Journal of Pharmaceutical Sciences



Evaluating the Effects of Hinge Flexibility on the Solution Structure of Antibodies at Concentrated Conditions

Journal:	Journal of Pharmaceutical Sciences
Manuscript ID	Draft
Article Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Blanco, Marco; Merck and Co Inc, Merck Research Laboratory Hatch, Harold; National Institute of Standards and Technology, Chemical Informatics Group, Chemical Sciences Division, Material Measurement Laboratory Curtis, Joseph; National Institute of Standards and Technology, NIST Center for Neutron Research Shen, Vincent; National Institute of Standards and Technology, Chemical Informatics Group, Chemical Sciences Division, Material Measurement Laboratory
Keywords:	IgG antibody(s), High concentration, Molecular modeling, Monte Carlo Simulation(s), Physical stability



Evaluating the Effects of Hinge Flexibility on the Solution Structure of Antibodies at Concentrated Conditions

Marco A. Blanco,^{1, 2, *} Harold W. Hatch,¹ Joseph E. Curtis,³ and Vincent K. Shen¹

¹ Chemical Informatics Group, Chemical Sciences Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.

² Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850, USA.

³NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, MD 20899,

USA

* Correspondence to: Marco A. Blanco.

E-mail address: <u>mblanco@udel.edu</u>

Current address for MAB: Merck & Co., Inc., Kenilworth, NJ 07033, USA

Abstract

Employing two different coarse-grained models, we evaluated the effect of intramolecular domain-domain distances and hinge flexibility on the general solution structure of antibodies (mAbs), within the context of protein-protein steric repulsion. These models explicitly account for the hinge region, and represent antibodies at either domain or subdomain levels (i.e., 4-bead and 7-bead representations, respectively). Additionally, different levels of mAb flexibility are also considered. When evaluating mAbs as rigid structures, analysis of small-angle scattering (SAS) profiles showed that changes in the relative internal distances between Fc and Fab domains significantly alter the local arrangement of neighboring molecules, as well as the molecular packing of the concentrated mAb solutions. Likewise, enabling hinge flexibility in either of the mAb models led to qualitatively similar results, where flexibility increases the spatial molecular arrangement at elevated concentrations. This occurs because fluctuations in mAb quaternary structure are modulated by the close proximity between molecules at elevated concentrations (> 50 mg/mL), yielding an increased molecular packing and osmotic compressibility. However, our results also showed that the mechanism behind this synergy between flexibility and packing strongly depends on both the level of structural detail and the number of degrees of freedom considered in the coarse-grained model.

Keywords

Antibodies, Small-Angle Scattering, Protein Stability, Hinge Flexibility, High Concentration, Molecular Simulation, Coarse-Grained Model

Introduction

Monoclonal antibodies (mAbs) represent the largest class of biotherapeutics in the current pipeline of the pharmaceutical industry, as their high specificity and efficacy make them ideal candidates for treating high-impact illnesses such as various type of cancers¹ and autoimmune diseases.² Currently, more than 30 antibody or antibody-derived therapeutics are marketed worldwide, and a larger number of similar proteins are under development or going through clinical trials.^{3,4} However, the success or failure of a given candidate is directly related to our ability to identify suitable formulations that provide long-term (1 year or longer) stability for these molecules. If mAbs are not adequately stabilized at the formulation conditions, their efficacy and potency as therapeutics is compromised, as well as they fail to comply with commercial and regulatory requirements.^{5,6}

Nonetheless, achieving such levels of stability remains an outstanding challenge. Changes in pH, ionic strength, temperature, and the type/concentration of any excipient in solution may alter protein-protein interactions and protein conformation, causing mAbs to undergo unfolding and/or aggregation.^{7.9} These stability problems may further be exacerbated for high-concentration formulations (> 50 mg/mL), where additional solution non-idealities (e.g., molecular crowding and multi-body interactions) may result in other stability issues such as critical opalescence, liquid-liquid phase separation, and anomalous high viscosity.¹⁰⁻¹³ As such, identifying optimal mAb formulations requires probing the effects of a significant large number of factors on the behavior and stability of antibody solutions. By contrast, from a practical standpoint, an exhaustive study of solution conditions is problematic due to the generally limited amount of available protein material during early stages of drug development, as well as to the intrinsic limitations of experimental biophysical and biochemical techniques for probing concentrated protein solutions.^{12, 14} In this regard, there is an increasing interest in developing computational mAb models to narrow this search space, if not, to predict the behavior of mAb solutions from low to high concentrations.¹⁵⁻²⁰ Notably, these models are limited to physically relevant, but computationally efficient coarse-grained representations, since fully atomistic simulations of concentrated mAb solutions remain out of reach due to the time- and length-scales of most of the protein instability phenomena.

At a molecular level, protein-protein interactions are the result of an assortment of different residueresidue and residue-solvent/excipient forces, which include solvophobic, van der Waals, dipole-dipole, hydrogen bonding, and screened electrostatic interactions.²¹⁻²⁴ These forces can be classified as: (i) short-range attractions; (ii) long-range attractions/repulsions; and (iii) steric repulsions. Interactions of type (i) and (ii) are due to the specific arrangement of hydrophobic, hydrophilic, and charged amino acids, which results in a heterogeneous mixture of attractive and repulsive surface "patches" with a variety of interacting ranges.^{25, 26} These types of interactions are the sole energetic contribution to the stability of the solution, and they are responsible for most of

Journal of Pharmaceutical Sciences

the solution structure (i.e., the arrangement of protein molecules at mid-to-long range length-scales). More importantly, they can be controlled by changes in the solution conditions, where different molecular attributes such as the protonated state of some residues, the likelihood of hydrogen-bond formation, and the effective range of charge-charge interactions can be modulated by the nature and concentration of the other species in solution.^{27, 28} On the other hand, steric repulsions are related to the protein's excluded volume (i.e., the physical space occupied by the protein); therefore, they are only affected by changes in protein structure through conformational fluctuations or unfolding, with minimal to no effect from the solution conditions. While interactions of type (iii) are not as readily controllable as the other types, they provide most of the entropic contributions to the free-energy of solution (i.e., the thermodynamic stability) due to packing effects.^{29, 30} Furthermore, steric interactions are considered the dominant force at elevated protein concentrations, where they govern the overall protein arrangement in crowded environments.^{31, 32} As such, any computational or theoretical model must appropriately incorporate these different types of interactions in order to accurately capture conformational and thermodynamic instability issues in proteins.

In the specific case of antibodies, many coarse-grained models have been developed during the last decade to study different instability phenomena of mAb solutions;¹⁵⁻²⁰ however, most of these models assume mAbs as rigid-bodies, focusing on capturing short- and long-range interactions but neglecting any conformational fluctuation that may affect the overall mAb structure. Such models contrast with recent experimental studies employing high-resolution structural techniques such as electron tomography³³ and small-angle scattering,^{34, 35} which have highlighted the highly flexible nature of mAbs. Due to the flexible linker connecting the Fc and Fab domains (i.e., the hinge), mAbs exhibit large conformational fluctuations in their quaternary structure, with a broad distribution of separation distances between Fc and Fab domains.²⁰ On the other hand, in those models where flexibility has been considered, it has been suggested that structural fluctuations in mAbs may significantly affect the osmotic compressibility of the solution at protein concentrations above 140 mg/mL.¹⁷ Structural fluctuations might also be responsible for the somewhat narrow and asymmetric phase-equilibrium coexistence curves of mAbs when compared to those of globular proteins.^{8, 16, 20} Surprisingly, we are unaware of any studies that have systematically and explicitly evaluated the role of the hinge flexibility on the behavior of antibodies, and thus it remains an open question how these inherent conformational fluctuations might affect the physical stability of concentrated mAb solutions.

A long-term objective is to develop an accurate coarse-grained computational model that accounts for the roles of all types of protein-protein interactions in order to efficiently predict the macroscopic behavior of mAb solutions. Therefore, this report takes a first step in that direction by assessing how the behavior of antibodies in solution is affected by hinge flexibility, within the context of steric repulsion or excluded volume effects. To that

end, two different coarse-grained mAb models are considered which represent antibodies at either the domain or subdomain level, and incorporate different types of hinge flexibility (e.g., stretching, bending, and torsion). Furthermore, small-angle scattering profiles are also employed to evaluate the solution structure at different length-scales, as well as to provide a direct link between the working models and the experimental behavior of mAbs.

Models and Methods

Coarse-grained antibody models

In order to evaluate the general solution structure of antibodies, we employed two different molecular coarse-grained mAb models: a 4-bead and a 7-bead representation. The use of two different models is intended to factor out artifacts that might arise as a result of the particular choice of coarse-graining. These models simplify mAbs at either a domain-level (e.g., Fab, Fc) or subdomain-level (e.g., C_H3, C_H2). That is, each of the mAb domains or subdomains is represented by spherical particles located at the center-of-mass of the corresponding structures. Moreover, the hinge region is also considered and modeled by a single bead. These considerations lead to the two different mAb models used in this work. The 4-bead representation is the coarser of the two, where the Fc domain, the two Fab domains, and the hinge are each represented by a single bead (Figure 1a). Similarly, the 7-bead model incorporates slightly more detail (Figure 1b), and it consists of: (i) two beads representing the Fc domain (i.e, the C_H3 and C_H2 subdomains of both heavy chains); (ii) two beads modeling the Fab domain, which account for the C_H1/C_L and the V_L/V_H subdomains, respectively (labeled as beads C_H1 and F_v in Figure 1b); and (iii) an additional bead for the hinge region. Unlike those coarse-grained mAb models developed elsewhere, ¹⁵⁻²⁰ we introduce the hinge region, which allows for explicitly probing mAb flexibility and the independent movement of the Fc and Fab domains with respect to each other.

