A multi-scale numerical approach to study monoclonal antibodies in solution

Special Collection: Structure and Mechanics of Biofluids, Biomaterials, and Biologics

Marco Polimeni ; Emanuela Zaccarelli ; Alessandro Gulotta ; Mikael Lund ; Anna Stradner ;
Peter Schurtenberger ✉

Check for updates

View Online

Export Citation

03 June 2024 06:29:54

APL Bioengineering

Special Topic:

Bioengineering of the Brain

Guest Editors: Roberto Portillo Lara, Elena De Momi, Giuseppe Baselli

Submit Today!

AIP Publishing

AIP Publishing

A multi-scale numerical approach to study monoclonal antibodies in solution

View Online   Export Citation   CrossMark

Marco Polimeni,[1,a)] (iD) Emanuela Zaccarelli,[2,b)] (iD) Alessandro Gulotta,[1] (iD) Mikael Lund,[3,c)] (iD) Anna Stradner,[1] (iD) and Peter Schurtenberger[1,d)] (iD)

AFFILIATIONS

[1]Division of Physical Chemistry, Lund University, Lund, Sweden
[2]Institute for Complex Systems, National Research Council (ISC-CNR), Piazzale Aldo Moro 5, 00185 Rome, Italy
[3]Division of Computational Chemistry, Lund University, Lund, Sweden

Note: This paper is part of the special issue on Structure and Mechanics of Biofluids, Biomaterials, and Biologics.
[a)]Electronic mail: marco.polimeni@fkem1.lu.se
[b)]Also at: Department of Physics, Sapienza University of Rome, Piazzale Aldo Moro 2, 00185 Rome, Italy.
[c)]Also at: LINXS Institute of Advanced Neutron and X-ray Science, Lund University, Lund, Sweden.
[d)]Author to whom correspondence should be addressed: peter.schurtenberger@fkem1.lu.se

ABSTRACT

Developing efficient and robust computational models is essential to improve our understanding of protein solution behavior. This becomes particularly important to tackle the high-concentration regime. In this context, the main challenge is to put forward coarse-grained descriptions able to reduce the level of detail, while retaining key features and relevant information. In this work, we develop an efficient strategy that can be used to investigate and gain insight into monoclonal antibody solutions under different conditions. We use a multi-scale numerical approach, which connects information obtained at all-atom and amino-acid levels to bead models. The latter has the advantage of reproducing the properties of interest while being computationally much faster. Indeed, these models allow us to perform many-protein simulations with a large number of molecules. We can, thus, explore conditions not easily accessible with more detailed descriptions, perform effective comparisons with experimental data up to very high protein concentrations, and efficiently investigate protein–protein interactions and their role in phase behavior and protein self-assembly. Here, a particular emphasis is given to the effects of charges at different ionic strengths.

I. INTRODUCTION

Therapeutic proteins, such as monoclonal antibodies (mAbs), represent an important tool for the treatment of numerous diseases, such as cancer,[1] autoimmune disorders,[2] and viral infections.[3] An example is given by the anti-SARS-CoV-2 mAbs, recently authorized by the Food and Drug Administration (FDA), to be used for the treatment of COVID-19 symptoms, but more than 100 mAbs have already been approved[4] in the last 35 years and many more are under investigation.[3] However, antibody-drug development faces several challenges,[5] such as protein aggregation, high viscosity, and phase separation. Protein aggregation, especially in the form of insoluble aggregates or high-molecular-weight species, can lead to reduced efficacy and potential immunogenicity.[6] High viscosity and phase separation can pose challenges during manufacturing and storage processes.[7]

Therefore, understanding and controlling the mechanisms driving these phenomena is of fundamental importance. Several experimental techniques, such as static light scattering (SLS), dynamic light scattering (DLS), and small-angle x-ray scattering (SAXS), are commonly used to study these systems. SLS provides information about the size and molecular weight of aggregates or oligomers present in the solution. Furthermore, by analyzing the scattering pattern at different protein concentrations, SLS gives insight into protein–protein interactions and solution stability in terms of the osmotic compressibility κ_T and the second osmotic virial coefficient, b_{22}. DLS measures the intensity fluctuations of the scattered light caused by the Brownian motion of particles in a solution and is commonly used to determine the collective diffusion coefficient, the hydrodynamic radius, and the size distribution of monomers and aggregates. SAXS involves exposing a sample

03 June 2024 06:29:54

to x rays and analyzing the scattering pattern produced by the interaction of x rays with the sample. Therefore, SAXS provides structural information at much smaller length scales than SLS and gives valuable insights into the overall shape, size, conformational changes, and intermolecular interactions of proteins in solution. The combination of these approaches is usually adopted by taking a systematic variation of the formulation components, such as buffer conditions, temperature, pH, ionic strength, excipient concentrations, and surfactants, to identify the optimal conditions for mAb stability and solubility. Ideally, by monitoring changes in the aggregation state and size distribution of the aggregates, one can aim to optimize formulation conditions.

Although laboratory experiments are essential to investigate protein solutions, computational modeling is an invaluable tool to make progress in a multi-scale, challenging problem such as the study of antibody behavior in concentrated solutions. The combined use of computer simulations at different length scales can, thus, deliver important predictions on the best solution conditions. In particular, atomistic models combined with molecular dynamics (MD) and/or Monte Carlo (MC) simulations provide details on protein conformational changes and can be used to access time scales and length scales not directly observable through experiments. Furthermore, more coarse-grained models can be used to speed up the simulations and study the system at extended length and time scales and higher protein concentrations. Overall, computational modeling allows us to have control over properties, such as temperature, pressure, pH, and protein concentration. This level of control enables the investigation of the effects of specific conditions on protein behavior and helps to elucidate the underlying mechanisms and driving forces governing the solution properties. This strategy can be used to make predictions on structural and dynamic quantities inaccessible through experiments alone. Furthermore, computational modeling can be more cost-effective and time-efficient than certain experimental techniques. Performing experiments, especially those requiring specialized equipment, difficult sample preparation procedures, and large sample volumes, can be expensive and time-consuming. Furthermore, in the early stages of biologics development, the available quantities of material are usually limited, which makes systematic studies of the solution properties at high concentrations difficult or impossible. Computational modeling,

thus, allows for virtual experiments and exploration of a wide range of conditions without physical materials and laboratory resources.

In a recent work,[8] we investigated the solution properties of an IGg1 mAb, whose behavior is largely controlled by repulsive electrostatic interactions, by means of several experimental techniques (SLS, DLS, SAXS, and microrheology) and colloidal theory. For this system, we have shown that, even though the mAb charge distribution is heterogeneous, the dominant electrostatics is controlled by the mAb high net positive charge at low protein concentrations and low ionic strength. This suggested the use of a simple colloidal model for this mAb, where the protein is represented by a hard sphere with a homogeneous surface charge density resulting in an effective charge Z_{eff} and an additional weak short-range attraction. The model was, indeed, capable of reproducing experimental properties, such as the apparent molecular weight or the collective diffusion coefficient over a large range of mAb concentrations and at different ionic strengths. However, this model failed to describe the more detailed structural properties of the mAb solution given by the static structure factor $S(q)$ from SAXS for mAb concentrations higher than 50 mg/mL, because of the non-realistic charge treatment and the missing shape anisotropy of the underlying colloid model. We indeed showed in Ref. 8 that, to this end, the mAb shape anisotropy is needed to explore a wider range of solution conditions as previously done in several works.[9–12]

Inspired by these considerations, in the present paper, we aim to overcome the limitations of using too simplified models while avoiding the computational costs of near atomistic simulations and develop the basis for a general protocol leading to efficient models able to describe solutions at high protein concentrations and high ionic strength. To do so, we go one step forward with respect to previous studies by working across different length scales. We, thus, use a multi-scale approach, described in detail below, which involves all-atom MD simulations, coarse-grained MC simulations, and mesoscale models, as shown in Fig. 1. By varying the mAb concentration, pH, and added salt, we explore several solution conditions and, for each of them, we can quantitatively reproduce the experimentally measured effective structure factor, $S_{eff}(q)$, obtained by SAXS. Such a comparison is highly non-trivial because it provides microscopic information about the spatial arrangement, conformational changes, and specificity of the
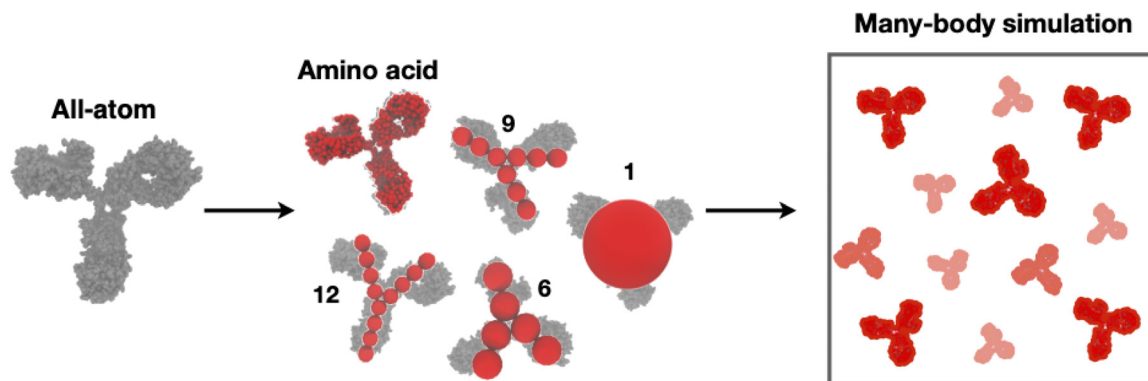
FIG. 1. The experimental mAbs structure at atomic resolution (leftmost, gray) is first relaxed using MD simulations, where-after it is coarse-grained at five different levels shown in the middle, in red. This includes models with 1, 6, 9, or 12 interaction sites, as well as a model where each amino acid is treated as a (charged) sphere. The coarse-grained models are subsequently used in many-body MC simulations to explore the effect of varying mAb concentration and salt concentrations (rightmost).

