HHS Public Access

Author manuscript

J Pharm Sci. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

J Pharm Sci. 2018 May ; 107(5): 1269–1281. doi:10.1016/j.xphs.2017.12.015.

Predicting protein-protein interactions of concentrated antibody solutions using dilute solution data and coarse-grained molecular models

Cesar Calero-Rubio[1], **Ranendu Ghosh**[1,a], **Atul Saluja**[2,b], and **Christopher J. Roberts**[1,†]

[1]Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, DE 19716, USA

[2]Department of Drug Product Science and Technology, Bristol-Myers Squibb, New Brunswick, NJ 08901, USA

Abstract

Protein-protein interactions for solutions of an IgG1 molecule were quantified using static light scattering (SLS) measurements from low to high protein concentrations (c_2). SLS was used to determine second osmotic virial coefficients (B_{22}) at low c_2, and excess Rayleigh profiles (R^{ex}/K *vs* c_2) and zero-q structure factors ($S_{q=0}$) as a function of c_2 at higher c_2 for a series of conditions (pH, sucrose concentration, and total ionic strength (*TIS*)). Repulsive (attractive) interactions were observed at low *TIS* (high *TIS*) for pH 5 and 6.5, with increasing repulsions when 5% w/w sucrose was also present. Previously developed and refined coarse-grained (CG) antibody models were used to fit model parameters from B_{22} *vs TIS* data. The resulting parameters from low-c_2 conditions were used as the sole input to multi-protein Monte Carlo simulations to predict high-c_2 R^{ex}/K and $S_{q=0}$ behavior up to 150 g/L. Experimental results at high-c_2 conditions were quantitatively predicted by the simulations for the CG models that treated antibody molecules as either 6 or 12 (sub)domains, which preserved the basic shape of a monoclonal antibody. Finally, preferential accumulation of sucrose around the protein surface was identified *via* high-precision density measurements, which self-consistently explained the simulation and experimental SLS results.

INTRODUCTION

The limited stability of protein solutions currently requires extensive experimental evaluation of solution properties and quality attributes during the development stages of biopharmaceuticals.[1–6] For instance, physicochemical properties and key quality attributes such as protein solubility and aggregation, and solution opalescence and viscosity are

†corresponding author; tel: 302-831-0838; fax: 302-831-1048; cjr@udel.edu.
[a]Present address: Biocon Research Laboratory, University of Calcutta, Bengaluru, Karnataka, India
[b]Present address: Global Pharmaceutical Development Biologics, Sanofi, Framingham, MA 01701, USA

difficult to predict for highly concentrated protein solutions (upwards of 100 g/L).[2,7–9] Such high-concentration conditions have been increasingly targeted for monoclonal antibody (MAb) solutions when developing self-injectable products.[2,4] Protein physicochemical properties are sensitive to changes in solution formulation conditions such as pH, ion concentration and chemical identity, and the presence of other excipients. They also depend on protein identity (sequence, structure, binding affinity, etc.), storage temperature, and mechanical stresses (shaking, stirring, etc.).[2–11] This leads to an exceedingly large experimental space to be contend with during the development stages of any MAb, or more generally any protein product. It would be beneficial to have molecular- and/or sequence-based models that could reduce the formulation space by predicting at least a subset of these properties while minimizing the amount of experimental data needed as inputs.

So called “colloidal” or “weak” protein-protein interactions have been shown to correlate, in some cases, with protein phase separation, opalescence, aggregation rates, and elevated viscosities.[3,7,8,12] These interactions can be characterized experimentally using laser scattering (LS), analytical ultracentrifugation, small-angle neutron and/or x-ray scattering (SANS and SAXS, respectively), and osmometry.[3,12–18] Each of these methods allows one to measure second osmotic virial coefficients (B_{22}) as a function of solution formulation, but this interaction parameter is limited to dilute conditions (typically below ~10 g/L). Recent work has highlighted that changes in protein-protein interactions as a function of protein concentration (c_2) can be quantified using at least a subset of these experimental techniques.[3,7,12,19] Static light scattering (SLS) experiments allow one to directly measure protein interactions as a function of c_2 up to high concentrations, in the form of Kirkwood-Buff (KB) integrals, particularly the protein-protein KB integral, G_{22}.[3,12,20] This is also effectively the same as the static structure factor at the zero-q limit ($S_{q=0}$).[20] SLS provides $S_{q=0}$ and B_{22} values more easily and/or with greater accessibility or higher throughput compared to other techniques. This makes SLS an attractive method to experimentally quantify protein-protein interactions for a range of solution formulations (see also, Methods).[3,12,21]

Colloidal or “weak” protein-protein interactions arise from three main contributions: steric repulsions (*i.e.*, excluded volume effects); short-ranged non-electrostatic attractions such as van der Waals interactions and hydration/solvation effects; and long- or short-ranged electrostatic attractions and repulsions.[12,22,23] The balance between these forces is mediated by the solution formulation and the protein of interest.[2,3,6,21,24–27] While these forces are relevant at both low- and high-c_2 conditions, the average distance between protein molecules at higher c_2 is necessarily much smaller than at low c_2. As the distance-dependence of each of the forces listed above is different, it is expected that greatly changing c_2 can considerably affect the balance of these forces.[12,21,23,28,29] Additionally, most simplified molecular models, such as coarse-grained (CG) models, treat the solvent implicitly. This may lead to incorrect estimations of inter-protein forces when the average distance between protein molecules is of the order of a few water (hydration) layers (~1 nm).[30–33] Previous work with colloidal models applied to globular proteins showed that protein interactions at high c_2 can be predicted accurately from only low-c_2 behavior under certain solution conditions.[12,34] However, that approach was tested only for the case of a simpler model protein, and it has

yet to be assessed for highly asymmetric proteins such as MAbs, where packing effects may be greatly different from globular proteins and colloids.[16,23,35]

Previous work often has used a colloidal description of inter-particle interactions to describe protein solutions.[12,22,23,34] For instance, CG models with implicit solvent were developed to self-consistently capture physical volume and excluded volume effects of MAb solutions.[23] Such models were also developed to at least qualitatively capture trends in protein-protein interactions at low- and high-c_2 conditions.[12,23] Although simplified descriptions of MAbs in solution have been historically employed as a minimalist approach, the validity of using CG models for protein solutions has been raised in some cases.[12,22,23,34] This arises from the complex structural nature of antibodies, compared to traditional colloidal particles, and the fact that protein-protein interactions are potentially strongly influenced by factors such as pH, total ionic strength (*TIS*), added excipients, protein concentration (c_2), protein sequence, and the three-dimensional structure and heterogeneous surface chemistry of proteins. Recent examples raised the question of whether purely CG descriptions of protein behavior are practically useful for capturing or predicting high-c_2 properties of protein solutions.[12]

The present work considers the challenge of using CG molecular models to predict experimental protein interactions (*via* excess Rayleigh scattering) for highly concentrated antibody solutions using simple structural/sequence information and limited low-c_2 experimental data as the only model inputs. The excess Rayleigh profiles (R^{ex}/K *vs* c_2) of an IgG1 protein were experimentally determined as a function of pH, *TIS* and sucrose content. B_{22} values were determined from low-c_2 experiments and used to parameterize a previously developed CG antibody model as a function of pH, in terms of only two model parameters. These parameters capture the balance between short-ranged non-electrostatic attractions and the ratio of the true and theoretical charges on the domains of an IgG1 molecule at a given pH, without prior knowledge of the high-c behavior. Experimental high-c_2 R^{ex}/K results were then predicted using the low-c_2 parameters using multi-protein Monte Carlo simulations, which included upwards of 500 MAb molecules interacting simultaneously. The results are discussed from both qualitative and quantitative perspectives, highlighting strengths and weaknesses of the approach for net-repulsive and net-attractive conditions. Finally, partial specific volume measurements and protein-excipient KB integrals were used to rationalize the effect of adding sucrose to the protein solution from the perspective of preferential solute and solvent interactions, protein-protein interactions, and the model parameters determined from experimental and simulated B_{22} data.

MATERIAL AND METHODS

Sample preparation

Sodium acetate buffer stock solutions were prepared by dissolving glacial acetic acid (Fisher Scientific) in deionized water (MilliQ, Millipore-Sigma) to reach 10 mM acetic acid, and titrated to pH 5.1 ± 0.05 (termed pH 5 below) using a 5 M sodium hydroxide solution (Fisher Scientific). Similarly, 10 mM histidine buffer stock solutions were prepared by dissolving histidine hydrochloride (Sigma) in deionized water and titrating to pH 6.5 ± 0.05 (termed pH 6.5 below). A stock IgG1 solution was provided by Bristol-Myers Squibb at a

starting c_2 of ~50 g/L. pH 5 and pH 6.5 protein stock solutions were filtered and dialyzed using 10 kDa molecular weight cutoff (MWCO) Spectra/Por dialysis membrane (Spectrum Laboratories, Rancho Dominguez, CA) with the desired buffer using four 12-hr buffer exchanges at 4 °C to remove any undesired solutes from the original protein solution.

Excipient stock solutions were prepared by dissolving sucrose (HPLC grade, Sigma) and/or NaCl (Fisher Scientific) in 10 mM buffer solutions (acetate for pH 5 and histidine for pH 6.5) to obtain final solutions of 30% w/w sucrose and/or 1.3 M NaCl. Those solutions were titrated to the respective pH with small volumes of a 5 M sodium hydroxide solution. Final protein solutions were prepared gravimetrically by combining (1) protein stock solution, (2) pH-adjusted buffer, (3) excipient stock solution with a matched buffer. The proportions of (1), (2), and (3) were selected to achieve a constant excipient concentration and pH as specified in Table 1. This was done for a series of increasing c_2 values every 0.25 g/L up to a maximum of 1.5 g/L (for density measurements) or every 1 g/L up to a maximum of 10 g/L (for low-c_2 SLS measurements) to ensure sufficiently dilute protein conditions.[3,12]

For c_2-conditions above 10 g/L, concentrated protein stock solutions were prepared through membrane centrifugation at ~ 3200 RCF using 10 kDa MWCO Amicon-Ultra centrifugal tubes (Millipore-Sigma) and two buffer exchange steps. A pH shift was observed as the protein solution was concentrated from ~35 g/L to ~160 g/L, therefore starting pH values of 4.3 and 5.9 were selected for the dialysis with resulting pH values of 5.06 ± 0.05 and 6.49 ± 0.05, respectively, for final 160 g/L solutions after centrifugation with two buffer exchange steps. Lower-c_2 samples were then prepared by gravimetrically diluting the concentrated protein solutions in the desired buffer to obtain c_2 values ranging from 10 to 150 g/L. Final excipient and c_2 values were later calculated and corrected with measured density values (see below) and UV-VIS spectrophotometry (Agilent 8453, Santa Clara, CA) at 280 nm using an extinction coefficient of 1.54 L g^{-1} cm^{-1}, before and after dilution from the concentrated stock solutions.[3] Less than 0.1% variation between targeted and actual values for the protein and excipient concentrations was achieved in all cases.

Static light scattering (SLS)

Batch SLS experiments were conducted using a Wyatt Technology (Santa Barbara, CA) DAWN HELEOS II instrument with laser wavelength (λ) of 658.9 nm at 25.0 ± 0.1°C. In SLS, the average scattered intensity at 90° can be determined and used to calculate the excess Rayleigh ratio, represented as R^{ex}, as previously reported.[3,20] Measurements of R^{ex} as a function of c_2 can be used to estimate protein-protein interactions in the form of the protein-protein Kirkwood-Buff integral, G_{22}, as shown in equation 1. K is the optical constant and equal to $4\pi^2 n^2 (dn/dc_2)^2 N_A^{-1} \lambda^{-4}$, with n denoting the solution refractive index, (dn/dc_2) is the change in solution refractive index as a function of protein concentration, and N_A is Avogadro's number. Mw,app is the protein apparent molecular weight, and M is the protein true molecular weight.[3,20]

$$\frac{R^{ex}}{K} = M_{w,app}\, c_2 + M_w\, G_{22}\, c_2^2 \quad (1)$$

The zero-q limit for the structure factor ($S_{q=0}$) can be obtained from static light scattering at a single angle (90° in the present work) if there is no angle-dependence of the scattered light.[20] This is usually valid for protein solutions as proteins are much smaller than the laser wavelength, and this holds in the absence of long-ranged density fluctuations such as those observed near critical points and phase boundaries more generally.[3,12,20,23] $S_{q=0}$ can be obtained by dividing the right hand side of equation 1 by $c_2 M_w$, with the canonical simplification that $M_{w,app} \approx M_w$.[3,12,20,23] In this case, $S_{q=0}$ is equal to $1 + c_2 G_{22}$ and is dimensionless.[3,12,20,23] Values of ($d\pi/dc_2$) were determined using a J157HA Refractometer (Rudolph Scientific, Hackettstown, NJ) for c_2 values up to 10 g/L for each formulation. The values were 0.203 ± 0.03 mL/g for all buffer-only and NaCl formulations, and 0.220 ± 0.04 mL/g for all formulations with 5% w/w sucrose, for both pH values.

In the limit of dilute protein conditions (*i.e.*, c_2 below approximately 10 g/L and $\|c_2 G_{22}\| < 0.1$),[3,12,20] it holds that $G_{22} \approx -2B_{22}$, with B_{22} denoting the protein second osmotic virial coefficient.[20,23,29,36,37] As B_{22} is independent of protein concentration, B_{22} values were obtained by fitting experimental excess Rayleigh profiles to equation 1 for low-c_2 conditions, and setting $B_{22} = -½G_{22}$.[3,12] Additionally, KB theory and the corresponding analysis and Eq. 1 are applicable at higher c_2, and can be used to quantify protein-protein interactions at high c_2 from SLS data. A negative (positive) G_{22} value is equivalent to $S_{q=0}$ values below (above) 1, and corresponds to net repulsive (attractive) interactions. Correspondingly in dilute solutions, positive (negative) B_{22} values indicate net repulsions (attractions).

High-precision density measurements

Solution density (ρ) measurements were carried out using a DDM 2911 Plus density meter (Rudolph Scientific, Hackettstown, NJ). All measurements were done at 25.00 ± 0.02 °C and ambient pressure. Partial specific volume values ($\hat{V}_i$) were determined from high precision density measurements as a function of weight fraction (w_i) using equation 2, as previously described.[38,39] Linear regression was used to obtain the intercept and the slope as needed in equation 2. 95% confidence intervals for $\hat{V}_2$ were obtained from the corresponding t-value and standard error analyses of the slope and the intercept with error propagation for a given component *i*.[38,39]

$$\hat{V}_i = \frac{1}{\rho_o} + \left[\frac{d\left(\frac{1}{\rho}\right)}{dw_i}\right]_{T,P,m_{j\neq i}} \quad (2)$$

Preferential interactions were obtained by calculating protein partial specific volumes ($\hat{V}_2$) as a function of sucrose concentration (c_3) at a constant concentration of buffer (c_4) and comparing the results to equation 3, where κ_T represents the isothermal compressibility of the solution, R is the gas constant and T is the solution temperature. G_{i2} values represent the interaction between a given molecular species i and the protein (subscript 2).

$$\hat{V}_2 = \frac{RTk_T}{M_{w,2}} - G_{12} + c_3\hat{V}_3(G_{12} - G_{23}) + c_4\hat{V}_4(G_{12} - G_{24}) \quad (3)$$

Keeping with standard notation, subscripts 1 and 3 denote water and excipient (sucrose), respectively. The difference (G_{12}-G_{23}) is equivalent to the preferential interaction between water and protein, relative to sucrose and protein molecules.[38–41] A positive difference is characteristic of preferential exclusion or desolvation of sucrose, while a negative difference is characteristic of preferential accumulation, binding or solvation of the excipient to the protein surface.[24,30,38–42] This behavior can be observed by plotting $\hat{V}_2$ as a function of $c_3\hat{V}_3$, with $\hat{V}_3$ being the partial specific volume of the excipient in the same buffer solution (*i.e.*, $c_2 \rightarrow 0$ g/L).[39] The first term on the right-hand side in equation 3 ($RTk_T/M_{w,2}$) is negligible compared to the other terms for liquids far from their critical point.[38,39] Similarly, previous work has shown that the last term on the right-hand side ($c_4\hat{V}_4(G_{12} - G_{24})$) can be assumed to be negligible and/or c_3-independent for low concentrations of buffer (<30 mM) due to $c_4\hat{V}_4$ values being ~ 10^{-3} and with physically reasonable (G_{12}-G_{24}) values.[39]

Coarse-grained MAb models and interaction parameters

Two different coarse-grained (CG) molecular models were used to model low-c_2 behavior and predict high-c_2 SLS experimental behavior. These were a subset of a larger group of possible CG molecular models that were tested in earlier work that showed these two models can provide an optimal balance between accuracy and computational burden.[23] Figure 1 shows a schematic of the geometric constraints in these CG models, referred to as the HEXA and DODECA models in the remainder of this report. These models were developed in previous work to resemble the overall shape of a MAb molecule, and used 6 (HEXA) or 12 (DODECA) beads per protein. Additional details of the development of the models can be found elsewhere.[23] A modification to the previously proposed description of short-ranged non-electrostatic attractions was made here to achieve an effective attractive range of ~1 nm for both the HEXA and DODECA models (compared to equation 9 in Ref. 23). This was achieved by modifying the range of attractions for the DODECA model in comparison to previous work as shown in equation 4 and explained below. In equation 4, ε_{SR} represents the strength of the short-ranged non-electrostatic attractions, n represents the range of the attractions, and is equal to 6 for the DODECA model and 10 for the HEXA model. c is a normalization factor to make the interaction energy equal to $-\varepsilon_{SR}$ at its minimum value, and is equal to 1.2196 for the DODECA model and 1.3464 for the HEXA model.[23]

$$\frac{u_{SR}(r_{ij})}{k_BT} = \frac{\varepsilon_{SR}}{k_BT}c\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{128} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{n}\right] \quad (4)$$

Similarly, a modification to the previously proposed electrostatic model was made to better model experimental data. This was achieved by changing from a Yukawa potential (equation 10 in Ref. 23) to a modified screened-Coulomb potential as shown in equation 5.[12,23,43]

$$\frac{u_{el}(r_{ij})}{k_B T} = \zeta \psi_i \psi_j q_i q_j \frac{e^{-k(r_{ij} - \sigma_{ij})}}{r_{ij}\left[1 + \frac{1}{2}(k\sigma_{ij})\right]^2} \quad (5)$$

ζ corresponds to the Bjerrum length and is equal to $(4\pi\varepsilon\varepsilon_o k_B T)^{-1}$, with ε representing the solution relative permittivity at a given temperature, ε_o is the vacuum permittivity (in units of $qe^2N^{-1}\ m^{-2}$, with q_e representing the elemental charge of an electron),[43] k_B is the Boltzmann constant, and T is the absolute temperature. For solutions considered here, ζ was equal to 7.15 Å for any buffer + NaCl formulations at 25 °C, and 7.26 Å for formulations with 5% w/w added sucrose at 25 °C.[44] q_i and q_j are the theoretical valences of domain/fragment/amino acid i and j, respectively, as calculated from the protein sequence (see below), while ψ_i (or ψ_j) is used to scale the theoretical charge such as $\psi_i q_i$ (or $\psi_j q_j$) is equal to the effective valence in solution, $q_{i,eff}$ (or $q_{j,eff}$). σ_{ij} is the average diameter of beads i and j, equal to ½(σ_i + σ_j), where σ_i and σ_j are the diameters of the ith and jth bead or (sub)domain, respectively. κ is the Debye screening length based on the *TIS* of the solution.[12,23,43,45,46] r_{ij} is the center-to-center distance between the interacting beads i and j. The interaction potential in equation 5 allows one to better capture electrostatic phenomena, and can allow for easier comparison with electrophoresis measurements.[45]

Theoretical valence values (q_i) were calculated using the standard Henderson-Hasselbach equation.[47] The protein sequence was provided by Bristol-Myers Squibb, including a homology model for capturing the molecule geometry (see Fig. 1). This sequence was partitioned into equal-chain-length units to compute the charge of each HEXA or DODECA model bead. For the HEXA model calculations, the Fv domain was composed of the upper half of the light chain (residues 1-107) and the upper quarter of the heavy chain (residues 1-118) (*i.e.*, combining both V_H and V_L domains); the C1 domain was composed of the lower half of the light chain (residues 108-214) and the second quarter of the heavy chain (residues 119-234) (*i.e.*, combining both C_H1 and C_L domains); the C2 domain was composed of residues 244-357 of each heavy chain (*i.e.*, both C_H2 domains); and the C3 domain was composed of last quarter of each of the heavy chains (residues 358-474) (*i.e.*, both C_H3 domains). For the DODECA model calculations, the heavy chains were portioned into four units (residues 1-118 for the V_H, 119-234 for the C_H1, 244-357 for the C_H2, and 358-474 for the C_H3) and the light chains into two units (residues 1-107 for the V_L, and 108-214 for the C_L) for a total of 12 beads, each with its respective net charge.[23] In what follows, the terms valence and charge will be used interchangeably. Examples of the theoretical charge distribution for the DODECA model are shown in Figure 2, while the theoretical charge values are shown in Table 2 for pH 5 and 6.5 for both CG models.