The diameters of the beads are selected to match the excluded volume of the domains or subdomains that they represent. To that end, we use as reference the crystal structure of an IgG2 antibody (pdb: 1IGT). Given the high homology between different monomeric antibodies^{36, 37} (>95% sequence and structural homology, with most of the structural differences located at the hinge region), it is anticipated that the size of the coarse-grained (sub)domains are general and transferable between antibodies. The excluded volume of each (sub)domain is obtained by calculating their corresponding second virial coefficients ($B_{22,i}^{HS}$, with *i* denoting the evaluated domain or subdomain); values of $B_{22,i}^{HS}$ are based on the atomistic, crystal structure of the reference antibody, where it is assumed that each atom behaves as a hard sphere. Details of the calculation of $B_{22,i}^{HS}$ are provided in the Supporting Information (SI). The diameter of the *i*-th coarse-grained (sub)domain σ_i is readily given by $B_{22,i}^{HS} = 2\pi\sigma_i^3/3$. Because of its small size, the effective diameter of the hinge bead cannot be determined in the same manner. Instead, σ_{hinge} is estimated from its radius of gyration (i.e., $\sigma_{hinge} = 2R_g$). Here, R_g is calculated from the

 variance of the atomic coordinates; that is, $R_g^2 = \sum \mathbf{r}_k \cdot \mathbf{r}_k / n - \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the geometric center of the hinge and \mathbf{r}_k is the position of the *k*-th atom out of *n* that constitutes the hinge in the rigid, crystal structure of the reference antibody. Table 1 summarizes the resulting domain or subdomain excluded volumes and corresponding hard-sphere diameters. For illustrative purposes, Figure 1 also depicts the resulting mAb models in comparison against the crystal structure of the reference IgG2. Note that each of the beads encompasses most of the crystal structure; however, fragments of the corresponding structure are over- or underestimated as none of the antibody (sub)domains are perfectly spherical.

For each of the considered mAb representations, the interaction potential E in an implicit solvent system with N mAb molecules is given by the sum of the intra- and intermolecular energy contributions via

$$E = \sum_{n=1}^{N} \sum_{i$$

where r_i^n corresponds to the position vector for the *i*-th bead in the *n*-th molecule. *K* represents the number of beads per molecule in a given molecular representation (i.e., K = 4 or 7 for the 4-bead or 7-bead mAb models, respectively). Thus, $u_{ij}^{intra}(r)$ indicates the intramolecular interaction energy between beads *i* and *j* separated by a distance *r*, while $u_{ij}^{inter}(r)$ is the interaction energy between a pair of beads that belong to different molecules (i.e., the intermolecular energy contribution).

Here, intermolecular interactions are modeled via a hard-sphere (u_{ij}^{HS}) potential for each pair of beads,

$$u_{ij}^{HS}(r_{ij}) = \begin{cases} \infty, & \text{if } r_{ij} < \sigma_{ij} \\ 0 & \text{otherwise} \end{cases}$$
(2)

where r_i^n is the distance between the centers of the *i*-th and *j*-th beads, and $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ is the average diameter for the pair *i*-*j*.

In the case of intramolecular contributions, bonded and non-bonded internal interactions are considered in these models to probe the effect of the hinge flexibility. Although intramolecular interactions between domains and/or subdomains are affected by the same physics that control intermolecular forces, the flexibility of the hinge region is rather dominated by the steric constrains on the backbone of those residues that link the Fc and Fab domains.³⁸ Different experimental studies have shown that the mobility of the mAb domains is seemingly

random, with no indication of any influence from short- and/or long-ranged interactions between domains.^{39, 40} Thus, as a first approximation, intramolecular contributions are modeled here by simple steric effects via a hardsphere potential akin to eq. 2a, but using an effective diameter that is 5% smaller than that of intermolecular interactions (i.e., $\sigma_{ij}^{intra} = 0.95\sigma_{ij}$, with σ_{ij} defined in eq. 2). This reduced size is intended to compensate for the overestimation of the (sub)domain sizes as they are assumed to be spherical particles. That is, due to the nonglobular geometry of domains and subdomains, the resulting bead representation misrepresents the shape of these mAb fragments (cf. Figure 1). The use of this reduced intramolecular diameter provides a better representation of the internal mobility of the antibodies. Consequently, Fc and Fab domains are permitted to freely move (and rotate in the case of the 7-bead model) so long as they are subjected to bond and steric constraints, which we describe below.

Additionally, stretching of the links between the hinge region and the other domains is also freely allowed for the 4-bead mAb model (i.e., without any energetic penalty), although it is limited to a separation distance *d* between the center of the hinge bead and the center of the Fc or Fab domain so that it does not lead to overlaps and is no larger than 7 nm (i.e., $\sigma_{hinge,i}^{intra} \le d \le 7$ nm, with the subscript *i* indicating the Fc or Fab domains). This range of separation distances between linked domains is intended to probe the effect of hinge flexibility under extreme cases beyond those reported in the literature. Moreover, it provides a broad perspective regarding how the flexibility at the hinge region affects the behavior of mAbs in solution from dilute to concentrated conditions. For the 7-bead model, the bond distances between any pair of linked beads (i.e., hinge–C_H2, hinge–C_H1, C_H1–F_V, and C_H3–C_H2) are fixed and set to values equal to those of the reference IgG2 molecule, as summarized in the Table S1 in the SI.

Flat-histogram Monte Carlo

Flat-histogram sampling methods are used to evaluate the solution behavior of antibodies. Specifically, Wang-Landau-Transition Matrix Monte Carlo^{41, 42} (WL-TMMC) simulations are performed in the grand-canonical ensemble. WL-TMMC is a powerful method that provides all thermodynamic properties (e.g., free-energy, potential energy, pressure) and structural information as a function of concentration at fixed volume and temperature. As simulations are performed in the grand-canonical ensemble, density fluctuations directly related to macroscopic behavior can be readily obtained from a single simulation, with minimal system size effects.⁴³

Grand canonical WL-TMMC is used to calculate the probability of observing *N* particles for a given chemical potential μ_0 (e.g., activity), temperature *T*, and volume *V* (i.e., the macrostate probability distribution, Π (*N*; μ_0, T, V)). Specific implementation details of WL-TMMC are provided elsewhere^{25, 44} and briefly summarized here. Simulations are initialized by running Wang-Landau with the update factor initially set to unity. Every time

a flatness criterion of 80% is met, this factor is multiplied by 0.5 and the histogram of visited macrostates is reset. The collection matrix, needed for Transition Matrix Monte Carlo, is updated when the update factor is smaller than 10⁻⁶. The simulation completely switches to Transition Matrix Monte Carlo when the update factor is below 3×10^{-8} , with the TMMC biasing function calculated from the collection matrix and subsequent updates every 10⁶ trial moves. Canonical ensemble averages for the quantities used in this work (see below) are accumulated and update every 10³ trials after a simulation has swept at least two times. A sweep is achieved if each macrostate (i.e., each *N* value) has been visited from a different macrostate at least 100 times. In order to reduce computational time, the entire concentration range for a given isotherm (i.e., the range of *N* macrostates, from 0 to N_{max}) is divided into 12 windows, with overlap between neighboring windows of no less than five particle numbers. The size of each window is selected to balance computational load by decreasing window size with increasing concentration range by a power-scaling factor of 3. Furthermore, the chemical potential μ_0 of each window is selected to ensure that the most likely macrostate falls within the corresponding concentration range.

 $\Pi(N;\mu_0,T,V)$ for the complete N range is reconstructed by matching the free-energy of neighboring windows at the middle of the overlapping region, while the two largest and smallest macrostates at each window are discarded when neighbor windows are present. To do so, all the resulting macrostate distributions are reweighted to the same thermodynamic state (μ , *T*, *V*) via

$$ln \Pi(N;\mu) = ln \Pi(N;\mu_0) + \beta(\mu - \mu_0)N \quad (3)$$

where $1/\beta = k_B T$ is the thermodynamic temperature, and k_B denotes Boltzmann's constant. Eq. 3 corresponds to a histogram reweighting scheme, and it provides the framework to evaluate a wide range of thermodynamic states once the entire macrostate distribution has been reconstructed (see below).

In order to achieve an efficient sampling of the configurational space of mAbs, the following MC trial moves are implemented: (i) rigid-body translation and rotations; (ii) particle insertion/deletion; (iii) multi-first particle insertion⁴⁵ (MFI); and (iv) configurational-bias⁴³ (CBMC). Both MFI and CBMC moves are employed to ensure that the most likely state is selected (from a list of proposed configurations) during the trial move. Insertions and deletions are carried out by coupling MFI with CBMC. MFI is used to select the position for inserting a new particle (based on steric interactions). The inserted particle corresponds to one bead of any of the arms in the mAb molecules. CBMC is then employed to rebuild the molecule one bead at a time. Note that after inserting the hinge region and one of the arms, the positions of the remaining arms in the molecule are not fully independent,⁴⁶ hence a couple-decouple scheme⁴⁷ is employed to simultaneously insert these arms and to correctly

map the configurational phase space of the molecule. A total of 10 proposed configurations are evaluated during either MFI or CBMC.

Additionally, CBMC is also performed on canonical moves for partial 're-growth' steps, as follows. A mAb molecule and one of its beads are randomly selected. If the selected bead is the hinge region, a direction in which the molecule is going to start regrowing is also randomly selected; otherwise the molecule is regrown towards the hinge region first. The molecule is then partially reconstructed using the scheme outlined here, where the position of the selected bead is kept fixed. Table 2 summarizes the different MC trials used in this work, as well as their corresponding probabilities for attempting one of these moves. Additional details involving Monte Carlo simulations of the 7-bead mAb model are provided in the SI.

During a simulation, canonical ensemble averages are collected for the bond and angle distributions between mAb (sub)domains, as well as site-site distribution functions. In general, for an observable A, the canonical average (denoted \tilde{A}_N) is calculated as

$$\tilde{A}_N = \frac{1}{M_N} \sum_{j=1}^{\tau} A_j \delta(N - N_j) \qquad (4)$$

where N_j is the number of particles at the trial $j, \delta(x)$ is the Dirac delta function, and A_j is the value of the observable *A* at the trial *j*. τ is the total number of sampled configurations during the course of the simulation, and $M_N = \sum_j \delta(N - N_j)$ is the number of times the macrostate *N* is sampled.

