Assessing and engineering antibody stability using experimental and computational methods

Cheng Zhang and Paul Dalby*

Department of Biochemical Engineering, University College London, Gordon Street, London, WC1E 7JE, UK

* Corresponding Author: p.dalby@ucl.ac.uk

1   Abstract

Engineering increased stability into antibodies can improve their developability. While a range of properties need to be optimised, thermal stability and aggregation are two key factors that affect the antibody yield, purity and specificity throughout the development and manufacturing pipeline. Therefore, an ideal goal would be to apply protein engineering methods early-on, such as in parallel to affinity maturation, to screen out potential drug molecules with the desired conformational and colloidal stability. This chapter introduces our methods to computationally characterise an antibody Fab fragment, propose stabilising variants, and then experimentally verify these predictions.

1.1   Key words

Antibody, Fab, Rosetta, Molecular dynamics (Gromacs), B-factor, T_m, Entropy, Aggregation

2   Introduction

The antibody market has doubled in the past 5 years[1]. Monoclonal antibodies (mAb) continued to be the leading drug candidates in clinical trials, while other formats of antibodies were emerging, including bispecific antibodies (BsAb) and antibody fragments (Fab)[2, 3]. The developability of antibodies contributes significantly to the time and cost required to convert lead molecules into a marketable product. Multiple properties, such as their expression level, solubility, thermal stability and aggregation propensity, are addressed in early-stage development so as to minimise the risk in late large-scale manufacture, as well as during pharmacokinetics and pharmacodynamics studies[4, 5].

Both directed evolution and rational design strategies, have been applied to improve a range of essential antibody properties, with aggregation and thermal stability considered as the most critical parameters[6]. To aid in these approaches, the aggregation-prone regions (APRs) can be estimated by a range of sequence or structure-based algorithms (e.g. PASTA, TANGO, Aggrescan)[7, 8], or using a consensus of several of these, such as via AmylPred[9]. For rational design in particular, the change in thermal stability due to point mutations, determined as the change in Gibbs free energy of unfolding ($\Delta\Delta G$), can be predicted with algorithms that use physical-based potentials, knowledge-based potentials, or a combination of both, to evaluate the protein stability according to their structure[10, 11]. In addition, molecular dynamics has matured significantly, to provide detailed atomistic understanding of protein dynamics and

interactions that contribute to stability, at any defined pH, temperature and ionic strength[(12, 13)]. The wider adoption of these approaches for guiding the rational design of proteins, will improve as their accuracy evolves. Towards this goal, the predictions from any *in-silico* ("dry lab") modelling need to be validated by appropriate *in-vitro* ("wet lab") experiments[(14)]. The *in-vitro* data could then be used by developers to further optimise the parameters in the *in-silico* predictions.

The Fab A33 is an anti-TNF-α antibody fragment, similar to those finally approved to treat colorectal cancers[(15–17)]. As a non-commercialised variant, Fab A33 provides a good model platform for protein engineering, and this was recently conducted to reduce its aggregation propensity[(18)]. This chapter aims to elaborate on the procedures used, and to provide more detailed insights that others can then use. Fab A33 has a relatively high thermostability (T_m ca. 70 °C), but any therapeutic formulation must also meet highly stringent standards for aggregation kinetics on the order of <1% per year of storage[(19–21)]. It was hypothesised that as the protein would be fully native under typical storage conditions (4-25 °C), then the local dynamics of the native protein could be a highly influential factor for aggregation. Thus, our aim was to explore the design of rigidifying mutations in particularly flexible regions, and examine their impact on both thermostability and aggregation kinetics. *RosettaCM* was used to construct a reliable starting structure with complete residues, followed by Rosetta *relax* to refine the local conformational space. Next, molecular dynamics simulations and crystallographic B-factors were collectively used to identify the most flexible sites of the Fab. Rosetta *cartesian_ddg* was then used to predict mutations at these sites with the greatest $\Delta\Delta G$, and therefore a likelihood of reducing local flexibility. Several predicted variants were made, and their T_m, unfolding entropy and aggregation kinetics determined experimentally. The stabilising mutations were mapped onto the Fab structure, for comparison against the heavy chain hinge region which was identified as the most flexible region. The substitution of suboptimal residues within the hinge region were found to greatly rigidify the native ensemble, and also suppress the aggregation kinetics. Due to the focus of this chapter, some of the procedures will not be fully covered, but can be found in the previous paper[(18)]. All commands, file names, directories and terminologies are formatted in italics.

3   Materials and Methods

3.1   Using *RosettaCM* to obtain a starting structure model

The *RosettaCM* tutorial (See Note 4.2.2) was referred to in designing this protocol, with modifications made to customise it for Fab A33.

3.1.1   Alignment

Several residues were not resolved in the original crystal structure used (*FAB_ChainB&L.pdb*, SI)[(22)], including residue 214 in the light chain, and residues 315, 346-351, and 432-442 in the heavy chain. Furthermore, residue E215 in the heavy chain was only partially resolved and so incorrectly assigned as A215. Since a full-length PDB is required in the molecular modelling and simulation, *RosettaCM*[(23)] was used to fill the missing residues followed by energy minimisation.

We used the raw crystal structure as the template structure. A fasta file *FabMissResi.fasta* (See Note 4.2.2) can be generated (https://zhanglab.ccmb.med.umich.edu/pdb2fasta/) based on the raw crystal structure, and by inserting the chain separator "/" to combine the two chains together.

The full-length Fab sequence (See Note 4.2.4) is aligned to the template structure sequence *FabMissResi.fasta* at https://www.ebi.ac.uk/Tools/msa/clustalo/, and the job submitted with default settings. The alignment can then be downloaded to a file called *Fab_FabMissResi_raw.aln* showing "-" as the missing residues (SI). Because the last residue cysteine in the LC of the crystal structure is missing, the alignment incorrectly places the alanine as the last residue of LC, as *clustalo* could not recognise the chain separator "/".

```
Fab            LSKADYEKHKVYACEVTHQGLSSPVTKSFNRGECEVQLVESGGGLVQPGGSLRLSCAASG    240
FabMissResi    LSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEA-VQLVESGGGLVQPGGSLRLSCAASG    239
               *******************************.  .************************
```

This needs to be adjusted as

```
Fab            LSKADYEKHKVYACEVTHQGLSSPVTKSFNRGECEVQLVESGGGLVQPGGSLRLSCAASG    240
FabMissResi    LSKADYEKHKVYACEVTHQGLSSPVTKSFNRGE-AVQLVESGGGLVQPGGSLRLSCAASG    239
               *******************************.  ************************
```

The *Fab_FabMissResi_raw.aln* was further edited as *Fab_FabMissResi_adjusted.aln* (SI) to remove the unnecessary stars (*).

3.1.2 Fragment files

Fragment files are used as the "building blocks" to construct the target protein structure based on the template structure. They are particularly used to predict any unresolved regions in the template structure.

Three-residue (3MER) and nine-residue (9MER) libraries were generated at the robetta.bakerlab.org.

- After signing in, go to *robetta.bakerlab.org/fragmentsubmit.jsp*.
- Put *Fab* under *Target name*.
- Copy/paste all text from *Fab.fasta* into the provided field.
- Click *Submit* and wait for the job to complete
- Once completed, fragment files can be downloaded and should be saved as *Fab_3.frags* and *Fab_9.frags* (SI).

3.1.3 Prepare the alignment in grishin format

Edit the *Fab_FabMissResi_adjusted.aln* and save it as *Fab_FabMissResi.grishin* (SI). There is no need to add the chain separator "/". The "/" will always be automatically added in the subsequent *FabMissResi.pdb.pdb* [*sic*] file (first line).

3.1.4 Thread the target sequence over the template sequence

run command

- *<rosetta_directory>/main/source/bin/partial_thread.default.linuxgccrelease -in:file:fasta Fab.fasta -in:file:alignment Fab_FabMissResi.grishin -in:file:template_pdb FabMissResi.pdb*

output

- *FabMissResi.pdb.pdb*, rename it as *Fab_on_FabMissResi.pdb*
- In the *Fab_on_FabMissResi.pdb* file, change the chain ID of the heavy chain as *B* or *H*. Visualise it in Pymol, make sure there is no connection between the end of light chain and the head of heavy chain.

3.1.5   Obtain the weights files

As the Fab is not a membrane protein, obtain the *stage1.wts*, *stage2.wts*, *stage3.wts* (SI) files from the Rosetta directory *<rosetta_directory>\demos\tutorials\rosetta_cm\3_hybridize*

3.1.6   Define hybridize script: rosetta_cm.xml

Obtain the *rosetta_cm.xml* file from the local Rosetta directory
<rosetta_directory>\demos\tutorials\rosetta_cm\3_hybridize

Edit it based on the Fab files (SI), including changing the names for the fragment and template files to use.

3.1.7   Define options: *rosetta_cm.options*

Obtain the *rosetta_cm.options* from the local Rosetta directory
<rosetta_directory>\demos\tutorials\rosetta_cm\3_hybridize

Edit it based on the Fab files, including changing the names for the *fasta* and *xml* files, and deleting the last membrane section (SI)

3.1.8   Run RosettaCM hybridize

Put all the files in one folder. The name and path of the folder are not important. The user needs to run the command under that folder.

Run command

- *<rosetta_directory>/main/source/bin/rosetta_scripts.default.linuxgccrelease @rosetta_cm.options*
- This is ideally carried out on an HPC to accelerate the structure generation. In total, for Fab, more than 40,000 structures were generated.

3.1.9   Post-treatment

To score the generated structures, use *score_jd2.linuxgccrelease*. Run command:

- *<rosetta_directory>/main/source/bin/score_jd2.linuxgccrelease -scorefile score.sc -s <name>.pdb*

From the top structures (lowest scores by *score_jd2*), choose only the PDBs for which all the 5 disulfide bonds are intact, i.e. the first five lines in the PDB files were *SSBOND*.

The best structure with all the 5 disulfide bonds, was selected with score of -1195.553 (*S_00041.pdb*, SI, See Note 4.2.5).

3.2   Using Rosetta *relax* to refine the local conformational space

The initial models generated by *RosettaCM* were not energy minimised. Thus, the Rosetta *relax* was conducted to refine the local conformational space while retaining most of the raw crystal structure.

3.2.1   Define disulfide bond: *Fab.disulfide*

Create a Rosetta disulfide bond file *Fab.disulfide* (SI), with 5 lines, each line representing a disulfide bond:

23 88

134 194

214 434

236 310

358 414

3.2.2   Define options: *relax.options*

The *relax.options* file can be obtained from
<rosetta_directory>\demos\tutorials\rosetta_cm\4_relax, with the dualspace options
disabled (SI).

3.2.3   Run relax

<rosetta_directory>/main/source/bin/relax.linuxgccrelease @relax.options

- This is ideally carried out on an HPC to accelerate the process.
- More than 20,000 pdbs were generated. A python script was used to identify the pdb
  with lowest score (-1464.373, *S_00041_0007.pdb*, SI)

The structure with the lowest score was then used in the following protein modelling.

3.3   Molecular dynamics in Gromacs

Molecular dynamics simulations can be used to identify the flexible regions of Fab as potential
targets for protein engineering. The Gromacs software was used by following the online
tutorial (See Note 4.3.1). The initial preparation steps, from *pdb2gmx* to *genion*, were carried
out in local PC, where the *Gromacs 2020-beta2* was installed at Windows Subsystem for Linux
(Ubuntu). The subsequent steps were conducted at UCL High-performance Computing, where
gromacs/2019.3/intel-2018 was used.

3.3.1   Determine the protonation states

To define the protonation states of residues in the newly relaxed structure, go to http://apbs-rest-test.westus2.cloudapp.azure.com/pdb2pqr (See Note 4.3.2), and upload the PDB file that
scored best above.

Set the desired pH (pH 7 as default). Choose "Use PROPKA to assign protonation states at
provided pH".