Monte Carlo simulations for both low- and high-c_2 conditions

Low c_2—*Low c_2:* the HEXA and DODECA models were used to compute B_{22} for a given pH and *TIS* using the Mayer Sampling method employing the Overlap Sampling algorithm (MSOS) developed by Kofke and coworkers.[12,23,48] A similar methodology to the one employed before[23] was used here: MSOS simulations were performed at 25 °C with 10^7

Monte Carlo (MC) attempts for both the reference system and the model of interest.[12,23] Each MC attempt consisted of either a translation or a rotation around the center of mass of the first protein molecule using the center of the second molecule as the origin. The maximum displacement and rotation were obtained with a pre-equilibration step of 10^5 MC attempts where those values were adjusted to obtain an acceptance ratio of 50%. The steric-only behavior of the protein was used as a reference, so the simulation directly returned $B_{22}/B22,ST$, where $B22,ST$ represents the steric-only second osmotic virial coefficient (*i.e.,* the value due to only protein excluded volume contributions) as explained elsewhere and below.[12,23] The following simulations were performed: $B_{22}/B22,ST$ was calculated for ε_{SR} values between 0 and 2 k_BT, ψ_i values between 0 and 1.5, and TIS values between 0 and 510 mM for both CG models. The obtained $B_{22}/B22,ST$ values were compared to experimental values for further parameter tuning (see Discussion below). Statistical uncertainties were estimated by performing 5 independent simulations for each model parameter set and a given solution condition. The standard deviation was used as the estimate of statistical uncertainty, including error propagation.

High c_2—Transition Matrix Monte Carlo (TMMC) was used to compute R^{ex}/K *vs* c_2 profiles for c_2 values above 10 g/L using the methods described below. The particular parameter values were based on the MSOS simulations and experimental B_{22} values described above and in the next section.[12] The simulations were carried out in a grand-canonical (osmotic) system. An initially uniform concentration probability distribution was used, which was subsequently reconstructed at the end of each cycle until it converged to the equilibrium probability distribution, with each cycle being defined as 10^6 MC attempts. A MC attempt consisted of one of the following randomly selected moves: a translation, a rotation or a molecule insertion or deletion. Translations and rotations represented 30% of all MC attempts, while deletions and insertions represented the remaining 70%. Temperature was held constant at 25 °C. Preliminary simulations were used to find an adequate value of the reference chemical potential, depending on the parameter value(s) (see below). Due to boundary effects, G_{22} was observed to depend on the box size for $c_2 > 30$ g/L and box lengths below 50 nm.[49,50] Consequently, a box length from 60 nm to 180 nm was used, where simulated values of G_{22} were not found to significantly depend on the box length, and larger box sizes were used for low-c_2 conditions to decrease the noise on simulated G_{22} values. The simulation box was started with an empty system and G_{22} values were calculated by using histogram reweighting on the c_2 probability distribution using equation 6, where $\langle N_2 \rangle$ represents the average number of protein molecules in the system, and $\langle N_2^2 \rangle - \langle N_2 \rangle^2$ represents the average fluctuations in the number of protein molecules for a given choice of protein chemical potential.

$$c_2 G_{22} = \frac{\langle N_2^2 \rangle - \langle N_2 \rangle^2}{\langle N_2 \rangle} - 1 \quad (6)$$

Excess Rayleigh profiles and $S_{q=0}$ values were obtained by inserting simulated values of $c_2 G_{22}$ in equation 1 with a M value of 146.5 kDa and assuming that $M_{w,app}$ and M_w are equal.[3,12,23]

Steric-only behavior as a reference state and the corresponding equation of state (EoS)

The steric-only behavior can be used as a reference state, as this corresponds to the baseline level of interactions any macromolecule would experience in solution if non-steric repulsions counter-balance any attractions.[23,28] In the case of low-c_2 behavior, steric interactions are c_2-independent. This results in a B_{22} value of ~0.01 L/g, termed B22,ST in the remainder of this work.[23] This value can be used to normalize B_{22} values across different solution formulation. Consequently, B_{22}/B22,ST values above 1 are representative of additional repulsions (beyond sterics), which will be termed "net-repulsive" in the remainder of this report. Similarly, $B_{22}/B_{22,ST}$ values below 1 are representative of attractions that overcome the steric-only behavior of the protein, and this will be termed "net-attractive" in the remainder of this report.

For higher c_2 values, it is necessary to develop expressions to compute the c_2-dependent steric interactions. Previous work computed those interactions as a function of c_2 using several CG models, including those used in this work.[23] Grand Canonical MC simulations were carried out to obtain values of c_2 *vs* μ_2, with μ_2 representing the protein chemical potential. These results can be used to compute a steric-only EoS to analytically calculate the high-c_2 behavior due to only steric interactions. This can then be used as a c_2-dependent reference state instead of the ideal gas or non-interacting behavior (*i.e.*, $B_{22} = 0$ or $S_{q=0} = 1$). The analytical steric EoS was based on the virial expansion as is done in the McMillan-Mayer solution theory (referred to as VE below, as short-hand for Virial Expansion).[28,29,36,46] The VE-EoS provides a simple 4th-order polynomial as shown in equation 7, which can be used to analytically calculate thermodynamic properties. This approach is expected to strongly deviate from real multi-body behavior as c_2 increases beyond the range of simulated concentrations (*i.e.*, above 180 g/L).

$$\frac{\Pi_{ST} M_w}{c_2 RT} = A_1 + A_2\eta + A_3\eta^2 + A_4\eta^3 + A_5\eta^4 \quad (7)$$

In equation 7, Π_{ST} represents the osmotic pressure of the protein in solution, R is the gas constant and T is the absolute temperature. η corresponds to the protein volume fraction in solution (= $v_2 c_2$). v_2 is the protein molecular volume and was previously computed using atomistic simulations and found equal to 0.924 mL/g (see Table 2 in Ref. 23) for a series of MAb molecules.[23] The analytical steric-only EoS models was fitted to simulated data by minimizing the error in both the isothermal compressibility ($\kappa_{T,ST} = (\partial c_2/\partial \Pi_{ST})_T * c^{-1}2$) and the osmotic pressure (Π_{ST}) as a function of protein volume fraction ($\kappa_{T,ST}$ *vs* η and Π_{ST} *vs* η, respectively) for values of $\eta < 0.165$ (*i.e.*, $c_2 < 180$ g/L). The resulting parameters obtained from error minimization are shown in Table 3. The reader must take into account that extrapolating to higher volume fractions (or c_2 values) above 180 g/L is discouraged as additional parameters might be required to capture more crowded environments.[20,23,29,46] Final steric-only G_{22} values (G22,ST) as a function of c_2 were calculated using equation 8 combined with equation 7 and the parameters in Table 3.

$$c_2 G_{22,ST} = \frac{RT}{M_w} c_2 k_{T,ST} - 1 = \frac{RT}{M_w}\left(\frac{\partial c_2}{\partial \Pi_{ST}}\right)_T - 1 \quad (8)$$

Average Relative Deviation (ARD) calculations and model validation

To evaluate the effectiveness of the present CG models to model or predict experimental SLS behavior, the average relative deviation (ARD) values were calculated for any given data set using equation 9,

$$ARD\,(\%) = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{x_i^{experimental} - x_i^{predicted}}{x_i^{experimental}}\right| \quad (9)$$

where n in Eq. 9 represents the number of data points (as opposed to its usage in Eq. 4) and x_i is the experimental or simulated value to be evaluated (*e.g.*, $B_{22}/B_{22,\mathrm{ST}}$ *vs TIS*, and R^{ex}/K *vs* c_2 in this work). As the ARD is a measurement of the average deviation between the model and the experimental data, a cutoff value between 10% and 20% was used below as a criterion for considering a prediction to be quantitatively accurate, as this average deviation can be considered a conservative estimate of the model prediction uncertainty, particularly given typical experimental uncertainties for SLS data.

Parameter tuning from low-c_2 data

To predict high-c_2 Rayleigh profiles from low-c_2 measurements using the formulated CG models, $B_{22}/B_{22,\mathrm{ST}}$ *vs TIS* experimental data were used to tune two model parameters: the strength of short-ranged non-electrostatic attractions ($\varepsilon_{\mathrm{SR}}$), and the correction factor to the theoretical charges ψ_i (see above).[12] Additional model parameters were refined in previous work based only on the geometry of multiple MAbs from their published crystal structures, as well as the homology model for this molecule (see above).[23] Under high-*TIS* conditions, electrostatic interaction are expected to be heavily screened (according to the Debye-Hückel theory), so $B_{22}/B_{22,\mathrm{ST}}$ values under these conditions can be used to set the value of $\varepsilon_{\mathrm{SR}}$ by combining experimental and simulated data (see Results and Discussion). Conversely, low-*TIS* conditions are expected to be dominated by electrostatic interactions, and this can be used to determine an optimized value of ψ_i. For simplicity, all ψ_i values were assumed to be equal for all the domains. Consequently, ψ_i will be referred as ψ in the remainder of this work, as an average correction factor for all theoretical charges at a given pH. Similarly, $\varepsilon_{\mathrm{SR}}$ will be used as an averaged short-ranged attraction strength (for solvation and hydration effects, and van der Waals attractions) and equal for all the molecule domains. This could be modified in future work to account for additional domain-specific information such as relative hydrophobicity, but was not used in the present work.

The following methodology was employed for the parameter tuning exercise: $B_{22}/B_{22,\mathrm{ST}}$ *vs TIS* values were simulated using both the HEXA and DODECA models using the MSOS algorithm for a range of [$\varepsilon_{\mathrm{SR}}$, ψ] pairs. ARD values were computed for each pair by

comparing experimental and simulated $B_{22}/B_{22,ST}$ *vs TIS* results. Experimental $B_{22}/B_{22,ST}$ values between - 0.05 and 0.05 were excluded from any ARD calculation to avoid heavy biasing on final ARD values. This was done to avoid the inherently larger experimental uncertainties in $B_{22}/B_{22,ST}$ results when those values are near zero. Surface plots of ARD *vs* [ε_{SR}, ψ] were constructed, where a funnel-like behavior is expected if there is a unique subset of [ε_{SR}, ψ] pairs that minimizes the ARD results. As there is uncertainty in experimental $B_{22}/B_{22,ST}$ values, the previous exercise would result in a parameter space of [ε_{SR}, ψ] pairs that can accurately mimic the experimental data, as shown below and in previous work.[12] Consequently, all simulated [ε_{SR}, ψ] pairs that resulted in ARD values below 20% were subsequently used to predict high-c_2 Rayleigh scattering behavior, creating a predicted "envelope" for R^{ex}/K rather than a single curve.

RESULTS

SLS was used to determine excess Rayleigh profiles (R^{ex}/K as a function of c_2) for a series of solution conditions. At low c_2, these measurements were used to determine B_{22} values as a function of *TIS* by changing the NaCl molarity. Figure 3 shows the results of B_{22} *vs TIS* for two series of formulations (buffer + NaCl, and buffer + 5% w/w sucrose + NaCl) and two pH values, 5 (panel A) and 6.5 (panel B), all measured for c_2 < 10 g/L. B_{22} values were normalized using the steric-only behavior from a 3D-homology model ($B_{22,ST}$ = 0.01 L/g) as a reference state and for easier comparison with MSOS simulations.

$B_{22}/B_{22,ST}$ *vs TIS* profiles differ quantitatively between pH 5 and pH 6.5, and between both sucrose concentrations. At pH 5 and low *TIS*, protein-protein interactions were relatively large and net-repulsive ($B_{22}/B_{22,ST}$ >> 1). Increasing *TIS* by adding NaCl decreases $B_{22}/B_{22,ST}$ until reaching a constant value for *TIS* values above approximately 300 mM. At pH 6.5 and low *TIS*, protein-protein interactions were net-attractive, relative to steric-only interactions ($B_{22}/B_{22,ST}$ < 1). Increasing *TIS* by adding NaCl decreases the magnitude of $B_{22}/B_{22,ST}$ monotonically until reaching a constant value for *TIS* values above approximately 300 mM in all situations. $B_{22}/B_{22,ST}$ values at high *TIS* (>300 mM) were the same for both pH values but less attractive (less negative) with the addition of sucrose, with similar high-*TIS* results in the presence of sucrose for both pH values. Conversely, $B_{22}/B_{22,ST}$ values differ significantly across pH at low *TIS* (below 50 mM), where pH 5 resulted in larger (more repulsive) $B_{22}/B_{22,ST}$ values than pH 6.5. The addition of sucrose did not result in statistically distinguishable behavior at low *TIS* across pH values.

Insets in Figs. 3A and 3B show the experimental R^{ex}/K *vs* c_2 (high-c_2) results for the formulations presented in Table 4 and that correspond to the low-c_2 measurements in the main panels.[3] Additionally, the steric-only behavior for this molecule is shown as a reference, as computed using the VE EoS model (equations 7 and 8, and Table 3).[23] The experimental values in the insets were shown previously,[3] but with a new correction in the present work by using updated dn/dc_2 values from the refractive index measurements (see Methods). In particular, the results for formulations with 5% w/w sucrose (red symbols) significantly changed based on the updated dn/dc_2 values. In those cases, the dn/dc_2 values were found to be 15% higher in the presence of sucrose than previously reported, so results

in Fig. 3 at high c_2 are ~30% lower in magnitude than previously reported.[3] The results without sucrose are effectively unchanged from prior work.

In summary, the results in Fig. 3A (pH 5) show that R^{ex}/K profiles for both buffer-only and sucrose formulations are net-repulsive (R^{ex}/K values below the steric-only behavior) without added NaCl, with R^{ex}/K profiles for sucrose below (more repulsive than) those for buffer-only. Adding 100 mM NaCl results in a large increase in R^{ex}/K values and brings the high-c_2 values well above (more attractive than) the steric-only behavior. For Fig. 3B (pH 6.5), the buffer-only results overlap with the steric-only behavior at low c_2, and adding 5% w/w sucrose results in a decrease in the R^{ex}/K profiles (increase in repulsions) while adding 100 mM results in an increase in R^{ex}/K values (increase in attractions). In each formulation (buffer-only, buffer + sucrose, and buffer + NaCl), R^{ex}/K values at pH 5 are lower in magnitude than those at pH 6.5 for equal c_2 values.

Figures 4 and 5 show a comparison of $B_{22}/B_{22,ST}$ *vs TIS* between experiments and simulations for the HEXA and DODECA models coupled with MSOS simulations, respectively. The experimental data and formulations are the same as those presented in Fig. 3 and Table 1. Shaded areas in the main panels represent the simulated $B_{22}/B_{22,ST}$ *vs TIS* profiles obtained from ARD values below 20% (gray minima in surface plots in the insets). The insets show colored surface plots of ARD as a function of ε_{SR} and ψ values. From those parameter-response surfaces, one can identify a narrow parameter space (values for ε_{SR} and ψ, also referred as [ε_{SR}, ψ] pairs) that accurately captures the low-c_2 experimental behavior with a given CG model. That range of [ε_{SR}, ψ] pairs are further used to predict high-c_2 SLS behavior.

Figures 6 and 7 show a comparison of the experimental and predicted high-c_2 excess Rayleigh scattering results as a function of protein concentration. The results from the HEXA and DODECA models are based on the TMMC simulations for the parameter space obtained by fitting low-c_2 data (*cf.*, Figs. 4 and 5). Fig. 6 corresponds to parameters from Fig. 4 and the HEXA model. Fig. 7 corresponds to parameters from Fig. 5 and the DODECA model. The formulation conditions are the same as shown in Table 4. Shaded areas in the main panels in Figs. 6 and 7 represent the confidence intervals of the predicted R^{ex}/K vs c_2 profiles using model parameters that resulted in an ARD value below 20% from the low-c_2 parameter tuning (gray regions in insets of Figs. 4 and 5). The symbols in Figs. 6 and 7 represent the same experimental data shown in the insets in Fig. 3, including 95% confidence intervals as error bars. By visual inspection, the parameters obtained at low-c_2 allow the CG models to be predictive of the high-c_2 behavior within a 20% average deviation from 10 to 150 g/L. The steric-only behavior at high c_2 is also included as a reference in Figs. 6 and 7. None of the predictions in Figs. 6 and 7 utilize experimental data from high c_2 as inputs to the models, but do require knowledge of the $B_{22}/B_{22,ST}$ values at each given solution condition. For reference, parameter ranges that predicted the high-c_2 experimental behavior within 20% uncertainty are shown in Table 4 for both CG models.