interacting proteins, contributing to a better understanding of the underlying molecular interaction mechanisms. Analyzing these results, we rationalize the mAb behavior as a function of the different solution conditions. The starting point for our simulations is the all-atom conformation obtained with homology modeling using a commercial software package [molecular operating environment (MOE),[13] see Sec. IV]. From this structure, we first build a coarse-grained version of the mAb at an amino-acid level and then perform constant-pH MC simulations to estimate the amino-acid charge distribution. Once we know the protonation state of each amino acid for the investigated conditions, we go on to perform all-atom MD simulations of a single mAb in explicit water and added salt to select characteristic molecular conformations to be used in many-protein MC simulations to study protein–protein interactions in more detail. Based on the information collected at both all-atom and amino-acid levels, we finally construct mesoscopic models made of simple beads which, while maintaining the Y-shaped geometrical anisotropy of the mAb, can be used to study the effect of coarse-graining on the resulting solution structure more efficiently. We, then, systematically vary the number of beads in the model to assess the optimal description of the molecules as a balance of accuracy and numerical cost. A schematic representation of the models used is presented in Fig. 1. The outcomes of these simulations are finally compared with SAXS data, as well as with the predictions from a simple and frequently used colloid model, to provide guidelines on the effects of coarse-graining on the resulting best-fit parameters used to define the potential of mean force between the mAbs.

II. RESULTS AND DISCUSSION

A. Calculation of amino acid charges

By using the one-protein MC simulations at the amino acid level, we estimate the mAb net charge, Q_{net}, the mAb charge distributions, and the mAb electrostatic iso-surface at pH 6 and either 7 or 57 mM ionic strength. As described in Sec. IV, Q_{net} is equilibrated using the titration moves. To start with, each amino acid has a negative charge. After a few steps of equilibration, the net charge quickly reaches a plateau, as reported in Fig. S1 of the supplementary material. The average values of the net charge of the mAb obtained discarding the first 200 equilibration steps are $Q_{net} = 31.3e$ for the $I = 7$ mM condition and $Q_{net} = 36.3e$ for $I = 57$ mM. This difference is expected due to salt screening effects which allow for a higher net-charge. In addition, we obtain the individual protonation state of the amino acid charge distribution that is then used as input in the all-atom MD simulations. In Figs. S2–S5, the charge iso-surface of the mAb is reported, illustrating the overall positive charge of the molecule. We remark on the different electrostatic properties of this antibody compared to previously studied mAbs,[14] which results in a very different concentration behavior dominated by repulsive interactions. This feature is partially mitigated at high ionic strength where, despite an overall higher net charge, a larger degree of heterogeneity among the charges at the amino-acid level is estimated due to screening effects in different regions of the protein. This may also explain why self-assembly and an increased viscosity were observed experimentally at high ionic strength and very high mAb concentrations (see Ref. 8).

B. mAb conformations obtained from MD simulations

Next, we carry out all-atom MD simulations to obtain representative conformations of the mAb at both studied ionic strengths. In Fig. 2,

we report the root-mean-squared deviation (RMSD) from the initial conformation, and the radius of gyration, R_g, for the two investigated ionic strengths, I. For $I = 7$ mM, the RMSD converges approximately after 800 ns (part highlighted in red). Similarly, for $I = 57$ mM, the convergence starts approximately after 600 ns. Using the algorithm in Daura et al.[15] the two MD trajectories are reduced into 11 and 20 groups of conformations, respectively, for the conditions $I = 7$ and $I = 57$ mM. The most probable conformations of these two groups are shown in Fig. 3 (center and rightmost) together with their representation at the amino acid level. We can see that in both cases, the average conformations are more compact compared to the initial structure obtained from homology modeling (see Fig. 3, leftmost). This is also manifested in their low R_g value, which is found to be ∼4.8 nm for the condition at 7 mM and slightly lower, ∼4.5 nm, for the condition at 57 mM. It is interesting to compare the calculated form factors for the initial structure and the most probable conformations at low and high ionic strengths (both coarse-grained at the amino acid level) with experimental SAXS data obtained at low mAb concentrations ($C_p = 2$ mg/mL), where structure factor effects should be minimal. The corresponding data are shown in Fig. S11 in the supplementary material and demonstrate that the agreement between the form factor from the relaxed structure and the experimental data is very good at low ionic strength except for very low q-values, where some non-negligible contributions from $S(q)$ exist even at this low mAb concentration. Quite in contrast, the calculated form factor from the initial structure is in clear disagreement with the measured data, in particular, at q-values around $0.05 - 1$ nm^{-1}. For the higher ionic strength, the most probable conformation results in slightly less good agreement, in particular, at lower q-values, indicating that the simulated structure is slightly too compact, but the agreement with the experimental data is still much better than for the initial structure.

C. Effective structure factor calculations from many-protein MC simulations at the amino-acid level

Using the two most representative conformations obtained from the MD simulations, we build a coarse-grained representation at an amino-acid level and perform many-protein MC simulations. We can, thus, calculate the $S_{eff}(q)$ of the mAb at different concentrations to be compared with experimental results. This is reported in Fig. 4 for the two series of simulations carried out at the two investigated ionic strengths. For each condition, we explored the effect of increasing the attractive well of Lennard–Jones potential, ε_{ij}, from $0.05 k_B T$, which is commonly used in coarse-grained simulations at the amino acid level as a starting value, to $0.085 k_B T$. For the low ionic strength condition, left panels of Fig. 4, the simulation data obtained with the two most probable conformations show a very good quantitative agreement with the SAXS data at all studied mAb concentrations. We further note that the effect due to the increase in ε_{ij} is almost negligible at low C_p where all replicas, calculated for different values of ε_{ij}, yield very similar results among each other. Hence, under these conditions, $S_{eff}(q)$ is completely dominated by long-range Coulomb interactions. This was expected, as by considering the electrostatic isosurface potential shown in Figs. S2 and S3 of the supplementary material it is clear that despite the mAb heterogeneous charge distribution the predominant influence is from the positive charges. Consequently, when other mAbs approach, they will encounter a relatively spherical $+1 k_B T/e$ repulsive
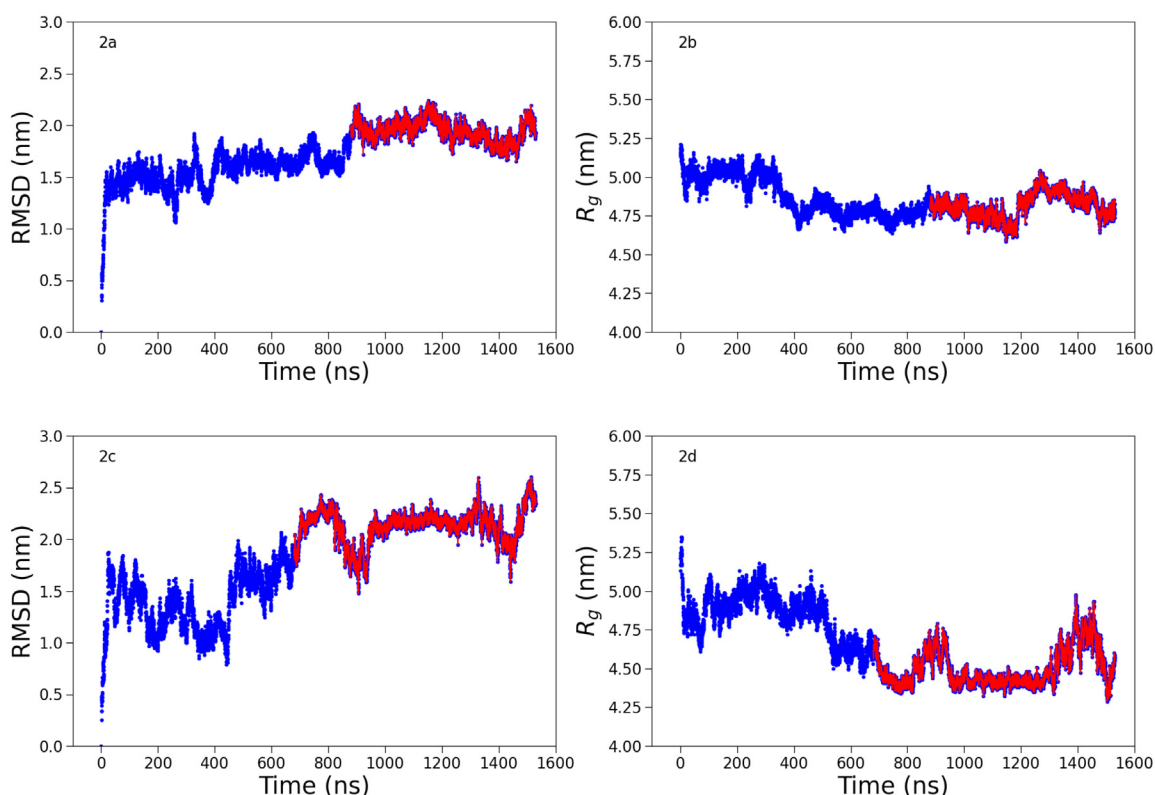
FIG. 2. Root mean squared deviation (RMSD) and radius of gyration (R_g) obtained from all-atom MD simulations for the condition I = 7 mM [(a) and (b)] and I = 57 mM [(c) and (d)]. The equilibrated part of the trajectory over which the average R_g is calculated is highlighted in red.

isopotential surface, resulting in a quite globular centrosymmetric and soft repulsive interaction potential.