Keep the default PARSE forcefield.

Keep the default naming scheme as "Internal naming scheme".

For additional options, select

- Ensure that new atoms are not rebuilt too close to existing atoms
- Optimize the hydrogen bonding network
- Create an APBS input file
- Add/keep chain IDs in the PQR file
- Remove the waters from the output file

Submit the job, which only takes a few seconds to complete.

There are several output files. The most important is the *xxx.pqr* file, which summarises the
total charge and a breakdown of atom charges. Use a script to summarise the charges for
individual residue and C/N terminus (See Note 4.3.3).

3.3.2 Load the PDB structure into Gromacs

Convert the structure from PDB to Gromacs format by running the command (Note 4.3.4):

gmx pdb2gmx -f Fab.pdb -o Fab_processed.gro -water spce -inter -ignh -merge interactive

When prompted, select the *OPLS-AA/L all-atom force field*, merge chains, determine the protonation states based on the summary of the PQR file, confirm the five disulfide bonds, determine the terminus types based on Note 4.3.3. In the end, the prompt summarises the total charge, which should correspond to the PQR file. At pH 7, the Fab has 7 positive charges (i.e. 7 e) (Note 4.3.3).

3.3.3 Define the solvent box

Place the protein structure into a cubic box by running the command:

gmx editconf -f Fab_processed.gro -o Fab_newbox.gro -c -d 1.0 -bt cubic

(See Note 4.3.5)

3.3.4 Solvate the box with water

Fill the box with water molecules by running the command:

gmx solvate -cp Fab_newbox.gro -cs spc216.gro -o Fab_solv.gro -p topol.top

(See Note 4.3.6)

3.3.5 Add ions to neutralize the box

Add ions to the solvated box by running the command (See Note 4.3.7):

gmx grompp -f ions.mdp -c Fab_solv.gro -p topol.top -o ions.tpr -maxwarn 1

gmx genion -s ions.tpr -o Fab_solv_ions.gro -p topol.top -pname NA -nname CL -nn 7 -neutral -conc 0.05

When prompted, select the *SOL* group, so that water molecules will be randomly selected and replaced by the ions.

3.3.6 Energy minimisation, NVT, NPT, grompp for production run

The system needs to be energy minimised (EM) to avoid steric clashes. Afterwards, the solvent then needs to be equilibrated around the solute protein, under an NVT (constant Number of particles, Volume, and Temperature) and an NPT (constant Number of particles, Pressure, and Temperature) ensemble. The system temperature, density and pressure will be stabilised so that it will not collapse during the production run. As EM, NVT and NPT require relatively more computational resources, it is recommended to perform them at a high-performance computer (HPC).

Run the commands at HPC (See Note 4.3.8):

gmx grompp -f minim.mdp -c Fab_solv_ions.gro -p topol.top -o em.tpr

gerun mdrun_mpi -v -deffnm em

gmx grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr -r em.gro

gerun mdrun_mpi -deffnm nvt

gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr -r nvt.gro

gerun mdrun_mpi -deffnm npt

gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr -r npt.gro

3.3.7 Production run

The production run should also ideally be performed using an HPC to accelerate the process. The command for starting the simulation run is (See Note 4.3.9):

gerun mdrun_mpi -deffnm md_0_1 -cpi -append

3.3.8 Analyse the simulation

There are a variety of methods to examine the simulation, after the system reaches an equilibrium.

Run the commands below to superimpose the trajectory to the reference structure (*gmx trjconv*), followed by the average distance (*gmx rms*) and fluctuation (*gmx rmsf*) analysis (See Note 4.3.10):

echo 1 1 | gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_fit-rot_trans.xtc -ur compact - fit rot+trans -e 100000

echo 1 1 | gmx rms -s md_0_1.tpr -f md_0_1_fit-rot_trans.xtc -o rmsd.xvg -tu ns

echo 1 | gmx rmsf -s md_0_1.tpr -f md_0_1_fit-rot_trans.xtc -o rmsf.xvg -oq bfac.pdb -res -b 50000 -e 100000

Figure 1 of the structure-averaged RMSD during the simulation shows that the RMSD reached equilibrium within the first 50 ns. In this case, the RMSF can be analysed based on the 50-100 ns (*-b 50000 -e 100000*) window. The Figure 2 of RMSF for each residue, shows that the most flexible regions are the hinge region of the heavy chain, and the residues at around position 350. Given their flexibility, these regions could potentially be engineered, and thus rigidified, to reduce their local dynamics.

3.4 B-factor analysis

An analysis of crystallographic B-factors may also be used as an alternative measure of flexibility, and this can be based on one structure, or a consensus from several related crystal structures*(24)* as will be used here below.

3.4.1 Obtain the raw B-factors from homologous PDBs

The raw PDB files of homologous human Fab were downloaded from http://www.rcsb.org/ (Note 4.4.1). PDB files were edited manually in a text editor, so that only one set of light and heavy chains were kept (Note 4.4.2). PDB2FASTA was used to extract the sequences from the PDB for heavy and light chains of homologous structures (Note 4.4.3). After that, the extracted

sequences were aligned with that of Fab A33 using BioEdit*(25)*, and the alignments saved as the files *LC_aligned.txt* and *HC_aligned.txt* (SI).

3.4.2   Convert the atomic B-factors to residue level B-factors

In the PDB profiles, each atom has its own B-factor. In order to have an overall inference for the whole residue, all the atomic B-factors of a same residue were averaged and assigned to their corresponding residues. These residual B-factors were then tabulated into the sequence alignment file so that they were aligned according to the Fab A33 residues (Note 4.4.4). In addition, the residual B-factors within each protein were normalised into a distribution with average 0 and standard deviation 1*(24)*. In the end, only the B-factors that accounted for Fab A33 residues were retained. The average B-factors of Fab A33 residues were calculated by averaging the B-factors from individual homologous structures. In order to reduce the scattering noise, the B-factors were further processed by window-averaging across 5 residues. The entire procedure to process the B-factor is shown in Figure 3. The normalised B-factor after window averaging is shown in Figure 4.

3.5   Prediction of mutational △△ G by Rosetta *cartesian_ddg*

3.5.1   Perform the *in-silico* prediction of mutational △△ G

Single point mutations were evaluated using the Rosetta *cartesian_ddg(26)* method. Each of the 442 residues in the PDB structure were mutated into the other 19 amino-acid residues, totalling 8398 variants of single point mutations. This could be run at only the sites pre-selected for flexibility, but we chose to evaluate all possible mutations at all sites given that we had access to an HPC. For each mutation, a *mutfile* was supplied to specify the target mutation. An example of the *mutfile* and job file were in SI. Three iterations (i.e. repeats) were run for both the wild type and mutant. A python script was used to facilitate the generation of mutation files for all the 8398 variants (*single_mut_file_prep.py*, SI). Jobs were submitted to the UCL Myriad High-Performance Computing Facility (Myriad@UCL) with Rosetta Version 2018.48.60516-mpi. After mutating, a python script (*read_ddg.py*) was used to calculate the average change in stability (ΔΔG) induced by each point mutation with reference to the original wild-type model (Note 4.5.1).

3.5.2   Analysis of the ΔΔG prediction

After *in-silico* mutagenesis, 8398 mutations were generated. Their ΔΔG values (in arbitrary Rosetta Energy Units) are plotted in Figure 5 and their frequency distribution is plotted in Figure 6. For more detailed analysis, see Note 4.5.2.

3.6   Selection of mutations for stability improvement

3.6.1   Correlation between RMSF and B-factor

Residue-level flexibilities, as measured by MD simulation and from crystal structure analysis, can be compared by plotting the RMSF-values against corresponding B-factors, along with a simple linear regression as shown in Figure 7 (Note 4.6.1).

3.6.2   Most flexible residues based on RMSF and B-factor

We chose to perform a product between the RMSF and B-factor, and then ranked the 442 residues based on their corresponding product values (Note 4.6.2).

3.6.3   Stable mutant candidates

For picking mutations predicted to be more stable (Note 4.6.3), Figure 8 shows the schematic procedure to select those for construction. Mutants were designed based on a combined

analysis of the B-factors, RMSF and predicted ΔΔG, with a finite number of total mutations to be made and tested experimentally.

The hinge region had the highest RMSF but did not have B-factor values due to the inability of the crystal structures to resolve it. The last two hinge residues, 441 and 442, were selected for mutation. For the other regions (residues 1 to 436), both the B-factors and RMSF were considered. For the seven selected residues, each was mutated into the three amino acids predicted by Rosetta to have the lowest ΔΔ*G* values from across all 19 candidates. As we were interested in the effects of mutations within flexible regions, compared to other regions, additional mutations were selected based only on the lowest ΔΔ*G* values predicted by Rosetta, regardless of the flexibility of the target site.

3.6.4   Unstable mutant candidates

As an additional control in our study, several mutants predicted to be destabilising, were selected from only those candidates with the highest ΔΔ*G* values as predicted by Rosetta. As above, the designed mutants avoided mutations from and to cysteine, salt bridge modifications, and CDR regions of Fab.

3.6.5   Production of the variants

When the mutations are selected, *in vitro* mutagenesis (Note 4.6.4), and expression by fermentation (Note 4.6.5), purification (Note 4.6.6) and storage (Note 4.6.7), can then be performed. The exact procedure for this varies in different labs, and this is not the emphasis of this chapter. Nevertheless, several experiences are shared in the *Note* Section.

3.7   Formulating the Fab

3.7.1   Filtering

Fab samples were filtered through Anotop 25 0.02 µm syringe filters (GE Healthcare, Buckinghamshire, UK) to remove any aggregates generated during storage or freeze-thawing (Note 4.7.1).

3.7.2   Concentrating

Concentration of the 47 kDa Fab was performed using 30 kDa MWCO Vivaspins (Generon Ltd., Bershire, UK) on a benchtop centrifuge. As Fab A33 is stable in water, the retentate on the Vivaspins was diluted with MilliQ water several times to fully replace the storage buffer with pure water. The concentration needs to be precisely adjusted to the desired value as measured by absorbance at 280 nm, such as in a Thermo Scientific Nanodrop (Wilmington, USA). The Fab monomer in water was temporarily stored at 4 °C until final use (Note 4.7.2).

3.7.3   Buffer selection

A scouting study was conducted to identify the conditions that could complete one full kinetics study in a single working day (Note 4.7.3).

Figure 9A shows the retention of Fab monomer over time at 65 °C, in 20 mM acetate, pH 5, with NaCl added to bring the total ionic strength to 200 mM. This indicated that the monomer population decreased by less than 20% in two days. Based on this trend, it would have taken a week to fully aggregate the monomer at this condition. A more extreme condition of pH 4 (20 mM citrate) but with a more moderate incubation temperature of 45 °C, was then tested as shown in Figure 9B. However, this condition was even milder, and no detectable monomer was observed for the first 2 days. After that, a low pH at 4 with 65 °C incubation temperature was further tested as shown in Figure 9C. The required curve for monomer decay was

obtained such that more than 90% monomer was lost in 5 hours. Based on these results, this condition was selected to evaluate the stability of mutants in the aqueous phase. In order to obtain a more accurate record for the monomer loss, a sampling interval of at least every 30 mins was also used.

3.7.4 Formulating the Fab into the desired buffer

The various Fab variants were previously concentrated at 2 mg/ml (Section 3.7.2). For the buffer, a double-concentrated solution was prepared (Note 4.7.4). Then the 2 mg/ml Fab solution was mixed with the double-concentrated buffer at a 1:1 volume ratio. As a result, 1 mg/ml Fab with 1-fold buffer concentration was achieved. As a general note, direct addition of buffers by volume or by mass is preferable to dialysis, to avoid unexpected changes in protein concentration. The actual volume ratios need to be prepared according to different formulation studies (see Section 3.8, 3.9).

