The Role of Amino Acid Sequence in the Self-Association of Therapeutic Monoclonal Antibodies: Insights from Coarse-Grained Modeling

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JANUARY 2013

Impact Factor: 3.3 · DOI: 10.1021/jp3108396 · Source: PubMed

6 AUTHORS, INCLUDING:

Anuj Chaudhri
Lawrence Berkeley National Laboratory
14 PUBLICATIONS **87** CITATIONS
SEE PROFILE

Isidro E Zarraga
Genentech
25 PUBLICATIONS **420** CITATIONS
SEE PROFILE

Sandeep Yadav
Genentech
21 PUBLICATIONS **503** CITATIONS
SEE PROFILE

Steven J Shire
Genentech
100 PUBLICATIONS **4,574** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

The Role of Amino Acid Sequence in the Self-Association of Therapeutic Monoclonal Antibodies: Insights from Coarse-Grained Modeling

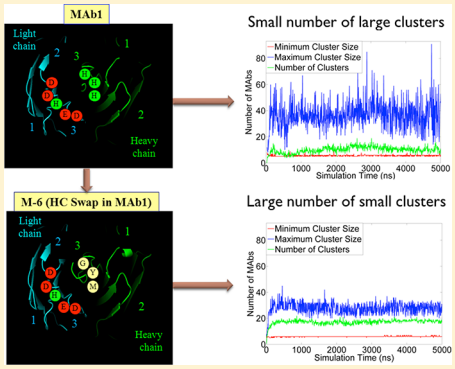
Anuj Chaudhri,[†] Isidro E. Zarraga,*[,‡] Sandeep Yadav,[‡] Thomas W. Patapoff,[§] Steven J. Shire,[‡] and Gregory A. Voth*[,†]

[†]Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics and Computation Institute, University of Chicago, Chicago Illinois 60637, United States

[‡]Late Stage Pharmaceutical Development, Genentech Inc., 1 DNA Way, S. San Francisco, California 94080, United States

[§]Early Stage Pharmaceutical Development, Genentech Inc., 1 DNA Way, S. San Francisco, California 94080, United States

ABSTRACT: Coarse-grained computational models of therapeutic monoclonal antibodies and their mutants can be used to understand the effect of domain-level charge−charge electrostatics on the self-association phenomena at high protein concentrations. The coarse-grained models are constructed for two antibodies at different coarse-grained resolutions by using six different concentrations. It is observed that a particular monoclonal antibody (hereafter referred to as MAb1) forms three-dimensional heterogeneous structures with dense regions or clusters compared to a different monoclonal antibody (hereafter referred to as MAb2) that forms homogeneous structures without clusters. The potential of mean force (PMF) and radial distribution functions (RDF) plots for the mutants (hereafter referred to as M1, M5, M7, and M10) show trends consistent with previously reported experimental observation of viscosities. The mutant referred to as M6 shows strongly attractive interactions that are consistent with previously reported negative second virial coefficients (B_{22}) obtained from light-scattering experiments (Yadav et al. *Pharm. Res.* **2011**, *28*, 1750−1764; Yadav et al. *Mol. Pharmaceutics.* **2012**, *9*, 791−802). Clustering data on MAb1 reveal a small number of large clusters that are hypothesized to be the reason for the high experimental viscosity. This is in contrast with M6 (that differs from MAb1 in only a few amino acids), where cluster analysis reveals the formation of a large number of smaller clusters that is hypothesized to be the reason for the observed lower viscosity. The coarse-grained representations are effective in picking up differences based on local charge distributions of domains to make predictions on the self-association characteristics of these protein solutions.



INTRODUCTION

Targeting diseases through monoclonal antibodies[1] has achieved considerable success with 29 U.S. Food and Drug Administration (FDA) approved MAbs in the market[2] and many more in the pipeline.[3] These monoclonal antibodies require high doses (>1 mg/kg) to be administered to a patient due to potency issues. While a subcutaneous (SC) route of administration is more convenient than the usual intravenous delivery, it poses an upper limit on the dosage volume that can be administered at a given time, typically <1.5 mL, and thus necessitates the development of formulations for SC administration. High concentration protein solutions pose formulation challenges such as high viscosity during manufacturing,[4,5] protein stability issues leading to association-based aggregation,[6,7] and degradation[8,9] and may also cause undesired immunogenic responses in the body.[10] Understanding the issues of self-association and aggregation can lead to more stable therapeutic drugs.

High concentration protein formulation development requires an understanding of protein−protein interactions (PPI)[11] and nonideal solution behavior. Protein−protein interactions include hydrogen bonding, excluded volume, electrostatic, hydrophobic, and van der Waals dispersion forces. A number of experimental studies have characterized the behavior of these high concentration solutions and analyzed the most important PPI that govern the behavior of concentrated antibody solutions. Liu et al.[12] studied the viscosity behavior of three monoclonal antibodies as a function of concentration, pH, and ionic strength. They found that one of the antibodies (MAb1, same as the one studied here) showed sharp viscosity changes with concentration and pH. MAb1 at 125 mg/mL was found to be 60-fold more viscous than the solution without protein. The high viscosity of MAb1 was attributed to the reversible self-association behavior of these protein molecules. Based on the decreasing viscosity with addition of salt, electrostatic charge−charge interactions were hypothesized to be the most important PPI in solution. In contrast to MAb1, MAb2 (same as the one studied here)

Received: November 2, 2012
Revised: January 8, 2013

   dx.doi.org/10.1021/jp3108396 | *J. Phys. Chem. B* XXXX, XXX, XXX−XXX

viscosity as a function of concentration could be described by a modified version of the extended Mooney equation,[12] which only takes into account excluded volume effects. Kanai et al. performed a series of titration studies on the MAb1 solutions and concluded that the Fab-Fab interaction between MAbs was the key contributor to the formation of an organized multivalent network.[13]

Yadav et al.[14−16] have done extensive experimental studies on different mAb solutions at different pH, ionic strengths, and concentrations using dynamic light scattering to probe interactions between the MAbs. The intermolecular interactions were characterized using interaction parameter k_D and second virial coefficient B_{22} determined by using dynamic and static light scattering, respectively. The interaction parameter k_D approaches a minimum suggesting a relative increase in attraction between molecules at pH 6.0. Ultrasonic shear rheometry experiments were used to measure the solution storage modulus, which exhibits a sharp increase above 80 mg/mL at pH 6.0. The storage modulus peaked at pH 6.0 for high concentrations indicating strong intermolecular interactions that conferred rigidity to the solution at pH 6.0. On comparing different MAbs, Yadav et al. observed that the viscosity profiles for the IgG1MAbs cannot be explained by electroviscous effects or by effective molecular volume.[15] The observed viscosity behavior was attributed to short-range attractive and repulsive potentials between specific domains of the MAbs, which may result in a self-associating network resulting in large viscosity changes.

Additionally, Yadav et al.[17,18] studied the effect of amino acid sequence on the self-association and viscosity characteristics of high concentration protein formulations by performing mutations in the variable light (VL) and variable heavy (VH) chains of MAb1 and MAb2. The mutants were designed so that the sequence variation between the two MAbs can be minimized. The study attempted to determine the importance of specific residues in the CDR region responsible for self-association and viscosity properties of MAb1. Swapping charged residues out resulted in loss of viscosity for mutants M5, M6, and M7 (defined in Methods section). Mutations only in the VL domain (M5), or both in the VL and VH domains (M7) of MAb1, decreased the charge asymmetry of these MAbs resulting in loss of self-associating behavior. In the case of M6 where mutations are done in the VH domain, the net change in attractive interactions is only minimal whereas the viscosity decrease is significant. Replacement of specific residues[18] or aglycosylation[17] outside the CDR did not show any change in viscosity behavior. However, not all possibilities of changes were tested and other changes could lead to differences in behavior as well. The change in the electrostatic surface potential following mutation correlated fairly well with the self-associating and viscosity behavior of the MAbs. Theoretical and computational models of the MAbs can thus play a vital role in testing the hypothesis that the charge distribution affects the self-associating behavior more than the net charge on the protein.[19]

Numerical simulations using molecular dynamics (MD)[20,21] are useful in small liquid systems but advancements in computational power allow simulations of more complex systems. Enhanced computational power has allowed for simulations of very large biomolecular systems having a million atoms at a considerable increase in computational speed.[22] However, issues still remain as most of the processes that are interesting in nature span over multiple length and time scales and often traditional MD cannot treat this disparity in time and length scales. For example, one mAb in a solvated environment requires running simulations for over 300 000 atoms. Simulating this system by using all-atom MD is intractable. Hence alternative techniques may help understand processes at greater length and longer time scales. Coarse-grained (CG) and structurally reduced models of biosystems help bridge the long time and length scale gap and make biological processes more accessible.[23−25] These CG methodologies have been applied to a large number of systems and have shed light on many fundamental processes (see a recent book[26] and the examples within).