At high mAb concentrations, an increase in ε_{ij} has a visible effect as the different replicas give slightly different results. For the higher ionic strength (57 mM), right panels of Fig. 4, the first thing to notice is that the measured $S_{eff}(q)$ are overall less repulsive, compared to the low salt condition, as visible from the significant increase in the $q \rightarrow 0$ limit, that is directly related to the osmotic compressibility of the

suspension. In addition, a small change in the attraction strength leads to a significant difference in the resulting $S_{eff}(q)$ even at the lowest studied mAb concentration (10 mg/mL). This effect increases with increasing C_p. Under these concentrated conditions, the agreement between experiments and simulations is still good but it starts to deteriorate, especially in the low q limit. This could be due to the need to introduce flexibility in the model or to the absence of explicit counterions that could effectively change the screening conditions upon
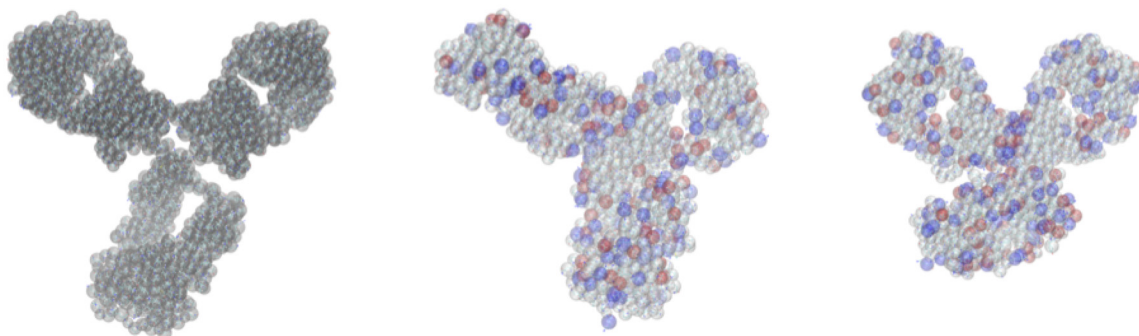
FIG. 3. mAb structure before and after MD simulations. At the leftmost, the initial mAb structure at atomic resolution overlaps with its coarse-grained representation at the amino acid level. On the center and rightmost, the representative mAb conformations obtained from MD simulations, respectively, for the conditions at 7 and 57 mM of ionic strength. Red, blue, and gray beads represent negatively charged, positively charged, and neutral amino acids, respectively.
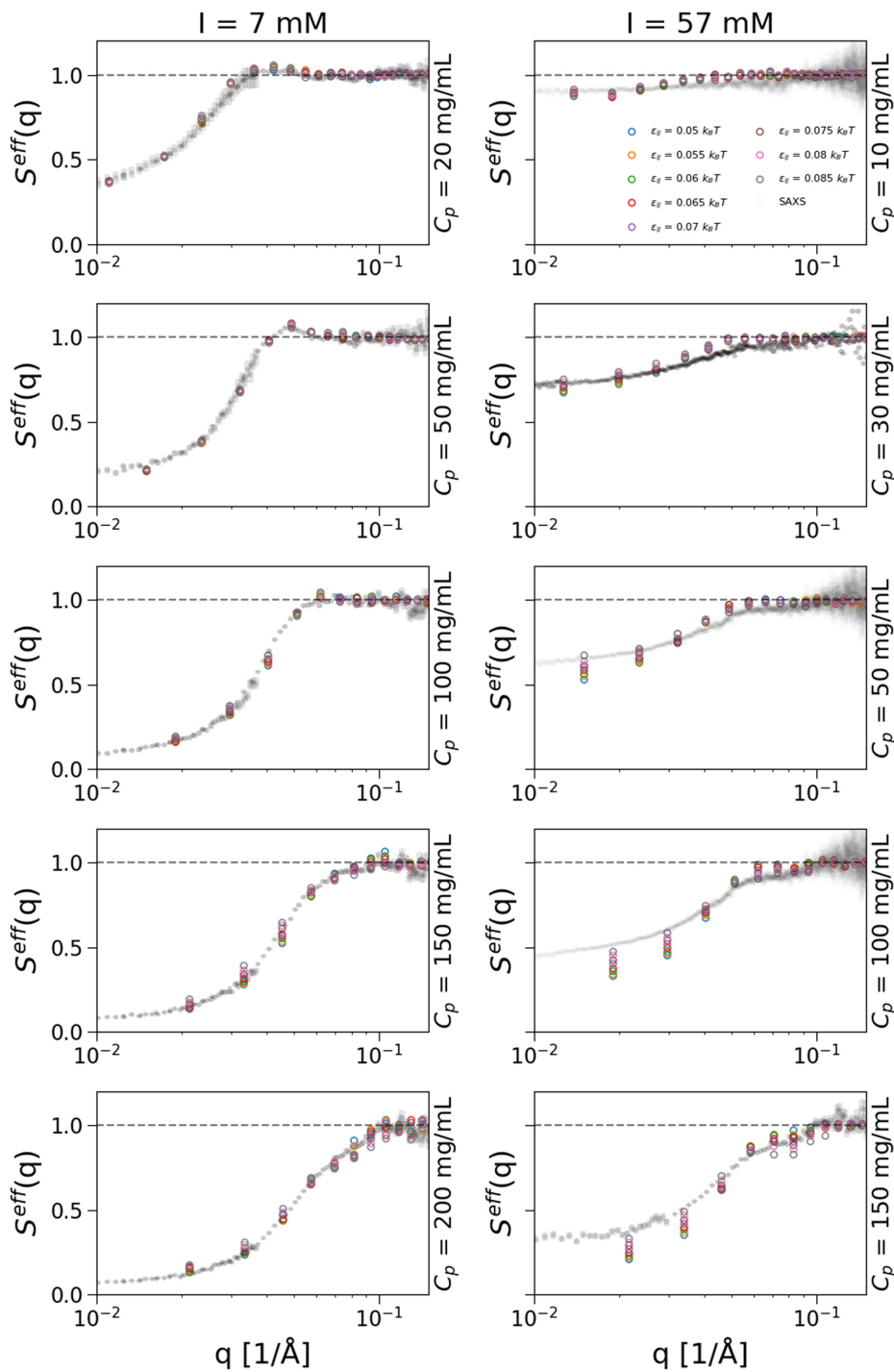
FIG. 4. Effective structure factors, $S_{eff}(q)$, obtained from simulations of $N_p = 20$ mAbs coarse-grained at an amino acid level in comparison with experimental data. The left and right columns refer to, respectively, the 7 and 57 mM ionic strength conditions, while on each graph the mAb concentration is indicated. The gray circles in the background represent the experimental data obtained by SAXS experiments together with their error bars, while the open symbols are the simulation results. Each of the colors indicates a different value of the depth of the attractive well in the Lennard–Jones term of Eq. (3), ε_{ij}. The explored range is $\Delta\varepsilon_{ij} = 0.05$–$0.085\,k_BT$ with steps of $0.005\,k_BT$.

reversible interactions with the mAbs at this high ionic strength. In Fig. 5, we compare experimental results with simulation results obtained either using the initial non-relaxed PDB conformation or the relaxed conformation obtained by the all-atom MD simulations. The comparison is reported for two representative cases, respectively, at low (20 mg/mL on the left) and high mAb concentration (150 mg/mL on the right), both for the low ionic strength of 7 mM. For low C_p, we find quantitative agreement with experimental data by using either the relaxed or the non-relaxed conformation. This result is likely because the mAb carries a high positive net charge and a relatively
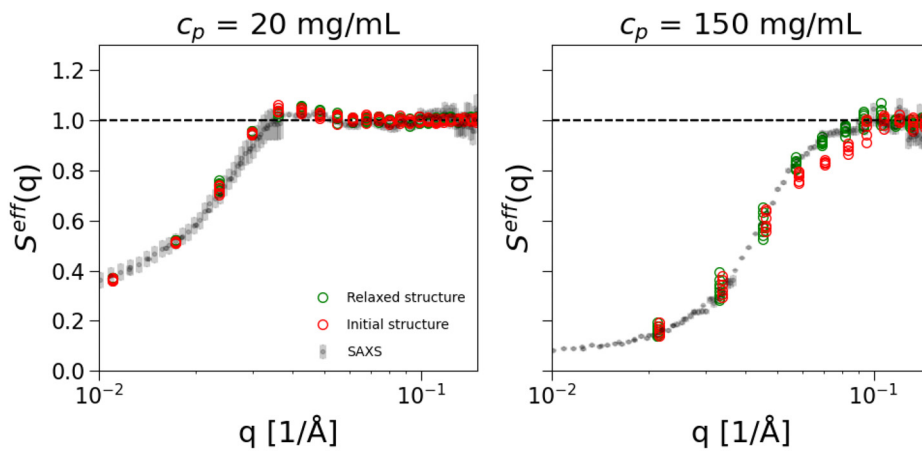
FIG. 5. Effective structure factors, $S_{eff}(q)$, obtained from simulations of $N_p = 20$ mAbs coarse-grained at an amino acid level and at low ionic strength (7 mM). Comparison with initial and relaxed structure. The explored range is $\Delta\varepsilon_{ij} = 0.05$ $-0.085 k_B T$ with steps of $0.005 k_B T$ for both cases.

