PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Impact of deglycosylation and thermal stress on conformational stability of a full length murine IgG2a monoclonal antibody: Observations from molecular dynamics simulations

Xiaoling Wang, Sandeep Kumar,* Patrick M. Buck, and Satish K. Singh

Pharmaceutical Research and Development, BioTherapeutics Pharmaceutical Sciences, Pfizer Inc, 700 Chesterfield Parkway West, Chesterfield, Missouri 63017

ABSTRACT

With the rise of antibody based therapeutics as successful medicines, there is an emerging need to understand the fundamental antibody conformational dynamics and its implications towards stability of these medicines. Both deglycosylation and thermal stress have been shown to cause conformational destabilization and aggregation in monoclonal antibodies. Here, we study instabilities caused by deglycosylation and by elevated temperature (400 K) by performing molecular dynamic simulations on a full length murine IgG2a mAb whose crystal structure is available in the Protein Data bank. C^{α}-atom root mean square deviation and backbone root mean square fluctuation calculations show that deglycosylation perturbs quaternary and tertiary structures in the C_H2 domains. In contrast, thermal stress pervades throughout the antibody structure and both Fabs and Fc regions are destabilized. The thermal stress applied in this study was not sufficient to cause large scale unfolding within the simulation time and most amino acid residues showed similar average solvent accessible surface area and secondary structural conformations in all trajectories. C_H3 domains were the most successful at resisting the conformational destabilization. The simulations helped identify aggregation prone regions, which may initiate cross-β motif formation upon deglycosylation and upon applying thermal stress. Deglycosylation leads to increased backbone fluctuations and solvent exposure of a highly conserved APR located in the edge β-strand A of the C_H2 domains. Aggregation upon thermal stress is most likely initiated by two APRs that overlap with the complementarity determining regions. This study has important implications for rational design of antibody based therapeutics that are resistant towards aggregation.

Proteins 2013; 81:443–460.

Key words: monoclonal antibody; biotherapeutics; molecular dynamics; simulation; structure.

INTRODUCTION

Recent years have witnessed an emergence of monoclonal antibodies (mAbs) as the most important class of biopharmaceuticals.[1,2] These molecules offer several advantages over small molecule pharmaceuticals, such as high specificity and affinity toward the target, and minimal nonmechanism toxicity. At an average molecular weight of ~150 kDa, full length mAbs are considerably bigger than small molecules (~0.4–1 kDa) and possess highly complex heterogeneous three-dimensional structures that are vulnerable to various physico-chemical liabilities such as deamidation, oxidation, hydrolysis, and aggregation.[3–5] Thus, commercialization of a therapeutic mAb drug product requires development of appropriate chemistry, manufacturing and control (CMC) strategies to assure molecular stability during manufacturing as well as during shipping and storage over shelf-life of the drug product.[4] Understanding of structural and dynamic behavior of mAbs is essential for reliable prediction of potential physico-chemical degradation vulnerabilities

Additional Supporting Information may be found in the online version of this article.

Abbreviations: APR, aggregation prone region; BB-RMSF, backbone root mean square fluctuation; C^{α}-RMSD, C^{α} atom root mean square deviation; CDR, complementarity determining region; CMC, chemistry, manufacturing and control; Fab, fragment antigen binding; Fc, fragment crystallizable; HC, heavy chain; LC, light chain; mAb, monoclonal antibody; MD, molecular dynamics; PDB, protein data bank; SASA, solvent accessible surface area.

*Correspondence to: Sandeep Kumar, BioTherapeutics Pharmaceutical Sciences Pfizer Inc, C2482, 500 Technology Drive #1, Weldon Spring, MO 63044, USA.
E-mail: sandeep.kumar@pfizer.com
Received 10 August 2012; Revised 2 October 2012; Accepted 4 October 2012
Published online 8 October 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24202

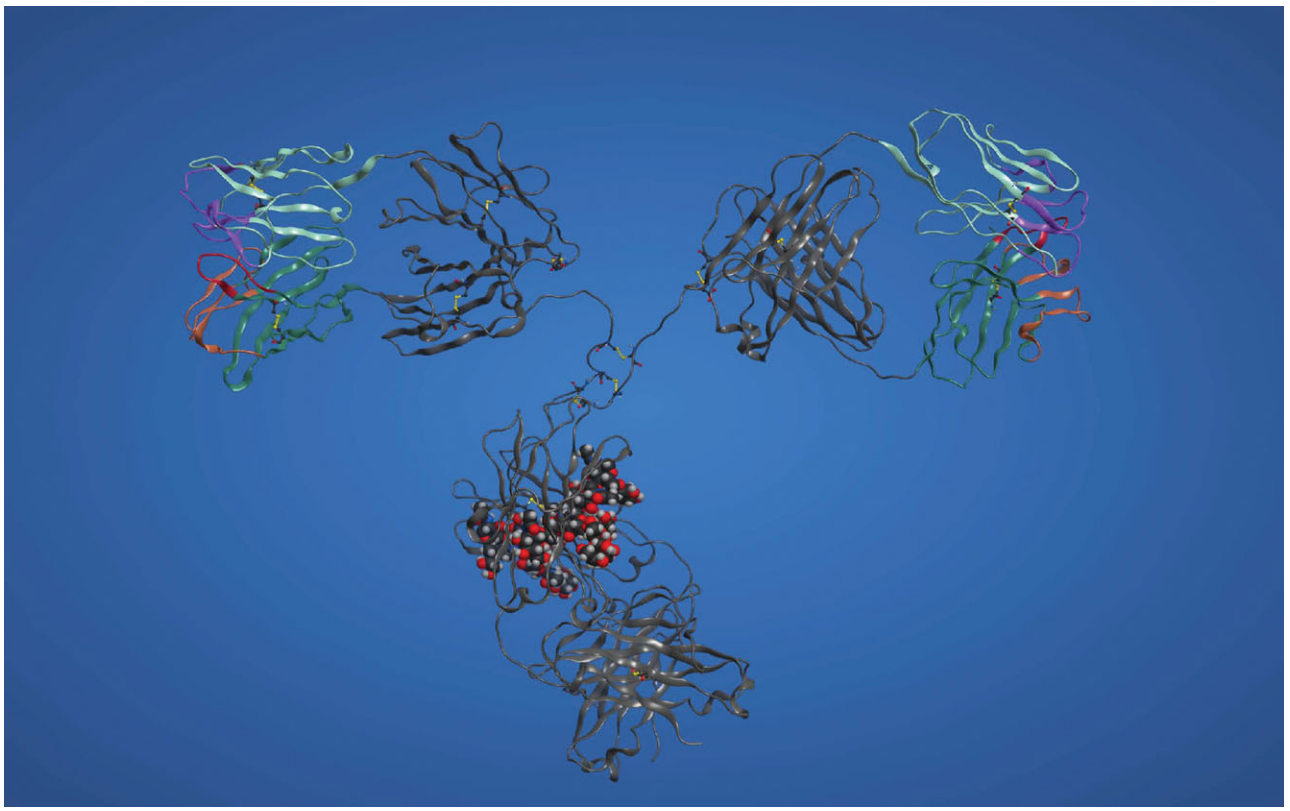


Figure 1

Ribbon representation of crystal structure of the full length murine mAb (PDB entry, 1IGT) whose MD simulations were performed in this study. The glycan atoms are shown in CPK representation. The constant regions of the mAb (C_L, C_H1 Hinge, C_H2, and C_H3 domains) are shown in gray color ribbons. The variable region is shown in green (V_H) and cyan (V_L) ribbons. The CDRs in the light chain are shown in magenta ribbons. The CDRs 1 and 2 in the heavy chain are shown in the brown ribbons and the CDR3 is shown in the red colored ribbon.

associated with them at early stages of drug discovery and formulation. For example, exposure to elevated temperatures during shipping and storage of these products can destabilize their conformation and lead to aggregation. This may impact potency and safety of these products.[3,4,6,7]

Asn-linked glycans contribute towards stability of the antibodies, play important roles in Fc-mediated cell-killing effector functions and their heterogeneity has been correlated with human diseases.[8–13] Deglycosylation has been shown to cause conformational destabilization and lead to increased aggregation in human/humanized IgG1 mAbs[9,11] and in human IgG1 Fc.[10] Deglycosylation is not a major degradation route for therapeutic mAbs. However, insights gained by studying the impact of deglycosylation on antibody conformation can be used to design novel aglycosylated mAbs and Fc/C_H2 domain containing novel modalities for applications where the cell killing aspects of Fc effector functions may not be required.[14]

Here, we asked the following questions: Do the conformational destabilizations in a full length mAb caused by thermal stress and deglycosylation differ from each other? If yes, what are the differences and how they may impact potential aggregation liabilities associated with the mAb? To understand these questions, we have performed all atom explicit water molecular dynamics (MD) simulations on a full length murine IgG2 mAb (Mab231)[15] (Fig. 1) whose crystal structure is available in the Protein Data Bank (PDB).[16] The simulations were performed in presence and absence of glycans at 298 K. The glycans studied here are the same as the ones reported in the crystal structure. The deglycosylated murine mAb was also simulated at 400 K. The elevated temperature simulations performed here were meant to apply a thermal stress to perturb but not unfold the mAb structure. This situation is analogous to brief periods of temperature incursions or brief interruptions in cold supply chains potentially experienced by biotherapeutics during storage and/or shipping, even though the time scales explored by these simulations are much smaller.

Our study indicates that conformational destabilizations of murine IgG2a mAb caused by deglycosylation and by elevated temperature are different. The Asn-linked glycans in this murine mAb are located at the C_H2

domain dimer interface[8] and deglycosylation causes structural perturbations in this region. The two C_H2 domains move with respect to each other during the simulations, thereby, perturbing the quaternary structure of the deglycosylated mAb. This is accompanied by tertiary structure perturbations as the conformations of the individual C_H2 domains are also destabilized because deglycosylation solvent exposes several aromatic and hydrophobic residues that lie in structural vicinity to the glycans.[8] The residues in the individual C_H2 domains show greater backbone root mean square fluctuations (BB-RMSF) values in 298 K simulation trajectories of deglycosylated mAb along with increased average solvent accessible surface area (SASA) values for the glycan vicinal hydrophobic and aromatic residues. On the other hand, the conformational destabilization caused by elevated temperature pervades all over the mAb's quaternary and tertiary structures and causes greater conformational change as suggested by greater C^{α}-atom root mean square deviation (C^{α}-RMSD) and BB-RMSF values for the 400 K simulation trajectories. The backbone secondary structural features did not show significant changes within the simulation times for all trajectories, consistent with the purpose of this study.

Both deglycosylated and thermally stressed mAbs have been shown to aggregate.[4,9,11,17] The MD simulations helped identify aggregation prone regions (APRs) that can initiate formation of cross-β motif and promote aggregation upon deglycosylation and upon applying thermal stress. The APRs located in and around complementarity determining regions (CDRs) may actively promote aggregation upon thermal stress. Two such APRs were identified by the simulations for this mAb: 29-INVWLSWY-36, overlapping with CDR1 in the light chains, and 32-YYMYWV-37, overlapping with CDR1 in the heavy chains. Deglycosylation solvent exposes an APR (240-SVFIFP-245) located in the edge β-strand A of each C_H2 domain. All these three APRs are strongly predicted by a TANGO/PAGE combination.[18–20] This study has considerably improved our fundamental understanding of conformational dynamics of a full length mAb, and thereby, suggested rational design/selection strategies to overcome the challenges faced during development of antibody based drug candidates as successful products.