Computational studies on MAb1 and MAb2 with several different CG models have been performed in a previous study.[27] CG models were developed for both rigid and flexible antibodies by including intramolecular constraints. The simulations corroborated the role of Fab-Fab interactions in MAb1 but also pointed to the equally dominant Fab-Fc interactions. MAb2, on the other hand, interacts primarily via Fab-Fc interactions. Fc-Fc interactions play the least significant role in self-association. The study also showed for the first time the formation of dense clusters in MAb1 at 120 mg/mL compared to MAb2 where no such clusters were present. Six different clusters of varying sizes were found all over the simulation domain in MAb1. It was concluded that the dense clusters found in MAb1 might be responsible for the high viscosity found in experiments of high concentration MAb1 solutions. The homogeneous distribution of the MAbs for the MAb2 solution might be responsible for the lower viscosity at high concentrations compared to MAb1. In addition, the charge distribution at the domain level plays a vital role in determining the equilibrium structures formed by the different MAbs. At the present level and resolution, the model captured appreciable differences in network arrangements between MAb1 and MAb2 due to specific sequence differences in CDR regions, which is consistent with experimental viscosity observations.

In this paper, domain-level charge−charge electrostatics is studied to test the self-association of a number of variants of MAb1 and MAb2. Experiments have pointed to electrostatics being important in forming networks and leading to differences in solution properties for these monoclonal antibodies. To test this hypothesis, as a first approximation, a novel simulation methodology has been developed to understand the highly complex multiscale problem of self-association. Two reduced CG models (12-site and 26-site) of these antibodies have been developed and CG molecular dynamics (CGMD) simulations were performed as a function of concentration. The systems were tested at a pH of 6.0 to minimize chemical degradations. Specifically the hypothesis of enhanced electrostatic interactions is tested in these simulations. The following sections of this article are organized as follows. In the Methods section, the techniques developed and used to construct the reduced mAb model are explained, utilizing a normal mode elastic network analysis of a proposed solution structure based on MD simulations.[28] The Results section presents the CGMD simulations performed on the CG models of the antibodies showing the effects of the CG resolution and geometry on the association process. The results are assessed further in the Discussion section, where the differences in structures formed by the antibodies are clearly shown and corroborated with experimental data wherever possible.

METHODS

Description of Monoclonal Antibodies and Charge-Swap Mutants. The general approach to identify and construct

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

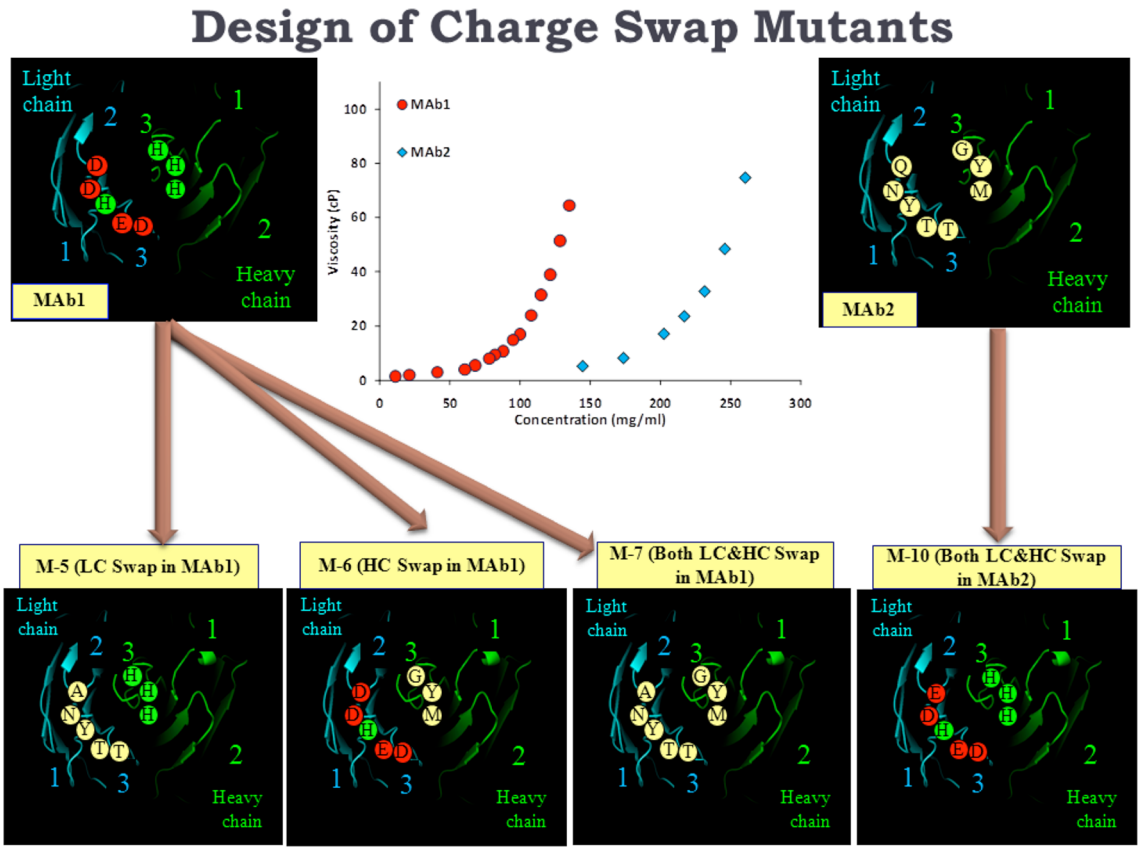


Figure 1. Schematic of charge-swap design strategy. Mutants M5, M6, and M10 were designed to make specific changes to the CDR regions in MAb1. Mutant M10 incorporates changes to the CDR regions of MAb2. The graph shows the variation of experimental viscosities with protein concentration for MAb1 and MAb2. MAb1 shows dramatic increases in viscosity at concentrations greater than 100 mg/mL whereas MAb2 shows gradual changes up to 250 mg/mL.

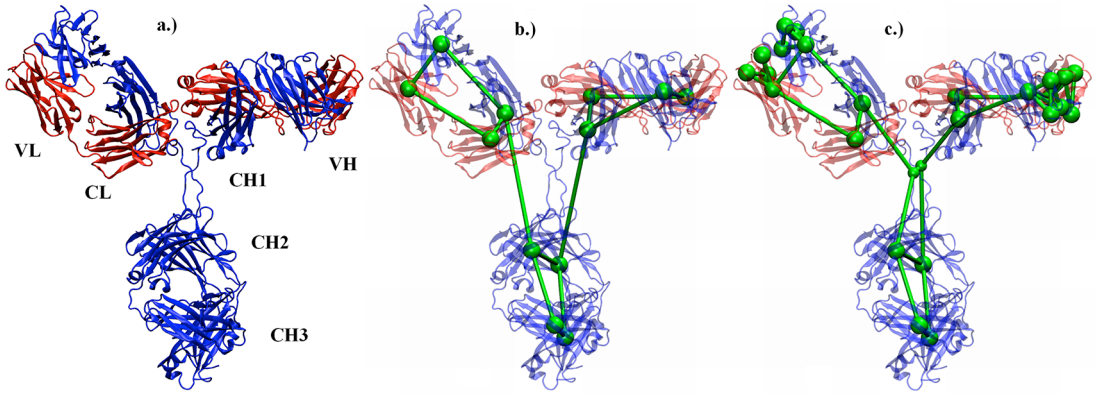


Figure 2. Monoclonal antibody structure and CG models. (a) The 3D structure of the IgG1 monoclonal antibody.[28] The antibody contains two identical heavy and light chains. The heavy chain contains four domains: VH, CH1, CH2, and CH3. The light chain contains two domains: VL and CL. (b) The 12-site CG model overlaid on the 3D structure of the antibody. (c) The 26-site CG model overlaid on the 3D structure of the antibody. The figures were rendered by using VMD.[47]

