

**SELF-INTERACTIONS AND AGGREGATION
OF THERAPEUTIC PROTEINS**

by

James K. Forder

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

Winter 2024

© 2024 James K. Forder
All Rights Reserved

**SELF-INTERACTIONS AND AGGREGATION
OF THERAPEUTIC PROTEINS**

by

James K. Forder

Approved: _____
Millicent O. Sullivan, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Levi T. Thompson, Ph.D.
Dean of the College of Engineering

Approved: _____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Christopher J. Roberts, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Arthi Jayaraman, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Eric M. Furst, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Erinc Sahin, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

First and foremost, I am deeply appreciative to my significant other Naorin for her unwavering love and encouragement, which helped me stay grounded in times when I was consumed by the ups and downs of research. I was fortunate to have someone who could provide moral and technical support as well as unbiased opinions when I needed an outside perspective. I want to acknowledge other friends and family as well for their enduring support.

I am grateful to my advisor Chris Roberts for his thoughtful guidance, mentorship, and for the fruitful opportunities that he provided. I would also like to extend my appreciation to Erinc Sahin who has been an invaluable collaborator and mentor. Thanks also to the two other members of my thesis committee, Eric Furst and Arthi Jayaraman, for their helpful feedback. My sincere gratitude to the many current and former members of the Roberts group who have supported me over the years when I had experimental or computational issues, or when I needed an understanding friend. Special thanks to Hassan Shahfar and Veer Palakollu who I collaborated with in parts of this thesis.

The Drug Product Science and Technology department at Bristol-Myers Squibb and the Discovery Pharmaceutical Sciences department at Merck & Co., Inc. are gratefully acknowledged for providing protein materials and financial support. The Biomolecular Interaction Technologies Center (BITC) is also acknowledged for financial support.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xii
ABSTRACT	xxix

Chapter

1	INTRODUCTION	1
1.1	Motivation	1
1.2	Protein-protein self-interactions	3
1.2.1	Coarse-grained molecular simulations	6
1.3	Physical instability.....	9
1.3.1	Conformational stability	10
1.3.2	Reversible self-association	11
1.3.3	Irreversible aggregation.....	12
1.4	Objectives	16
1.5	Applications in industrial drug development	17
1.5.1	Candidate selection.....	17
1.5.2	Formulation development.....	19
1.6	Therapeutic proteins of interest	20
1.6.1	Monoclonal antibodies	21
1.6.2	Fc-fusion proteins.....	22
1.7	Dissertation outline.....	24
2	COARSE-GRAINED MODELS FOR MOLECULAR SIMULATIONS OF MONOCLONAL ANTIBODIES AND THERAPEUTIC PROTEINS.....	28
2.1	Introduction	28
2.2	Methods	32

2.2.1	Monte Carlo simulations	32
2.2.2	CG models and interaction potentials.....	35
2.2.3	Average relative deviation error calculations	40
2.2.4	Adjustable model parameters	41
2.3	Results and discussion.....	41
2.3.1	Higher order virial coefficients and higher protein concentrations.....	53
2.3.2	Comparison of computational burden	57
2.3.3	Considerations in selecting among different CG models	59
2.4	Summary and conclusions.....	60
3	ELECTROSTATICALLY MEDIATED ATTRACTIVE SELF- INTERACTIONS AND REVERSIBLE SELF-ASSOCIATION OF FC- FUSION PROTEINS.....	62
3.1	Introduction	62
3.2	Materials and methods.....	64
3.2.1	Sample preparation	64
3.2.2	Static light scattering	65
3.2.3	Dynamic light scattering.....	66
3.2.4	1bAA CG molecular simulations and parameter optimization ...	67
3.2.5	Mayer-weighted average pairwise electrostatic energies	69
3.3	Results and discussion.....	69
3.3.1	SLS and DLS measurements of self-interactions and reversible self-association	69
3.3.2	CG molecular simulations to model self-interactions	81
3.3.3	Identification of specific electrostatic interactions involved in attractive self-interactions	86
3.4	Summary and conclusions.....	92
4	SIMULATION OF HIGH-CONCENTRATION SELF-INTERACTIONS FOR MONOCLONAL ANTIBODIES FROM WELL-BEHAVED TO POORLY-BEHAVED SYSTEMS.....	94
4.1	Introduction	94
4.2	Materials and methods.....	97

4.2.1	Sample preparation	97
4.2.2	Static and dynamic light scattering.....	99
4.2.3	CG models, interaction potentials, B_{22} simulations and CG model parameterization	100
4.2.4	Simulations at high protein concentrations	102
4.2.5	Explicit charge equilibria algorithm.....	104
4.2.6	Mayer-averaged electrostatic energies	105
4.3	Results and discussion.....	106
4.3.1	Experimental SLS and DLS measurements	106
4.3.2	High-concentration predictions of self-interactions from CG simulations.....	118
4.3.3	Methods to improve representation of charge equilibria.....	127
4.3.4	Context-driven CG model and charge equilibria method selection.....	136
4.3.5	Ranking average electrostatic energies of charged sites for strong electrostatic attractions	137
4.4	Summary and conclusions.....	140
5	DECONVOLUTING THE BIOPHYSICAL PHENOMENA THAT INFLUENCE LONG-TERM AGGREGATION RATES OF HIGH- CONCENTRATION MONOCLONAL ANTIBODY FORMULATIONS ..	142
5.1	Introduction	142
5.2	Materials and methods.....	145
5.2.1	Backgrounded membrane imaging.....	145
5.2.2	Differential scanning calorimetry.....	147
5.2.3	Size-exclusion chromatography	148
5.2.4	Isothermal stability studies	149
5.2.5	Aggregation rate calculations	150
5.2.6	Elastic net regression models	151
5.3	Results and discussion.....	153
5.3.1	Stability study design	153
5.3.2	Results from stability studies.....	155
5.3.3	Impacts of conformational stability and self-interactions on aggregation rates.....	170
5.3.4	ML models for feature selection and prediction of high- concentration aggregation rates.....	189
5.3.5	Subvisible particle formation	200

5.3.6	Outlook and industrial applications.....	206
5.4	Summary and conclusions.....	209
6	CONCLUSIONS AND FUTURE WORK.....	212
6.1	Conclusions	212
6.2	Future work	218
6.2.1	Characterization of oligomers and aggregates	218
6.2.2	Advanced algorithms for high-concentration molecular simulations.....	221
6.2.3	Experimental and simulation methods to screen excipients.....	225
	REFERENCES.....	227
Appendix		
A	SUPPORTING INFORMATION FOR CHAPTER 2	254
B	SUPPORTING INFORMATION FOR CHAPTER 3	267
B.1	Preliminary computational charge swap mutations.....	272
C	SUPPORTING INFORMATION FOR CHAPTER 4	275
D	SUPPORTING INFORMATION FOR CHAPTER 5	281
D.1	Feature screening.....	283
D.1.1	Net valence at a given pH.....	286
D.1.2	Net self-interactions.....	287
D.1.3	Conformational stability	288
D.1.4	Features derived from other stability studies (accelerated).....	289
D.1.5	Features derived from other stability studies (30 °C).....	290
D.1.6	Features derived from other stability studies (refrigerated)	292
D.1.7	Final feature set	293
E	REPRINT PERMISSION LETTERS.....	296

LIST OF TABLES

Table 2.1:	Optimized values of εSR (units of kBT) and ψ (dimensionless) for each MAb, pH and CG model and the resulting percent ARD values. ^α . 47
Table 2.2:	Order-of-magnitude relative CPU time for calculating virial coefficients for each CG model, normalized by the CPU time using the HEXA model ^α 59
Table 3.1:	Parameter values for 1bAA simulations..... 83
Table 4.1:	Summary of solution conditions for samples at high protein concentration 107
Table 4.2:	Optimized values of εsr and ψ for each MAb, pH and CG model and qualitative agreement with high- c_2 experimental data 120
Table 4.3:	Optimized values of ψ for each MAb and charge equilibria model and qualitative agreement with low- c_2 experimental data 132
Table 4.5:	Ranking of the 10 amino acids with the most negative ϕ_i values for 1bC/D MSOS simulations of MAb 2 at pH 6.5 139
Table 5.1:	Formulation conditions for stability studies in this chapter ^α 145
Table 5.2:	Subtype, pI, and incubation temperature (T_{inc}) for accelerated condition of each MAb..... 154
Table 5.3:	High molecular weight aggregate formation case by formulation ^α 168
Table 5.4:	Onset unfolding temperature and midpoint unfolding temperatures for each formulation via DSC 176
Table 5.5:	Feature coefficients and model performance for elastic net regression models of 130 mg/mL aggregation rates 190
Table 5.6:	Results from BMI measurements of subvisible particle formation ^α 201
Table 5.7:	Feature coefficients and model performance for elastic net logistic regression models of 130 mg/mL subvisible particle formation 204

Table A.1:	Valence values for HEXA beads for all conditions. The HEXA domains are F_v ($V_L + V_H$), C1 ($C_L + C_{H1}$), C2 ($C_{H2} + C_{H2}$) and C3 ($C_{H3} + C_{H3}$), where there are two domains for both F_v and C1. ^{49,50}	255
Table A.2:	Valence values for DODECA beads for all conditions. The last column denotes the total net charge of the MAb at the given pH value. ^{49,50}	256
Table A.3:	Relative hydrophobicity values (ϵ_i , dimensionless) and bead diameters (σ_i , angstroms) for the 1bAA model, where i denotes the chemical identity of the amino acid. Reproduced from Blanco et al. ⁷¹	256
Table A.4:	Residue indices for each domain in the DODECA and 1bC/D models. Indices start at the N-terminal amino acid for the heavy or light chain. ^{69,298}	257
Table A.5:	Charged site diameters (σ, i , angstroms) for charged sites in a given amino acid for the 1bC/D model, where i denotes the chemical identity of the amino acid. ¹⁸³ For data shown in the main text, ΔH was 3 Å (e.g., the final bead diameter for the charged oxygen site in Glu was 6.8 Å for $\sigma, i + \Delta H$).	258
Table A.6:	Domain bead diameters (σ_i , nm) for the HEXA and DODECA models (independent of MAb identity) and the 1bC/D model for each MAb. ^{28,50}	259
Table C.1:	Valence values for DODECA models.....	275
Table D.1	Aggregation rate calculation case by formulation ^a	281
Table D.1.1:	Initial set of input features that were considered for ENR models	284
Table D.1.1.1:	Features in the net valence at a given pH subgroup and the feature(s) that were selected for use in ENR models	286
Table D.1.2.1:	Features in the net self-interactions subgroup and the feature(s) that were selected for use in ENR models.....	287
Table D.1.3.1:	Features in the conformational stability subgroup and the feature(s) that were selected for use in ENR models	289
Table D.1.4.1:	Features in the stability studies (accelerated) subgroup and the feature(s) that were selected for use in ENR models	290

Table D.1.5.1:	Features in the stability studies (30 °C) subgroup and the feature(s) that were selected for use in statistical models	291
Table D.1.6.1:	Features in the stability studies (refrigerated) subgroup and the feature(s) that were selected for use in ENR models	292
Table D.1.7.1:	Final set of input features that were considered for statistical modeling	293

LIST OF FIGURES

Figure 1.1:	Comparison of published $B22$ and kD values for MAbs, where $B22$ and kD are normalized by $B22, ST$, the steric-only contribution to $B22$ (see section 2.2.1). ⁴²⁻⁴⁸	5
Figure 1.2:	Generalized aggregation diagram with MAbs used as an example. Gray ovals represent natively folded domains, blue ovals represent unfolded or misfolded domains, and domains in gold are linked to form an irreversible aggregate. Reversible steps are represented by \rightleftharpoons symbols and irreversible steps are represented by \rightarrow symbols. The mechanism shown here is an illustrative example.	10
Figure 1.3:	Semi-quantitative relationship between the change in net self-interactions ($G22$) and aggregation rate ($kobs$) when comparing buffer-only solutions (denoted with the superscript “ref”) and solutions with buffer and 100 mM NaCl. Results at pH 5 and pH 6.5 conditions are shown as filled and open blue squares, respectively. This figure is reproduced from Ghosh et al. ⁴³	14
Figure 1.4:	Schematic diagrams of the proteins of interest in this thesis, roughly to scale. The hinge region of the MAb and the linkers for the Fc-fusion proteins are shown for illustrative purposes.....	24
Figure 2.1:	Schematic structures of CG models for monoclonal antibodies shown roughly to scale. Domain beads for 1bC/D are deliberately shown as slightly smaller to show the location of explicit charged sites for positive (red) and negative (blue) charges. Linkers shown between the Fc and Fab for HEXA, DODECA, and 1bC/D are putative and not represented with explicit beads in those models. Portions of this figure are reproduced from Calero-Rubio et al. ²⁸	31
Figure 2.2:	Experimental $B22B22, ST$ values as a function of TIS for MAb 1 at pH 5 (indigo squares), MAb B at pH 5 (green circles), and MAb 2 at pH 6.5 (yellow triangles) to illustrate characteristic behavior for Cases (i), (ii), and (iii), respectively. ^{42,49}	42

- Figure 2.3: $B22B22, ST$ as a function of ϵSR in the limit where electrostatic interactions are not included (i.e., $\psi = 0$) for the 1bAA model for MAb 1 (solid black), 2 (dashed indigo), A (dotted green), B (dash-dotted yellow), and C (dash-double dotted purple). 46
- Figure 2.4: (A) Experimental $B22B22, ST$ values (black symbols and curve as guide to the eye) and simulated $B22B22, ST$ values using ψ values that minimized ARD for different models for MAb 1 at pH 5: HEXA (indigo curve), DODECA (green curve), 1bC/D (yellow curve) and 1bAA (purple curve); (B) ARD vs. ψ for the experimental data in panel A for HEXA (indigo curve), DODECA (green curve), 1bC/D (yellow curve) and 1bAA (purple curve), illustrating how the ψ value for a given model was selected for a given experimental data set. 48
- Figure 2.5: Comparison of experimental $B22B22, ST$ values (black symbols and line) and simulated predictions with ψ values that minimized ARD for HEXA (indigo), DODECA (green), 1bC/D (yellow) and 1bAA (purple) models for: (A) MAb B at pH 5; (B) MAb 2 at pH 6.5; (C) MAb C at pH 5. Values in panel (C) are designated as $-G222B22, ST$ instead of $B22B22, ST$ because some values are so large that they indicate multi-body interactions at the experimental protein concentrations, rather than two-body interactions that are captured by $B22$. The corresponding ARD plots are shown in Figure A.7 (Appendix A). 49
- Figure 2.6: Illustrative steric-only virial coefficient values (A_i, ST , $i = 2$ to 5) for MAb B, normalized by the value of A_i, ST for the all-atom model (for a given virial coefficient, $i=2,3,4,5$), for HEXA (indigo), DODECA (green), 1bC/D (yellow), and 1bAA (purple). Steric-only virial coefficients $i = 2$ to 5 for the DODECA model were reported in prior work.⁵⁰ The values of A_i, ST for $i = 2$ to 5 for the all-atom model were $(9.46 \pm 0.07) \times 10^{-3}$ L/g, $(5.45 \pm 0.08) \times 10^{-5}$ (L/g)², $(2.06 \pm 0.02) \times 10^{-7}$ (L/g)³, and $(6.29 \pm 0.02) \times 10^{-9}$ (L/g)⁴, respectively. Error was calculated as the 95% confidence interval from three independent simulations. The terms $B22, ST$ and $A2, ST$ are used interchangeably... 54
- Figure 2.7: Virial coefficients with respect to steric-only behavior ($A_i A_i, ST$) illustrated for MAb B at pH 5 as a function of TIS using the 1bC/D model: $B22B22, ST$ (black triangles and line), $A3A3, ST$ (indigo circles and line), $A4A4, ST$ (green squares and line) and $A5A5, ST$ (yellow diamonds and line) coefficients. The black horizontal line represents net interactions equivalent to steric-only interactions for a given virial coefficient ($i=2, 3, 4, 5$). 57