homogeneous charge distribution as demonstrated with the electrostatic potential iso-surfaces shown in Fig. S2 in the supplementary material. Therefore, electrostatic repulsions prevent contribution from short-range interactions where the actual mAb conformation plays a role. However, the situation is different at high mAb concentration, where, by using the initially not-relaxed conformation, $S_{eff}(q)$ is significantly underestimated in the q-range corresponding to the main (or nearest neighbor) peak, i.e., when mAbs are close to each other, suggesting that this conformation is overestimating the mAbs excluded volume. It is interesting to note that for these q-values around $0.05 - 1$ nm^{-1} we also find significant discrepancies between the calculated form factor of the initial conformation and the measured SAXS data at low mAb concentrations (Fig. S11 in the supplementary material). These results put forward that an appropriate treatment of initial structures of the mAbs, although computationally demanding, is needed to correctly take into account the excluded volume, particularly at contact when attractive interactions play a role. This is confirmed by the fact that at high ionic strength, the comparison with experimental results further deteriorates in quality (not shown), which is the primary reason why we carried out all-atom MD simulations. On the other hand, a reassuring result is that, by using the most probable conformation, the coarse-grained many-body simulations are capable of reproducing the experimental results under the present conditions, except for the small discrepancy at high I and high C_p discussed above.

D. Effective structure factor calculation with bead models

By using a model at an amino acid level for the mAb representation, we have so far obtained a relatively accurate description of the interactions at play in the solution. However, due to the computational cost, we are limited in the number of mAbs we can simultaneously simulate, which is also directly related to the lowest q-value we can sample. To overcome this limitation, we further coarse-grain the mAb using the bead models shown in Fig. 1. For each of these models, we then perform many-protein MC simulations with different values for the effective charge, Q_{eff}, and ε_{ij}.

In Fig. 6, we report for each of the studied models with different numbers of beads (1-bead, 6-bead, 9-bead, and 12-bead), the χ^2 tables, measuring the deviation of the calculated $S_{eff}(q)$ from the

experimental ones, to assess the ability of the chosen potential parameters (Q_{eff}, ε_{ij}) to reproduce the experimental data at low mAb concentration (20 mg/mL) and low ionic strength (7 mM). Each χ^2 value is defined as

$$\chi^2 = \sum_i \frac{(S_{eff}^{sim}(q_i) - S_{eff}^{exp}(q_i))^2}{S(q_i)_{eff}^{exp}}, \qquad (1)$$

where $S(q_i)_{eff}^{sim}$ and $S(q_i)_{eff}^{sim}$ refer to the effective structure factor obtained, respectively, from simulations and SAXS experiments, and calculated in the point $q = q_i$. The number of q points goes from $q_0 = 0.01$ to $q_{50} = 0.12$ Å^{-1} with a Δq between each point equal to 0.0022 Å^{-1}.

We also report the tables for all the other studied conditions in the supplementary material for completeness (Figs. S6–S9). These tables provide us with a graphical indication of the uniqueness of a given combination of Q_{eff} and ε_{ij} to best reproduce the experimental data, where the lower the χ^2 value, the better that particular combination of charge and attractive strength reproduces the experimental data. Figure 7 clearly shows that while Q_{eff} is well defined under these conditions of low C_p and ionic strength, ε_{ij} has no systematic effect and can, thus, not be determined accurately for these solutions. Moreover, Q_{eff} strongly depends on the model. For the 1-bead model, the value of Q_{eff} best reproducing the experimental results is found to be in the range $\Delta Q_{net} = 20 \pm 1.5e$ (see Fig. 6), while for the 6-bead model, this value shifts to $26 \pm 2e$ (see Fig. 7).

The $S_{eff}(q)$ obtained with these effective charges in combination with $\varepsilon_{ij} = 0.8 k_B T$ are reported in Fig. 7 for the conditions at low ionic strength (7 mM) and both low (20 mg/mL) and high (150 mg/mL) mAb concentrations. At low mAb concentration (Fig. 7, left), both the 1-bead and the six-bead models are in quantitative agreement with the experimental data. However, the net charge we have to assign to match the experimental data is significantly smaller than that expected from the model at the amino acid level and our constant-pH Monte Carlo simulations. At high mAb concentration (Fig. 7, right), instead the 1-bead model is unable to reproduce the experimental behavior. Indeed, $S_{eff}(q)$ takes the shape that is typical for hard-sphere charged colloids at low ionic strength, with a highly enhanced first peak. Differently, the 6-bead model yields quite good agreement with experimental data, except for the q-range $0.03 \lesssim q \lesssim 0.07$ Å^{-1}, where $S_{eff}(q)$ is slightly
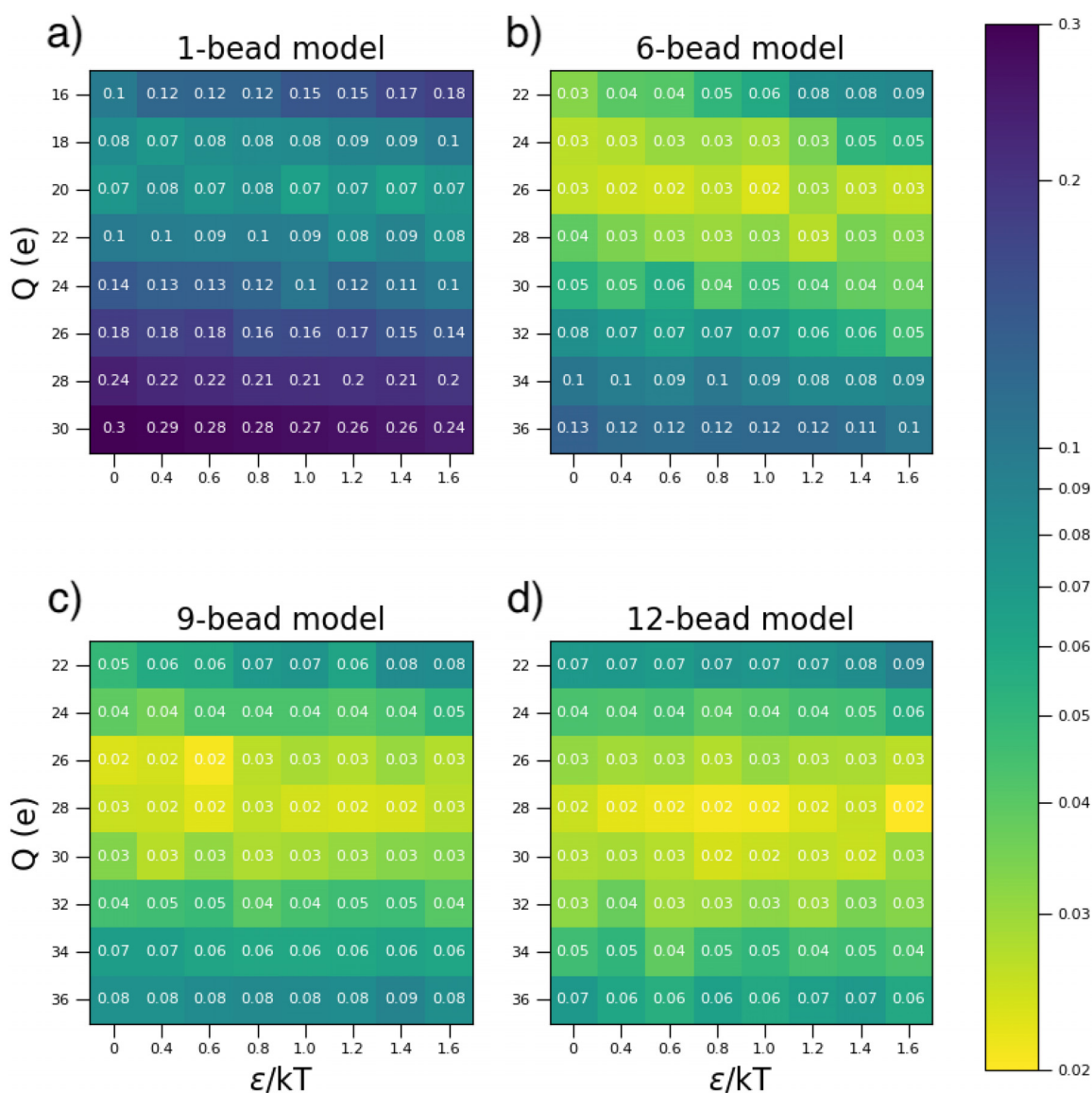
FIG. 6. χ^2 table for the 1-bead (a), 6-bead (b), 9-bead (c), and 12-bead (d) models related to the condition at low mAb concentration (20 mg/mL) and low ionic strength (7 mM) for several combinations of the potential parameters (Q_{eff}, ε_{ij}). Each χ^2 value is associated with a color. A logarithmic scale is used to highlight differences among the values which are rounded to two decimal places.

