

Prediction of Aggregation Prone Regions of Therapeutic Proteins

Naresh Chennamsetty,[†] Vladimir Voynov,[†] Veysel Kayser,[†] Bernhard Helk,[‡] and Bernhardt L. Trout^{*,†}

Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, and Novartis Pharma AG, CH-4002, Basel, Switzerland

Received: December 10, 2009; Revised Manuscript Received: April 6, 2010

Therapeutic proteins such as antibodies are playing an increasingly prominent role in the treatment of numerous diseases including cancer and rheumatoid arthritis. However, these proteins tend to degrade due to aggregation during manufacture and storage. Aggregation decreases protein activity and raises concerns about an immunological response. We have recently developed a method based on full antibody atomistic simulations to predict antibody aggregation prone regions [*Proc. Natl. Acad. Sci.* 2009, 106, 11937]. This method is based on “spatial-aggregation-propensity (SAP)”, a measure of the dynamic exposure of hydrophobic patches. In the present paper, we expand on this method to analyze the aggregation prone regions over a wide parameter range. We also explore the effect of different hydrophilic mutations on these predicted aggregation prone regions to engineer antibodies with enhanced stability. The mutation to lysine is more effective than serine but less effective than glutamic acid in enhancing antibody stability. Furthermore, we show that multiple simultaneous mutations on different SAP peaks can have a cumulative effect on enhancing protein stability. We also investigate the accuracy of various cheaper alternatives for SAP evaluation because the full antibody atomistic simulations are highly computationally expensive. These cheaper alternatives include antibody fragment (Fab, Fc) simulations, implicit solvent models, or direct computations from a static structure (i.e., a structure from X-ray or homology modeling). The SAP evaluation from the static structure is 200 000 times faster but less accurate compared to the SAP from explicit atom simulations. Nevertheless, the SAP from a static structure still predicts most of the major aggregation prone regions, making it a potential approach for use in high-throughput applications. Thus, the SAP technology described here could be employed either in high-throughput developability screening of therapeutic protein candidates or to improve their stability at later stages of manufacturing.

Introduction

Therapeutic proteins such as antibodies currently constitute the most rapidly growing class of human therapeutics for various cancers and chronic inflammatory diseases.¹ Their market has been growing yearly at the highest rate (~35%) among all categories of biotech drugs and far exceeding that of small molecule drugs (6–8%).² One of the major problems encountered in antibody-based therapies is that these antibodies tend to aggregate under the high concentration formulations required for disease treatment.³ Aggregation leads to a decrease in antibody activity and raises concerns about causing an immunological response.^{4–8} Even aggregation of a few percent can be extremely significant. However, the antibody regions prone to aggregation are unknown. Stabilization of therapeutic antibodies is generally performed during the development phase using trial and error methods, which are both costly and time-consuming. Therefore, there is a tremendous need to understand the antibody aggregation prone regions and to devise methods to prevent aggregation.

Much of the earlier work in understanding protein aggregation has been done on amyloid fibril formation or the aggregation of small proteins where β -sheet formation is predominant. In contrast, there is relatively little work on predicting the aggregation prone regions of antibodies. While there are many

possible mechanisms for aggregation^{9–11} or the kind of interactions involved, hydrophobic interactions were shown to be the predominant interactions in extensive studies of protein folding and protein–protein binding.^{12–19} The prior work in predicting protein aggregation prone regions can be broadly divided into two categories: (1) phenomenological models and (2) molecular simulation techniques. The phenomenological models are based on the application of physicochemical properties such as hydrophobicity, β -sheet propensity, etc. to attempt to identify aggregation prone regions from protein primary sequences.^{20–26} While these phenomenological models were shown to perform well for small peptides and denatured proteins, aggregation propensities might differ for globular proteins such as antibodies where the tertiary structure and the stability of the native state are important. In contrast, molecular simulation techniques use three-dimensional structures and the dynamics of proteins to locate the regions prone to aggregation.^{27–35} For example, detailed atomistic models were used to study the aggregation of heptapeptide GNNQQNY³⁴ and A β _{16–22} amyloid peptide KLVFFAE.³⁵ These systems, however, involved simulation of very small peptides that contain seven residues each. In the case of antibodies that contain around 1300 residues, employing these techniques to simulate the aggregation of multiple antibodies would be extremely computationally demanding.

We have recently developed a tool called “spatial-aggregation-propensity (SAP)” based on full antibody atomistic simulations to predict antibody aggregation prone regions.^{36–38} SAP gives the effective dynamically exposed hydrophobicity of a

* Corresponding author. E-mail: trout@mit.edu.

[†] Massachusetts Institute of Technology.

[‡] Novartis Pharma AG.

certain patch on the protein surface. We applied SAP to therapeutic antibodies and determined their critical regions for aggregation. Mutating the residues in these critical regions experimentally from hydrophobic to hydrophilic residues enhanced the stability of these therapeutic antibodies significantly.³⁶ We also used the SAP tool to identify the aggregation prone motifs that are common to IgG1, IgG2, IgG3, and IgG4 subclasses of antibodies.³⁷

In the current work, we expand the SAP tool to perform a detailed analysis of the antibody aggregation prone regions. We vary the SAP parameters over a wide range to examine the aggregation prone regions at different resolutions. Whereas the SAP at low resolution identifies the major aggregation prone regions, the SAP at high resolution identifies the sites to be mutated for protein stabilization. We explore the effect of different hydrophilic mutations (lysine, serine, or glutamic acid) on protein stability experimentally. We also prepare a triple-site mutant to examine if multiple simultaneous mutations can have a cumulative effect on antibody stability. Since the full antibody explicit solvent simulations are very computationally demanding, we explore fragment simulations and various implicit solvent models to speed up the SAP calculations. We also show that an approximate value of SAP can be obtained directly from the static structure itself for high-throughput applications, albeit with some loss of accuracy. Thus the SAP tool could be employed either in high-throughput developability screening of antibody candidates or to improve their stability at later stages of manufacturing.

Methods

Molecular Simulations of Full Antibody. Molecular dynamics simulations are performed for a full antibody using an all-atom model with explicit solvent. The starting structure for simulation is obtained by attaching the X-ray structures of individual Fab and Fc fragments of the antibody. The X-ray structure of the Fab fragment is obtained from Novartis Pharma AG. The X-ray structure of the Fc fragment is obtained from that of another IgG1 antibody of similar sequence, 1HZH.³⁹ The structure of a full antibody is then obtained by aligning the Fab and Fc fragments using 1HZH structure as a model template. We call this antibody-A. To align the fragments at the correct distance and orientation, we minimize the rmsd (root mean square deviation) between the common CYS residues of the fragments and the full antibody template (1HZH). This structure is then used to perform explicit atom simulations for 30 ns. The CYS residues in the resulting antibody-A are all involved in disulfide bonds, including the ones in the hinge region. The glycosylation of the Fc fragment of the antibody is usually a mixture of G0, G1, and G2 etc. depending on the cell type and the growth conditions used in producing the antibodies. We use a G0 glycosylation pattern for our simulations since this is the most common glycosylation pattern observed in antibodies.

We use the CHARMM simulation package⁴⁰ for setup and analysis and the NAMD package⁴¹ for performing simulations. The CHARMM fully atomistic force field⁴² is used for the protein and TIP3P⁴³ solvent model for water. The simulations are performed at 298 K and 1 atm in the NPT ensemble. The parameters for the sugar groups involved in glycosylation of the Fc fragment are derived in consistence with the CHARMM force field, following from the CSFF force field.⁴⁴ The protonation states of Histidine residues at pH-7 are decided based on the spatial proximity of electro-negative groups from the Histidine ring. The proton is assigned to the nitrogen of the Histidine ring with the closest electron donor within 3 Å

distance, which has a potential to form a hydrogen bond. If none of the nitrogens have electron donors within 3 Å, then the proton is assigned to ND1 (delta nitrogen). The full antibody is solvated in an orthorhombic box since this minimizes the number of water molecules required and thus minimizes the computational time required. Periodic boundary conditions are used in all three directions. A water solvation shell of 8 Å is used in each direction of the orthorhombic box. The resulting total system size is 202 130 atoms. Simulating this system for 30 ns involved a total computational cost of 1536 CPU days. However, it was accomplished within 30 days using 128 processors in parallel on supercomputer resources (Teragrid, NCSA Tungsten cluster). The orthorhombic box remains stable during the 30 ns simulation without any significant change in box dimensions on all three axes. The initial box dimensions are 161.9, 145.4, and 83.2 Å, respectively, and they change very little during the 30 ns simulation, ending at 161.2, 144.7, and 82.8 Å, respectively. The antibody does not rotate significantly during the 30 ns simulation, thereby maintaining the minimum distance between the antibody and its periodic images of more than 14 Å. Sufficient ions are added to neutralize the total charge of the system. The charge neutrality is required by the Ewald summation technique that is used to calculate contribution due to the electrostatic interactions.