MATERIALS AND METHODS

Molecular dynamics simulation setup

The crystal structure of a full length IgG2a murine antibody (MAb231; PDB entry, 1IGT[15]) was used as the starting structure for MD simulation. All the initial structure preparation was performed using the AMBER Tleap program. This included generation of coordinates for missing atoms. The N- and C-termini of all chains in 1IGT were acetylated and amidated to neutralize terminal charges. 1IGT was then solvated with TIP3P model water in a cubic box.[21] The size of water box was chosen such that the minimum distance between solute and its periodic image is greater than the cutoff for potential energy calculation throughout the simulation. Cl^- ions were added to neutralize the total charge on the system. All dynamic simulations were performed using NAMD simulation package[22] and AMBER99 all-atom force field.[23] The 1IGT crystal structure contains glycans in its Fc portion. The glycans were either removed before solvation or kept throughout the simulation. In the simulations containing the glycans, the Glycam_06 carbohydrate force field was used for the glycan atoms.[24]

The simulation of the solvated 1IGT involved the following steps. First, the system was subjected to energy minimization with all solute heavy atoms fixed until the final root mean square gradient (RMSG) fell below 1. This is to optimize water molecules, especially those surrounding the solute. Second, a 5000 step energy minimization without any restraints was performed to optimize the whole system. Thirdly, an equilibration run was then performed for 1 ns. The temperature, pressure, and energy profiles of the system were equilibrated to the target values. The last step was production run with varying durations. The isothermal-isobaric (NPT) ensemble at $P = 1$ bar was used in all equilibration and production runs. The temperature was set at either $T = 298$ K or $T = 400$ K. The initial velocities for the equilibration run were generated from Maxwell–Boltzmann distribution. All bonds involving hydrogen atoms were restrained and a 2.0 fs integration time step was used in all simulations. Periodic boundary conditions were employed. A cutoff of 10 Å was applied to nonbonded interactions with switch starting at 8.5 Å. Nonbonded atom pairs were searched within 11.5 Å and were updated every 10 time steps. Particle mesh ewald summation was performed at every other steps for long range electrostatic interactions.

Collection of simulation trajectories

MD simulations were performed on the murine mAb and its components in presence/absence of the glycans at 298 K and at elevated temperature (400 K). In each case, we obtained two independent MD simulation trajectories. A short name was given to each set for convenience of presentation. Table I lists the 9 sets of simulations, including the short name, description of the simulated system, simulation temperature and production run time of each trajectory. An identifier "1" or "2" is used after the simulation short name to refer to the two independent simulations in the set.

In all simulations, the atomic coordinates were saved every 10 ps during the production run. The length of simulation refers to the duration of production run. The first 5 ns of production run were discarded in all time-averaged analyses. However, the data for the whole production run was included in time-series plots. The simulations on full

Table I
List of Simulations Performed in this Study

Number	Short name	System description	Temperature (K)	Simulation time (ns)	Number of trajectories
1	Full	Full length murine mAb (1IGT) without glycans	298	40	2
2	FullGly	Full length murine mAb (1IGT) with glycans	298	40	2
3	Full400K	Full length murine mAb (1IGT) without glycans	400	40	2
4	FabAOnly	FabA only	298	20	2
5	FabBOnly	FabB only	298	20	2
6	FcOnly	Fc only without glycans	298	20	2
7	FcGlyOnly	Fc only with glycans	298	20	2
8	CH2Only	C_H2 domain dimer	298	20	2
9	CH3Only	C_H3 domain dimer	298	20	2

length 1IGT are of comparable duration as the previously reported simulations on IgG molecules.[25–27] On a 32 CPU cluster, 1.5 computer days were required to simulate 1 ns of full 1IGT system dynamics. The total simulation time for all the trajectories obtained in this study is 480 ns.

Definition of CDRs and residue numbering

Andrew Martin's definition of CDRs in antibody is followed.[28] For convenience, the continuous residue numbering is used in this study. For clarity, the two light chains named as chain A and chain C in 1IGT PDB are renamed as light chain 1 (LC1) and light chain 2 (LC2), respectively, throughout this study. Similarly, the two heavy chains, chain B and chain D, are renamed as heavy chain 1 (HC1) and heavy chain 2 (HC2), respectively. The Fab containing LC1, V_H, and C_H1 domains of HC1 is named as FabA and the other Fab is named as FabB.

Calculation of residue-wise BB-RMSF

Residue-wise backbone root mean square fluctuations (BB-RMSF) were computed taking the average of root mean fluctuations (RMSF) in positions of four backbone atoms, namely, N, C^{α}, C′, and O. The RMSF calculations were performed in VMD.[29]

Prediction of aggregation prone regions

Potential APRs in the murine mAb light and heavy chain sequences were identified using a combination of TANGO[18] and PAGE[19] as described earlier.[20] The full aggregation spectra for the light and heavy chains of this mAb are presented in the Supporting Information [Fig. S1(a,b)].

Calculation of percent solvent exposed surface area (SASA)

The percent SASA value for each residue, *X*, is its solvent accessibility as defined below.

$$SASA_X\ (\%) = SASA_X\ (Å^2)/SASA_{ALA\text{-}X\text{-}ALA}\ (Å^2)$$

where, $SASA_X$ is the SASA ($Å^2$) of the residue X in the mAb's three-dimensional structure. $SASA_{ALA\text{-}X\text{-}ALA}$ is the SASA ($Å^2$) of the residue *X* in a tripeptide Ala-X-Ala in fully extended conformation. The solvent accessibility was calculated using the algorithm of Lee and Richards.[30] The solvent probe radius is 1.4 Å.

RESULTS AND DISCUSSION

The goal of formulation development of antibody based therapeutics is to assure pharmaceutical stability of the drug substance during shipping and storage of the drug products within their shelf life times. An optimal formulation protects the drug substance against all potential physico-chemical degradations, which may be initiated during normal storage and/or in response to stress conditions such as temperature and pH extremes, agitation, etc. To achieve this goal, it is essential to gain an understanding of structural and dynamical aspects of the drug substance at the beginning of formulation development process. In particular, gathering information on all potential degradation routes associated with a candidate biotherapeutic molecule can help formulators devise appropriate cost effective strategies for successful development of the biotherapeutic drug products. Recurrent manufacturability issues encountered at developmental stages can also be mitigated via rational design/selection at the early lead discovery stages.

In this study, we explore the conformational destabilization of a full length IgG2a murine mAb, MAb231 (PDB entry, 1IGT) due to thermal stress and deglycosylation via MD simulations. MAb231 was used as a model system in this study because almost all heavy atom coordinates including those for the glycans are available to a reasonably high resolution (2.8 Å) via crystallography.[15] There are only three crystal structures for full length mAbs in the PDB (the entries, 1HZH, 1IGT, and 1IGY) and homology based full length models for therapeutic mAb candidates are made using one of these structures as the overall templates. To the best of our knowledge, this is a first report on MD simulations of a full length murine mAb.

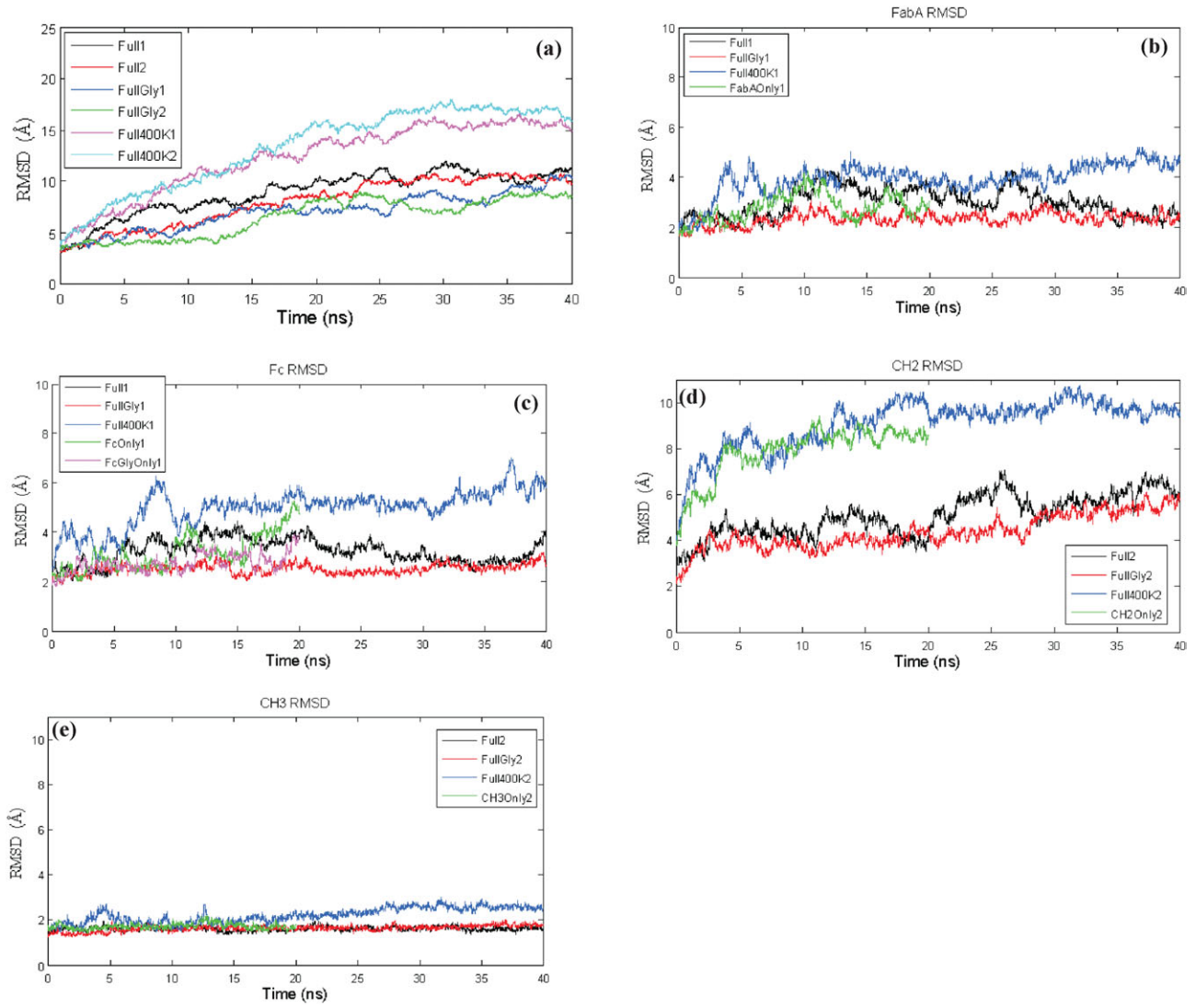


Figure 2

C^{α} root mean square deviation (C^{α}-RMSD) plotted with respect to time in different simulation trajectories for (**a**) full length mAb, (**b**) Fab region, (**c**) Fc region, (**d**) C_H2 domain dimer, and (**e**) C_H3 domain dimer.

Impact of thermal stress and deglycosylation on overall conformational stability of the murine mAb

Figure 2(a–e) show plots of C^{α}-RMSD with respect to time in different simulation trajectories for the full length murine mAb and its various regions. Figures 3 and 4 present BB-RMSF and average SASA values for each residue in the first set of full length mAb simulation trajectories, respectively. As stated earlier, the names for various simulation trajectories are given in Table I. Briefly, the names Full1 and Full400K1 denote the first simulation trajectories of deglycosylated mAb at 298 K and at 400 K, respectively; and the name FullGly1 indicates first simulation trajectory of glycosylated mAb at 298 K. A movie made from FullGly1 trajectory is also presented in the Supporting Information.