CG representations requires the identification of CG site positions within each protein domain and the CG effective potential parameters. The strategies for the two tasks are discussed below. The mAb 3D structure framework was based on an MD model for MAb2 that is the most likely conformation in solution.[28] The MAb1 representation uses the same framework. It is important to note that it is extremely difficult to obtain experimentally the crystal structure of a full length mAb due to poor crystallization or extreme flexibility of the hinge region (see ref 28 and references therein). In their study, Brandt et al.[28] pieced together a full structure of the IgG1 antibody from crystal structures of the fragments and equilibrated it in-silico to a relaxed conformation. This relaxed structure was used to construct the representative CG models.

To generate the structure of the mutants, homology models of MAb1 and MAb2 were constructed followed by energy minimization to avoid steric clashes.[18] The aim of designing the mutants was to determine the impact of surface charge distribution in MAb1 on self-association and viscosity. This was accomplished by swapping charged residues in the CDR of

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

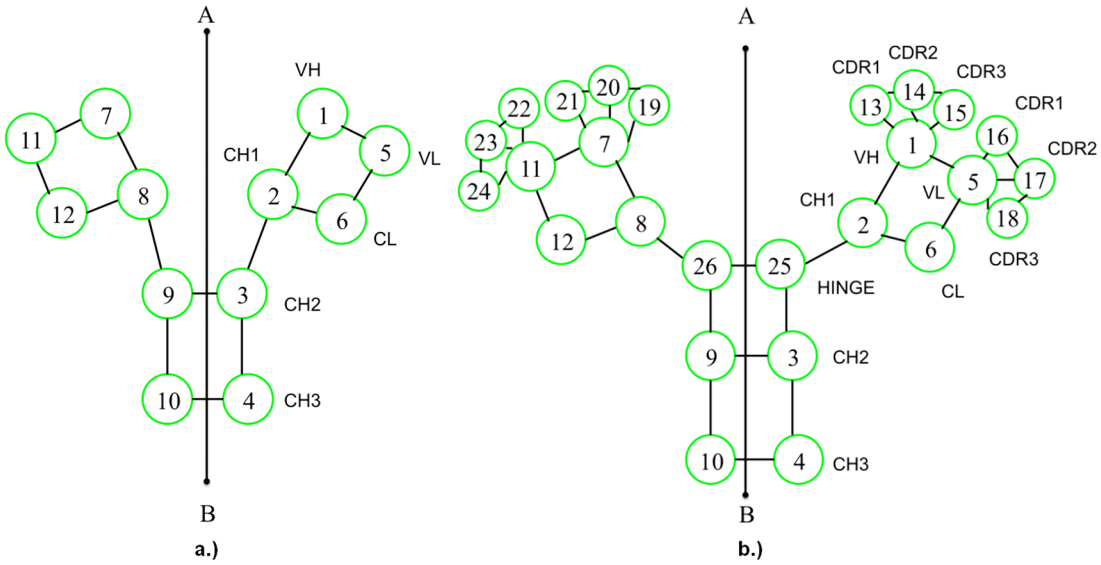


Figure 3. CG representations of the monoclonal antibody. The schematic shows all the sites with numbers used to identify the different regions of the mAb. The antibody is symmetric with respect to the heavy and light chains through the plane AB that divides the antibody in half. (a) The 12-site lower resolution model and (b) the 26-site higher resolution model.

MAb1 with the corresponding noncharged residues in the CDR of MAb2. Thus mutants M5, M6, and M7 were designed by varying the residues in the VL and VH regions of MAb1 as shown in Figure 1. The mutants were designed to test the role of specific interactions that arise from the residues in the CDR loops. The mutant M10 was designed by swapping the charged residues in the CDR of MAb1 into the corresponding positions in the CDR of MAb2 and seeing the impact of these changes on MAb2 self-association and viscosity. The mutant M1 was created to assess the impact on viscosity of differences in length of the CDR between MAb1 and MAb2. Since the CDR of MAb1 has four additional residues in the first loop of the light chain compared to MAb2 the four additional residues were deleted. More details about the description of these mutants can be obtained from Yadav et al.[17,18]

Identification of CG Sites. Two kinds of mAb CG models were developed—one with 12 sites and another with 26 sites as shown in Figure 2. The 12-site model was developed by performing an elastic network normal-mode analysis[29] that places the CG sites based on the long wavelength collective motion of the protein domains that play an important role in their biological function.[30,31] In the elastic network model of the protein the low-frequency modes of collective dynamics are calculated that are then used to define the dynamically correlated domains and the CG representation. By using this procedure, it was found that each domain of the antibody moved as a whole unit in the low-frequency limit. Hence the center of mass of each domain was chosen for the placement of the corresponding CG site. The sites were numbered from 1 to 12 as shown in Figure 3. The 26-site model uses the same structure as the 12-site model except that extra sites are added to the CDR regions and a CG site is added to the center of mass of the hinge region as shown in Figure 3. The total mass and charge of each CG site was calculated by summing up the masses and partial charges respectively of the underlying residues that the site represents. The partial charges at each site were based on the CHARMM force field.[32] It was assumed that each Histidine residue contributes a partial charge of +0.5 units at pH 6.0. The masses and charges for each domain at pH 6 are shown in Table 1. In the 12-site model, the mass and charge of the hinge region are divided equally between sites 2 and 3. In the 26-site model, sites 25 and 26 represent the underlying hinge residues.

CG Force Field. The CG variables represent collective degrees of freedom of multiple atoms in a protein domain where the interactions between the CG sites have to be chosen to represent large-scale protein motion in an averaged, effective manner. An effective potential can be written as a sum of intraprotein and interprotein interactions as follows:

$$U_{\text{total}} = U_{\text{intra}} + U_{\text{inter}} \tag{1}$$

The bond, angle and dihedral potential functions constitute the intraprotein interactions whereas the electrostatic and dispersion forces constitute the interprotein interactions. The effect of the topology of the mAb and electrostatics on self-association is by itself a very important problem to study in protein solutions.[33] The effect of the internal degrees of freedom on the self-association characteristics has been tested separately[27] by including the intramolecular interactions to the already developed CG models using atomistic trajectory information from MD simulations. It was found that the intramolecular interactions did not change the overall equilibrium distribution of the mAb solutions significantly. Therefore, in this study, only the rigid, compact Y model was used.