After the simulation is finished and $\Pi(N;\mu)$ is calculated, grand-canonical averages of A are readily obtained as a continuous function of molecular concentration $\langle N/V \rangle_{\mu VT}$ via

$$\langle A \rangle_{\mu VT} = \frac{\sum_{N} \tilde{A}_{N} \Pi(N;\mu)}{\sum_{N} \Pi(N;\mu)}$$
$$\langle N \rangle_{\mu VT} = \frac{\sum_{N} N \Pi(N;\mu)}{\sum_{N} \Pi(N;\mu)}$$

Note that the state condition of the average properties in Eq. 5 can be controlled by reweighting the macrostate distribution to different values of μ via eq. 3. Additionally, explicit calculation of $\Pi(N;\mu)$ in

Journal of Pharmaceutical Sciences

combination with $\langle N/V \rangle_{\mu VT}$ allow one to fully define any thermodynamic state variable such as entropy.⁴² Particularly, one is often interested in knowledge of the osmotic pressure *P*, as it provides a direct connection between the molecular and the macroscopic behavior of proteins in solutions.¹² *P* is calculated as

$$\beta PV = ln \left[\sum_{N=0}^{\infty} \Pi(N;\mu) \right] - ln \Pi(0) \quad (6)$$

Second virial coefficients

Effective interactions between proteins are typically characterized by the second virial coefficient B_{22} . This quantity can be experimentally probed for most systems of interest,⁴⁸ and thus it plays a central role in many molecular models and theoretical approaches for evaluating the behavior of proteins in solution.^{49, 50} B_{22} is formally related to the strength of the intermolecular interaction between a pair of proteins via⁵¹

$$B_{22} = \frac{1}{2\Omega^2} \iint \langle e^{-\beta W_{22}} - 1 \rangle dr_{12} d\Omega_{12}$$
(7)

where the brackets $\langle ... \rangle$ denotes that the corresponding quantity is ensemble averaged over all internal degrees-offreedom, as well as over the position/orientation of any other species in solution. W_{22} is the infinite-dilution potential of mean force between two protein molecules, which corresponds for the working implicit-solvent mAb models to the intermolecular energy contribution (i.e., the second summation on the right hand side of eq. 1 when N = 2). r_{12} is the center-of-mass separation vector between a pair of molecules, while Ω_{12} represents the molecular orientation of one molecule with respect to that of the second molecule. $\Omega = \int dx \Omega_i$ indicates the complete orientational space for each molecule.

Qualitatively, as described by eq. 7, negative values of B_{22} indicate net attractive interactions, while positive values represent net repulsive pair interactions. Given that the working mAb representations only interact through a steric (repulsive) potential, B_{22} can only take positive values. Notably the above definition of B_{22} is only a function of the intermolecular interaction energy; nonetheless, for the antibody models studied in this work, eq. 7 implicitly depends on the conformational degrees of freedom, where the degree of molecular flexibility affects the resulting value.

To obtain theoretical values of B₂₂, direct Mayer-Sampling Monte Carlo (MSMC) was implemented.^{52, 53}

This method calculates the integral in eq. 7 using a biased Monte Carlo approach on a system with only two molecules in an infinite volume. Specifically, MSMC performs an importance-sampling Monte Carlo simulation based on those configurations relevant to the integral for the system of interest, while simultaneously sampling a reference system for which the second virial coefficient is known. Thus, B_{22} for the system of interest is obtained as a free-energy perturbation from a simpler system. Herein, the reference system is taken as a spherical particle with a diameter of $\sigma_{ref} = 4$ nm located at the hinge of the mAb, which interacts through a hard-sphere potential as that described in eq. 2. The particle size of the reference system was optimized from preliminary simulations to ensure resulting values of B_{22} are independent of the choice of the reference system.

MSMC is performed for a system of only two molecules in an infinite volume, where attempted trial moves consist of rigid-body translations and rotations. Furthermore, when molecular flexibility is considered, rigid-body moves are supplemented by molecular regrowing trials, where each bead in a molecule is relocated. For this latter trial move, an attempted new configuration is selected with a probability p_{select} that depends on the Boltzmann factor of the intramolecular energy (i.e., $p_{select} \propto exp(-\beta u_{intra})$, with u_{intra} being the intramolecular contribution in eq. 1). Each MSMC simulation consists of a short "equilibration" period (5×10⁶ configurations) to adjust step sizes for the trials to achieve a 50% acceptance rate, followed by a sampling period of 5×10⁸ configurations. In previous works, similar operating parameters were found to provide well converged B_{22} values on similar and more complex systems.^{17, 54, 55}

Small-angle scattering profiles

Small-angle scattering (SAS) intensities I(q) are used here to probe the solution structure of antibodies. From a computational standpoint, I(q) consists of Fourier transforms of intra- and inter-molecular spatial distribution functions between pairs of atoms or scattering sites, hence providing information regarding the internal structure of the molecules and their local arrangement with respect to themselves.⁵⁶ As a result, they allow one to identify whether intermolecular spatial correlations influence the conformational behavior of antibodies. More importantly, evaluation of I(q) from molecular simulations provides an explicit connection with an experimental observable, as this quantity can be assessed from SAXS or SANS techniques.

Following a previously developed methodology,⁵⁷ SAS intensities are readily calculated as $I(q) = I_{intra}$ (q) + $I_{inter}(q)$. $I_{intra}(q)$ and $I_{inter}(q)$ are the intra- and intermolecular scattering contributions, respectively, and are evaluated via

 $I_{intra}(q) = \langle N \rangle_{\mu VT} \sum_{a=1}^{K} f_a^2 + \sum_{a \neq b}^{K} f_a f_b \langle N \hat{\omega}_{ab}(q;N) \rangle_{\mu VT} \quad (8a)$

$$I_{inter}(q) = \sum_{a,b=1}^{K} f_a f_b \langle (N - \delta_{ab}) [S_{ab}(q;N) - 1] \rangle_{\mu VT} \quad (8b)$$

where the subscripts *a* and *b* identify the different beads that constitute a mAb molecule, and thus the summations in the above expressions go over the total number of beads *K* in a given mAb representation. δ_{ab} corresponds to Kronecker delta function, which is equal to 1 if a = b and zero otherwise. f_i represents the scattering amplitude for each of the beads in a molecule. Since these beads are coarse-grained representations of the mAb domains or subdomains, f_i is given by the scattering form factor of the corresponding beads, which are calculated from the crystal structure of the reference antibody. Details on the calculation of f_i are provided in the SI. The brackets $\langle ... \rangle_{\mu VT}$ indicate that those quantities are obtained from grand-canonical ensemble averages (see eq. 5). The grandcanonical ensemble accurately captures the experimental measurement of I(q), where molecules are exchanged between the scattering volume *V* and the bulk solution.

In eq. 8, $S_{ab}(q;N)$ corresponds to the site-site static structure factor for the pair of beads *a-b* from different molecules, and it is measured in a canonical ensemble (i.e., at fixed *N*, *V*, and *T*). $\hat{\omega}_{ab}(q;N)$ is an equivalent quantity to $S_{ab}(q;N)$ for intramolecular pairs. These two functions are evaluated from the following Fourier transforms

$$\hat{\omega}_{ab}(q;N) = \rho \int \omega_{ab}(r;N) J_0(qr) 4\pi r^2 dr \quad (9a)$$

$$S_{ab}(q;N) - 1 = \rho \int [g_{ab}(r;N) - 1] \times J_0(qr) 4\pi r^2 dr \quad (9b)$$

where $J_0(x) = \frac{\sin(x)}{x}$ is the zeroth spherical Bessel function. $\rho = N/V$ is the number density of molecules for a given thermodynamic macrostate. $\omega_{ab}(r;N)$ corresponds to the site-site intramolecular distribution function, and $g_{ab}(r;N)$ represents the site-site radial distribution function. These two functions account for the probability of finding a bead *b* at a distance *r* from a bead *a*, and in a spherical shell of volume $4\pi r^2 dr$. Conceptually, the only difference between these two functions is that $\omega_{ab}(r;N)$ only accounts for the pair of beads *a*-*b* belonging to the same molecule, while $g_{ab}(r;N)$ considers the same pair but with each site belonging to different molecules. The

symmetry of the system ensures that
$$g_{ab}(r;N) = g_{ba}(r;N)$$
 and $\omega_{ab}(r;N) = \omega_{ba}(r;N)$.

Following standard practices in the analysis of SAS profiles, calculated I(q) curves from eq. 8 are processed for smearing effects via^{59, 60}

$$I(q) = \int_{0}^{\infty} \mathcal{N}(t \mid q, \sigma_q) I'(t) dt \quad (10a)$$

$$\mathcal{N}(t \mid q, \sigma_q) = \frac{1}{(2\pi\sigma_q^2)^{1/2}} exp\left[\frac{-(t-q)^2}{2\sigma_q^2}\right] \quad (10b)$$

where I'(t) is the calculated SAS intensity (eq. 8), while I(q) represents the smeared scattering intensity. $\mathcal{N}(t \mid q, \sigma_q)$ is the smearing function and corresponds to a Gaussian probability distribution function with characteristic moments q and σ_q^2 . The value of σ_q is taken here as 10% of q.

Results and Discussion

The local structural arrangement of antibody solutions is studied here as a function of protein concentration and in the context of the hinge flexibility. To that end, SAS profiles, as obtained from grand canonical WL- TMMC simulations, are probed for the working implicit solvent mAb models at a range of concentrations of 0–200 mg/mL, where mass concentrations are calculated by assuming a typical mAb size of 145 kDa. Simulations are performed as described in the Models and Methods section on a cubic box with length L = 90 nm under periodic boundary conditions. A series of systems with different internal mAb structures are probed, including fully and partially flexible mAbs, as well as rigid-body mAbs with different domain-domain distances. These rigid vs. flexible systems are probed in the context of molecular excluded volume (i.e., under only steric interactions).

Effects of intramolecular domain-domain distances

To evaluate the role of hinge flexibility on the solution structure of antibodies, it is critical to determine whether or not changes in the relative intramolecular distances domains can influence mAb behavior, as well as the outcome of any macroscopic structural observable. As an initial step, a series of different rigid body mAb systems are probed. Each of these systems consists of mAb molecules (either 4-bead or 7-bead mAb representations) with different internal but rigid configurations. Specifically, the relative position of all domains with respect to each other is characterized by the angles θ_1 , θ_2 , and θ_3 , which correspond to the angles Fc–hinge– Fab1, Fc–hinge–Fab2, and Fab1–hinge–Fab2, respectively. The different sets of { θ_1 } considered here spans configurations where all domains are in contact with each other (i.e., at "closed"-contact) to those where the

Journal of Pharmaceutical Sciences

domains are at the largest separation (i.e., in an "expanded" geometry). For simplicity, the distance of any of the (sub)domains bonded to the hinge is fixed and equal to that of the reference antibody (see Table S1), which allows one to control Fab–Fab and Fab-Fc distances by $\{\theta_1\}$ alone.