- Figure 3.1: Schematic diagrams (left) and 1bAA models (right), shown roughly to scale, compared to the 1bAA structures, of the three proteins in this chapter. 64
- Figure 3.2: Excess Rayleigh ratio values (R_{90exK}) as a function of protein concentration (c_2) for the fusion partner protein (A) and the monovalent Fc-fusion (B). Data are shown for a subset of the experimental conditions: pH 5 buffer-only (solid black squares), pH 5 + 300 mM NaCl (open black squares), pH 6.5 buffer-only (solid gray circles), and pH 6.5 + 300 mM NaCl (open gray circles). The curves are fits to Equation 3.1 used to calculate G_{22} and M_w, app . Illustrations in the corners are included to orient the reader when comparing between the different proteins (see Figure 3.1). 71
- Figure 3.3: $-G_{22}/2B_{22}, ST$ as a function of ionic strength for the fusion partner protein (A), and the monovalent Fc-fusion (B) at pH 5 (black squares) and pH 6.5 (gray circles). The horizontal dashed line at $-G_{22}/2B_{22}, ST = 1$ corresponds to a steric-only reference state. Apparent molecular weight (M_w, app) as a function of ionic strength for the fusion partner protein (C), monovalent Fc-fusion (D), following the same color scheme. The horizontal dashed line corresponds to the true molecular weight. Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1). 73
- Figure 3.4: DLS correlograms for 2 mg/mL solutions of the FP protein (A) and the monovalent Fc-fusion (B). The correlograms are normalized by the value of the autocorrelation function at the shortest delay time ($g\tau_0 - 1$) for ease of comparison between conditions. Data are shown for a subset of the experimental conditions: pH 5 buffer-only (solid black lines), pH 5 + 300 mM NaCl (dashed black lines), pH 6.5 buffer-only (solid gray lines), and pH 6.5 + 300 mM NaCl (dashed gray lines). Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1). 74
- Figure 3.5: kD as a function of ionic strength for the fusion partner protein (A), and the monovalent Fc-fusion (B) at pH 5 (black squares) and pH 6.5 (gray circles). Hydrodynamic radius (rH) as a function of ionic strength for the fusion partner protein (C), monovalent Fc-fusion (D), following the same color scheme. Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1). 76

- Figure 3.6: SLS and DLS measurements for the bivalent Fc-fusion at pH 5 buffer-only (solid black squares), pH 5 + 300 mM NaCl (open black squares), pH 6.5 buffer-only (solid gray circles), and pH 6.5 + 300 mM NaCl (open gray circles). A) Excess Rayleigh ratio values (R_{90exK}) as a function of protein concentration (c_2). The curves are fits to Equation 3.1 used to calculate Mw, app values. B) Apparent molecular weight, Mw, app , as a function of ionic strength, where the horizontal dashed line corresponds to the true molecular weight. C) Collective diffusion coefficient values, DC , as a function of protein concentration (c_2). The lines are linear fits, and D_0 (the y-intercept of a given linear fit) was used to calculate hydrodynamic radius, rH , shown in panel D. Illustrations in the corners are included to orient the reader when comparing between the different proteins (see Figure 3.1)..... 79
- Figure 3.7: DLS correlograms for 2 mg/mL solutions of the bivalent Fc-fusion. The correlograms are normalized by the value of the autocorrelation function at the shortest delay time ($g\tau_0 - 1$) for ease of comparison between conditions. Data are shown for a subset of the experimental conditions: pH 5 buffer-only (solid black lines), pH 5 + 300 mM NaCl (dashed black lines), pH 6.5 buffer-only (solid gray lines), and pH 6.5 + 300 mM NaCl (dashed gray lines). Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1)..... 81
- Figure 3.8: $B_{22}/B_{22}, ST$ as a function of ionic strength calculated by 1bAA simulations for the FP protein (A) and the monovalent Fc-fusion (B). Experimental measurements are reproduced from Figure 3.3 where data at pH 5 is shown as black squares and at pH 6.5 as gray circles. The shaded regions correspond to predictions from the 1bAA simulations at pH 5 (indigo) and pH 6.5 (green) based on uncertainty in the ψ parameter. 1bAA diagrams in the corners are included to orient the reader when comparing between the different proteins (see Figure 1). 84
- Figure 3.9: Simulated contour plots of $B_{22}/B_{22}, ST$ as a function of ψ and ionic strength for the bivalent Fc-fusion at pH 5 (A) and pH 6.5 (B). 86

- Figure 3.10: Mayer-weighted average pairwise electrostatic energy values (ϕ_{ij}) for charged sites in the FP protein (A), the monovalent Fc-fusion (B), and the bivalent Fc-fusion (C) at pH 6.5 and 10 mM ionic strength. Pairwise repulsions (positive ϕ_{ij} values) are shown in blue, and attractions (negative ϕ_{ij} values) are shown in red. The heatmaps for the Fc-fusion proteins are ordered based on the structural domains. 87
- Figure 3.11: Mayer-weighted average pairwise electrostatic energy values (ϕ_{ij}) for charged sites in the FP protein (A), the monovalent Fc-fusion (B), and the bivalent Fc-fusion (C) at pH 5 and 10 mM ionic strength. Pairwise repulsions (positive ϕ_{ij} values) are shown in blue, and attractions (negative ϕ_{ij} values) are shown in red. The heatmaps for the Fc-fusion proteins are ordered based on the structural domains. 88
- Figure 3.12: Mayer-weighted average pairwise electrostatic energies (ϕ_{ij}) for charged sites in the monovalent Fc-fusion (A) and bivalent Fc-fusion (B) at pH 6.5 and 30 mM ionic strength. Repulsions are shown in blue, and attractions are shown in red. The heatmaps are ordered based on the structural domains. 89
- Figure 4.1: Schematic diagrams of the DODECA and 1bC/D models shown to scale. Linkers shown between the Fc and Fab are for illustrative purposes only. Charged sites in the 1bC/D model are blue (red) for positively (negatively) charged atoms. The DODECA model is reproduced from Calero-Rubio et al.²⁸ 96
- Figure 4.2: Solution density as a function of MAb concentration (c_2) for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D) in the following buffer solutions: pH 5 low IS (filled purple circles), pH 5 high IS (open purple circles), pH 6.5 low IS (filled green triangles) and pH 6.5 high IS (open green triangles). 99
- Figure 4.3: $B_{22}/B_{22}, ST$ as a function of ionic strength for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) at pH 5 (black squares) and pH 6.5 (gray circles), with lines to guide the eye.^{49,50} The y-axis scale for MAb 2 (panel b) is extended to fit all of the data. 108
- Figure 4.4: kD as a function of ionic strength for MAb 3 (a), MAb 4 (b), MAb 5 (c), and MAb 6 (d) at pH 5 (black squares) and pH 6.5 (gray circles), with lines to guide the eye. 111

- Figure 4.5: Mw, app as a function of ionic strength for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) at pH 5 (black squares) and pH 6.5 (gray circles).^{49,50} Mw, app values at other ionic strength conditions for MAb 1 were not reported in prior work. 113
- Figure 4.6: Polydispersity index (PDI) as a function of c_2 for MAb 3 (a), MAb 4 (b), MAb 5 (c), and MAb 6 (d) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles). 114
- Figure 4.7: $Sq = 0$ as a function of c_2 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles).^{49,50} The purple dot-dashed line is a steric repulsion equation of state.⁵⁰ The y-axis scale for MAb 2 (panel b) is extended to fit all of the data. 116
- Figure 4.8: DC as a function of c_2 for MAb 3 (a), MAb 4 (b), MAb 5 (c), and MAb 6 (d) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles). Data for MAb 3 at pH 5 at low ionic strength above 75 mg/mL are omitted due to polydispersity values significantly above 0.1. 117
- Figure 4.9: $B_{22}/B_{22}, ST$ as a function of ϵsr where no electrostatic interactions are present (i.e., $\psi = 0$) for the DODECA (a) and 1bC/D (b) models. The DODECA model is not MAb-specific, so one curve is shown in black. The results for the 1bC/D model are specific to the MAb structure and are shown in indigo for MAb 1, green for MAb 2, purple for MAb 3, yellow for MAb 4, light blue for MAb 5, and gray for MAb 6. Points in both figures correspond to the chosen ϵsr value for each MAb and match the previously listed colors. 118

- Figure 4.10: MSOS simulations of $B22/B22, ST$ as a function of ionic strength at pH 5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 (black squares), DODECA results are dashed green lines, and 1bC/D results are solid indigo lines. Shaded regions correspond to uncertainty in the ψ parameter from Table 4.2. 122
- Figure 4.11: High- c_2 predictions of $Sq = 0$ as a function of MAb concentration (c_2) at pH 5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 as filled (open) black squares for low (intermediate) ionic strength conditions. Predictions with the optimal ψ value by the 1bC/D model are filled (open) indigo triangles for low (intermediate) ionic strength. Predictions with the optimal ψ value by the DODECA model are solid (dashed) green lines for low (intermediate) ionic strength. Shaded regions for DODECA and error bars for 1bC/D correspond to uncertainty of the ψ parameter (see Table 4.2). The purple dash-dotted line is a steric-only equation of state.⁵⁰ 123
- Figure 4.12: MSOS simulations of $B22/B22, ST$ as a function of ionic strength at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 (gray circles), DODECA results are dashed green lines, and 1bC/D results are solid indigo lines. Shaded regions correspond to uncertainty in the ψ parameter from Table 4.2. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data..... 125
- Figure 4.13: High- c_2 predictions of $Sq = 0$ as a function of MAb concentration (c_2) at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 as filled (open) gray circles for low (intermediate) ionic strength conditions. Predictions with the optimal ψ value by the 1bC/D model are filled (open) indigo triangles for low (intermediate) ionic strength. Predictions with the optimal ψ value by the DODECA model are solid (dashed) green lines for low (intermediate) ionic strength. Shaded regions for DODECA and error bars for 1bC/D correspond to uncertainty of the ψ parameter (see Table 4.2). The purple dash-dotted line is a steric-only equation of state.⁵⁰ The y-axis scale for MAb 2 (panel b) is extended to fit all of the data. 126

- Figure 4.14: MSOS simulations of $B22/B22, ST$ as a function of ionic strength at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are gray circles (reproduced from Figure 4.3), static nominal results are solid blue lines (reproduced from Figure 4.12), static PROPKA results are dashed orange lines, and fluctuating nominal results are dash-dotted red lines. Shaded regions correspond to uncertainty in the ψ parameter from Table 4.3. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data. 131
- Figure 4.15: High- $c2$ predictions of $Sq = 0$ as a function of MAb concentration ($c2$) at the pH 6.5 low ionic strength formulation for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are filled gray circles (reproduced from Figure 4.3) and predictions with the static nominal method are filled blue triangles (reproduced from Figure 4.13). Predictions from the static PROPKA and fluctuating nominal methods are shown as open orange squares and open red diamonds, respectively. Lines connecting the predictions are guides to the eye and match the formatting from Figure 4.14. Error bars correspond to uncertainty of the ψ parameter (see Table 4.3). The purple dash-dotted line is a steric-only equation of state.⁵⁰ The y-axis scale for MAb 2 (panel b) is extended to fit all of the data. 134
- Figure 4.16: Mayer-weighted average electrostatic energy (ϕ_i) for charged sites in MAb 2 at pH 5 at low ionic strength (a) and pH 6.5 at low ionic strength (b). Positively (negatively) charged atoms are shown in blue (red). Domain boundaries are vertical dashed grey lines with labels on the x-axis. The black dashed line at $\phi_i = 0$ denotes zero net energy. Error bars are 95% confidence intervals from 3 independent simulations..... 138
- Figure 5.1: Observed aggregation rate coefficient (k_{obs}) values as a function of MAb concentration ($c2$) at accelerated conditions (see Table 5.2) for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D), with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. Solid (dashed) lines connecting the data points correspond to low (high) IS and are guides to the eye. 157

- Figure 5.2: Observed aggregation rate coefficient (k_{obs}) values as a function of MAb concentration (c_2) at 30 °C for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D), with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. Solid (dashed) lines connecting the data points correspond to low (high) IS and are guides to the eye. The gray shaded region for $\log_{10}k_{obs}days - 1 < -4.86$ represents the measurement noise floor and any formulations with data not shown were below the measurement noise floor. 159
- Figure 5.3: Observed aggregation rate coefficient (k_{obs}) values at refrigerated conditions for 130 mg/mL solutions, with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. Symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. MAb identities are listed on the x-axis. Aggregation rates for MAb 3 at pH 5 low IS, MAb 3 at pH 5 high IS, and MAb 6 at pH 5 low IS were below the measurement floor (the shaded region for $\log_{10}k_{obs}days - 1 < -4.86$) after at least 1 year. Error bars are 95% confidence intervals. 160
- Figure 5.4: Relative stability rankings, where a higher ranking was a slower k_{obs} value, of each 130 mg/mL formulation as a function of T_{inc} (nonlinear in x-axis; “Acc.” is short for accelerated and “Ref.” is short for refrigerated) for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D). The legend is the same for all panels where symbols for each formulation from Table 5.1 and solid (dashed) lines connecting the data points correspond to low (high) IS. When the k_{obs} values of multiple formulations were statistically indistinguishable, the rankings were averaged. 162
- Figure 5.5: Parity plots comparing aggregation rates for accelerated (blue triangles), 30 °C (orange circles), and refrigerated (red squares, 130 mg/mL only) conditions between: (A) pH 5 and pH 6.5 formulations, and (B) low and high ionic strength formulations. Aggregation rates for MAb 3 at accelerated temperature conditions are omitted from panel A as the incubation temperatures were different between the two pH conditions (see Table 5.2). The black dashed line is the parity line, and the gray region represents the measurement floor of the study ($\log_{10}k_{obs}days - 1 = -4.86$). The numbers shown in the measurement floor regions correspond to formulations where one or both of the aggregation rates were less than the measurement floor at 30 °C or refrigerated conditions. 163