After the antibody is solvated, the energy is initially minimized with SD (steepest descents) by fixing the protein to allow the water to relax around the protein. Then the restraints are removed and the structure is further minimized with SD and ABNR (adopted basis Newton–Raphson). The system is then slowly heated to room temperature with 5 °C increments every 0.5 ps using a 1 fs time step. The system is then equilibrated for 1 ns before we start computing the various properties from simulation. The configurations are saved every 0.1 ps during the simulation for further statistical analysis.

Fragment and Implicit Solvent Simulations. The Fab and Fc fragments of antibody-A and antibody-B are simulated for 30 ns each. The X-ray structure of the Fab fragment of antibody-A is obtained from Novartis Pharma AG, whereas that of antibody-B is built using homology modeling (described in the next section). The X-ray structure of Fc fragments of both antibodies is similar because they both belong to the IgG1 subclass. Therefore, their Fc fragment structure is obtained from that of another IgG1 antibody of similar sequence, 1HZH.³⁹ These Fab and Fc fragment structures are solvated in a cubic box with a water solvation shell of 8 Å in the longest dimension. The rest of the setup is similar to that of full antibody-A simulation. Simulating these fragments for 30 ns involved a total computational cost of 810 CPU days each. However, it was accomplished within 15 days using 64 processors in parallel on supercomputer resources. The implicit solvent models, GBSW, ACE, and EEF1, were all run on a single processor (2.4 GHz, 2GB RAM) as implemented in the CHARMM simulation package.

Homology Modeling. The X-ray structure of antibody-B is not available, and its structure is built using homology modeling. The sequence is obtained from Novartis Pharma AG. In the first step, the variable fragment (F_V) structure is built from its sequence using the canonical structure method^{45–47} in combination with homology modeling. This involved identifying the respective canonical structures for each of the CDR loops and modeling the rest of the antibody framework through homology modeling. We use the modeling tool WAM⁴⁸ to identify the canonical structures. To validate the homology method as giving us the accurate structure, we use this method to initially obtain

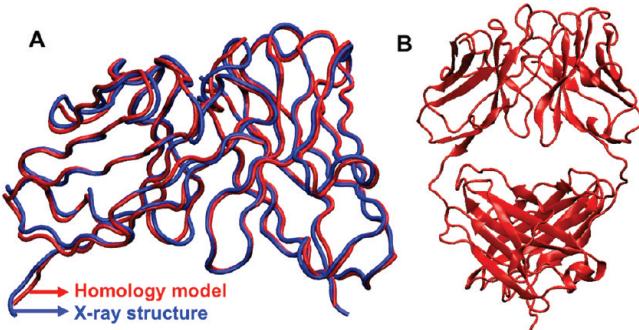


Figure 1. (A) Structure of the variable fragment (F_v) for antibody-A obtained using homology modeling compared with its X-ray structure. (B) The final structure for the Fab fragment of antibody-B obtained using homology modeling.

a structure for antibody-A, whose X-ray structure is already known. The resulting homology model is compared with the known X-ray structure in Figure 1A. The two structures in Figure 1A are almost identical with very minor deviations. Therefore, this homology modeling approach gives structure with good accuracy. The structure of the antibody-B variable fragment is then determined using the same homology modeling approach to generate the final Fab fragment structure as shown in Figure 1B.

Protein Expression. Vectors that carry the light chain or the heavy chain genes of the IgG1 antibodies antibody-A and antibody-B were obtained by subcloning the genes from proprietary vectors (Novartis) into a gWIZ vector (Genlantis), optimized for high expression from transiently transfected mammalian cells. Antibody variants were generated using site-directed mutagenesis by PCR. All constructs were confirmed by DNA sequencing. Plasmid DNA at the milligram scale was purified from bacterial cultures with DNA Maxi Prep columns (Invitrogen). Tissue culture and transient transfection of FreeStyle HEK 293 cells were carried out following the manufacturer's protocols (Invitrogen), except that polyethyleneimine (Polysciences) was used as the transfection reagent. Transfected cells were incubated in a CO_2 incubator at 37 °C for 7–9 days.

Protein Purification. Antibody wild type and variants were purified from the tissue culture supernatant on a Protein A column (GE Healthcare). Antibodies were eluted from the column with 50 mM citrate buffer pH 3.5 and equilibrated to pH 6.6–7.0 with 1 M Tris-HCl pH 9.0 and further purified on a Q Sepharose column (GE Healthcare). Solutions of antibody-A wild type and variants were further concentrated with 30K MWCO filters and buffer exchanged with 20 mM His buffer pH 6.5 to a final concentration of 150 mg/mL. Solutions of antibody-B wild type and variants were also concentrated with 30K MWCO filters and buffer exchanged with 10 mM His buffer pH 6.0 to a final concentration of 60 mg/mL. As a quality control, aliquots of the purified and concentrated samples were analyzed by nonreducing and reducing SDS-PAGE and by circular dichroism.

SEC-HPLC. Size-exclusion high-performance liquid chromatography (SEC-HPLC) was used to determine monomer loss over time in accelerated aggregation experiments. Antibody-A wild type and variants were incubated at 58 °C at 150 mg/mL and antibody-B wild type and variants at 52 °C at 60 mg/mL. We used several criteria in selecting these particular conditions of concentration, temperature, and time of incubation for stressing antibody-A and antibody-B for SEC analysis. Since our primary interest is in understanding aggregation at the high protein concentration that therapeutic antibodies are usually

used, we stressed antibody-A and antibody-B at concentrations of 150 and 60 mg/mL that are the respective concentrations of formulations for these antibodies. We balanced the choice of temperature and time for stressing antibodies by using low enough temperatures that are at least 10 degrees lower than the lowest melting transition of each antibody. At the same time, we used high enough temperatures to obtain a few percent of aggregates within a day or two of incubation to minimize the effect of water evaporation on the overall protein concentration of the samples. Stressed samples were diluted in 15 mM potassium phosphate buffer pH 6.5 to 10 mg/mL. Monomers were resolved from nonmonomeric species by SEC-HPLC (Agilent 1200 LC) on a TSKgel Super SW3000 column (TOSOH Bioscience), kept at 22 °C, with mobile phase 150 mM potassium phosphate, pH 6.5, and flow rate of 0.2 mL/min. Percent monomer was calculated as the area of the monomeric peak divided by the total area of all peaks detected at 280 nm by an in-line UV-signal detector. These percent monomer values are averaged over multiple runs (typically two to four runs), and the standard deviations are reported as error bars within the figures. Typical SEC-HPLC chromatograms are given for two variants in the Supporting Information, Figure S1 and Figure S2. These figures show that the aggregate peaks are well separated from the monomer peaks, enabling us to quantify the aggregation amount with good accuracy.

Results and Discussion

The hydrophobic residues are usually buried within the protein core, while a few of them might be exposed. The dynamic fluctuations that protein goes through from its native state could further expose more hydrophobic residues. These dynamically exposed residues can combine with the surface residues to form larger hydrophobic patches on the surface of the antibody. It is unfavorable for the hydrophobic patches to be exposed to water. Thus, they could combine with the hydrophobic patches on other antibodies to form aggregates. We perform simulation of a full antibody to identify such hydrophobic patches prone to aggregation on the surface of the antibody. One common way to find the surface exposure of different residues is through solvent accessible area (SAA). SAA gives an estimate of the surface area of each residue that is in contact with the solvent. A high SAA would indicate that the residue is exposed. However, SAA by itself does not provide the correct estimate of hydrophobic patches because of the following shortcomings:

- (1) SAA does not distinguish between hydrophobic and hydrophilic regions.