Residue wise BB-RMSF plots shown in Figure 3 identify the flexible regions in the mAb. As expected, the residues in N- and C-termini of the light and heavy chains, CDR loops, hinge regions and loops connecting different domains show greater BB-RMSF values in all simulation trajectories. In all the simulation trajectories, residue wise average SASA plots (Fig. 4) show negligible changes in mean solvent exposure of the residues, except for a few residues in CDR3 of heavy chain and the residues vicinal to glycans in the C_H2 domain. However, the average C^{α}-RMSD values for Fab and Fc regions in the 400 K simulation trajectories increase by more than 30% over the corresponding regions in the 298 K simulation trajectories (Table II). These observations indicate that large scale unfolding of this mAb and its various domains did not

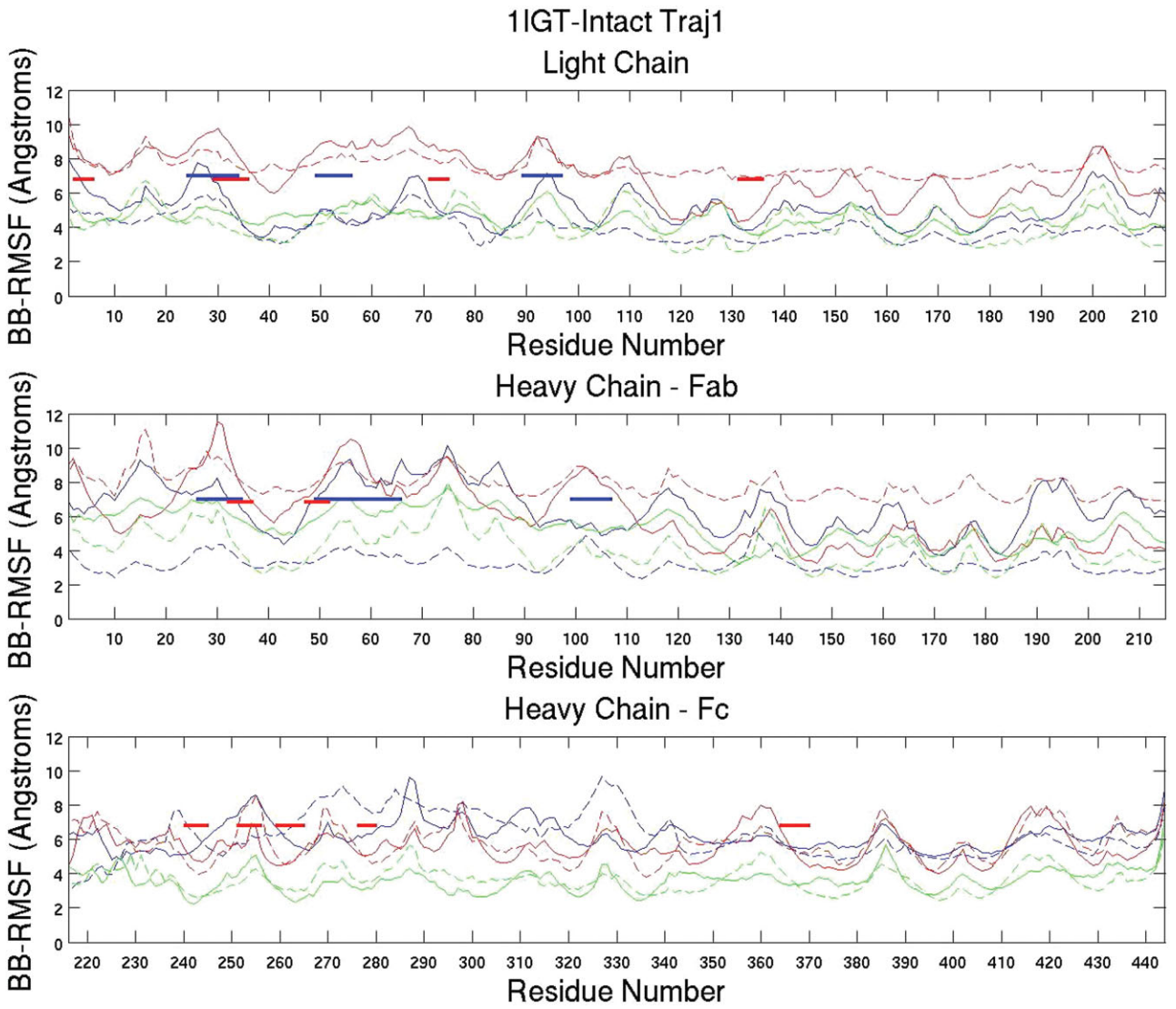


Figure 3

Backbone root mean square fluctuation (BB-RMSF) values plotted with respect to the residue number for (**a**) light chains, (**b**) Fab part of heavy chains, and (**c**) Fc part of the heavy chains for the first set of full length mAb simulation trajectories: Green, 298 K simulation on glycosylated mAb (FullGly1); blue, 298 K simulation on deglycosylated mAb (Full1); red, 400 K simulation on deglycosylated mAb (Full400K1). Solid lines show HC1 and LC1 pair and dashed lines show HC2 and LC2 pair. In each subplot, blue and red horizontal bars show positions of CDRs and APRs, respectively, along the sequences of light and heavy chains.

begin within the simulation times, but the mAb structure was perturbed. Furthermore, the overall secondary structural features of the mAb did not change significantly over the course of all simulation trajectories. As an example, secondary structural features of all the mAb residues in 800 frames selected over the duration of Full400K1 trajectory are shown in Fig. S2 in the Supporting Information. These results are in line with our purpose of destabilizing but not denaturing the mAb via thermal stress. The BB-RMSF values in the second simulation trajectories yielded similar trends, but with larger magnitudes, especially, in the Fc region (Supporting Information Fig. S3).

Thermal stress at 400 K has a greater impact on overall conformational stability of the murine mAb than deglycosylation. Figure 2(a) shows the C^{α}-RMSD plots for the simulation trajectories for full length mAb in presence/absence of glycans (at 298 K) and for deglycosylated mAb at 400 K. The 298 K simulation trajectories in presence (red color) and absence (black colors) of glycans show similar overall C^{α}-RMSD profiles, but these values are higher by ~1–2 Å for trajectories of the deglycosylated mAb. The simulation trajectories obtained at 400 K (blue color) show substantially higher C^{α}-RMSD values (close to 5 Å), indicating greater change in the mAb

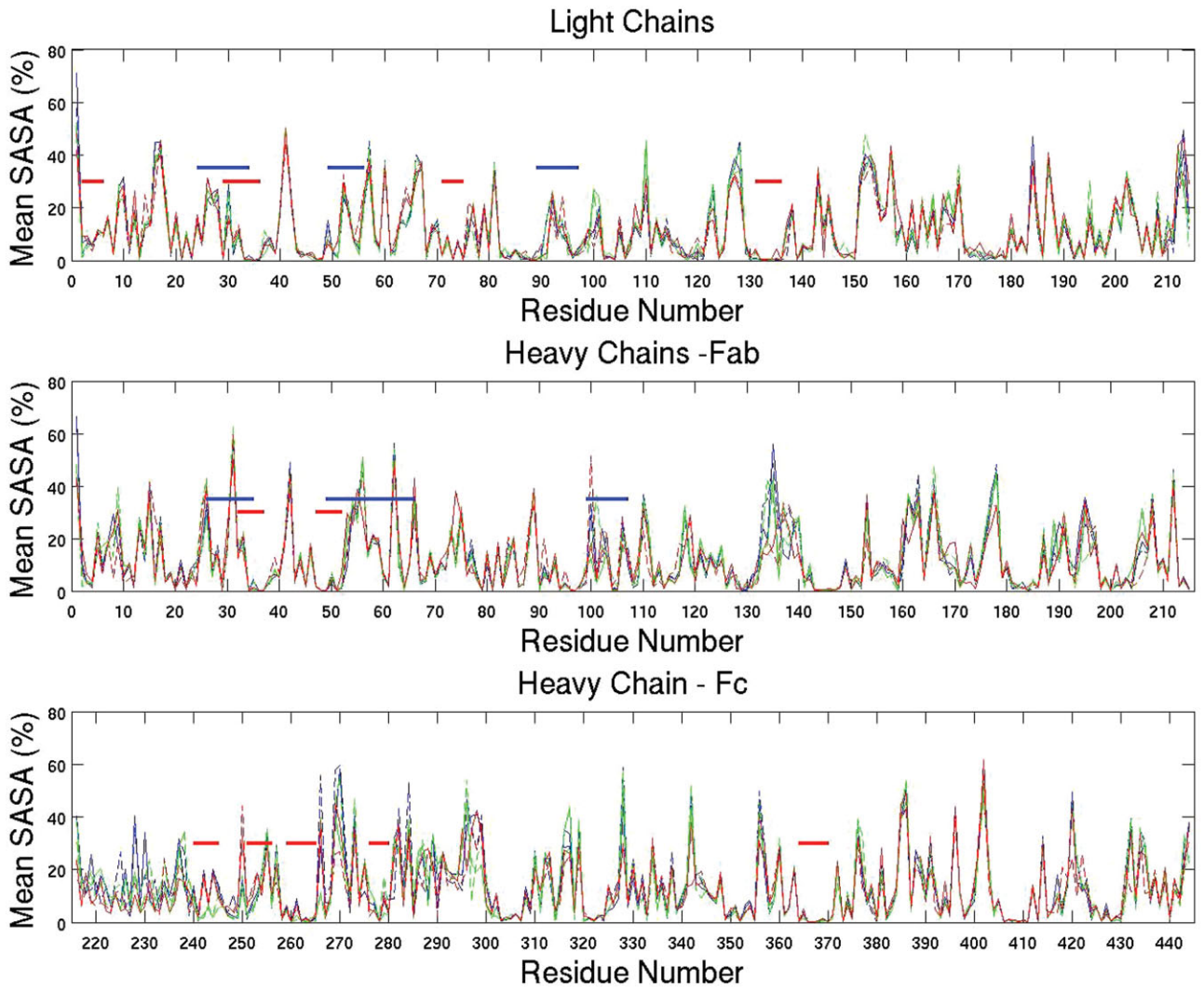


Figure 4

Average solvent accessible surface area (SASA) values plotted with respect to the residue number for (**a**) light chains, (**b**) Fab part of heavy chains, and (**c**) Fc part of the heavy chains for the first set of full length mAb simulation trajectories. For the purpose of plotting this figure, the hinge region was included in Fc part of the heavy chains. The graph plotting scheme followed here is the same as the one in Figure 3.

structure. By comparing the simulation trajectories for the deglycosylated mAb at 298 K and 400 K shown in this figure, it can be inferred that destabilization of the mAb in 400 K trajectories is mostly due to the elevated temperature. This observation is supported by Figure 3 where residue-wise BB-RMSF values are greater by 4–8 Å for the 400 K trajectory (Full400K1, red colored solid and dashed lines). At 298 K, simulation trajectories for full length mAb in presence (green colored solid and dashed lines) and absence of glycans (blue colored solid and dashed lines) show similar BB-RMSF values in all regions except for the two C_H2 domains (discussed later). The mAb structure used to obtain the 400 K simulation trajectories was deglycosylated to maximize the protein destabilization. Table III presents the residue-wise BB-RMSF values averaged over different regions of the murine mAb such as Fv, Fab, and Fc regions and their components. These average BB-RMSF values are different

Table II

Average and Standard Deviation of C^{α} RMSD Values for Different Regions in the Full Length mAb Simulations

Simulation name	FabA (Å)	FabB (Å)	Fc (Å)
Full1	3.1 ± 0.5	2.2 ± 0.3	3.3 ± 0.4
Full2	2.7 ± 0.4	2.1 ± 0.2	4.1 ± 0.6
FullGly1	2.4 ± 0.2	1.9 ± 0.2	2.5 ± 0.2
FullGly2	2.4 ± 0.3	1.9 ± 0.2	3.8 ± 0.5
Full400K1	4.1 ± 0.4	4.0 ± 0.7	5.2 ± 0.6
Full400K2	3.7 ± 0.9	4.1 ± 0.8	7.8 ± 0.8