- Figure 5.6: Parity plot comparing k_{obs} values at refrigerated conditions for 130 mg/mL formulations that were measured experimentally (y-axis) and k_{obs} values predicted by Arrhenius extrapolation (x-axis) via rates at 30 °C (Figure 5.2) and accelerated conditions (Figure 5.1). The black dashed line is the parity line. Symbols for each formulation correspond to Table 1 and error bars are 95% confidence intervals. 165
- Figure 5.7: SEC chromatograms that illustrate differences in HMW formation as a function of T_{inc} . The 130 mg/mL MAb 4 at pH 6.5 low IS formulation was used as the example case. Each panel is a time course of samples from (A) accelerated, (B) 30 °C, and (C) refrigerated studies. Incubation times are shown in the legend, where T_0 represents the unstressed sample. 167
- Figure 5.8: Observed aggregation rate coefficient values (k_{obs}) as a function of the difference between the pH and pI ($|pH-pI|$) for 130 mg/mL formulations at refrigerated (red squares), 30 °C (orange circles), and accelerated (blue triangles) conditions. For the accelerated conditions, 50 °C rates are shown as blue upwards triangles, and 45 °C rates are shown as light blue downwards triangles. Filled and open symbols correspond to low IS and high IS conditions, respectively. The gray region represents the measurement floor of the study ($\log_{10}k_{obs}days - 1 = -4.86$). 171
- Figure 5.9: Observed aggregation rate coefficient values (k_{obs}) as a function of the difference between the pH and pI ($|pH-pI|$) for (A) 10 mg/mL and (B) 35 mg/mL formulations at 30 °C (orange circles) and accelerated (blue triangles) conditions. For the accelerated conditions, 50 °C rates are shown as blue upwards triangles, and 45 °C rates are shown as light blue downwards triangles. Filled and open symbols correspond to low IS and high IS conditions, respectively. The gray region represents the measurement floor of the study ($\log_{10}k_{obs}days - 1 = -4.86$). 173
- Figure 5.10: DSC thermograms for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D) in the following buffer solutions: pH 5 low IS (blue), pH 5 high IS (green), pH 6.5 low IS (purple) and pH 6.5 high IS (yellow)..... 175

- Figure 5.11: Aggregation rates (*kobs* values) at accelerated conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the *Tinc* (45 °C or 50 °C, see Table 5.2) and the first midpoint unfolding temperature (*Tm, app1*). Rates at *c2* values of 10, 35, and 130 mg/mL are shown as red squares, orange circles, and blue triangles, respectively. Dashed lines are linear fits for a given *c2* and error bars are 95% confidence intervals. 177
- Figure 5.12: Aggregation rates (*kobs* values) at 30 °C conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the *Tinc* (30 °C) and the first midpoint unfolding temperature (*Tm, app1*). Rates at *c2* values of 10, 35, and 130 mg/mL are shown as red squares, orange circles, and blue triangles, respectively. Dashed lines are linear fits for a given *c2* and error bars are 95% confidence intervals. The gray region represents the measurement floor of the study ($\log_{10}kobsdays - 1 = -4.86$). 179
- Figure 5.13: Aggregation rates (*kobs* values) at refrigerated conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the *Tinc* (4 °C) and the first midpoint unfolding temperature (*Tm, app1*). Rates at 130 mg/mL are shown as blue triangles. Dashed lines are linear fits for a given *c2* and error bars are 95% confidence intervals. The gray region represents the measurement floor of the study ($\log_{10}kobsdays - 1 = -4.86$). 180
- Figure 5.14: Aggregation rates (*kobs* values) at accelerated conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the *Tinc* (45 °C or 50 °C, see Table 5.2) and the second midpoint unfolding temperature (*Tm, app2*). Rates at *c2* values of 10, 35, and 130 mg/mL are shown as red squares, orange circles, and blue triangles, respectively. Error bars are 95% confidence intervals. 182
- Figure 5.15: Observed aggregation rate coefficient (*kobs*) values as a function of MAb concentration (*c2*) at 50 °C for MAb 1, with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. Symbols correspond to Table 5.1, with two other solution conditions Lines connecting the data points are guides to the eye and error bars are 95% confidence intervals. Adapted from Ghosh et al.⁴³ 183

- Figure 5.16: Aggregation rates at accelerated temperature conditions for MAb 1 plotted against the difference between the *Tinc* (50 °C) and the first midpoint unfolding temperature (*Tm, app1*). Dashed lines are linear fits for a given MAb concentration and error bars are 95% confidence intervals. Note that the x-axis is narrowed compared to Figure 5.11 in the main text. Data was adapted from Ghosh et al.⁴³ 184
- Figure 5.17: Comparison of the change in *G22* and the change in the aggregation rate (*kobs* values) at (A) accelerated, (B) 30 °C, and (C) refrigerated conditions due to addition of 100 mM NaCl (referred to as high IS in this chapter). Data for MAb 3, MAb 4, MAb 5, and MAb 6 are shown as light blue squares, green circles, purple triangles, and yellow diamonds, respectively. pH 5 conditions are filled symbols, and pH 6.5 conditions are open symbols. Data for MAb 1 (gray pentagons, panel A only) at 50 °C are reproduced from Ghosh et al.⁴³ 186
- Figure 5.18: Correlation coefficients (R values) for quantities that were considered to represent net self-interactions at 130 mg/mL. *Dc* and *G22* were measured at 130 mg/mL and all others listed on the axes were not *c2* dependent..... 188
- Figure 5.19: Correlation coefficients (R values) between input features and 130 mg/mL aggregation rates (*kobs* values) at accelerated (blue), 30 °C (orange), and refrigerated (red) conditions. *Tm, app1* – *Tinc* is shortened to *Tm, app1* for readability. 192
- Figure 5.20: Statistical framework applied in this section. Green boxes are the inputs and the light blue box is the output..... 193
- Figure 5.21: ENR model predictions for 130 mg/mL aggregation rates for models with feature(s) derived from stability studies at other conditions at: (A) accelerated conditions, (B) 30 °C, and (C and D) refrigerated conditions. Panels A, B, and C are parity plots of measured aggregation rates (x-axis) and aggregation rates predicted by a linear ENR model (y-axis). Panel D shows predicted aggregation rates by the logistic ENR model with ($\log_{10}kobsdays - 1 > -4.56$) as the boundary criterion, where observed aggregation rates are shown as filled black squares. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. 196

- Figure 5.22: ENR model predictions for 130 mg/mL aggregation rates for models without feature(s) derived from stability studies at other conditions at: (A) accelerated conditions, (B) 30 °C, and (C and D) refrigerated conditions. Panels A, B, and C are parity plots of measured aggregation rates (x-axis) and aggregation rates predicted by a linear ENR model (y-axis). Panel D shows predicted aggregation rates by the logistic ENR model with $(\log_{10}k_{obs}days - 1 > -4.56)$ as the boundary criterion, where observed aggregation rates are shown as filled black squares. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. 197
- Figure 5.23: Logistic models for 130 mg/mL subvisible particle formation for the following conditions: (A) accelerated, (B) 30 °C, (C) 30 °C, low IS only, (D) 30 °C, high IS only, and (E) refrigerated. The black squares are the experimental results (see Table 5.6 above) and symbols for predicted values correspond to Table 5.1..... 203
- Figure 5.24: Subvisible particle areas for 130 mg/mL, pH 6.5 MAb 3 formulations at accelerated (45 °C) conditions. The low and high IS conditions are shown as filled and open green triangles, respectively (to match Table 5.1). Solid and dashed lines connecting the data points correspond to low and high IS, respectively, and are guides to the eye..... 206
- Figure 6.1: Domain-domain radial distribution functions between 1bC/D domain beads in 130 mg/mL simulations of MAb 4 at the following formulations: (A) pH 5 low ionic strength, (B) pH 5 high ionic strength, (C) pH 6.5 low ionic strength, and (D) pH 6.5 high ionic strength. pH 5 results are shown in purple and pH 6.5 results are shown in green to match the color scheme in Chapter 5..... 223
- Figure A.1: Experimental B_{22}, ST values from static light scattering for MAb 1 (A), MAb 2 (B), MAb A (C), MAb B (D) and MAb C (E).^{42,49,50} Values in panel (E) are designated as $-G_{22}, ST$ instead of B_{22}, ST because some values are so large that they indicate multi-body interactions at the experimental protein concentrations, rather than two-body interactions that are captured by B_{22} .¹⁹¹ 254
- Figure A.2: Left: The number of atoms in MAb B that have SASA above a given cutoff value. 50.3% of the atoms have $SASA > 0.2 \text{ \AA}^2$. Right: Percent deviation of all-atom steric-only simulations of the second virial coefficient as a function of SASA. Percent deviation was defined as $100 * B_{22}, ST_{all - atom} - B_{22}, ST_{cutoff} / B_{22}, ST_{all - atom}$ 255

Figure A.3: Geometric features for the HEXA and DODECA models. Reproduced from Calero-Rubio et al. ⁵⁰	258
Figure A.4: DODECA (left) and HEXA (right) surface response of ARD as a function of ϵSR and ψ for MAb 1 at pH 5 as an example. Because HEXA and DODECA are much more computationally efficient, a global optimization was performed for both ϵSR and ψ which takes into account the contribution from short-ranged non-electrostatic interactions at intermediate TIS as well (e.g., 100 mM). For monotonic profiles of $B22B22, ST$ vs. TIS, the difference in the resulting ϵSR values between the two procedures was negligible.	259
Figure A.5: $B22, ST$ as a function of the domain bead diameter for MAb 1 (black), 2 (indigo), A (green), B (yellow) and C (purple) for the 1bC/D model. Circles on each curve show the bead diameter that results in the correct value of $B22, ST$. The curves for MAb 1 and MAb 2 overlap due to domain beads covering more charged sites as bead diameter is increased. The effect of charged beads and non-planar structure of domain beads can also be seen in the lower value of bead diameter for all MAbs compared to 3.5 nm for DODECA.....	260
Figure A.6: $B22B22, ST$ as a function of ϵSR , where electrostatic interactions are ignored (i.e., $\psi = 0$) for (left) HEXA (red) and DODECA (blue), and (right) 1bC/D for MAbs 1 (solid black), 2 (dashed indigo), A (dotted green), B (dash-dotted yellow), and C (dash-dot-dotted purple).	261
Figure A.7: Left: Experimental $B22B22, ST$ measurements (black) and simulated predictions with ψ that minimized ARD for HEXA (indigo), DODECA (green), 1bC/D (yellow) and 1bAA (purple). Right: ARD for HEXA, DODECA, 1bC/D and 1bAA as a function of ψ . The qualitative experimental self-interaction behavior (i.e., Case i, ii or iii) is listed in the headers of each panel.	264
Figure A.8: ARD as a function of ψ for different values of ΔH for the 1bC/D model, with ΔH values: 0 Å (black), 0.75 Å (indigo), 1.5 Å (green), 2.25 Å (yellow), and 3 Å (purple). The qualitative experimental self-interaction behavior (i.e., Case i, ii or iii) is listed in the headers of each panel.	266
Figure B.1: Excess Rayleigh ratio ($R90exK$) as a function of protein concentration ($c2$) for the FP protein at pH 5 (A) and pH 6.5 (B) and the monovalent Fc-fusion at pH 5 (C) and pH 6.5 (D). The curves are fits to Equation 3.1 used to calculate Mw, app	267

Figure B.2:	DLS results for the FP protein. Collective diffusion coefficient, DC at pH 5 (A) and pH 6.5 (B), and polydispersity index at pH 5 (C) and pH 6.5 (D) as a function of protein concentration (c_2).	268
Figure B.3:	DLS results for the monovalent Fc-fusion protein. Collective diffusion coefficient, DC at pH 5 (A) and pH 6.5 (B), and polydispersity index at pH 5 (C) and pH 6.5 (D) as a function of protein concentration (c_2).	269
Figure B.4:	Polydispersity as a function of protein concentration (c_2) for the bivalent Fc-fusion protein at pH 5 buffer-only (solid black squares), pH 5 + 300 mM NaCl (open black squares), pH 6.5 buffer-only (solid gray circles), and pH 6.5 + 300 mM NaCl (open gray circles).	270
Figure B.5:	RMSD as a function of ψ for the FP protein (A) and the monovalent Fc-fusion (B) at pH 5 (indigo) and pH 6.5 (green).	270
Figure B.6:	Simulated contour plots of $B_{22}/B_{22}, ST$ as a function of ψ and ionic strength for the FP protein at pH 5 (A) and pH 6.5 (B), and the monovalent Fc-fusion at pH 5 (C) and pH 6.5 (D).	271
Figure B.7:	A) Simulated $B_{22}/B_{22}, ST$ from 1bAA simulations of the wild-type protein (gray), the Fc positive-to-negative charge swap variant (indigo), and the fusion partner negative-to-positive charge swap variant (green) for the monovalent Fc-fusion and bivalent Fc-fusion. B-E) Corresponding Mayer-weighted average pairwise electrostatic energy (ϕ_{ij}) heatmaps for the monovalent Fc-fusion (B: Fc swap, D: fusion partner swap) and bivalent Fc-fusion (C: Fc swap, E: fusion partners swap). The heatmaps are ordered based on the structural domains with fusion partner protein abbreviated as FP.	274
Figure C.1:	G_{22} as a function of c_2 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles). ^{49,50} The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.	276
Figure C.2:	RMSD as a function of ψ from MSOS simulations of $B_{22}/B_{22}, ST$ at pH 5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (e). DODECA results are green lines, and 1bC/D results are indigo lines.	277

Figure C.3: RMSD as a function of ψ from MSOS simulations of <i>B22/B22, ST</i> at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (e). DODECA results are green lines, and 1bC/D results are indigo lines. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.....	278
Figure C.4: RMSD as a function of ψ from MSOS simulations of <i>B22/B22, ST</i> at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (e). Static nominal results are blue lines, static PROPKA results are orange lines, and fluctuating nominal results are red lines. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.	279
Figure C.5: High- c_2 predictions of $Sq = 0$ as a function of MAb concentration (c_2) at pH 6.5 for MAb 4 with the fluctuating nominal charge representation method. Low ionic strength results (reproduced from Figure 4.14) are shown as open red diamonds with dash-dotted lines and intermediate ionic strength results are shown as filled black diamonds with dashed lines. Lines connecting the predictions are guides to the eye and match the formatting from Figure 4.14. Error bars correspond to uncertainty of the ψ parameter (see Table 4.3). The purple dash-dotted line is a steric-only equation of state. ⁵⁰	280
Figure D.1: Parity plot of k_{obs} calculated by change in monomer concentration (case 1; x-axis) and change in aggregate peak area (case 2; y-axis). Formulations where case 1 was chosen are shown in green, and formulations where case 2 was chosen are shown in purple, matching the color scheme in Table D.1. The gray shaded region for $\log_{10}k_{obs}days - 1 < -4.86$ represents the measurement noise floor.	282
Figure D.1.1.1: Correlation coefficients for features in the net valence at a given pH subgroup.	286
Figure D.1.2.1: Correlation coefficients for features in the net self-interactions subgroup. This figure is largely the same as what is shown in Figure 5.18 but is included here for completeness.	287
Figure D.1.3.1: Correlation coefficients for features in the conformational stability subgroup.	288
Figure D.1.4.1: Correlation coefficients for features in the stability studies (accelerated) subgroup.	289
Figure D.1.5.1: Correlation coefficients for features in the stability studies (30 °C) subgroup.	290