Figure 8 shows $\hat{V}_2$ values as a function of $c_3\hat{V}_3$ in acetate, pH 5 (panel A), and in histidine, pH 6.5 (panel B); with sucrose as the added excipient (component 3). For both panels, the addition of sucrose leads to a monotonic and linear decrease in protein partial specific

volume (*i.e.*, a negative slope for $\hat{V}_2$ *vs* $c_3\hat{V}_3$). The values for fitted intercepts using equation 3 are 0.708 ± 0.006 mL/g for pH 5 and 0.709 ± 0.004 mL/g for pH 6.5. Similarly, the slopes are equal to −1.0 ± 0.1 mL/g for pH 5, and −1.08 ± 0.06 mL/g for pH 6.5. Both pH values returned the same slope and intercept values within their statistical uncertainties. The negative slopes clearly indicate a similar level of preferential accumulation or solvation by sucrose around the protein surface for both pH values.

DISCUSSION

Protein-protein "weak" interactions are mediated by the solution environment the protein is subjected to. These interactions have three main contributions: sterics or excluded volume effects (repulsive); short-ranged non-electrostatic van der Waals interactions and hydration/solvation effects (net attractive or repulsive); and electrostatic interactions (both attractive and repulsive). Among these, only the latter should be significantly affected by the solution ionic environment (*via* charge screening) if one neglects ion binding effects.[12,23,43,45] Consequently, the decrease in $B_{22}/B_{22,ST}$ values and plateau behavior that are observed as *TIS* increases in Fig. 3 can be attributed to a screening effect of strong charge-charge repulsions with the addition of NaCl, as described at least qualitatively by the Debye-Hückel theory.[12,23,46] This agrees with previously published experimental behavior of a number of proteins as a function of ionic environment.[7,11,12,51,52] Since $B_{22}/B_{22,ST}$ converges towards values less than 1 at high-*TIS* conditions, there should be short-ranged non-electrostatic attractions present in the molecule to overcome the steric repulsions.[23] Additionally, both pH 5 and 6.5 results converged towards equal $B_{22}/B_{22,ST}$ values for *TIS* > 300 mM, suggesting that electrostatic contributions are completely screened and the solvation effects and van der Waals attractions present between the molecules are not affected by the differences in buffer chemistry and pH.[12,39] Conversely, the difference in interactions at lower *TIS* values suggests different electrostatic behaviors with the change in both buffer-type and pH, going from strongly repulsive to mildly attractive (relative to steric-only interactions, $B_{22}/B_{22,ST} = 1$) as pH increases.[12]

It is commonly accepted that, for most proteins, the total effective protein charge approaches zero as the pH of the solution approaches the isoelectric point (pI) of the molecule. Consequently, the strength of electrostatic repulsions (caused by strong charge-charge repulsions) will decrease as the pH approaches the pI of the molecule. The pI of several IgG1 molecules have been reported to lie between 7.5 and 8.5, and the present molecule has a theoretical pI of 7.9 (calculated as explained in the Methods). Consequently, the decrease in repulsions with increased pH is expected based on a decrease in the total effective charge of the molecule (mostly due to deprotonated histidine residues at pH 6.5 in comparison to pH 5) and the change in the ion clouding/de-clouding that this might lead to.[45,53,54] This is highlighted in Fig. 2, where the values of the theoretical charges decrease from pH 5 to 6.5, and in some case (as in the C_H3 domain) this can cause a shift in sign. Since short-ranged non-electrostatic attractions are present at both pH values (net-attractions dominate at high *TIS*), this decrease in the effective protein charge with increasing pH would lead to stronger net-attractions (relative to sterics) at low *TIS*, as seen in Fig. 3.

Collecting $B_{22}/B_{22,ST}$ data as a function of *TIS* allows one to gain insights into two of the main contributions to protein-protein solution interactions: (a) the strength and sign of net electrostatic interactions (observed at low *TIS*) and (b) the strength of short-ranged non-electrostatic attractions (observed at high *TIS*).[12,23] This is better visualized in the insets of Figs. 4 and 5, where data shown in Fig. 3 were used to tune the model parameters for the HEXA (Fig. 4) and DODECA (Fig. 5) models as described above. In Figs. 4 and 5, all surface response plots show a funnel-like behavior, where a small subset of [ε_{SR}, ψ] pairs is capable of accurately modeling the experimental data within a 20% ARD. Additionally, a smaller subset of [ε_{SR}, ψ] pairs was also found to model the experimental data to much higher resolution in some situations (ARD ~ 0%) when the uncertainty in the experimental data is taken into account. These results showed the capability of the present CG models to quantitatively capture two-particle behavior as a function of *TIS* and at low c_2. Additionally, Figs. 4 and 5 show that the currently proposed electrostatic interaction model is capable of accurately modeling the $B_{22}/B_{22,ST}$ behavior from low to high *TIS* as well as the plateau in $B_{22}/B_{22,ST}$ values that occurs at high *TIS* (above 300 mM).

Differences in the values of the parameters within the gray regions are observed when comparing insets in Figs. 4 and 5 across pH (panels A *vs* C, and B *vs* D), added sucrose (panels A *vs* B, and C *vs* D) and model-type (Fig. 4 vs Fig. 5). By comparing results across pH values for both Figs. 4 and 5, one observes that the only parameter that is significantly affected is ψ, as it shifts from ~0.35 at pH 5 to ~0.65 at pH 6.5 for the HEXA model, and from ~0.65 at pH 5 to ~1.0 at pH 6.5 for the DODECA model. This increase in ψ can potentially be explained by a decrease in ion binding due to smaller net charges in the protein molecule (see Fig. 2 and discussion above) and the possible changes in binding affinity of the ions.[45] Consequently, the accumulation of ions around the protein may change with changes in pH, causing an increase in ψ as the solution charges approach their theoretical value since $\psi \rightarrow 1$ as $q_{i,eff} \rightarrow q_i$. From the results in Fig. 4 and 5, this change in ψ is only observed across changes in pH, as ψ remains constant when comparing across sucrose content (panels A *vs* B, and C *vs* D). This suggests that the addition of sucrose should only induce a significant non-electrostatic effect to the protein solution behavior. This is also observed in the experimental data (Fig. 3), where the values of $B_{22}/B_{22,ST}$ are equal for both buffer-only and buffer + sucrose formulations at low *TIS*, but diverge as *TIS* increases. Comparing panels A and B, and C and D in Fig 3., the addition of sucrose correlates with a decrease in the value of ε_{SR}. For the HEXA model, ε_{SR} goes from ~1.1 k_BT for buffer-only conditions to ~1.0 k_BT for buffer + sucrose conditions for both pH values (Fig. 4). Similarly, for the DODECA model the resulting value shifts from ~0.72 k_BT to ~0.64 k_BT (Fig. 5). This decrease in ε_{SR} and increase in $B_{22}/B_{22,ST}$ at high *TIS* with added sucrose suggests changes in the hydration shells of the protein in the form of protein-sucrose interactions, and this will be discussed further below.

Comparing Figs. 4 and 5 also shows that the ε_{SR} values within the gray areas are always lower in magnitude for the DODECA model (0.62 - 0.78 k_BT) than for the HEXA model (1.0 -1.2 k_BT) for all simulated formulations in Table 1. This is due to the decrease in the number of beads/domains by moving from the DODECA to the HEXA model, analogous to previous work with globular proteins.[23,55] Conversely, the magnitude of ψ increases in the DODECA model in comparison to the HEXA model (see numbers above). Although the

discussion for ε_{SR} in terms of the differences in number of domains for HEXA *vs* DODECA also applies to ψ, changes in the values of the charges also play a relevant role in this case. As expected, charges in the HEXA model are effectively twice the magnitude of those in the DODECA model (see Table 2 and Fig. 2). Charge-charge interactions were modeled *via* equation 5, where the electrostatic potential energy is proportional to the product of the charges. Consequently, doubling the value of the charges (by going from DODECA to HEXA) would induce an increase in potential energy by a factor of 4, which can compensate and overcome the decrease in the number of simulated domains/beads (from 12 to 6 in this case).

The present CG models also locate the charges in the center of the bead/domain in contrast to the real protein charges on the surface. A larger charge can be conducive of stronger counter ion accumulation due to strong charge-charge attractions and territorial ion accumulation, so lower ψ values are expected for more coarse-grained (less structurally detailed) models. Although this is intended to represent ion accumulation on protein solutions, the reader must be cautious that both ε_{SR} and ψ values used in this work and for these CG models are model specific, and likely will differ if one changes the structural resolution of the models (either higher or lower resolution).[12,23,55] In addition, one must realize that ψ is a “lumped” parameter that attempts to capture deviations from theoretical and true net charge that could have origins from multiple sources (e.g., not only territorial ions but also hydration effects or ion-specific effects). As such, caution is needed to avoid over-interpreting the underlying causes for ψ values that deviate significantly from 1.

It is anticipated that the same qualitative behavior discussed above might apply at higher c_2. At pH 5, both buffer-only and buffer + sucrose excess Rayleigh and $S_{q=0}$ profiles are observed to lie below the steric-only behavior (net-repulsive, see Figs. 3, 6 and 7). This agrees with the low-c_2 behavior as $B_{22}/B_{22,ST} \sim 1.8$ (*cf.*, Fig. 3). Nevertheless, buffer + sucrose conditions were observed to be more repulsive (lower $S_{q=0}$ and R^{ex}/K *vs* c_2 values) than buffer-only conditions, and this deviation is more pronounced as c_2 increases. This does not correlate with low-c_2 measurements, as both pH conditions resulted in equal $B_{22}/B_{22,ST}$ values within their experimental uncertainties (Fig. 3A). At pH 6.5, the buffer-only conditions remain net-attractive at high c_2, but sucrose conditions are net-repulsive between 10 and 120 g/L, converging towards the steric-only behavior at higher c_2. These results could not be predicted from low-c_2 information alone, as all measured $B_{22}/B_{22,ST}$ results were net-attractive at pH 6.5, with equal $B_{22}/B_{22,ST}$ results for sucrose and buffer-only formulations at low *TIS* (see discussion above). For all formulations with 100 mM NaCl, the excess Rayleigh profiles show net-attractive behavior for both pH values, with stronger attractions at pH 6.5 in comparison to pH 5 and in good agreement with the results at low c_2.

Although there are an increasing number of studies that focus on experimentally correlating low-c_2 measurements with high-c_2 protein physicochemical behavior, results in Fig. 3 highlight some of the short-comings of these approaches, as interactions and solution behavior might change as the solution transitions from low to high c_2. As shown in Figs. 6 and 7, the changes in $S_{q=0}$ are not monotonic (not constant G_{22}). This leads to weaker net-attractions relative to those seen only at low c_2. This is of greater relevance during screening of drug candidates and formulations during early stages of development, where limited

access to protein material necessitates measurements at low-c_2 (dilute solution conditions). As the solution is concentrated, the solution behavior is expected to be influenced heavily by the steric (shape-based) contributions based on general arguments from the statistical mechanics of liquids.[23,28,56] Thus, the shape of the molecule is expected to greatly affect the way MAb molecules interact under concentrated conditions.[23] The addition of short-ranged interactions (either attractive or repulsive) mediates preferentially interacting domains, which might lead to increases in viscosity as reported in previous work.[7,12,57,58] These two effects (enhanced short-ranged interactions and packing behavior) are reasonably well captured by the HEXA and DODECA model.

The results in Figs. 4 and 5 allow one to obtain a small family of [ε_{SR}, ψ] pairs that can be used to evaluate the predictive capabilities of the HEXA and DODECA models at high-c_2 conditions as shown in Figs. 6 and 7. This is highlighted in Table 4, where the [ε_{SR}, ψ] pairs that best predict the high-c_2 SLS data lie within the parameter space obtained during the low-c_2 parameter tuning (*cf.*, insets in Figs. 4–5). The results in Figs. 6–7 show that both models are capable of accurately predicting, not simply regressing, excess Rayleigh scattering profiles up to 150 g/L of protein concentration. Small qualitative difference can be observed between the results for the HEXA (Fig. 6) and DODECA (Fig. 7) models. While the HEXA model results in smaller deviations at pH 5 than the DODECA model, the opposite is observed at pH 6.5. At pH 5, there is a noticeable deviation for values above 120 g/L for the DODECA model, where the predicted R^{ex}/K profiles qualitatively deviate with increasing c_2 from the experimental data (Fig. 7A). This behavior might be caused by the geometry of the models and the ease of packing of each model. The HEXA model locates all of its beads on a single plane while the DODECA model increases the complexity of the model by extending it to two planes (see Fig. 1). This increase in geometrical complexity potentially adds stronger packing limitations at high c_2 for the DODECA model. Additionally, these two models were simulated by neglecting the flexibility of the hinge region due to limited access to data that can be used to refine such behavior (*e.g.*, SANS or SAXS). The hinge flexibility might correct for these discrepancies at higher c_2 for the DODECA model by easing the packing constraints of such model as suggested in previous work on simpler MAb models.[16,23] However, the addition of a flexible hinge would pose additional computational challenges in terms of convergence or precision of the simulations. That would be further exacerbated if one permitted full chain flexibility and local unfolding in the simulations.[16,59,60]

Finally, sucrose has been historically categorized as a preferentially excluded excipient.[38,42] Preferential exclusion from the protein surface causes a reduction in the available free-volume for the protein molecules. This induces stronger attractions between proteins driven by the steric repulsion between protein and sucrose molecules (in the form of depletion forces). This behavior has been identified as the main mechanism behind protein flocculation and “salting-out” in the presence of preferentially excluded excipients. However, results presented in Figs. 3–7 show the opposite behavior. Stronger repulsions were observed for solution with sucrose than those when sucrose was absent. This increase in repulsions was caused by weaker short-ranged non-electrostatic attractions, as exemplified by the fact that the magnitude of ε_{SR} obtained from parameter tuning was effectively lower in magnitude for sucrose conditions than those without sucrose (see

discussion above) at both tested pH values. Similarly, results in Fig. 8 (negative slopes at both pH 5 and 6.5) suggest preferential accumulation or solvation by sucrose around the protein surface, in agreement with previously reported results where sucrose was also found to solvate proteins.[39]

Consequently, by combining the results from Figs. 3–8 along with the simulations results offered by the CG modeling, one can hypothesize how addition of sucrose induces stronger repulsions between protein molecules. The accumulation of sucrose around the protein surface (*i.e.*, protein solvation by sucrose molecules) might be affecting the way proteins interact through two different mechanism. First, the solvation by sucrose around the protein surface displaces water molecules from the hydration layers (protein dehydration), decreasing the gain in solution free energy upon protein-protein close contact.[31] Second, sucrose has a larger molecular diameter than water,[39] and a protein solvated by sucrose might experience an increase in its effective excluded volume, increasing the strength of apparent steric repulsions between proteins. Both contributions are expected to be present upon addition of sucrose, but the current experimental data does not allow one to resolve which mechanism might be dominating the observed solution behavior.

SUMMARY

Static light scattering was used to quantify “weak” protein-protein interactions of an IgG1 molecule as a function of protein concentration for a range of pH, *TIS* values, and sucrose concentration. Experimental measurements showed both net-repulsive and net-attractive protein interactions at low *TIS*, and at low- to high-c_2 conditions. Two coarse-grained molecular models were tested to evaluate their potential to predict excess Rayleigh profiles and zero-q structure factors at high c_2. Additional high-precision density measurements were used to further assess the non-electrostatic effect of adding sucrose to the solution. Low-concentration results showed that the IgG1 molecule exhibits net-repulsive behavior at low *TIS* and pH 5, which transitions to net-attractive behavior as the solution *TIS* increases. At pH 6.5, the antibody showed weakly net-attractive behavior from low to high *TIS*, resulting in statistically equal values at both pH conditions for high *TIS* (> 300 mM). For all measured pH and *TIS* conditions, the addition of 5% w/w sucrose to the solution induced weaker net-attractions with increasing *TIS*. This behavior was also observed at high c_2, where formulations with 5% w/w sucrose always resulted in net-repulsive behavior at low *TIS*. This is counter to traditional expectations based on depletion interactions, in that preferentially excluded solutes drive protein-protein attractions, not repulsions. For conditions without sucrose present, buffer-only formulations shifted from net-repulsive behavior (relative to steric-only interactions) at pH 5 to net-attractive at pH 6.5, while formulations with 100 mM NaCl resulted in net-attractive behavior at both pH values.

In terms of model predictions from low to high c_2, the quantitative differences were not statistically significant, and therefore both models could be used to accurately predict high-c_2 behavior depending on the requirements of the user (*e.g.*, computational burden and molecular features). Finally, the combination of experimental and simulated SLS data with experimental partial specific volume data indicated preferential accumulation (solvation) of sucrose around the protein surface. This led to stronger repulsion between protein molecules

observed at both low and high c_2. The simulations results showed that both CG models, the HEXA and DODECA models, were able to quantitatively or semi-quantitatively predict the experimental data based solely on parameters obtained by combining $B_{22}/B_{22,ST}$ *vs TIS* experimental and simulated data collected at low c_2.

Acknowledgments

C.C.R. and C.J.R. gratefully acknowledge Bristol-Myers Squibb, the National Science Foundation (CHEM 1213728), and the National Institutes of Health (R01 EB006006) for financial support.