underestimated. Now increasing the number of beads, the χ^2 tables in Fig. 6 suggest that $\Delta Q_{net} = 28 \pm 2e$ and $\Delta Q_{net} = 29 \pm 2e$, respectively, for the 9-bead and the 12-bead models. These results indicate a clear trend: as the number of beads increases, the window of effective charges able to reproduce the experimental data increases, approaching the expected value estimated at the amino acid level. This is illustrated in Fig. 9, which also shows that for a number of beads greater than 9, Q_{eff} saturates, suggesting no major differences between the two models. This is confirmed by Fig. S10 (in the supplementary material) where the comparison of $S_{eff}(q)$ between the 9-bead and the 12-bead models is reported for some of the studied conditions, showing

indistinguishable results for the two bead models. Hence, we can conclude that the 9-bead is the best compromise in balancing computational cost and accuracy. It is also interesting to note that the agreement between simulated and measured $S_{eff}(q)$ at high mAb concentrations is significantly better for the much more strongly coarse-grained 6 or 9-bead models than for the amino acid-level coarse-grained model when using the initial non-relaxed structure (see Figs. 5, 7, and 10. While this seems at first counterintuitive, at high mAb concentrations excluded volume interactions and the actual dimensions of the mAb structure become important. A direct comparison between the initial and the relaxed configuration at the full amino-acid level
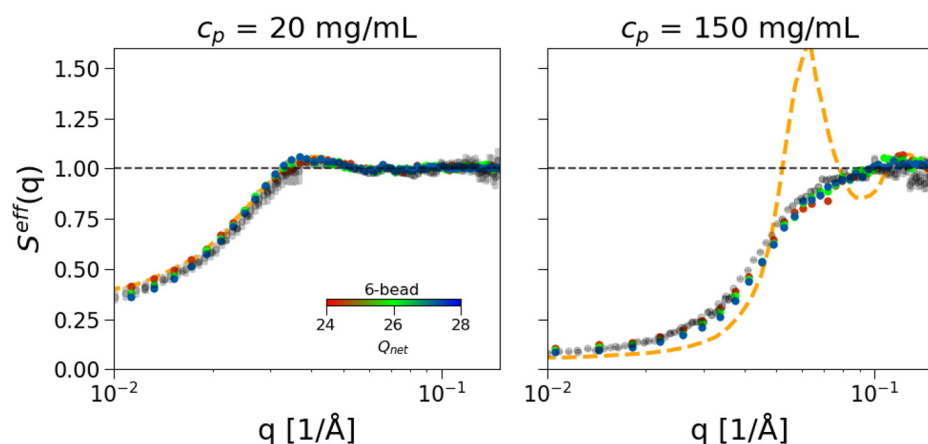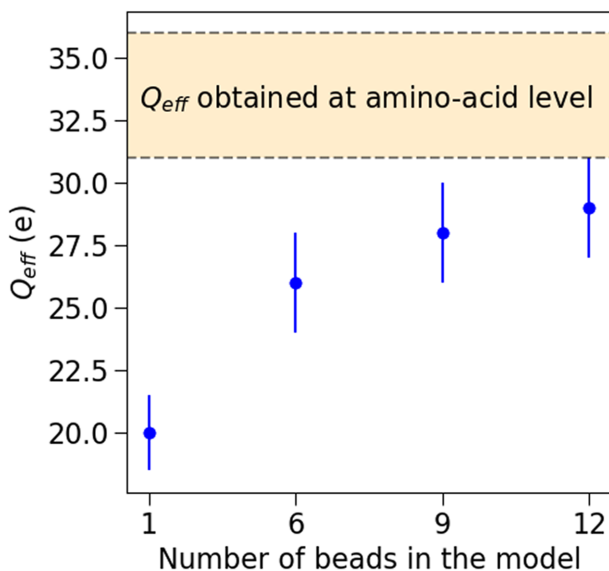
FIG. 7. Effective structure factors, $S_{eff}(q)$, obtained from experimental data (gray circles) at low ionic strength (7 mM) and 20 mg/mL (left) and 150 mg/mL (right) mAb concentrations in comparison with the 1-bead model (dashed orange line) and an effective charge of $Q_{eff} = 20\,e$ and the 6-bead model (open circles) and an effective charge in the range $\Delta Q_{eff} = 26 \pm 2e$.

and the 6-bead model as shown in Fig. S12 (in the supplementary material) indeed indicates that the agreement between the 6-bead model and the relaxed structure is improved when compared with the initial non-relaxed structure. As a result, the simpler Y-shaped bead models perform better than the initial configuration in terms of the local structural correlations at very high mAb concentrations. It appears that the appropriate choice of the radius of gyration is crucial to correctly reproduce the overall dimensions and local packing constraints when using highly coarse-grained bead models. Having focused on the low ionic strength, we now turn to the $I = 57$ mM case, where repulsion is much more screened, and also the amino acid model showed some discrepancy from the experimental results at high mAb concentrations. It is important to note that to tackle this case, we should, in principle, allow for a different net charge since this is also found to be differe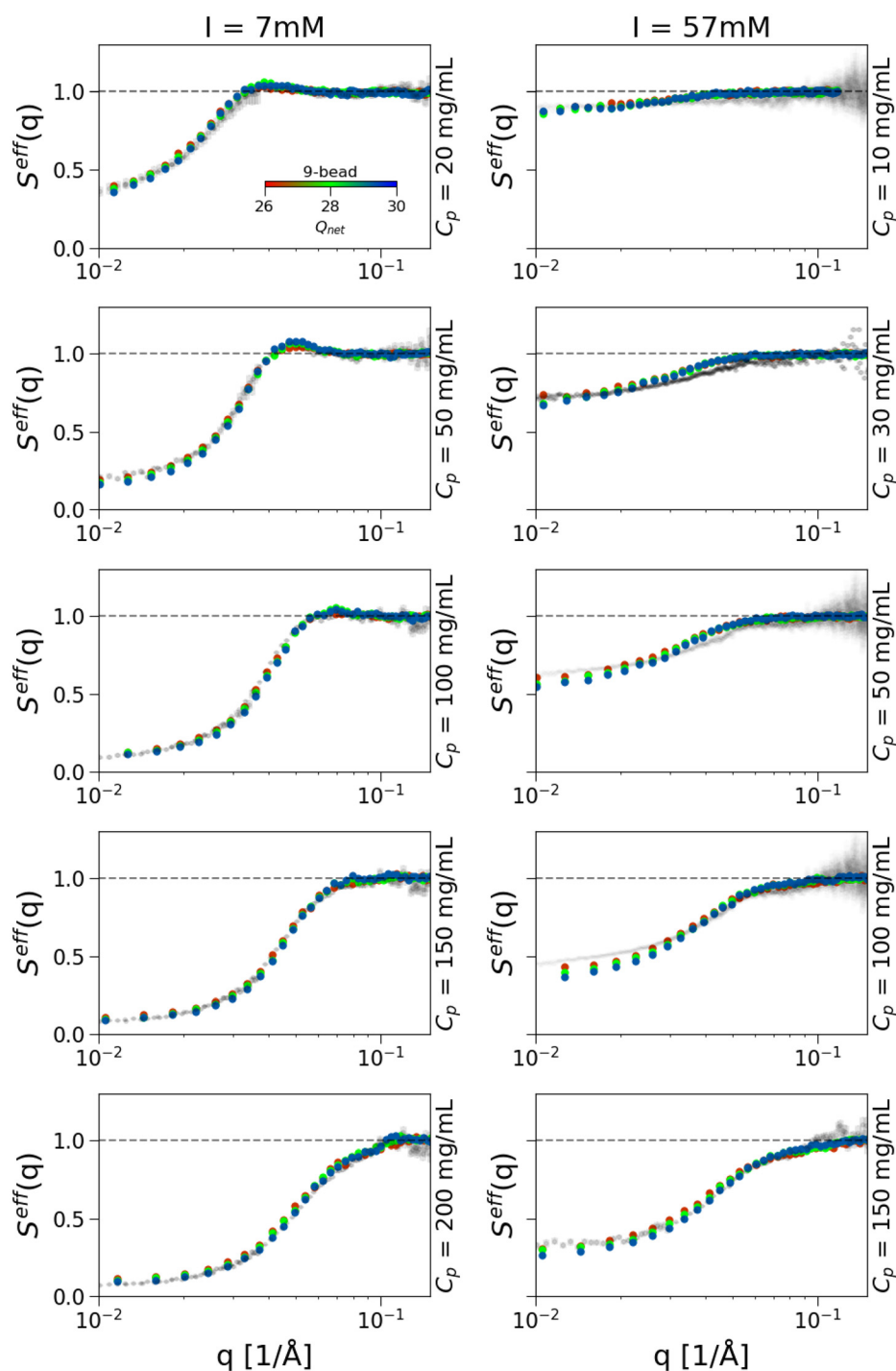nt in the constant-pH simulations. We then need to adjust the Debye length according to Eq. (3) to take into account the added NaCl as well as the counterions from the charged beads. On the other hand, we expect the attractive strength ε_{ij} to only weakly depend on the ionic strength. We, thus, fix it to the same value that provided the best agreement for $I = 7$ mM, namely, $\varepsilon_{ij} = 0.8\,k_B T$. In Fig. 10, we report $S_{eff}(q)$ for the 9-bead model in comparison with the SAXS data for both ionic strength conditions (7 and 57 mM) and different mAb concentrations. We compare results for different Q_{eff}, finding no major differences between the two cases. Overall, we find that the agreement between experimental and numerical structure factors is quantitative at all studied conditions with small discrepancies observed at high mAb concentration and high ionic strength. This is probably because, under these conditions of high screening, attractive interactions become more important and these may be anisotropic and/or localized in the more hydrophobic spots of the mAb, something we cannot properly capture by assigning an overall identical charge to all beads and an isotropic long-range attraction. However, the quality of the agreement is comparable to that obtained for the amino acid level description (see Fig. 4).