Table III
Residue Wise BB-RMSF Values Averaged Over Different Regions of the Full Length Murine mAb

mAb region[a]	Average BB-RMSF (Å)					
	First set of simulations[b]			Second set of simulations[b]		
	Full1	FullGly1	Full400K1	Full2	FullGly2	Full400K2
V_L LC1	5.22 ± 1.09	4.77 ± 0.50	8.14 ± 0.91	3.63 ± 0.65	7.26 ± 1.16	7.91 ± 0.91
V_H HC1	7.13 ± 1.39	6.10 ± 0.65	7.39 ± 1.50	4.84 ± 0.89	5.85 ± 0.84	7.19 ± 0.80
FvA	6.22 ± 1.58	5.47 ± 0.88	7.74 ± 1.31	4.27 ± 0.99	6.52 ± 1.23	7.53 ± 0.93
C_L LC1	4.93 ± 0.82	4.36 ± 0.54	6.04 ± 1.05	3.72 ± 0.71	5.23 ± 1.09	6.52 ± 1.33
C_H1 HC1	5.82 ± 1.17	4.54 ± 0.70	4.47 ± 0.70	3.90 ± 0.43	4.54 ± 0.90	5.59 ± 0.76
FabA	5.81 ± 1.43	4.98 ± 0.93	6.58 ± 1.75	4.05 ± 0.86	5.75 ± 1.41	6.84 ± 1.28
Hinge HC1	5.57 ± 0.57	3.88 ± 0.48	6.14 ± 0.70	5.06 ± 0.84	4.03 ± 0.78	3.84 ± 0.40
C_H2 HC1	6.58 ± 0.92	3.34 ± 0.62	5.58 ± 0.68	4.13 ± 0.36	3.65 ± 0.65	5.66 ± 1.38
C_H3 HC1	5.96 ± 0.58	3.83 ± 0.62	5.60 ± 1.19	3.69 ± 0.26	5.28 ± 1.51	5.99 ± 2.24
Fc HC1	6.27 ± 0.83	3.59 ± 0.67	5.59 ± 0.97	3.91 ± 0.38	4.46 ± 1.42	5.82 ± 1.86
First half	5.94 ± 1.27	4.51 ± 1.06	6.25 ± 1.59	4.04 ± 0.77	5.29 ± 1.53	6.42 ± 1.62
V_L LC2	4.42 ± 0.73	4.47 ± 0.87	7.83 ± 0.59	5.84 ± 0.88	5.27 ± 0.74	7.50 ± 0.70
V_H HC2	3.37 ± 0.54	4.57 ± 1.10	8.26 ± 0.74	5.45 ± 1.23	4.27 ± 1.19	7.28 ± 1.21
FvB	3.87 ± 0.83	4.52 ± 1.00	8.05 ± 0.71	5.64 ± 1.10	4.75 ± 1.12	7.38 ± 1.00
C_L LC2	3.57 ± 0.37	3.91 ± 1.04	7.25 ± 0.35	6.78 ± 0.70	3.80 ± 0.75	5.82 ± 0.74
C_H1 HC2	3.12 ± 0.51	3.63 ± 0.88	7.41 ± 0.49	6.17 ± 1.37	2.65 ± 0.62	4.88 ± 0.49
FabB	3.63 ± 0.74	4.17 ± 1.05	7.71 ± 0.69	6.04 ± 1.18	4.04 ± 1.26	6.43 ± 1.35
Hinge HC2	4.98 ± 1.27	4.10 ± 0.49	6.27 ± 0.72	5.54 ± 1.46	3.83 ± 0.83	5.06 ± 0.84
C_H2 HC2	7.29 ± 1.06	3.62 ± 0.61	5.61 ± 1.10	4.15 ± 0.54	3.64 ± 0.43	6.91 ± 1.96
C_H3 HC2	5.55 ± 0.58	3.63 ± 0.82	5.87 ± 0.91	3.67 ± 0.27	4.93 ± 1.30	9.06 ± 1.65
Fc HC2	6.42 ± 1.22	3.63 ± 0.72	5.74 ± 1.01	3.91 ± 0.49	4.29 ± 1.16	7.99 ± 2.10
Second half	4.55 ± 1.59	4.00 ± 0.98	7.04 ± 1.22	5.36 ± 1.42	4.11 ± 1.22	6.87 ± 1.80
Whole mAb	5.25 ± 1.60	4.25 ± 1.05	6.65 ± 1.47	4.70 ± 1.32	4.70 ± 1.51	6.64 ± 1.72

[a]mAb region indicates the different sequence and structural regions in the murine mAb: V_L LC1, Variable domain of Light chain 1 (residues 1–107); C_L LC1, Constant domain in the light chain 1 (residues 108–214); V_H HC1, variable domain in heavy chain 1 (residues 1–118); C_H1 HC1, constant domain 1 in heavy chain 1 (residues 119–215); FvA, V_L LC1 + V_H HC1; FabA, FvA+C_L LC1+C_H1 HC1 (residues 1–215 in heavy chain 1 and 1–214 in light chain 1); Hinge HC1, residues 216–238 in heavy chain 1; Fc HC1, residues 239–444 in heavy chain 1; C_H2 HC1, C_H2 domain in heavy chain 1 (residues 239–341); C_H3 HC1, C_H3 domain in heavy chain 1 (residues 342–444). First half, First light chain and heavy chain pair in the mAb (residues 1–214 in light chain 1 and residues 1–215 in heavy chain 1). The definitions for mAb regions in the second mAb are analogous. Whole mAb, all amino acid residues in the murine mAb taken together.
[b]Simulation names are as described in Table I. Full1 and Full2 are the first and second 298 K simulation trajectories for deglycosylated mAb. FullGly1 and FullGly2 are the first and second 298 K simulation trajectories for glycosylated mAb. Full400K1 and Full400K2 are the first and second simulation trajectories for the deglycosylated mAb obtained at 400 K.

between the two sets of simulation trajectories but show similar overall trends.

Conformational stability of the Fab regions

Thermal stress, but not deglycosylation, lowers conformational stability of the Fab regions. Figure 2(b) shows C^α-RMSD profiles of a Fab (FabA) extracted from the simulation trajectories for the full length mAb. An independent 20 ns trajectory obtained for FabA alone (simulation of FabA in absence of the rest of mAb structure) is also included in the plot (green color). It can be seen that C^α-RMSD profiles for all the Fab trajectories are similar, with the C^α-RMSD values for the simulation at 400 K (Full400K1, blue color) being relatively higher by 1–2 Å. Table II shows the average C^α-RMSD values for different regions of the mAb computed from all the full length mAb simulation trajectories. Fabs show higher average C^α-RMSD values for the 400 K trajectories, Full400K1 (C^α-RMSD is 4.1 ± 0.4 Å for FabA and 4.0 ± 0.7 Å for FabB) and Full400K2 (C^α-RMSD is 3.7 ± 0.9 Å for FabA and 4.1 ± 0.8 Å for FabB) as compared to the 298 K trajectories, Full1(C^α-RMSD is 3.1 ± 0.5 Å for FabA and 2.2 ± 0.3 Å for FabB) and Full2 (C^α-RMSD is 2.7 ± 0.4 Å for FabA and 2.1 ± 0.2 Å for FabB) (Table II). The C^α-RMSD profiles for Fab portion of the full length mAb in presence and absence of glycosylation (red and black colors) appear to converge by the end of simulation time [Fig. 2(b)]. The average C^α-RMSD values for Fab portions of the glycosylated and deglycosylated mAb trajectories are also similar (Table II). For example, the average C^α-RMSD values for FabA are 3.1 ± 0.5 Å and 2.4 ± 0.2 Å in Full1 and FullGly1 trajectories, respectively. The corresponding values for Fab B are 2.2 ± 0.3 Å and 1.9 ± 0.2 Å. The above results are also in good agreement with the residue wise BB-RMSF plots [Figs. 3(a,b)] and the average BB-RMSF values for V_L, V_H, F_v, and Fab regions shown in Table III. These regions show the greatest BB-RMSF values in the Full400K1 and Full400K2 trajectories. The two Fab regions of this murine mAb do not contain any glycosylation sites and are not glycosylated in the crystal structure.

Table IV
C^{α} RMSD (Å) Obtained by Superposing for the Conformations of C_H2 Domains from the Crystal Structure of the Murine mAb on to Those of the C_H2 Domains in the Last Frames of Various Simulation Trajectories[a]

	Crystal structure 1IGT chains B and D	FullGly1 (FullGly2)	Full1 (Full2)	Full400K1 (Full400K2)
Superposition of the C_H2 domain dimer				
Crystal structure 1IGT chains B and D		3.67(5.87)	5.21(5.86)	6.56(9.89)
FullGly1 (FullGly2)	3.67(5.87)		5.56(6.60)	6.62(9.20)
Full1 (Full2)	5.21(5.86)	5.56(6.60)		6.31(6.39)
Full400K1 (Full400K2)	6.56(9.89)	6.62(9.21)	6.31(6.39)	
Superposition of the single C_H2 domain from heavy chain 1 (HC1)				
Crystal structure 1IGT chain B		2.06(2.40)	2.46(2.61)	3.42(3.96)
FullGly1 (FullGly2)	2.06(2.40)		2.17(2.59)	3.31(3.79)
Full1 (Full2)	2.46(2.61)	2.17(2.59)		3.54(3.37)
Full400K1 (Full400K2)	3.42(3.96)	3.31(3.79)	3.54(3.37)	
Superposition of the single C_H2 domain from heavy chain 2 (HC2)				
Crystal structure 1IGT chain D		1.83(2.28)	2.74(2.32)	3.94(4.17)
FullGly1 (FullGly2)	1.83(2.28)		3.20(3.00)	4.26(4.23)
Full1 (Full2)	2.74(2.32)	3.20(3.00)		3.71(3.91)
Full400K1 (Full400K2)	3.94(4.17)	4.26(4.23)	3.71(3.91)	

[a]The numbers for the second set of simulation trajectories are shown within brackets. This table indicates that both tertiary and quaternary structures of the murine mAb are perturbed upon deglycosylation and thermal stress.

Conformational stability of the Fc region

Both thermal stress and de-glycosylation lower conformational stability of the Fc region. Figure 2(c) compares C^{α}-RMSD profiles for Fc region extracted from the 298 K trajectories of the full length mAb in presence/absence of glycans and at elevated temperature. Two 20 ns each trajectories obtained by simulating Fc alone (independent of the rest of the mAb) in presence (magenta) and absence (green) of glycans are also shown. Absence of the glycans leads to 1–2 Å higher C^{α}-RMSD values in deglycosylated Fc trajectories (Full1 and FcOnly1), while thermal stress leads to 2–4 Å higher C^{α}-RMSD values in 400 K trajectory (Full400K1). The above observations are supported by the data shown in Table II. The average C^{α}-RMSD values are higher for the Fc portion in deglycosylated mAb trajectories at 298 K (by 0.8 Å for Full1 and by 0.3 Å for Full2) and at 400 K (by 2.7 Å for Full400K1 and by 4.0 Å for Full400K2) when compared with the glycosylated mAb trajectories at 298 K, FullGly1 and FullGly2 (Table II). Figure 3(c) also shows that BB-RMSF values for the Fc region in the deglycosylated mAb are higher by 2–4 Å in Full1 (solid and dashed blue lines) and Full400K1 (solid and dashed red lines) trajectories when compared with the BB-RMSF values for the Fc region of glycosylated mAb (FullGly1, solid and dashed green lines). The average BB-RMSF values for the Fc region of the deglycosylated mAb simulations are also greater than those in glycosylated mAb simulations (Table III). For example, average BB-RMSF values for Fc region in heavy chain 1 are 6.27 ± 0.83 Å and 5.59 ± 0.97 Å for Full1 and Full400K1 trajectories versus 3.59 ± 0.67 Å in FullGly1 trajectory. Note that the average BB-RMSF values for the Fc region are similar between 298 K and 400 K trajectories of the deglycosylated mAb (Table III).

Conformational stability of the C_H2 domain

C_H2 domain region of the murine mAb is susceptible to conformational destabilization upon deglycosylation and thermal stress. Figure 2(d) shows the C^{α}-RMSD profiles for the C_H2 domain dimers extracted from different trajectories for full length mAb along with a 20 ns simulation trajectory for deglycosylated C_H2 domain dimer alone (green color). At 298 K, the C^{α}-RMSD profiles for C_H2 domain in presence (red color) and in absence of glycans (black color) follow similar trends, but the C^{α}-RMSD values are higher by 0.5–2.0 Å for the deglycosylated mAb for the most part [Fig. 2(d)]. The C_H2 domain dimer is significantly more destabilized (4–6 Å higher C^{α}-RMSD) in the 400 K trajectory of deglycosylated mAb (blue color). Note that 20 ns trajectory for deglycosylated C_H2 domain dimer alone (green color) obtained at 298 K shows C^{α}-RMSD profile that is similar to 400 K trajectory (blue color). Thus, in absence of the rest of the mAb structural context (C_H3 domains, Fabs, and hinge regions) deglycosylation causes perturbation in C_H2 domain dimer at 298 K that is of similar extent as the perturbation at 400 K with the rest of mAb structure present.