To analyze the small-angle scattering profiles of antibodies (i.e. I(q)), one needs to independently consider three main q regions: (i) the high-q region for $q > 1 \text{ nm}^{-1}$; (ii) the intermediate-q region for 0.3 nm⁻¹ $\leq q \leq 1$ nm⁻¹; and (iii) the low-q region for q < 0.3 nm⁻¹. Note that the wavevector q is associated with a real-space distance d via $d = 2\pi/q$. As such, a SAS intensity at a given q value is related to the magnitude of the spatial correlations of the system at distances around d. Therefore, the high-q region is related to structural correlations within each (sub)domain (e.g., the secondary and tertiary protein structure). Given that both mAb representations considered here are coarse-grained at that level, the resulting SAS intensities in this region are unaffected by changes in intra-domain distances, but they linearly scale with protein concentration (see below). On the other hand, the intermediate-q region involves short-ranged structural correlations, which include spatial correlations between domains (e.g., the quaternary mAb structure) and the local arrangement between neighboring molecules. In the case of the low-q region, these SAS intensities depend on long-ranged correlations, and thus they provide information regarding the macroscopic behavior of antibodies. Since only steric interactions are evaluated for the working mAb representations, these long-ranged correlations translate into how mAbs molecules arrange or 'pack' with respect to each other. For simplicity, all depicted scattering profiles are normalized by the scattering intensity of a single molecule at q = 0 (I_0). I_0 is calculated as the sum of the scattering lengths of all the atoms in the reference antibody (see SI).

Figure 2 illustrates the resulting small-angle scattering profiles for the different mAb models under a variety of rigid-body configurations at an average protein concentration of 175 mg/mL. Similar I(q) curves at different concentrations are also provided in the SI. The results in Figure 2 clearly show that changes in the relative intramolecular distance between domains (i.e., molecular conformation) significantly affect the intermediate- and low-*q* regions, with similar outcomes for both mAb representations. That is, decreasing interdomain distances leads to higher SAS intensities at small *q*, which implies more efficient molecular packing for mAbs in the closed-contact configuration. Similarly, smaller domain-domain separations yield more pronounced interaction peaks at intermediate-*q*, which suggests an increase in short-ranged spatial correlations between nearest-neighbor molecules. However, over the same set of mAb configurations, there is also a decrease in the correlations at distances equivalent to the size of the domains (i.e., $q \approx 1 \text{ nm}^{-1}$), which is the dominant feature in the scattering profiles for the extended configuration. Since these mAb models in Figure 2 are rigid, larger I(q) at this latter *q*-value corresponds to a higher likelihood that two (sub)domains from different mAb molecules are at contact. Notably, these differences in I(q) are more pronounced at high mAb concentrations (cf. Figs. 2, S2, and S3), with any effect from changes in inter-domain separations being negligible at dilute conditions (< 10 mg/mL).

The above results are consistent with a previous study from Castellanos et al.,³⁴ which suggested that for flexible molecules like antibodies, molecular structures with a smaller characteristic size yield higher SANS intensities. When one compares the effective size σ_{eff} of the working mAb systems via their B_{22} values (where $\sigma_{eff} = (3B_{22}/2\pi)^{1/2}$), the magnitude of the low-q SAS intensities of these systems inversely correlates with their excluded volume (see Table S3 in the SI). For the 4-bead mAb model, σ_{eff} ranges between 10.4 nm and 11.2 nm for the 'closed' and 'extended' configurations, respectively; similarly, in the 7-bead representation, 10.9 nm ≤ $\sigma_{eff} \leq 11.1$ nm over the same range of mAb structures. While the direct correlation between molecular size and SAS intensity may seem trivial from the perspective of globular proteins, it is not obvious for a protein with a "Y" shape such as mAbs in a rigid configuration. For instance, one might have anticipated that the 'extended' configuration (with the largest value for σ_{eff}) would favor phenomena like inter-digitation, and thus it would yield an improved molecular packing and higher scattering intensity at low-q. Instead, the closed configuration, rather than the extended one, yields more efficient molecular packing (cf. Figure 2). That is, at fixed protein concentrations, solutions composed of mAb structures with smaller σ_{eff} have larger inter-protein distances, which means mAb molecules with smaller σ_{eff} are less restricted for diffusing or tumbling in solution (i.e., improved mobility). This relation between σ_{eff} and molecular packing is also generally reflected in the intermediate-q region of the SAS profiles (cf. Figure 2). By changing the quaternary structure of the rigid mAb representations, there is clearly a shift of the interaction peak towards lower *q*-values as the structure changes from an 'extended' to a 'closed' configuration. The change is noticeable (e.g., from 0.95 nm⁻¹ to 0.5 nm⁻¹ for the 4-bead mAb), reflecting a change from strong (sub)domain-(sub)domain correlations to protein-protein correlations. Consequently, from an excluded-volume standpoint, this implies that structures with shorter intermolecular domain-domain distances are anticipated to have more favorable properties at higher concentrations, which could have rheological implications or improved solubility than their 'extended' counterparts.

Effects of flexibility on the solution structure of mAbs

The rigid-body results shown above demonstrate that both the local arrangement between neighbor molecules (e.g., domain–domain contacts) and the macroscopic behavior at high mAb concentrations are strongly affected by the intramolecular domain distances (within a rigid mAb structure). However, the structure of antibodies is far from rigid. Specifically, their quaternary structure is inherently flexible, and it is expected that the separation between domains is constantly oscillating between 'closed' and 'extended' configurations. To assess the effects of these intramolecular structural fluctuations on the behavior of mAbs, we incorporate different levels of flexibility to both mAb models. In the case of the 4-bead mAb, flexibility is accounted for by allowing the distance of any domain to both the hinge and the other domains to freely fluctuate (i.e., 'bond' and 'angle'

Journal of Pharmaceutical Sciences

flexibility, respectively) with respect to the rigid structure of the reference IgG2. On the other hand, for the 7-bead mAb, 'bond' flexibility is not considered, but instead we incorporate 'angle' flexibility. That is, the relative positions of those beads belonging to the same domain are fixed with respect to that of the hinge, but they are allowed to fluctuate with respect to the position of the other domains. Additionally, a 'fully' flexible configuration is also evaluated for the 7-bead mAb, which allows for the independent rotations of each of the domains. These latter degrees of freedom arise from the torsion of the linkers that connect the hinge to each domain, where the F_V and C_H3 subdomains freely rotate about the axis defined by the hinge and the C_H1 and C_H2 , respectively. Note that the movement of the C_H1 and C_H2 subdomains is controlled by the 'angle' flexibility. Figure 3 depicts the SAS intensities for both mAb representations under different degrees of flexibility, and at a protein concentration of 175 mg/mL. Similar figures at different mAb concentrations are also provided in the SI (Figs. S4 and S5).

As one might anticipate from the previous results, by enabling mAb flexibility, significant differences are observed in the corresponding I(q) profiles with respect to those of the reference, rigid antibody. In general, allowing for mAb domains to freely move (and rotate for the 7-bead mAb) leads to higher SAS intensities in the low- to intermediate-*q* range. The magnitude and nature of these differences are similar for both levels of coarse-graining. In the case of the 4-bead mAb (Figure 3a), differences in scattering intensities between rigid and flexible configurations are limited to q values smaller than that of the interaction peak (i.e., $q \le 1.0 \text{ nm}^{-1}$). That is, flexibility increases spatial correlations between neighboring molecules, which indicates a larger likelihood of finding two mAb molecules in contact at high protein concentrations, while reducing domain–domain correlations related to some degree of interdigitation. This increase in short-ranged correlations also propagates to larger length-scales, which suggests changes in the solution structure are due to more efficient molecular packing. Note that these effects are observed regardless of the type of flexibility considered (i.e., 'bond' *vs*. 'angle' flexibility). However, there are some minor differences in the resulting I(q) profiles by allowing for either of these degrees of freedom (cf. teal and green curves in Figure 3a). Conversely, there is a synergistic effect in the low-*q* region when both types of intramolecular degrees-of-freedom are considered, which results in even larger scattering intensities and indicates that a 'fully' flexible 4-bead mAb yields an improved arrangement of molecules.

In the case of the 7-bead mAb, the effects of flexibility on the solution structure are mainly limited to the intermediate-*q* region, while only a slight increase in I(q) is observed at low-*q* values (Figure 3b). Within the intermediate-*q* range, two different changes in the scattering profiles are prominent. At $q \approx 0.6$ nm⁻¹ (i.e., at the interaction peak), there is a significant increase in the scattering intensity for the 'fully' flexible configuration. As discussed above and elsewhere,⁶¹ a larger interaction peak is associated with an increase in the number of closely packed molecules (i.e., stronger short-ranged spatial correlations or short mAb-mAb distances). As such, by fully allowing for hinge flexibility, mAb molecules (in a 7-bead representation) exhibit a higher propensity to stay in

close proximity to each other. By contrast, at the characteristic distance of the size of the individual domains (i.e., $q \approx 1 \text{ nm}^{-1}$), there is a decrease in I(q) as a result of incorporating flexibility, which implies a loss in molecular and/or solution structure at that length scale. Nonetheless, this latter reduction in short-ranged correlations is somewhat anticipated, as the 'fully' flexible configuration allows for the individual domains to independently reorient. That is, given that no interaction forces beyond sterics are considered, it is expected that the movement of each domain depends only on the presence of neighboring molecules but is largely uncorrelated from the position/orientation of the other domains in the molecule. Notably, substantial changes in I(q) only occur when the Fc and Fab domains freely rotate, as allowing for 'angle' flexibility alone yields small changes with respect to the reference, rigid-body antibody. From the perspective of steric interactions, these results suggest that "floppiness" in the mAb domains (i.e., torsions at the hinge) is a key factor of the hinge flexibility that favors the close proximity and packing between antibodies at high concentrations.