Figure D.1.5.2: Comparison of k_{obs} values from 130 mg/mL accelerated studies (x-axis; either 45 °C or 50 °C), and $k_{obs, acc}$, where rates at 45 °C were converted to estimated rates at 50 °C via Arrhenius extrapolation (y-axis). The black dashed line is the parity line.....	291
Figure D.1.6.1: Correlation coefficients for features in the stability studies (refrigerated) subgroup.....	292
Figure D.1.7.1: Correlation matrix for input features for models of accelerated conditions. The last entry is the $\log_{10}k_{obs}$ value for 130 mg/mL accelerated conditions. $T_m, app1 - T_{inc}$ is shortened to $T_m, app1$ for readability.....	294
Figure D.1.7.2: Correlation matrix for input features for models of 30 °C conditions. The last entry is the $\log_{10}k_{obs}$ value for 130 mg/mL 30 °C conditions.	295
Figure D.1.7.3: Correlation matrix for input features for models of refrigerated conditions. The last entry is the $\log_{10}k_{obs}$ value for 130 mg/mL refrigerated conditions.....	295

ABSTRACT

Protein-based therapies are a prominent class of drug products used in the treatment of a broad range of chronic illnesses such as cancers and immune-related disorders, and more recently infectious diseases such as SARS-CoV-2 and RSV. Many of the highest selling drugs globally are protein-based therapeutics, typically monoclonal antibodies (MAbs) or structurally derivative proteins such as Fc-fusion proteins and bispecific antibodies. The development process for therapeutic proteins is particularly uncertain, expensive, and resource-intensive compared to small molecule drugs, so there is sizable interest in the biopharmaceutical industry in methods that can improve predictions of how likely a protein drug candidate is to be successfully developed into a commercial product (also known as “developability”), and in methods that can streamline the development process.

Many of the challenges that are faced during drug development of therapeutic proteins arise from protein-protein self-interactions, which are “weak” intermolecular forces (i.e., weak in comparison with “lock-and-key” specific binding events) between proteins of the same species in solution. The influence of self-interactions on solution nonidealities and problematic behaviors is increased at elevated protein concentrations, which is of particular relevance as the preferred liquid dosage form for many protein-based therapies is at relatively high protein concentration (on the order of 100 mg/mL). Static light scattering (SLS) and dynamic light scattering (DLS) are commonly used to measure net self-interactions in early-stage development to screen for attractive self-interactions that are fundamentally associated with a host of

challenging behaviors and properties such as reversible self-association, irreversible aggregation, elevated viscosity, liquid-liquid phase separation, opalescence, and low solubility. Irreversible aggregation is especially problematic because proteins have a common tendency to aggregate and methods to predict changes in aggregation rates or mechanisms between different proteins or as a function of solution conditions are not well-developed. The presence of aggregates can be a liability in a number of manufacturing processes, reduce the efficacy and shelf-life of the product, and elicit a dangerous immunogenic response when administered to a patient. This thesis is focused on the development and assessment of methods to characterize and predict self-interactions and aggregation rates for therapeutic proteins with emphasis on practical applications in streamlining industrial drug development. The experimental datasets are for solution conditions and proteins similar to those in commercial protein-based therapies, fairly diverse in the behaviors they represent, and large compared to many other publicly-available datasets.

Coarse-grained (CG) molecular simulations are applied throughout this thesis to model self-interactions, predict net self-interactions at high-concentration conditions, and probe specific electrostatic interactions between charged residues that were involved in attractive self-interactions. A range of coarse-grained models for therapeutic proteins were evaluated based on the tradeoffs between computational efficiency and accuracy in calculating net self-interactions. A dataset of previously reported experimental values of the second osmotic virial coefficient (B_{22}) from SLS for five MAbs at multiple solution conditions (i.e., different pH and ionic strength conditions) were used as a test case. Lower resolution (e.g., domain-level) models allowed for higher throughput and more intensive simulation algorithms (e.g.,

simulations with many protein molecules to simulate high concentrations) but were limited in their representation of interactions between specific sites in the protein, such as attractive electrostatic interactions between specific charged amino acids. Higher resolution models were able to capture specific electrostatic attractions, but at great cost to computational efficiency. A hybrid model that combines features from the domain-level and higher resolution models was introduced that can capture specific electrostatic attractions and was tractable for simulations at high-concentrations like those representative of commercial therapeutic protein products.

Net self-interactions via SLS and DLS experiments were measured systematically for four MAbs, two Fc-fusion proteins and the associated fusion partner (FP) protein as a function of solution pH and ionic strength. The measurements for the Fc-fusion proteins and FP protein were confined to low-concentration (e.g., B_{22} values), while for the four MAbs, the measurements scaled from low to high protein concentration. The solution conditions were chosen to represent fundamental features of commercial drug products within typical bounds of each feature (i.e., pH, ionic strength, and protein concentration). The proteins displayed a broad range of net self-interactions from strong repulsions to strong attractions that were sensitive to the changes in solution conditions that were assessed.

The two Fc-fusion proteins and FP protein displayed reversible self-association at some conditions, which is linked to many industrial development challenges and is also a possible precursor to irreversible aggregation. The reversible self-association appeared to be related to attractive electrostatic self-interactions, so a high-resolution CG model was used to investigate the origins of attractive electrostatic self-interactions for the two Fc-fusion proteins. The results indicated that they were due to

cross-domain interactions between the FP and Fc domains, which suggests that reversible self-association was due to those interactions as well.

A previously developed method to combine low-concentration experimental values of B_{22} with CG molecular simulations to make predictions of high-concentration net self-interactions was improved by the integration of the hybrid CG model. The domain-level and hybrid CG models were directly compared based on how well they predicted high-concentration net self-interactions, using B_{22} values from SLS for six MAbs (two from prior work) to parameterize the CG models for a given MAb and pH. The predicted net self-interactions were compared against high-concentration SLS measurements. The findings and guidance from the CG model comparison described above with respect to low-concentration net self-interactions were also generally applicable for high-concentration net self-interactions where domain-level CG models were only able to reliably capture net repulsions and weak non-electrostatic attractions, while the hybrid CG model could capture strong electrostatic attractions as well. Inaccurate predictions of high-concentration behavior with the hybrid CG model at certain conditions were improved by methods that represented charge equilibria more precisely.

The four MAbs were also used for studies with the overall goal of understanding and predicting MAb aggregation rates between different MAbs and as a function of solution conditions. Conformational stability of the four MAbs at four different formulations was measured by differential scanning calorimetry, and aggregation rates were measured via isothermal stability studies at four formulations (varying both pH and ionic strength) as a function of protein concentration and incubation temperature. Prediction of aggregation rates for solutions at high protein

concentration stored at refrigerated conditions was of particular interest as it was intended to directly represent the protein concentration and storage condition of many commercial products. Studies at higher temperatures, where aggregation rates were generally faster, were judged by how they might relate to aggregation rates at refrigerated conditions and how similar the fundamental factors that mediated aggregation rates were. Overall, studies at elevated temperatures were poor predictors of aggregation rates at refrigerated storage conditions. Interpretable machine learning models were developed to rigorously deconvolute the impacts of fundamental phenomenon on aggregation rates, which included the net self-interactions and conformational stability measurements. At the highest temperatures, conformational stability was the most influential phenomenon, while at lower and refrigerated temperatures, net valence was the most influential, perhaps due to the influence of repulsive electrostatic self-interactions. The ML methods were also used to more thoroughly assess whether results from stability studies at higher incubation temperatures or lower protein concentrations could be useful for predicting aggregation rates. Another goal in developing the ML models was to provide a robust platform for predicting aggregation rates with the vast datasets that are not publicly available but presumably exist in the archives of many pharmaceutical companies.

The studies in this thesis developed computational and statistical methods that were validated by or trained with fairly large, systematic datasets of experimental biophysical characterization, especially with respect to self-interactions. The results demonstrate how to 1) select a CG molecular model for a given application, 2) use CG molecular simulations in close connection with experimental measurements to extract additional knowledge about self-interactions and predict net self-interactions at other

conditions (e.g., higher protein concentrations), and 3) understand and predict MAb aggregation rates as a function of protein concentration, incubation temperature, and solution conditions. These findings can be applied to various phases of industrial drug development for MAbs, Fc-fusion proteins, or other therapeutic proteins to improve selection of protein candidates (i.e., candidate selection) and optimization of formulation conditions (i.e., formulation development).

Chapter 1

INTRODUCTION

1.1 Motivation

Modern healthcare relies in part on drug products that are developed and commercialized by pharmaceutical companies in a competitive marketplace. The drug development process is exceptionally expensive; per approved new compound, R&D costs are well over \$1 billion when factoring in costs for unsuccessful drug candidates.¹ Currently, most of the highest selling drugs globally are biologics, either for the treatment or prevention of SARS-CoV-2, or protein-based therapeutics for oncology and treatment of immune-related disorders.² The highest selling protein-based therapeutics are typically monoclonal antibodies (MAbs) or their related derivatives, e.g., Fc-fusion proteins, bispecific antibodies, and antibody-drug conjugates.²⁻⁵ As of 2023, there have been over 100 MAb products approved by the US Food and Drug Association (FDA).^{3,6-9}

Many different factors contribute significantly to development costs for therapeutic proteins, such as performing clinical trials, early stage drug discovery, product development, and commercialization.^{1,5,10,11} Not only must drug candidates under consideration be safe and efficacious, but a balance must be maintained between investing in screening a large number of candidates for how likely they are to be successful (i.e., risk assessment), and moving forward quickly with a relatively small number of candidates.^{12,13} Naturally, much emphasis has been placed on streamlining the development process for therapeutic proteins, and in recent years industrial

capabilities have improved significantly, allowing for a more broad range of disease targets such as infectious diseases.^{4,5,14} Therapeutic proteins to treat many novel disease targets such as HIV, Ebola, Hepatitis B, and Alzheimer's are now in clinical development, with a few recently approved by the FDA.^{5,9,15} Still, numerous challenges persist that limit the developability of a protein candidate, or the baseline potential, based on its physicochemical properties, to be successfully developed into a safe, efficacious product that can be consistently manufactured and is stable at its intended storage conditions.^{11,16-18} Improving predictions of developability or expanding the space of developable proteins is of great industrial interest to reduce costs and shorten development timelines.¹⁷

Protein therapies involving MAbs and similar proteins require dosages on the order of 10 or 100 mg for the desired therapeutic effect. The large majority of commercially-available therapeutic proteins are administered either via intravenous (IV) infusion or subcutaneous (SC) injection.³ Typical IV infusions are 100 mL or 250 mL, while SC injection volumes are generally less than 2 mL.^{3,19} Administration via SC injection greatly reduces the burden on health care providers and patients, and is considered the preferred administration route, whenever possible.^{15,19,20} From a biophysical perspective, proteins are fairly dilute in IV infusion (on the order of 1 mg/mL) and quite concentrated when used for SC injection (on the order of 100 mg/mL).^{8,15,21} High-concentration protein solutions are particularly prone to a number of problematic solution behaviors such as reversible and irreversible aggregation, elevated viscosity, phase separation, low solubility, and opalescence.^{16,22-25}

1.2 Protein-protein self-interactions

Proteins in liquid solution participate in intermolecular interactions with the solvent, cosolutes, and crucially, other proteins. This thesis examines topics that are directly or fundamentally related to “weak” intermolecular interactions between proteins of the same species in solution, known as protein-protein self-interactions (referred to hereafter as “self-interactions”). In the context of this thesis, those interactions are distinct from stronger or “lock-and-key” specific binding events whose dissociation constants (K_D values) are typically on the order of nanomolar to micromolar.^{26,27} Self-interactions are commonly referred to elsewhere as protein-protein interactions or colloidal stability. Self-interactions include contributions from at least three fundamental interaction categories: steric repulsions, electrostatic interactions between charged residues, and short-ranged non-electrostatic attractions such as those due to hydrophobic, hydrogen bonding, and van der Waals forces.²⁸ Those fundamental forces are influenced by the protein (e.g., the amino acid sequence, and higher-order structure including whether the protein is in its native state), the solution conditions (i.e., protein concentration, pH, concentration of any cosolutes), and environmental conditions (e.g., temperature and pressure). Electrostatic interactions intrinsically act on a longer length scale compared to short-range non-electrostatic attractions, evidenced by the difference in how analytical models for force scale with distance (r): electrostatic interactions (i.e., Coulomb’s law) scale with $1/r^2$, while e.g., van der Waals forces scale with $1/r^6$.²⁹ However, the relative length scales of electrostatic interactions and short-ranged non-electrostatic attractions can shift as the range of electrostatic interactions can be reduced by electrostatic screening due to the concentration of ions in solution (which can include other proteins), commonly expressed via the Debye screening length (κ).²⁹

Electrostatic repulsions are in some cases due to general “simple-colloidal” repulsions between proteins with high net charge. In the other extreme, electrostatic repulsions or attractions can be highly specific to a certain subset of amino acids and therefore can be specific to the anisotropy of charged sites on the protein surface.³⁰ The protein surface charge distribution is dependent on the amino acid sequence, the pH of the solution, interactions with cosolutes (especially ions), and the pK_a of each ionizable (or titratable) residue. Nominal pK_a values for each ionizable residue serve as a useful starting point and as a surrogate for the average pK_a of a given residue type in a protein, but the pK_a value of an ionizable residue is also influenced by its local environment (e.g., the influence of other proximal residues in the protein).^{31,32} Short-ranged non-electrostatic attractions are also dependent on the distribution of surface-exposed residues. One general approach to estimating the strength of those attractions is via relative hydrophobicity scales, although some interactions within this category are not so generalizable, such as hydrogen bonding or π - π interactions.³³ Short-ranged non-electrostatic attractions can be of increased relevance to the overall self-interactions in higher ionic strength conditions where the longer-ranged electrostatic interactions are screened, at high protein concentration where proteins are forced to be close to one another, or if specific interaction orientations are highly attractive.