- (2) SAA is not directly proportional to a residue's hydrophobicity (for example, MET has more surface area than LEU but is less hydrophobic).

- (3) SAA does not indicate whether several hydrophobic residues are in close proximity and thus could enhance the hydrophobicity of a certain region. These residues could be nearby either in primary sequence or in the tertiary structure. Either way, their proximity enhances the hydrophobicity in that region on the antibody surface.

The values of SAA for different residues in the antibody Fc fragment are presented in Figure 2C. These SAAs are determined from the Lee and Richard surface approach as implemented in the CHARMM simulation package. This is defined as the surface resulting from rolling a probe sphere on the van der Waals spheres of a CPK model of the molecule. We use a probe sphere of radius 1.4 Å, which is equivalent to that of a water molecule. We observe many peaks in SAA, most of which

are for hydrophilic residues. However, it would be difficult to obtain a measure of hydrophobic patches from these SAA values.

SAP Tool to Predict Aggregation Prone Regions. We account for the shortcomings in SAA by defining a new parameter called spatial-aggregation-propensity (SAP) that gives the effective dynamically exposed hydrophobicity of a certain patch on the protein surface.³⁶ The new parameter, SAP, is illustrated in Figure 2A and is defined as

$$(spatial-aggregation-propensity (SAP))_{atom i} =$$

$$\sum_{\text{simulation av}} \left\{ \sum_{\substack{\text{residues with at least} \\ \text{one side chain atom} \\ \text{within } R \text{ from atom } i}} \frac{\text{SAA of side chain atoms} \\ \text{within radius } R}{\text{SAA of side chain atoms} \\ \text{of fully exposed residue}} \times \right. \\ \left. \text{residue hydrophobicity} \right\}$$

Here

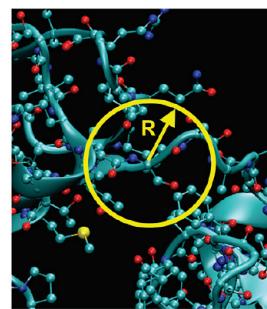
(1) SAA of side chain atoms within radius R is computed at each simulation snapshot.

(2) SAA of the side chain of fully exposed residue (say for amino acid "X") is obtained by calculating the SAA of side chains of the middle residue in the fully extended conformation of tripeptide "Ala-X-Ala". These SAA values for side chains of fully exposed residues are given in Table 1.

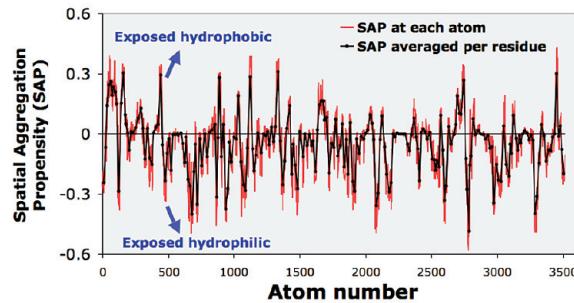
(3) Residue hydrophobicity is obtained by normalizing the hydrophobicity scale of Black and Mould.⁴⁹ The scale is normalized such that Glycine has a hydrophobicity of zero (values shown in Table 1). Therefore, amino acids that are more hydrophobic than Glycine are positive and less hydrophobic than Glycine and are negative on the hydrophobic scale.

The spatial-aggregation-propensity (SAP) is calculated for spherical regions with radius R centered on every atom in the antibody. This gives a unique SAP value for each atom. Then the SAP for a residue is obtained by averaging the SAP of all its constituent atoms. The atomic and residue-averaged SAP values thus evaluated at $R = 5 \text{ \AA}$ for the Fc fragment of antibody-A are shown in Figure 2B. Here the positive SAP values stand for hydrophobic regions, while negative values stand for hydrophilic regions. Within the positive SAP regions, a high peak indicates a highly exposed hydrophobic region, while a low value indicates a buried hydrophobic region. Similarly for the negative SAP regions, a huge dip indicates an exposed hydrophilic region, while a low value indicates a buried hydrophilic region. We notice more negative dips than positive peaks, indicating that most exposed regions are hydrophilic. This is as expected since most of the exposed protein surface is usually hydrophilic. The few SAP peaks that are positive indicate exposed hydrophobic regions. It is unfavorable for them to be exposed to water because of their hydrophobic nature. One way they can decrease their exposure is by interacting with similar regions on other proteins and getting buried, thereby leading to protein aggregation. The values of SAP are compared with that of SAA (solvent accessible area) in Figure 2C. The plot shows a clear difference between the SAP and SAA values and peak positions. We observe many positive peaks in SAA,

A



B



C

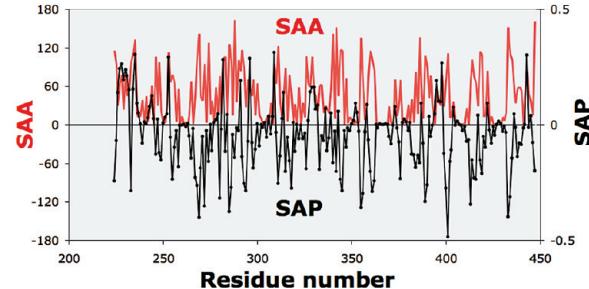


Figure 2. (A) Definition of spatial-aggregation-propensity (SAP) based on atoms within radius (R) from a given atom. (B) SAP values for the Fc fragment of the antibody with the radius of patch $R = 5 \text{ \AA}$. This is an average SAP value obtained from the 30 ns full antibody simulation. (C) The comparison of solvent-accessible-area (SAA) and spatial-aggregation-propensity (SAP) values for the Fc fragment of the antibody.

which are absent in SAP. Most of these SAA peaks are for the hydrophilic residues that do not have any correlation with the hydrophobic patches on the surface. The peaks in SAP, however, correctly indicate the hydrophobic patches. Therefore, SAP is a better parameter than SAA in finding the potential aggregation prone hydrophobic patches.

The residue averaged values of SAP thus evaluated for two different antibodies, antibody-A and antibody-B, are shown in Figure 3A and Figure 4A, respectively. These values are for one of the Fab fragments and chains of Fc, while the values for their identical sequence counterparts in the full antibody are similar. For antibody-A, the SAP values are obtained from 30 ns simulation of the full antibody structure that was obtained by assembling the X-ray structures of Fc and Fab fragments. For antibody-B, the SAP values are calculated from the Fab fragment simulations for 30 ns. The SAP values for the Fc fragment of antibody-B are taken from those of antibody-A since they both belong to the IgG1 class and share the same sequence.

These SAP values from simulation are mapped onto the antibody structure in Figure 3B and Figure 4B, respectively. In these figures, the antibody surface is colored according to the values of SAP. Positive values of SAP (hydrophobic) are colored in red, while negative values (hydrophilic) are in blue. The

TABLE 1: Hydrophobicity, Normalized Hydrophobicity, and SAA of Fully Exposed Side Chains for Different Residues^a

residue	hydrophobicity	normalized hydrophobicity	SAA of fully exposed side chains
PHE	1	0.499	186.7
ILE	0.943	0.442	151.242
LEU	0.943	0.442	139.524
TYR	0.880	0.379	200.306
TRP	0.878	0.377	229.619
VAL	0.825	0.324	124.237
MET	0.738	0.237	164.674
PRO	0.711	0.210	111.533
CYS	0.680	0.179	95.2439
ALA	0.616	0.115	64.7809
GLY	0.501	0	23.1338
THR	0.450	-0.051	111.597
SER	0.359	-0.142	81.2159
LYS	0.283	-0.218	177.366
GLN	0.251	-0.250	147.855
ASN	0.236	-0.265	113.187
HSD ^b	0.165	-0.336	146.449
HSE ^b	0.165	-0.336	149.451
GLU	0.043	-0.458	143.924
ASP	0.028	-0.473	110.209
ARG	0	-0.501	210.02

^a The hydrophobicity values are taken from Black and Mould.⁴⁹ The normalized hydrophobicity values used in SAP calculation are obtained by normalizing all hydrophobicity values with respect to Glycine. SAA stands for solvent accessible area, which is calculated according to the Lee and Richard approach within the CHARMM simulation package. ^b HSD is Histidine with hydrogen on the delta nitrogen, whereas HSE is Histidine with hydrogen on the epsilon nitrogen.

intensity of color is proportional to the magnitude of SAP. Therefore, a highly exposed hydrophobic patch would be deep red, and similarly a highly exposed hydrophilic will be deep blue. For both antibodies A and B, we observe that the surface is predominantly blue indicating that the surface is mostly hydrophilic. Again, this is as expected since most of the protein surface is usually hydrophilic. However, we also notice a few red areas indicating exposed hydrophobic regions. These red areas with high SAP are the hydrophobic patches prone to aggregation.