References

1. Wang, W., Roberts, CJ. Aggregation of Therapeutic Proteins. John Wiley & Sons, Inc; 2010.
2. Uchiyama S. Liquid formulation for antibody drugs. Biochim Biophys Acta - Proteins Proteomics. 2014; 1844(11):2041–2052.
3. Ghosh R, Calero-Rubio C, Saluja A, Roberts CJ. Relating Protein-Protein Interactions and Aggregation Rates From Low to High Concentrations. J Pharm Sci. 2016; 105(3):1086–1096. [PubMed: 26928400]
4. Shire SJ, Shahrokh Z, Liu J. Challenges in the development of high protein concentration formulations. J Pharm Sci. 2004; 93(6):1390–1402. [PubMed: 15124199]
5. Weiss IVWF, Young TM, Roberts CJ. Principles, approaches, and challenges for predicting Protein Aggregation Rates and Shelf Life. J Pharm Sci. 2009; 98(4):1246–1277. [PubMed: 18683878]
6. Mitragotri S, Burke PA, Langer R. Overcoming the challenges in administering biopharmaceuticals: formulation and delivery strategies. Nat Rev Drug Discov. 2014; 13(9):655–672. [PubMed: 25103255]
7. Connolly BD, Petry C, Yadav S, Demeule B, Ciaccio N, Moore JMR, Shire SJ, Gokarn YR. Weak interactions govern the viscosity of concentrated antibody solutions: high-throughput analysis using the diffusion interaction parameter. Biophys J. 2012; 103(1):69–78. [PubMed: 22828333]
8. Raut AS, Kalonia DS. Pharmaceutical Perspective on Opalescence and Liquid-Liquid Phase Separation in Protein Solutions. Mol Pharm. 2016; 13(5):1431–1444. [PubMed: 27017836]
9. Raut AS, Kalonia DS. Opalescence in Monoclonal Antibody Solutions and Its Correlation with Intermolecular Interactions in Dilute and Concentrated Solutions. J Pharm Sci. 2015; 104(4):1263–1274. [PubMed: 25556561]
10. Neergaard MS, Kalonia DS, Parshad H, Nielsen AD, Møller EH, van de Weert M. Viscosity of high concentration protein formulations of monoclonal antibodies of the IgG1 and IgG4 subclass - prediction of viscosity through protein-protein interaction measurements. Eur J Pharm Sci. 2013; 49(3):400–410. [PubMed: 23624326]
11. Lilyestrom WG, Yadav S, Shire SJ, Scherer TM. Monoclonal antibody self-association, cluster formation, and rheology at high concentrations. J Phys Chem B. 2013; 117(21):6373–6384. [PubMed: 23560896]
12. Woldeyes MA, Calero-Rubio C, Furst EM, Roberts CJ. Predicting Protein Interactions of Concentrated Globular Protein Solutions Using Colloidal Models. J Phys Chem B. 2017; 121(18): 4756–4767. [PubMed: 28422503]
13. Yearley EJ, Godfrin PD, Perevozchikova T, Zhang H, Falus P, Porcar L, Nagao M, Curtis JE, Gawande P, Taing R, Zarraga IE, Wagner NJ, Liu Y. Observation of small cluster formation in concentrated monoclonal antibody solutions and its implications to solution viscosity. Biophys J. 2014; 106(8):1763–1770. [PubMed: 24739175]
14. Liu Y, Chen W-R, Chen S-H. Cluster formation in two-Yukawa fluids. J Chem Phys. 2005; 122(4): 44507. [PubMed: 15740267]
15. Yearley EJ, Zarraga IE, Shire SJ, Scherer, Thomas M, Gokarn Y, Wagner NJ, Liu Y. Small-angle neutron scattering characterization of monoclonal antibody conformations and interactions at high concentrations. Biophys J. 2013; 105(3):720–731. [PubMed: 23931320]

16. 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(49):12511–12518. [PubMed: 27973814]
17. Courtenay ES, Capp MW, Anderson CF, Record MT Jr. Vapor Pressure Osmometry Studies of Osmolyte - Protein Interactions : Implications for the Action of Osmoprotectants in Vivo and for the Interpretation of " Osmotic Stress " Experiments in Vitro. Biochemistry. 2000; 39(15):4455–4471. [PubMed: 10757995]
18. Schuck P. Analytical Ultracentrifugation as a Tool for Studying Protein Interactions. Biophys Rev. 2013; 5(2):159–171. [PubMed: 23682298]
19. Li L, Kumar S, Buck PM, Burns C, Lavoie J, Singh SK, Warne NW, Nichols P, Luksha N, Boardman D. Concentration dependent viscosity of monoclonal antibody solutions: Explaining experimental behavior in terms of molecular properties. Pharm Res. 2014; 31(11):3161–3178. [PubMed: 24906598]
20. 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. [PubMed: 21682538]
21. Blanco MA, Perevozchikova T, Martorana V, Manno M, Roberts CJ. Protein-protein interactions in dilute to concentrated solutions: α-Chymotrypsinogen in acidic conditions. J Phys Chem B. 2014; 118(22):5817–5831. [PubMed: 24810917]
22. 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. [PubMed: 24289039]
23. 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(27):6592–6605. [PubMed: 27314827]
24. Xie G, Timasheff SN. Temperature dependence of the preferential interactions of ribonuclease A in aqueous co-solvent systems: thermodynamic analysis. Protein Sci. 1997; 6(1):222–232. [PubMed: 9007994]
25. Barnett GV, Razinkov VI, Kerwin BA, Hillsley A, Roberts CJ. Acetate- and Citrate-Specific Ion Effects on Unfolding and Temperature-Dependent Aggregation Rates of Anti-Streptavidin IgG1. J Pharm Sci. 2016; 105(3):1066–1073. [PubMed: 26886346]
26. Ferrenberg A, Swendsen R. New Monte Carlo technique for studying phase transitions. Phys Rev Lett. 1988 Dec.61:2635–2638. [PubMed: 10039183]
27. Awotwe-Otoo D, Agarabi C, Wu GK, Casey E, Read E, Lute S, Brorson KA, Khan MA, Shah RB. Quality by design: Impact of formulation variables and their interactions on quality attributes of a lyophilized monoclonal antibody. Int J Pharm. 2012; 438(1–2):167–175. [PubMed: 22944306]
28. McQuarrie, DA. Statistical Mechanics. University Science Books; 2000.
29. Ben-Naim, A. Statistical Thermodynamics for Chemists and Biochemists. Plenum Press; 1992.
30. Schellman JA. The thermodynamics of solvent exchange. Biopolymers. 1994; 34(8):1015–1026. [PubMed: 8075384]
31. Priya MH, Merchant S, Asthagiri D, Paulaitis ME. Quasi-chemical theory of cosolvent hydrophobic preferential interactions. J Phys Chem B. 2012; 116(22):6506–6513. [PubMed: 22574766]
32. Ben-Naim A, Navarro AM, Leal JM. A Kirkwood-Buff analysis of local properties of solutions. Phys Chem Chem Phys. 2008; 10(18):2451–2460. [PubMed: 18446245]
33. Shimizu S. Estimation of excess solvation numbers of water and cosolvents from preferential interaction and volumetric experiments. J Chem Phys. 2004; 120(10):4989–4990. [PubMed: 15267361]
34. 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(2):278–291. [PubMed: 26707135]
35. Castellanos MM, McAuley A, Curtis JE. Investigating Structure and Dynamics of Proteins in Amorphous Phases Using Neutron Scattering. Comput Struct Biotechnol J. 2017; 15:117–130. [PubMed: 28138368]

36. Vafaei S, Tomberli B, Gray CG. McMillan-Mayer theory of solutions revisited: Simplifications and extensions. J Chem Phys. 2014; 141(15):154501. [PubMed: 25338903]
37. Neal BL, Asthagiri D, Lenhoff AM. Molecular Origins of Osmotic Second Virial Coefficients of Proteins. Biophys J. 1998; 75(5):2469–2477. [PubMed: 9788942]
38. Barnett GV, Razinkov VI, Kerwin BA, Blake S, Qi W, Curtis RA, Roberts CJ. Osmolyte Effects on Monoclonal Antibody Stability and Concentration-Dependent Protein Interactions with Water and Common Osmolytes. J Phys Chem B. 2016; 120(13):3318–3330. [PubMed: 27007711]
39. Calero-Rubio C, Strab C, Barnett GV, Roberts CJ. Protein Partial Molar Volumes in Multi-Component Solutions From the Perspective of Inverse Kirkwood-Buff Theory. J Phys Chem B. 2017; 121(24):5897–5907. [PubMed: 28525711]
40. Barnett GV, Razinkov VI, Kerwin BA, Blake S, Qi W, Curtis RA, Roberts CJ. Reply to Comment on “Osmolyte Effects on Monoclonal Antibody Stability and Concentration-Dependent Protein Interactions with Water and Common Osmolytes ” Reply to Comment on “Osmolyte Effects on Monoclonal Antibody Stability and Concentration- Dependent. J Phys Chem B. 2016; 120(43): 11333–11334. [PubMed: 27715047]
41. Rösgen J, Auton M. Comment on “Osmolyte Effects on Monoclonal Antibody Stability and Concentration-Dependent Protein Interactions with Water and Common Osmolytes. J Phys Chem B. 2016; 120(43):11331–11332. [PubMed: 27715051]
42. Kendrick BS, Chang BS, Arakawa T, Peterson B, Randolph TW, Manning MC, Carpenter JF. Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: role in restricted conformational mobility and compaction of native state. Proc Natl Acad Sci U S A. 1997; 94(22):11917–11922. [PubMed: 9342337]
43. Bratko D, Striolo A, Wu JZ, Blanch HW, Prausnitz JM. Orientation-averaged pair potentials between dipolar proteins or colloids. J Phys Chem B. 2002; 106(10):2714–2720.
44. Malmberg CG, Maryott AA. Dielectric constants of aqueous solutions of dextrose and sucrose. J Res Natl Bur Stand (1934). 1950; 45(4):299.
45. Laue T, Demeler B. A postreductionist framework for protein biochemistry. Nat Chem Biol. 2011; 7(6):331–334. [PubMed: 21587250]
46. Sandler, SI. An Introduction to Applied Statistical Thermodynamics. John Wiley & Sons, Inc; 2010.
47. Po HN, Senozan NM. The Henderson-Hasselbalch Equation: Its History and Limitations. J Chem Educ. 2001; 78(11):1499–1503.
48. Schultz AJ, Kofke DA. Virial coefficients of model alkanes. J Chem Phys. 2010; 133(10):104101. [PubMed: 20849158]
49. Frenkel, D., Smit, B. Understanding Molecular Simulation. In: Press, A., editor. Vol Second Edi. 2002.
50. Shen V, Errington J. Metastability and instability in the Lennard-Jones fluid investigated by transition-matrix Monte Carlo. J Phys Chem B. 2004; 108(51):19595–19606.
51. Kim N, Remmele RL, Liu D, Razinkov VI, Fernandez EJ, Roberts CJ. Aggregation of anti-streptavidin immunoglobulin gamma-1 involves Fab unfolding and competing growth pathways mediated by pH and salt concentration. Biophys Chem. 2013; 172:26–36. [PubMed: 23334430]
52. Arzenšek D, Kuzman D, Podgornik R. Hofmeister Effects in Monoclonal Antibody Solution Interactions. J Phys Chem B. 2015; 119(33):10375–10389. [PubMed: 26207819]
53. Zhang F, Skoda MWA, Jacobs RMJ, Zorn S, Martin RA, Martin CM, Clark GF, Weggler S, Hildebrandt A, Kohlbacher O, Schreiber F. Reentrant condensation of proteins in solution induced by multivalent counterions. Phys Rev Lett. 2008; 101(14):148101. [PubMed: 18851577]
54. Roberts D, Keeling R, Tracka M, Van Der Walle CF, Uddin S, Warwicker J, Curtis R. The role of electrostatics in protein-protein interactions of a monoclonal antibody. Mol Pharm. 2014; 11(7): 2475–2489. [PubMed: 24892385]
55. 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(3): 763–770. [PubMed: 23245189]
56. Zimmerman S, Minton A. Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys. 1993:27–65.

57. Schmit JD, He F, Mishra S, Ketchem RR, Woods CE, Kerwin BA. Entanglement model of antibody viscosity. J Phys Chem B. 2014; 118(19):5044–5049. [PubMed: 24758234]

58. Tomar DS, Kumar S, Singh SK, Goswami S, Li L. Molecular basis of high viscosity in concentrated antibody solutions: Strategies for high concentration drug product development. MAbs. 2016; 8(2):216–228. [PubMed: 26736022]

59. Harris LJ, Skaletsky E, McPherson A. Crystallographic structure of an intact IgG1 monoclonal antibody. J Mol Biol. 1998; 275(5):861–872. [PubMed: 9480774]

60. Bongini L, Fanelli D, Piazza F, De Los Rios P, Sandin S, Skoglund U. Freezing immunoglobulins to see them move. Proc Natl Acad Sci U S A. 2004; 101:6466–6471. [PubMed: 15082830]

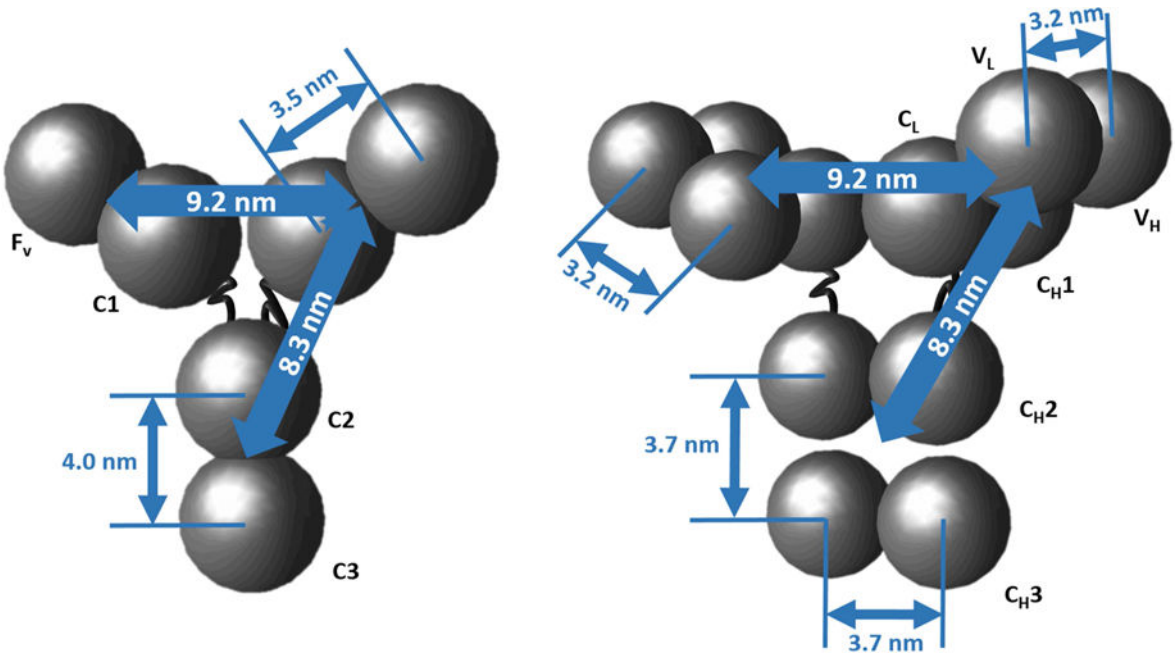


Figure 1.

Geometries of HEXA and DODECA coarse-grained models. Domain diameters (σ_i) were equal to 3.5 nm for the DODECA and 4.4 nm for the HEXA model.

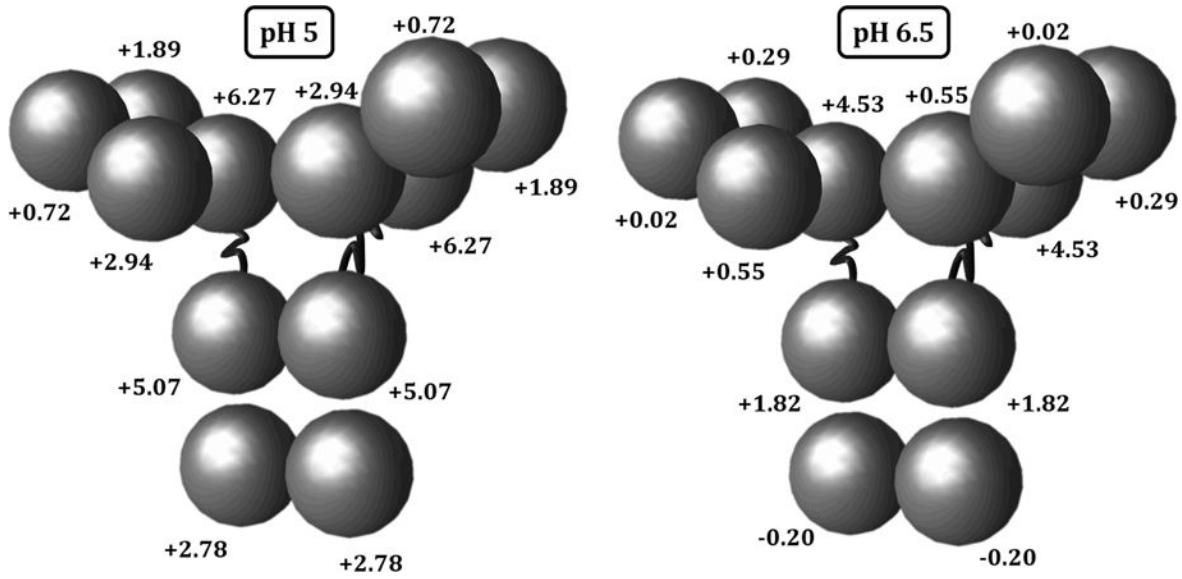
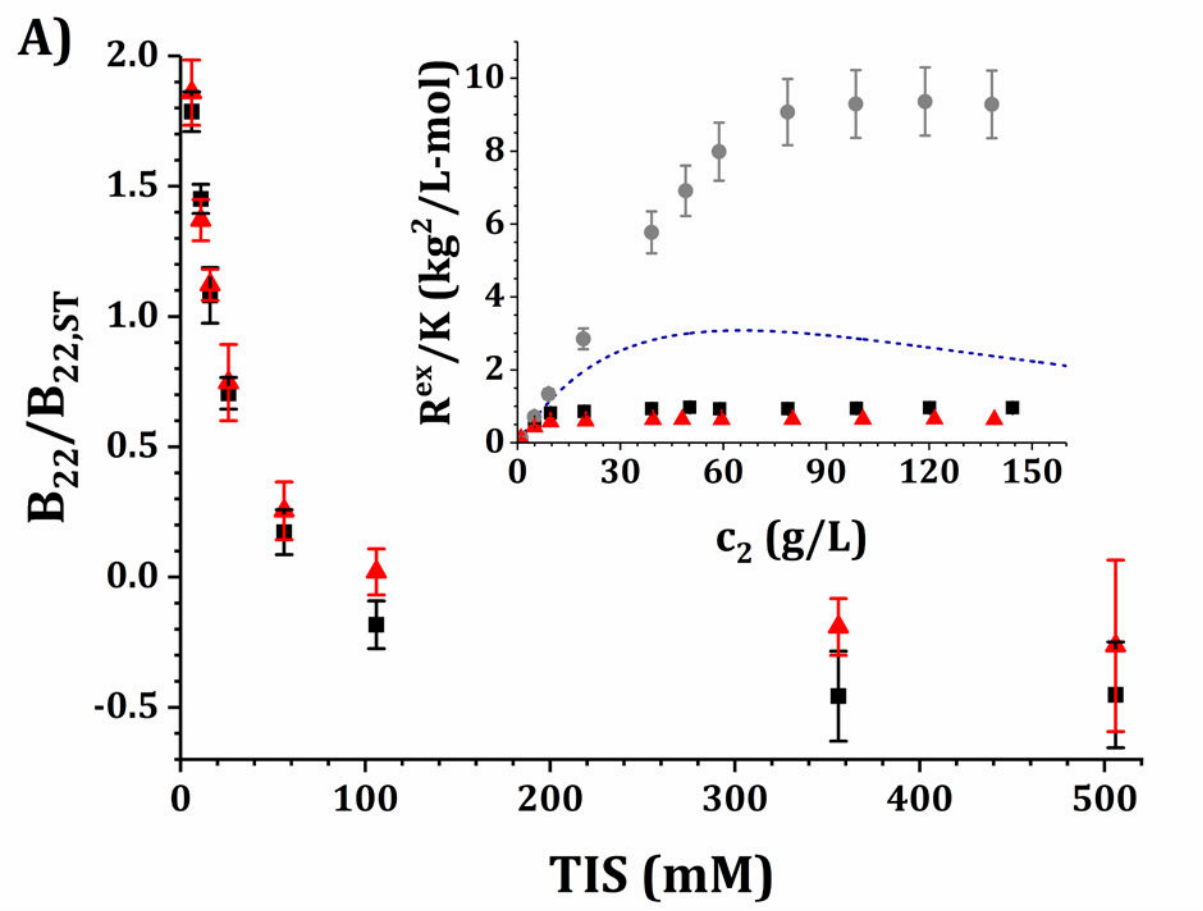


Figure 2.