Experimentally, self-interactions can be quantified in terms of the second osmotic virial coefficient (B_{22}) via techniques such as static light scattering (SLS), analytical ultracentrifugation (AUC), self-interaction chromatography, and small-angle x-ray or neutron scattering.³⁴⁻³⁹ An analog to B_{22} is the interaction parameter from dynamic light scattering (DLS), k_D , which is well-known to be correlated with B_{22} for therapeutic proteins, as shown in Figure 1.1 which is a collection of published

data for MAbs.^{40,41} These methods all quantify net self-interactions, the overall tendency of a protein to be attracted or repelled by another molecule of the same protein species in solution.

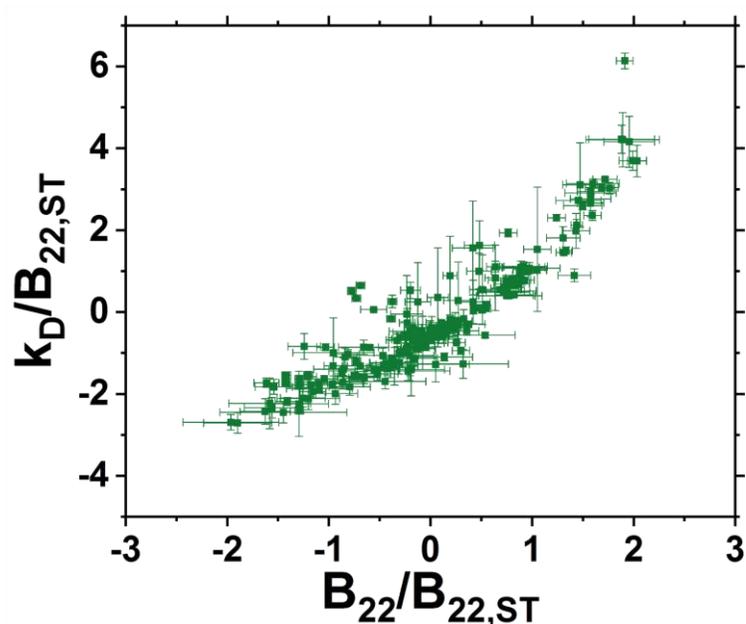


Figure 1.1: Comparison of published B_{22} and k_D values for MAbs, where B_{22} and k_D are normalized by $B_{22,ST}$, the steric-only contribution to B_{22} (see section 2.2.1).^{42–48}

B_{22} and k_D are dilute solution quantities that are generally measured at protein concentrations up to ~ 10 mg/mL. In dilute conditions, intermolecular distances between proteins are relatively large compared to the range of their self-interactions, thus on average a given protein will only interact significantly with one other protein. That is, the experimental behavior captures two-body interactions with reasonable fidelity.⁴⁹ However, at elevated protein concentrations, intermolecular distances

decrease and, on average, proteins experience significant interactions with multiple neighboring proteins.⁴⁹ As such, although measurements at low protein concentration provide a fundamental quantity related to how proteins interact with one another, they are not necessarily generalizable predictors of high-concentration behavior.^{50,51} Net self-interactions from low- to high-concentration can be quantified more generally by the Kirkwood-Buff protein-protein integral, G_{22} , which is related to B_{22} via $B_{22} = -\frac{1}{2} \lim_{c_2 \rightarrow 0} G_{22}$.^{35,52} Experimental methods used in this work to measure net self-interactions are described in detail in the relevant Chapters.

Measurements of net self-interaction parameters such as B_{22} and k_D are commonplace in the development of therapeutic proteins to screen different drug candidates and formulation conditions for attractive self-interactions that might cause potential poor biophysical properties (discussed in more detail in section 1.5).^{53,54} In some cases (e.g., elevated viscosity or liquid-liquid phase separation), that strategy can be effective as low-concentration measurements of net self-interactions are correlated with those properties, even at significantly higher protein concentrations.^{54,55} However, some phenomena such as reversible self-association and irreversible aggregation (discussed in more detail in section 1.3) have less clear relationships with measurements of net self-interactions despite a straightforward phenomenological link.^{53,56-61}

1.2.1 Coarse-grained molecular simulations

Molecular simulations of proteins are rising in popularity as tools to replace or augment experiments to investigate, e.g., self-association, protein folding, phase separation, solution viscosity, and molecular packing.^{28,42,49,50,62-79} Continuing advancements in high-performance computing infrastructure have made these tools

more accessible and efficient, enabling integration into industrial therapeutic protein drug development. Of particular interest in this thesis are coarse-grained (CG) Monte Carlo (MC) molecular simulations that are applied to model and predict self-interactions of therapeutic proteins in low- to high-concentration systems.

CG models are created by “lumping” multiple atoms or amino acids into a single “bead”, which reduces the number of interaction sites in the model compared to an atomistic model and thus also reduces the computational burden. CG models for proteins can vary from a single sphere to models where each amino acid has one or multiple explicit beads.²⁸ While atomistic models are the most accurate way to model a protein’s molecular structure, large proteins such as MAbs require a sizable amount of computing power to simulate at that resolution, especially when considering quantities that require many configurations to converge, or intermolecular phenomena that require multiple proteins in the simulation.²¹ Explicit representation of solvent molecules also greatly increases the computational burden.⁸⁰⁻⁸² The use of atomistic molecular models with an explicit solvent is therefore generally impractical for industrial applications, where efficient, high-throughput analysis is paramount. Similar reasoning can be applied to the choice of an MC algorithm over a molecular dynamics (MD) algorithm for certain applications: MD simulations are more computationally burdensome in that they probe shorter time scales than what can be sampled with MC simulations and sometimes require an explicit solvent.²⁸

CG MC molecular simulations are needed for simulations pertaining to self-interactions because by definition self-interactions are between multiple proteins and the configurations needed to properly calculate B_{22} or G_{22} can be quite rare in some cases, so many configurations must be generated. In some cases, the interactions

between proteins are complex or specific to particular amino acids or atoms, so the necessary level of structural can shift depending on the system in question. CG simulations in this thesis are in part a continuation of previous work that has developed several CG models and the underlying algorithms and interactions potentials that are used for simulations of B_{22} and G_{22} .^{28,70,71,83} Those interaction potentials divide the fundamental interactions of self-interactions in the same manner as was described in the previous section (section 1.2): steric repulsions, short-ranged non-electrostatic attractions, and electrostatic attractions and repulsions. Chapter 2 goes into more extensive detail on the different levels of coarse-graining within molecular simulations of proteins and the justifications for the specific CG models used throughout this thesis. Prior work also developed methods to use experimental measurements of B_{22} to parameterize CG simulations that predicted G_{22} at high-concentration, with data for two MAbs and one globular protein as test sets.^{49,50,84} The CG models used were low resolution, domain-level models with at most 12 beads per protein. The predictions of G_{22} were poor for strongly attractive conditions, which are typically caused by interactions between specific subsets of amino acids, which in those CG models had been lumped into beads with many other amino acids. Methods to deconvolute the specific, pairwise interactions between sites (e.g., amino acids) in high resolution CG models were also developed in prior work, where analysis of interactions at amino acid resolution provided insight into the specific pairwise interactions that were most responsible for the net self-interactions, particularly those that were electrostatic in nature.^{44,48,69}

1.3 Physical instability

Physical instability is used herein to denote structural changes in either the native, folded state of the protein, or the monomeric state of the protein (e.g., whether the protein is a constituent of an oligomer).²⁵ The thermodynamic stability of the folded state is also known as conformational stability and is principally an intramolecular phenomenon that is not necessarily related to self-interactions.^{56,85} The formation of higher order structures (e.g., oligomers) by bonding or strong association with one or more other proteins is intrinsically related to self-interactions, although often those associations occur between partially or fully unfolded proteins.^{25,85-87} Other forms of physical instability include adsorption to bulk interfaces, and fragmentation (i.e., peptide bond cleavage that separates a proteins into smaller parts, particularly relevant for multidomain proteins like MAbs).^{88,89} The following subsections highlight a few key interrelated biophysical behaviors that are relevant to physical instability. Figure 1.2 shows an illustrative diagram of the aspects of physical instability discussed in this section, with MAbs as an example case. The exact mechanism that is shown is one of a large number of possible pathways, as discussed below.

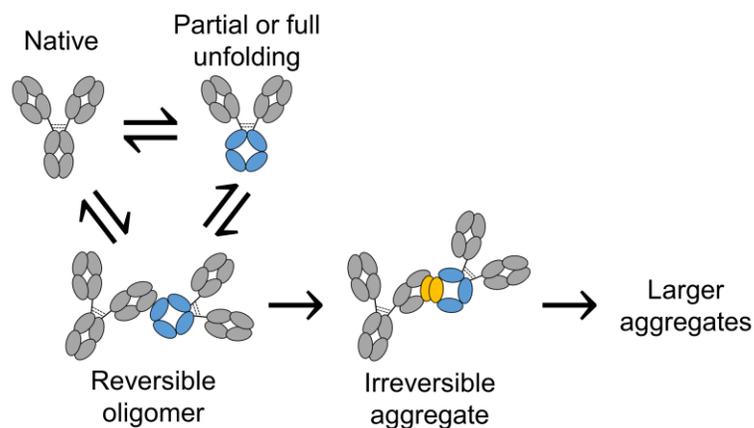


Figure 1.2: Generalized aggregation diagram with MAbs used as an example. Gray ovals represent natively folded domains, blue ovals represent unfolded or misfolded domains, and domains in gold are linked to form an irreversible aggregate. Reversible steps are represented by \rightleftharpoons symbols and irreversible steps are represented by \rightarrow symbols. The mechanism shown here is an illustrative example.

1.3.1 Conformational stability

The native folded state of a protein is stabilized by interactions between buried hydrophobic residues (and the resulting structural motifs like β -sheets), disulfide bonds, glycans (where applicable), among many other thermodynamic contributions.⁹⁰ Partial or full unfolding can occur in thermodynamic equilibrium with the folded state (as a function of temperature, pressure, solution conditions, etc.), or due to external factors such as interfacial or mechanical stress.^{91–93} Unfolded or misfolded proteins may already have lost their desired activity, and regions prone to reversible self-association and/or irreversible aggregation (discussed in more detail in the following subsections) could be exposed.^{92,94} Conformational stability is commonly approximated with differential scanning calorimetry (DSC; see Chapter 5) or isothermal chemical denaturation, where the fundamental quantity that is analyzed is

the Gibbs free energy of unfolding, ΔG_{un} . The underlying forces are thus both enthalpic and entropic in nature, and can depend on pH, ionic strength, and interactions with any solvent or cosolute molecules. That perhaps includes self-interactions particularly when the protein concentration is not dilute, but the dependence of conformational stability on protein concentration is challenging to probe experimentally. Protein concentration has been theorized to have an impact on ΔG_{un} and has been shown experimentally to impact the structure or stability of the folded state.^{43,95-98} Some computational methods have been developed to predict ΔG_{un} for proteins but were beyond the scope of this thesis.^{59,99,100}

1.3.2 Reversible self-association

In the context of this thesis, reversible self-association refers to self-interactions between two or more proteins of the same species in solution that result in the formation of transient, reversible oligomers.^{24,101-105} Reversible self-association has been linked to elevated viscosity^{34,106-108}, larger scale self-association phenomena such as liquid-liquid phase separation¹⁰⁹, and in some cases, changes in *in vivo* pharmacokinetic properties.¹¹⁰⁻¹¹³ Reversible oligomers are distinct from, but potential precursors of, net irreversible, non-native protein aggregates, as shown in Figure 1.2 and discussed in more detail in the following subsection.^{24,114-116} Partial or full unfolding of at least one of the proteins is usually necessary to expose aggregation prone regions (APRs) that make the protein “reactive” in the context of reversible self-association or irreversible aggregation.^{100,117} In comparison to the illustrative example shown in Figure 1.2, reversible oligomerization could plausibly occur via reversible self-association involving any number of native or (partially) unfolded monomers and

there could be multiple reversible steps to form larger reversible oligomers before an irreversible aggregate species is formed.

The experimental techniques mentioned in section 1.2 with respect to self-interactions can also be used for characterizing reversible oligomers, such as SLS, DLS, and AUC.^{13,101,118–121} In comparison to irreversible aggregates, reversible oligomers can be more challenging to characterize due to their transient, reversible nature as techniques that are commonly used to characterize irreversible aggregates such as SEC-MALS and capillary electrophoresis are limited by filtration through a chromatographic column, large extents of dilution, or other changes of solution conditions.^{13,107,122–124}

1.3.3 Irreversible aggregation

One of the most problematic and ubiquitous challenges in drug development of therapeutic proteins is irreversible protein aggregation (hereafter referred to as aggregation), which can create challenges in many processes within development and manufacturing, limit product shelf-life, and trigger harmful immunogenicity if aggregates are introduced to a patient.^{114,125–128} Regulatory agencies like the FDA strictly monitor the presence of physical impurities like aggregates not only when the product is first produced, but throughout its shelf-life.^{33,90} The mechanisms of protein aggregation are diverse and methods to predict the aggregation rates of different therapeutic proteins in different formulations (e.g., protein concentration, pH, identity and concentration of excipients) are not well-established.^{85,129}

The smallest or initial irreversible aggregates (also known as “nuclei”) are formed by relatively strong and stable non-covalent interactions between APRs that might occur after preliminary (partial) unfolding or reversible oligomerization

step(s).^{94,100} Those irreversible nuclei can form larger aggregates by the incorporation of monomeric proteins and/or other aggregates through many different mechanisms such as monomer addition, chain-polymerization, aggregate-aggregate coalescence or condensation.^{117,125,130} Generally speaking, multiple different mechanisms are likely occurring simultaneously with relative rates that are dependent on solution and environmental conditions as well as the overall extent of aggregation or oligomerization in solution.⁸⁶ In comparison with thermodynamically-driven reversible self-association, aggregation is a kinetic process, so the rate of aggregation is mediated by the rate limiting step of the mechanism, which can be reversible or irreversible, and can (and often does) change in different conditions. The aggregation mechanism also determines the resultant aggregate size distribution, which can be an important factor in potential immunogenicity.^{116,126,127,131–133} In some cases aggregates can grow into subvisible (or larger) particles which are under specific scrutiny by regulatory agencies (see USP <787> and <788>).^{134–136}

Irreversible aggregation is directly linked to self-interactions in the sense that two or more proteins must interact with each other (i.e., by definition it is a kind of self-interaction) and also indirectly in that self-interactions impact contact probabilities between proteins and influence the relative orientations of proteins when they come into close proximity with one another. For instance, long-ranged electrostatic repulsions reduce the probability that two proteins will approach each other close enough to potentially form an aggregate or reversible oligomer. Prior studies have shown that net repulsive electrostatic self-interactions can sometimes reduce aggregation rates.^{43,85,137–140} Chapter 5 was particularly motivated by prior work that identified an example of this subcase for a MAb that related measured G_{22}

values to aggregation rates at elevated temperature conditions, reproduced in Figure 1.3.⁴³ In that work, the relationship was specific to the difference in the aggregation rate (quantified by the observed rate coefficient, k_{obs} ; see also Chapter 5) between buffer-only solutions (denoted with the superscript “ref”) and solutions with 100 mM added NaCl. The increased ionic strength in the solutions with NaCl reduced the net strength of electrostatic repulsions via Debye screening, which was semi-quantitatively related to the relative increase in the aggregation rate.