Interprotein Interactions. The screened electrostatics and van der Waals interactions are assumed to be independent of each other and additive as in the Derjaguin−Landau−Verwey−Overbreek theory of colloidal particle interactions.[34] The electrostatic interactions were modeled by the Coulomb potential with Debye−Hückel screened electrostatics[34] having an exponential Yukawa term.

$$U_{\text{Coulomb}} = \frac{q_i q_j}{4\pi\varepsilon_r r} \exp(-\kappa r) \tag{2}$$

Here, q_i and q_j represent the net charges on the CG sites, ε_r is the effective dielectric constant, and κ is the Debye screening parameter. Calculating the effective dielectric constant of protein solutions has been the subject of many investigations.[35,36] However, no consensus has been established on the best way to

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 1. Masses and Charges Used in the CG Simulations[a]

(a) 12-Site Model							
site no.	MAb1	M1	M5	M6	M7	MAb2	M10
				mass (amu)			
1, 7	13506	13506	13506	13446	13446	13536	13596
2, 8	10854.5	10854.5	10854.5	10854.5	10854.5	11027.5	11027.5
3, 9	12552.5	12552.5	12552.5	12552.5	12552.5	12552.5	12552.5
4, 10	12047	12047	12047	12047	12047	12047	12047
5, 11	12389	11925	12332	12389	12332	12277	12291
6, 12	11398	11398	11398	11398	11398	11162	11162
				charge at pH 6			
1, 7	+3.5	+3.5	+3.5	+2.0	+2.0	+3.5	+5.0
2, 8	+2.75	+2.75	+2.75	+2.75	+2.75	+3.75	+3.75
3, 9	+3.25	+3.25	+3.25	+3.25	+3.25	+3.25	+3.25
4, 10	−0.5	−0.5	−0.5	−0.5	−0.5	−0.5	−0.5
5, 11	−1.5	−0.5	+2.0	−1.5	+2.0	+4.5	+2.0
6, 12	−2	−2	−2	−2	−2	−2	−2
(b) 26-Site Model							
site no.	MAb1	M1	M5	M6	M7	MAb2	M10
				mass (amu)			
1, 7	8936	8936	8936	8936	8936	8882	8882
2, 8	9997	9997	9997	9997	9997	10170	10170
3, 9	11695	11695	11695	11695	11695	11695	11695
4, 10	12047	12047	12047	12047	12047	12047	12047
5, 11	8997	8997	8997	8997	8997	9455	9455
6, 12	11398	11398	11398	11398	11398	11162	11162
13, 19	1159	1159	1159	1159	1159	1132	1132
14, 20	1756	1756	1756	1756	1756	2016	2016
15, 21	1655	1655	1655	1595	1595	1506	1566
16, 22	1688	1224	1646	1688	1646	1113	1112
17, 23	721	721	721	721	721	756	756
18, 24	983	983	968	983	968	953	968
25, 26	1715	1715	1715	1715	1715	1715	1715
				charge at pH 6			
1, 7	+1	+1	+1	+1	+1	+2	+2
2, 8	+3	+3	+3	+3	+3	+4	+4
3, 9	+3.5	+3.5	+3.5	+3.5	+3.5	+3.5	+3.5
4, 10	−0.5	−0.5	−0.5	−0.5	−0.5	−0.5	−0.5
5, 11	+3	+3	+3	+3	+3	+4	+4
6, 12	−2	−2	−2	−2	−2	−2	−2
13, 19	0	0	0	0	0	+0.5	+0.5
14, 20	0	0	0	0	0	+2	+2
15, 21	+2.5	+2.5	+2.5	+1	+1	−1	+0.5
16, 22	−2	−1	0	−2	0	0	−1
17, 23	−1	−1	−1	−1	−1	0	0
18, 24	−1.5	−1.5	0	−1.5	0	+0.5	−1
25, 26	−0.5	−0.5	−0.5	−0.5	−0.5	−0.5	−0.5

[a]The masses and charges were calculated as a cumulative sum of all the underlying residues in each CG site. The charges at pH 6 assume HIS to be 50% protonated and have a net positive charge of +0.5.

calculate the effective dielectric constant within and between two protein molecules especially for implicit solvent models.[37,38] In this study, a screened implicit solvent model with an effective dielectric constant of 1 is assumed. A value of 2.5 nm is used for the Debye screening length, which is the inverse of the Debye screening parameter κ. The value was picked based on experimental investigations, using 15 mM salt solutions.[14,15] The dispersion and repulsive forces due to excluded volume were modeled by using a Lennard-Jones potential with a force-smoothing function.

$$U_{LJ} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right];\ r < r_{in} \tag{3}$$

$$F = C_1 + C_2(r - r_{in}) + C_3(r - r_{in})^2 + C_4(r - r_{in})^3;\ r_{in} < r < r_c \tag{4}$$

where ε_{ij} is the well depth for the ijth pair of CG sites, and σ_{ij} is the radius at which LJ potential is exactly zero. In eq 4, r_{in} is the inner cutoff radius and r_c is the outer cutoff radius for pairs of CG site interactions. The polynomial coefficients C_1, C_2, C_3, and C_4 are computed on the fly to smoothly vary the force from the inner

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

cutoff to the outer cutoff.[39] The σ_{ij} value for the LJ potential (shown in Table 2) was chosen based on the average size of each domain. The well depth is chosen to be much weaker than the strength of the hydrogen bond.

Table 2. Pair Interaction Potential Parameters for the CG Models Used in the Simulations

simulation parameter	value
ε_r	1.0
κ	0.04 Å^{-1}
ε_{ij}	1.0 kcal/mol
σ_{ij}	20.0 Å
r_{in}	100 Å
r_c	200 Å
time step	1 ps for rigid MAbs
Langevin bath damping parameter	5 ps
Langevin bath temp	300 K
total no. of CG time steps	5 000 000 for rigid MAbs

Simulation Details. The CG sites and the force field described in the above sections were used to perform CG Langevin dynamics simulations of the 12- and 26-site models with use of the LAMMPS package.[39] The simulation parameters for the models are listed in Table 2. Initially 1000 mAb molecules were arranged in a cubic lattice and the system was allowed to run for 5 μs until it reaches equilibrium. The final configuration from the equilibrated trajectory was then heated to an elevated temperature of 1000 K to ensure that all the MAbs have random orientations and are not stuck in a metastable state and subsequently used as the initial condition for all the simulations. Periodic boundary conditions were applied in all three directions. The CG simulations were performed under NVT conditions with use of a Langevin thermostat with the temperature set to 300 K. The CG simulations for the rigid antibodies were run for 5 million iterations, using a time step of 1 ps. This corresponds to 5 μs of CGMD simulation time with the first 1 million iterations chosen as the equilibration time. The remaining 4 million iterations were chosen to calculate all of the properties and time-averaged data. To understand the distribution of MAbs after equilibrium has been reached in the system, the domain was divided into cubic boxes of side 100 Å at each snapshot of 5 ns over 4 μs of simulation time. The number of mAb center of masses is counted within each cube and the cubes are colored accordingly. Cubes with one mAb center are called hereafter monomers, cubes with two mAb centers dimers, and cubes with three mAb centers trimers. It is important to note that dimers and trimers here refer to mAb centers that are "self-associated" spatially within the boxes due to electrostatic interactions rather than chemically stable structures.

RESULTS

The CG models described above were used to study the effect of the charge distribution on the self-association of antibodies under different conditions. Langevin dynamics simulations were performed on these MAbs to obtain equilibrium structures at the end of the simulations. The results from the simulation studies are discussed below.

Equilibrium Structure Distribution. The average numbers of monomeric, dimeric, and trimeric structures using both 12- and 26-site rigid, compact Y models are shown in Table 3. As seen in Table 3, the number of dimers in MAb1 is greater than that in MAb2 for both 12- and 26-site models. This was confirmed in a previous study.[27] For the 12-site model, the number of dimers for all the systems varies and is ranked as M6 > MAb1 > M1 > M7 > M5 > MAb2 > M10. For the 26-site model, the number of dimers rank as M6 > M7 > MAb1 > M5 > M1 > M10 > MAb2. It is interesting to note that for the 12-site model, except for M6, the number of dimers follows the same trend as the viscosity differences between the systems. Within standard error listed in parentheses in Table 3, the number of dimers in the 26-site model also follows the same trends as the 12-site model.

Table 3. Average Number of Monomeric, Dimeric, and Trimeric Self-Associated Structures Calculated from Equilibriated Structures at 120 mg/mL[a]

	no. of monomers	no. of dimers	no. of trimers
12-site rigid compact Y model			
MAb1	800(10)	97(5)	2(1)
M1	864(12)	68(6)	0
M5	946(9)	27(4)	0
M6	675(12)	152(6)	7(2)
M7	928(12)	36(6)	0
MAb2	972(7)	14(3)	0
M10	978(5)	11(3)	0
26 site rigid compact Y model			
MAb1	894(9)	53(4)	0
M1	918(8)	41(4)	0
M5	912(10)	44(5)	0
M6	829(9)	84(5)	1(1)
M7	886(12)	57(6)	0
MAb2	960(9)	20(4)	0
M10	959(7)	21(4)	0

[a]The standard deviation from the mean is shown in parentheses.