Mutating SAP Predicted Regions for Protein Stabilization. We have demonstrated earlier that mutations engineered experimentally on the peaks of SAP changing the hydrophobic residues to hydrophilic residues led to antibodies of enhanced stability.³⁶ The sites chosen for mutation in antibody-A and antibody-B are shown in Figure 3 and Figure 4, respectively (blue circles). Five mutants were generated for antibody-A (A_1 -L235K, A_2 -I253K, A_3 -L309K, A_4 -L235K L309K, A_5 -L234K L235K) and five mutants for antibody-B (B_1 -W94K, B_2 -W100K F101K, B_3 -L235K, B_4 -W94K W100K F101K, B_5 -W94K L235K). All these mutants involved mutation to lysine, a hydrophilic residue. The mutants are then tested for their aggregation behavior using accelerated aggregation experiments under heat stress. SEC-HPLC (size-exclusion high-performance liquid chromatography) was used to determine monomer loss over time after heat stress. The SEC-HPLC results for antibody-A indicate monomer increase from 91% for wild type to 93–96% for the variants, indicating enhanced stability of the mutants.³⁶ Similarly, for antibody-B, the SEC-HPLC results show monomer increases from 95% for wild type to 96–100% for the variants, except for two mutations (B_1 , B_5). These mutants B_1 and B_5 contain 3% aggregates upon purification as opposed to 1% for wild type,

possibly indicative of some folding or expression perturbation. Apart from this exception, the variants in general were more stable than the wild type validating that peaks in SAP represent the regions prone to aggregation.

In this report, we generate three additional mutants for antibody-A and one for antibody-B. These mutants are A_6 -L235S, A_7 -V282K, A_8 -(L235K V282K L309K), and B_6 -L235E. Whereas the earlier mutations (A_1 to A_5 and B_1 to B_5) were all to lysine (K), these new mutations also involved serine (S) and glutamic acid (E). These new mutants are intended to examine whether mutation to different hydrophilic residues (K, S, or E) has a different effect on aggregation propensity. The monomer loss determined from SEC-HPLC for these new mutants is shown in Figure 5. The SEC-HPLC results for antibody-A indicate a monomer increase from 91% for wild type to 92–97% for the new variants (A_6 to A_8), indicating enhanced stability of the mutants. There is a small increase in stability for antibody-B variant B_6 compared to its wild type (0.7% in 36 h). Thus, all the new variants are more stable than wild type providing additional validation that the peaks in SAP correspond to regions prone to aggregation. Furthermore, the mutation to serine (A_6 -L235S) gives less improvement in stability compared to the mutation to lysine (A_1 -L235K). This is probably because serine is smaller and less hydrophilic compared to lysine and thus could not inhibit the hydrophobic patch as much as lysine. This is also consistent with the change in the SAP value due to the mutation. SAP is proportional to the residue fraction exposed and the residue hydrophobicity. A mutation from a hydrophobic to a hydrophilic residue causes a large change in the residue hydrophobicity and only a minor or no change in the fraction exposed. Thus, we would expect a larger change in SAP and thereby a larger change in stability for the mutation A_1 -L235K compared to A_6 -L235S. As expected, we saw a larger improvement in stability for the mutation A_1 -L235K compared to A_6 -L235S. Thus, the change in SAP value is consistent and related to the change in protein stability.

In antibody-B, the mutation to glutamic acid (B_6 -L235E) led to a small increase in stability. Although this is small, it is still better than the mutation to lysine (B_3 -L235K).³⁶ In general, the mutation at the site L235 in antibody-B led to a smaller improvement in stability compared to mutation at the same position in antibody-A. This is probably because there are other larger hydrophobic patches in antibody-B (within CDR regions as shown by SAP) that are predominantly contributing to aggregation. We saw a huge improvement in antibody-B stability when we performed mutations within these CDR regions (e.g., mutations B_2 -W100K F101K and B_4 -W94K W100K F101K).³⁶ Thus, the hydrophobic patches in the antibody-B CDR regions are the most aggregation prone. In the case of antibody-A, however, SAP calculations showed that the CDR regions are not as hydrophobic compared to some of its Fc fragment regions. Therefore, mutations performed within the Fc fragment region of antibody-A (A_1 to A_8) led to a significant improvement in antibody-A stability. The triple mutant (A_8 -L235K V282K L309K) is the most stable of all antibody-A mutants. It is more stable than any of the single mutations at similar sites (A_1 -L235K, A_3 -L309K, and A_7 -V282K). The triple mutant's stability appears to be a cumulative effect of these three separate mutations. We would expect the change in SAP value due to mutation to be highest for the triple mutant, A_8 , compared to other mutants. Thus, the stability increase is consistent with and quantitatively related to the overall change in SAP value for the triple mutant, A_8 . This also shows that performing multiple mutations simultaneously on all the peaks of SAP could lead

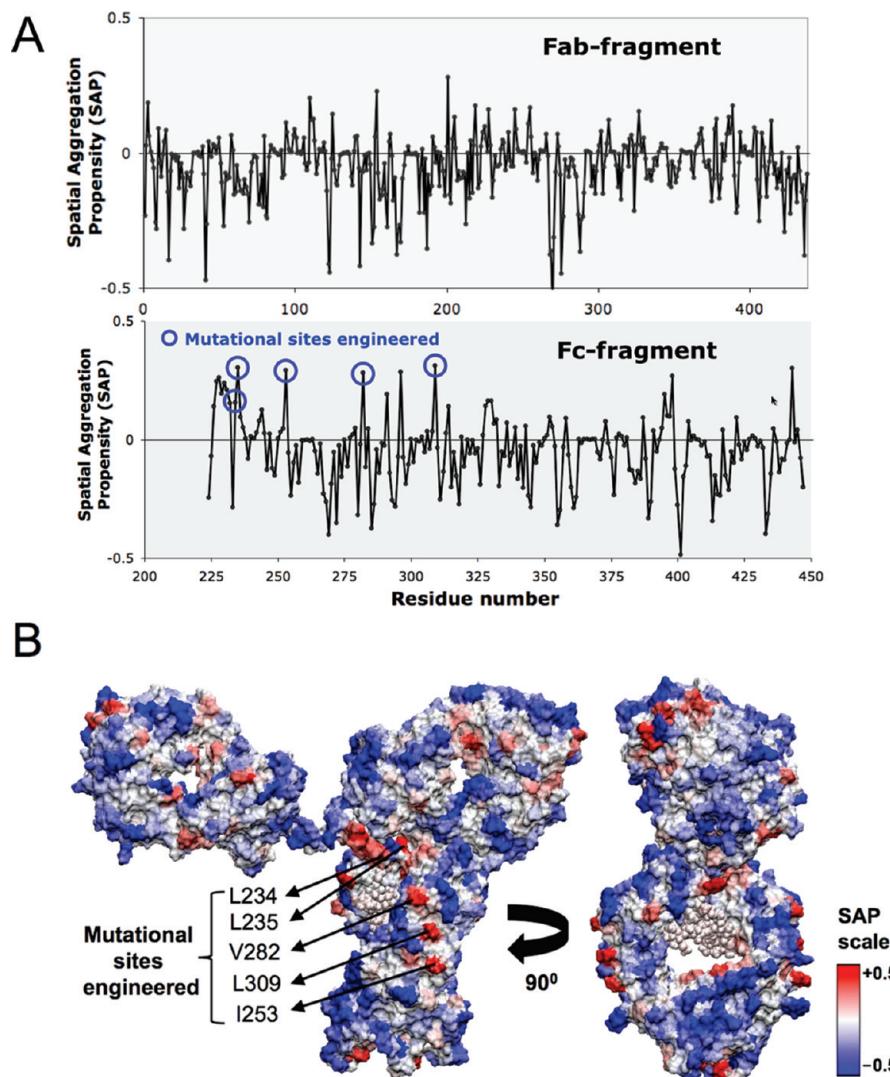


Figure 3. (A) Spatial-aggregation-propensity (SAP) values at $R = 5 \text{ \AA}$ for the Fab and Fc fragments of antibody A. The sites chosen for mutations are also indicated (blue circles: L234, L235, I253, V282, and L309). (B) The SAP values at $R = 5 \text{ \AA}$ values are mapped onto the antibody-A structure where red regions represent positive peaks and blue regions are negative dips. Again, the sites chosen for mutation are indicated.