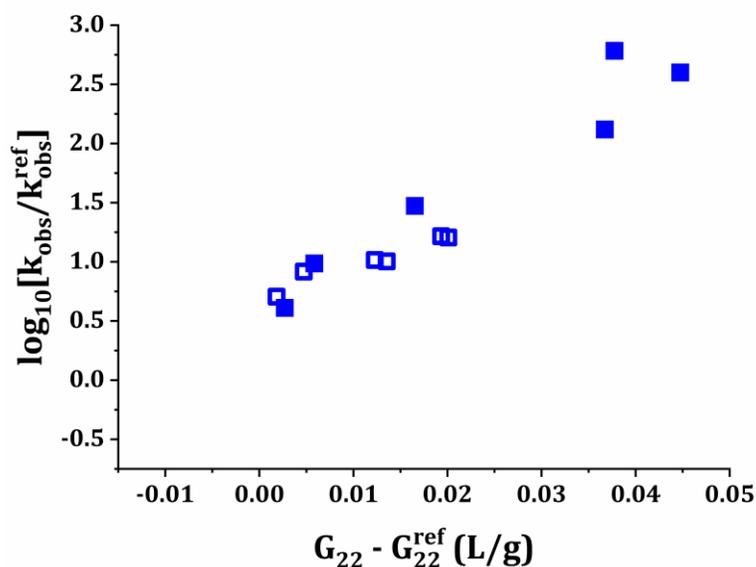


Figure 1.3: Semi-quantitative relationship between the change in net self-interactions (G_{22}) and aggregation rate (k_{obs}) when comparing buffer-only solutions (denoted with the superscript “ref”) and solutions with buffer and 100 mM NaCl. Results at pH 5 and pH 6.5 conditions are shown as filled and open blue squares, respectively. This figure is reproduced from Ghosh et al.⁴³

Solution conditions have an indirect influence on aggregation rates by mediating changes in more fundamental phenomena that can be specific to the protein in question. For example, pH is not expected to be directly relevant to aggregation in a quantitative sense, but rather how the pH impacts the spatial distribution of charged residues and the resultant intermolecular and intramolecular electrostatic interactions. Protein aggregation can be triggered by many other environmental factors as well, including elevated temperatures, interfacial stress (e.g., due to agitation and freeze-thawing), mechanical stress, chemical instability (e.g., oxidation and deamidation), light exposure, and metal contamination.^{25,141-143} Many of these potential issues can be reasonably screened with computational methods or with “accelerated” or “forced degradation” stability studies: studies at elevated temperatures, repeated freeze-thaw cycles, shaking, isothermal chemical denaturation (ICD), light exposure, and computational algorithms that predict residues that may be prone to various chemical instabilities.^{25,144} With the exception of elevated temperature, these stress conditions were not considered in this thesis, except that they were controlled for when applicable (e.g., stability studies were performed in quiescent conditions without light exposure).

Although many of the important factors that drive protein aggregation have been identified (e.g., self-interactions, conformational stability, and APRs), and there are some reports that can quantitatively or semi-quantitatively connect one or more of those features to measured aggregation rates, a “unified theory” of protein aggregation has not been developed and is perhaps not feasible due to the vast number of possible pathways.^{43,86,145} Some phenomenological models have been developed, however they lack generality and are difficult to validate with existing experimental

capabilities.^{117,130,145,146} This is particularly problematic for industrial drug development, where aggregation can occur in many different stages (e.g., expression, manufacturing, or long-term storage), and stable conditions need to be determined quickly (see also section 1.5). Aggregation has also been shown to impact viscosity, which is another critical parameter in a protein drug product.¹⁴⁷ Chapter 5 of this thesis is focused on long-term aggregation rates relevant to the shelf-life of therapeutic protein drugs, particularly for high-concentration liquid solutions such as those intended for subcutaneous administration that are typically stored at refrigerated conditions.

1.4 Objectives

This thesis is focused on experimental and computational biophysical characterization of therapeutic proteins as a function of pH, ionic strength, and protein concentration, and on the development of tools and methods to understand and make useful predictions of the underlying phenomena. Evaluating and predicting net self-interactions and assessing their relationship with reversible self-association and irreversible aggregation is of particular interest. The objectives of this thesis are as follows:

1. Define a suite of CG models for therapeutic proteins and guide model selection based on the balance between accuracy of simulated self-interactions and relative computational burdens.
2. Characterize self-interactions and reversible self-association of two Fc-fusion proteins and the corresponding fusion partner protein and apply CG molecular simulations to investigate the relevant electrostatic interactions.
3. Develop a comprehensive framework to model and make predictions of net self-interactions from low- to high-concentration via CG molecular simulations and low-concentration SLS experiments.

4. Assess the influence of self-interactions and conformational stability on aggregation rates for MABs as a function of pH, ionic strength, MAB concentration, and incubation temperature.
5. Develop interpretable machine learning methods to predict long-term aggregation rates for high-concentration MAB formulations, and deconvolute and quantify the influential biophysical phenomena.

This thesis aims to address challenges in and inform rational design of experimental and computational analyses in industrial drug development of therapeutic proteins. Large, systematic experimental datasets were used to illustrate a broad range of possible behaviors, robustly validate the methods, and corroborate the overall findings. The relevant drug development phases and the therapeutic proteins that were used in this thesis are described in the subsequent sections.

1.5 Applications in industrial drug development

Biophysical behaviors such as reversible self-association and aggregation can cause difficulties throughout the industrial drug development process. Two specific drug development phases that are of particular relevance to this thesis are discussed below in the context of self-interactions, reversible self-association, and irreversible aggregation.

1.5.1 Candidate selection

During candidate selection (also known as lead identification), a relatively large number of proteins that have been identified as having potential to be developed into successful drug products are screened and ranked based on a host of properties via computational and experimental methods. Some of the most important of those properties are related to how well the drug will perform after administration, e.g., pharmacokinetics, toxicology, and the ability to bind to the desired epitope with high

affinity, potency, and specificity.^{129,148} There are many other biophysical properties that often receive less attention or are difficult to predict when physical protein material is limited, even though suboptimal properties may jeopardize future development phases.

Reversible self-association and aggregation occur due to interactions between a relatively small number of residues (i.e., APRs). Thus, identification and mitigation of those problematic interactions during candidate selection can mitigate significant risks in later stages of development, where poor biophysical behaviors can lead to increased costs and higher attrition rates.^{11,100,149} Current computational methods to predict APRs are limited in applicability as they are not typically trained against physical stability data at intended storage conditions and representative formulations.^{100,150} Many of these models rely on data from proteins that are quite distinct from leading therapeutic proteins, like much smaller peptides or amyloid proteins, and it is not clear how translatable those models are to MAbs and other related therapeutic proteins.^{100,151} Another aspect of this challenge is understanding how APRs might be more or less active as a function of solution conditions (see also the subsequent subsection about formulation development).

Besides being phenomenologically linked to reversible self-association and irreversible aggregation, attractive self-interactions are associated with a number of other problematic behaviors, such as elevated viscosity, low solubility, phase separation, and opalescence.^{22,33,55,125,143,152–154} Many of those behaviors can be predicted with reasonable accuracy by experimental measurements of net self-interactions^{53–55,155}, which could be streamlined via methods that can predict net self-interactions with fewer or no experiments (see Chapter 4). Methods that reduce or

eliminate the need for experimental data are of particular utility in earlier stages of development like candidate selection where protein material and time can be scarce quantities. Sequence optimization is often an important part of candidate selection, where mutations in the amino acid sequence are made to improve the properties of the protein. Rational design of point mutations of charged residues was of particular interest in this thesis, with the goal of attenuating potentially problematic attractive electrostatic self-interactions (see Chapters 3 and 4).

1.5.2 Formulation development

Formulation development is the process in which the final configuration of a drug product is determined, including the dosage form. This thesis focuses on liquid solutions, whose composition (i.e., protein concentration, pH, and concentration of excipients) is optimized during formulation development.⁹¹ Other dosage forms such as lyophilized powders for reconstitution are also an option, although high-concentration liquid dosage forms are generally preferred for protein-based therapies (see Chapter 1.1).^{15,19,20} Solution conditions are determined such that the drug product is stable throughout long-term storage to ensure safety and efficacy when administered to a patient. The final product configuration and its physiochemical properties are under strict scrutiny by regulatory agencies like the FDA.⁹⁰

Formulations that are prone to physical or chemical instability, or that have poor biophysical properties (e.g., elevated viscosity) introduce a host of challenges during development.²⁵ The manufacturing process is harder to optimize if e.g., non-platform processes are needed.¹¹ High-concentration formulations cannot be administered via SC injection if the solution viscosity is too high, solubility is low, or if the solution is prone to aggregation or phase separation.^{47,55,90} Aggregation is of

particular concern because it is one of the most common degradation pathways for proteins and the potential mechanisms of aggregation are diverse and difficult to predict (see also section 1.3.3).

Self-interactions, reversible self-association, and aggregation are all dependent on solution conditions. pH and ionic strength impact charge states of ionizable residues and Debye screening, respectively, which both mediate electrostatic self-interactions. Many excipients are screened to mitigate specific degradation routes or physiochemical properties e.g., to stabilize against oxidation or reduce viscosity. Protein concentration is another important factor that can exacerbate many instabilities and poor biophysical properties, often in a manner not easily extrapolatable from low-concentration studies.^{21,40,45,53,156} The impact of solution conditions can be different between proteins, so a rational screening process with relevant characterization experiments and representative forced degradation studies are needed to determine suitable conditions.^{91,144} The effects of specific formulation components are often in conflict with each other, e.g., the ideal pH is often different for minimizing different degradation pathways like aggregation, fragmentation, and oxidation.^{25,89}

Ultimately, efficient determination of suitable solution conditions is vital for successful drug development. Designing studies to identify a stable formulation with the fewest number of experiments and/or in the shortest amount of time is of great interest to pharmaceutical companies.

1.6 Therapeutic proteins of interest

This thesis uses computational and experimental methods to study two classes of therapeutic proteins: monoclonal antibodies and Fc-fusion proteins. They are two of the most common kinds of proteins used in protein-based therapeutics, as mentioned

in section 1.1. The following subsections provide general background information on the two protein classes, and more specific information is described as needed in the pertinent chapters.

1.6.1 Monoclonal antibodies

MAbs are naturally-occurring Y-shaped immunoglobulin G (IgG) glycoproteins (~150 kDa) that bind to a specific antigen. In the last several decades, they have become the leading class of therapeutic proteins.^{14,115} MAbs are applied in the treatment of immune-mediated disorders, many different forms of cancer, and more recently infectious diseases.⁶⁻⁸ Methods to develop and manufacture MAb drug products are fairly well-established, although there are some differences between pharmaceutical companies and sometimes specific to a particular MAb or dosage form.^{3,14,157}

MAbs are comprised of three fragments composed of four domains each: one fragment crystallizable (Fc) fragment connected to two antigen-binding (Fab) fragments via a “hinge” region. Standard notation for each domain is used throughout this thesis: the Fc fragment contains two C_H3 and two C_H2 domains, while each Fab fragment contains one C_H1, one C_L, one V_H, and one V_L domain (see also Figure 1.4 below).¹⁵⁸ The Fc domain is stabilized by glycans attached to the C_H2 domains.¹⁵⁹ Both Fab fragments contain complementary defining regions (CDRs) within the two variable domains (V_L and V_H) that vary between different MAbs and bind to a specific epitope.¹⁵⁸ The large majority of the amino acid sequence besides these CDR regions is conserved within a given MAb subclass.¹⁵⁸ MAbs of the IgG1 and IgG4 subclass are used in this work, which differ in their biophysical properties in large part due to

differences in their hinge regions.¹⁶⁰ The IgG1 subclass is the most common in commercial drug products, but IgG4 and IgG2 MAbs are also in use.¹¹⁵

In Chapter 2, previously published data for 5 MAbs are used as test cases and are referred to with the same nomenclature as in prior work: MAb 1^{43,49,50}, MAb 2⁴⁹, MAb A^{42,69}, MAb B^{42,69} and MAb C^{42,69}. Data for MAb 1 is also referenced in Chapter 4 and Chapter 5, and data for MAb 2 is also referenced in Chapter 4. In Chapter 4 and Chapter 5, four MAbs that have not been reported on previously are introduced, referred to as MAb 3, MAb 4, MAb 5, and MAb 6. Those four MAbs were provided by Bristol Myers-Squibb (New Brunswick, NJ). MAb A, MAb C, MAb 1, MAb 3, MAb 5, and MAb 6 are of the IgG1 subclass, while MAb B, MAb 2, and MAb 4 are of the IgG4 subclass. MAb 1 is also the protein from prior work shown in Figure 1.3.

1.6.2 Fc-fusion proteins

Fusion proteins are a growing class of biopharmaceuticals where the sequences of two or more proteins or peptides are genetically combined and expressed recombinantly to yield a single multi-domain protein molecule with fusion partners (FPs) covalently bonded by a peptide linker. Often, one domain of the fusion protein performs a desired biological function, while the other improves other pharmacological properties such as physical stability, half-life, or solubility.¹⁶¹ Fc-fusion proteins, where the Fc domain of a MAb is attached to a protein or peptide via a peptide linker, are of particular interest due to their favorable pharmacokinetic properties, ability to solubilize hydrophobic proteins, and adaptability for use with established processing methods for MAbs.¹⁶¹⁻¹⁶⁴ Since the first approval of etanercept by the US Food and Drug Administration (FDA) in 1998, several Fc-fusion proteins have been developed and approved by the FDA for use in treating cancer, autoimmune

disorders, and a variety of other disease targets.^{162,165} As of November 2023, there are approximately 44 ongoing clinical trials with Fc-fusion proteins, according to clinicaltrials.gov. Several different classes of FPs have been used in Fc-fusion proteins such as extracellular domains¹⁶⁶, ligands¹⁶¹, peptides¹⁶⁷, cytokine traps, and enzymes.¹⁶⁸

Two Fc-fusion proteins (monovalent and bivalent) and their corresponding FP protein are studied in Chapter 3 and were provided by Merck & Co., Inc. (Rahway, NJ). Schematic diagrams of the two Fc-fusion proteins, the FP, and a representative MAb are shown in Figure 1.4. Linkers that connect the Fc to the FP in the Fc-fusion proteins in this work were flexible and are represented with curved lines, while MAb hinge regions are generally more rigid (although still somewhat flexible) and are represented with straight lines. MAb hinge regions are also linked by disulfide bonds that vary by subclass, which are shown as dashed lines.