III. CONCLUSIONS

In this study, we aimed to develop a computationally efficient protocol for investigating protein solutions, specifically monoclonal antibodies (mAbs), and reproduce their structural properties accurately. We exploited a multi-scale approach, spanning from all-atom MD simulations and coarse-graining at amino acid level, to build up mesoscale bead models which turned out to be very efficient in reducing the system complexity while preserving the essential structural characteristics. Our investigations focused on a mostly repulsive mAb with an isotropic charge distribution and a high net charge. By using a coarse-graining at the amino acid level in combination with constant-pH MC simulations, we calculated the mAb charge distribution for the selected solution conditions. Then, we used the knowledge acquired to perform all-atom MD simulations of a single mAb and extrapolate representative mAb conformations. This step turned out to be necessary since we have shown that performing many-protein simulations at high concentrations with a non-relaxed mAb conformation leads to an incorrect estimate of the excluded volume. With such representative conformations coarse-grained at an amino acid level, we then carried out many protein MC simulations to reproduce a wide range of

FIG. 8. Values of Q_{eff} best reproducing the experimental data for the different bead models used as judged from the χ-tables shown in Figs. 6 and S6–S9 in the supplementary material.

FIG. 9. Effective structure factors $S_{eff}(q)$, obtained from simulations with the 9-bead model and an effective charge in the range of $\Delta Q_{eff} = 26 - 30e$ in comparison with experimental data (gray circles). Here, the attractive term is van der Waals-like with $\epsilon_{ij} = 0.8 k_B T$.

experimental structure factors up to the mAb concentration of 200 mg/mL. However, simulating mAb solutions at this level of detail is computationally demanding, especially for studying behavior in the high-concentration regime. Therefore, based on the information extrapolated at both all-atom and amino-acid levels, we have built up

bead models much easier to handle and at the same time very accurate in describing the structural properties of the mAb solution.

The bead models rely on two key parameters: the effective charge Q_{eff}, which is isotropically distributed among the beads, and the strength of attraction ε_{ij}, which takes into account short-range

interactions on top of the excluded volume, that is taken care of by the number of beads in the model as explained above. To select Q_{eff} for each bead model, we focused on low mAb concentration and low ionic strength, where electrostatic repulsions are dominant. By minimizing the difference between experimental data and model predictions, we determined the appropriate Q_{eff} value, which is also found to approach the value obtained in the one-body constant-pH simulations with an increasing number of beads. For ε_{ij}, we then found the best value that, in combination with Q_{eff}, can reproduce the experimental data under all studied conditions of mAb concentration and ionic strength. It is, however, important to note that in contrast to Q_{eff}, ε_{ij} cannot accurately be determined independently from a measurement of $S_{eff}(q)$ as these parameters are strongly coupled as demonstrated by the χ^2-tables shown in the supplementary material.

By systematically increasing the level of detail, we have identified the minimum bead model capable of fully reproducing solution structure factors for mAb concentrations up to 200 mg/mL and ionic strengths up to 57 mM. In more detail, the 1-bead model was able to reproduce the experimental structure factor only at low mAb concentration and low ionic strength. This is because, in this regime, the interactions are mainly driven by strong long-range electrostatic repulsions such that short-range interactions and the actual shape of the molecule only play a minor role. On the other hand, the Y-shaped 6-bead model provided a significant improvement, confirming the need to include shape anisotropy to account for local structural correlations at mAb concentrations exceeding 50 mg/mL. However, such a model failed to reproduce the effective structure factor at mAb concentrations above 150 mg/mL, due to overestimation of excluded volume effects. Both the 9-bead and 12-bead models were shown to be able to resolve this problem with almost identical accuracy, so we can conclude that the 9-bead is the best compromise in terms of gain of computational cost, in agreement with recent work on a different, more attractive and cluster-forming mAB.[14]

The validation of the coarse-grained bead models was carried out against small-angle x-ray scattering (SAXS) experiments and the comparison with amino-acid level descriptions highlighted the computational advantage of the former models. In conclusion, our developed bead models offer a computationally affordable yet sufficiently accurate approach to studying mAb solutions' structural properties. Their efficiency allows for extended investigations under various conditions, facilitating predictions of structure factors for challenging concentration regimes not easily accessible in experimental setups. Future work will need to extend and generalize such simplified treatments to more complicated mAbs where aggregation and/or self-assembly also take place, taking into account electrostatics in an accurate way to include heterogeneous charge distributions and to allow for, e.g., directional attraction, which often drives cluster formation in mAbs.[14,16]

IV. METHODS
A. Initial structure calculation

We use the IGg1 mAb called Actemra (or Tocilizumab) as the basis of our work, for which we have described a detailed characterization of key solution parameters over a large range of concentrations and solution conditions elsewhere.[8] The primary amino acid sequence was retrieved from patent US20120301460. A homology model was prepared using the Antibody Modeler module in Molecular Operating Environment (MOE) 2020.[8,13] Briefly, the primary sequence was used to identify suitable existing structures for the framework

and variable domains upon which the model was built. The complementarity-determining regions (CDRs) were modeled individually based on known loop structures and were then grafted onto the antibody framework. The structure then underwent energy minimization using "LowModeMD" to eliminate steric clashes.

B. One-protein MC simulations at the amino acid level

Based on this structure, and using the same protocol as in Ref. 14, we construct a coarse-grained representation of the mAb by replacing each amino acid with a spherical bead of diameter $\sigma_{bead}^{aa} = (6M_W/\pi\rho)^{1/3}$, where M_W is the amino acid molecular weight (in g mol^{-1}), $\rho = 1$ (in g mol^{-1} Å^{-3}) is an average amino acid density,[17] and the suffix aa stands for "amino acid" (see Fig. 3, leftmost).

With the amino-acid-based coarse-grained model, we perform Metropolis-Hastings Monte Carlo (MC) simulations of the mAb solution using Faunus (v2.9.1 git 3edf85cf),[18] which is a software allowing for several types of MC simulations, to estimate the mAb charge distribution (as performed here[19]). In particular, we carry out constant-pH MC simulations with a single rigid mAb and titration moves only,[20] which allow the amino-acid charges to fluctuate and to reach an equilibrium distribution at a given pH and the ionic strength. We consider two ionic strength conditions as in experiments:[8] 7 and 57 mM. For each of them, we perform a simulation of 10^4 MC sweeps where, on each sweep, $N = 10$ titration moves were attempted. The full interaction potential is defined as

$$\beta V = \underbrace{\sum_{i}^{n-1}\sum_{j=i+1}^{n} \frac{\lambda_B q_i q_j}{r_{ij}} e^{-r_{ij}/\lambda_D}}_{Electrostatic}$$

$$\underbrace{- \sum_i \ln\left(\frac{N_i!}{(N_i+\nu)!}V^{\nu_i}\right) - \sum_i \ln(10)(pK_{a,i}-pH)}_{Titration}, \quad (2)$$

where the first term accounts for electrostatic interactions; the second and the third terms account for the energy due to the titration of charged amino acids. Here, $\beta = 1/k_B T$ is the reciprocal of the thermal energy, where k_B is the Boltzmann constant and T is the system temperature; $\lambda_B = \beta e^2/4\pi\varepsilon_0\varepsilon_r = 7.1$ Å is the Bjerrum length, where e is the electron charge, ε_0 is the vacuum permittivity, and $\varepsilon_r = 78.7$ is the relative dielectric constant of water at room temperature (25 °C) (as the solvent is treated as a continuum); q_i and q_j are the electrical charges on the ith and jth amino acid, and r_{ij} is the distance between them; and λ_D is the Debye length, which accounts for charge screening effects as the salt is treated implicitly. As we consider single molecule behavior only, i.e., work in the limit of infinite dilution, λ_D is set equal to $3.04/\sqrt{I(M)}$ Å,[21] where $I(M)$ is the molar ionic strength. The titration terms propagate the reaction $AH \rightleftarrows A + H$ back and forward using a reactive MC scheme.[20] Here, AH is the protonated form of the amino acid, while $A + H$ is its dissociated form. In addition, N_i is the number of charged amino acids of species i, ν_i is the stoichiometric coefficient (positive for the products and negative for the reagents), V is the volume of the system, and $K_{a,i}$ is the acid dissociation constant of the ith amino acid residue. In this way, the MC moves shift the equilibrium reaction based on the solution conditions (such as pH and salt concentration). The $pK_{a,i}$ values are taken from the literature[22] and reported in Table S1 of the supplementary material.

C. All-atom molecular dynamics simulations and conformational analysis

Next, we perform two molecular dynamics (MD) simulations of a single mAb, one for each ionic strength condition (7 mM or 57 mM), with the GROMACS package (version 2022.2),[23] using the Amber99SB force field[24] and the SPC/E water model.[25] The protonation state of charged amino acids, histidine, arginine, lysine, aspartic acid, glutamic acid, and glutamine, have been assigned based on the average charges obtained for each amino acid from the one-protein MC simulations at pH 6 and either 7 mM or 57 mM ionic strengths.