Potential of Mean Force. To understand the effect of higher concentrations on the association characteristics of the MAbs, the radial distribution function (RDF) was calculated by using the equilibrium center of mass (COM) distribution. The RDF was then used to define a potential of mean force (PMF in kcal/mol) by using the equation:

$$\mathrm{PMF} = -k_B T \ln[g(r)] \tag{5}$$

The PMF for MAb1 and MAb2 at pH 6 for both 12- and 26-site models is plotted in Figure 4 at 120 mg/mL. The M6 systems have more short-range attraction as evidenced by the sharp minima in the PMF shown in Figure 4. In both the 12- and 26-site models, M6 is the most attractive among the MAbs with significant short-ranged structure at 60 and 100 Å. This is followed by MAb1 that shows an attractive well close to 100 Å. All other MAbs have attractive potential wells between 130 and 150 Å. MAb2 and M10 are almost indistinguishable as seen in Figure 4. For the 26-site model, the attractive wells are shifted by 20 due to additional sites for the CDR loops that help move the MAbs slightly away from each other.

Density-Based Clustering. To further understand the behavior of the MAbs at higher concentrations, clustering calculations were done on the mAb simulations at 120 mg/mL with use of DBSCAN, a density-based clustering algorithm.[40,41] The clusters were defined based on the criteria that a minimum of 4 mAb centers are found within a neighborhood radius of 100 Å. The number of clusters found in MAb1 and M6 at snapshots of 5 ns is plotted in Figures 5 and 7 as a function of simulation time. The minimum and maximum cluster size is also plotted in these figures. It is interesting to note that an average of 10 clusters is found in MAb1 at any given point of time. The clusters are

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

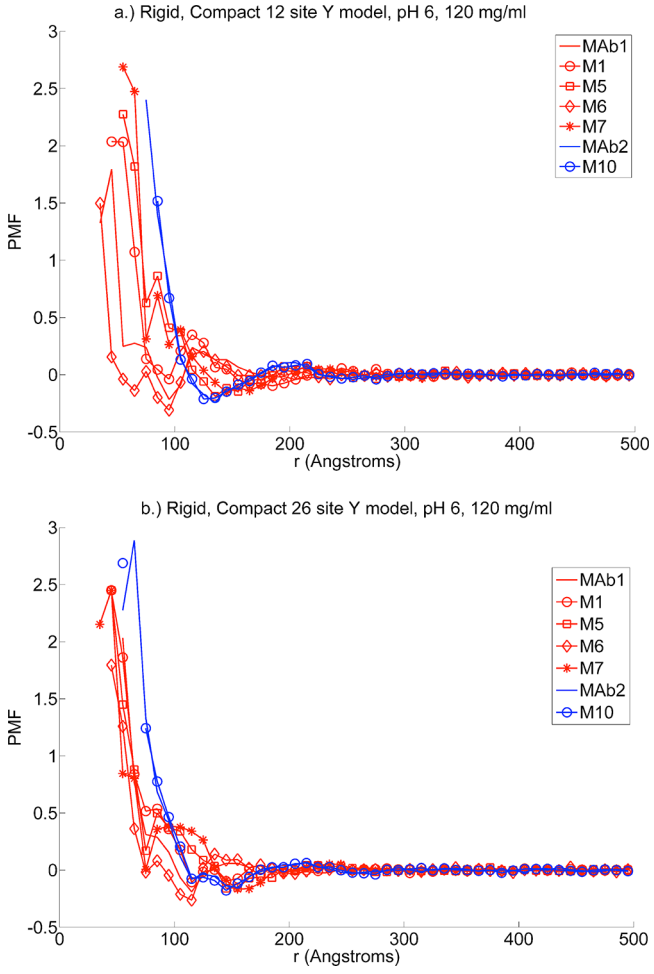


Figure 4. Potential of mean force (PMF) plots (in kcal/mol) for the monoclonal antibodies MAb1 and MAb2 and their mutants. Panel a shows results for the compact Y-shaped 12-site rigid model at pH 6 and panel b shows the results for the compact Y-shaped 26-site rigid model. The PMF is calculated by using eq 5 from the equilibrium center of mass (COM) distribution of the MAbs.

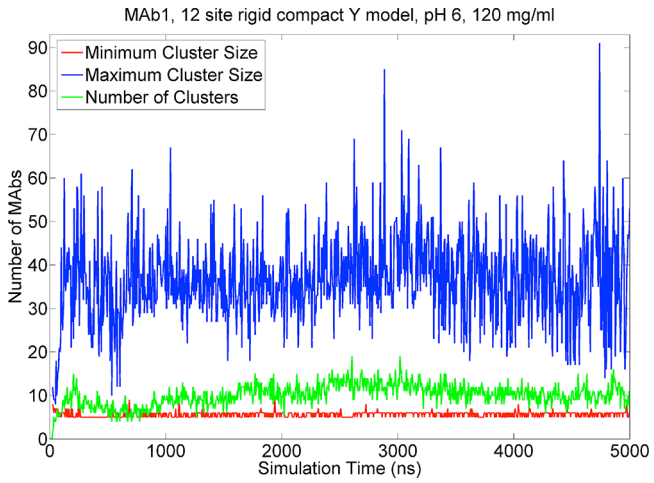


Figure 5. Number of clusters of MAb1 at 120 mg/mL as a function of simulation time. The minimum and maximum cluster sizes at each step are also shown. On average, 15 clusters exist during the simulation. The clusters range in average sizes from 5 to 45. The maximum size of any cluster during the simulation is 91 MAbs.

dynamic in nature, as seen in Figures 6 and 8, where the snapshots are plotted every 625 ns of CG simulation time because the MAbs are associated spatially and not via chemical bonds. Since a mAb can enter and leave the cluster or merge with another cluster at any point of time, there appears to be large fluctuations in the maximum and minimum cluster sizes. The average cluster sizes can range from 5 to 45 in MAb1. The same algorithm gave no evidence of clusters in MAb2, M1, M5, M7, and M10. For M6, the algorithm returned a larger number of clusters but smaller in size all over the domain as seen in Figures 7 and 8. The average number of clusters in M6 is close to 18 with average cluster sizes in the range of 5 to 25.

Intersite Radial Distribution Functions. Radial distribution functions (RDFs) for site−site interactions were calculated to understand the most prominent interactions between the MAbs. The CG site numbers are given in Figure 3. The site−site interactions can be broken down into Fab-Fab, Fc-Fc, and Fab-Fc. The Fab-Fab interactions between the MAbs were calculated for site numbers 1−1, 7−7, 5−5, 11−11, 1−5, 7−11, 1−11, 1−7, 5−11, and 5−7. The Fab-Fc interactions between the MAbs were calculated for site numbers 4−1, 4−7, 4−5, 4−11, 10−1, 10−7, 10−5, and 10−11. The Fc-Fc interactions were calculated for sites 4−4, 10−10, and 4−10. To compare the strength of the different interactions, the area under the RDF plots was calculated for all the systems at 120 mg/mL by using numerical quadrature. The data for all the models are reported in Table 4. It is interesting to note that both Fab-Fab interactions are equally dominant in MAb1 and M6 compared to other MAbs. The importance of Fab-Fab interactions in MAb1 at high concentrations has also been confirmed experimentally.[13] For both the 12- and 26-site models, the relative strength of Fab-Fab interactions (compared to other interactions) can be arranged as MAb1 ∼ M6 > M1 > M7 > M5 ∼ M10 > MAb2. Again, except for M6, the strength for Fab-Fab type interactions is the same as the number of dimers formed as shown in Table 3 and the viscosity trends found in experiments. The Fab-Fc interactions in the mAb systems can be arranged as MAb2 ∼ M10 > M5 > M7 > M1 > M6 > MAb1. The Fab-Fc interactions are equally important in these mAb systems and seem to follow the opposite trend to all the other trends listed above. The trends found by the RDF calculations corroborate the fact that Fab-Fab interactions are consistent with high viscosity trends found experimentally and Fab-Fc interactions are consistent with more homogeneous fluid mAb structures.

