Z mean
dipole	dipole Z std
dipole	dipole T mean
dipole	dipole T std
lj14	LJpotential 1-4 mean
lj14	LJpotential 1-4 std
ljSR	LJpotential short-range mean
ljSR	LJpotential short-range std
coloumb14	coulomb 1-4 mean
coloumb14	coloumb 1-4 std
coloumbSR	coulomb short-range mean
coloumbSR	coloumb short-range std
potential	potential energy mean
potential	potential energy std
kinetic	kinetic energy mean
kinetic	kinetic energy std
energy	total energy mean
energy	total energy std
temperature	temperature mean
temperature	temperature std
pv	pressur volume mean
pv	pressure volume std

Continued on next page

Group Name	**Feature**
enthalpy	enthalpy mean
enthalpy	enthalpy std
charge	charge mean
charge	charge std
rmsf	rmsf mean
rmsf	rmsf std

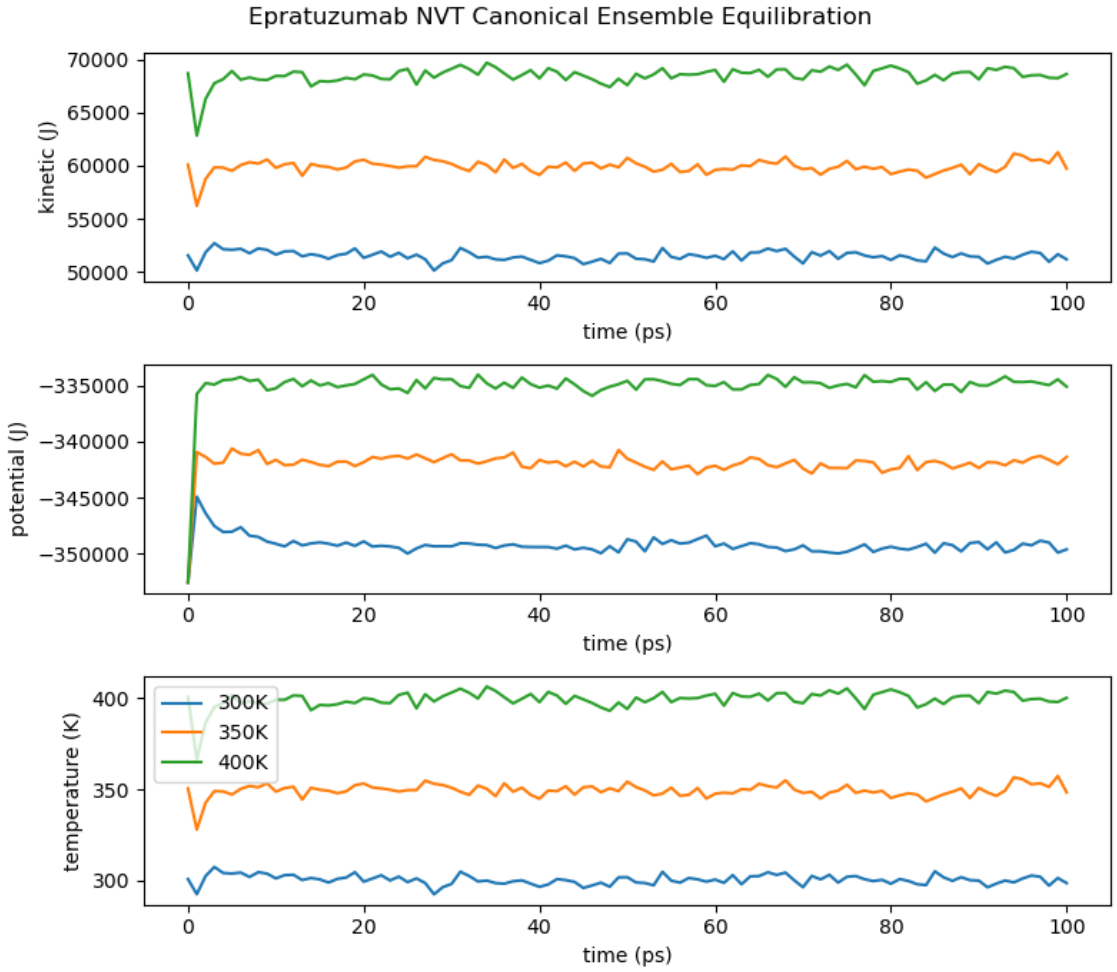

Figure S3: Epratuzumab Martini3 coarse grained equilibration is stable for 100 ps with the NVT canonical ensemble at each temperature. (A) Kinetic energy. (B) Potential energy. (C) Temperature.