3.8 Thermal stability analysis

3.8.1 Sample preparation

The samples were prepared on ice to minimise protein degradation. Each measurement requires 9 µl protein sample. For triplicates measurement, 40 µl Fab was prepared by mixing 2 mg/ml 20 µl Fab variant and double-concentrated citrate buffer, to achieve a final concentration of 1 mg/ml Fab in single-concentrated citrate buffer. For each sample, 9 µl was pipetted into the cuvette (UNi) slot. The loaded cuvettes were then placed onto the equipment for thermal ramping measurement (Note 4.8.1).

3.8.2 Measurement parameters

The thermal stability of proteins was characterised by their intrinsic fluorescence using the UNit (Unchained Laboratories, Wetherby, UK). A step-wise ramping mode was used to heat the protein from 20 to 90 °C at 1 °C per step (Note 4.8.2).

3.8.3 Fitting the thermal measurement result

The Barycentric Mean (BCM) was outputted to *csv* format files after the run completed. Each set of data was fitted to the Equation 1 to Equation 4[18] (Note 4.8.3).

3.8.4 Analysis of the thermal stability

The thermal stability data of the Fab variants was shown in Figure 10. Detailed discussion of our results can be found in the previous paper[18]. The T_m values did not vary significantly between the stable variants. However, six variants, namely HC-A227E, HC-A228H, HC-T135Y, HC-S219Y, LC-L154A and HC-G194H, resulted in significant increases in ΔH_{vh} and ΔS_{vh}. This indicated decreased conformational flexibility for the native ensemble than in the wild type, assuming that both wild type and the variants retained a similar S_U for the unfolded ensemble. This demonstrates that targeted stabilisation to rigidify flexible regions, can reduce the number of conformational states populated in the native ensemble, as measured by an increase in unfolding cooperativity[27].

3.9 Aggregation kinetics

3.9.1 Sample preparation

The samples were prepared on ice to minimise the protein degradation. For each variant, 24 RNase-free PCR tubes (Fisher Scientific, UK) were labelled with sample ID, sampling time and repeat ID (e.g. WT, 4-2), for 8 sampling time points and 3 repeats per time point. The tubes were placed on a tube rack for the ease of handing. 500 µl stock solution was prepared by

mixing 250 µl 2 mg/ml Fab in water and 250 µl double-concentrated buffer to achieve a final concentration of 1 mg/ml Fab in pH 4, 20 mM sodium citrate, with NaCl to bring the total ionic strength to 200 mM. This 500 µl would be enough for 24 samples with 20 µl per sample (24*20=480 µl). 20 µl was taken from the 500 µl stock solution and individually pipetted into the bottom of the 24 tubes, making sure no bubble was generated. The same procedure was carried out for all variants.

3.9.2   Thermal incubation

A PCR thermal cycler (Bio-Rad C1000 Touch, UK) was pre-set at 4 °C. After the sample preparation, the PCR tubes were placed on to the thermal cycler, except the samples required for measurements at the first sampling time (Time 0). Once loaded, the PCR lid is closed and the instrument temperature set to 65 °C. It should take 10-20 seconds for the instrument to reach 65 °C from 4 °C, at which point a timer is started.

3.9.3   Sampling

Samples were sacrificed in triplicates for each variant every 5 min (Note 4.9.1). After taking out from the thermal cycler, samples were immediately quenched on ice for at least 5 min. The samples were then centrifuged at 15,000 rpm, 4 °C for at least 15 min. After that, 15 µl of supernatant was transferred into a previously labelled HPLC vial insert, capped and sent for HPLC analysis of the monomer retained.

3.9.4   HPLC analysis for monomer retention

The monomer retention was analysed by using an Agilent Zorbax Bio Series GF-250 SEC-HPLC column (Agilent, Berkshire, UK) on an Agilent 1200 HPLC system (Cheshire, UK). The sample chamber was set 4 °C. Calibration curves were established prior and after each batch of analyses, in which close calibration coefficients indicating the column stability throughout the analysis. 5 µl sample was injected and the mobile phase was at a flow rate of 1 ml/min with 200 mM sodium phosphate, pH 7. Each cycle took 4.5 min to complete with the Fab A33 monomer eluted at 2.6 min. The resulted peak area was referred to the calibration curve to determine the monomer retention.

3.9.5   Fitting the monomer retention kinetics

An exponential function was used to fit the kinetics of monomer retention using Equation 5, in which A and k are the coefficients, y is the monomer retention normalized from 0 to 1, and t is the incubation time. The first derivative of Equation 5 is shown in Equation 6, in which the absolute value of initial aggregation rate is "$A*k$" when $t = 0$. The OriginLab software (Version 2018 or above) was used for the fitting. A single exponential function is appropriate to Fab aggregation, which appears to have a rate order close to 1 in most conditions. Other proteins may require more complex kinetic fitting to obtain rate constants.

3.9.6   Analysis of the aggregation kinetics

The aggregation kinetics data of the Fab variants are shown in Figure 11. Detailed discussion of the results can be found in the previous paper[18]. HC-S219Y and HC-G194H reduced the initial rate by more than 12%, while HC-A228H was slower by 6.5% compared to the wild type. However, their T_m values were even 1 ° C lower than that of wild type. This implied that they could be in, or close to, particular hotspots of structure that affect the aggregation mechanism. To visualize this, the locations of all stable variants with $\ln(v)$ = 8.53 to 8.86% day^{-1} are highlighted in Figure 12. It shows that the three top mutations were located close together at the C-terminal end of the heavy chain. Two of them, HC-A228 and HC-G194, were mutated to

histidine, and the other one HC-S219 was mutated to tyrosine. The substitution from aliphatic side chains to imidazole or aromatic ones could potentially provide more interactions to lockdown the flexible hinge region. This led to the increased unfolding cooperativity as shown by their decrease in S_N (increased ΔS_{vh}) compared to the wild type.

3.9.7   Correlation between $\Delta\Delta G$, T_m, and $\ln(v)$

Figure 13 shows the correlation between the $\Delta\Delta G$ and the T_m, $\ln(v)$. A strong correlation (R^2 =0.945) between T_m and the $\Delta\Delta G$ indicates that Rosetta could reasonably predict the change in free energy upon point mutation, for the subset of mutations selected. Its scoring function (i.e. ref2015**(28)**) has been highly optimised to approximate the biomolecular conformation. The correlation between $\Delta\Delta G$ and $\ln(v)$, however, is less significant, with R^2 at 0.504. The poor correlation implies that additional factors need to be accounted for when predicting aggregation kinetics, and that these are unlikely to have been considered by Rosetta. For example, local dynamics at the protein surface could influence the overall colloidal stability, or reveal specific motifs that self-interact between proteins, and thus lead to the aggregation. In this work, the hinge region had been identified as an aggregation-prone region, which could be rigidified by introducing aromatic side chains (Note 4.9.3).

4   Note

4.1   Efficient file management and batch operation

4.1.1   Efficient text editor

1. In order to efficiently generate, access and modify the large number of files relevant for protein modelling, various batch operation tricks were discussed here.

A good text editor is essential to visualise the exact information and understand the fundamental knowledge. The text editor *Notepad++* is used in this work. The software has many features that enable productive processing of the text content. In particular, it can

- Assign different colours to different type of words based on the file extension. For example, for a text file called *script.py*, a python script file, the numbers, strings, variable names and syntax words (e.g. *if*, *print*) are coloured differently.
- At the bottom of the software, the row/column position of the cursor and the length of the highlighted words are displayed. This is useful when protein residues at certain positions need to be retrieved.
- Column-wise editing is very helpful. By pressing *Shift + Alt*, the cursor is shown column-wised. Users can delete or add the content column-wised without affecting the others. This is particularly useful for tabulated PDB files. For instance, when the chain ID is missing, one can add the chain ID for all the *ATOM* rows at one time.

4.1.2   Batch operations in text files and spread sheets

2. Occasionally, hundreds or thousands of files/folders need to be processed. It is better to contain the content in simple text format or spread sheet (*xls, xlsx, csv*), as they are convenient to be processed by script. Python, R, MatLab could be used to achieve this batch processing. For python, the *PyCharm* software is recommended for coding. The commonly used modules are

- xlsxwriter: create xlsx files, but cannot read from xlsx files
- xlrd, xlwt, xlutils: read and write xlsx files
- csv: edit csv files

- numpy, scipy: mathematical operations

4.1.3 High-performance Computing

High-performance Computing becomes necessary when personal computers do not have sufficient cores and memory to handle large amount of calculation, or do not have sufficient storage space. We have benefited greatly from our university research computing facility (https://www.rc.ucl.ac.uk/docs/). The centralised computing facility enables the researchers to accelerate their computational work to a great extent.

When a job is submitted, it will be queued before executed. The queue time depends on the computational resources requested, including cores, memory and time. It takes longer to queue for a job requesting more resources. To facilitate immediate execution with more cores, a strategy is to reduce the time while requesting more cores. For example, to run a production simulation, submit *hold-jobs* that will only be executed when its previous job has completed. Another issue to consider is the number of cores in a node. It is better to request the number of cores that within a same node. For example, if the architecture is 40 cores in a node, it is suggested to request a multiple of 40 (i.e. 40, 80, 120) cores.

4.2 Using *RosettaCM* to obtain a starting structure model

4.2.1 The residue numbering method for the Fab A33

The Fab has 442 residues, with 214 residues in the light chain (LC) and 228 residues in the heavy chains. In most cases, the heavy chain residues were numbered from 215 to 442. But when *HC* is specified, it is numbered from 1. For example, HC-G194 refers the same residue as G408.

4.2.2 RosettaCM tutorial

https://www.rosettacommons.org/demos/latest/tutorials/rosetta_cm/rosetta_cm_tutorial

4.2.3 Fasta file "FabMissResi.fasta" with missing residues of Fab A33
>FabMissResi

DIQMTQSPSSLSASVGDRVTITCKASQNVRTVVAWYQQKPGKAPKTLIYLASNRHTGVPSRFSGSGSGTD
FTLTISSLQPEDFATYFCLQHWSYPLTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRE
AKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGE
/AVQLVESGGGLVQPGGSLRLSCAASGFAFSTYDMSWVRQAPGKGLEWVATISSGGSYTYYLDSVKGRF
TISRDSSKNTLYLQMNSLRAEDTAVYYCAPTTVPFAYWGQGTLVTVSSASTKGPSVFPLAPSGTAALGCLV
KDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEP

4.2.4 The amino acid sequence of the Fab A33: Fab.fasta
Light chain (Residue 1-214)

Heavy chain (Residue 215-442)

>Fab

DIQMTQSPSSLSASVGDRVTITCKASQNVRTVVAWYQQKPGKAPKTLIYLASNRHTGVPSRFSGSGSGTD
FTLTISSLQPEDFATYFCLQHWSYPLTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRE
AKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGE
C/EVQLVESGGGLVQPGGSLRLSCAASGFAFSTYDMSWVRQAPGKGLEWVATISSGGSYTYYLDSVKGR
FTISRDSSKNTLYLQMNSLRAEDTAVYYCAPTTVVPFAYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGG

TAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNT
KVDKKVEPKSCDKTHTSAA

4.2.5   Variations in different batches
Different values are expected at different batches.

4.3   Molecular dynamics in Gromacs

4.3.1   Gromacs tutorial
For beginners, it is recommended to follow the tutorial at http://www.mdtutorials.com/gmx/.
The official website is http://www.gromacs.org/, where users can download the software and
read the protocols. The online forum is useful at https://mailman-1.sys.kth.se/mailman/listinfo/gromacs.org_gmx-users.

4.3.2   PDB2PQR
The old website was http://nbcr-222.ucsd.edu/pdb2pqr_2.1.1/, which was retired on April
30th, 2020.

The pH value assigned is the most important input as it affects the selection of protonation
states in the Gromacs.

It is advised to keep most of the default settings. But "Add/keep chain IDs in the PQR file" is
recommended as it helps to summarise the residue charge in the later stage.