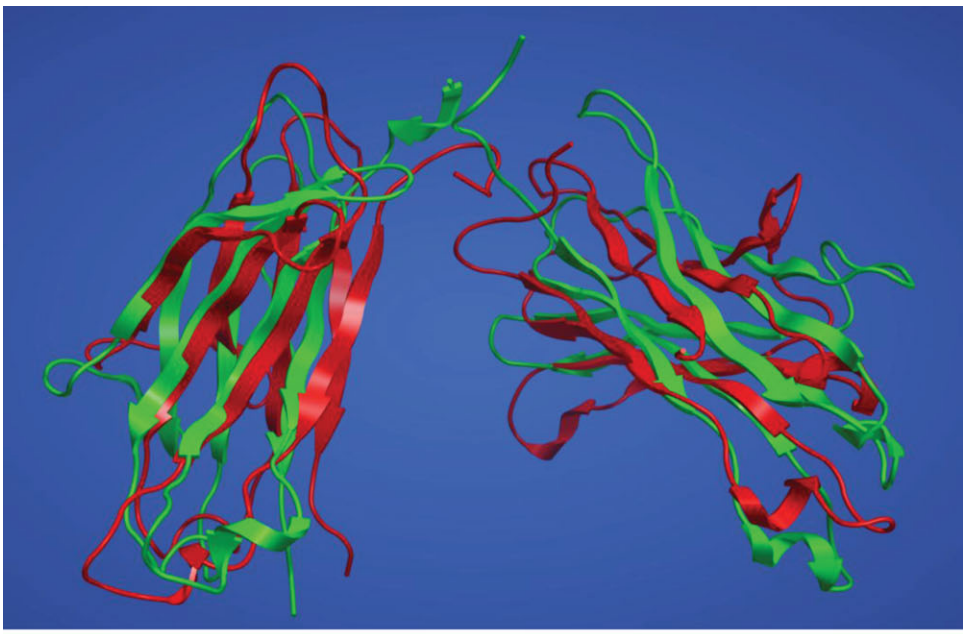
Quaternary structure of the murine mAb is perturbed upon deglycosylation and thermal stress as the two C_H2 domains move with respect to each other. Table IV shows C^{α}-RMSD values obtained by superposing the C_H2 domain conformations (dimer and monomers, residues: A232-K341 from the heavy chains. Note that residues from the lower hinge region are included in the C_H2 domain for the purpose of these calculations) in the crystal structure of the murine mAb (1IGT) on to the C_H2 domain conformations in the last frames at the end of various simulations. The C_H2 domain dimer in the last frame of first simulation trajectory of the deglycosylated

mAb at 298 K (Full11) shows a C^{α}-RMSD value of 5.2 Å when superposed on the C_H2 domain dimer in the crystal structure. In comparison, the analogous C^{α}-RMSD value is 3.7 Å for the C_H2 domain dimer in the last frame of first simulation trajectory of the glycosylated mAb at 298K (FullGly1). The C_H2 domain dimer in the last frame of first 400 K simulation trajectory of the deglycosylated mAb follows the same trend (Full400K1, C^{α}-RMSD = 6.6 Å). The second set of simulation trajectories yield similar overall trends as the first set and the larger deviations were observed consistently (Table IV). Significant deviations were also observed when C_H2 domain dimers were compared among the last frames from different simulation trajectories. Figure 5(a) shows superposition of the C_H2 domain dimers taken from the last frames of the second set of 298 K simulation trajectories for glycosylated (FullGly2, green ribbons) and deglycosylated (Full2, red ribbons) murine mAb (C^{α}-RMSD = 6.6 Å, Table IV). This figure shows that C_H2 domain monomers are oriented differently in the presence and absence of glycans, indicating that individual C_H2 domains in the deglycosylated mAb have moved with respect to each other during simulations. These above results are consistent with our previous simulations on $Hinge^{++}$ system derived from middle portions of this antibody,[27] the C^{α}-RMSD values were found to be high because the two C_H2 domain monomers in the C_H2 dimer moved with respect to each other. The $Hinge^{++}$ system did not contain glycans, Fv and C_H3 regions.[27]

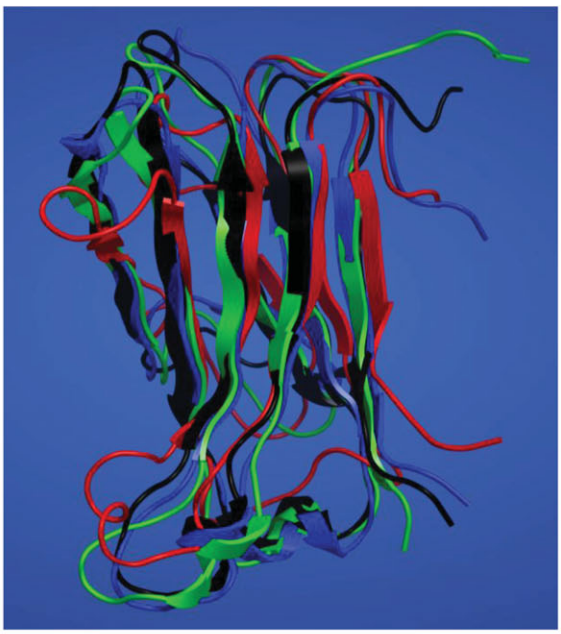
The quaternary structure perturbation is accompanied by perturbations in the tertiary structures of the individual C_H2 domains upon deglycosylation and thermal stress (Table IV). The BB-RMSF values in the 298 K simulation trajectories (Full1 and FullGly1) of the mAb in presence and absence of glycosylation are of similar magnitudes (3–6 Å) in all regions except in the C_H2 domains [see green and blue solid and dashed lines in Fig. 3(C)]. For most of the residues in the C_H2 domains, the BB-RMSF values in 298 K deglycosylated mAb simulation trajectory (Full1, blue solid and dashed lines) reach magnitudes that are comparable to the BB-RMSF values in 400 K deglycosylated mAb simulation trajectory, Full400K1 [red solid and dashed lines in the Fig. 3(c)], indicating significant destabilization of the individual C_H2 domains. The average BB-RMSF values for the C_H2 domains in the deglycosylated mAb are similar for 298 and 400 K trajectories (Table III). For example, average BB-RMSF values for the two C_H2 domains in heavy chains 1 and 2 are 6.58 ± 0.92 Å and 7.29 ± 1.06 Å, respectively, in the first 298 K deglycosylated mAb simulation (Full1). The corresponding values are 5.58 ± 0.68 Å and 5.61 ± 1.10 Å, respectively, in the first 400 K deglycosylated mAb simulation (Full400K1). In comparison, the average BB-RMSF values C_H2 domains in the heavy chains 1 and 2 are 3.34 ± 0.62 Å and 3.62 ± 0.61 Å, respectively, in the first 298 K glycosylated mAb simulation (FullGly1).

These above observations are supported by the C^{α}-RMSD values obtained from superposing the individual C_H2 domains extracted from the crystal structure and the last frames of the simulation trajectories (Table IV). The conformations of the individual C_H2 domains extracted from the last frames of deglycosylated mAb simulation trajectories (Full1, Full2, Full400K1, and Full400K2) show greater deviations (range of C^{α}-RMSD values = 2.5–4.2 Å) from individual C_H2 domain conformations in the crystal structure than the deviations shown by the individual C_H2 domain conformations extracted from the last frames of glycosylated mAb trajectories (FullGly1 and FullGly2, range of C^{α}-RMSD values = 1.8–2.4Å). For example, the C_H2 domain of heavy chain 2 (HC2) from the last frame in the first 298 K trajectory from glycosylated murine mAb (FullGly1) superposes with C^{α}-RMSD value of 1.83 Å on the corresponding C_H2 domain in the chain D of the crystal structure. The analogous C_H2 domains in the last frames from the first 298 K and 400 K trajectories of the deglycosylayed mAb (Full1 and Full400K1) show C^{α}-RMSD values of 2.74 and 3.94 Å, respectively. Figure 5(b) shows the superposition of all these four C_H2 domain conformations. It can be seen that the C_H2 domains show increasing deviations from the conformation in the crystal structure (black ribbon) as we go from FullGly1 (green ribbon) to Full1 (blue ribbon) to FullGly400K1 (red ribbon), indicating greater perturbations of the tertiary structure. Deglycosylation solvent exposes several hydrophobic and aromatic residues, which lie in structural vicinity of the glycans.[8] The average SASA values for several C_H2 domain residues are changed significantly between the glycosylated and the deglycosylated mAb simulation trajectories [Fig. 4(c)]. This may lead to conformational destabilization of the C_H2 domain monomers. However, no large scale secondary structural unfolding of the individual C_H2 domains was observed within the duration of the 298 K and 400 K simulations of the deglycosylated mAb.

Our observations on destabilization of C_H2 domain can be rationalized by taking into account the location of glycans in antibody structure. The Asn-linked glycans in this antibody are attached to each C_H2 domain monomer and form an integral part of C_H2 domain dimer interface (Fig. 1). These observations are also in good agreement with the previous experimental and computational studies on other mAbs. Zheng *et al.*[11] have studied the impact of deglycosylation on conformational stability of three humanized mAbs. Thermal and chemical denaturation as well as accelerated stability studies showed that deglycosylation lowered the thermal stability of C_H2 domains. Furthermore, the mAbs became more susceptible to chemical denaturation by Guanidium hydrochloride and showed higher levels of aggregation.[11] The NMR experiments performed on Fc portion of human IgG mAbs also indicate that aglycosylated Fc has lower



(a)



(b)

Figure 5

(**a**) Superposition of the C_H2 dimers taken from the two heavy chains (HC1 and HC2) of conformations contained in the last frames in the second set of 298 K simulation trajectories for glycosylated (FullGly2, green ribbons) and deglycosylated (Full2, red ribbons) full length mAb. Note that the individual C_H2 domains have significantly moved with respect to each other, thereby, perturbing the quaternary structure upon deglycosylation. (**b**) Superposition of the monomeric C_H2 domain (sequence regions A232–K342 in HC2) taken from the last frames in the first set of simulation trajectories for 298 K glycosylated (FullGly1, green ribbons), 298 K deglycosylated (Full1, blue ribbons) and 400 K deglycosylated (Full400K1, red ribbons) of the murine mAb on to the C_H2 domain in Chain D of the crystal structure (1IGT, black ribbons). Note that the individual C_H2 domain conformations in the simulation trajectories for deglycosylated mAb (blue and red ribbons) have deviated further from the crystal structure (black ribbon).

conformational stability and greater aggregation propensity.[10] The MD simulations of a human IgG1 mAb in presence of the glycans[9] show that the aromatic residues, Phe 241 and Phe243, in C_H2 domains form cation-pi interactions with the glycans.

Conformational stability of the C_H3 domain

Both thermal stress and deglycosylation have smaller impact on the conformational stability of the C_H3 domain. Figure 2(e) shows corresponding C^{α}-RMSD profiles for the C_H3 domain dimer. For the duration of simulation, the C^{α}-RMSD values for C_H3 domain portion of the full length mAb remain constant at ~2 Å for both glycosylated and deglycosylated simulation trajectories (red and black colors). A trajectory from C_H3 domain dimer alone also shows similar C^{α}-RMSD profile (green color). The C^{α}-RMSD profile for the C_H3 domain dimer shows only small increase (~0.5 Å higher) in the 400 K full length mAb simulation trajectory (blue color). Taken together these observations indicate that among all domains, C_H3 domains of this mAb were the most successful at resisting conformational destabilization caused by thermal stress and deglycosylation.

Consequences of conformational destabilization of murine mAb towards aggregation: Activation of different APRs

Among all the physico–chemical degradation routes associated with biotherapeutics, aggregation is the most common one.[4,5,31] Protein aggregation is also an important research area in fundamental protein science and, in recent years, several computational tools have become available for prediction of potential APRs.[5,31] Most of these APR prediction tools identify short (5–9 residues) sequence regions as APRs based on their ability to form cross-β motif.[32,33] This motif has been observed in amyloid fibrils and plaques formed by the proteins involved in neurodegenerative diseases[34] and also by the proteins elsewhere.[35,36] Even in amorphous aggregates formed by large proteins, the cross-β motif can be detected using marker dyes such as Thioflavin T and congo red.[37] The cross-β motif containing irreversible aggregates have been detected in the expired biotherapeutics[38] and in acid-denatured mAbs.[39] Such aggregates are of concern because of their potential to elicit immune response.[7,38,40,41] In this report, APRs in the murine mAb were predicted using a combination[20] of TANGO[18] and PAGE.[19] Both TANGO and PAGE have been extensively validated by the developers of these programs.[18,19] Additionally, we have independently collected experimental examples that show APR-disruption mitigates aggregation in antibodies and other globular proteins to validate that APRs predicted by these programs do contribute towards aggregation of large proteins.[42,43] Furthermore, disruption of a TANGO predicted strong APR in the light chain framework 2–CDR loop 2 (FR2-L2) region of an IgG2 mAb by a single point mutation reduced its aggregation propensity and improved solubility as indicated by biophysical experiments (our internal efforts, data not shown). A double mutant was designed to simultaneously disrupt two strong TANGO predicted APRs present in the human IgG C_H2 domain. The double mutant of C_H2-based binder showed improved *E. coli* expression as detected by western blot experiments and remarkably improved monomer levels in size exclusion chromatography experiments (Q. Zhao, R. Gong, and D.S. Dimitrov, personal communication). In both the cases, the binding affinities of the molecules toward their cognate partners were preserved.