DISCUSSION

The present CG simulations suggest that domain-level charge−charge electrostatics can help differentiate between the equilibrium structures formed by the different antibodies studied here. Both the high-resolution and low-resolution CG models of the MAbs were able to predict the differences in self-association characteristics between the charge-swap mutants. Experiments have shown that the MAbs show very different viscosity behavior with an increase in concentration. The MAb1 molecules show sharp viscosity changes as a function of increasing concentration, whereas the MAb2 systems show more gradual changes in viscosity with increase in concentration.[12] The viscosity behavior of the charge-swap mutants lies in between MAb1 and MAb2.[17,18] According to the CG simulations, these differences can be attributed to the kind of self-associating structures that are formed with an increase in concentration as confirmed in previous studies.[27] Although equilibrium structures are not

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

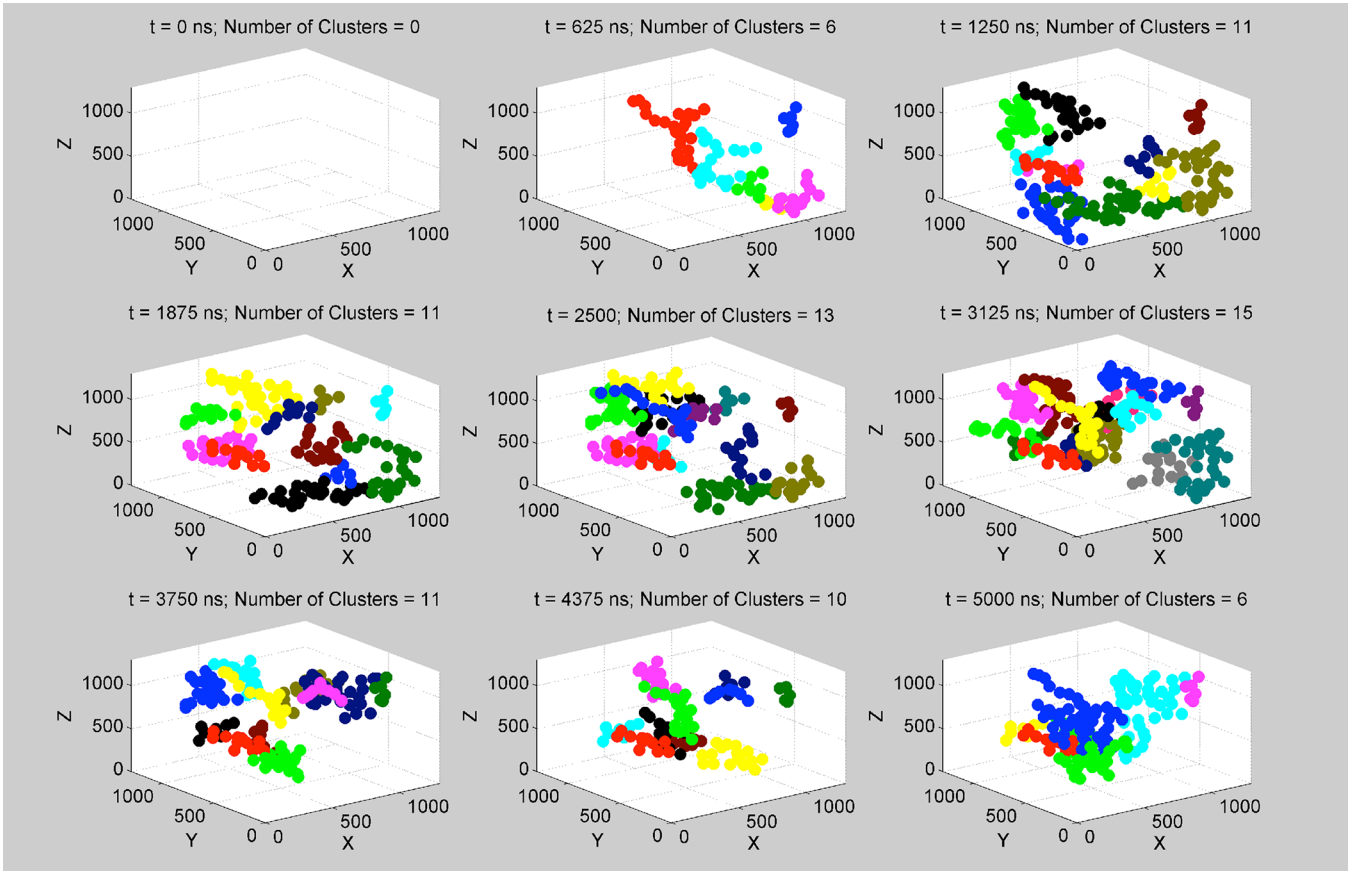


Figure 6. Cluster formation in MAb1 at 120 mg/mL. The plot shows the number of clusters at snapshots of 625 ns starting initially with no clusters at t = 0 up to the end of the simulation at t = 5000 ns. MAb1 shows fewer clusters containing a large number of MAbs that dynamically break and form as the simulation progresses.

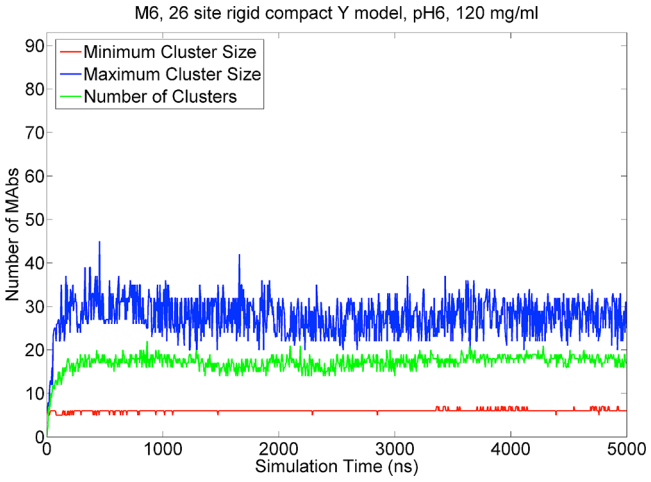


Figure 7. Number of clusters of M6 at 120 mg/mL as a function of simulation time. The minimum and maximum cluster size at each step is also shown. On average, 18 clusters exist during the simulation. The clusters range in average sizes from 5 to 25. The maximum size of any cluster during the simulation is 45 MAbs.

enough to predict rheological behavior of these solutions, some inferences can be made.

On the basis of the equilibrium distribution and the number of dimers formed in the system, the antibodies follow the same trend as the viscosity behavior found in experiments.[17,18] The relative number of dimers found in each mutant can be ranked as M6 > MAb1 > M1 > M7 > M5 > MAb2 > M10. The equilibrium behavior of the mutants M1, M7, M5, and M10 follows the expected trends. The only exception to this is the mutant M6 that deviates from the observed trend. From Table 1, it is interesting to note that there is a net decrease in negative charge in mutants M1, M5, and M7 compared to MAb1. In M1, the net change in charge is +1 at sites 5, 11 in the 12-site model and 16, 22 in the 26-site model. In M5, the change in net charge is +3.5 at sites 5, 11 in the 12-site model and 16, 22, 18, 24 in the 26-site model. For M7, the net change in charge is +2.0 at sites 5, 11, 1, 7 in the 12-site model and sites 15, 21, 16, 22, 18, 24 in the 26-site model. In the case of M10, there is a net change in charge of +4.5 at sites 1, 7, 5, 11 in the 12 site-model and 15, 21, 16, 22, 18, 24 in the 26-site model. Since Fab-Fab interactions are primarily responsible for the self-association characteristics of the antibodies,[13−15,27] the net reduction of attractive charges in the Fab region of the above antibodies should lead to a decrease in self-association and reduction in viscosity at high concentrations. This trend has been confirmed by the CG simulations on these mutants and further helps to strengthen the argument.