to a most stable antibody. On the other hand, we note that control cases are needed to show whether the SAP identifies some or all of the aggregation prone regions. These control cases involve mutating to add charge on a low SAP neutral amino acid or region and testing whether it is charge rather than hydrophobicity that affects the aggregation rates. Such control cases are the subject of future work, as here the objective has been to develop a method, SAP, which correlates with experimental rates of aggregation. We have shown that placing the same charge at different locations (e.g., A₁-L235K, A₂-I253K, A₃-L309K, and A₇-V282K) leads to a difference in aggregation propensity. This shows that certain sites are more aggregation prone than others, and placing the charge at different sites does not have the same effect on improving stability. SAP identified some of these sites with high aggregation propensity where mutating to a charged residue leads to a high improvement in stability.

Effect of Changing SAP Resolution on the Prediction of Aggregation Prone Regions. Until now, we computed SAP values using a radius of 5 Å. This gives information about hydrophobic patches with a size of radius 5 Å. Increasing the SAP radius will decrease SAP resolution and will indicate if there are larger aggregation prone patches on the antibody.

Figure 6 and Figure 7 show the plots and the mapping of SAP values calculated with different radii ($R = 5, 7.5, 10, 15$, and 20 \AA). In Figure 6, we observe that going from lower radii of patches (5 Å) to the higher radii (20 Å) eliminates some peaks, whereas some other peaks are enhanced. Some of these peaks are eliminated at higher SAP radii because the small hydrophobic patches representing these peaks are usually surrounded by hydrophilic patches. Thus, averaging SAP over higher radii leads to an effective increase in hydrophilicity for the region, resulting in lesser peaks. In certain other regions, however, SAP is enhanced at higher radii because of hydrophobic patches surrounding a similar hydrophobic patch. We also notice that the SAP values become increasingly negative with an increase in the SAP radius. At a radius of $R = 20 \text{ \AA}$, almost all SAP values are negative except for one positive peak. This is because the protein surface is predominantly hydrophilic. Therefore, averaging over a larger radius makes the contribution from hydrophilic regions (with negative SAP) dominate over the hydrophobic regions (with positive SAP).

Figure 7 shows the mapping of SAP values calculated with different radii ($R = 5, 7.5, 10, 15$, and 20 \AA). Figure 7A also shows the direct mapping of the hydrophobicity scale⁴⁹ onto the antibody. Using this hydrophobicity scale to directly predict

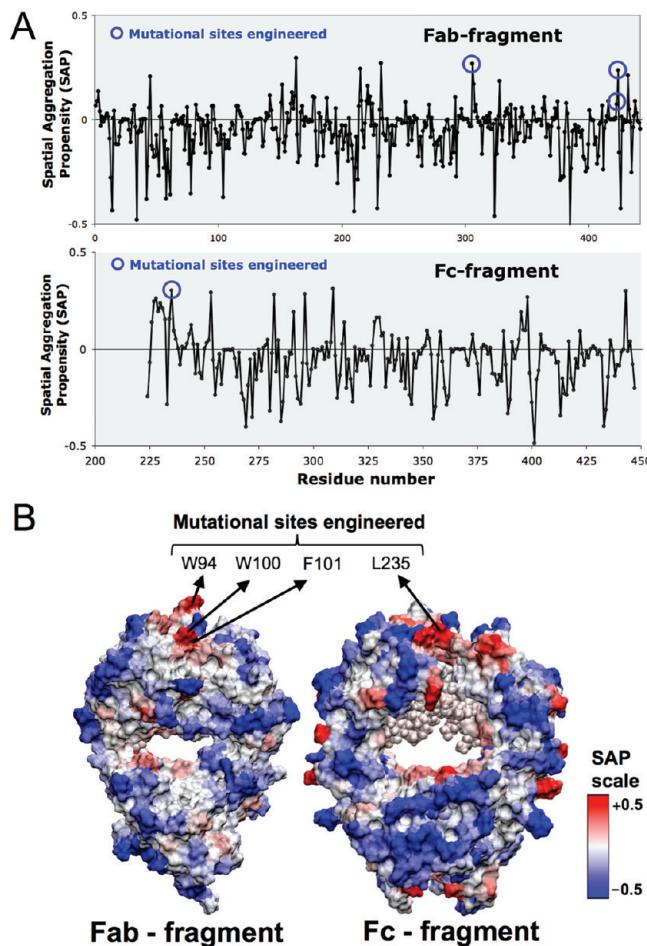


Figure 4. (A) Spatial-aggregation-propensity (SAP) values at $R = 5 \text{ \AA}$ for the Fab and Fc fragments of antibody-B. The sites chosen for mutations are also indicated (blue circles: W94, W100, F101, and L235). (B) The SAP values at $R = 5 \text{ \AA}$ are mapped onto the antibody-B structure where red regions represent positive peaks, and blue regions are negative dips. Again the sites chosen for mutation are indicated.

the major hydrophobic patches (Figure 7A) would be difficult because the hydrophobic regions appear to be homogeneously distributed. In contrast, SAP readily identifies the major aggregation prone regions as shown in Figure 7B to Figure 7F. While the SAP at low radius ($R = 5 \text{ \AA}$) gives a detailed view of the aggregation prone hydrophobic patches, the SAP values at higher radii ($R = 7.5$ and 10 \AA) show the regions of bigger patches. Going to an even higher radius ($R = 15$ and 20 \AA) makes the hydrophobic patches disappear because of the averaging over a larger area where the contribution from hydrophilic groups dominates. Thus, SAP loses resolution for $R > 10 \text{ \AA}$ and is no longer able to clearly identify the aggregation prone patches. SAP at $R = 10 \text{ \AA}$ is ideal for finding the bigger aggregation prone patches, whereas SAP at $R = 5 \text{ \AA}$ is ideal for a more detailed view of these patches. Whereas the SAP at low resolution (high SAP radius) identifies the major aggregation prone regions, the SAP at high resolution (low SAP radius) identifies the sites to be mutated for protein stabilization.

SAP from Fragment Simulations. The SAP values discussed above are calculated from the full antibody simulation, which is extremely computationally demanding. The SAP value can also be estimated with less computational cost through the Fc and Fab fragment simulations, which are much faster than the full antibody simulations. In what follows, we test whether the fragment simulations can predict the aggregation prone regions with as much accuracy as the full antibody simulations. The

SAP values from the full antibody and fragment simulations are compared by mapping them onto the full antibody structure in Figure 8. Note that in the case of fragment SAP mapping, while the mapping is done on the full antibody for easier comparison, the simulation data were obtained from the Fab and Fc fragment simulations separately. We notice that the SAP mapping looks essentially identical between the two, with some minor differences as marked in Figure 8. The differences are due to the difference in dynamic fluctuations between fragment and full antibody simulations. Fragment simulations lack the connection between Fab and Fc fragments and thus undergo different dynamic fluctuations and conformational changes compared to the full antibody simulations. These differences in fluctuations give rise to variations in the surface exposure of certain residues, leading to the differences in SAP values observed in Figure 8. On the other hand, we notice that the differences are minor, and the aggregation prone regions validated earlier through mutations (in Figure 3) are still predicted accurately from fragment simulations. Therefore, fragment simulations can be used to identify the aggregation prone regions with good accuracy but with less precision than that of the full antibody simulations.