4.3.3   Post-treatment for the PQR file
The PQR file lists the total charges (Line 11) and a breakdown of atomic charges (starting from
Line 13). The Gromacs requires the residue-level charges. So the atomic charges need to be
summarised into residue level.

The PQR file uses the PDB format, so atomic property information (atom ID, residue ID,
coordinates, etc.) is tabulated at different columns. Column 55-62 are used to position the
atomic charges. A script is useful to read the charges line by line and summarise the changes
into residue level. In the end, make sure the total charge (Line 11) corresponds to the sum of
individual residual charges. Additional caution needs to be emphasised:

- Gromacs requires the terminus charges. For a protein of two chains (e.g. Fab), there
  are two N terminus and two C terminus. The N terminus is protonated (i.e. +1 e) when
  Atom *H*, *H2* and *H3* are present. The C terminus is protonated (i.e. 0 e) when Atom
  HO is present. Particular care is needed when the terminus residue is chargeable. For
  example, the total charge of N terminus residue could be 0 when N terminus is
  protonated, which implies the side chain of the chargeable residue is negatively
  charged (i.e. -1 e).
- There are four types of charge states for histidine to choose in Gromacs, including "H
  on ND1 only (HISD)", "H on NE2 only (HISE)", "H on ND1 and NE2 (HISH)" and "Coupled
  to Heme (HIS1)". The presence of Atom *HD1* or *HE2* in the histidine refers the type of
  HISD or *HISE*, respectively; or *HISH* when both *HD1* and *HE2* exist. The *HIS1* is a rare
  case, which refers a delta-protonated histidine.
- For the Fab A33, it has +7 charges at pH 7, including +26 lysine, +12 arginine, 0
  glutamine, -16 aspartic acid, -15 glutamic acid, 0 histidine, 0 terminus.

4.3.4   Load the PDB structure to Gromacs
The Gromacs is run in Linux environment. So common Linux commands are applicable.

Use *dos2unix* to make the PDB file compatible in Unix system.

There are variations for the options used in the *gmx pdb2gmx*. In our case, the *-inter* is to interactively handle the options, including assigning protonation status to each chargeable and termini residue, and selecting disulfide bonds. The protonation status is referred from Note 4.3.3. *-merge* is to merge multiple chains into one molecule. The Fab has two chains, which need to be considered as one molecule. The *-merge* ensures the intactness of the interchain disulfide bond.

The user can also choose other force fields that are not listed in the default selections. Force fields can be downloaded at http://www.gromacs.org/Downloads/User_contributions/Force_fields. If needed, the force field directory (suffixed with *.ff*) needs to be put at the working directory, or in the */usr/local/gromacs/share/gromacs/top*.

4.3.5   Define the solvent box
The *-bt* defines the geometry of the box, and *cubic* box is a common option for the ease of visualisation. Other types, e.g. *triclinic, dodecahedron, octahedron*, can be used when a large system needs to be constructed with limited computational resource. The *-d 1.0* sets 1 nm distance between the protein and the box. It is noted that, the *-box* could directly sets the box dimension. It is needed if the protein needs to be put into a box with defined solute concentration.

4.3.6   Solvate the box with water
Water is commonly used to solvate the solution box. The configuration of the solvent (*-cs spc216.gro*) matches the water type (SPC/E) previously used in the *pdb2gmx* command.

4.3.7   Add ions to neutralize the box
The *-maxwarn 1* is to prevent fatal error.

The *gmx genion* is to neutralise the box by adding 7 negative Cl^- ions. In addition, *-conc 0.05* will add additional Na^+ and Cl^- ions to make the total ionic strength at 50 mM.

4.3.8   Energy minimisation, NVT, NPT, grompp for production run
The *minim.mdp, nvt.mdp, npt.mdp, md.mdp and job_em_nvt_npt_prod_grompp.sh* are supplied in SI.

The *.mdp* files are used for temperature at 300 K. For other temperatures, change the *ref_t* and *gen_temp* in *nvt.mdp*, change the *ref_t* in *npt.mdp*. The *md.mdp* file sets the frequency to save the coordinate (*nstxout*), velocity (*nstvout*) and energy (*nstenergy*). One can customise these parameters to save the disc space on the HPC.

4.3.9   Production run
The *job_conti.sh* file is in SI. It is suggested to put only the *job_conti.sh* and *md_0_1.tpr* files in a new folder, so as to distinguish with the previous preparation files.

When submitting a job in HPC, it will queue before actually executing (Note 4.1.3). To maximise the use of the computational resources, the time needed in the *job_conti.sh* file is reduced to 1 hour (*h_rt=01:00:0*), but with relatively more cores (*mpi 12*). It aims to execute the job short after submission. After 1 hour, the same job will be submitted. Because *-cpi -append* is used, the next job will continue from the previous checkpoint.

To facilitate more efficient job submission with continuation feature, one can submit a job that will only be run when its dependent job has finished. For example, *qsub -hold_jid 12345 myscript.sh* means to submit a job that will only run when the Job ID 12345 has finished. This is very useful when we use *-cpi -append* in the Gromacs production run.

Repeats are always required. At least three repeats of MD should be carried out, preferably more replicates if necessary. The analysis for each MD batch should be performed separately to see the difference within the same condition. The starting model needs to be improved if the batches of same condition deviate significantly from each other. When an identical condition (e.g. same PDB, pH, temperature, ionic strength) is performed in repeats, the same *em.tpr* can be used with different subsequent NVT, NPT, etc.

It is usually more likely to have 10 jobs running with each job occupying 12 cores, than to have 1 job running with 120 cores requested. So the strategy is to split the resources requested to have the maximum overall resources acquired. The *job_conti_batch.sh* shown in SI is an example. Two MD conditions are prepared, with each condition having 3 repeats. 30 dependent jobs (*for i in 2 3 4 5* …) are submitted, each is held to run until its previous one completes. By typing *./job_conti_batch.sh* in the current working directory, 180 jobs (2 conditions x 3 repeats x 30 dependent ones) will be submitted.

4.3.10 Analyse the simulation
The *gmx trjconv* is used to align (*-fit rot+trans*) the raw trajectory to the reference structure, so that it is easier to visualise the relative dynamics of the protein compared to the initial state. *-e 100000* is used to only consider the first 100 ns. *echo 1 1* in the *gmx trjconv* is to use the all-atom protein as the reference for least-squares fit alignment and output only the all-atom protein entries; *echo 11* in the *gmx rms* is to use the all-atom protein for both the least-square fit alignment and the RMSD group calculation; the *echo 1* in the *gmx rmsf* is to compute the root mean square fluctuation in the all-atom protein.

4.4   B-factor analysis
4.4.1   Homologous PDBs
1A6T, 1B2W, 1C5D, 1DFB, 1DN0, 1DQD, 1FGN, 1IT9, 1L7I, 1OPG, 1T3F, 2Z4Q, 2ZKH, 3D69, 3G6A, 3HC0, 3HI5, 3VG0, 4GSD, 4HBC, 4HH9, 4HIE, 4LKX, 4OCY, 4OSU, 7FAB.

4.4.2   Get unique chains
When *Download Files* for a particular PDB entry, choose *Biological Assembly* instead of *PDB Format* can download only one set of unique chains.

4.4.3   PDB2FASTA
Due to the crystallisation uncertainties for excessive thermal motion of particular residues, some residues were not displayed in their PDB files but shown in their sequence files. As a result, only the residues that displayed in PDB files could be used for sequence alignment with Fab A33 and B-factor cross comparison. Therefore, instead of using the FASTA sequence file, the actual sequence information was extracted from the PDB for heavy chain and light chain separately. An online server could be https://zhanglab.ccmb.med.umich.edu/pdb2fasta/.

4.4.4   Tabulate the B-factors to excel
All the input and output files are in the SI. Use *B_factor_calculation_excel_with_alignment.py* to process the PDBs in the Folder *pdb_rename_chain_ID* based on the sequence alignment in

the Folder *sequence_aligned*. Adjust the first three lines in the python script. The output is *B-factor_raw.xlsx*, which is further processed and saved as *B-factor.xlsx*.

In this case, the B-factor of each residue is averaged from their atoms, without considering the atom mass. Alternatively, mass-weighted average can also be performed.

4.5 ΔΔ *G* prediction by Rosetta *cartesian_ddg* method

4.5.1 Perform the *in-silico* prediction

Based on the Fab A33 structure, the Rosetta *cartesian_ddg(26)* method, which was developed in 2016, could perform the calculation more rapidly, more accurately, more reproducibly and was more straightforward to learn compared to the previous ΔΔ*G* method in 2011, *ddg_monomer(26)*. The usage of *ddg_monomer* on the Fab A33 structure could refer to our previous paper*(18)*.

The *cartesian_ddg* requires an optimised initial structure for ΔΔ*G* prediction. As our structure had already been relaxed (Section 3.2) after the RosettaCM, further relaxation was not conducted as instructed on the *cartesian_ddg* website (https://www.rosettacommons.org/docs/latest/cartesian-ddG). If the user is not sure about the necessity of the *relax*, they could run with and without the *relax* step, and compare the correlation with *in vitro* data (Note 4.9.3).

The *cartesian_ddg* optimises the neighbouring packing around the target residue. As a result, three iterations for the wild type of one mutation could deviate from those of another mutation. This is an acceptable phenomenon when the protein involved was large.

4.5.2 Analysis of the ΔΔ*G* prediction

It can be seen that most of the mutations had ΔΔ*G* values of close to 0, which implies that most single mutations exert only a limited impact on the protein stability. Moreover, negative ΔΔ*G* values were all greater than -10, whereas positive ΔΔ*G* values extended much further with a maximum value of 69. This result implies that the Fab A33 sequence already had a relatively stable form, which is not surprising given that it is already the result of significant selection and engineering as a potential therapeutic. However, there remained some room to further stabilise the protein even though the extent was very limited. As the Δ*G* is around -1700, a -10 ΔΔ*G* improvement would only provide less than 1% stabilising impact on the free energy of the global structure. On the contrary, the destabilising mutants may exhibit great detrimental effect with the maximal exerting more than 4% loss in Δ*G*, which include polar or chargeable mutations in the protein core, hydrophobic mutations on the outer surface, hindrance caused by large amino acid substitutions like aromatic ones and disruption of salt bridge and hydrogen bond.

4.6 Selection of mutations for stability improvement

4.6.1 Correlation between RMSF and B-factors

The B-factor values from the human Fab structures had a similar sequence dependence as the RMSF. This implies that B-factors from the crystal structures of Fab reflect a disorder *in vitro* that can be simulated by Gromacs *in silico*. In order to verify that both independent measurements achieve similar flexibility results for the protein residues, it was useful to find out if they correlated with each other to some extent. Figure 7 shows the correlation between RMSF and B-factor. It is shown that nearly 50% variation could be explained by the linear regression model. Therefore, the RMSF and B-factor values were statistically correlated.

4.6.2 Most flexible residues based on RMSF and B-factors

The product calculation is shown in rmsf-bfactor.xlsx (SI). The top 3 residues (433-434) are near the hinge region. Residues 347-351 are also top ranked, which have very high RMSF.

4.6.3 Stable mutant candidates

The hinge regions in the heavy chains had no available B-factor values, but their RMSF values were very high. It was suggested that the hinge regions accounted for the flexibility and instability of an IgG, and switching it to another subclass could potentially improve the formulation stability while maintaining its binding affinity**(29)**.

Three mutations were selected for each residue because Rosetta is expected to fit with 69% prediction accuracy**(30)**, so theoretically, three mutations should yield at least one stable mutant.

In order to design the mutants only based on B-factors, RMSF and ΔΔG values without losing or introducing new features, several additional filter criteria were applied. Mutations were avoided that would introduce a cysteine, remove a disulphide bond, or disrupt salt bridges. In addition, mutations were avoided in the Complementarity Determining Regions (CDR) of the Fab as this would affect Fab function, and guide mutations to regions that could be potentially useful in any Fab generated by industry. The CDR regions are listed in Table 1.