Most therapeutic mAbs currently available in the clinic contain several strongly predicted cross-β motif forming APRs distributed all over their sequence and structure.[7,42,44] However, many of the predicted APRs may have been effectively blocked from initiating and promoting aggregation by their neighboring gate keeping residues[45] and by their protein structural context such as location, solvent exposure and flexibility.[42,44,46] Therefore, sequence-based APR predictions need to be further refined via additional computations as experimentally testing each predicted APR would be a tedious task. We have used MD simulations to make knowledge based distinction between potentially "active" and "inactive" APRs within the structural context of antibodies.[46] Supporting experimental data on such large systems is generally unavailable.

All atom MD simulations study aggregation in peptides and proteins in several different ways.[5,47] A common approach has been to investigate aggregation seeding requirements for a chosen small peptide (often, less than 10 residues) and its sequence or charge variants. Usually, the chosen peptides are segments from amyloidogenic proteins or designed peptides that have been experimentally shown to form amyloid fibrils. Such studies enable teasing out sequence and charge effects that underpin aggregation initiation/propagation kinetics and thermodynamics, along with providing details about seed configurations and copy numbers. A second approach pioneered by Vendruscolo and coworkers is to combine APR predictions, NMR experiments and restrained MD simulations on small single domain proteins to describe their aggregation competent transition state ensembles.[48,49] In a third completely different approach, Trout and coworkers have used full length mAb MD simulations to identify APRs via their SAP (spatial aggregation propensity) approach.[26] The APRs identified by SAP are conformational APRs and the residues involved constitute surface exposed hydrophobic patches. Such APRs are not relevant to cross-β motif formation and may or may not overlap in the protein structure with the cross-β motif forming sequentially contiguous APRs predicted by TANGO, PAGE, and other such programs. Along with

Table V
Potential APRs in MAb231 Predicted by a TANGO/PAGE Combination[a]

APR	Location	Description
Light chain: 2-IVLTQ-6	Edge β strand at N-terminal (Framework 1 region)	Weak predictions by TANGO and PAGE
Light chain: 29-INVWLSWY-36	CDR1 (L1)-Framework 2 region	Strongly predicted by TANGO
Light chain: 71-FTLTI-75	Framework 3 region	Strongly predicted by TANGO
Light chain: 131-SVVCFL-136	C_L	Strongly predicted by TANGO and PAGE
Heavy chain: 32–YYMYWV-37	CDR1 (H1)-Framework 2 region	Strongly predicted by TANGO
Heavy chain: 47–WVAYIS-52	Framework 2–CDR 2 (H2)	Strongly predicted by TANGO
Heavy chain: 240-SVFIFP-245	C_H2 domain	Strongly predicted by PAGE
Heavy chain: 251-VLMISL-256	C_H2 domain	Strongly Predicted by TANGO
Heavy chain: 259-IVTCVVV-265	C_H2 domain	Strongly predicted by TANGO
Heavy chain: 276-ISWFV-280	C_H2 domain	Strongly predicted by TANGO
Heavy chain: 364-VTLTCMV-370	C_H3 domain	Strongly predicted by TANGO

[a]The APRs which may initiate and actively promote aggregation of the murine mAb upon thermal stress and deglycosylation are shown in bold font.

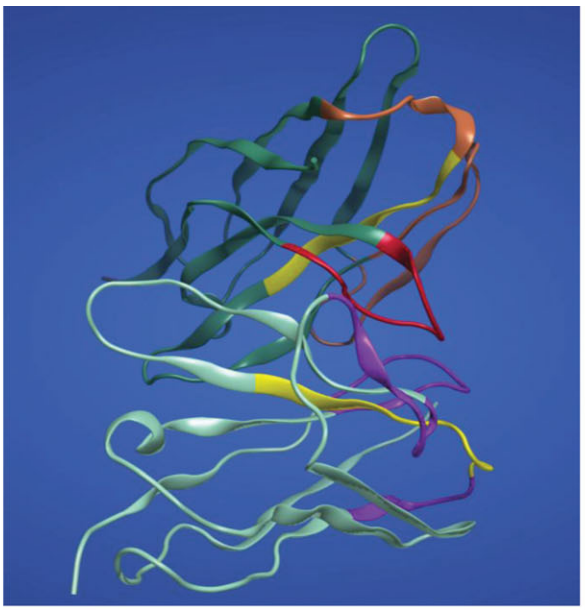
the all atom MD simulations, coarse-grained simulations have been also used to study aggregation in peptides and proteins.[5,50,51]

Can short time scale MD simulations yield useful information on aggregation, which occurs on long time scales? This question has macroscopic aggregation kinetics and microscopic protein structural dynamics components: (1) Can we predict long term kinetics of aggregation process based on short elevated temperature MD or other simulations? Although beyond scope of this report, this aspect is being intensively investigated by others.[39,52,53] (2) Do the structural perturbations observed during short elevated temperature MD simulation runs inform us about the initial events in aggregation process? Several reports in literature have found interesting correspondences between experimentally studied aggregation mechanisms and the initial structural perturbations observed during elevated temperature (400–500 K) MD simulations of proteins. Salient examples include Acylophosphatase, human transthyretin, human lysozyme, immunoglobulin light chain κ, TGF-β3, and γδ-Crystallin.[54–60] Recently, De Simone *et al.* have used a combination of all atom and coarse grain simulations to understand aggregation of fungal hydrophobin EAS at air–water interface.[57] While acknowledging that full length mAbs are considerably bigger and more complex than the proteins cited above, these examples provide confidence that MD simulations presented in this work may yield useful information on their "active" APRs.

Potential cross-β motif forming APRs identified in MAb 231 are listed in Table V and the overall aggregation spectrum for both light and heavy chains of this mAb is shown in the Supporting Information [Fig. S1(a,b)]. BB-RMSF and mean SASA values computed from the simulation trajectories helped identify the APRs, which are more likely to initiate cross-β motif formation and promote aggregation upon deglycosylation and upon applying thermal stress. Besides a strong prediction for the ability to form cross-β motif, identification of a potentially active APR in this study is based on the following additional structural requirements[42]: First, it should lie at or close to protein surface so that it becomes solvent exposed upon perturbations in the native protein conformation. Second, it should be either flexible itself or lie in surroundings that are flexible so that it can seek out binding partners from other copies of the protein molecule when exposed to the solvent.

CDR overlapping APRs may initiate aggregation upon thermal stress

Aggregation due to the thermal stress mostly likely comes from the APRs that overlap with the CDRs in variable region of the mAb [see Figs. 3(a,b) and 4(a,b)], the CDRs and APRs are indicated by blue and red bars, respectively, along the *X*-axis). The second APR in the light chains (29-INVWLSWY-36), overlapping with the CDR1 (L1), lies in the region of high BB-RMSF [Fig. 3(a)] and several residues in and around this APR are also solvent exposed [Fig. 4(a)]. Similarly, the first APR in the heavy chains (32-YYMYWV-37), which overlaps with first CDR (H1 loop), also lies in high BB-RMSF region [Fig. 3(b)] and a few residues in this region are also solvent exposed [Fig. 4(b)]. The 400 K simulation trajectories further amplify the BB-RMSF values of these regions. The second set of simulation trajectories show similar results for these APRs [Fig. S3(a,b) in Supporting Information]. Both of these APRs are strongly predicted by TANGO[18] and their locations in the Fv region are shown in Figure 6(a). Tables VI and VII list the average SASA and backbone torsion angles (ϕ, ψ) values for all the residues involved in these APRs in the first set of simulation trajectories (FullGly1, Full1, and Full400K1). It can be seen that average SASA values are similar for all residues in these APRs, but the standard deviations are greater for the 400 K trajectory (Table VI). The backbone torsion angles also show similar behavior as SASA values in all but two residues in these APRs (Table VII). The average ϕ and ψ values for Asn 30, in the APR Ile 29–Tyr



(a)

(b)

Figure 6

Different APRs which may initiate and promote aggregation upon conformational destabilization due to thermal stress and deglycosylation. (**a**) Two APRs, viz. 29-INVWLSWY-36 in light chain (cyan ribbon) and 32-YYMYWV-37 in heavy chain (green ribbon) that can initiate aggregation in response to thermal stress is shown in yellow colored ribbons. The APRs lie in the variable region (Fv) of the murine mAb and overlap with the CDRs. The CDRs are shown in magenta, brown, and red colored ribbons. (**b**) An APR 240-SVFIFP-245 (yellow colored ribbons) that lies in structural vicinity to the Asn-linked glycans (shown in line representation) attached to each C_H2 domain (brown and magenta ribbons) becomes solvent exposed up deglycosylation. The atoms in the residues belonging to this APR are shown in ball and stick representation.

36 in the light chains, and Tyr 32, in the APR Tyr 32–Val 37 in the heavy chains, change significantly in Full400K1 simulations. Taken together, the above observations indicate that the two CDR overlapping APRs have a greater chance at becoming solvent exposed, seeding cross-β motif formation and promoting aggregation upon thermal stress.

The APR 47-WVAYIS-52 in the heavy chain also overlaps with CDR2 (H2 loop) and residues in and around this region also show high BB-RMSF values

Table VI
Average SASA Values for APRs that can Initiate Cross-β Motif Formation and Actively Promote Aggregation of Murine mAb upon Conformational Destabilization due to Thermal Stress and Deglycosylation

Residue in APR	Full1[a]		FullGly1[a]		Full400K1[a]	
	Chain 1 SASA(%)	Chain 2 SASA(%)	Chain 1 SASA(%)	Chain 2 SASA(%)	Chain 1 SASA(%)	Chain 2 SASA(%)
LC: I29	0.84 ± 0.76	0.15 ± 0.33	0.36 ± 0.55	0.28 ± 0.53	0.44 ± 0.60	0.76 ± 0.86
LC: N30	28.69 ± 6.58	22.85 ± 8.83	26.43 ± 4.88	25.20 ± 5.70	16.77 ± 9.27	14.93 ± 10.39
LC: V31	2.40 ± 1.15	5.22 ± 1.72	3.67 ± 1.61	4.57 ± 1.69	6.36 ± 2.66	7.89 ± 3.41
LC: W32	10.80 ± 3.10	9.07 ± 3.33	9.78 ± 2.17	10.24 ± 2.57	11.69 ± 4.36	13.38 ± 5.38
LC: L33	0.50 ± 0.44	0.51 ± 0.56	0.18 ± 0.24	0.04 ± 0.09	0.46 ± 0.63	0.17 ± 0.33
LC: S34	1.80 ± 1.62	0.62 ± 0.80	0.62 ± 0.77	0.13 ± 0.24	0.59 ± 1.01	0.66 ± 1.20
LC: W35	0.03 ± 0.09	0.04 ± 0.11	0.06 ± 0.15	0.12 ± 0.18	0.09 ± 0.20	0.16 ± 0.36
LC: Y36	0.60 ± 0.56	0.28 ± 0.33	0.55 ± 0.68	0.68 ± 0.56	0.72 ± 0.84	0.83 ± 1.06
HC: Y32	12.60 ± 2.35	15.36 ± 2.18	13.69 ± 2.73	15.18 ± 2.60	16.18 ± 6.41	11.70±4.10
HC: Y33	21.31 ± 3.21	18.77 ± 2.19	16.92 ± 3.25	22.61 ± 3.10	14.94 ± 5.19	21.80 ± 3.42
HC: M34	0.26 ± 0.62	0.16 ± 0.37	0.14 ± 0.34	0.08 ± 0.22	0.93 ± 1.33	0.21 ± 0.44
HC: Y35	0.39 ± 0.61	4.52 ± 2.23	0.40 ± 0.64	3.79 ± 1.96	1.28 ± 1.54	0.84 ± 1.28
HC: W36	0.01 ± 0.05	0.01 ± 0.04	0.02 ± 0.07	0.01 ± 0.05	0.12 ± 0.31	0.06 ± 0.16
HC: V37	0.08 ± 0.21	0.10 ± 0.24	0.15 ± 0.25	0.01 ± 0.05	0.18 ± 0.32	0.29 ± 0.51
HC: S240	12.64 ± 7.04	13.91 ± 8.09	9.06 ± 6.70	10.81 ± 7.14	5.74 ± 6.93	11.83 ± 6.47
HC: V241	1.17 ± 1.41	0.56 ± 0.91	0.71 ± 0.80	1.34 ± 1.68	2.32 ± 2.32	4.05 ± 3.58
HC: F242	19.83 ± 3.55	16.94 ± 2.35	3.08 ± 3.10	1.74 ± 2.55	16.68 ± 4.16	18.72 ± 3.57
HC: I243	6.44 ± 1.34	7.66 ± 1.69	5.45 ± 1.71	6.41 ± 1.35	6.84 ± 2.10	7.80 ± 2.54
HC: F244	20.05 ± 2.71	19.82 ± 2.51	1.99 ± 1.51	2.90 ± 2.20	18.17 ± 4.52	18.94 ± 3.68
HC: P245	13.02 ± 2.35	14.10 ± 2.54	5.40 ± 1.88	10.62 ± 3.01	9.40 ± 3.68	10.02 ± 3.51

[a]FullGly1, the first simulation trajectory of glycosylated murine mAb at 298 K; Full1, the first simulation trajectory of deglycosylated mAb at 298 K; Full400K1, the first simulation trajectory of deglycosylated mAb at 400 K.