SAP from Implicit Solvent Simulations. Until now, we have discussed the explicit solvent simulations where the water molecules are explicitly accounted for. Unlike these explicit solvent simulations, implicit solvent simulations treat the water as a continuum solvent and thus are faster. We tested three implicit solvent models, GBSW, ACE, and EEF1, as implemented in the CHARMM simulation package. Of the three models, GBSW is the most accurate, followed by ACE and EEF1. The computational demand is in the same order. The time taken for 1 ns simulation of the Fab fragment using these implicit models is compared with that of explicit models in Table 2.

We observe from Table 2 that the explicit model using the NAMD simulation package is up to 3 times faster than that using the CHARMM simulation package. The implicit solvent models are in general faster than the explicit solvent models. However, the GBSW implicit model is actually a bit slower than the explicit simulations using NAMD. ACE is faster but only twice as much. The fastest model is the EEF1 implicit solvent model, which is up to 20 times faster than the explicit model. We therefore test the EEF1 model to see if it predicts the same aggregation prone regions as that of the explicit solvent model. The SAP values estimated for the EEF1 implicit model and the explicit model are compared by mapping them onto the Fab fragment structure in Figure 9. We observe many differences between the EEF1 and the explicit model as marked in Figure 9(a–h). Therefore, if mutations were designed for enhancing antibody stability using the EEF1 model based in Figure 9, we would have missed successful mutations in some regions (a, b, and d) and would have performed wrong mutations in other regions (c, e, f, g, and h). The reason for the difference in SAP between the EEF1 implicit model and the explicit model is due to the difference in dynamic exposure of residues in each case. The dynamic exposure is dependent on the dynamic fluctuations and conformational changes that occur during the EEF1 implicit simulations. We observed that the protein undergoes significant conformational changes using the EEF1 model, different from that of the explicit model. Whereas the root-mean-square deviation (rmsd) in 30 ns is 4.5 \AA with the explicit model, it is 7 \AA with the EEF1 model. These higher fluctuations and conformation changes in the EEF1 model lead to the differences in the SAP values observed in Figure 9.

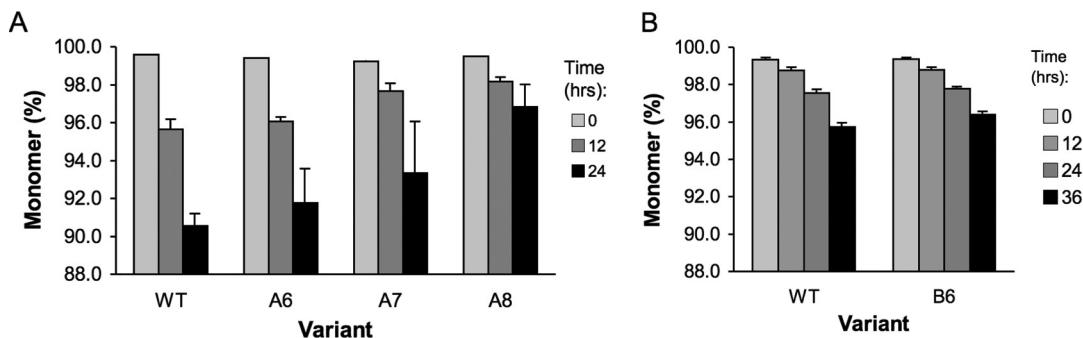


Figure 5. Stability comparison of antibody-A and antibody-B wild type and variants. (A) Monomer loss determined from SEC-HPLC upon heat stress at 58 °C for antibody-A wild type and variants (A₆-L235S, A₇-V282K, and A₈-(L235K V282K L309K)). (B) Monomer loss upon heat stress at 52 °C for antibody-B wild type and variant (B₆-L235E). Data are mean \pm SD ($n = 3$ experiments).

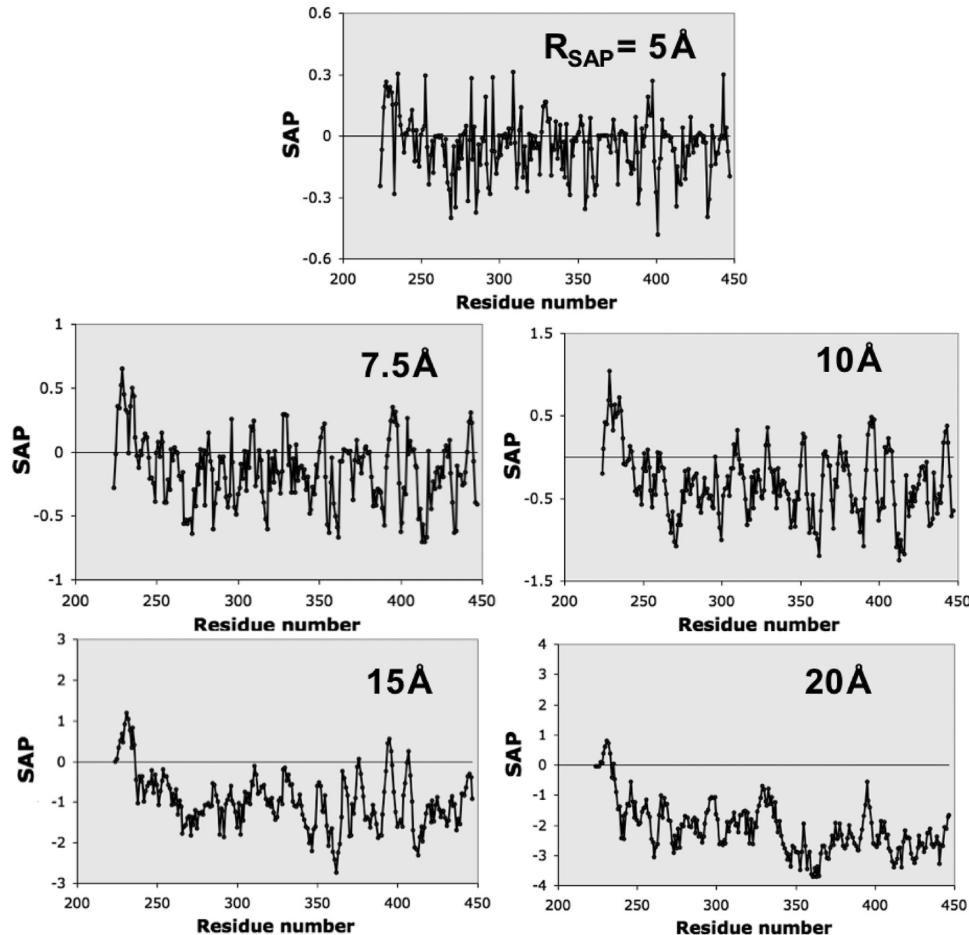


Figure 6. Spatial-aggregation-propensity (SAP) values at different radii ($R = 5, 7.5, 10, 15$, and 20 \AA) for the Fc fragment of antibody-A.

SAP from the Static Structure. Simulations in general are extremely computer intensive for even a single antibody. This makes it prohibitive for high-throughput applications such as for antibodies in the drug discovery phase. Here, rapid screening of many antibody molecules is desired to select the one with the least propensity for aggregation. Therefore, it is highly desirable in such situations to have a tool that predicts the aggregation prone regions without employing the full-scale molecular simulations. One solution is to apply the SAP tool on the static structure directly to see if it matches the value from molecular simulation. The static structure could be either from X-ray or from homology modeling. In our tests, the SAP calculation from the static structure was 200 000 times faster than from full antibody explicit atom simulations. To check the feasibility of this approach, we compare the values of SAP (5

\AA) from simulation for antibody-A with the value from the X-ray structure in Figure 10. In this figure, the SAP values obtained from simulation and from the X-ray structure are mapped onto the antibody structure. We observe that while the predicted aggregation prone regions (colored red) look very similar between the simulation and X-ray values there are certain differences as indicated in the figure. These differences are more prominent in the loop regions, especially in the CDR loops. This is due to the high dynamic fluctuations that are expected in the flexible loop regions during simulations that expose buried hydrophobic residues. In spite of these differences, we note that the aggregation prone regions validated earlier (mutational sites engineered in Figure 3) are still identified through SAP obtained from the X-ray structure. Thus, SAP from the X-ray did not miss major aggregation prone regions. However, care should