4.6.4 Production of the variants: mutagenesis

The amino acid and DNA sequences of Fab A33 are provided in the SI. When designing the primers for site-direct mutagenesis, the codons used mostly by *E. coli* were adopted from a previous codon usage work**(31)**.

4.6.5 Production of the variants: fermentation

Depending on the equipment availability, the users can choose a suitable scale of the bioreactor to express the protein. It lasts commonly one week to perform the fermentation including autoclave and cleaning. It is suggested to run several bioreactors in parallel so as to complete 15-20 variants in 1-2 months. Each bioreactor does not need to be very large as long as decent quantity of protein can be produced. Shake flasks were tested, but the yield was very low so it is not recommended.

In our case, the DASbox® Mini Bioreactor (working volume 60-250 ml) (Eppendorf, Germany) was used as four different mutants could be run in parallel reactors. A successful cycle can result in more than 30 mg Fab, allowing initial screening for stability.

4.6.6 Production of the variants: purification

Though the total quantity of Fab is low from the DASbox bioreactor, the 1 or 5 ml HiTrap Protein G column was not used, due to the concern of losing precious Fab during the flow-through. Instead, a 90-ml Protein G column was used to maximise the protein capture. After acidic elution, neutralised buffer was immediately added to preserve the Fab stability. After purifying one variant, the Protein G column was kept at 70% ethanol overnight to fully remove the residual Fab.

4.6.7 Production of the variants: storage

The Fab A33 is stable at neutral pH, 4 °C for at least one week. So, most of the purified Fab solution needs to be stored at sub-zero condition for long-term usage. The protein concentration needs to be measured to allow proper labelling and aliquoting. Concentrate the protein solution if it is very low concentration (e.g. less than 0.5 mg/ml) (e.g. Vivaspin, see

Note 4.7.2). A -80°C freezer could preserve the protein better than the -20°C. Snap-freezing (e.g. liquid nitrogen) is highly recommended to minimise the denaturation from freezing. When the protein is needed, thaw the frozen samples in room temperature, wait for them to fully melt in aqueous phase, then put on ice or 4 °C fridge for subsequent usage.

4.7   Formulating the Fab

4.7.1   Filtering

The filtering and concentrating process will lose some sample. So it is advisable to take 30-50% more sample than the final estimated protein amount required, to avoid having to repeat the preparation steps.

The Fab sample could contain minor amount of aggregates and fragments as a result of freeze-thawing. An alternative approach to obtaining the pure monomer, is through preparative size-exclusion chromatography (SEC) on an AKTA system. When multiple Fab mutants were needed, it took excessively long period to prepare individual mutants through size-exclusion chromatography. Therefore, where possible, filtering and concentrating was used to remove aggregates and fragments, respectively, and SEC used only in more challenging cases.

The molecular weight of Fab A33 is around 47 kDa. Filtering can either be carried out with 0.02 μm syringe filters, or 100 kDa molecular weight cut-off (MWCO) concentrating tubes (e.g. vivaspin). It is not advisable to use 0.2 μm filters, as small aggregates can still be retained.

4.7.2   Concentrating

The light or heavy chain of Fab is around 24 kDa. So passing through a 30 kDa MWCO Vivaspins was able to effectively remove the Fab fragment. When using the centrifuge, do not operate at an excessive rotation speed, as this will deteriorate the membrane and cause protein loss. Always keep the flow-through solution in case of the membrane fouling.

This process could also be achieved by using dialysis cassettes, e.g. Slide-A-Lyzer Dialysis cassettes (Fisher Scientific, Leicestershire). When more than 10 ml sample needs to be processed, a dialysis cassette is recommended as it saves the time in repeated centrifugation. However, some residual liquid will unavoidably be lost within dialysis cassettes. So for minor volume, e.g. less than 5 ml, concentrating tubes are recommended.

It is a very laborious work to tuning the concentration to the desired value. Over-diluting the protein means to do the centrifuging again to raise the concentration. Always make sure the measurement is within the linear range of the Nanodrop equipment. If a high-concentration sample is required, dilute the sample aliquot to within the linear range before measuring.

4.7.3   Buffer selection

To determine the buffer used for formulation, the buffer should have clinic or scientific relevance. Physiological condition (e.g. pH 7.4) is desired, but Fab will be very stable at this condition, and significant variation in stability could not be readily observed among the variants. When the protein is incubated in Eppendorf tubes for a long time, like several weeks or months, controlling against evaporation becomes challenging in such tubes. The user needs to render the Fab to degrade in a manageable time manner, or use an alternative with tight fitting screwcaps.

Based on previous work *(19)*, low pH, high salt concentration and high temperature would cause rapid aggregation. In this work, the Fab wild type was firstly tested to provide a benchmark for the designed variants.

4.7.4   Formulating the Fab at desired buffer

The https://www.biomol.net/en/tools/buffercalculator.htm website was used to calculate the buffer recipe. Note down the values of acid, base and NaCl amount required by setting the pH 4 citrate with NaCl to bring the total ionic strength to 200 mM. These values need to be doubled to prepare the double-concentrated buffer solution. It is suggested to prepare stock solutions of citric acid and NaOH, filtered by 0.22 µm filters, before mixing the two stock solutions at the desired volume ratio.

The citrate buffer has three pKa values, which makes it difficult to obtain a precise solution. After preparing the double-concentrated buffer, its pH needs to be measured. Take an aliquot of the double-concentrated buffer, add equal volume of water to dilute it into one-concentration, then use the pre-calibrated pH meter to measure the pH. Add high-concentration NaOH to the buffer so that there is negligible volume change in the buffer. Alternative buffer calculators can be used, like http://clymer.altervista.org/buffers/cit.html. For other buffers except citrate, https://www.liverpool.ac.uk/pfg/Research/Tools/BuffferCalc/Buffer.html can be used.

Alternatively, a buffer exchange method can be used to replace the water with the desired buffer, using a dialysis cassette (Note 4.7.2). However, this would take several hours to complete and the dialysed Fab may still need to be concentration adjusted afterwards. This method is also not favoured for formulation work with excipients where the excipients may interact with the protein and result in an increased molar concentration.

4.8   Thermal stability analysis

4.8.1   Sample preparation

Theoretically, 27 µl protein sample is necessary for triplicate measurement. 40 µl is suggested to prepare to have sufficient volume for pipetting. The UNi slots are open at the two ends. When loading the sample, aspirate 9 µl Fab solution using a 10 µl tip, place the tip to the centre of the slot, then dispense the solution, so that the solution will spread to both ends of the slot. During dispensing, gradually lift the tip until the tip comes out from the slot. This operation could minimise the bubble within the slot. If any bubble is seen after loading, aspirate the solution and redo the loading. Though 9 µl was indicated in the manual, 8.5 µl would be sufficient without causing the droplet reaching out of the slot.

4.8.2   Measurement parameters

The UNit instrument can accommodate 3 cuvettes (UNis), with each cuvette having 16 slots to load sample, which makes a maximum of 48 samples measured in one batch. The instrument characterises the unfolding of proteins by capturing the shift of fluorescence emission wavelength. The fluorescence emission, typically from tryptophan, centres around 330 nm when the protein is folded, and shifts to 350 nm when it is unfolded.

Step-wise and linear are the two ramping modes that can be selected. The step-wise mode takes reading for all the samples at the same temperature, before it increases to a higher temperature. Therefore, it takes longer to keep the sample at a particular temperature when more samples are measured. This is particularly influential when the temperature approaches the transition point, as proteins would be stressed differently when the sample number varies. As a result, the number of samples affects the reproducibility of the reading outputs.

Linear ramping mode was introduced when the second-generation of the UNit instrument had been developed. In this mode, the temperature keeps increasing at constant rate (1 °C/min),

regardless of the number of samples. This could provide a better reproducibility when a high reading frequency is met.

In our case, the step-wise mode was still used, as we want to compare the data to the first-generation instrument we previously owned. The sample number was kept at 12 so as to give more consistent result by retaining the same overall sample run timings.

4.8.3 Fitting the thermal measurement result

The van't Hoff thermal parameters ΔH_{vh}, ΔS_{vh} and thermal unfolding temperature (T_m) at which 50% of protein population was unfolded, were determined from the barycentric mean (BCM) (i.e. I_T in Equation 1) of protein intrinsic fluorescence spectra at 280-460 nm, at each temperature T, by fitting to the van't Hoff equation (Equation 1 and Equation 2)*(32, 33)*. The fraction of unfolded protein (f_T) at any temperature T was calculated from Equation 3. T_{on} was defined as the temperature T at which 2% protein was unfolded (i.e. f_T = 0.02) using Equation 4 derived from both Equation 1 and Equation 3.

The *OriginLab* software (Version 2018 or above) was used for the BCM fitting. A customised equation was built by clicking *Tools - Fitting Function Organizer*. For Equation 1, temperature *T* and BCM were set as *Independent Variables* and *Dependent Variables*, respectively. All the other parameters were set as *Parameter Names*.

4.9 Aggregation kinetics

4.9.1 Sampling

The 5 min sampling interval can be an initial test for the variants. It should be adjusted for proteins with varied stability.

Be very careful when aspirating the 15 μl supernatant after centrifugation.

4.9.2 HPLC analysis for monomer retention

5 μl was tested as a safe volume to be taken from the 15 μl sample. Make sure the resulted peak area was within the calibration range. For high concentration analysis, reduce the injection volume accordingly.

4.9.3 Correlation between $\Delta\Delta G$ and the T_m, ln(v)

Figure 13 has clearly shown the importance of having destabilising variants as negative controls. Actually, when only stabilising variants were used in the linear regression, an opposite correlation was obtained. It is also important to select the proper variables for correlation. For example, aggregation rates should be plotted as logarithmic values to create potential linear free-energy relationships with $\Delta\Delta G$ or T_m. When using OriginLab software to fit the linear regression curve, the results would be different with or without incorporating the errors. The variants analysed in this work were still very limited, and mostly clustered at the stable region. The performance might be improved if more variants, which cover a wider range of $\Delta\Delta G$, were included in the analysis.

5 Acknowledgement

6 Supporting Information

Supporting document can be downloaded from the book websites. *C226S* refers to the wild type of Fab A33.