Theoretical charge distribution for the DODECA model at pH 5 and 6.5 for the IgG1 molecule in this work.



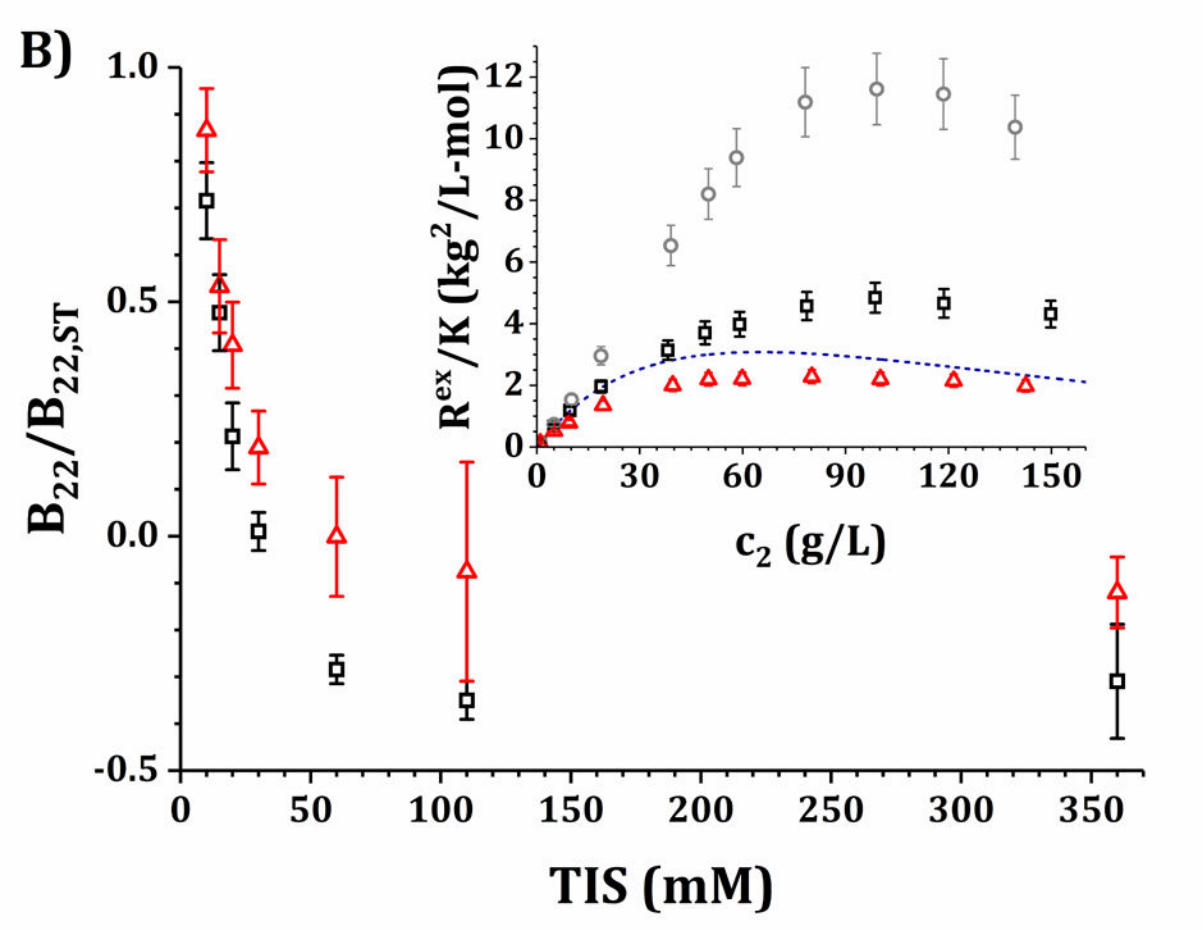
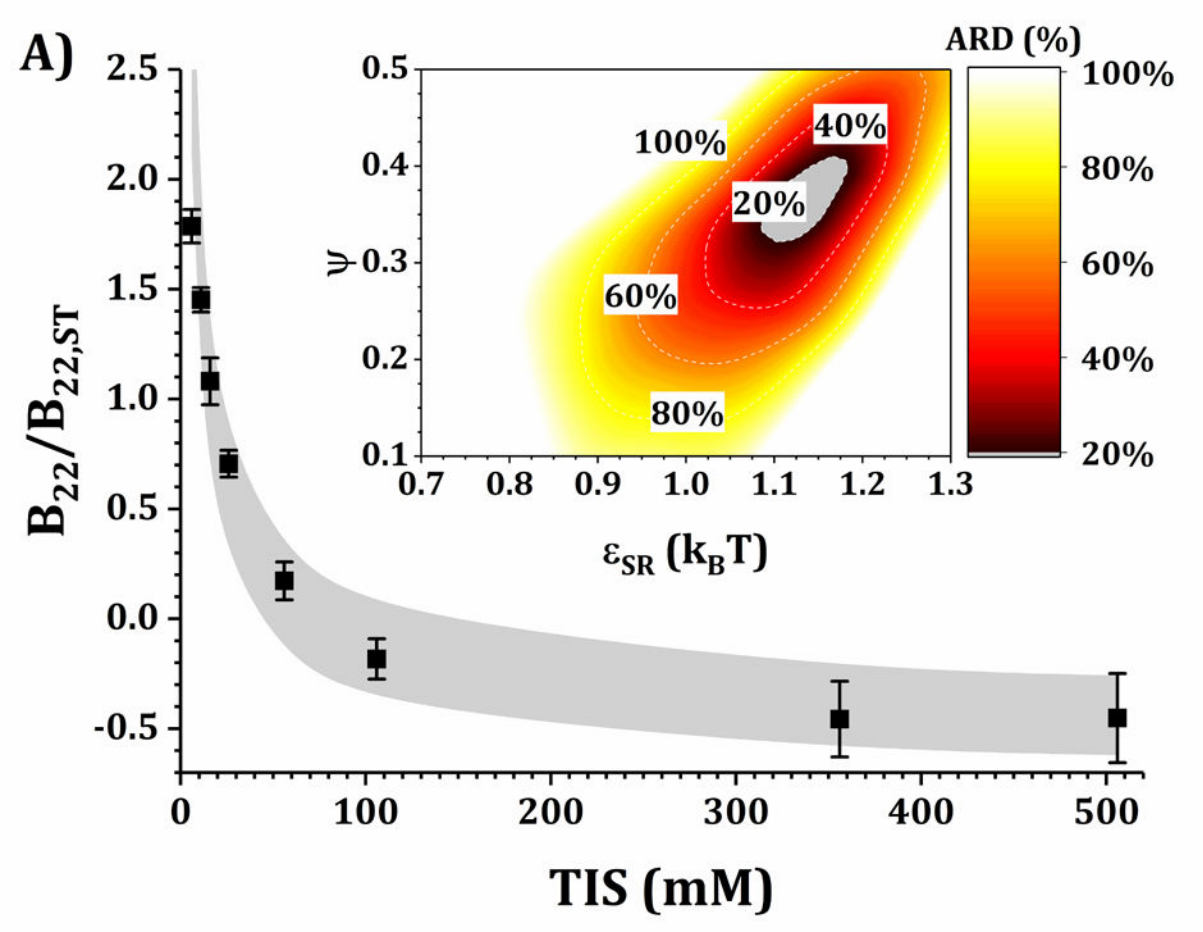
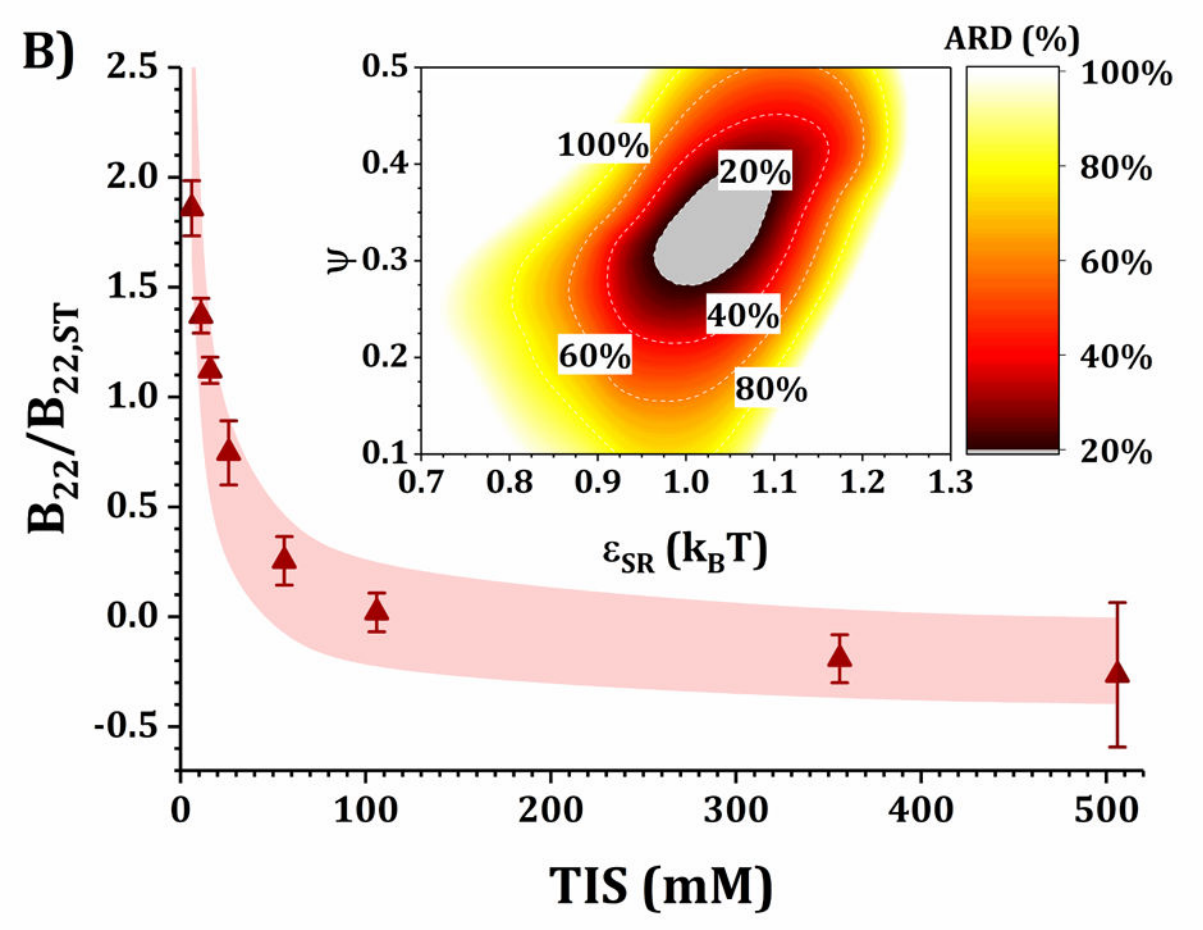
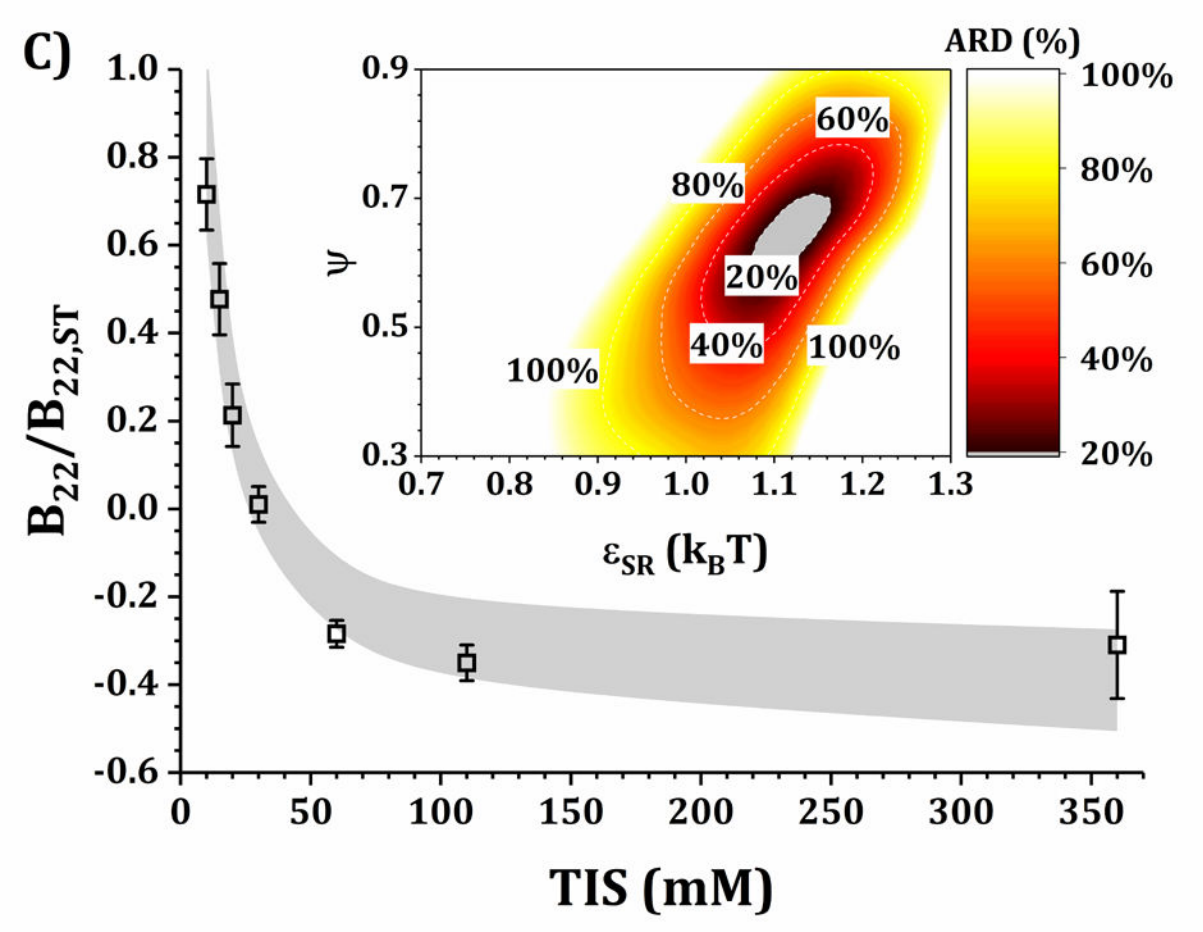


Figure 3. Main panels

$B_{22}/B_{22,ST}$ values as a function of *TIS* for the IgG1 molecule in this work at pH 5 and 10 mM acetate buffer (panel A) and pH 6.5 and 10 mM histidine buffer (panel B) with added NaCl from 0 to 500 mM. Black symbols represent data with only buffer and added NaCl while red symbols represent the same solutions with 5% w/w added sucrose. **Insets:** high-c_2 data as shown in Ref. 3 but corrected as mentioned in the main text for pH 5 (panel A) and pH 6.5 (panel B) for buffer-only (black squares), 5% w/w sucrose (red triangles) and 100 mM NaCl (gray circles). The blue dashed line corresponds to the steric-only behavior calculated using the VE EoS (equation 7).