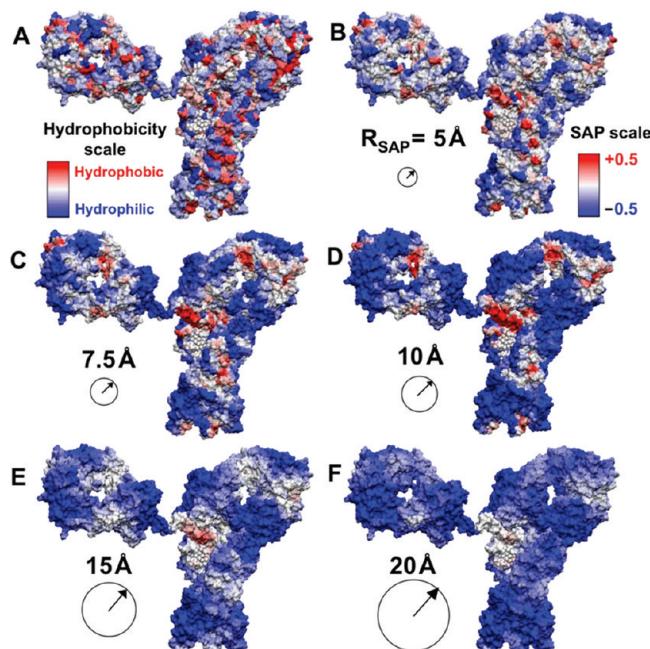


Figure 7. Spatial-aggregation-propensity (SAP) values calculated from the 30 ns simulation at different radii ($R = 5, 7.5, 10, 15$, and 20 \AA) mapped onto the structure of antibody-A. Also shown are the sizes of patches used in the calculation of SAP. For comparison, part A shows the mapping of the hydrophobicity scale (46) directly onto the antibody. Thus, part A shows the surface exposed hydrophobic residues.

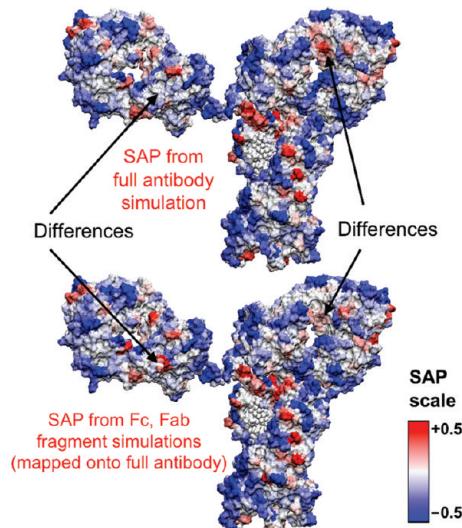


Figure 8. Comparison of SAP from the full antibody simulation and from the Fc and Fab fragment simulations. The SAP in both cases is mapped onto the full antibody-A structure, and the differences are marked.

be taken in interpreting SAP from the X-ray near the loop regions where it could miss the aggregation prone regions or could wrongly identify stable regions as aggregation prone. Therefore, in the case of high-throughput applications where fast screening is required, SAP from the static structure could be used to get a first pass estimate of the aggregation prone regions, keeping in mind its limitations. A more accurate prediction will however require a full-scale atomistic simulation of the protein.

Conclusions

Protein aggregation is a major problem encountered in the manufacture and storage of therapeutic proteins. We have

TABLE 2: Comparison of Time Taken with Different Simulation Models

simulation model employed	model package or type	time taken (for 1 ns simulation of Fab fragment on a single computer)
explicit solvent	CHARMM simulation package	80 days
explicit solvent	NAMD simulation package	27 days
implicit solvent	GBSW	32 days
implicit solvent	ACE	14 days
implicit solvent	EEF1	1.2 days

developed a tool based on molecular simulations, called spatial-aggregation-propensity (SAP), to predict protein aggregation prone regions. The predicted regions are confirmed to be responsible for aggregation by performing target mutations on two different therapeutic antibodies, leading to variants that are less aggregation prone and more stable. Thus, the SAP tool can be used to design stable therapeutic proteins as illustrated in Figure 11. We also showed that mutating residues in the aggregating regions to lysine or glutamic acid is more effective than mutating to serine. Furthermore, we prepared a triple mutant to show that simultaneous mutations on different SAP peaks can have a cumulative effect on enhancing protein stability, more than any of the separate single mutations. This shows that multiple regions contribute simultaneously to protein aggregation. We also studied the aggregation prone regions at different resolutions by varying the radius of SAP calculation, R . Whereas the SAP at $R = 10 \text{ \AA}$ is suitable for finding the larger aggregation prone patches, the SAP at $R = 5 \text{ \AA}$ is suitable for a more detailed view of these patches.

We have shown that antibody fragment simulations predict the aggregation prone regions with good accuracy while taking a fraction of the computational time as that of the full antibody simulations. We have also compared the various implicit solvent models (GBSW, ACE, EEF1) with that of the explicit solvent

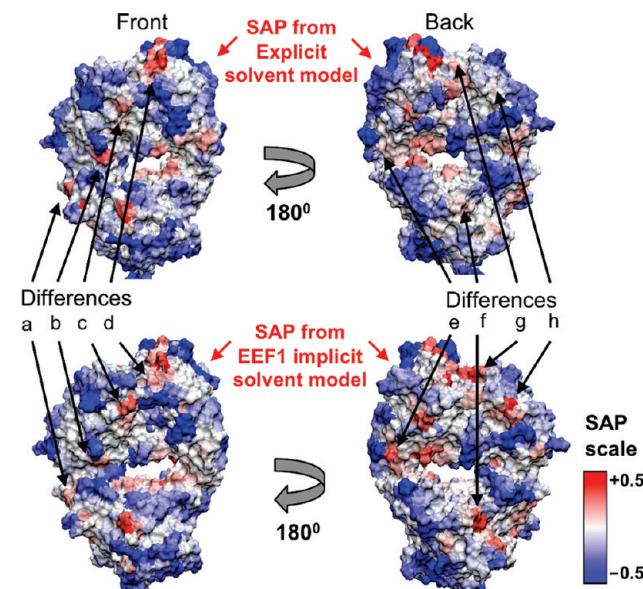


Figure 9. Comparison of spatial-aggregation-propensity (SAP) for the Fab fragment from the explicit solvent model and the EEF1 implicit solvent model. The SAP in both cases is mapped onto the Fab fragment structure, and the differences are marked.

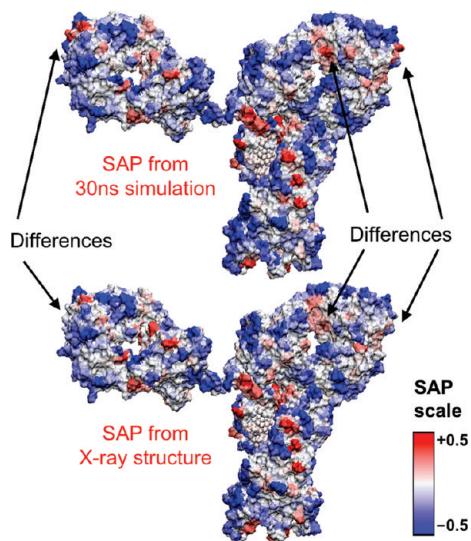


Figure 10. Comparison of SAP from 30 ns simulation average and the X-ray structure. The SAP in both cases is mapped onto the antibody-A structure, and the differences are also marked.

computationally demanding, the fragment simulations can be used to predict the aggregation prone regions with good accuracy. The next in accuracy is the ACE implicit solvent model, which is twice as fast as the explicit model. We also showed that an approximate estimate of SAP could be obtained from the static structure itself such as the X-ray structure, which enables it to predict protein aggregation prone regions in high-throughput applications. Even if the X-ray structure is not available, a structure generated through homology modeling from the protein primary sequence can be used to calculate SAP. In case of antibody-B whose X-ray structure was not available, we have shown that SAP based on such homology modeling predicts aggregation prone regions with good accuracy.

The SAP simulation tool developed could be used to improve the stability of potentially all therapeutic antibodies against aggregation. Apart from antibodies, the SAP tool could be employed to identify aggregation prone regions on other proteins and peptides as well. With the mounting number of protein therapeutics, this technology could greatly improve the developability screening of candidate biopharmaceuticals or further stabilize the selected candidates.