7 Reference

1. Grilo AL and Mantalaris A (2019) The Increasingly Human and Profitable Monoclonal Antibody Market. Trends Biotechnol 37:9–16

2. Ecker DM, Jones SD, and Levine HL (2015) The therapeutic monoclonal antibody market. MAbs 7:9–14

3. Wu X, Sereno AJ, Huang F, et al (2015) Fab-based bispecific antibody formats with robust biophysical properties and biological activity. MAbs 7:470–482

4. Carter PJ and Lazar GA (2018) Next generation antibody drugs: Pursuit of the "high-hanging fruit." Nat Rev Drug Discov 17:197–223

5. Chiu ML and Gilliland GL (2016) Engineering antibody therapeutics. Curr Opin Struct Biol 38:163–173

6. Arslan M, Karadağ D, and Kalyoncu S (2019) Protein engineering approaches for antibody fragments: Directed evolution and rational design approaches. Turkish J Biol 43:1–12

7. Thusberg J and Vihinen M (2009) Pathogenic or not? and if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum Mutat 30:703–714

8. Meric G, Robinson AS, and Roberts CJ (2017) Driving Forces for Nonnative Protein Aggregation and Approaches to Predict Aggregation-Prone Regions. Annu Rev Chem Biomol Eng 8:139–159

9. Tsolis AC, Papandreou NC, Iconomidou VA, et al (2013) A Consensus Method for the Prediction of "Aggregation-Prone" Peptides in Globular Proteins. PLoS One 8:1–6

10. Potapov V, Cohen M, and Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng Des Sel 22:553–60

11. Khan S and Vihinen M (2010) Performance of protein stability predictors. Hum Mutat 31:675–684

12. Nimrod G, Fischman S, Austin M, et al (2018) Computational Design of Epitope-Specific Functional Antibodies. Cell Rep 25:2121-2131.e5

13. Sutton EJ, Bradshaw RT, Orr CM, et al (2018) Evaluating Anti-CD32b F(ab) Conformation Using Molecular Dynamics and Small-Angle X-Ray Scattering. Biophys J 115:289–299

14. Kuroda D and Tsumoto K (2020) Engineering Stability, Viscosity, and Immunogenicity

of Antibodies by Computational Design. J Pharm Sci 109:1631–1651

15. Welt S, Ritter G, Williams C, et al (2003) Phase I study of anticolon cancer humanized antibody A33. Clin Cancer Res 9:1338–1346

16. King DJ, Antoniw P, Owens RJ, et al (1995) Preparation and preclinical evaluation of humanised A33 immunoconjugates for radioimmunotherapy. Br J Cancer 72:1364–1372

17. Weir N, Athwal D, Brown D, et al (2006), A new generation of high-affinity humanized PEGylated Fab' fragment anti-tumor necrosis factor-α monoclonal antibodies

18. Zhang C, Samad M, Yu H, et al (2018) Computational-design to reduce conformational flexibility and aggregation rates of an antibody Fab fragment. Mol Pharm 15:3079–3092

19. Chakroun N, Hilton D, Ahmad SS, et al (2016) Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. Mol Pharm 13:307–319

20. Amin S, Barnett G V., Pathak JA, et al (2014) Protein aggregation, particle formation, characterization & rheology. Curr Opin Colloid Interface Sci 19:438–449

21. Roberts CJ (2014) Therapeutic protein aggregation : mechanisms , design , and control. 32

22. Tang J and Kozielski F (2020) Crystallization and synchrotron diffraction analysis of humanised A33 Fab, an immunotherapy candidate to colorectal cancer, (working paper)

23. Song Y, Dimaio F, Wang RYR, et al (2013) High-resolution comparative modeling with RosettaCM. 21:1735–1742

24. Reetz MT and Carballeira JD (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat Protoc 2:891–903

25. Hall T (1999), BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT

26. Park H, Bradley P, Greisen P, et al (2016) Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. J Chem Theory Comput 12:6201–6212

27. Shakhnovich EI and Finkelstein A V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers 28:1667–1680

28. Alford RF, Leaver-Fay A, Jeliazkov JR, et al (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput 13:3031–3048

29. Neergaard MS, Nielsen AD, Parshad H, et al (2014) Stability of monoclonal antibodies at high-concentration: Head-to-head comparison of the IgG1 and IgG4 subclass. J Pharm Sci 103:115–127

30. Kellogg EH, Leaver-Fay A, and Baker D (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins Struct Funct Bioinforma 79:830–838

31.   Sharp PM, Cowe E, Higgins DG, et al (1988) Codon usage patterns in Escherichia coli, Bacillus subtiiis, Saccharomyces certvisiae, Schizjosaccharomycespombt, Drosophila mclanogaster and Homo sapiens; a review of the considerable within-species diversity. Nucleic Acids Res 16:8207–8211

32.   Santoro MM and Bolen DW (1988) Unfolding Free Energy Changes Determined by the Linear Extrapolation Method. 1. Unfolding of Phenylmethanesulfonyl a-Chymotrypsin Using Different Denaturants. Biochemistry 27:8063–8068

33.   Consalvi V, Chiaraluce R, Giangiacomo L, et al (2000) Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima. Protein Eng 13:501–507

8   Figure
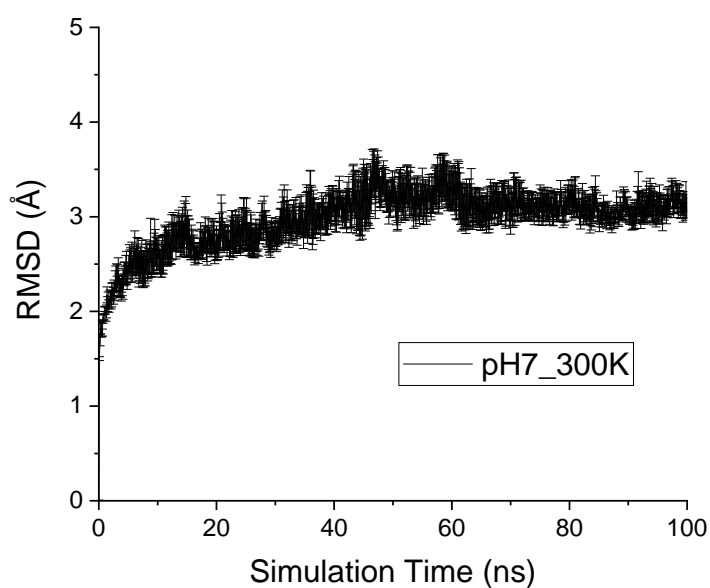
Figure 1. RMSD of the protein with simulation time at pH 7, 300 K. The average of six independent simulations is shown with the SEM as error.
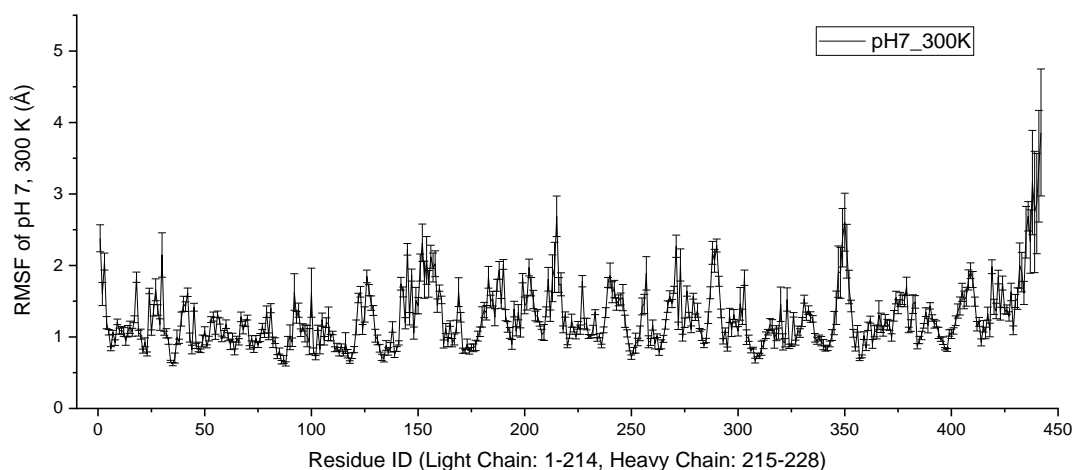
Figure 2. RMSF of the protein at pH 7, 300 K, for the 50-100 ns. The average of six independent simulations is shown with the SEM as error. Reproduced from C. Zhang 2018 with permission from ACS Publications(18).
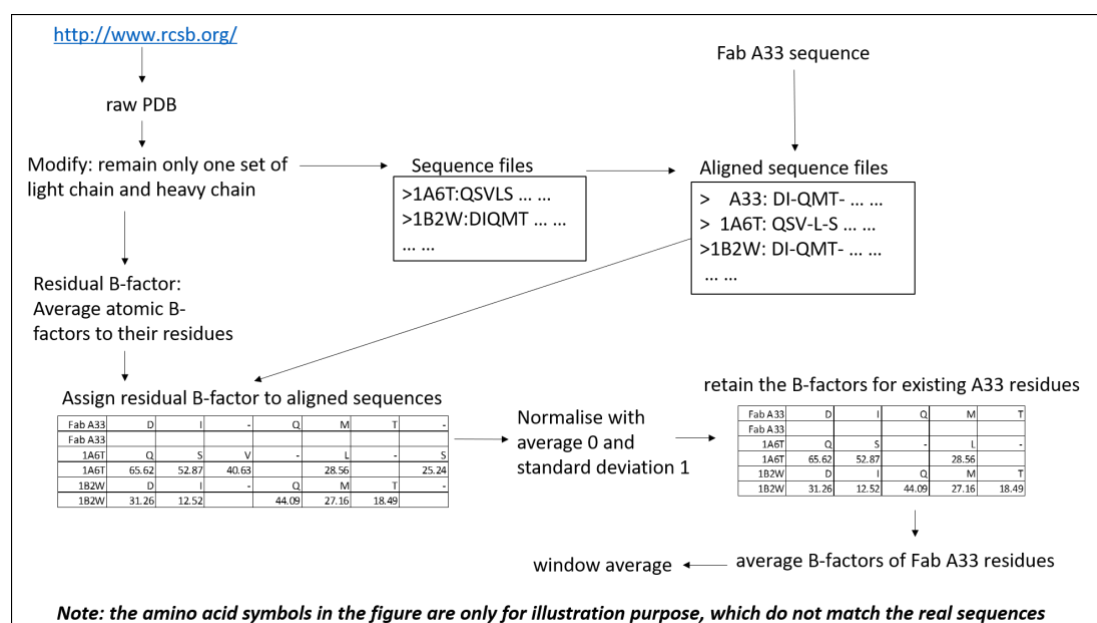
Figure 3. The schematic diagram for processing the B-factor

Raw PDB profiles were downloaded from the website, and only one set of light chain and heavy chain was retained for each profile. Amino acid sequence files were extracted from these modified PDB, and were then aligned against the Fab A33 sequence file. In addition, the residual B-factors were obtained by averaging their atomic B-factors, and tabulated along the amino acid sequence alignment. Normalisation was conducted for the tabulated "B-factor/sequence alignment" and the averaged B-factors accounting for Fab A33 residues were retained, and were further processed by window-averaging across 5 residues.
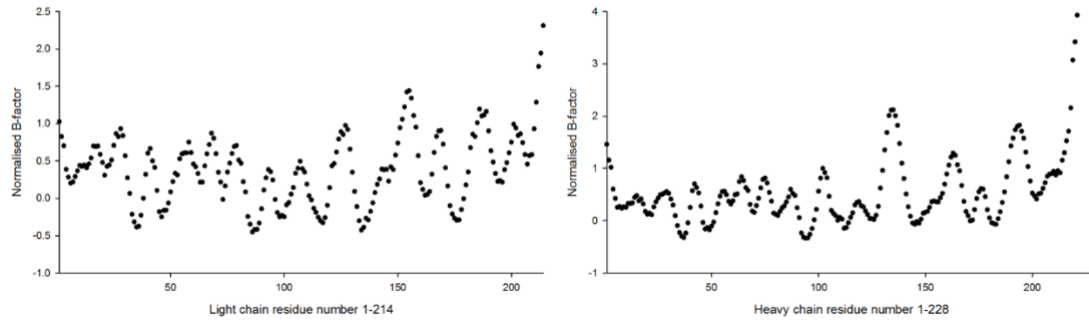
Figure 4. The normalised B-factor after window averaging with 5 neighbouring residues