The mAb is inserted into a cubic simulation box of volume 17.85 × 17.85 × 17.85 nm³, containing either 183 792 or 183 436 water molecules, based on the ionic strength condition. In each simulation, the net charge is neutralized by adding an appropriate number of counterions. Furthermore, we add the necessary amount of Na^+ and Cl^- ions to reach the desired ionic strength. Before each production run, the system energy of the whole system is minimized using the steepest descent method with a tolerance of 1000 kJ mol^{-1} nm^{-1}. We equilibrate the temperature and pressure of the system in two phases. First, using the Berendsen thermostat,[26] we stabilize the system temperature at 300 K performing a run of 100 ps in the canonical ensemble. Then, we equilibrate the pressure with the Parrinello–Rahman method[27] to the target value of 1 bar. After the equilibration process, we carry out simulations at each condition using the leap-frog integrator[28] for a total time of 1.5 µs. All simulations are performed with an integration step of 2 fs and with a 1 nm cutoff for the short-range van der Waals (vdW) interactions, while electrostatics are treated using the particle mesh Ewald[29] (PME) method with a real space cutoff of 1 nm. Periodic boundary conditions (PBCs) are used. To satisfy the so-called minimum image convention, i.e., to avoid the mAb seeing its periodic image, which would lead to spurious forces, we specify a solute-box distance of 1.0 nm. This ensures at least 2.0 nm between any two periodic images of the mAb.

By analyzing the MD trajectories, we selected representative conformations of the mAb using the algorithm described by Daura et al.[15] This is a well-established method to analyze MD trajectories and to identify distinct conformational states of a bio-molecule. It allows the grouping of similar protein configurations based on their structural similarity. The algorithm is mainly based on three different steps: (i) measuring the structural differences between all conformation pairs by calculating the root mean squared deviation (RMSD); (ii) constructing a similarity matrix where it assigns higher similarity values to pairs of conformation with lower RMSD values. In this process, a cutoff is selected to determine the maximum distance at which two conformations are considered to be similar. Here, we use a cutoff of 6 nm; and (iii) grouping similar conformations, i.e., selecting those with relative distance in the RMSD smaller than the cutoff, to find distinct structural states. For each simulation (one for each ionic strength), the MD trajectory is then reduced to groups of conformations. Each group has a central structure that is the closest (the most similar) to all the others in the same group (in terms of RMSD). We chose the central structure of the most populated group of each simulation as the representative for the entire MD trajectory.

D. Many-protein MC simulations at the amino-acid level

The most representative conformation obtained from the analysis of the all-atom MD trajectories was coarse-grained at an amino-acid

level for each of the two ionic strengths investigated (see Fig. 3, center and rightmost) and used to perform many-protein MC simulations. In more detail, $N_p = 20$ rigid translating and rotating mAbs are inserted in a cubic box of volume $V = N_p M_w / (c_p N_a 1 \times 10^{-27})$, where V is in Å³, $M_w = 148$ kDa (or, equivalently, 148 000 g/mol) is the mAb molecular weight, N_a is Avogadro's number, and C_p is the mAb concentration that we chose equal to 20, 50, 100, 150, and 200 mg/mL to compare with experimental data.[8] The interactions are ruled by the pair potential between the ith and jth amino acids, $V(r_{ij})$, defined as

$$\beta V(r_{ij}) = \underbrace{\lambda_B q_i q_j \frac{e^{-r_{ij}/\lambda_D}}{r_{ij}}}_{Electrostatic} + \underbrace{4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{vdW}. \quad (3)$$

The amino acid potential is composed of the sum of two accounting terms for electrostatic and van der Waals interactions. For a solution containing charged mAbs, dissociated counterions, monovalent salt, and buffer ions, $1/\lambda_D^2$ is calculated as

$$1/\lambda_D^2 = 4\pi\lambda_B \left[\left(\frac{1}{1-\phi} \right) Z_{mAb}\rho_{mAb} + 2\rho_{salt} + 2\rho_{buffer} \right], \quad (4)$$

where ϕ is the excluded volume of a single mAb and the factor $1/(1-\phi)$ accounts for the free accessible volume, with ϕ being calculated as the excluded volume of a hard sphere of diameter σ_{HS}, equal to

$$\phi_{HS} = \rho_{mAb} \frac{\pi \sigma_{HS}^3}{6}, \quad (5)$$

where ρ_{mAb} is the mAb number density. For our simulations, we choose $\sigma_{HS} = 2R_g$, where R_g is the average radius of gyration of the mAb obtained from MD simulations. For the two ionic strength conditions considered, R_g is equal to 48.2 Å for $I = 7$ mM and 45 Å for $I = 57$ mM.

Furthermore, ρ_{salt} and ρ_{buffer} are the number densities of salt and of the dissociated buffer ions, respectively. The van der Waals term accounts for short-range interactions, which is here represented by the Lennard–Jones potential, where ε_{ij} is the depth of the attractive well, and $\sigma_{ij} = (\sigma_{bead,i}^{aa} + \sigma_{bead,j}^{aa})/2$ is the characteristic interaction distance.[30] For each of the solution conditions, we perform eight replicas with the replica exchange method,[31] which allows us to reach a faster sampling, especially for high mAb concentration conditions. Each of the replicas is characterized by the assigned value of ε_{ij}. By increasing ε_{ij} among the replicas by 0.005 $k_B T$, we explore the range $\Delta\varepsilon_{ij} = 0.05$–0.085 $k_B T$. This interval has been selected from previous studies of coarse-grained models for both globular proteins[32–36] and antibody solutions.[12,16,19] For simplicity, at this stage, we neglect the hydrophilic and/or hydrophobic nature of the single amino acid and we assume that each bead contributes with the same attractive strength. Each simulation is carried out for 10^4 MC sweeps where, on each sweep, each protein is attempted to be translated and rotated.

E. Many-protein MC simulations with bead-models

In parallel, we also develop a different strategy to investigate the mAb solutions, which consists of simplifying the mAb representation with simple models made by a limited number of beads. Specifically, we use models with one, six, nine, and twelve beads, each constituted

by either one or several identical hard spheres with an effective diameter, σ_{bead} (see Fig. 1). Except for the 1-bead model, we assume a fixed angle of 60° between the two arms of the Y-shaped mAb (the FAB domains) and we choose σ_{bead} to reproduce the same average radius of gyration, R_g, obtained from all-atom MD simulations. Since R_g is slightly different based on the two ionic strength conditions, for each of the bead models, we select two σ_{bead} values, which are reported in Table S2 of the supplementary material.

The bead–bead pair potential, $V(r_{ij})$ is the sum of a hard-sphere contribution plus electrostatic and van der Waals interactions taking place at bead–bead distances, r_{ij}, longer than σ_{bead}, as

$$
\begin{aligned}
\beta V(r_{ij}) &= \infty & (r_{ij} < \sigma_{\text{bead}}) \\
&= \underbrace{\lambda_B q_{bead}^2 \left(\frac{e^{\sigma_{\text{bead}}/2\lambda_D}}{1 + \sigma_{\text{bead}}/2\lambda_D} \right)^2 \frac{e^{-r_{ij}/\lambda_D}}{r_{ij}}}_{Electrostatic} \\
&\quad \underbrace{- \epsilon_{ij} \left(\frac{\sigma_{bead}}{r_{ij}} \right)^6}_{vdW} & (r_{ij} > \sigma_{\text{bead}}), \quad (6)
\end{aligned}
$$

where, in the electrostatic term, $q_{bead} = Q_{eff}/N_{beads}$ is the electrical charge carried by each bead, with N_{beads} being the number of beads present on each model, and Q_{eff} being the net effective charge of the mAb. The inverse of the Debye screening length, $1/\lambda_D$, is again given by Eq. (4). Since the beads are treated as hard spheres, only the attractive part of the Lennard–Jones potential is included to account for short-range interactions. As for the simulations at the amino-acid level, for each condition, we use a cubic simulation box of volume $V = N_p M_w/(c_p N_a)$, where now N_p is set to 500. Each simulation is carried out for 10^5 MC sweeps, where on each sweep, each molecule is attempted to be rotated and translated.

F. Effective structure factor calculation

The purpose of all our many-body simulations is to obtain the effective structure factor, $S_{eff}(q)$ of the mAb solution under different conditions. $S_{eff}(q)$ is calculated explicitly as[31,37]

$$
S_{eff}(q) = \frac{1}{\langle F(q) \rangle} \frac{1}{N_{tot}} \left\langle \left(\sum_i^{N_{tot}} \sin(\mathbf{q}\mathbf{r}_i) \right)^2 + \left(\sum_i^{N_{tot}} \cos(\mathbf{q}\mathbf{r}_j) \right)^2 \right\rangle,
\tag{7}
$$

where N_{tot} is the product of the number of mAbs and the number of beads used in each model (i.e., 1, 6, 9, ...); r_i and r_j identify the position of the ith and jth bead, and $F(q)$ is the form factor that is obtained from an MC simulation in a cubic box of volume V where only a single molecule is inserted ($N_p = 1$). The averages are performed over all directions obtained by permuting the crystallographic index [100], [110], and [111] of the cubic box to define the scattering vector $\mathbf{q} = 2\pi p/L(h, k, l)$, where $p = 1, 2, \ldots, p_{max}$, and $p_{max} = 25$.