The only exception to this observed trend is mutant M6 that shows a higher than expected number of dimers due to a net reduction in repulsive charge of 1.5 units, though its viscosity vs concentration profile is in between that of MAb1 and the other mutants. The net reduction in repulsions in sites 1, 7 in the 12-site model and 15, 21 in the 26-site model should indeed lead to an increase in the net increase in Fab-Fab interactions and hence lead to greater self-association. This is well predicted by both the CG models as seen in Figure 4, where the PMF plots show a

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

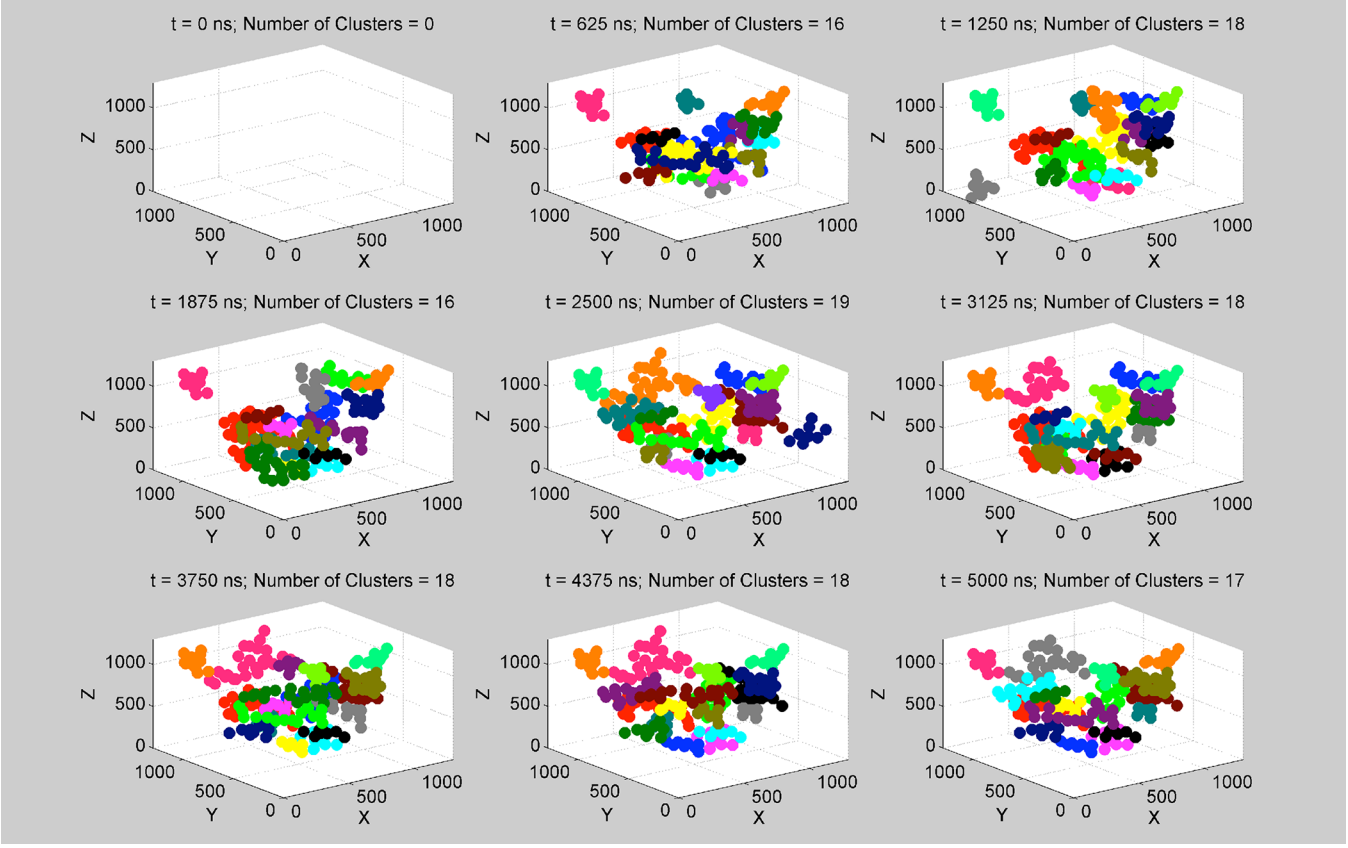


Figure 8. Cluster formation in M6 at 120 mg/mL. The plot shows the number of clusters at snapshots of 625 ns starting initially with no clusters at $t = 0$ up to the end of the simulation at $t = 5000$ ns. M6 shows a large number of smaller clusters that seem to be more compact in size. The number and placement of clusters does not vary significantly as the simulation progresses indicating a stronger net attraction between the MAbs.

Table 4. Quantitative Assessment of the Fab-Fab, Fab-Fc, and Fc-Fc Interactions in the CG Simulations from the Site−Site Radial Distribution Functions at 120 mg/mL

	MAb1	M1	M5	M6	M7	M10	MAb2
	Fab-Fab/(Fab-Fab + Fab-Fc + Fc-Fc)						
12-site, rigid compact Y	0.46	0.39	0.34	0.46	0.36	0.34	0.33
26-site, rigid compact Y	0.47	0.44	0.37	0.47	0.38	0.37	0.36
	Fab-Fc/(Fab-Fab + Fab-Fc + Fc-Fc)						
12-site, rigid compact Y	0.41	0.45	0.49	0.43	0.46	0.5	0.51
26-site, rigid compact Y	0.4	0.39	0.45	0.41	0.44	0.46	0.46
	Fc-Fc/(Fab-Fab + Fab-Fc + Fc-Fc)						
12-site, rigid compact Y	0.13	0.17	0.17	0.11	0.18	0.17	0.16
26-site, rigid compact Y	0.13	0.17	0.18	0.12	0.18	0.18	0.18

deeper attractive well for M6 compared to MAb1. There is also more short-range structure in M6 compared to the other MAbs. However, the experimental viscosity behavior is quite different than what the trends in the number of dimers and PMF plots suggest. Therefore, stronger short-range attractive interactions may not lead to a higher degree of self-association and thus to a higher viscosity. Indeed viscosity behavior at high concentrations will also depend on the type of clusters that can be found in the system. For example, in attractive colloids, the transition from fluid to solid-like behavior is manifested by an increase in viscosity resulting from the formation of larger clusters or network.[42,43] Hence, the smaller number of clusters found in the domain could lead to less than expected viscosities. Comparing Figures 5 and 7, it is interesting to see that M6 forms a large number of smaller clusters all across the domain compared to MAb1 where a small number of large clusters are found. It is important to remember that the clusters are dynamic in nature, since the network formation is dynamic and these molecules are self-associated spatially. The increase in attractive interactions in M6 forms smaller, more compact clusters that are difficult to break and hence lead to a deeper attractive potential well and a large number of dimers. However, the more compact M6 clusters have a lower hydrated volume per protein than the more extended MAb1 clusters, leading to both lower intrinsic viscosities and a smaller effective volume fraction per protein.[44] This smaller effective excluded volume could possibly explain the significantly lower M6 viscosities compared to MAb1 especially at high concentrations where packing effects are important. The largest cluster that was formed in M6 was half the size of the largest cluster formed in MAb1 (Figure 7). On average, M6 forms clusters that range in size from 5 to 25 MAbs compared to MAb1 where cluster sizes range from 5 to 45 MAbs. Hence, the CG simulations are helpful in predicting trends found in mutants and corroborate qualitatively to experimental results.

Since viscosity is typically measured under flow conditions, an accurate description of how the mAb network resists deformation or disruption under shear requires nonequilibrium shear flow computations performed as a function of concentration, which will be the subject of future studies. In addition, a

dx.doi.org/10.1021/jp3108396 | J. Phys. Chem. B XXXX, XXX, XXX−XXX

more complex model may be needed to describe behavior at very high concentrations (>100 mg/mL). Factors such as near-field hydrodynamics, hydrogen bond networks,[45] and specific ion-binding effects[46] could have a significant contribution to the interactions when the protein surfaces are close to each other. The major difference between the antibodies in the present work lies in the net charge of the Fab domains, and this difference is enough for the MAbs to show very different self-association and clustering characteristics. The results therefore predict that effective domain-level electrostatic interactions can play a dominant role in the self-association of antibodies leading to effective clustering.