Interestingly, comparison of the effects of the different levels of flexibility at different mAb concentrations indicates that there is a limiting concentration for which a rigid mAb representation is no longer capable to capture the solution behavior of antibodies. That is, at low protein concentrations, the resulting SAS profiles from flexible mAb representations exhibit minimal difference from those of the rigid configuration. As protein concentration increases, however, the differences between I(q) for rigid and flexible models become evident, which highlights the inherent limitations in probing the behavior of mAbs at elevated concentrations by neglecting hinge flexibility. Nonetheless, the concentration at which such difference is observed strongly depends on the number of internal degrees of freedom incorporated in the model. Thus, for instance, by including only 'angle' flexibility to the 7-bead representation, the resulting SAS profile is indistinguishable from that of the rigid model up to a concentration of 100 mg/mL (cf. Fig. 3b and Fig. S5 in the SI); on the other hand, allowing for rotation of the domains in the 7-bead mAb (in addition to 'angle' flexibility) leads to I(q) that significantly differs from the rigid representation at concentrated conditions. Given that the 'fully' flexible configurations are expected to provide a more realistic representation of the behavior of antibodies, the remainder of this report focuses on comparing the solution behavior of mAbs when all possible degrees of freedom are allowed.

When considering 'fully' flexible mAb representations, the effect of flexibility on the solution structure of both mAb models (as observed by SAS profiles) is only prominent at protein concentrations above 50 mg/mL (Figure 4). At dilute conditions and in the absence of protein-protein interactions beyond sterics, the average separation between antibodies is of several molecular diameters. As a result, changes in the quaternary mAb structure due to hinge flexibility yield negligible effects on the low- to intermediate-*q* SAS intensities. That is, from the solution structure perspective, there is no distinction between rigid and flexible molecules under dilute conditions, where their arrangement in solution is governed by the characteristic molecular size. On the other

Page 17 of 44

Journal of Pharmaceutical Sciences

hand, as protein concentration increases, the average mAb–mAb separation decreases, and thus the quaternary mAb structure is highly influenced by the close proximity with neighboring molecules. The ubiquitous movement of the mAb domains (resulting from the hinge flexibility) becomes restricted to favor more efficient packing between nearest-neighbor molecules, as it is evidenced by the increase in the scattering intensity at $q < 1 \text{ nm}^{-1}$.

Although the effects of increasing concentration on the structure of flexible mAbs are qualitatively equivalent for both levels of coarse-graining (i.e., an increase in spatial correlations), the extent of these effects vary by the number of degrees of freedom that are considered. That is, depending on both the number of beads and the types of flexibility moves used to model antibodies, one might obtain a different resulting behavior for these molecules at elevated concentrations. To better understand the underlying causes for this variability, one needs to analyze the distribution of molecules in real-space; specifically, we focus our attention on the hinge– hinge radial distribution function $g_{HH}(r)$. This distribution function is defined as the number of system configurations, relative to that of an ideal gas, where a hinge particle is within a spherical shell separated by a distance r from the hinge of another mAb molecule. Given that the hinge is buried inside the mAb, $g_{HH}(r)$ can then provide information about the spatial arrangement of mAb molecules, as well as changes in their internal structure. Figure 5 illustrates $g_{HH}(r)$ as a function of protein concentration for rigid and fully flexible mAbs (both 4-bead and 7-bead representations).

The results in Figure 5 show how flexibility affects the structure solution of antibodies within the limitations of each coarse-grained model. Thus, by enabling flexibility in the 4-bead mAb representation (Figure 5a), there is a significant increase in the likelihood of finding two mAb molecules at distances far below the effective molecular size (i.e., the position of the first peak in $g_{HH}(r)$) when compared to those of rigid mAbs. Furthermore, flexibility also yields a decrease in the height of the first peak, which can be related to a loss in solution structure around the first neighbor shell. Two different internal structural features in flexible 4-bead mAbs were observed from sampled configurations at high concentrations that may explain these differences in $g_{HH}(r)$. (i) mAbs acquire a 'closed' structure (i.e., Fabs and Fc in closed-contact), which exposes the hinge; and (ii) mAbs take a configuration where at least one of the internal angles is extended, facilitating inter-penetration or inter-digitation between molecules. Both of these types of molecular structures can increase the probability of finding two hinge beads separated by a distance r < 11 nm, which also reduce the number of configurations within a distance *r* around the first neighbor shell.

In the case of the fully flexible 7-bead mAb (Figure 5b), changes in $g_{HH}(r)$ as a result of flexibility are completely opposite to those observed for the 4-bead representation. That is, hinge flexibility reduces the likelihood that two hinge domains are at short distances (e.g., r < 10 nm), and it yields larger oscillations around

the first neighbor shell. These changes are consistent with those observed from I(q) and are the result of the rotation of the outer subdomains in the Fc and Fab arms. While analysis of sampled configurations at high concentrations suggests a structural preference of flexible mAb molecules towards extended inner angles (i.e., $C_H 1$ -hinge- $C_H 2$ and $C_H 1$ -hinge- $C_H 1$), the continuous re-orientation of the mAb domains minimizes interdigitation between molecules; hence the reduced $g_{HH}(r < 11 \text{ nm})$ and increased solution structure at 10 nm < r < 20 nm in comparison with $g_{HH}(r)$ of the rigid mAb.

In spite of the mechanistic differences on how flexibility affects the molecular behavior of the different mAb representations, there is a common feature among the $g_{HH}(r)$ of all coarse-grained mAb models: a shift of the effective molecular size (i.e., the position of the first peak in $g_{HH}(r)$) towards smaller size as protein concentration increases. For purely steric systems such as those considered in this section, the thermodynamics of the solution is completely controlled by the effective or characteristic size of the molecules.^{62, 63} As such, changes in the effective mAb size are anticipated to significantly impact the macroscopic behavior of antibodies at concentrated conditions. By evaluating the osmotic pressure P of the different mAb representations (Figure 6), the decrease in mAb size becomes evident when comparing P against that of hard-spheres with equivalent dimensions. Here, the hard-sphere pressure is calculated via the Carnahan-Starling approximation⁶⁴ for particles with diameter σ_{HS} of 10 nm and 11 nm. For both 4-bead and 7-bead models, the increase in P with increasing concentration is not as pronounced as that of hard-spheres, with the 7-bead mAb exhibiting lower osmotic pressures than their 4-bead counterparts at the same concentrations. Although this "reduced" P occurs for both rigid and flexible models, flexibility consistently leads to smaller values of P. From a physical chemistry standpoint, the osmotic pressure of steric systems is directly related to "crowding" effects, and thus lower values of P imply an increase in the amount of available volume in solution.⁶⁵ When considering the non-globular nature of antibodies, it is somewhat expected that the increase in osmotic pressure with concentration is not as stiff as that of hard-spheres (cf. Figure 6). The "Y" shape of mAb molecules facilitates their packing, which consequently results in lower values of P (or equivalently larger available volumes). Similarly, flexibility enables the rearrangement of the quaternary mAb structure, leading to further improved molecular packing and even lower osmotic pressures.

Summary and Conclusions

Employing two different coarse-grained models, we have investigated the effect of changes in intramolecular domain-domain distances on the solution structure of diluted to concentrated mAb solutions. These models represent mAbs at either domain- or subdomain-level (i.e., 4-bead or 7-bead mAb models). Regardless of the level of structural detail in these models, the results qualitatively highlighted how the quaternary structure of mAbs plays a critical role on modulating the organization of mAb molecules in solution, specifically at elevated

Page 19 of 44

Journal of Pharmaceutical Sciences

concentrations. When considering mAbs as rigid structures, analysis of SAS profiles showed that changes in the relative internal distances between Fc and Fab domains significantly alter local arrangement of neighboring molecules, as well as the molecular packing of the concentrated mAb solutions. Thus, rigid mAb representations with smaller effective molecular sizes (i.e., shorter separations between domains) yield increased spatial correlations between molecules at both short- and long-range length-scales. However, these results also suggest that one must carefully consider existing rigid mAb models, as they are based on a single molecular conformation and may misrepresent antibody behavior at high concentrations. This is particularly relevant for studying stability phenomena that strongly depend on an accurate description of molecular packing such as high viscosity and protein solubility.

Additionally, the solution behavior of mAbs was also probed in the context of the conformational fluctuations in the protein structure. Phenomenologically, enabling hinge flexibility in the working mAb models leads to increased spatial correlations at high concentrations in comparison to those of the reference, rigid antibody. This occurs because fluctuations in mAb quaternary structure are modulated by the close proximity between molecules at elevated concentrations (> 50 mg/mL), which results in a more efficient molecular packing and osmotic compressibility. However, our results also showed that the mechanism behind this synergy between flexibility and molecular arrangement strongly depends on both the level of structural detail and the number of degrees of freedom considered in the coarse-grained model. For instance, allowing for 'angle' flexibility leads to significant changes in molecular packing for the 4-bead model, but to small changes for the 7-bead model. When independent rotations of the mAb domains are considered, one observes larger differences between rigid and flexible representations in the 7-bead model. In this regard, it is expected that the 'fully' flexible 7-bead mAb model represents more realistically and accurately the solution structure of antibodies, as it accounts for all possible movements of the linkers at the hinge as well as the anisotropy of the individual domains. Nonetheless, if one is interested in qualitatively capturing the physical behavior of antibody solutions, the flexible 4-bead model might be considered in favor of computational efficiency.

We note that the present models do not quantitatively agree with measured profiles in confined environments.³⁴ This lack of quantitative agreement suggests that some details of the models considered here (e.g., spherical bead shape and hard core interactions) are not adequate to reproduce observed scattering profiles. In order to appropriately perform a comparison of the coarse-grained models against experimental SAS data, one needs to accurately introduce other inter-molecular forces to the working computational model, as well as to account for the anisotropy of the individual (sub)domains. Including the effects of these softer interactions and non-spherical bead shapes to represent the domains is the topic of future work. Additionally, collection of experimental SAS data for a stable mAb molecule up to elevated concentrations and different interacting

conditions is part of an ongoing study and will be the topic of an upcoming report. Such experimental data will facilitate parameterization and validation of a coarse-grained mAb model that will allow for accurate prediction of the solution behavior of antibodies at elevated concentrations. Finally, it is important to highlight that the above results and conclusions are solely based on steric repulsions, but they neglect both short- and long-range intermolecular attractions/repulsions. While these latter types of interactions are anticipated to dominate the macroscopic behavior of antibodies up to intermediate concentrations, the insights provided by the working models are representative of concentrated mAb solutions where molecular crowding due to excluded volume effects governs the solution properties.²⁹ Nonetheless, this work is a first step in the direction of developing accurate but simplified models of mAb solutions, and thus future reports will extend the present results to consider the role of other forces such as electrostatic and van der Waals interactions on the solution structure and macroscopic behavior of antibodies.