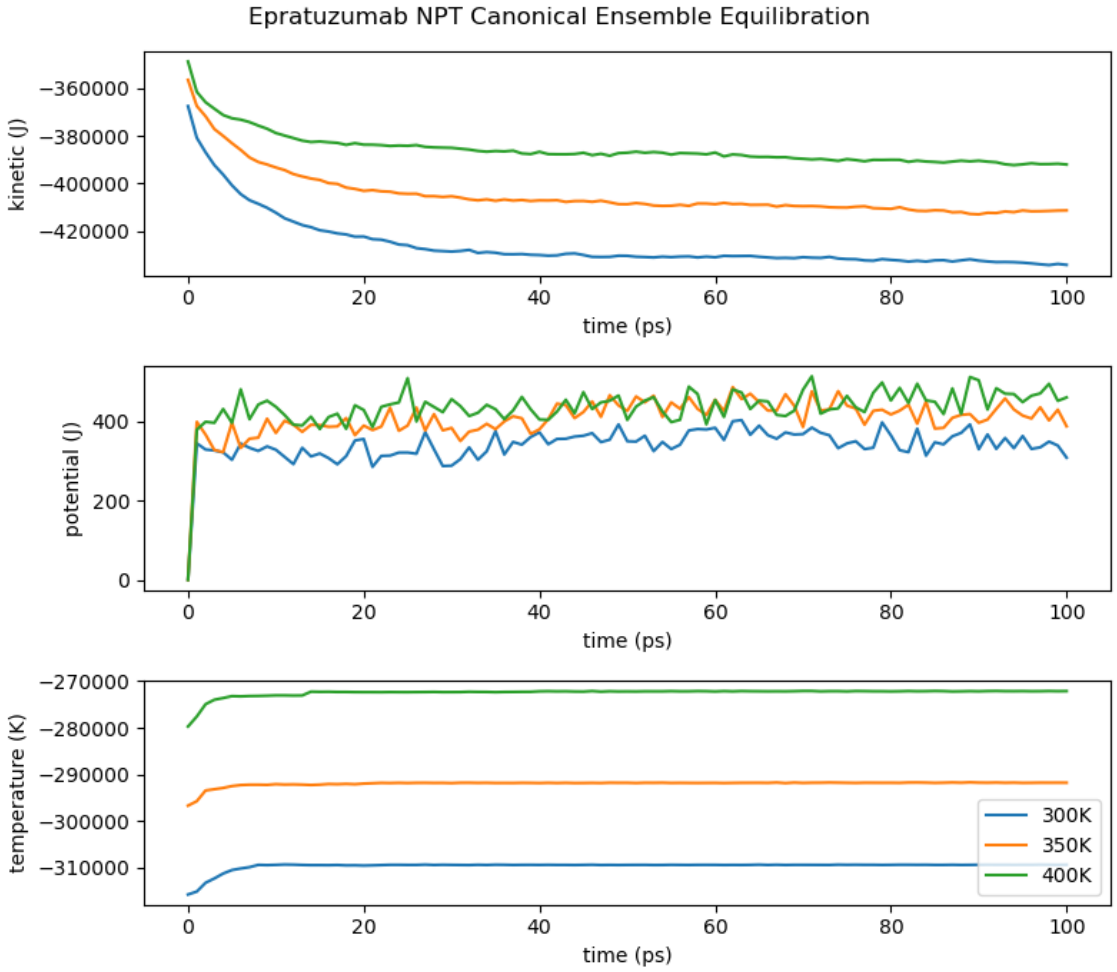

Figure S4: Epratuzumab Martini3 coarse grained equilibration converges over 100 ps with the NPT canonical ensemble at each temperature. (A) Kinetic energy. (B) Potential energy. (C) Temperature.