It is clear that these CG models can be useful in predicting qualitatively, specific sequence changes in different regions of the antibody that may or may not lead to clustering and self-associating behavior. However, the actual structures that are formed by the MAbs should be supplemented by experimental investigations, especially the clustering observed in M6. The present work suggests that the models developed here can be used to study the effect of changing the net charge on a domain of the antibody ultimately affecting the self-association characteristics of engineered antibodies. The CG models can thus be used as lead investigation probes to check the effect of changing the net sequence (charge) on the domains and to draw conclusions about their self-associating behavior based on equilibrium structure formation.

CONCLUSIONS

This study has examined the role of domain-level charge−charge electrostatics in the self-association of two engineered therapeutic monoclonal antibodies (MAb1, MAb2 and their charge-swap mutants). Reduced CG models of the antibody were developed by using an elastic network normal-mode analysis. The results suggest that the MAb1 tend to form dense clusters compared to MAb2. The trend in equilibrium behavior of M1, M5, M7, and M10 can be attributed to the decrease in Fab-Fab attractions and overall increase in net charge of the molecules. Though M6 does not follow the aforementioned viscosity trend, the clustering in M6 suggests that a large number of smaller clusters are found all over the domain contrasting with MAb1, which forms a small number of large clusters. The clustering behavior suggests that this might be one of the factors leading to lower than expected viscosities for M6. The results suggest that small changes in the amino acid sequence can cause appreciable changes in the equilibrium structures of mAb solutions. At the current resolution and approximation, the present CG model captures appreciable differences in network arrangements between antibodies with different sequences in the CDR regions and predicts trends in self-associating behavior and viscosity consistent with experimental observations that are at present intractable with all-atom MD simulations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: zarraga.isidro@gene.com (I.E.Z.) and gavoth@uchicago.edu (G.A.V.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported in part (A.C. and G.A.V.) by a National Science Foundation Grant Opportunity for Academic Liaison with Industry (GOALI) supplement (to grants CHE-0628257 and CHE-1047323) and in part by the National Science Foundation through the Center for Multiscale Theory and Simulation (grant CHE-1136709). The computational resources have been provided by a grant of supercomputing time provided by the U.S. Department of Defense (DoD) via project nos. AFPRD13233C3 V and AFPRD13233022. The computations were performed in part on Einstein (Cray XT5) at Navy DoD Supercomputing Resource Center (DSRC) and Jade (Cray XT4) at U.S. Army Engineer Research and Development Center (ERDC) DSRC.

REFERENCES

(1) Waldmann, T. A. *Nat. Med.* **2003**, *9*, 269.
(2) Aggarwal, S. *Nat. Biotechnol.* **2009**, *27*, 987.
(3) Chan, A. C.; Carter, P. J. *Nat. Rev. Immunol.* **2010**, *10*, 301.
(4) Shire, S. J.; Shahrokh, Z.; Liu, J. *J. Pharm. Sci.* **2004**, *93*, 1390.
(5) Shire, S. J. *Curr. Opin. Biotechnol.* **2009**, *20*, 708.
(6) Cromwell, M. E.; Hilario, E.; Jacobson, F. *AAPS J.* **2006**, *8*, E572.
(7) Shire, S. J.; Cromwell, M.; Liu, J. *AAPS J.* **2006**, *8*, E729.
(8) Frokjaer, S.; Otzen, D. E. *Nat. Rev. Drug Discovery* **2005**, *4*, 298.
(9) Daugherty, A. L.; Mrsny, R. J. *Adv. Drug Delivery Rev.* **2006**, *58*, 686.
(10) Rosenberg, A. S. *AAPS J.* **2006**, *8*, E501.
(11) Saluja, A.; Kalonia, D. S. *Int. J. Pharm.* **2008**, *358*, 1.
(12) Liu, J.; Nguyen, M. D. H.; Andya, J. D.; Shire, S. J. *J. Pharm. Sci.* **2005**, *94*, 1928.
(13) Kanai, S.; Liu, J.; Patapoff, T. W.; Shire, S. J. *J. Pharm. Sci.* **2008**, *97*, 4219.
(14) Yadav, S.; Liu, J.; Shire, S. J.; Kalonia, D. S. *J. Pharm. Sci.* **2010**, *99*, 1152.
(15) Yadav, S.; Shire, S. J.; Kalonia, D. S. *J. Pharm. Sci.* **2010**, *99*, 4812.
(16) Yadav, S.; Shire, S. J.; Kalonia, D. S. *J. Pharm. Sci.* **2012**, *101*, 998.
(17) Yadav, S.; Sreedhara, A.; Kanai, S.; Liu, J.; Lien, S.; Lowman, H.; Kalonia, D. S.; Shire, S. J. *Pharm. Res.* **2011**, *28*, 1750.
(18) Yadav, S.; Laue, T. M.; Kalonia, D. S.; Singh, S. N.; Shire, S. J. *Mol. Pharmaceutics* **2012**, *9*, 791.
(19) Laue, T.; Demeler, B. *Nat. Chem. Biol.* **2011**, *7*, 331.
(20) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford University Press: Oxford, UK, 1987.
(21) Haile, J. M. *Molecular dynamics simulation: elementary methods*; Wiley: New York, NY, 1992.
(22) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. *J. Mol. Graphics Modell.* **2010**, *29*, 116.
(23) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192.
(24) Tozzini, V. *Acc. Chem. Res.* **2010**, *43*, 220.
(25) Saunders, M. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2012**, *22*, 144.
(26) Voth, G. A. *Coarse-graining of condensed phase and biomolecular systems*; CRC Press: Boca Raton, FL, 2009.
(27) Chaudhri, A.; Zarraga, I. E.; Kamerzell, T. J.; Brandt, J. P.; Patapoff, T. W.; Shire, S. J.; Voth, G. A. *J. Phys. Chem. B* **2012**, *116*, 8045.
(28) Brandt, J. P.; Patapoff, T. W.; Aragon, S. R. *Biophys. J.* **2010**, *99*, 905.
(29) Zhang, Z.; Pfaendtner, J.; Grafmuller, A.; Voth, G. A. *Biophys. J.* **2009**, *97*, 2327.
(30) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164.
(31) Berendsen, H. J.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165.
(32) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd *J. Comput. Chem.* **2004**, *25*, 1400.
(33) Levy, Y.; Onuchic, J. N. *Acc. Chem. Res.* **2006**, *39*, 135.
(34) Larson, R. G. *The structure and rheology of complex fluids*; Oxford University Press: New York, NY, 1999.
(35) Matthew, J. B. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 387.
(36) Simonson, T. *Rep. Prog. Phys.* **2003**, *66*, 737.
(37) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(38) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400.

dx.doi.org/10.1021/jp3108396 | *J. Phys. Chem. B* XXXX, XXX, XXX−XXX

(39) Plimpton, S. J. *J. Comput. Phys.* **1995**, *117*, 1.
(40) Ester, M.; Kriegel, H.; Sander, J.; Xu, X. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, Portland, OR.
(41) Daszykowski, M.; Walczak, B.; Massart, D. L. *Chemom. Intell. Lab. Syst.* **2001**, *56*, 83.
(42) Trappe, V.; Prasad, V.; Cipelletti, L.; Segre, P. N.; Weitz, D. A. *Nature* **2001**, *411*, 772.
(43) Lu, P. J.; Zaccarelli, E.; Ciulla, F.; Schofield, A. B.; Sciortino, F.; Weitz, D. A. *Nature* **2008**, *453*, 499.
(44) Harding, S. E. *Prog. Biophys. Mol. Biol.* **1997**, *68*, 207.
(45) Kamerzell, T. J.; Kanai, S.; Liu, J.; Shire, S. J.; Wang, Y. J. *J. Phys. Chem. B* **2009**, *113*, 6109.
(46) Gokarn, Y. R.; Fesinmeyer, R. M.; Saluja, A.; Razinkov, V.; Chase, S. F.; Laue, T. M.; Brems, D. N. *Protein Sci.* **2011**, *20*, 580.
(47) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.