Acknowledgement

The authors would like to thank Nathan Mahynski and Monica Castellanos for many helpful and stimulating discussions, and Hongsuk Kang for revisions to the manuscript.

Supporting Information Available

See supporting information for additional details on the calculation of virial coefficients and scattering amplitudes for atomic structures, as well as for further results on SAS profiles of both rigid and flexible mAb models.

References

1. Smaglo BG, Aldeghaither D, Weiner LM. The development of immunoconjugates for targeted cancer therapy. Nat Rev Clin Oncol **2014;**11:637-648.

van Kasteren SI, Neefjes J, Ovaa H. Creating molecules that modulate immune responses. Nat Rev Chem
 2018.

3. Mullard A. 2017 FDA drug approvals. Nat Rev Drug Discov **2018**;17:81-85.

4. Cui Y, Cui P, Chen B, Li S, Guan H. Monoclonal antibodies: formulations of marketed products and recent advances in novel delivery system. Drug Dev Ind Pharm **2017;4**3:519-530.

5. Bye JW, Platts L, Falconer RJ. Biopharmaceutical liquid formulation: A review of the science of protein stability and solubility in aqueous environments. Biotechnol Lett **2014;**36:869-875.

6. Balcão VM, Vila MMDC. Structural and functional stabilization of protein entities: state-of-the-art. Adv Drug Deliv Rev **2015**;93:25-41.

1	
2	
3	
4	
5	
5	
0	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
17	
10	
19	
20	
21	
22	
23	
24	
25	
26	
27	
27	
20	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
30	
10	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
52	
55 F 4	
54	
55	
56	
57	
58	
59	

7. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of protein pharmaceuticals: An update. Pharm Res **2010;**27:544-575.

8. Wang Y, Lomakin A, Latypov RF, Benedek GB. Phase Separation in Solutions of Monoclonal Antibodies and the Effect of Human Serum Albumin. Proc Natl Acad Sci U S A **2011;**108:16606-16611.

9. Roberts CJ. Protein aggregation and its impact on product quality. Curr Opin Biotechnol **2014;**30:211-217.

10. Raut AS, Kalonia DS. Pharmaceutical Perspective on Opalescence and Liquid-Liquid Phase Separation in Protein Solutions. Mol Pharm **2016;**13:1431-1444.

11. Sarangapani PS, Weaver J, Parupudi A, Besong TMD, Adams GG, Harding SE, Manikwar P, Castellanos MM, Bishop SM, Pathak JA. Both Reversible Self-Association and Structural Changes Underpin Molecular Viscoelasticity of mAb Solutions. J Pharm Sci **2016**;105:3496-3506.

12. Schermeyer M-T, Wöll AK, Kokke B, Eppink M, Hubbuch J. Characterization of highly concentrated antibody solution - A toolbox for the description of protein long-term solution stability. mAbs **2017;**9:1169-1185.

13. Thompson RW, Latypov RF, Wang Y, Lomakin A, Meyer JA, Vunnum S, Benedek GB. Evaluation of effects of pH and ionic strength on colloidal stability of IgG solutions by PEG-induced liquid-liquid phase separation. J Chem Phys **2016**;145:185101.

14. Thiagarajan G, Semple A, James JK, Cheung JK, Shameem M. A comparison of biophysical characterization techniques in predicting monoclonal antibody stability. mAbs **2016**;8:1088-1097.

15. Corbett D, Hebditch M, Keeling R, Ke P, Ekizoglou S, Sarangapani P, Pathak J, Van Der Walle CF, Uddin S, Baldock C, Avendanio C, Curtis RA. Coarse-Grained Modeling of Antibodies from Small-Angle Scattering Profiles. J Phys Chem B **2017**;121:8276-8290.

16. Kastelic M, Vlachy V. Theory for the Liquid–Liquid Phase Separation in Aqueous Antibody Solutions. J Phys Chem B **2018**;122:5400-5408.

17. Calero-Rubio C, Saluja A, Roberts CJ. Coarse-Grained Antibody Models for "weak" Protein-Protein Interactions from Low to High Concentrations. J Phys Chem B **2016**;120:6592-6605.

18. Godfrin PD, Zarraga IE, Zarzar J, Porcar L, Falus P, Wagner NJ, Liu Y. Effect of Hierarchical Cluster Formation on the Viscosity of Concentrated Monoclonal Antibody Formulations Studied by Neutron Scattering. J Phys Chem B **2016**;120:278-291.

19. Chaudhri A, Zarraga IE, Yadav S, Patapoff TW, Shire SJ, Voth GA. The Role of Amino Acid Sequence in the Self-Association of Therapeutic Monoclonal Antibodies: Insights from Coarse-Grained Modeling. J Phys Chem B **2013**;117:1269-1279.

20. Sun G, Wang Y, Lomakin A, Benedek GB, Stanley HE, Xu L, Buldyrev SV. The phase behavior study of human antibody solution using multi-scale modeling. J Chem Phys **2016**;145:194901.

21. Chi EY, Krishnan S, Randolph TW, Carpenter JF. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. Pharm Res **2003;**20:1325-1336.

22. Ben-Naim A. On the driving forces for protein-protein association. J Chem Phys 2006;125 (2):024901.

23. Timasheff SN. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. Proc Natl Acad Sci USA **2002**;99 (15):9721-9726.

24. Cheung JK, Shen VK, Errington JR, Truskett TM. Coarse-Grained Strategy for Modeling Protein Stability in Concentrated Solutions. III: Directional Protein Interactions. Biophys J **2007**;92 (12):4316-4324.

25. Blanco MA, Shen VK. Effect of the surface charge distribution on the fluid phase behavior of charged colloids and proteins. J Chem Phys **2016**;145 (15):155102.

26. Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM. Toward a molecular understanding of protein solubility: Increased negative surface charge correlates with increased solubility. Biophys J **2012**;102:1907-1915.

27. Kamerzell TJ, Esfandiary R, Joshi SB, Middaugh CR, Volkin DB. Protein-excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. Adv Drug Deliv Rev **2011**;63:1118-1159.

28. Roberts CJ. Therapeutic protein aggregation: Mechanisms, design, and control. Trends Biotechnol **2014;3**2:372-380.

29. Minton AP. Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: Macromolecular crowding and protein stability revisited. Biophys J **2005**;88:971-985.

30. Tang KES, Bloomfield VA. Excluded volume in solvation: Sensitivity of scaled-particle theory to solvent size and density. Biophys J **2000;**79:2222-2234.

31. Hall D, Minton AP. Macromolecular crowding: Qualitative and semiquantitative successes, quantitative challenges. Biochim Biophys Acta, Proteins Proteomics **2003**;1649:127-139.

32. Miklos AC, Sarkar M, Wang Y, Pielak GJ. Protein crowding tunes protein stability. J Am Chem Soc **2011;**133:7116-7120.

33. Zhang X, Zhang L, Tong H, Peng B, Rames MJ, Zhang S, Ren G. 3D Structural Fluctuation of IgG1 Antibody Revealed by Individual Particle Electron Tomography. Sci Rep **2015**;5:9803.

34. Castellanos MM, Clark NJ, Watson MC, Krueger S, McAuley A, Curtis JE. Role of Molecular Flexibility and Colloidal Descriptions of Proteins in Crowded Environments from Small-Angle Scattering. J Phys Chem B **2016**;120:12511-12518.

35. Clark NJ, Zhang H, Krueger S, Lee HJ, Ketchem RR, Kerwin B, Kanapuram SR, Treuheit MJ, McAuley A, Curtis JE. Small-Angle Neutron Scattering Study of a Monoclonal Antibody Using Free-Energy Constraints. J Phys Chem B **2013**;117 (45):14029-14038.

36. Irani V, Guy AJ, Andrew D, Beeson JG, Ramsland PA, Richards JS. Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases. Mol Immunol **2015**;67:171-182.

1
2
3
4
5
6
7
/
8
9
10
11
12
13
14
15
16
17
18
10
עו 20
∠U 21
21
22
23
24
25
26
27
28
29
20
20 21
51
32
33
34
35
36
37
38
39
40
/1
יד ⊿ר
4∠ 12
43
44
45
46
47
48
49
50
51
52
52
57
54
55
56
57
58
59

37. Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: From structure to effector functions. Front Immunol **2014;5**:1-17.

38. Janeway CA, Travers P, Walport M, Shlomchik M. Immunobiology: The immune system in health and disease. 5th Ed.; New York: Garland Science, **2001**.

39. Stingaciu LR, Ivanova O, Ohl M, Biehl R, Richter D. Fast antibody fragment motion: Flexible linkers act as entropic spring. Sci Rep **2016;**6:1-13.

40. Sandin S, Öfverstedt LG, Wikström AC, Wrange Ö, Skoglund U. Structure and flexibility of individual immunoglobulin G molecules in solution. Structure **2004;**12:409-415.

41. Errington JR. Direct calculation of liquid–vapor phase equilibria from transition matrix Monte Carlo simulation. J Chem Phys **2003**;118:9915-9925.

42. Shen VK, Errington JR. Metastability and Instability in the Lennard-Jones Fluid Investigated by Transition Matrix Monte Carlo. J Phys Chem B **2004**;108:19595-19606.

43. Frenkel D, Smit B, *Understanding Molecular Simulations: from algorithms to applications*. 2nd ed.; New York: Academic Press, **2002**.

44. Hatch HW, Mittal J, Shen VK. Computational study of trimer self-assembly and fluid phase behavior. J Chem Phys **2015**;142:164901.

45. Vlugt TJH, Martin MG, Smit B, Siepmann JI, Krishna R. Improving the efficiency of the configurationalbias Monte Carlo algorithm. Mol Phys **1998;**94:727-733.

46. Vlugt TJH, Krishna R, Smit B. Molecular Simulations of Adsorption Isotherms for Linear and Branched Alkanes and Their Mixtures in Silicalite. J Phys Chem B **1999**;103:1102-1118.

47. Martin MG, Frischknecht AL. Using arbitrary trial distributions to improve intramolecular sampling in configurational-bias Monte Carlo. Mol Phys **2006**;104:2439-2456.

48. Blanco MA, Sahin E, Li Y, Roberts CJ. Reexamining protein-protein and protein-solvent interactions from Kirkwood-Buff analysis of light scattering in multi-component solutions. J Chem Phys **2011**;134 (22):225103.