[Fig. 3(b)]. But, average SASA values indicate that this APR remains solvent inaccessible [Fig. 4(b)] during the all the simulations, including the one performed at 400 K. Therefore, this APR has lower probability of becoming active upon short duration thermal stress.

Deglycosylation causes solvent exposure of an APR located in the C_H2 domain

Deglycosylation may activate a strongly predicted APR in the C_H2 domains. There are four predicted APRs in each C_H2 domains (Table V). All of these APRs lie in regions, which show higher BB-RMSF values for the

Table VII
Average ϕ and ψ Values for Residues in the APRs that can Initiate Cross-β Motif Formation and Actively Promote Aggregation of Murine mAb upon Conformational Destabilization Due to Thermal Stress and Deglycosylation

Residue in APR	Full1[a]		FullGly1[a]		Full400K1[a]	
	Chain 1 (ϕ, ψ)	Chain 2 (ϕ, ψ)	Chain 1 (ϕ, ψ)	Chain 2 (ϕ, ψ)	Chain 1 (ϕ, ψ)	Chain 2 (ϕ, ψ)
LC: I29	−123 ± 16, 13 ± 12	−130 ± 15, 23 ± 13	−126 ± 14, 22 ± 11	−129 ± 14, 24 ± 12	−128 ± 17, 7 ± 17	−125 ± 23, 7 ± 19
LC: N30	44 ± 29, −58 ± 18	8 ± 43, −61 ± 25	45 ± 14, −77 ± 13	35 ± 26, −80 ± 18	−30 ± 46, −34 ± 29	−35 ± 40, −39 ± 27
LC: V31	−148 ± 26, −8 ± 12	−128 ± 17, 3 ± 13	−141 ± 14, 6 ± 11	−135 ± 11, 9 ± 12	−129 ± 16, 7 ± 17	−86 ± 36, −10 ± 18
LC: W32	−89 ± 16, 72 ± 11	−92 ± 19, 34 ± 23	−83 ± 11, 52 ± 13	−81 ± 14, 38 ± 13	−116 ± 27, 34 ± 31	−122 ± 27, 9 ± 26
LC: L33	−143 ± 13, 148 ± 10	−122 ± 25, 146 ± 9	−135 ± 14,151 ± 12	−132 ± 14, 149 ± 8	−124 ± 33,149 ± 41	−76 ± 34,135 ± 24
LC: S34	−141 ± 9, 157 ± 16	−136 ± 10,158 ± 37	−143 ± 8, 158 ± 14	−143 ± 9, 147 ± 80	−139 ± 11,155 ± 35	−140 ± 12,150 ± 63
LC: W35	−134 ± 9, 152 ± 10	−128 ± 11, 149 ± 8	−136 ± 10, 146 ± 8	−135 ± 9, 152 ± 11	−135 ± 11,151 ± 12	−135 ± 12,149 ± 11
LC: Y36	−133 ± 9, 157 ± 43	−136 ± 9, 158 ± 48	−138 ± 9, 157 ± 57	−128 ± 12, 159 ± 33	−135 ± 11,149 ± 74	−131 ± 13,152 ± 58
HC: Y32	−135 ± 12, 136 ± 10	−133 ± 11,144 ± 14	−129 ± 14,133 ± 13	−127 ± 21, 136 ± 12	−113 ± 31, 31 ± 119	−126 ± 23,139 ± 17
HC: Y33	−76 ± 16, 148 ± 9	−102 ± 22, 153 ± 9	−75 ± 17, 156 ± 17	−74 ± 12, 145 ± 9	−91 ± 40, 150 ± 43	−77 ± 16, 146 ± 10
HC: M34	−135 ± 9, 156 ± 13	−141 ± 9,159 ± 38	−140 ± 9, 158 ± 25	−136 ± 9, 158 ± 20	−134 ± 17,152 ± 41	−136 ± 11,153 ± 16
HC: Y35	−141 ± 9, 159 ± 47	−142 ± 8, 159 ± 42	−135 ± 10,160 ± 39	−137 ± 11, 159 ± 45	−136 ± 12,149 ± 70	−143 ± 10,153 ± 54
HC: W36	−132 ± 8, 149 ± 8	−127 ± 9, 148 ± 7	−131 ± 8, 150 ± 7	−126 ± 9, 148 ± 8	−131 ± 11,145 ± 10	−129 ± 11,146 ± 11
HC: V37	−138 ± 9, 155 ± 58	−141 ± 8, 154 ± 15	−138 ± 9, 155 ± 46	−144 ± 8, 156 ± 39	−137 ± 12,150 ± 62	−137 ± 11,150 ± 61
HC: S240	−126 ± 19, 147 ± 55	−121 ± 22,147 ± 35	−146 ± 27,139 ± 83	−116 ± 32, 136 ± 43	−100 ± 34, 22 ± 49	−122 ± 31,117 ± 92
HC: V241	−137 ± 10, 145 ± 9	−135 ± 14, 141 ± 8	−140 ± 10,143 ± 10	−132 ± 20, 140 ± 8	−124 ± 23,135 ± 14	−125 ± 25,140 ± 11
HC: F242	−138 ± 9, 154 ± 16	−138 ± 9, 155 ± 14	−138 ± 9, 157 ± 23	−139 ± 9, 153 ± 12	−140 ± 14,150 ± 45	−142 ± 11,148 ± 52
HC: I243	−140 ± 10, 147 ± 8	−141 ± 10, 147 ± 8	−141 ± 10, 144 ± 8	−144 ± 10, 145 ± 7	−141 ± 14,146 ± 18	−138 ± 18,147 ± 28
HC: F244	−140 ± 8, 142 ± 9	−140 ± 8, 143 ± 9	−141 ± 8, 142 ± 9	−139 ± 8, 143 ± 9	−136 ± 18,137 ± 21	−126 ± 18,150 ± 17
HC: P245	−75 ± 9, 156 ± 61	−76 ± 9, 155 ± 63	−76 ± 9, 146 ± 86	−75 ± 9, 157 ± 56	−70 ± 13, 138 ± 88	−61 ± 12, 152 ± 41

[a]Simulation trajectory names are same as in Table VI.

deglycosylated mAb in at least one C_H2 domain monomer [Fig. 3(c)] and Fig. S3(C) in Supporting Information) both at 298 K and at 400 K. Besides the high BB-RMSF values in the deglycosylated mAb simulation trajectories, the first APR in the C_H2 domain 240-SVFIFP-245 also shows increased solvent exposure upon deglycosylation [Fig. 4(c) and Table VI)] and it is located in an edge β-strand (strand A) of the C_H2 domain. In the absence of opposing structural features,[61–63] the edge β-strands can be involved in uncontrolled run-away aggregation.[64] This APR was strongly predicted by PAGE[19] and it contains two Phe residues (F242 and F244), which interact with glycans and are solvent inaccessible in the glycosylated mAb [Fig. 6(b)]. In absence of the glycans, both F242 and F244 in this APR become solvent exposed to a greater extent as shown by their increased mean SASA values (Table VI). For example, the mean SASA values for Phe 242 and Phe 244 in the heavy chain 1 are 3.08 ± 3.10% and 1.99 ± 1.51%, respectively in the FullGly1, first 298K simulation trajectory of the glycosylated murine mAb. These mean SASA values for Phe 242 and Phe 244 increase to 19.83 ± 3.55% and 20.05 ± 2.71%, respectively in Full1, first 298 K simulation trajectory of the deglycosylated murine mAb. The solvent accessibilities of these residues in first 400K simulation trajectory of the deglycosylated murine mAb [Full400K1, mean SASA = 16.68 ± 4.16% for Phe242 in HC1 and 18.17 ± 4.52% for Phe244 in HC1, Table VI] remain similar to those in the Full1. These results are consistent with observations made by Voynov *et al.*[9] on homologous Phe residues in the C_H2 domain whose increased solvent exposure upon deglycosylation leads to increased aggregation of a human IgG1 mAb.

The average backbone torsion angle values of the Phe 242 and Phe 244 remain similar between Full1, FullGly1, and Full400K1 trajectories (Table VII). For example, average ϕ and ψ values for heavy chain 1 Phe 242 in the Full1 trajectory are −138 ± 9 and 154 ± 16, respectively. ϕ and ψ values for the same residue in FullGly1 trajectory are −138 ± 9 and 157 ± 23, respectively. In the Full400K1 trajectory, these values for heavy chain 1 Phe 242 are −140 ± 14 and 150 ± 45, respectively. Furthermore, all the residues in this β-strand do not show significant change in average backbone torsion angle values over the course of the Full1 and Full400K1 simulation trajectories, as compared to the FullGly1 trajectory (Table VII). Thus, deglycosylation and thermal stress did not lead to unfolding of the C_H2 domain strand A within the duration of simulations reported in this work.

The APR 240-SVFIFP-244 is conserved between murine and human mAbs (239-SVFLFP-244 in human IgG mAbs) and is also found in all commercially available therapeutic mAbs.[44] The scaffold engineering efforts for stable aglycosylated C_H2 domain based modalities[65,66] expressed in *E. coli*, involved creation of an additional disulfide bond between the two edge β-strands A and G. One of the two residues, that were mutated to create this disulfide bond, lies within this APR (Leu242 to Cys[66] in the 239-SVFLFP-244). It is conceivable that the additional disulfide bond would reduce the ability of this APR containing edge β-strand (strand A) to break loose from the rest of the C_H2 domain and initiate self-association by seeking out binding partners from other copies of the same molecule. Furthermore, the work of Price *et al.*[67] shows that interaction of Phe with Asn-linked Glycans contributes towards increased protein stability. The authors have proposed three "enhanced aromatic sequons," which can stabilize proteins via glycosylation when placed in the structural context of appropriate reverse turns and all three involve packing of Phe aromatic ring with the glycan. These motifs are Phe–Asn (linked to glycan)–X1–Thr, Phe–X2–Asn (linked to glycan)–X1–Thr, and Phe–X3–X2–Asn (linked to glycan)–X1–Thr.[67]

The other three APRs in the C_H2 domains also show increased BB-RMSF in the deglycosylated mAb simulation trajectories [Fig. 3(c) and S3(C)] but their mean SASA values do not change [Fig. 4(c)]. These APRs may have lower chance at being able to initiate cross-β motif formation and promote aggregation upon deglycosylation. However, the APR 251-VLMISL-256, which is located in the loop AB (the loop connecting β-strands A and B) and is partly solvent exposed, also appears to be another promising candidate site for initiating of β-aggregation [Figs. 3(c) and 4(c)]. This APR also shows the largest BB-RMSF values in deglycosylated mAb simulation trajectory at 298 K, similar to those in the 400 K simulation trajectory [Fig. 3(c)]. Furthermore, this APR is also strongly predicted by TANGO (Table V).