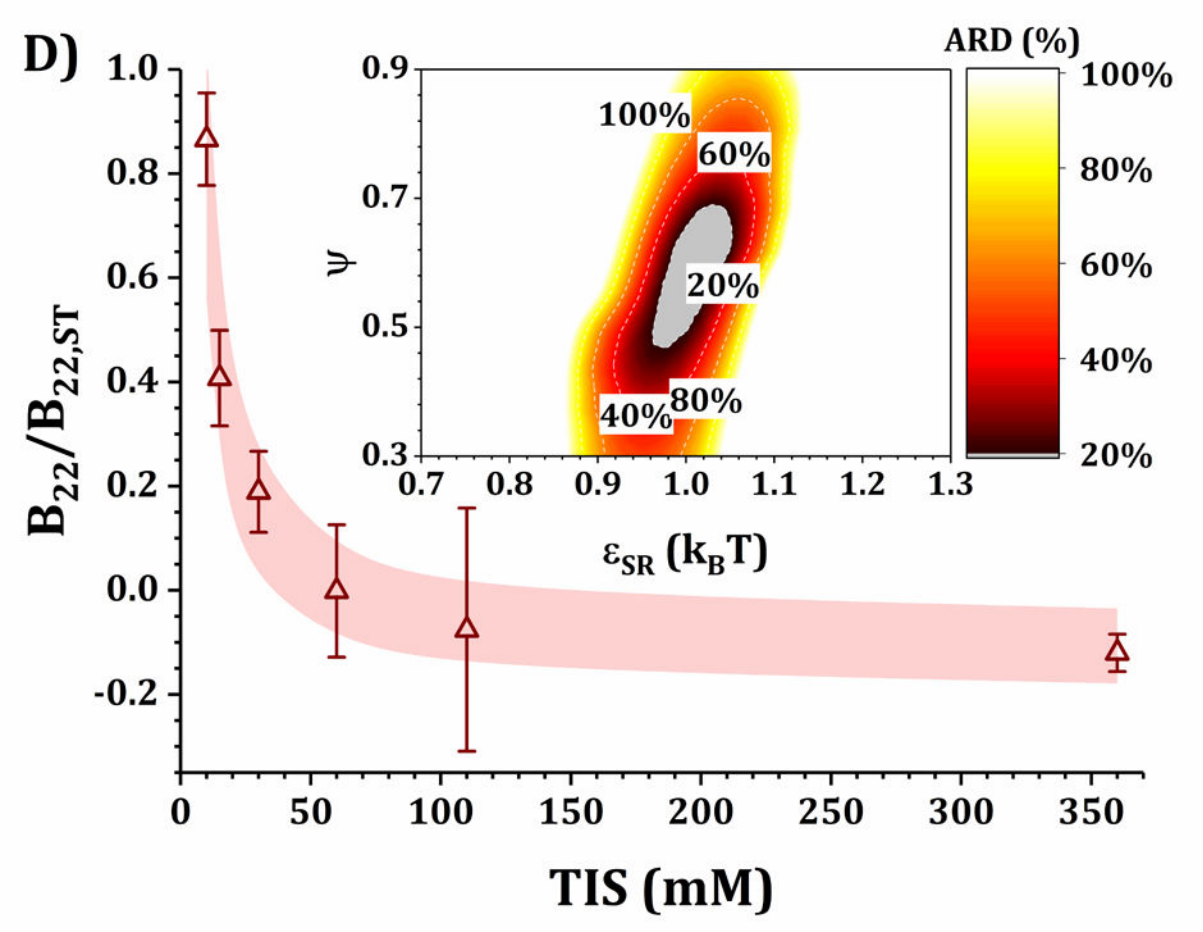
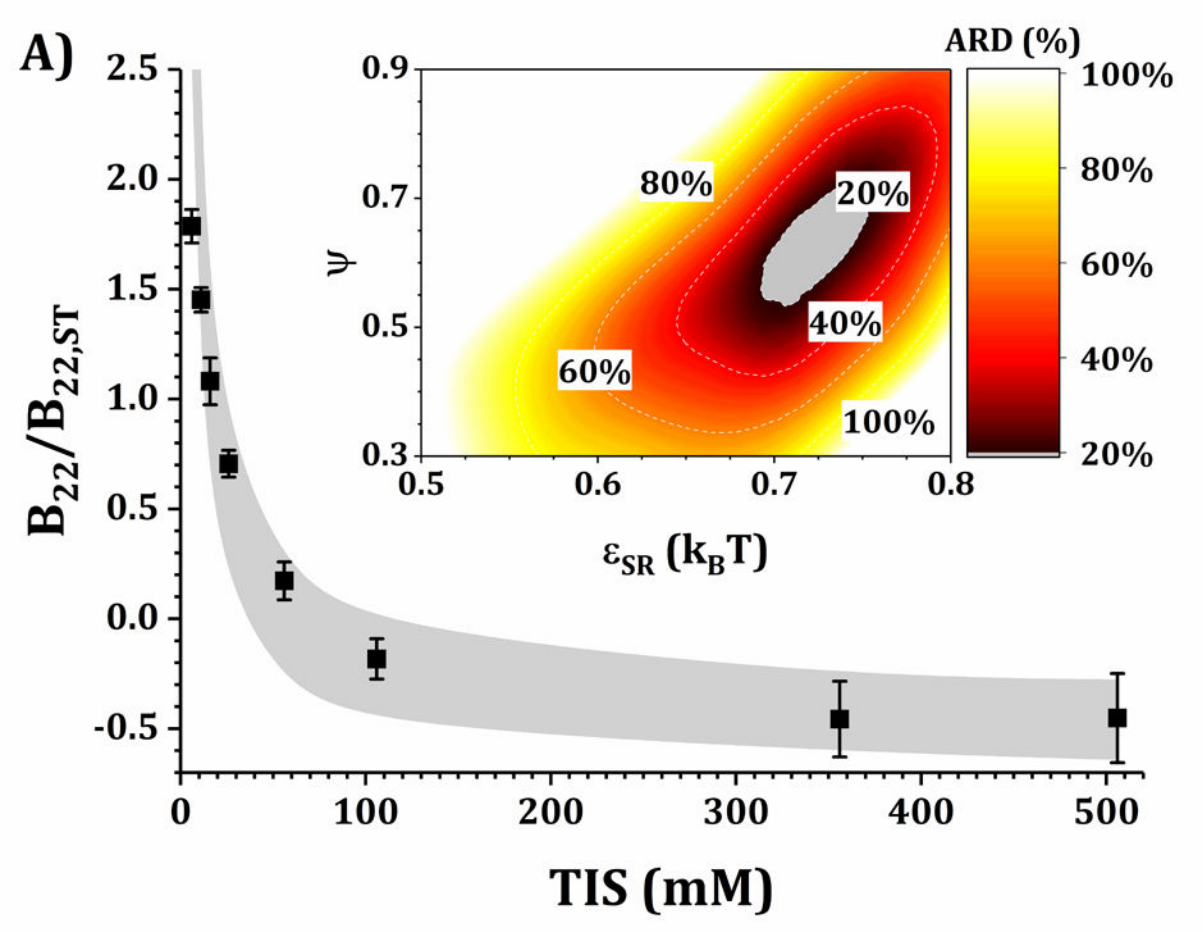
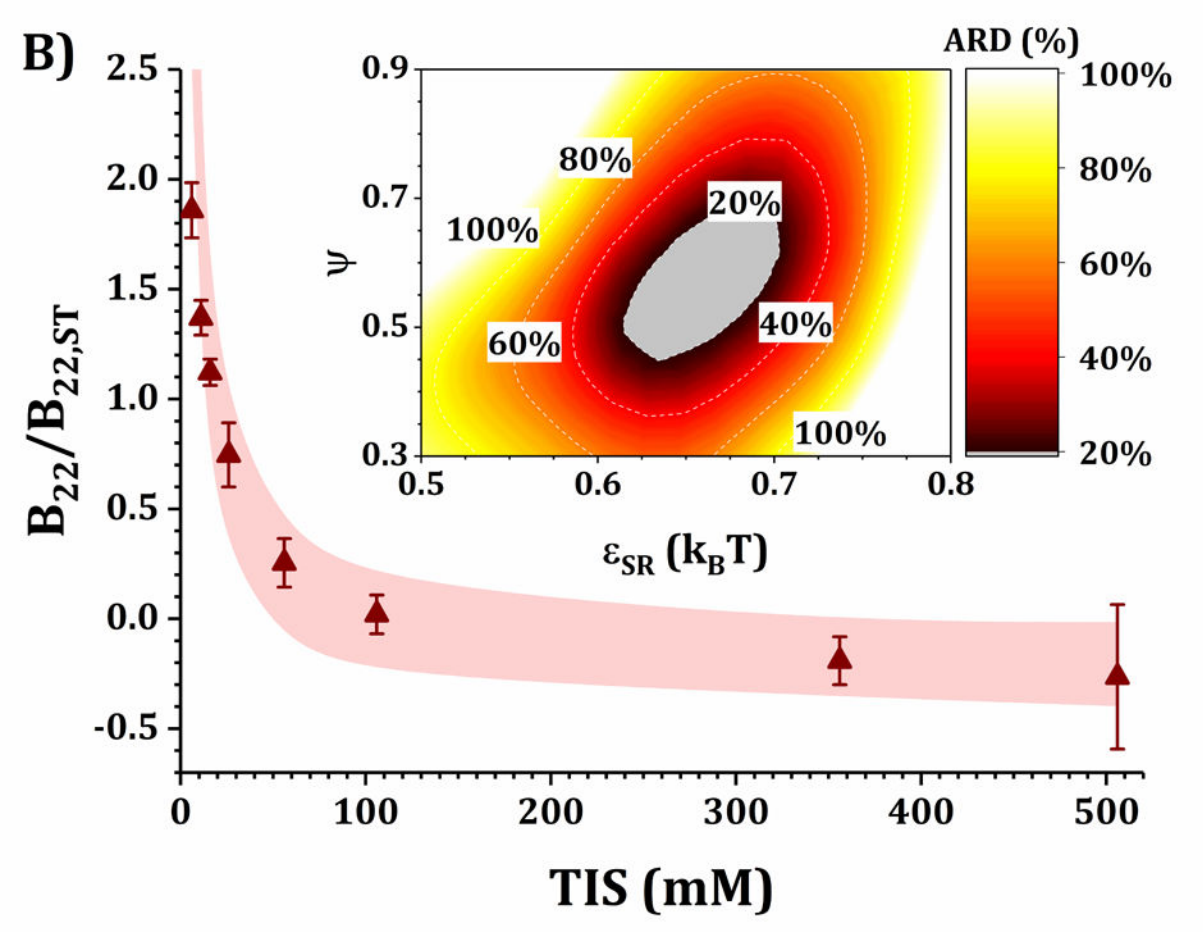
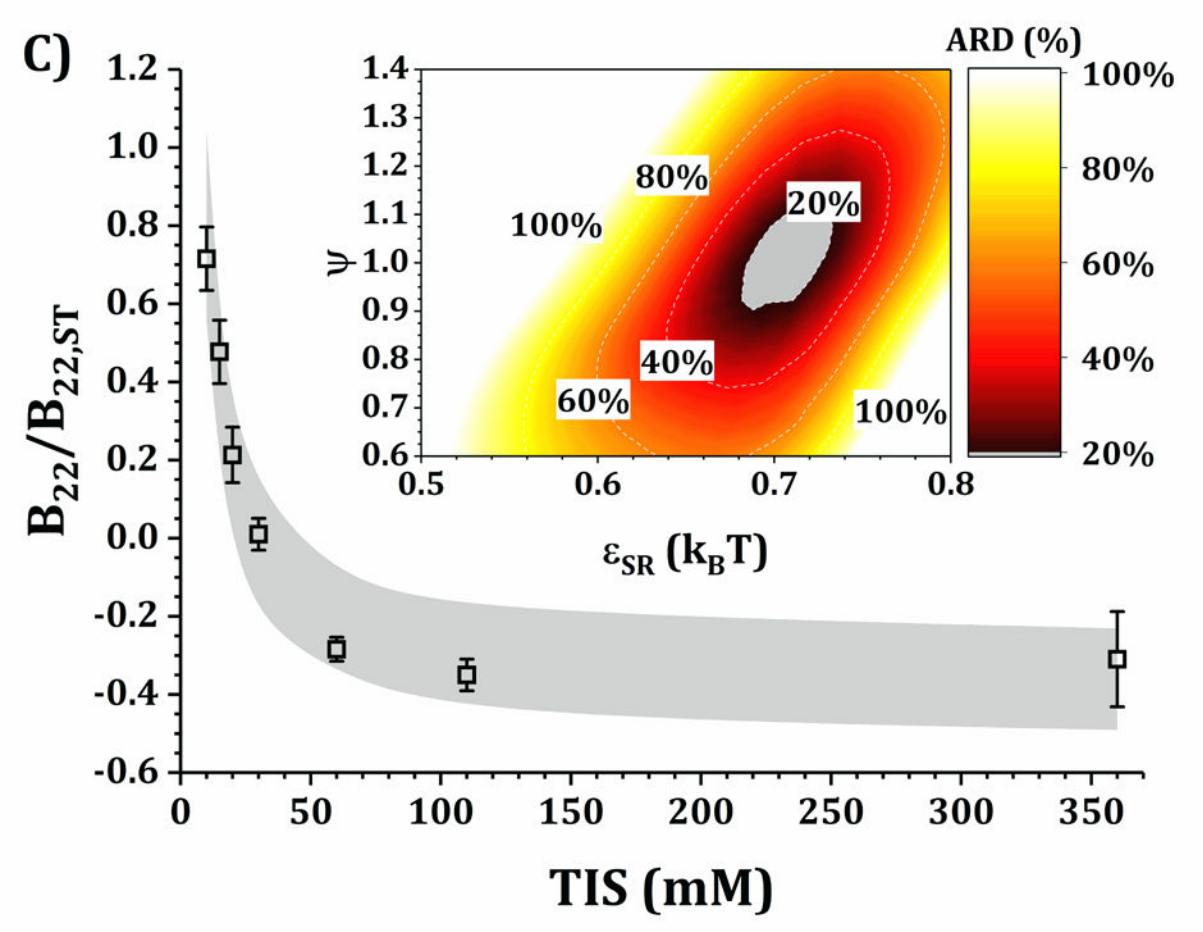


Figure 4. Main panels

comparison of $B_{22}/B_{22,ST}$ as a function of *TIS* between experimental (symbols) and simulated values (shaded areas) using the HEXA model at pH 5 for buffer-only (panel A) and 5% w/w added sucrose (panel B) and at pH 6.5 for buffer-only (panel C) and 5% w/w added sucrose (panel D). **Inset:** surface response of ARD values as a function of ε_{SR} and ψ for each respective formulation.