The B-factors were averaged, normalised and window-averaged across 5 residues. The last 7 residues of heavy chain have no B-factor values as this part of sequence has no consensus with any of the other human Fab used for sequence alignment. Reproduced from C. Zhang 2018 with permission from ACS Publications*(18)*.
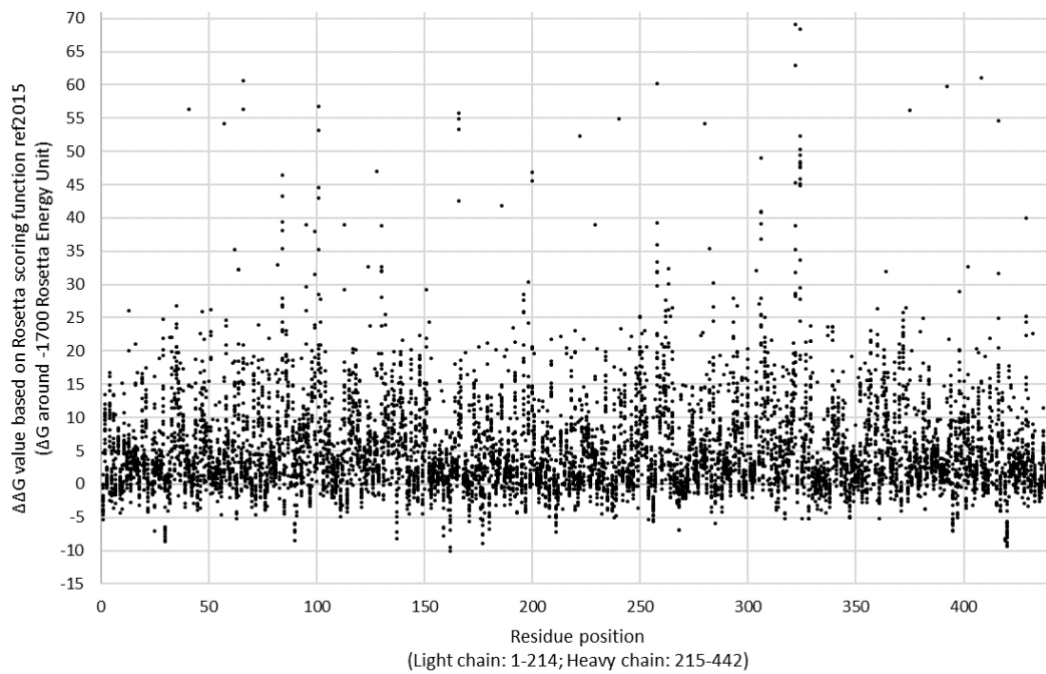
Figure 5. ΔΔG for 8398 variants of single point mutations. The Rosetta cartesian_ddg method was used to calculate the ΔΔG for 19 mutations per residue across the 442 residues.
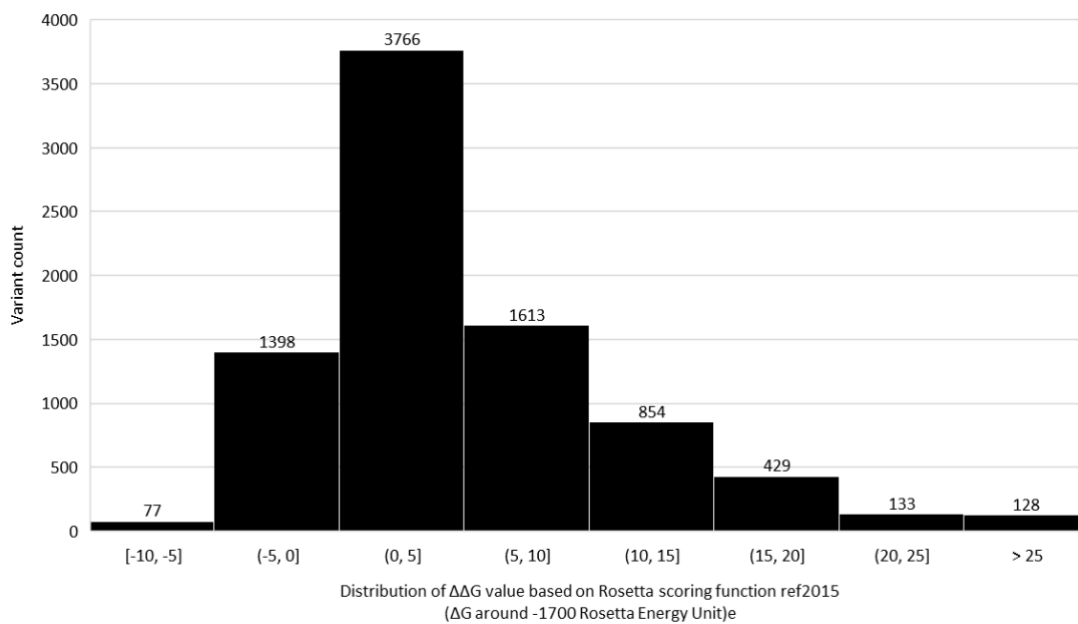
Figure 6. Histogram for the variant frequency distribution based on ΔΔG
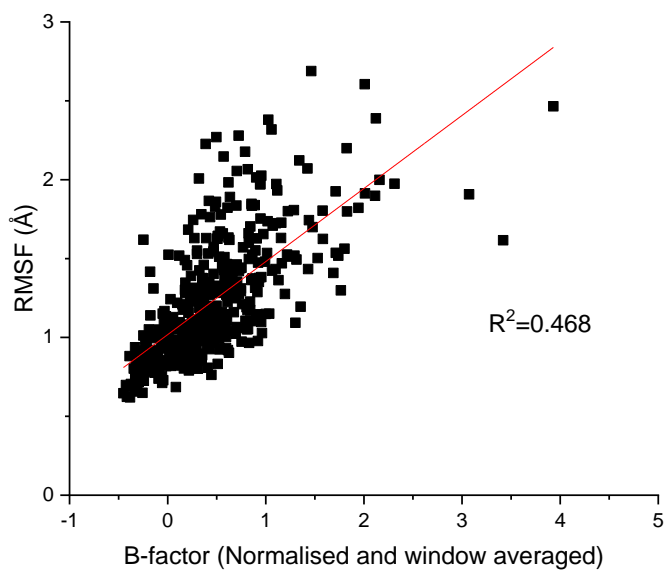
Figure 7. The correlation between RMSF and B-factor. RMSF and B-factor values were from Figure 2 and Figure 4, respectively. Reproduced from C. Zhang 2018 with permission from ACS Publications(18).
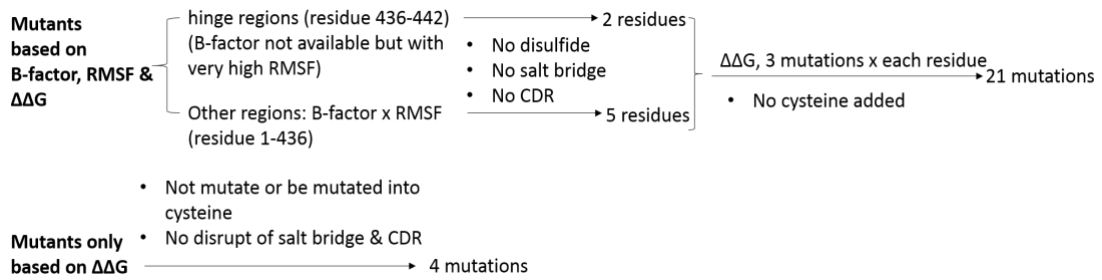
Figure 8. The schematic diagram to design stable mutants

Mutants were designed based on a combined analysis of B-factor, RMSF and ΔΔG. For hinge regions, residue 441 and 442 were selected for mutation. For other regions, the top residues with highest B-factor and RMSF were selected. For the selected residues, each was mutated into the three amino acids predicted by Rosetta to have the lowest ΔΔG values from across all 19 candidates. There were also mutations that only based on lowest ΔΔG. All the mutations had no influence on disulfide bond, salt bridge, CDR and cysteine addition. The number of mutations are used for planning, while the actual expression will depend on the real situation.

Table 1. The Complementarity Determining Regions (CDR) of C226S

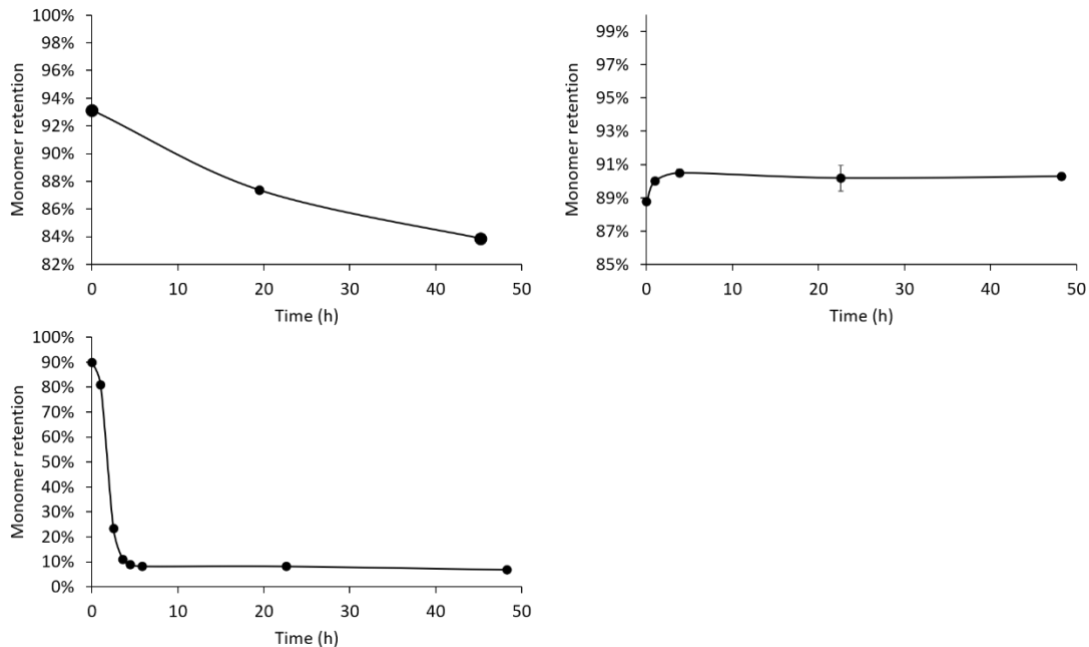| Light chain	Heavy chain
CDR-L1: 24-KASQNVRTVVA	CDR-H1: 26-GFAFSTYDMS
CDR-L2: 50-LASNRHT	CDR-H2: 50-TISSGGSYTYYLDSVKG
CDR-L3: 89-LQHWSYPLT	CDR-H3: 99-TTVVPFAY

Figure 9. Monomer retention kinetics of 1 mg/ml Fab at different formulation conditions, with error bars as standard error of the mean. (A) 65°C at 20 mM acetate, pH 5, with NaCl to bring the total ionic strength to 200 mM; (B) 45°C at 20 mM citrate, pH 4, with NaCl to bring the total ionic strength to 200 mM; (C) 65°C at 20 mM citrate, pH 4, with NaCl to bring the total ionic strength to 200 mM.

$$I_T = \frac{(I_N + aT) + (I_D + bT)\exp[\frac{\Delta H_{vh}}{R}(\frac{1}{T_m} - \frac{1}{T})]}{1 + \exp[\frac{\Delta H_{vh}}{R}(\frac{1}{T_m} - \frac{1}{T})]}$$

Equation 1

$$\Delta S_{vh} = \frac{\Delta H_{vh}}{T_m}$$

Equation 2

$$f_T = \frac{I_T - I_N - aT}{I_D + bT - I_N - aT} = \frac{\exp[\frac{\Delta H_{vh}}{R}(\frac{1}{T_m} - \frac{1}{T})]}{1 + \exp[\frac{\Delta H_{vh}}{R}(\frac{1}{T_m} - \frac{1}{T})]}$$

Equation 3

$$T_{on} = \frac{\Delta H_{vh}}{\Delta S_{vh} - R\ln\left(\frac{f_T}{1 - f_T}\right)} \quad (set\ f_T = 0.02)$$