CONCLUSIONS

This study indicates that consequences of the conformational destabilization of a murine mAb are distributed nonuniformly in its structure and different regions of the mAb may respond to thermal stress and deglycosylation differently. Thermal stress appears to destabilize the tertiary and quaternary structures of the mAb to greater extents than deglycosylation. Moreover, impact of thermal stress on conformational stability of the mAb is all pervasive and portions of Fv region, hinge and C_H2 domains show greater fluctuations in response to the thermal stress. Two APRs overlapping with the CDRs in this mAb may have greater chances at being able to initiate cross-β motif formation and promote aggregation upon thermal stress. The conformational destabilization caused by deglycosylation is limited to C_H2 domains. Removal of the glycans solvent exposes an APR located in the edge β-strand A of the individual C_H2 domains.

APRs that become active and underpin aggregation in response to different physicochemical stresses, can differ

in their sequence and structural locations. Thus, rational engineering efforts on the protein molecules via APR disruption need to be tailored towards specific stress types and protecting factors that resist conformational destabilization due to one stress may not work for the others.

ACKNOWLEDGMENTS

The authors wish to thank Drs. Kevin King, Sandeep Nema, and Ron Peeples for their interest. Pfizer research informatics and business technologies are thanked for providing the computational resources for these simulations. Xiaoling Wang and Patrick M. Buck thank Pfizer for postdoctoral fellowships. Two anonymous referees are acknowledged for their invaluable criticism and suggestion to improve this manuscript.

REFERENCES

1. Carter PJ. Introduction to current and future protein therapeutics: a protein engineering perspective. Exp Cell Res 2011;317:1261–1269.
2. Carter PJ. Potent antibody therapeutics by design. Nat Rev Immunol 2006;6:343–357.
3. Wang W, Singh S, Zeng DL, King K, Nema S. Antibody structure, instability, and formulation. J Pharm Sci 2007;96:1–26.
4. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of protein pharmaceuticals: an update. Pharm Res 2010; 27:544–575.
5. Agrawal NJ, Kumar S, Wang X, Helk B, Singh SK, Trout BL. Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions. J Pharm Sci 2011;100: 5081–5095.
6. Singh SK. Impact of product-related factors on immunogenicity of biotherapeutics. J Pharm Sci 2011;100:354–387.
7. Kumar S, Mitchell MA, Rup B, Singh SK. Relationship between potential aggregation-prone regions and HLA-DR-binding T-cell immune epitopes: implications for rational design of novel and follow-on therapeutic antibodies. J Pharm Sci 2012;101:2686–2701.
8. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. The impact of glycosylation on the biological function and structure of human immunoglobulins. Annu Rev Immunol 2007;25:21–50.
9. Voynov V, Chennamsetty N, Kayser V, Helk B, Forrer K, Zhang H, Fritsch C, Heine H, Trout BL. Dynamic fluctuations of protein-carbohydrate interactions promote protein aggregation. PLoS One 2009;4:e8425.
10. Latypov RF, Hogan S, Lau H, Gadgil H, Liu D. Elucidation of acid-induced unfolding and aggregation of human immunoglobulin IgG1 and IgG2 Fc. J Biol Chem 2012;287:1381–1396.
11. Zheng K, Bantog C, Bayer R. The impact of glycosylation on monoclonal antibody conformation and stability. mAbs 2011;3:568–576.
12. Beck A, Reichert JM. Marketing approval of mogamulizumab: a triumph for glyco-engineering. mAbs 2012;4:419–425.
13. Barb AW, Prestegard JH. NMR analysis demonstrates immunoglobulin G N-glycans are accessible and dynamic. Nat Chem Biol 2011; 7:147–153.
14. Labrijn AF, Aalberse RC, Schuurman J. When binding is enough: nonactivating antibody formats. Curr Opin Immunol 2008;20:479–485.
15. Harris LJ, Larson SB, Hasel KW, McPherson A. Refined structure of an intact IgG2a monoclonal antibody. Biochemistry 1997;36:1581–1597.
16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
17. Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL. Design of therapeutic proteins with enhanced stability. Proc Natl Acad Sci U S A 2009;106:11937–11942.
18. Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004;22:1302–1306.
19. Tartaglia GG, Cavalli A, Pellarin R, Caflisch A. Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. Prot Sci 2005;14:2723–2734.
20. Wang X, Singh SK, Kumar S. Potential aggregation-prone regions in complementarity-determining regions of antibodies and their contribution towards antigen recognition: a computational analysis. Pharm Res 2010;27:1512–1529.
21. William LJ, Corky J. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. J Comput Chem 1998;19:1179–1186.
22. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
23. Cheatham TEI, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J Biomol Struct Dyn 1999;16:845–862.
24. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 2008;29:622–655.
25. Brandt JP, Patapoff TW, Aragon SR. Construction, MD Simulation, and hydrodynamic validation of an all-atom model of a monoclonal IgG antibody. Biophys J 2010;99:905–913.
26. Chennamsetty N, Helk B, Voynov V, Kayser V, Trout BL. Aggregation-prone motifs in human immunoglobulin G. J Mol Biol 2009;391:404–413.
27. Wang X, Kumar S, Singh SK. Disulfide scrambling in IgG2 monoclonal antibodies: insights from molecular dynamics simulations. Pharm Res 2011;28:3128–3144.
28. MacCallum RM, Martin ACR, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. J Mol Biol 1996;262:732–745.
29. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:27–38.
30. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–380.
31. Kumar S, Wang X, Singh SK. Identification and impact of aggregation prone regions in proteins and therapeutic mAbs. In: Wang W, Roberts CJ, editors. Aggregation of therapeutic proteins. Hoboken: Wiley; 2010. pp 103–118.
32. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJW, McFarlane HT, Madsen AO, Riekel C, Eisenberg D. Atomic structures of amyloid cross-b spines reveal varied steric zippers. Nature 2007;447: 453–457.
33. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D. Structure of the cross-b spine of amyloid-like fibrils. Nature 2005;435:773–778.
34. Eisenberg D, Nelson R, Sawaya MR, Balbirnie M, Sambashivan S, Ivanova MI, Madsen AO, Riekel C. The structural biology of protein aggregation diseases: fundamental questions and some answers. Acc Chem Res 2006;39:568–575.
35. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R. Bacterial inclusion bodies contain amyloid-like structure. PLoS Biol 2008;6:1791–1801.
36. Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JWH, Rousseau F. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods 2010;7:237–242.

37. Ramshini H, Parrini C, Relini A, Zampagni M, Mannini B, Pesce A, Saboury AA, Nemat-Gorgani M, Chiti F. Large proteins have a great tendency to aggregate but a low propensity to form amyloid fibrils. PLoS One 2011;6:e16075.
38. Maas C, Hermeling S, Bouma B, Jiskoot W, Gebbink MFBG. A role for protein misfolding in immunogenicity of biopharmaceuticals. J Biol Chem 2007;282:2229–2236.
39. Brummitt RK, Nesta DP, Chang L, Kroetsch AM, Roberts CJ. Non-native aggregation of an IgG1 antibody in acidic conditions. II. Nucleation and growth kinetics with competing growth mechanisms. J Pharm Sci 2011;100:2104–2119.
40. Gebbink MF, Bouma B, Maas C, Bouma BN. Physiological responses to protein aggregates: fibrinolysis, coagulation and inflammation (new roles for old factors). FEBS Lett 2009;583:2691–2699.
41. van Beers MM, Sauerborn M, Gilli F, Brinks V, Schellekens H, Jiskoot W. Oxidized and aggregated recombinant human interferon beta is immunogenic in human interferon beta transgenic mice. Pharm Res 2011;28:2393–2402.
42. Wang X, Singh S, Kumar S. Potential aggregation-prone regions in complementarity-determining regions of antibodies and their contribution towards antigen recognition: a computational analysis. Pharm Res 2010;27:1512–1529.
43. Buck PM, Kumar S, Wang X, Agrawal NJ, Trout BL, Singh SK. Computational methods to predict therapeutic protein aggregation. Methods Mol Biol 2012;899:425–451.
44. Wang X, Das TK, Singh SK, Kumar S. Potential aggregation prone regions in biotherapeutics: a survey of commercial monoclonal antibodies. mAbs 2009;1:1–14.
45. Reumers J, Maurer-Stroh S, Schymkowitz J, Rousseau F. Protein sequences encode safeguards against aggregation. Hum Mutat 2009; 30:431–437.
46. Buck PM, Kumar S, Singh SK. Insights into the aggregation liabilities of a human Fab via high temperature molecular dynamic simulations. Protein Eng, in press.
47. Ma B, Nussinov R. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. Curr Opin Chem Biol 2006;10:445–452.
48. De Simone A, Dhulesia A, Soldi G, Vendruscolo M, Hsu ST, Chiti F, Dobson CM. Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. Proc Natl Acad Sci U S A 2011;108:21057–21062.
49. Neudecker P, Robustelli P, Cavalli A, Walsh P, Lundstrom P, Zarrine-Afsar A, Sharpe S, Vendruscolo M, Kay LE. Structure of an intermediate state in protein folding and aggregation. Science 2012; 336:362–366.
50. Wu C, Shea JE. Coarse-grained models for protein aggregation. Curr Opin Struct Biol 2011;21:209–220.
51. Cellmer T, Fawzi NL. Coarse-grained simulations of protein aggregation. Methods in molecular biology 2012;899:453–470.
52. Brummitt RK, Nesta DP, Roberts CJ. Predicting accelerated aggregation rates for monoclonal antibody formulations, and challenges for low-temperature predictions. J Pharm Sci 2011;100: 4234–4243.
53. Kayser V, Chennamsetty N, Voynov V, Helk B, Forrer K, Trout BL. Evaluation of a non-Arrhenius model for therapeutic monoclonal antibody aggregation. J Pharm Sci 2011;100:2526–2542.
54. Rodrigues JR, Simoes CJ, Silva CG, Brito RM. Potentially amyloidogenic conformational intermediates populate the unfolding landscape of transthyretin: insights from molecular dynamics simulations. Prot Sci 2010;19:202–219.
55. Chong SH, Lee C, Kang G, Park M, Ham S. Structural and thermodynamic investigations on the aggregation and folding of acylphosphatase by molecular dynamics simulations and solvation free energy analysis. J Am Chem Soc 2011;133:7075–7083.
56. Nowak M. Immunoglobulin kappa light chain and its amyloidogenic mutants: a molecular dynamics study. Proteins 2004;55:11–21.
57. De Simone A, Kitchen C, Kwan AH, Sunde M, Dobson CM, Frenkel D. Intrinsic disorder modulates protein self-assembly and aggregation. Proc Natl Acad Sci U S A 2012;109:6951–6956.
58. Bemporad F, De Simone A, Chiti F, Dobson CM. Characterizing intermolecular interactions that initiate native-like protein aggregation. Biophys J 2012;102:2595–2604.
59. Nayeem SM, Deep S. Rationalization of poor solubility of TGF-beta3 using MD simulation. Biochem Biophys Res Commun 2010; 401:544–547.
60. Moraitakis G, Goodfellow JM. Simulations of human lysozyme: probing the conformations triggering amyloidosis. Biophys J 2003; 84:2149–2158.
61. Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci U S A 2002;99:2754–2759.
62. Soldi G, Bemporad F, Chiti F. The degree of structural protection at the edge beta-strands determines the pathway of amyloid formation in globular proteins. J Am Chem Soc 2008;130:4295–4302.
63. Siepen JA, Radford SE, Westhead DR. Beta edge strands in protein structure prediction and aggregation. Prot Sci 2003;12:2348–2359.
64. Quintas A, Vaz DC, Cardoso I, Saraiva MJ, Brito RM. Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. J Biol Chem 2001; 276:27207–27213.
65. Gong R, Vu BK, Feng Y, Prieto DA, Dyba MA, Walsh JD, Prabakaran P, Veenstra TD, Tarasov SG, Ishima R, Dimitrov DS. Engineered human antibody constant domains with increased stability. J Biol Chem 2009;284:14203–14210.
66. Gong R, Wang Y, Feng Y, Zhao Q, Dimitrov DS. Shortened engineered human antibody CH2 domains: increased stability and binding to the human neonatal Fc receptor. J Biol Chem 2011;286: 27288–27293.
67. Price JL, Culyba EK, Chen W, Murray AN, Hanson SR, Wong CH, Powers ET, Kelly JW. N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability. Biopolymers 2012;98:195–211.