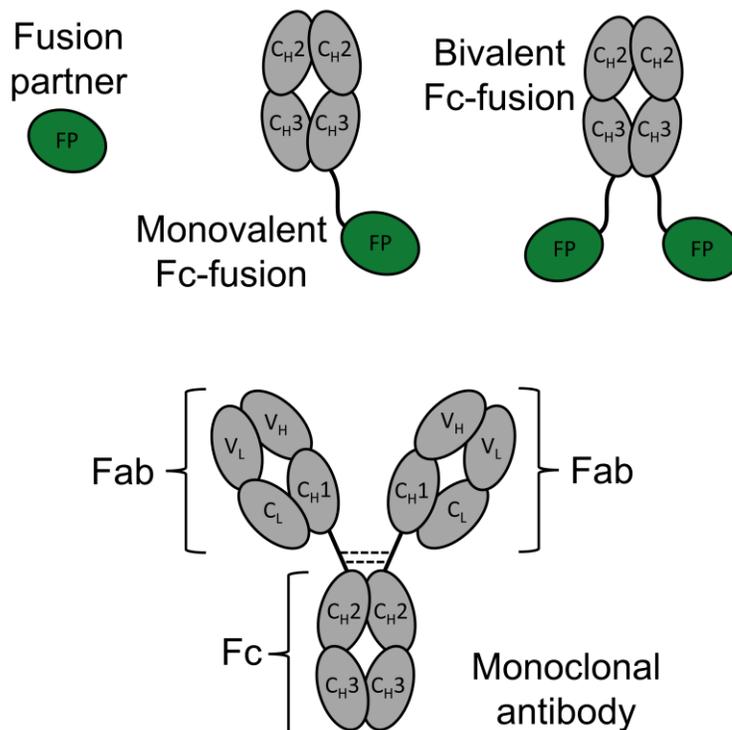


Figure 1.4: Schematic diagrams of the proteins of interest in this thesis, roughly to scale. The hinge region of the MAb and the linkers for the Fc-fusion proteins are shown for illustrative purposes.

1.7 Dissertation outline

Chapter 1 has provided motivation and background in terms of the relevant biophysical phenomena and resulting challenges in industrial drug development of therapeutic proteins. The objectives of this thesis are described in section 1.4. The remainder of the thesis is organized as follows.

Chapter 2 considers a suite of CG models for therapeutic proteins with emphasis on the prediction of net self-interactions for MAbs, with a test dataset of previously published values of the second osmotic virial coefficient (B_{22}) as a function of pH and ionic strength for five MAbs. The net self-interactions range from strong

attractions to strong repulsions. The CG models are compared based on the tradeoff between accuracy in their representation of self-interactions and their relative computational burdens, including when scaled to higher concentrations. Guidance is provided to inform CG model selection for molecular simulations of therapeutic proteins based on the resolution needed to capture the underlying interactions that make up self-interactions, from “simple-colloidal” electrostatic repulsions to specific attractive interactions between key amino acids.

Chapter 3 systematically characterizes net self-interactions and reversible oligomerization for a bivalent Fc-fusion, monovalent Fc-fusion, and the corresponding FP protein as a function of pH (5 and 6.5) and ionic strength (10 mM to at least 300 mM) experimentally via static and dynamic light scattering. The experimental results suggest that electrostatically mediated net attractive self-interactions are related to the formation of reversible oligomers at pH 6.5 conditions. CG molecular simulations are applied to model the self-interactions and investigate specific interactions between pairs of charged residues involved in attractive electrostatic self-interactions. Analysis of the strength of the specific electrostatic interactions suggests that cross-domain interactions between the FP and Fc domains are the source of the attractive electrostatic self-interactions and reversible self-association at pH 6.5 conditions.

Chapter 4 improves upon previously developed methodologies to combine low-concentration experimental measurements of net self-interactions (i.e., values of B_{22} or G_{22}) with CG molecular simulations to predict high-concentration net self-interactions. The predictions are improved by the application of a higher resolution CG model, 1-bead-per-charge-and-domain (1bC/D), which explicitly represents each charged site at increased resolution compared to the prior work. Low- to high-

concentration SLS data for six MAbs that represent a broad range of self-interactions are used as test cases and high-concentration predictions are compared between the 1bC/D model and a domain-level model. Challenges that stem from the representation of charge equilibria of ionizable residues are addressed by considering different approaches to determine the valence of ionizable residues and by explicit modeling of charge equilibria.

Chapter 5 expands the dataset of low- to high-concentration measurements of net self-interactions of four of the MAbs in Chapter 4 to also include conformational stability via DSC and aggregation rates via a broad range of stability studies. Stability studies were performed at two pH values (5 and 6.5), two ionic strengths (~10 and ~110 mM), three MAb concentrations (10, 35, and 130 mg/mL), and three incubation temperatures (4 °C, 30 °C, and 45-50 °C), representative of commercial MAb products and industrial studies during the candidate selection and formulation development phases. The studies were designed to systematically evaluate the impact of fundamental formulation features at multiple incubation temperatures on aggregation rates, with emphasis on how the rates were related to self-interactions and conformational stability. Accelerated studies (i.e., those at higher incubation temperatures) are shown to be poor indicators of quantitative or even qualitative aggregation rates at refrigerated (4 °C) conditions. Interpretable ML methods are developed that parse and quantify the phenomena relevant to high-concentration aggregation rates. The ML methods also provide a robust platform for predicting long-term aggregation rates for MAbs stored at typical refrigerated conditions.

Chapter 6 provides a summary of the individual conclusions from each chapter and how they are related to the broader goals of the thesis. Suggestions for future work are also provided based on the results and/or gaps in the thesis.

Chapter 2

COARSE-GRAINED MODELS FOR MOLECULAR SIMULATIONS OF MONOCLONAL ANTIBODIES AND THERAPEUTIC PROTEINS

2.1 Introduction

As discussed in Chapter 1, self-interactions are connected to many biophysical challenges in the development of therapeutic proteins, such as poor solubility, high viscosity, reversible self-association, and irreversible aggregation, which can negatively impact the developability of a therapeutic protein candidate during the drug discovery, purification and formulation stages.^{16,23,54,90,94,169,170} Experimental screening of self-interactions is often time-consuming and can require prohibitive amounts of protein material, which has prompted the development of computational tools such as CG molecular simulations to predict these phenomena, as discussed in section 1.2.1.

Net self-interactions are used as a relevant test case in this chapter, as they are readily measured experimentally^{41,121,171–174} and can be quantified through molecular simulation.^{28,63,70,71} Self-interactions are quantified in this chapter via osmotic virial coefficients.^{42,49,50} A given CG model should capture factors that are relevant to the net self-interactions, such as the solution pH (charge state of ionizable residues), ionic strength (Debye length for screened electrostatic interactions), co-solute concentrations (e.g., as it relates to preferential interactions), hydrophobicity, amino acid sequence and/or folded structure (location and identity of amino acids).^{28,70,140,175}

The second osmotic virial coefficient, B_{22} , can be used to quantify net self-interactions at low protein concentrations and is defined formally as

$$B_{22} = -\frac{1}{2} \int_{\mathbf{r}} \int_{\boldsymbol{\Omega}_1} \int_{\boldsymbol{\Omega}_2} [e^{-[W_{22}^{\infty}(\mathbf{r}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)/k_B T]} - 1] d\mathbf{r} d\boldsymbol{\Omega}_1 d\boldsymbol{\Omega}_2 \quad 2.1$$

where $W_{22}^{\infty}(\mathbf{r}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$ is the potential of mean force (i.e., including solvent-averaged interactions) between two proteins as a function of the center-to-center distance vector (\mathbf{r}) and relative orientation vectors ($\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2$), in the limit of protein concentration (c_2) approaching zero.¹⁷⁶ k_B is Boltzmann's constant, and T is absolute temperature. A value of B_{22} above (below) zero denotes net repulsions (attractions). The concentration-dependent analog of B_{22} is the Kirkwood-Buff protein-protein integral, G_{22} , which can also be used at higher concentrations (i.e., for multi-body self-interactions) and is related to B_{22} at low concentrations via $B_{22} = -\frac{1}{2} \lim_{c_2 \rightarrow 0} G_{22}$.¹⁷⁶

When protein concentration is increased, as is the case for MAb formulations for subcutaneous injection, proteins become more crowded and multibody interactions can influence self-interactions. In this case, self-interactions can be quantified by using higher-order terms in the virial expansion, shown in Equation 2.2. In Equation 2.2, Π_2 is the solution osmotic pressure, M_W is protein molecular weight, R is the gas constant and T is absolute temperature. A_i denotes the i th virial coefficient, and represents the net self-interactions between i protein molecules in solution ($A_2 = B_{22}$).^{49,177,178}

$$\frac{\Pi_2 M_W}{RT} = c_2 + B_{22} c_2^2 + A_3 c_2^3 + A_4 c_2^4 + A_5 c_2^5 + \dots \quad 2.2$$

MAbs are used as a test case in this chapter, which are discussed in more detail in section 1.6.1. Figure 2.1 schematically illustrates a series of representative coarse-grained models for MAbs, with different levels of coarse graining, and compared (roughly to scale) to an all-atom representation. Each of the CG models in Figure 2.1 was used in this chapter to predict a range of low-concentration net self-interactions compared to published experimental measurements (e.g., B_{22} values) as a function of pH and total ionic strength (TIS) for range of MAbs. The full sets of experimental data

are shown in Figure A.1 (Appendix A), with illustrative examples included below. The HEXA (6 beads) and DODECA (12 beads) models are from prior work^{28,50} and aim to approximate MAb excluded volume and broad charge anisotropy while maintaining computational efficiency. The one-bead-per-amino-acid (1bAA) model (~1300 beads for a typical MAb) was also developed in prior work^{70,71}, and provides better resolution of where each amino acid lies in the overall structure of the protein. However, simulating concentrated systems (e.g., $\sim 10^2$ – 10^3 proteins in the same simulation) is computationally impractical using the 1bAA model.^{49,50} An intermediate model is also considered here, termed one-bead-per-charged-site-and-per-domain (1bC/D). It uses domain level resolution for steric repulsions and short-ranged non-electrostatic attractions, and close to atomic-level detail for electrostatic interactions. The 1bC/D model contains much fewer CG sites relative to higher resolution models such as 1bAA, with typically ~ 5 -10 fold reduction depending on the number of charged sites in the MAb sequence. This allows for more efficient simulations and access to high-concentration conditions for protein systems that are strongly influenced by electrostatic interactions. While these models were applied to MAbs in this chapter, they or their analogues can be easily adapted for use with other proteins.

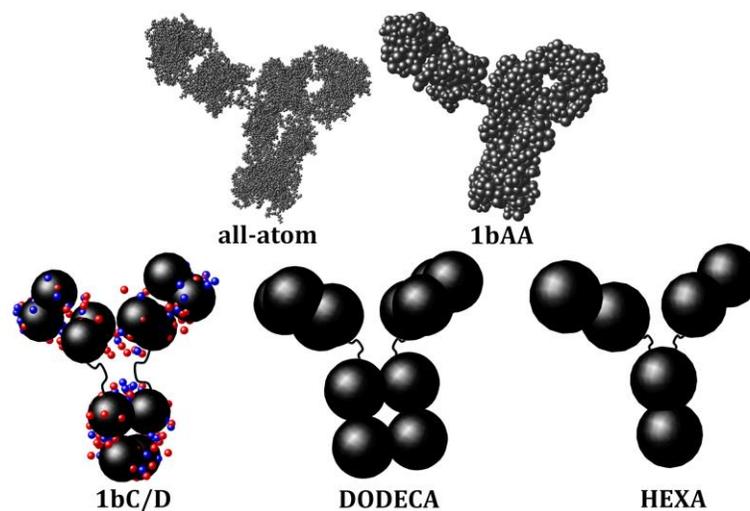


Figure 2.1: Schematic structures of CG models for monoclonal antibodies shown roughly to scale. Domain beads for 1bC/D are deliberately shown as slightly smaller to show the location of explicit charged sites for positive (red) and negative (blue) charges. Linkers shown between the Fc and Fab for HEXA, DODECA, and 1bC/D are putative and not represented with explicit beads in those models. Portions of this figure are reproduced from Calero-Rubio et al.²⁸

This chapter assesses a series of related CG models at different levels of coarse graining for molecular simulation of MABs using low-concentration self-interactions as the main test cases. The models were characterized based on their quantitative accuracy, relative speed on the same processor, and scalability for use in multibody simulations for higher protein concentrations (see also Chapter 4). Prior work has compared some of these models to experimental self-interaction parameters (specifically B_{22}) based on SLS over a range of pH values and TIS values, for MABs referred to here with the same nomenclature as in prior work: MAb 1^{49,50}, MAb 2⁴⁹, MAb A^{42,69}, MAb B^{42,69} and MAb C^{42,69}. The experimental net self-interactions varied from strongly repulsive to strongly attractive over the solution conditions and MAB

identities. The model comparisons provide data to inform a general approach to representing the fundamental interactions needed for CG models of therapeutic proteins (i.e., steric, electrostatic and short-ranged non-electrostatic interactions) based on available computational resources and the simulated quantity of interest.

Additionally, the intermediate model 1bC/D is an appealing option for systems that require site-level resolution for electrostatics but suffer from high computational burden especially at higher concentrations. Overall, this chapter provides systematic model comparisons that can be used to guide the selection or development of an appropriate CG model for molecular simulation of proteins and inform on general questions of what level of CG modeling is needed to capture the key attributes of self-interactions that span from colloid-like behavior to highly specific interactions between key residues.

The content in this chapter was published in a peer-reviewed journal and is reproduced here with permission (see Appendix E).¹⁷⁹ Hassan Shahfar was a co-first author in that article; he was responsible for the original design of the 1bC/D model as described in this chapter and performed a portion of the simulations presented in this chapter, which are included here for completeness.

2.2 Methods

2.2.1 Monte Carlo simulations

The Mayer sampling method with the overlap sampling algorithm (MSOS) was used to calculate B_{22} or higher order virial coefficients (i.e., $A_{i=3,4,5}$) for a given set of experimental conditions (MAb identity, pH value and TIS value). MSOS is an umbrella sampling method that biases the MC simulation towards configurations that

have a greater influence on the final value of a given virial coefficient. Virial coefficients are integrals where the integrands consist of one or more Mayer functions between protein molecules a and b , defined by $f_{ab} = (e^{-W_{ab}(r,\Omega)/k_B T} - 1)$. MSOS is a free-energy perturbation method that calculates integrals such as Equation 2.1 with respect to a reference state. In the present case, the reference state was the contribution to the i th virial coefficient from steric-only behavior of the given CG model, termed $B_{22,ST}$ or $A_{i,ST}$ (for $i > 2$) for higher order virial coefficients.^{50,177,178,180,181} Therefore, $B_{22}/B_{22,ST}$ or $A_i/A_{i,ST}$ values above (below) 1 indicate net repulsive (attractive) self-interactions with respect to steric-only behavior.