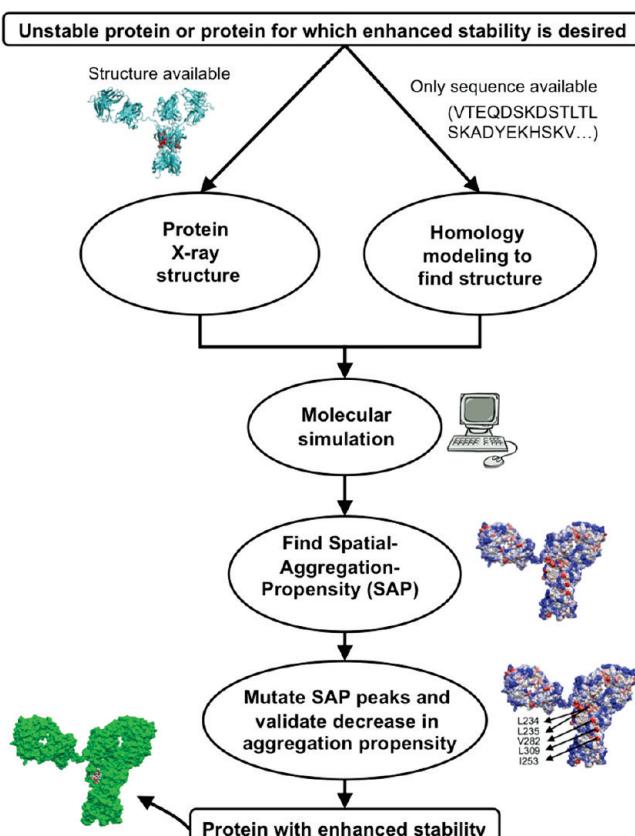


Figure 11. Use of spatial-aggregation-propensity (SAP) technology to design antibodies with enhanced stability.

model. While the GBSW model has the same computational demand as that of the explicit model, the ACE model is twice as fast, and the EEF1 implicit model is 20 times faster. Therefore, there is not much computational advantage of using either the GBSW model or the ACE model. Furthermore, they are less accurate than the explicit model. While the EEF1 model is quite fast, it gives some incorrect predictions of aggregation prone regions compared to the full antibody simulations. Therefore, it is recommended to use the explicit solvent simulations if feasible. If the full antibody simulations are too

Acknowledgment. The authors thank Dr. Holger Heine from Novartis for his expert advice and help with setting up the antibody expression and purification systems; Dr. Burkhard Wilms from Novartis for providing the DNA vectors for antibody-A and antibody-B; Dr. Jean-Michel Rondeau from Novartis for providing X-ray structures of the antibody Fab and Fc fragments; and Prof. Dane Wittrup from MIT for helpful discussions and for use of lab equipment. This work was supported by Novartis Pharma AG and by the National Center for Supercomputing Applications (NCSA) under National Science Foundation grant MCB060103 and used the NCSA tungsten cluster.

Supporting Information Available: SEC-HPLC chromatograms showing the separation of antibody monomers and aggregates. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Carter, P. J. *Nat. Rev. Immunol.* **2006**, *6*, 343–357.
- (2) Aggarwal, S. *Nat. Biotechnol.* **2007**, *25*, 1097–1104.
- (3) Shire, S. J.; Shahrokh, Z.; Liu, J. *J. Pharm. Sci.* **2004**, *93*, 1390–1402.
- (4) Rosenberg, A. S. *AAPS J.* **2006**, *8*, 501–507.
- (5) Kessler, M.; Goldsmith, D.; Schellekens, H. *Nephrol. Dial. Transplant* **2006**, *21* (suppl 5), v9–v12.
- (6) Maas, C.; Hermeling, S.; Bouma, B.; Jiskoot, W.; Gebbink, M. F. B. *G. J. Biol. Chem.* **2007**, *282*, 2229–2236.
- (7) Hermeling, S.; Crommelin, D. J. A.; Schellekens, H.; Jiskoot, W. *Pharm. Res.* **2004**, *21*, 897–903.
- (8) Braun, A.; Kwee, L.; Labow, M. A.; Alsenz, J. *Pharm. Res.* **1997**, *14*, 1472–1478.
- (9) Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F. *Pharm. Res.* **2003**, *20*, 1325–1336.
- (10) Roberts, C. J. *J. Phys. Chem. B* **2003**, *107*, 1194–1207.
- (11) Li, Y.; Roberts, C. J. *J. Phys. Chem. B* **2009**, *113*, 7020–7032.
- (12) Dill, K. A. *Biochemistry* **1990**, *31*, 7134–7155.
- (13) Ashbaugh, H. S.; Hatch, H. W. *J. Am. Chem. Soc.* **2008**, *130*, 9536–9542.
- (14) Guharoy, M.; Chakrabarti, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *43*, 15447–15452.
- (15) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13–20.
- (16) Young, L.; Jernigan, R. L.; Covell, D. G. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.* **1994**, *3*, 717–729.
- (17) Tsai, C.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *Protein Sci.* **1997**, *6*, 53–64.

- (18) Conte, L. L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177–2198.
- (19) DeLano, W. L.; Ultsch, M. H.; de Vos, A. M.; Wells, J. A. *Science* **2000**, *287*, 1279–1283.
- (20) Caflisch, A. *Curr. Opin. Chem. Biol.* **2006**, *10*, 437–444.
- (21) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 336–366.
- (22) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. *Nature* **2003**, *424*, 805–808.
- (23) Tartaglia, G. G.; Cavalli, A.; Pellarin, R.; Caflisch, A. *Protein Sci.* **2005**, *14*, 2723–2734.
- (24) Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. *Nat. Biotechnol.* **2004**, *22*, 1302–1306.
- (25) Trovato, A.; Chiti, F.; Maritan, A.; Seno, F. *PLoS Comp. Biol.* **2006**, *2*, 1608–1618.
- (26) Wang, X.; Das, T. K.; Singh, S. K.; Kumar, S. *mAbs* **2009**, *1*, 1–14.
- (27) Cellmer, T.; Bratko, D.; Prausnitz, J. M.; Blanch, H. W. *Trends Biotechnol.* **2007**, *25*, 254–261.
- (28) Harrison, P. M.; Chan, H. S.; Prusiner, S. B.; Cohen, F. E. *J. J. Mol. Biol.* **1999**, *286*, 593–606.
- (29) Tamamis, P.; Kasotakis, E.; Mitraki, A.; Archontis, G. *J. Phys. Chem. B* **2009**, *113*, 15639–15647.
- (30) Jang, S.; Yuan, J.; Shin, J.; Measey, T. J.; Schweitzer-Stenner, R.; Li, F. *J. Phys. Chem. B* **2009**, *113*, 6054–6061.
- (31) Dima, R. I.; Thirumalai, D. *Protein Sci.* **2002**, *11*, 1036–1049.
- (32) Broglia, R. A.; Tiana, G.; Pasquali, S.; Roman, H. E.; Vigezzi, E. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12930–12933.
- (33) Nguyen, H. D.; Hall, C. K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16180–16185.
- (34) Gsponer, J.; Haberthur, U.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5154–5159.
- (35) Klimov, D. K.; Thirumalai, D. *Structure* **2003**, *11*, 295–307.
- (36) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11937–11942.
- (37) Chennamsetty, N.; Helk, B.; Voynov, V.; Kayser, V.; Trout, B. L. *J. Mol. Biol.* **2009**, *391*, 404–413.
- (38) Voynov, V.; Chennamsetty, N.; Kayser, V.; Helk, B.; Trout, B. L. *mAbs* **2009**, *1*, 580–582.
- (39) Saphire, E. O.; Parren, P. W.; Pantophlet, R.; Zwick, M. B.; Morris, G. M.; et al. *Science* **2001**, *293*, 1155–1159.
- (40) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; et al. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; et al. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (42) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (44) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.
- (45) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 901–917.
- (46) Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smith-Gill, S. J.; et al. *Nature* **1989**, *342*, 877–883.
- (47) Al-Lazikani, B.; Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1997**, *273*, 927–948.
- (48) Whitelegg, N. R.; Rees, A. R. *Protein Eng.* **2000**, *13*, 819–824.
- (49) Black, S. D.; Mould, D. R. *Anal. Biochem.* **1991**, *193*, 72–82.

JP911706Q