SUPPLEMENTARY MATERIAL

See the supplementary material for the following additional information: A table with all the pK values of the different amino acids used in the Metropolis–Hastings Monte Carlo (MC) simulations performed to determine the charge distribution on the mAb for both ionic strengths (Table S1); an example of the net charge equilibration for both ionic strengths (Fig. S1); examples of the electrostatic iso-potential surfaces for ± 1 and $3 k_B/e$ at 7 and 57 mM ionic strengths (Figs. S2–S5); the χ^2-tables for all bead models and mAb concentrations investigated (Figs. S6–S9); a comparison of the effective structure factor $S_{eff}(q)$ obtained from simulations of the 9 and 12-bead models for two different mAb concentrations at both ionic strength values (Fig. S10); a comparison between the experimental SAXS data and the calculated form factor for the initial structure obtained from homology modeling and the relaxed structure from the MD simulations at 7 and 57 mM ionic strength (Fig. S11); and a comparison between the coarse grained (at the amino acid level) structure for the initial and the relaxed structure at 7 mM ionic strength and the 6-bead model (Fig. S12).

AUTHOR DECLARATIONS
Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

Ethics approval is not required.

Author Contributions

Marco Polimeni: Conceptualization (equal); Formal analysis (lead); Investigation (lead); Methodology (lead); Writing – original draft (lead). **Emanuela Zaccarelli:** Conceptualization (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Supervision (lead); Writing – original draft (lead). **Alessandro Gulotta:** Investigation (equal); Methodology (supporting). **Mikael Lund:** Investigation (equal); Methodology (equal); Supervision (lead). **Anna Stradner:** Funding acquisition (lead); Investigation (equal); Methodology (equal); Supervision (lead). **Peter Schurtenberger:** Conceptualization (lead); Formal analysis

(lead); Funding acquisition (lead); Investigation (equal); Methodology (equal); Supervision (lead); Writing – original draft (lead).

DATA AVAILABILITY

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.10478576, Ref. 38.

REFERENCES

[1] U. Hafeez, H. K. Gan, and A. M. Scott, "Monoclonal antibodies as immuno-modulatory therapy against cancer and autoimmune diseases," Curr. Opin. Pharmacol. **41**, 114–121 (2018).

[2] F. H. Du, E. A. Mills, and Y. Mao-Draayer, "Next-generation anti-CD20 mono-clonal antibodies in autoimmune disease treatment," Autoimmun. Highlights **8**, 12 (2017).

[3] G. Pantaleo, B. Correia, C. Fenwick, V. S. Joo, and L. Perez, "Antibodies to combat viral infections: Development strategies and progress," Nat. Rev. Drug Discovery **21**, 676–696 (2022).

[4] A. Mullard, "FDA approves 100th monoclonal antibody product," Nat. Rev. Drug Discovery **20**, 491–495 (2021).

[5] É. Kollár, B. Balázs, T. Tari, and I. Siró, "Development challenges of high concentration monoclonal antibody formulations," Drug Discovery Today: Technol. **37**, 31–40 (2020).

[6] E. M. Moussa, J. P. Panchal, B. S. Moorthy, J. S. Blum, M. K. Joubert, L. O. Narhi, and E. M. Topp, "Immunogenicity of therapeutic protein aggregates," J. Pharm. Sci. **105**, 417–430 (2016).

[7] W. Jiskoot, A. Hawe, T. Menzen, D. B. Volkin, and D. J. Crommelin, "Ongoing challenges to develop high concentration monoclonal antibody-based formulations for subcutaneous administration: Quo Vadis?," J. Pharm. Sci. **111**, 861–867 (2022).

[8] A. Gulotta, M. Polimeni, S. Lenton, C. G. Starr, J. S. Kingsbury, A. Stradner, E. Zaccarelli, and P. Schurtenberger, "Combining scattering experiments and colloid theory to characterize charge effects in concentrated antibody solutions," arXiv:2311.01986 (2023).

[9] A. Chaudhri, I. E. Zarraga, T. J. Kamerzell, J. P. Brandt, T. W. Patapoff, S. J. Shire, and G. A. Voth, "Coarse-grained modeling of the self-association of therapeutic monoclonal antibodies," J. Phys. Chem. B **116**, 8045–8057 (2012).

[10] N. Skar-Gislinge, M. Ronti, T. Garting, C. Rischel, P. Schurtenberger, E. Zaccarelli, and A. Stradner, "A colloid approach to self-assembling antibodies," Mol. Pharm. **16**, 2394–2404 (2019).

[11] S. Izadi, T. W. Patapoff, and B. T. Walters, "Multiscale coarse-grained approach to investigate self-association of antibodies," Biophys. J. **118**, 2741–2754 (2020).

[12] A. Chowdhury, J. A. Bollinger, B. J. Dear, J. K. Cheung, K. P. Johnston, and T. M. Truskett, "Coarse-grained molecular dynamics simulations for understanding the impact of short-range anisotropic attractions on structure and viscosity of concentrated monoclonal antibody solutions," Mol. Pharm. **17**, 1748–1756 (2020).

[13] Chemical Computing Group, Inc., see http://www.chemcomp.com for "Molecular Operating Environment (MOE) Software."

[14] N. Skar-Gislinge, F. Camerin, A. Stradner, E. Zaccarelli, and P. Schurtenberger, "Using cluster theory to calculate the experimental structure factors of antibody solutions," Mol. Pharm. **20**, 2738–2753 (2023).

[15] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, "Peptide folding: When simulation meets experiment," Angew. Chem., Int. Ed. **38**, 236–240 (1999).

[16] A. Chowdhury, N. Manohar, G. Guruprasad, A. T. Chen, A. Lanzaro, M. Blanco, K. P. Johnston, and T. M. Truskett, "Characterizing experimental monoclonal antibody interactions and clustering using a coarse-grained simulation library and a viscosity model," J. Phys. Chem. B **127**, 1120–1137 (2023).

[17] S. Kaieda, M. Lund, T. S. Plivelic, and B. Halle, "Weak self-interactions of globular proteins studied by small-angle x-ray scattering and structure-based modeling," J. Phys. Chem. B **118**(34), 10111–10119 (2014).

[18] B. Stenqvist, A. Thuresson, A. Kurut, R. Vácha, and M. Lund, "Faunus—A flexible framework for Monte Carlo simulation," Mol. Simul. **39**, 1233–1239 (2013).

[19] S. Mahapatra, M. Polimeni, L. Gentiluomo, D. Roessner, W. FrieSS, G. H. J. Peters, W. W. Streicher, M. Lund, and P. Harris, "Self-interactions of two monoclonal antibodies: Small-angle x-ray scattering, light scattering, and coarse-grained modeling," Mol. Pharm. **19**, 508–519 (2022).

[20] J. K. Johnson, A. Z. Panagiotopoulos, and K. E. Gubbins, "Reactive canonical Monte Carlo: A new simulation technique for reacting or associating fluids," Mol. Phys. **81**, 717–733 (1994).

[21] J. N. Israelachvili, *Intermolecular and Surface Forces*, 3rd ed. (Elsevier, 2011).

[22] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, and C. N. Pace, "pK values of the ionizable groups of proteins," Protein Sci. **5**, 1214–1218 (2006).

[23] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1–2**, 19–25 (2015).

[24] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," Proteins **65**, 712–725 (2006).

[25] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," J. Phys. Chem. **91**, 6269–6271 (1987).

[26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," J. Chem. Phys. **81**, 3684–3690 (1984).

[27] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182–7190 (1981).

[28] A. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed. (Pearson Education, 2009).

[29] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," J. Chem. Phys. **98**, 10089–10092 (1993).

[30] A. Hagler, E. Huler, and S. Lifson, "Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals," J. Am. Chem. Soc. **96**, 5319–5327 (1974).

[31] D. Frenkel, R. J. Vos, C. G. de Kruif, and A. Vrij, "Structure factors of polydisperse systems of hard spheres: A comparison of Monte Carlo simulations and Percus–Yevick theory," J. Chem. Phys. **84**, 4625 (1986).

[32] C. Pohl, M. Polimeni, S. Indrakumar, W. Streicher, G. H. Peters, A. Nørgaard, M. Lund, and P. Harris, "Electrostatics drive oligomerization and aggregation of human interferon alpha-2a," J. Phys. Chem. B **125**, 13657–13669 (2021).

[33] J. Hladilkova, T. H. Callisen, and M. Lund, "Lateral protein–protein interactions at hydrophobic and charged surfaces as a function of pH and salt concentration," J. Phys. Chem. B **120**, 3303–3310 (2016).

[34] A. Kurut, C. Dicko, and M. Lund, "Dimerization of terminal domains in spiders silk proteins is controlled by electrostatic anisotropy and modulated by hydrophobic patches," ACS Biomater. Sci. Eng. **1**, 363–371 (2015).

[35] A. Kurut, B. A. Persson, T. Åkesson, J. Forsman, and M. Lund, "Anisotropic interactions in protein mixtures: Self assembly and phase behavior in aqueous solution," J. Phys. Chem. Lett. **3**, 731–734 (2012).

[36] A. Kurut and M. Lund, "Solution electrostatics beyond pH: A coarse grained approach to ion specific interactions between macromolecules," Faraday Discussions **160**, 271–278 (2013).

[37] P. Lindner and T. Zemb, *Neutrons, X-Rays and Light: Scattering Methods Applied to Soft Condensed Matter* (Elsevier Science, 2002).

[38] M. Polimeni *et al.* (2024). "Electronic Notebook: A multiscale numerical approach to study monoclonal antibodies in solution," Zenodo. https://doi.org/10.5281/zenodo.10478576