49. Dumetz AC, Chockla AM, Kaler EW, Lenhoff AM. Protein phase behavior in aqueous solutions: Crystallization, liquid-liquid phase separation, gels, and aggregates. Biophys J **2008;**94:570-583.

50. Vliegenthart GA, Lekkerker HNW. Predicting the gas-liquid critical point from the second virial coefficient. J Chem Phys **2000**;112:5364-5369.

51. Ben-Naim A, *Statistical Thermodynamics for Chemists and Biochemists*. New York: Plenum Press, **1992**.

52. Singh J, Kofke D. Mayer Sampling: Calculation of Cluster Integrals using Free-Energy Perturbation Methods. Phys Rev Lett **2004;**92 (22):220601.

53. Shaul KRS, Schultz AJ, Kofke DA. Mayer-sampling Monte Carlo calculations of uniquely flexible contributions to virial coefficients. J Chem Phys **2011**;135 (12):124101.

54. Grünberger A, Lai PK, Blanco MA, Roberts CJ. Coarse-grained modeling of protein second osmotic virial coefficients: Sterics and short-ranged attractions. J Phys Chem B **2013**;117:763-770.

55. Blanco MA, Sahin E, Robinson AS, Roberts CJ. Coarse-Grained Model for Colloidal Protein Interactions, B₂₂, and Protein Cluster Formation. J Phys Chem B **2013**;117 (50):16013-16028.

56. Hansen J-P, McDonald IR, *Theory of Simple Liquids*. 3rd ed.; Amsterdam: Academic Press, 2006.

57. Blanco MA, Hatch HW, Curtis JE, Shen VK. A methodology to calculate small-angle scattering profiles of macromolecular solutions from molecular simulations in the grand-canonical ensemble. J Chem Phys **2018**; 149 (8):084203.

58. Watson MC, Curtis JE. Rapid and accurate calculation of small-angle scattering profiles using the golden ratio. J Appl Crystallogr **2013**;46:1171-1177.

59. Kline SR. Reduction and analysis of SANS and USANS data using IGOR Pro. J Appl Crystallogr **2006;**39:895-900.

60. Olds DP, Duxbury PM. Efficient algorithms for calculating small-angle scattering from large model structures. J Appl Crystallogr **2014;**47:1077-1086.

61. Blanco MA, Perevozchikova T, Martorana V, Manno M, Roberts CJ. Protein-Protein Interactions in Dilute to Concentrated Solutions: alpha-Chymotrypsinogen in Acidic Conditions. J Phys Chem B **2014;**118 (22):5817-5831.

62. Keshmirizadeh E, Modarress H, Eliassi A, Mansoori GA. A new theory for polymer/solvent mixtures based on hard-sphere limit. Eur Polym J **2003;**39:1141-1150.

63. Chang J, Sandler SI. An equation of state for the hard-sphere chain fluid-theory and Monte-Carlo simulation. Chem Eng Sci **1994;**49 (17):2777-2791.

64. Carnahan NF, Starling KE. Thermodynamic properties of a Rigid-sphere fluid. J Chem Phys **1970**;53:600.

65. Minton AP. The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. J Biol Chem **2001;**276:10577-10580.



Figure 1. Coarse-grained representation of the different antibody models considered here. (a) 4-bead mAb model. Antibodies are coarse-grained at the domain level, where the Fab, the Fc, and the hinge are represented by a single sphere each. (b) 7-bead mAb model. The different sub-domains on an antibody are modeled by a single sphere, and they correspond to the C_H3 , C_H2 , C_H1 , F_V , and hinge subdomains. (c) Crystal structure of the reference antibody (i.e., IgG2, pdb ID: 1IGT). For visualization purposes, coarsegrained models are overlaid over the structure of the reference antibody.

179x133mm (150 x 150 DPI)



Figure 2. Small-angle scattering profiles for the different mAb models as rigid-body representations under different internal configuration. (a) 4-bead mAb; and (b) 7-bead mAb. The black arrow in both panels indicates the direction in which domain-domain distances are decreased. These structures range from a configuration where Fab and Fc domains are at the farthest separation (red line) to that where all domains are in contact with each other (dark-blue line). <I>I(q)</I> profiles are shown at a concentration of 175 mg/mL, and they are normalized by the scattering intensity of a single molecule at q=0 (I_0). For comparison, graphical representations of the different mAb structures are also shown in the figures.

90x144mm (300 x 300 DPI)



Figure 3. Small-angle scattering profiles for the different mAb models under different levels of hinge flexibility. (a) 4-bead mAb; and (b) 7-bead mAb. The black arrow in both panels indicates the direction in which the degree of flexibility increases. The rigid configuration for both models corresponds to mAbs in the structure of the reference antibody (pdb: 1IGT). <I>I(q)</I> profiles are shown at a concentration of 175 mg/mL, and they are normalized by the scattering intensity of a single molecule at q=0 (I_0).

90x149mm (300 x 300 DPI)



Figure 4. Small-angle scattering profiles for the different mAb models at different mAb concentrations. (a) 4bead mAb; and (b) 7-bead mAb. In both panels, solid-lines represent fully flexible mAbs, while dashed-lines correspond to rigid mAbs (based on the structure of the reference antibody). Color-coding corresponds to different protein concentrations and is indicated in the legends of each panel. <I>I(q)</I> is normalized by the scattering intensity of a single molecule at q=0 (I_0) and by the average number of molecules <I> (N) in the scattering volume to facilitate comparison between profiles.

90x150mm (300 x 300 DPI)





Figure 5. Hinge-hinge radial distribution function ($<I>g_{HH}(r)</I>$) for the different coarse-grained mAb representations as a function of protein concentration. (a) 4-bead mAb; and (b) 7-bead mAb. Color-coding represents different mAb concentrations (20 – 180 mg/mL), and the black arrows indicate the direction in which concentration increases. In both panels, $<I>g_{HH}(r)</I>$ is shown for the rigid and fully flexible mAb configurations. For clarity, curves from the rigid configuration are shifted up by one unit.

90x128mm (300 x 300 DPI)



Figure 6. Osmotic pressure <I>P</I> as a function of mAb concentration for the different mAb models: (a) 4-bead mAb; and (b) 7-bead mAb. In both panels, blue solid-lines correspond to fully flexible mAbs, while red dashed-lines represent rigid mAbs (based on the structure of the reference antibody). For comparison, the pressure from hard-spheres with diameter <I> σ_{HS} </I> of 10 and 11 nm are also shown in the figures as dot-dashed lines. <I>P</I> is normalized by the pressure of an ideal-gas <I>P^{ig}=(Nk_BT)/V</I> at equivalent molecular concentration.

90x126mm (300 x 300 DPI)

Page 31 of 44

1	
2	
3 4	
5	
6 7	
8	
9 10	
11 12	
13	
14 15	
16	
17 18	
19	
20 21	
22	
23 24	
25 26	
20	
28 29	
30	
31 32	
33	
34 35	
36 37	
38	
39 40	
41	
42 43	
44 45	
45 46	
47 ⊿Ջ	
49	
50 51	
52	

Table 1. Excluded volume $B_{22,i}^{HS}$ and hard-sphere diameter σ_i for the different coarse-grained domains and subdomains.

Bead	$B_{22,i}^{HS}[nm^3]^{a,b}$	$\sigma_i[\mathbf{nm}]$
Hinge		1.52 ^c
Fc	527.87 ± 1.91	6.32
Fab	443.20 ± 0.26	5.96
F _v	208.13 ± 018	4.63
C _H 1	179.09 ± 0.06	4.41
C _H 2	316.83 ± 0.62	5.33
C _H 3	196.05 ± 0.14	4.54

^a Calculated via Mayer-Sampling from the crystal structure (i.e., all-atom representation) of the reference IgG2, where each atom is treated as a hard-sphere based on the Van der Waals radii. ^b Uncertainties of calculated values of $B_{22,i}^{HS}$ correspond to standard error over no less than four

independent Mayer-Sampling simulations.

lcura. ^c Diameter of the hinge region is calculated from the radius of gyration.

Trial	Probability
Rigid-body	0.35
translation	
Rigid-body rotation	0.20
Configuration-bias	0.15
Insertion or deletion ^{<i>a</i>}	0.30

Table 2. Monte Carlo trial moves and their probability of selection.

^{*a*} Insertion/deletion of molecules consists of the concatenation of Multi-first particle insertion and configurational-bias moves.

1 2 3	Evalua
4 5 6	Stru
7 8 9	Marco A. I
10 11 12	
13 14 15	
16 17 18	
19 20	National
21 22 23	Wattonat
24 25 26	
27 28 20	
30 31	
32 33 34	
35 36 37	
38 39 40	
41 42 42	
44 45	
46 47 48	
49 50 51	
52 53 54	
55 56	* Correspondence
57 58	Merck & Co., I

Evaluating the Effects of Hinge Flexibility on the Solution Structure of Antibodies at Concentrated Conditions

Marco A. Blanco,^{1,2,*} Harold W. Hatch,¹ Joseph E. Curtis,³ and Vincent K. Shen¹

¹Chemical Informatics Group, Chemical Sciences Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA ²Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850 ³NIST Center for Neutron Research,

National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

^{*} Correspondence to: Marco A. Blanco. E-mail address: mblanco@udel.edu; Current address to Blanco: Merck & Co., Inc., Kenilworth, NJ 07033, USA

I. SUPPORTING INFORMATION

A. Reference IgG2 Antibody

As model system for the design of the 4-bead and 7-bead mAb models, an IgG2 antibody (pdb: 1IGT) was used. In our coarse-grained models, each of the beads representing the different domains and subdomains are located at the center-of-mass of the corresponding mAb fragment (see Fig. 1 in the main text). These representations allow us to evaluate the different geometrical relations between domains/subdomains, within the context of the crystal structure of the reference antibody. Table S1 summarizes the bond distances and angles relating the Hinge to the different mAb domains/subdomains.

TABLE S1. Internal coordinates of the reference IgG2 antibody (pdb: 1IGT).