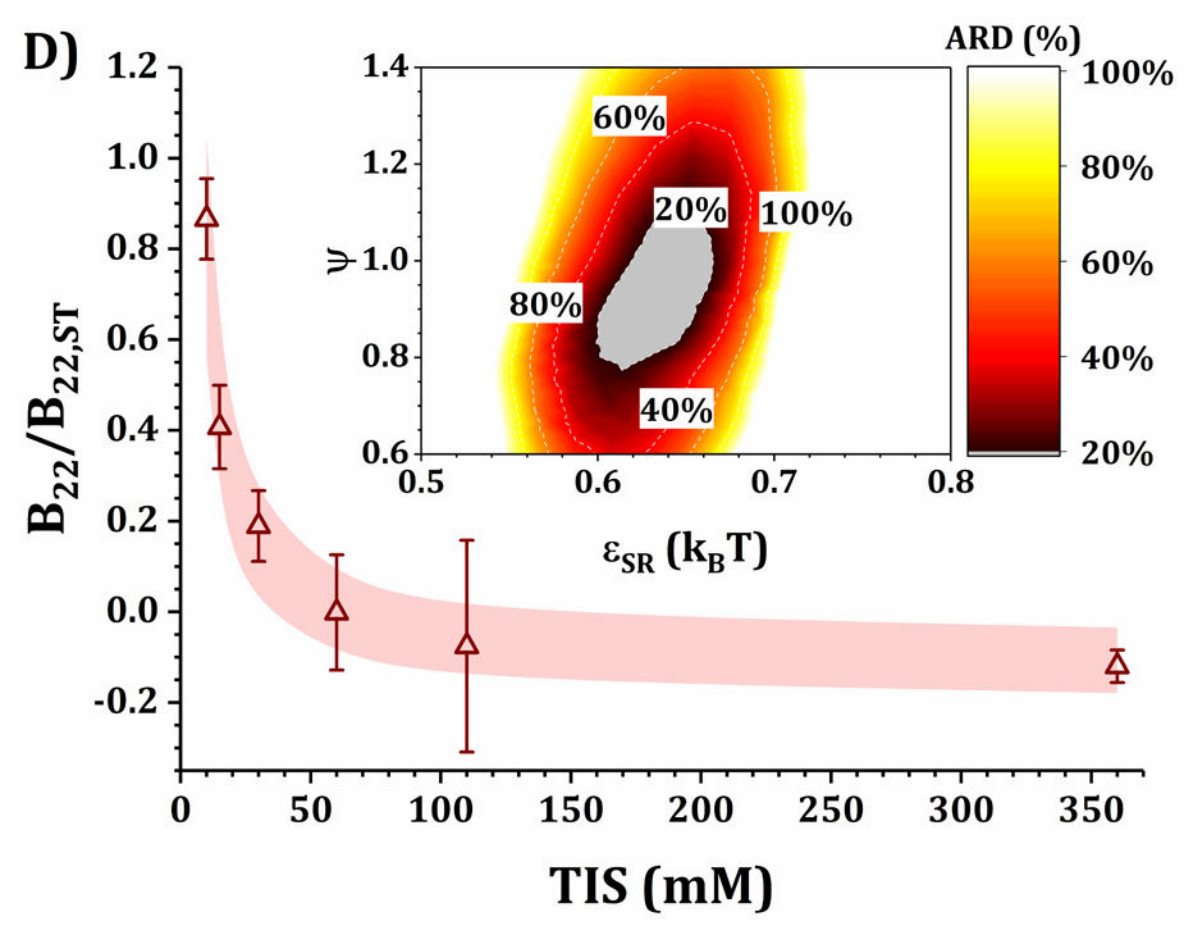
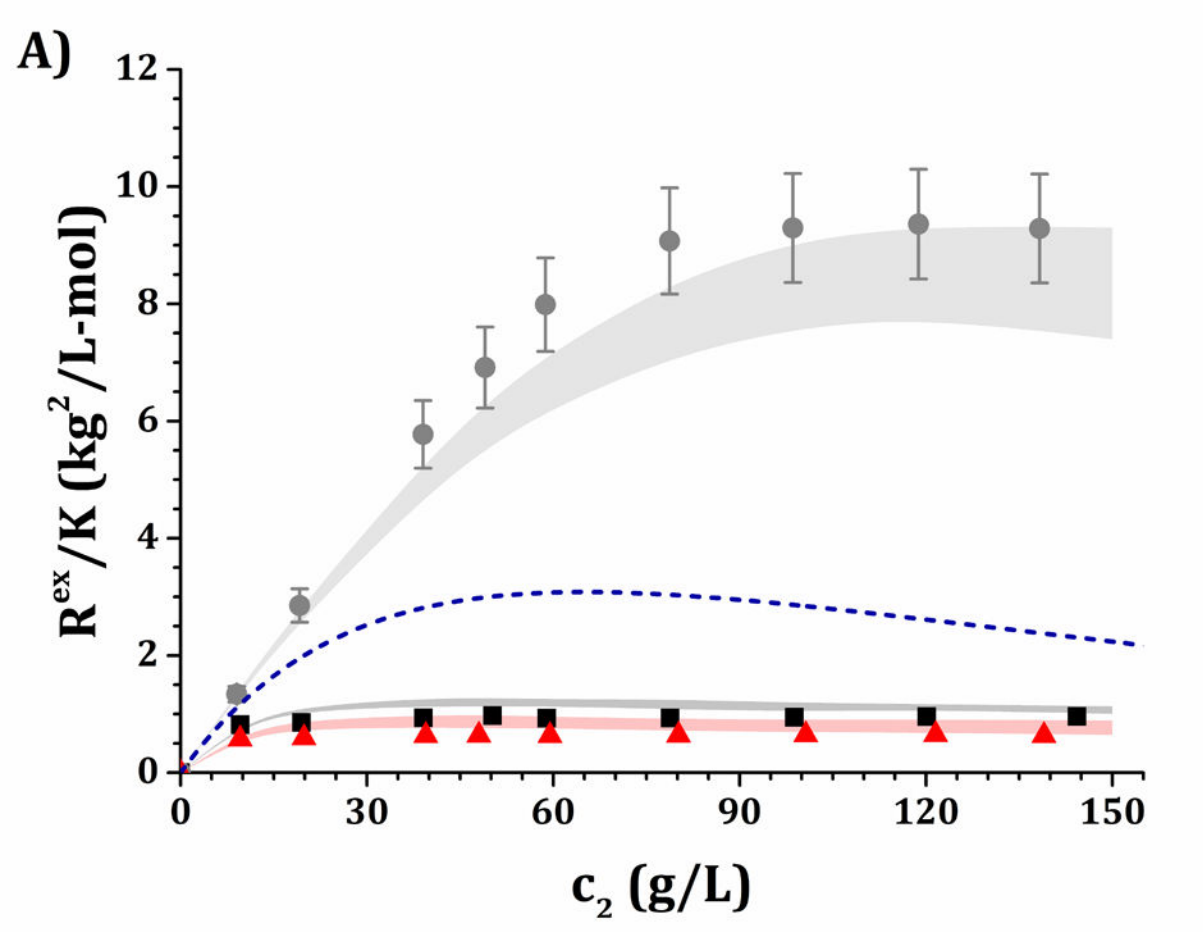
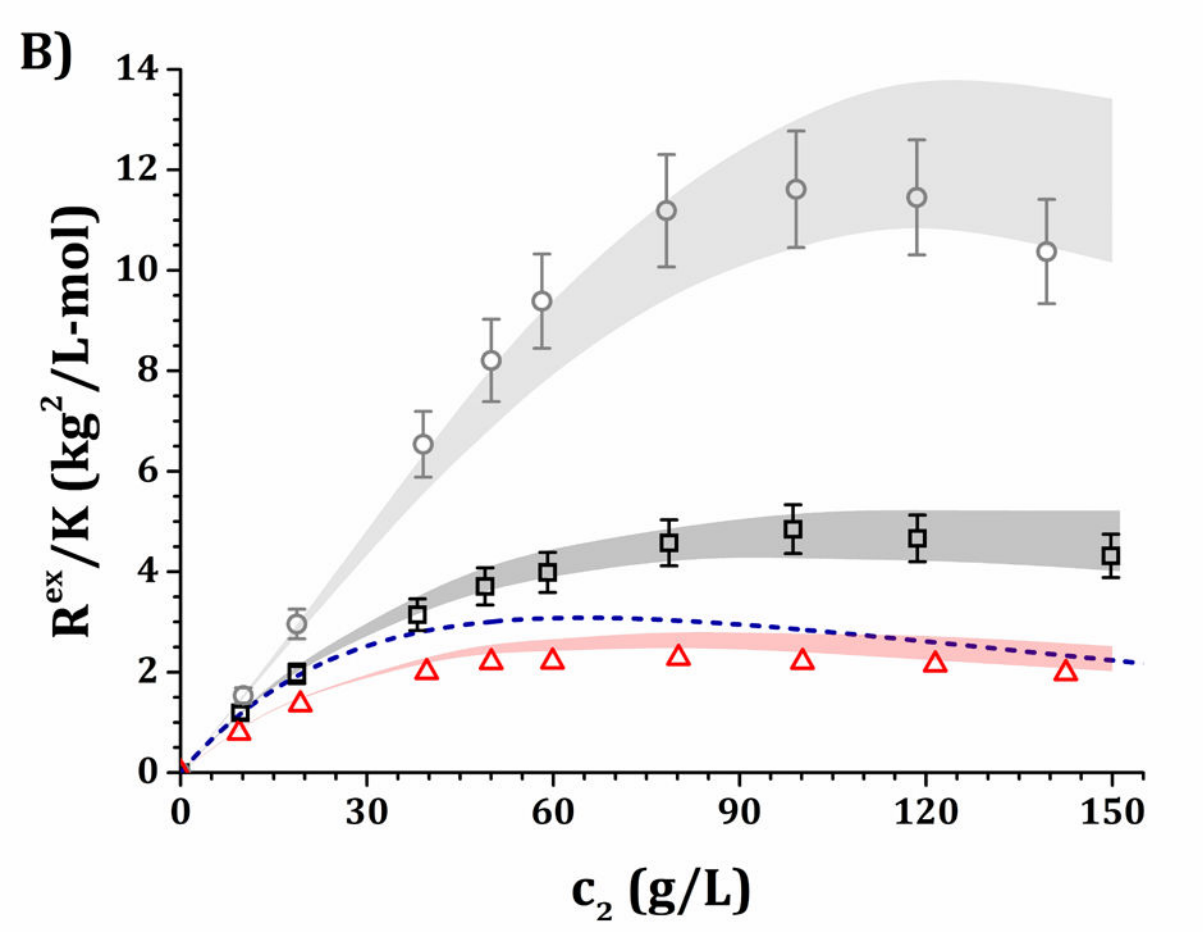
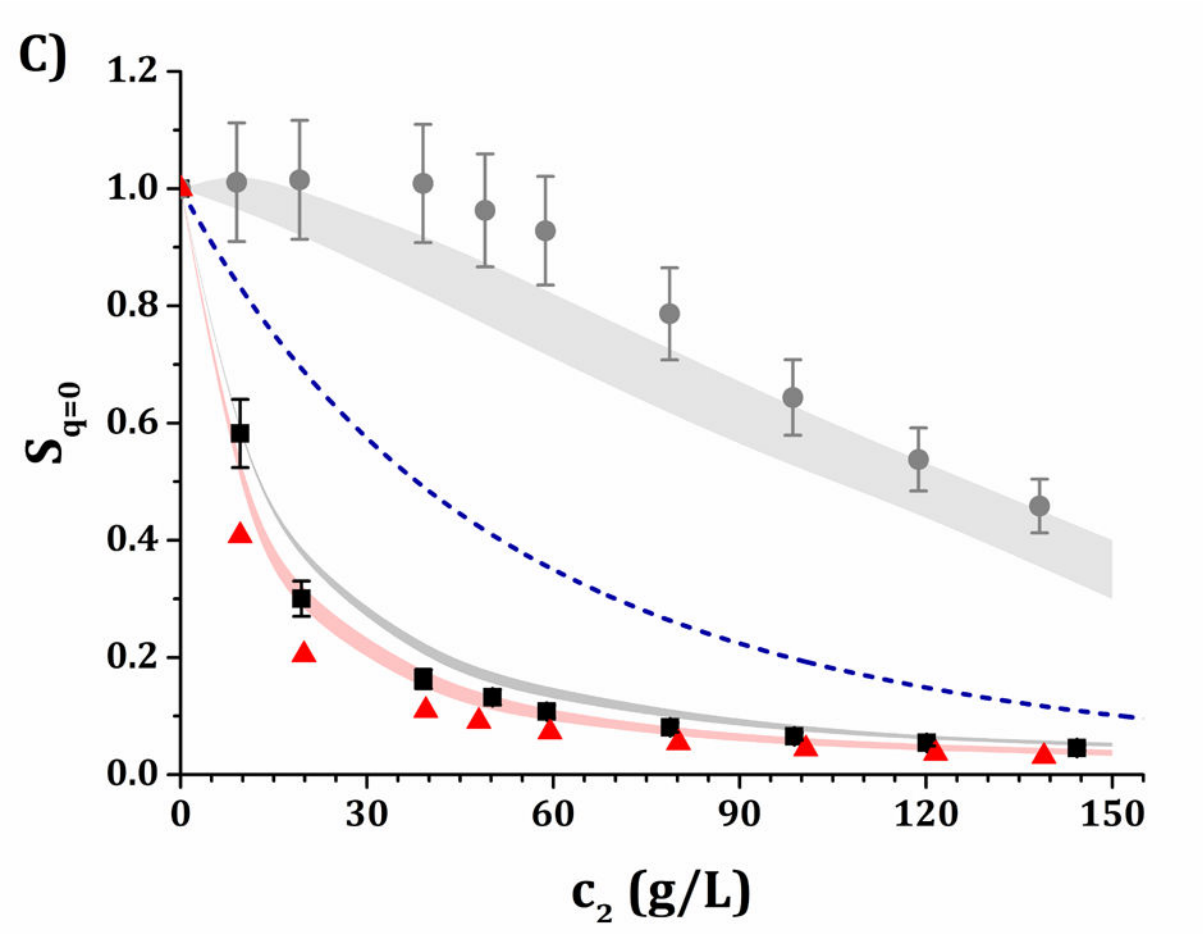


Figure 5. Main panels

comparison of $B_{22}/B_{22,ST}$ as a function of *TIS* between experimental (symbols) and simulated values (shaded areas) using the DODECA model at pH 5 for buffer-only (panel A) and 5% w/w added sucrose (panel B) and at pH 6.5 for buffer-only (panel C) and 5% w/w added sucrose (panel D). **Inset:** surface response of ARD values as a function of ε_{SR} and ψ for each respective formulation.







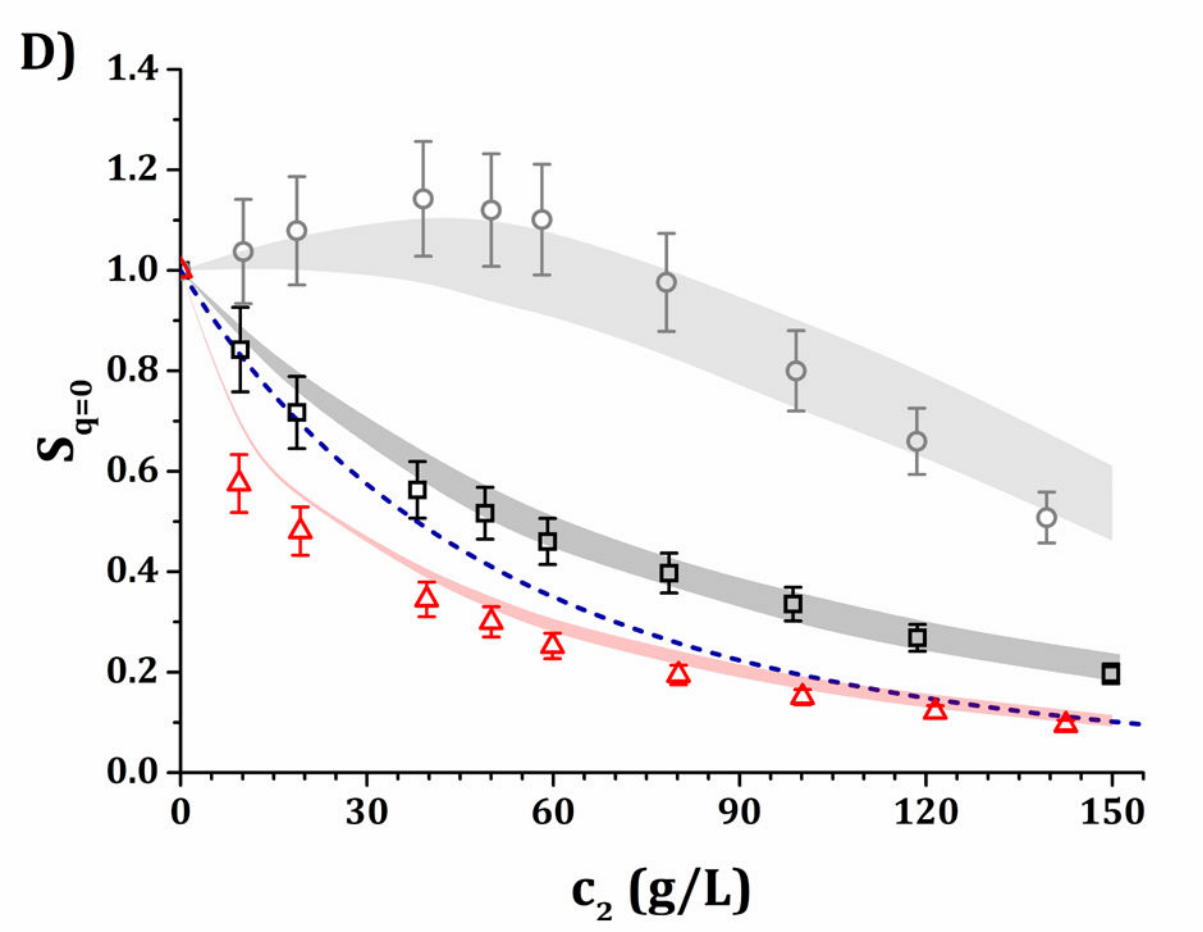
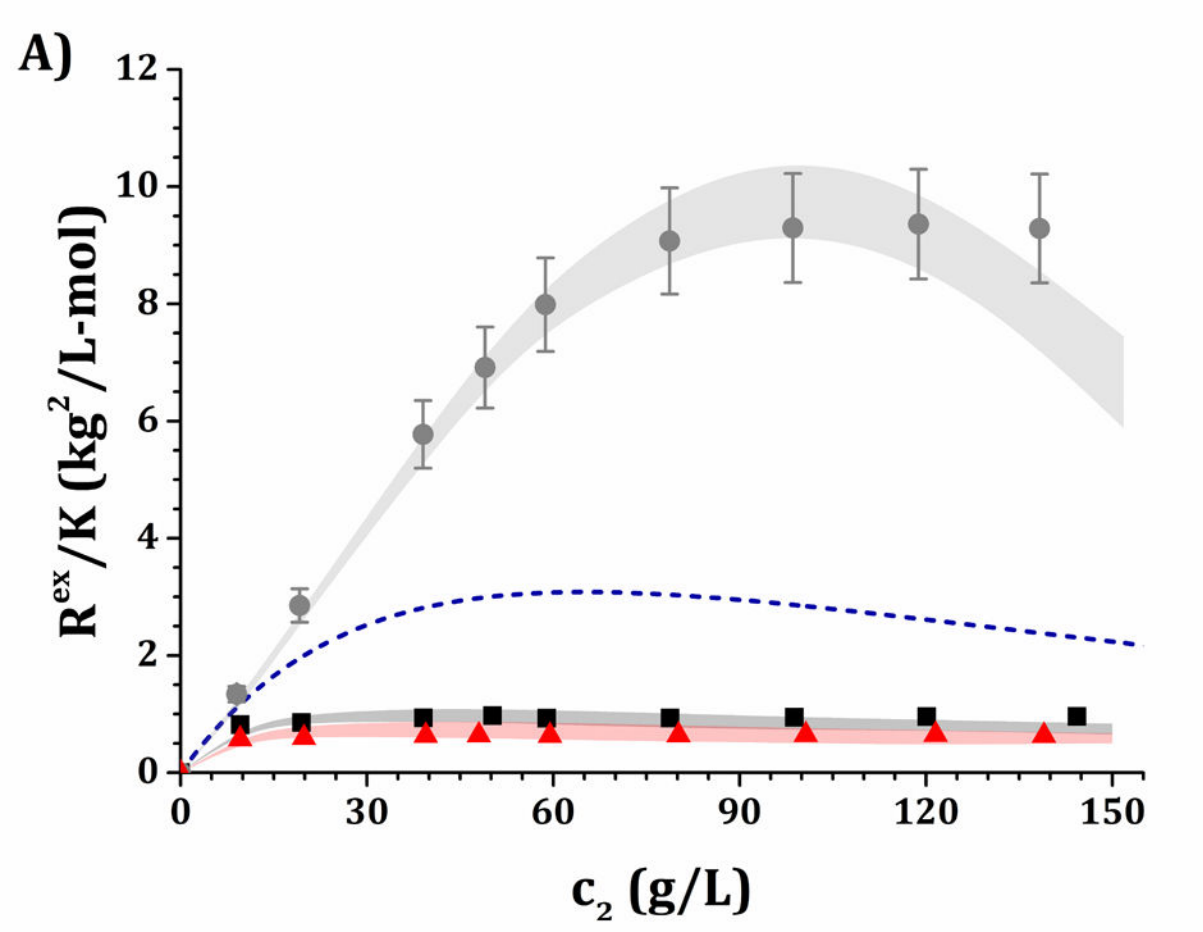
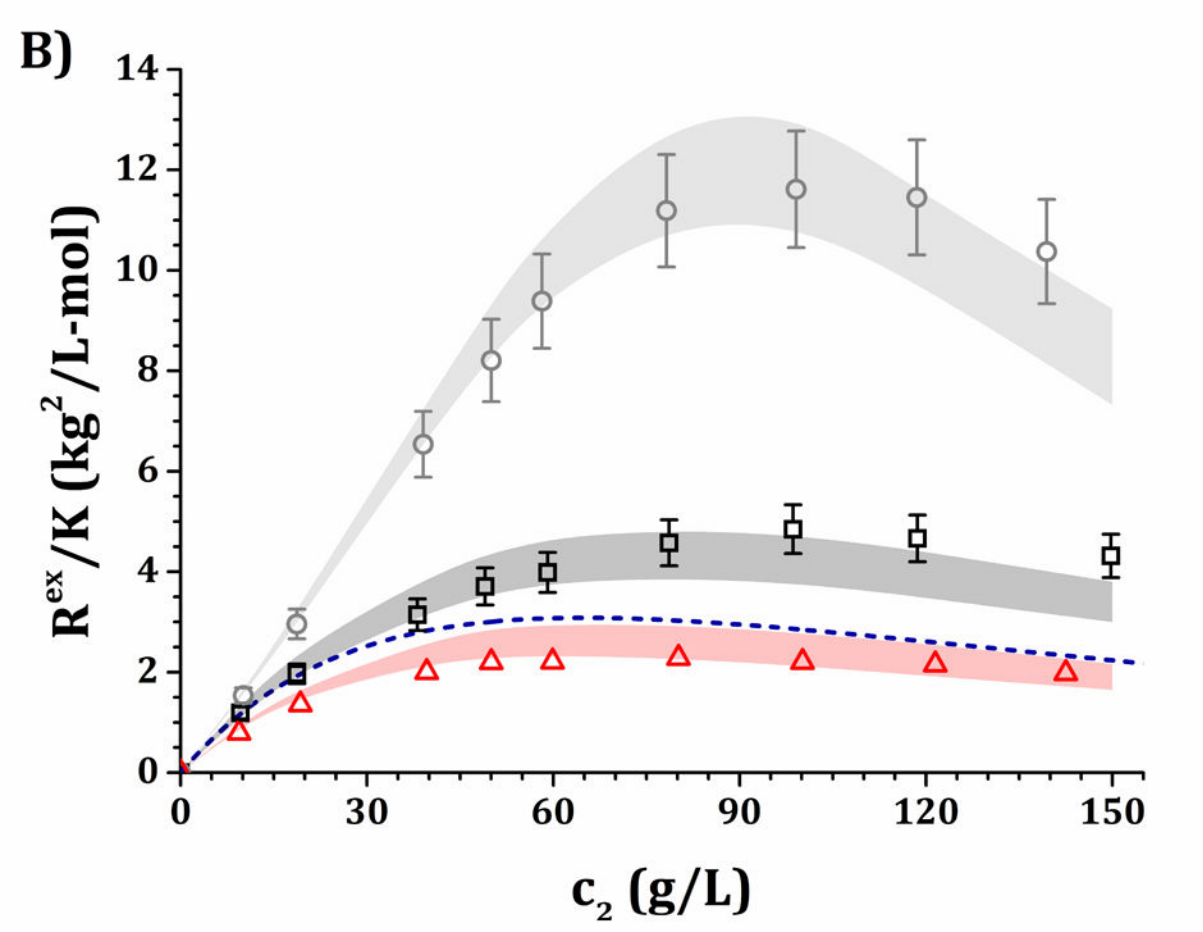
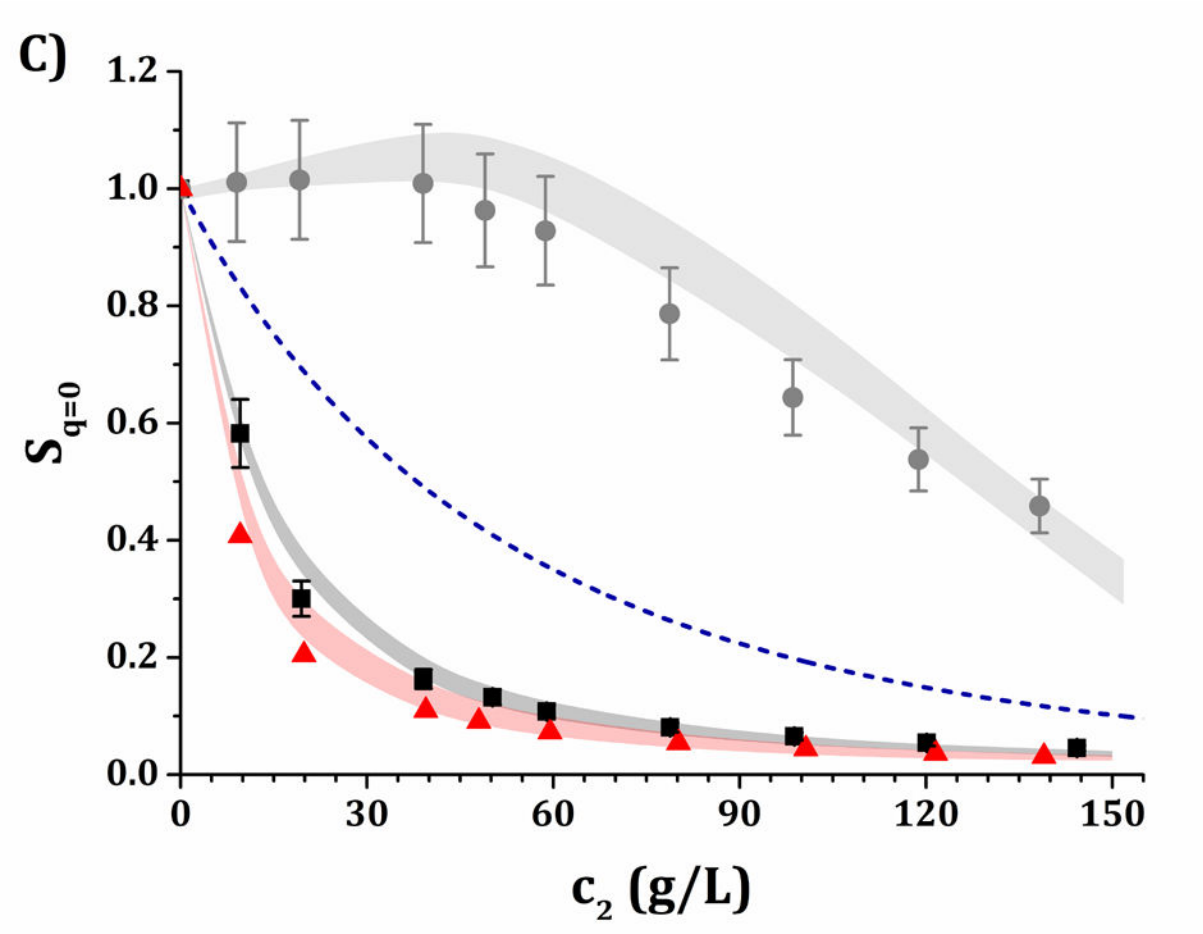