Equation 4

A

B
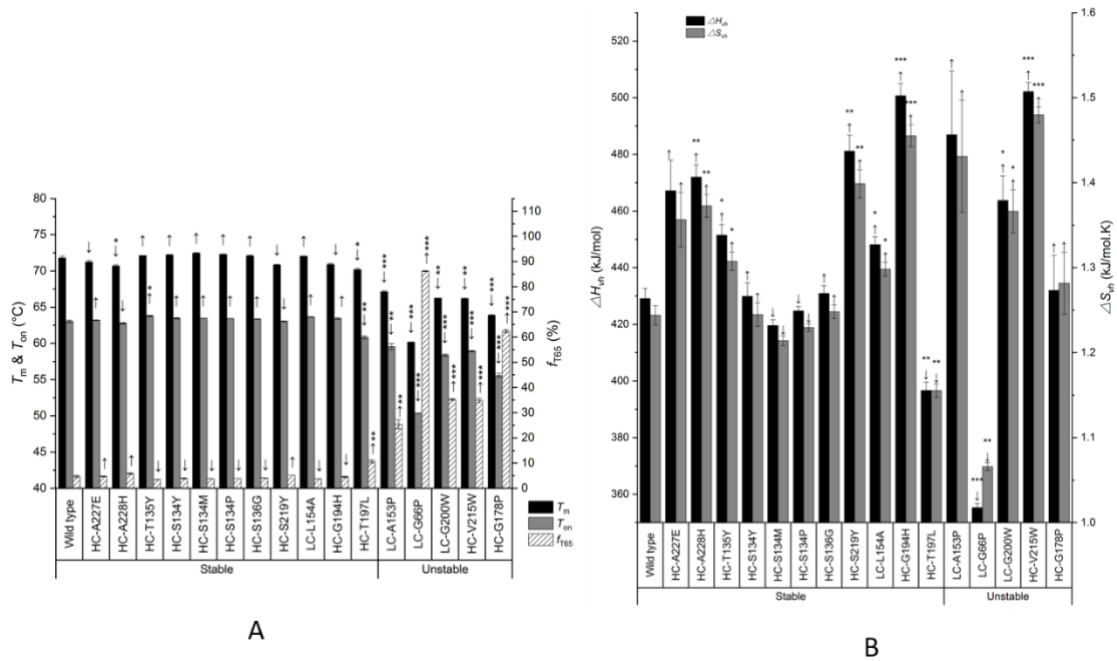
Figure 10. (A) The melting temperature (T_m, dark bars), temperature when 2% was unfolded (T_{on}, grey bars) and fraction of unfolding at 65°C (f_{T65}, tilt-line bars) as measured by UNit in triplicates at 1 mg/ml, pH 4 of 20 mM sodium citrate with NaCl to bring the total ionic strength to 200 mM. (B) The enthalpy and entropy changes at the midpoint of transition, ΔH_{vh} and ΔS_{vh}, derived from van't Hoff analysis. All error bars were standard error of the mean (SEM). Arrow ↑ or ↓ indicated an increase or decrease in average values of variants compared to the wild type. Two-sample t-test assuming unequal variances were performed between A33 wild type and other variants (*** $p<0.001$, ** $p<0.01$, * $p<0.05$). Reproduced from C. Zhang 2018 with permission from ACS Publications**(18)**.

$$y = A \times e^{(-kt)} \qquad\qquad \text{Equation 5}$$

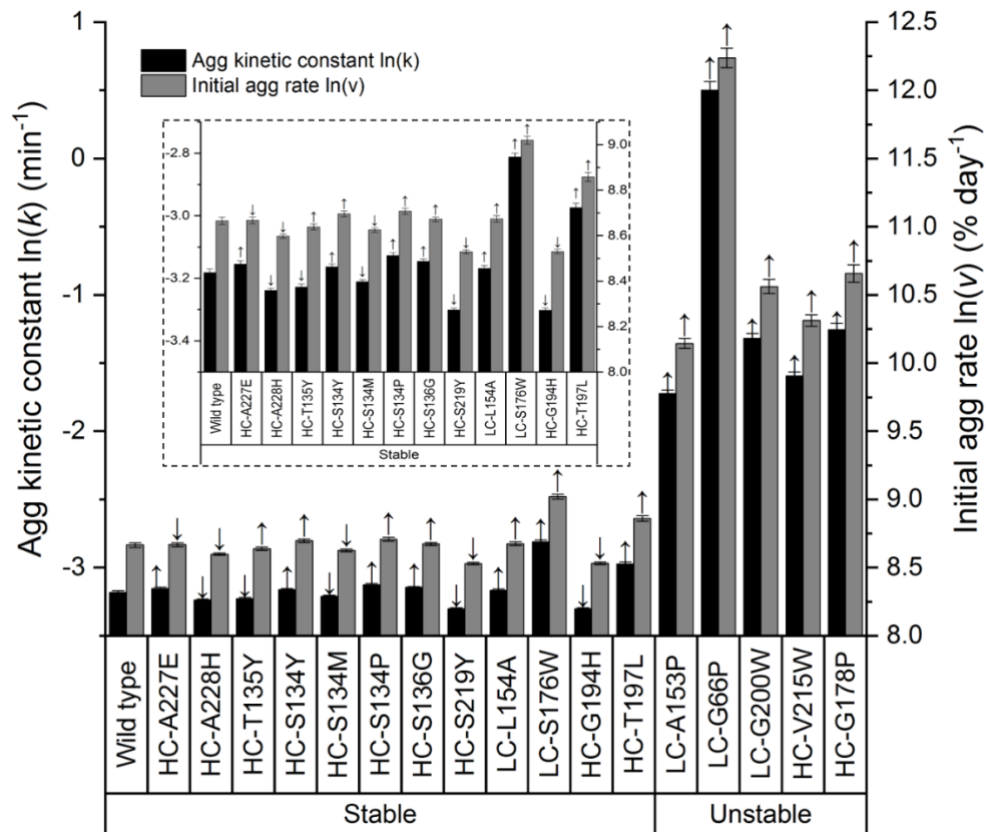$$\frac{d(y)}{d(t)} = -A \times k \times e^{(-kt)} \qquad\qquad \text{Equation 6}$$

Figure 11. The aggregation kinetic constants and initial aggregation rates of A33 Fab wild type and designed variants.

*The aggregation kinetic constant ln(k) and initial aggregation rates of A33 Fab wild type and designed variants. They were derived from the exponential equations, fitted from the monomer retention curves. Arrows ↑ or ↓ indicate increased or decreased values compared to those of wild type. Error bars were standard errors, calculated from curve-fitting derived parameters. Inset: Expanded view of stable variants only. Reproduced from C. Zhang 2018 with permission from ACS Publications**(18)**.*
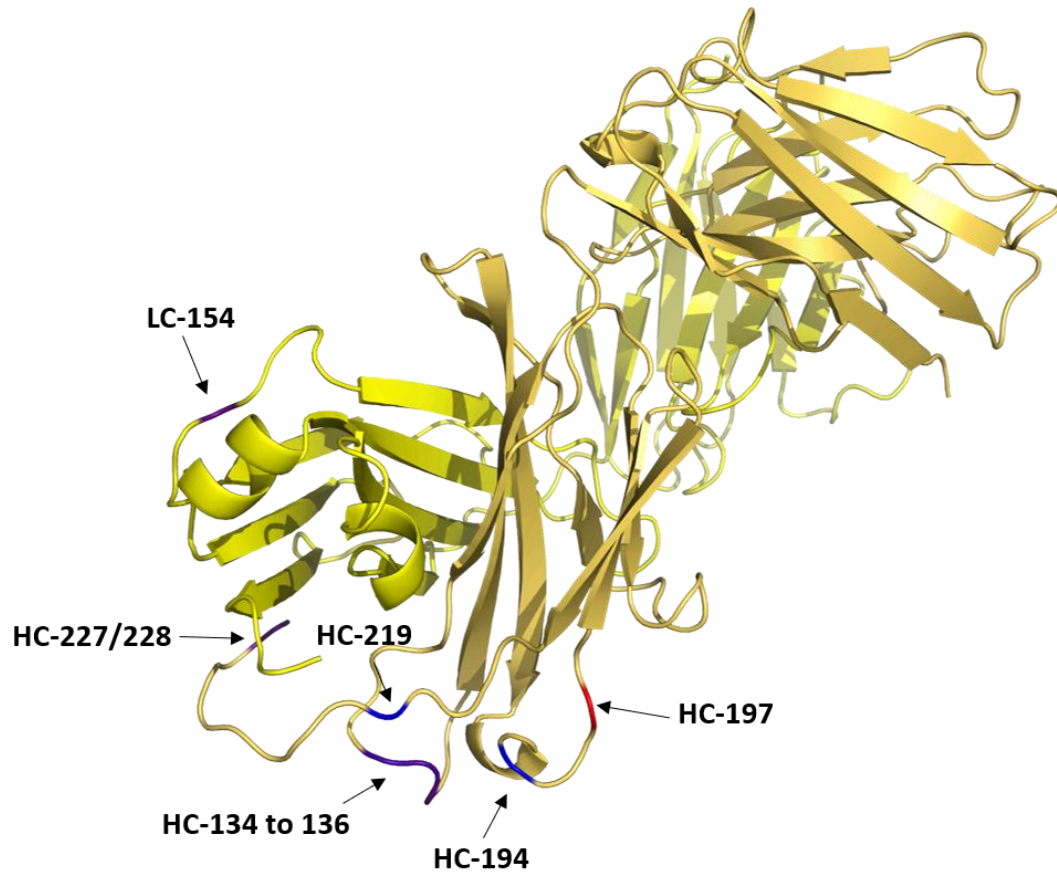
Figure 12. The locations of stable variants coloured by their aggregation rate

The locations of stable variants with aggregation rate ln(v) 8.53 to 8.86 % day^{-1} were coloured in gradient from blue to red, indicated by arrows. The light chain and heavy chain were coloured in yellow and orange, respectively. Reproduced from C. Zhang 2018 with permission from ACS Publications[(18)].
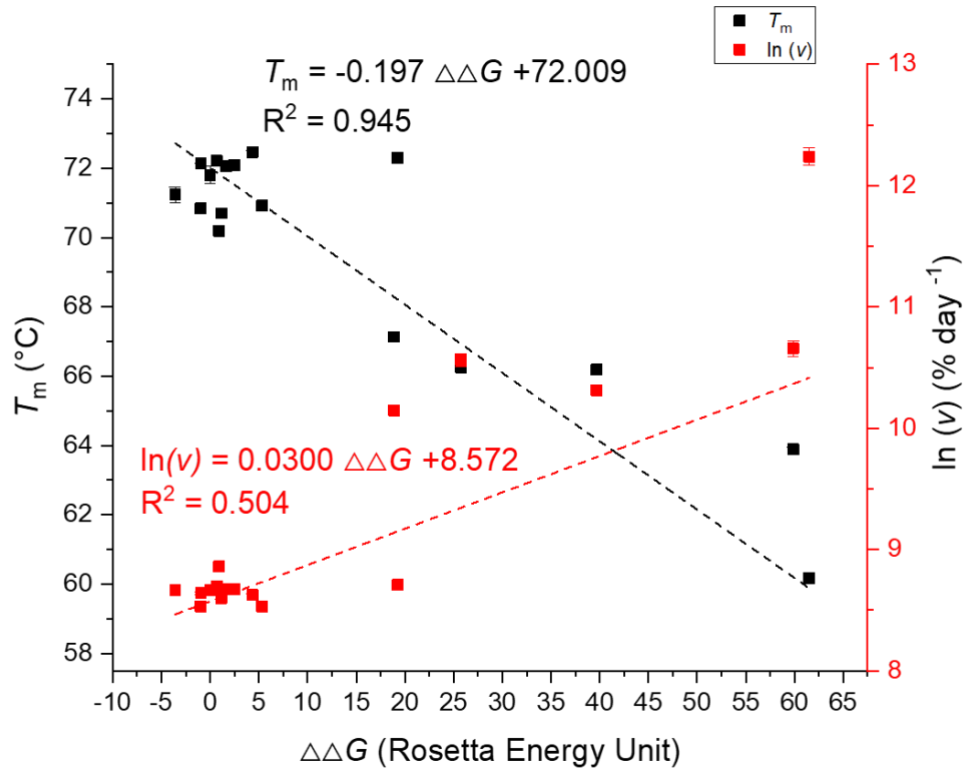
Figure 13. Correlations between △△G and Tₘ, ln(v). Error bars are standard error of the mean (SEM) for Tₘ and standard error (SE) for ln(v).