2-Body Relations		
Pair	Bond Distance [nm]	
Hinge–Fc	4.24	
Hinge–Fab-1	5.85	
Hinge–Fab-2	4.87	
$Hinge-C_H2$	2.98	
$Hinge-C_H1-1$	4.15	
$Hinge-C_H1-2$	3.42	
C_H2 – C_H2	2.83	
F_V -1– C_H 1-1	3.48	
F_V -2– C_H 1-2	3.15	
3-Body Relations		
Set	Angle [deg]	
Fc-Hinge-Fab-1	107.17°	
Fc–Hinge–Fab-2	108.98°	
Fab-1–Hinge–Fab-2	122.58°	
$C_{\rm H}2$ –Hinge– $C_{\rm H}1$ -1	122.31°	
$C_{\rm H}2$ –Hinge– $C_{\rm H}1$ -2	107.15°	
$C_{\rm H}$ 1-1–Hinge– $C_{\rm H}$ 1-2	98.30°	

B. All-Atom B_{22}^{HS} for Antibody Domains

Hard-sphere second virial coefficients for atomistic models of the mAb domains and subdomains ($B_{22,i}^{HS}$, with *i* denoting the evaluated domain or subdomain) were calculated using Mayer-Sampling Monte Carlo[1, 2] (MSMC), as described in the main text. Structures of the different antibody fragments were taken from the reference IgG2 antibody (pdb: 1IGT), as described in Table S2.

TABLE S2. Residue number for the amino acids forming the different mAb domains and subdomains. These ID numbers correspond to the internal identification of the reference antibody (pdb: 1IGT).

Antibody	Residue Number	
Fragment	Heavy Chain	Light Chain
Hinge	236-243	-
Fc	248-474	-
Fab	1-229	1 - 214
F_{v}	1–112	1 - 108
$C_{\rm H}1$	113–229	109 - 214
$C_{H}2$	248 - 360	-
$C_H 3$	361 - 474	

For MSMC simulations, interactions between atoms were represented by a hard-sphere potential. As such, atomic sizes were taken from the corresponding Van der Waals diameters,[3] and they are given by 0.37, 0.04, 0.33, 0.31, and 0.39 nm for C, H, N, O, and S, respectively. The reference system, required for MSMC calculations, consisted of a single hard-sphere particle located in the center-of-mass of the evaluated (sub)domain, with a diameter of 3 nm. The resulting values of $B_{22,i}^{HS}$ are shown in Table 1 in the main text.

C. Computational Method for 7-bead Model

For the 7-bead model simulations, five Monte Carlo trials were employed. These trials included rigid-body translations, rigid-body rotations, grand canonical insert or deletion with configurational bias, particle regrowth with configurational bias, and configurational

swaps between processors. The first four move types were attempted with equal probability of 0.24999375. The configuration swap was attempted with a probability of 2.5×10^{-5} as described in Ref. 4. For the grand canonical configurational-bias insertions and deletions, the hinge bead was first placed with 10 multiple first bead insertions. Then one of the arms was selected, and the bead furthest from the hinge was placed in 10 different trial positions. Then the bead on the same arm closer to the hinge was placed with a single trial. After this, the two beads of the other arms were placed that are furthest from the hinge simultaneously, in 10 trials. Finally, the two remaining beads were placed with a single trial. In the regrowth configurational bias trial, the hinge and one of the three branches was selected to remain, while the other two arms were regrown in a similar fashion as described above. To initialize the simulations, Wang Landau simulations were performed for 25 flatness checks, as described in Ref. 4. For the last 5 flatness checks, a collection matrix was initialized. This collection matrix was then used to initialize a transition matrix Monte Carlo simulation for 40 sweeps. The cubic periodic boundaries were 90 nanometers in size. The number of particles were constrained to be between 0 and 636. This density range was divided into 12 windows with a width scaled by an exponent of 1.5 for larger windows on the lower density. For the scattering calculation, histograms were computed every 1000 trials, with a 0.05nm bin spacing.

D. Domains and Subdomains Scattering Amplitudes

Scattering amplitudes (f_i) for the different domains and subdomains are calculated from the crystal structure of the reference IgG2. Briefly, f_i is calculated as [5]

$$f_i = \frac{1}{2} \sum_m b_m \int_0^\pi \cos\left[qr_m \cos\phi\right] \sin\phi d\phi \tag{D.1}$$

where b_m is the scattering contribution from each atom. For simplicity, b_m is assumed as the scattering length for neutrons, and it can take the values 6.651, 9.40, 5.804, 2.847, and -3.742 fm for C, N, O, S, and H, respectively [6].

In eq. D.1, the summation goes over all atoms in the evaluated structure, while the integral evaluates all possible relative orientations between the wavevector \mathbf{q} and the position $\mathbf{r_m}$ for each atom. As such, q is the magnitude of the wavevector, r_m is the distance of the *m*-th atom to the geometric center of the molecule, and ϕ is the angle form between both vectors.



FIG. S1. Scattering amplitudes f_i for the mAb domains and subdomains used to construct the different working coarse-grained models: (a) 4-bead mAb model; and (b) 7-bead mAb model.

Figure S1 shows the resulting scattering amplitudes.

Following eq. D.1 and the values of b_m , the q = 0 scattering intensity for a single molecule (I_0) can be calculated as $I_0 = \sum_m b_m^2$.

E. Rigid-Body mAbs

TABLE S3. Excluded volume for different mAb models and structures as measured by the second virial coefficient B_{22} .

4-bead	mAb	
Configuration	$\mathbf{B_{22}} \; [\mathrm{nm}^3]$	$\sigma_{\rm eff}~[\rm nm]$
Rigid-1IGT	2816.2 ± 4.4	11.04_{1}
Rigid–Closed contact	2376.5 ± 6.6	10.43_2
Rigid–Fc-Fab contact	2568.6 ± 6.0	10.70_2
Rigid–Fab-Fab contact	2727.1 ± 4.9	10.92_1
Rigid-Extended	2922.8 ± 5.7	11.17_{2}
Fully Flexible	2863.4 ± 2.8	11.10_{1}
7-bead	mAb	
Configuration	$\mathbf{B_{22}} \; [\mathrm{nm}^3]$	$\sigma_{\rm eff}$ [nm]
Rigid–1IGT	2821.5 ± 0.2	11.044_0
Rigid–Closed contact	2694.0 ± 0.8	10.875_1
Rigid–Extended	2881.2 ± 0.5	11.121_{1}
Angle Flexible	2800.9 ± 0.3	10.0170
Fully Flexible	2446.2 ± 0.8	10.531_{1}

1	
I	
2	
З	
2	
4	
5	
6	
2	
7	
8	
~	
9	
1	0
1	1
1	2
1	3
1	4
1	5
1	6
1	7
1	8
1	9
2	ი
~	0 4
2	1
2	2
2	2
2	З
2	4
2	5
2	2
2	0
2	7
2	R
-	v
~	~
2	9
2 3	9 0
2 3 3	9 0 1
2 3 3	9 0 1
2 3 3 3	9 0 1 2
2 3 3 3 3	9 0 1 2 3
2 3 3 3 3 3 3 3	9 0 1 2 3 4
2 3 3 3 3 3 3 3 3	9 0 1 2 3 4 5
2 3 3 3 3 3 3 3 3 3 3 3 3	9012345c
2 3 3 3 3 3 3 3 3	9 0 1 2 3 4 5 6
2 3 3 3 3 3 3 3 3 3 3 3 3 3 3	901234567
2333333333333	9012345678
2 3 3 3 3 3 3 3 3 3 3 2 2	90123456780
23333333333333	90123456789
2 3 3 3 3 3 3 3 3 3 4	901234567890.
2 3 3 3 3 3 3 3 3 3 4 4 .	90123456789012
233333333334444	901234567890122
2 3 3 3 3 3 3 3 3 3 4 4 4 4	901234567890123
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4	9012345678901234
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4	90123456789012345
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4	901234567890123456
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4	9012345678901234567
2 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4	90123456789012345678
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4	901234567890123456789
2 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4	9012345678901234567890
2333333333344444444444455	90123456789012345678901
233333333334444444444455	90123456789012345678901
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 5 5	901234567890123456789012
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4	9012345678901234567890123
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4	9012345678901234567890123
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4	90123456789012345678901234
2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4	901234567890123456789012345

- 56 57
- 58
- 59





FIG. S2. Small-angle scattering profiles for the 4-bead mAb models as a rigid-body representation at different protein concentrations: (a) 1 mg/mL; (b) 10 mg/mL; and (c) 100 mg/mL. Each curve corresponds to mAb structures with different separation distances between Fc and Fab domains. These configurations are controlled by the set of angles $\{\theta_1, \theta_2\}$, which are listed in each panel. θ_1 corresponds to the angle between Fab-hinge-Fc, and it is assumed to be identical for both Fab domains. θ_2 is the angle for Fab-hinge-Fab.





FIG. S3. Small-angle scattering profiles for the 7-bead mAb models as a rigid-body representation at different protein concentrations: (a) 1 mg/mL; (b) 10 mg/mL; and (c) 100 mg/mL. Each curve corresponds to mAb structures with different separation distances between Fc and Fab domains.

F. Flexible mAbs

tor per period



FIG. S4. Small-angle scattering profiles for the flexible 4-bead mAb models at different protein concentrations: (a) 1 mg/mL; (b) 10 mg/mL; and (c) 100 mg/mL. Each curve corresponds to mAb structures with different levels of flexibility.





FIG. S5. Small-angle scattering profiles for the flexible 7-bead mAb models at different protein concentrations: (a) 1 mg/mL; (b) 10 mg/mL; and (c) 100 mg/mL. Each curve corresponds to mAb structures with different levels of flexibility.

- [1] J. Singh and D. Kofke, Phys. Rev. Lett. (2004).
- [2] K. R. S. Shaul, A. J. Schultz, and D. A. Kofke, J. Chem. Phys. 135 (2011).
- [3] A. Grünberger, P. K. Lai, M. A. Blanco, and C. J. Roberts, J. Phys. Chem. B 117, 763 (2013).
- [4] H. W. Hatch, J. Mittal, and V. K. Shen, J. Chem. Phys. **142** (2015).
- [5] D. Corbett, M. Hebditch, R. Keeling, P. Ke, S. Ekizoglou, P. Sarangapani, J. Pathak, C. F. Van Der Walle, S. Uddin, C. Baldock, C. Avendaño, and R. A. Curtis, J. Phys. Chem. B 121, 8276 (2017).

[6] B. Jacrot, Rep. Prog. Phys. **39**, 911 (1976).