Figure 6.

High-c_2 predictions of R^{ex}/K and $S_{q=0}$ from low-c_2 parameters with the HEXA model shown in Fig. 4, for pH 5 (panels A and C) and pH 6.5 (panels B and D) and for buffer-only (black squares), 5% w/w sucrose (red triangles) and 100 mM NaCl (gray circles). The symbols represent the experimental data from insets in Fig. 3 while shaded areas represent the model predictions. The blue dashed line represents the steric-only behavior.







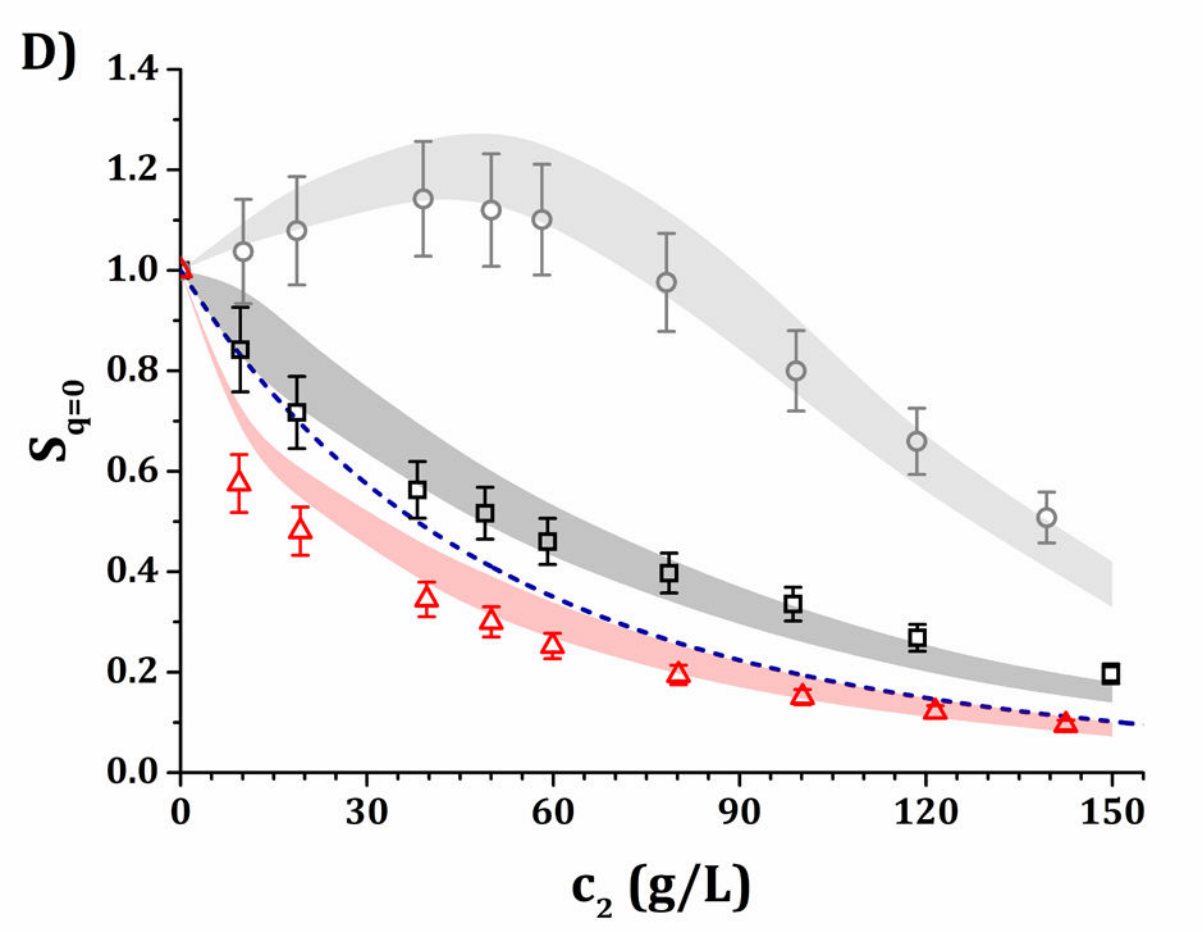
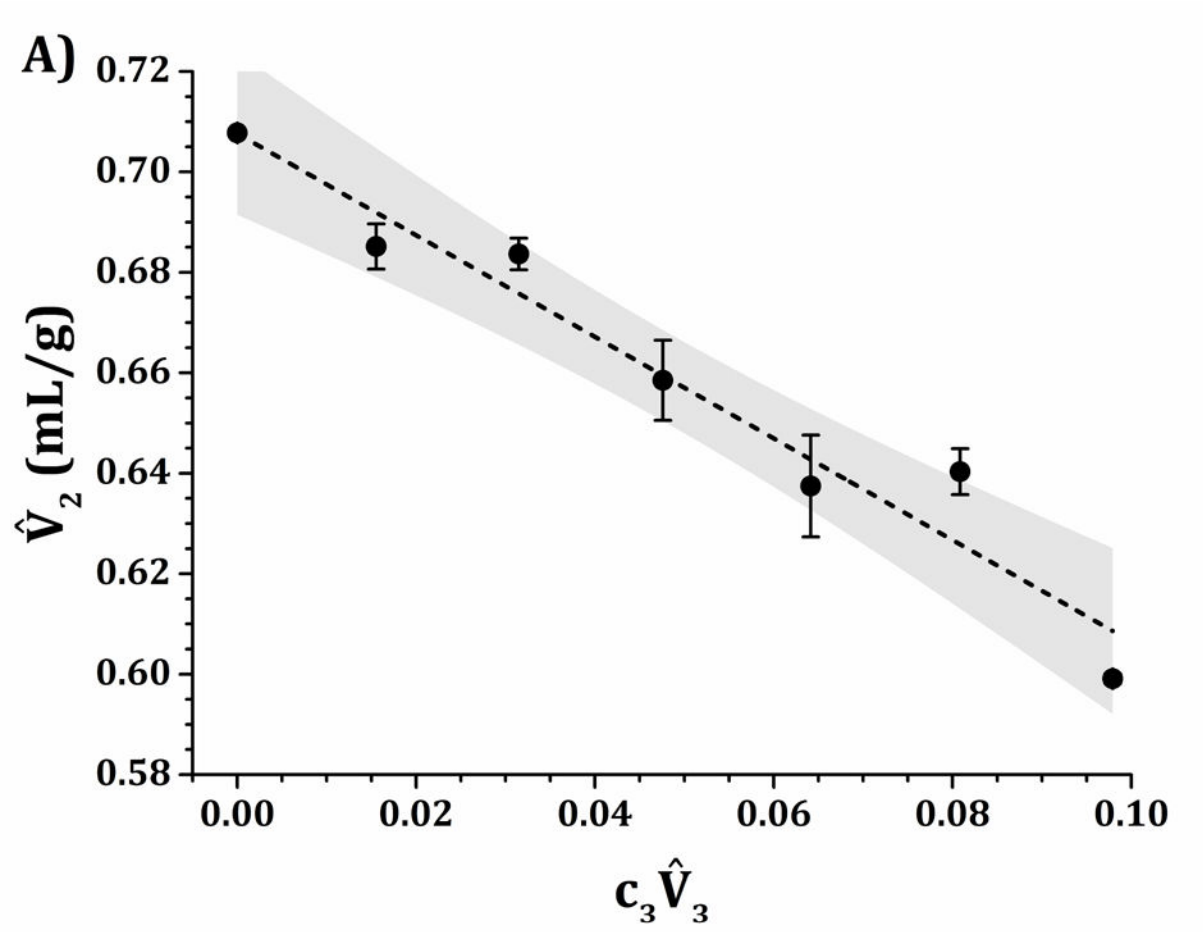


Figure 7.

High-c_2 predictions of R^{ex}/K and $S_{q=0}$ from low-c_2 parameters with the DODECA model shown in Fig. 5, for pH 5 (panels A and C) and pH 6.5 (panels B and D) and for buffer-only (black squares), 5% w/w sucrose (red triangles) and 100 mM NaCl (gray circles). The symbols represent the experimental data from insets in Fig. 3 while shaded areas represent the model predictions. The blue dashed line represents the steric-only behavior.



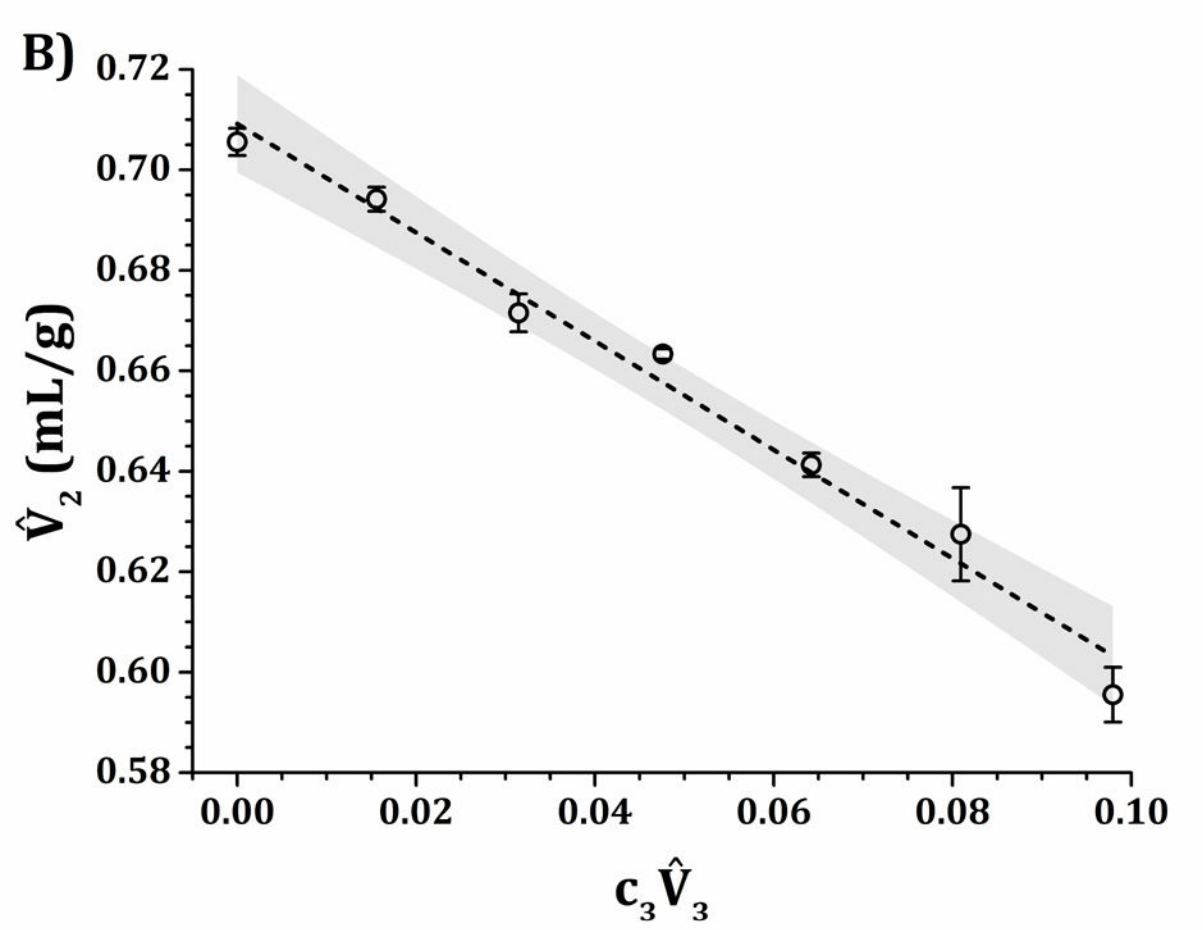


Figure 8.

IgG1 values as a function of sucrose concentrations at 25 °C and pH 5 (panel A) and pH 6.5 (panel B) for quaternary solutions of water, IgG1, sucrose and 10 mM acetate (panel A) or histidine (panel B) buffer. The dashed lines represent the linear fits to extract (G_{12}-G_{23}) from equation 3 with the assumption that $c_4\hat{V}_4$ is negligible. Shaded areas represent the 95% confidence level of each individual linear fit.

Table 1

Summary of formulations for low-c_2 data.

Formulation	Additional excipient (concentration range)	Experiment
pH 5, 10 mM acetate	NaCl (0 – 500 mM)	SLS
pH 5, 10 mM acetate + 5% w/w sucrose		SLS
pH 6.5, 10 mM histidine	NaCl (0 – 350 mM)	SLS
pH 6.5, 10 mM histidine + 5% w/w sucrose		SLS
pH 5, 10 mM acetate	Sucrose (0 – 24% w/w)	Densimetry
pH 6.5, 10 mM histidine		Densimetry

Table 2

Theoretical charges at pH 5 and 6.5.

Model	q_{VH}	q_{VL}	q_{CL}	q_{CH1}	q_{CH2}	q_{CH3}
DODECA, pH 5	1.89	0.72	2.94	6.27	5.07	2.78
DODECA, pH 6.5	0.29	0.02	0.55	4.53	1.82	−0.20
HEXA, pH 5	2.61		9.21		10.14	5.56
HEXA, pH 6.5	0.31		5.08		3.64	−0.40

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Model parameters for the steric-only EoS.

A_1	A_2	A_3	A_4	A_5
$1^{\dagger}$	10.551 ± 0.006	62.2 ± 0.2	136 ± 1	468 ± 4

†Preset to comply with dilute limit behavior (ideal gas EoS)

Table 4

Summary of formulation conditions and model parameter spaces that best predict the high-c_2 SLS data (ARD ≤ 20%) shown in Figs. 6 and 7.

Formulation	Short notation	HEXA		DODECA	
		e_{SR} (k_BT)	ψ	e_{SR} (k_BT)	ψ
pH 5, 10 mM acetate	pH 5, buffer-only	1.1–1.2	0.37–0.40	0.70–0.74	0.50
pH 5, 10 mM acetate + 5% w/w sucrose	pH 5, sucrose	0.90–1.0	0.40	0.64–0.70	0.50
pH 5, 10 mM acetate + 100 mM NaCl	pH 5, NaCl	1.15	0.30–0.40	0.70–0.74	0.45–0.66
pH 6.5, 10 mM histidine	pH 6.5, buffer-only	1.1–1.2	0.56–0.68	0.70–0.74	0.90
pH 6.5, 10 mM histidine + 5% w/w sucrose	pH 6.5, sucrose	0.90–1.0	0.62–0.68	0.64–0.70	0.83–1.0
pH 6.5, 10 mM histidine + 100 mM NaCl	pH 6.5, NaCl	1.15	0.56–0.68	0.70–0.74	0.90–1.0