To compare experimental and simulated B_{22} values across different models, experimental data were normalized by $B_{22,ST}$ as calculated for each MAb in prior work.^{42,49,50} MSOS simulations were performed in an infinite volume at 25 °C with the number of CG MAb molecules that corresponds to a given virial coefficient (e.g., two molecules for B_{22} , three molecules for A_3 , etc.). One MAb molecule is stationary at the origin while each new configuration is created by moving a different MAb molecule(s) relative to that origin. Each molecule is treated as a rigid body. Flexibility, particularly in the hinge region of MAbs, has been shown to play an important role for simulations with increased protein concentration, but is the subject of future work due to increased computational burden for high resolution models such as 1bC/D and 1bAA.^{28,76} New configurations were generated via translations and rotations for a given molecule based on its geometric center. An equilibration period was included that used 1% of the total configurations in the simulation to refine the maximum translational displacement and rotation angle so that the acceptance ratio was roughly 50% for the remainder of the simulation.^{28,71} For the HEXA and DODECA models,

$\sim 10^7$ MC configurations were generated for B_{22} simulations. For higher resolution models, the number of MC configurations for B_{22} simulations ranged from $\sim 10^6$ for net-repulsive systems to upwards of $\sim 10^7$ for some net attractive conditions, as more configurations were needed to converge B_{22} in conditions that yielded more strongly net protein-protein attractions.⁴² For higher-order virial coefficients, $\sim 10^7$ - 10^8 configurations were needed for convergence.

$B_{22}/B_{22,ST}$ was calculated with each model for all experimental conditions tested in this work. Higher order (third, fourth and fifth) virial coefficients were calculated for MAb B at pH 5 for a range of ionic strengths as a test case for the applicability of the 1bC/D model for high concentration conditions to compare with recent results.⁴⁹ Additionally, second and higher order steric virial coefficients (e.g., $B_{22,ST}$ and $A_{i,ST}$) were calculated for each model, including all-atom, to provide reference states and consider differences in packing in future work that will compare against high-concentration experimental results. For calculating the steric-only virial coefficients with the different levels of structural resolution, the reference state in MSOS was chosen to be a single hard sphere with a diameter of 6 nm, thus each simulation returns $A_{i,ST}/A_{i,HS}$. $A_{i,HS}$ is the virial coefficient for a system of i hard spheres which has been solved analytically.¹⁸² All-atom simulations were performed with the Cornell et al. model with implicit solvent and 10^7 configurations were generated.¹⁸³ Approximately 50 percent of the atoms in a MAb have no solvent accessible surface area (SASA) and were excluded from the all-atom simulations as they do not contribute to excluded volume effects between MAbs. For the fourth and fifth virial coefficients calculated for the all-atom model, atoms with SASA values below 5 \AA^2 were removed, reducing the total number of atoms from ~ 5000 to ~ 3000 .

Based on results for $B_{22,ST}$ shown in Figure A.2 (Appendix A), this produced an estimated error of 4 percent. Additionally, the relative speed of each CG model was compared using the MSOS algorithm to calculate each virial coefficient up to the fifth coefficient. For benchmarking CPU times, simulations were run on a single thread using an Intel Xeon E5-2680 v3 processor.

2.2.2 CG models and interaction potentials

Interaction potentials for the four models (HEXA, DODECA, 1bC/D and 1bAA) treat the solvent implicitly and consider only the sum of pairwise inter-protein interactions between beads. The potentials were parameterized by experimental data as described below. All four models use a similar mathematical form for the steric and electrostatic interaction potentials, but with different parameter values for the latter. Steric interactions were modeled by a hard sphere potential for all beads, shown in Equation 2.3, where $u_{ST}(r_{ij})$ is the steric potential between beads i and j , r_{ij} is the center-to-center distance between them, and σ_{ij} is the average diameter of the pair of beads, i.e., $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ where σ_i and σ_j are the diameters of beads i and j , respectively.^{28,71}

$$u_{ST}(r_{ij}) = \begin{cases} \infty, & r_{ij} \leq \sigma_{ij} \\ 0, & r_{ij} > \sigma_{ij} \end{cases} \quad 2.3$$

Electrostatic interactions were modeled with a modified screened-Coulomb potential via

$$\frac{u_{EL}(r_{ij})}{k_B T} = \zeta \cdot (\psi_i q_i) \cdot (\psi_j q_j) \frac{e^{-\kappa(r_{ij} - \sigma_{ij})}}{r_{ij} [1 + \frac{1}{2}(\kappa \sigma_{ij})]^2} \quad 2.4$$

where $u_{EL}(r_{ij})$ is the effective electrostatic potential between beads i and j , q_i is the valence assigned to bead i , and κ^{-1} is the Debye screening length based on the TIS of

the solution. ζ is the Bjerrum length, $(4\pi\epsilon\epsilon_0k_B T)^{-1}$, where ϵ is relative permittivity, ϵ_0 is vacuum permittivity, T is absolute temperature, and k_B is Boltzmann's constant. The underlying mean-field theory for this potential is most appropriate above very low ionic strength.^{184,185} In this chapter, it was restricted to TIS above 10 mM, and comparison to experimental data was focused on those conditions for a given pH and MAb. The net charge or valence assigned to bead i (denoted q_i) is scaled by an adjustable parameter ψ_i to better match its effective charge in solution, which can deviate from the theoretical value due to territorial ion binding or by the approximations noted in the Introduction regarding choice of charge location for CG models. That is, $\psi_i q_i$ is the effective charge in solution and $\psi_i = \frac{q_{i, effective}}{q_i}$.⁴² For simplicity, the value of ψ_i was assumed to be independent of the bead location or chemical identity for all beads in a given CG model and therefore a single ψ parameter was used in each model, although the numerical values of ψ were different for a given model.^{50,68} The terms valence and charge are used interchangeably in the remainder of this chapter.

For the HEXA and DODECA models, domain charges (q_i) are shown in Appendix A (see Table A.1 and Table A.2), and were calculated by applying the Henderson-Hasselbalch equation with nominal pK_a values to the amino acids combined or "lumped" into a given bead (domain).²⁸ The use of nominal pK_a values can impact the ψ parameter as pK_a values of residues are known to be influenced by interactions with other residues in the protein, therefore valences assigned to residues may be only approximate, leading to deviations from the effective charge in solution if the local chemical environment leads to shifted pK_a values (see also Chapter 4).^{175,186} For pH values in this chapter (5, 6.5 and 8), all 1bC/D or 1bAA beads representing

aspartic and glutamic acid residues were treated as having -1 valence, those for lysine and arginine have +1 valence, and histidine have +1 valence at pH 5 or no net charge at pH 6.5 and 8.⁷¹

The interaction potential for short-ranged non-electrostatic attractions such as those due to a combination of van der Waals and hydrophobic interactions is different for the domain-level models (HEXA, DODECA and 1bC/D) and the 1bAA model. For the domain-level models, the interaction potential, $u_{SR}(r_{ij})$ between bead i on a given protein and bead j on its neighboring protein, is shown in Equation 2.5, where n is 10 for HEXA and n is 6 for DODECA and 1bC/D, chosen so that the effective range of decay of attractive interactions is approximately 1 nm.⁵⁰ ϵ_{SR} is an adjustable parameter that represents the maximum strength (well depth) of short-ranged non-electrostatic attractions and c normalizes the potential such that the minimum energy is $-\epsilon_{SR}$ in units of $k_B T$.^{28,49,50}

$$\frac{u_{SR}(r_{ij})}{k_B T} = \frac{\epsilon_{SR}}{k_B T} c \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{128} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^n \right] \quad 2.5$$

Short-ranged non-electrostatic attractions for 1bAA are modeled as a continuous piecewise function^{71,72}

$$u_{SR}(r_{ij}) = \begin{cases} \infty & r_{ij} < \sigma_{ij} \\ 4\epsilon_{SR} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + (\epsilon_{SR} - \epsilon_{ij}\epsilon_{SR}) & \sigma_{ij} \leq r_{ij} \leq r_c \\ 4\epsilon_{SR}\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] & r_{ij} > r_c \end{cases} \quad 2.6$$

where $r_c = 2^{1/6}\sigma_{ij}$ and ϵ_{ij} is the geometric mean of the relative hydrophobic values of the two interacting amino acids, $\sqrt{\epsilon_i\epsilon_j}$.^{42,71,72} At $\sigma_{ij} > r_c$, short-ranged non-electrostatic interactions transition from being purely repulsive to attractive. Repulsive interactions are not specific to the interacting amino acids while attractive interactions

depend on $\sqrt{\varepsilon_i \varepsilon_j}$. Relative hydrophobicity values (ε_i) are normalized such that the scale is from 0 to 1, where the most hydrophobic residue (leucine) has $\varepsilon_{leucine} = 1$. The scale was derived by Bereau and Deserno based on analysis of crystallized protein residue-residue contacts based on the work of Miyazawa and Jernigan.^{72,187}

The 1bAA model places a bead at the geometric center of each amino acid, where each amino acid has its own diameter (σ_i), relative hydrophobicity (ε_i), and charge (q_i).^{42,71,72} Amino acid locations were determined from homology models or other sources of PDB files for each MAb. Bead diameters (σ_i) for the i th residue are based on the amino acid chemical identity, were determined previously, and are listed in Table A.3 (Appendix A).^{71,188} Each of the beads in the DODECA model corresponds to one MAb domain, where one bead comprises approximately 100 amino acids. The heavy and light chains are split into four and two approximately equal length segments, respectively, such that each bead is comprised of a similar number of amino acids. The HEXA model combines pairs of DODECA beads so that each of the six beads represents approximately 200 amino acids. HEXA beads include one bead for each of the F_v domains (each $V_L + V_H$), and one for each C1 domain (each $C_L + C_{H1}$), along with one bead for the C2 domain (combined C_{H2} from each heavy chain) and one for C3 (combined C_{H3} from each heavy chain). The default set of residue indices for DODECA domains are shown in Table A.4 (Appendix A), although they could be adjusted based on user preference and/or algorithm for delineating the domains. The models are geometrically constrained by distances between centers of the domains or regions (e.g., Fc and Fab) as shown in Figure A.3 (Appendix A). These distances were chosen to resemble existing crystal structures (IGGY, 1IGT, and from Padlan et al.) and are treated as constant regardless of MAb identity.^{28,50,189,190} Bead

diameters (σ_i) are 3.5 nm for each DODECA domain and 4.4 nm for each HEXA domain, and were chosen to match $B_{22,ST}$ of a model MAb as calculated by all-atom MSOS simulations.^{28,189}

Much like the DODECA model, the 1bC/D model introduced in the article this chapter is based on¹⁷⁹ uses one bead per domain with each bead having the same size. However, the domain beads are placed at the geometric center of each domain based on a given rigid protein 3-dimensional structure for that particular protein (e.g., homology model or x-ray structure) rather than using a single set of locations across all candidate MAbs that was the default for the HEXA and DODECA models. Domain bead locations for the 1bC/D model are specific to the given MAb homology model and the domain beads do not contribute to electrostatic interactions. The geometric center of a domain was calculated as the average of the coordinates from a given homology model of all non-hydrogen atoms in a domain, where domains consist of the same amino acid sequences as in the DODECA model (see Table A.4 in Appendix A). The domain-level beads in the 1bC/D model account for steric and short-ranged non-electrostatic interactions and use the same potential functions as DODECA, Equation 2.3 and Equation 2.5, with different parameter values.

The domain bead diameter for the 1bC/D model (same diameter for all domains) is MAb-specific and was chosen to match $B_{22,ST}$ of an all-atom model of the given MAb, as described in Figure A.5 (Appendix A). For the 1bC/D model, electrostatic interactions were modeled such that charged beads were placed at the location of each charged (united) atom. Bead diameters for charged sites were the sum of those from the Cornell et al. all-atom model ($\sigma_{c,i}$) with an added width of a putative hydration layer (Δ_H).¹⁸³ The value of Δ_H is the same for all charged sites in a given

simulation. The net diameter of a given charged bead was $\sigma_{c,i} + \Delta_H$. In the present work, the range of Δ_H was 0 to 3 Å, where 3 Å was used for the results shown in the main text below. This was chosen to capture a putative hydration layer of one water molecule, thus excluding very strongly interacting configurations that can occur when charged beads are unrealistically close and dehydrated, as the current model uses an implicit-solvent approximation. Results in this chapter are shown for Δ_H of 3 Å, although the qualitative and semi-quantitative conclusions hold for other values of Δ_H . Bead diameters ($\sigma_{c,i}$) for each charged site are shown in Table A.5 (Appendix A).

2.2.3 Average relative deviation error calculations

Model predictions of $B_{22}/B_{22,ST}$ are compared to experimental values using average relative deviation (ARD) values defined as follows,

$$ARD (\%) = \frac{100}{\eta} \sum_{i=1}^{\eta} \left| \frac{x_i^{experimental} - x_i^{predicted}}{x_i^{experimental}} \right| \quad 2.7$$

where η is the number of data points and x_i is the i th experimental or predicted value from a given data set.^{50,68} The experimental data can be found in Figure A.1 (Appendix A). For the purposes of this chapter, any differences between calculated (predicted) B_{22} values and the corresponding Kirkwood-Buff integral (G_{22}) values were neglected. That is only expected to be quantitatively significant for the experimental values for B_{22} for MAb C, and does not affect the conclusions below.¹⁹¹ ARD provides a quantitative measure of average error that is used to compare the prediction from each model for a given data set. ARD values when compared between data sets are not necessarily directly comparable as experimental values that are large in magnitude or close to zero will bias the resulting ARD values. For example, for MAb 2 at pH 5 and 107 mM TIS, $B_{22}/B_{22,ST}$ was measured to be -0.06 +/- 0.05. A

quantitatively accurate predicted $B_{22}/B_{22,ST}$ on the order of -0.1 (within experimental error) yields a relative deviation of 100% because the denominator is so close to zero, and that disproportionately biases the final ARD value. Note that this error definition is different than that for Chapters 3 and 4.

2.2.4 Adjustable model parameters

The short-ranged non-electrostatic and electrostatic interaction potentials each have an adjustable parameter that must be specified: ϵ_{SR} , the maximum strength (or well depth) of short-ranged non-electrostatic interactions for a given site/domain; ψ , an adjustment factor for the effective charge relative to the theoretical charge in solution for a given site/domain (see above). For the HEXA and DODECA models, ARD values were calculated from MSOS simulations of $B_{22}/B_{22,ST}$ vs. TIS for a range of ϵ_{SR} and ψ values. This generated surface plots (e.g., Figure A.4 in Appendix A) where the optimal range of values of both parameters were determined simultaneously. For the more computationally burdensome 1bC/D and 1bAA models, ϵ_{SR} was optimized for a given protein and pH value by matching the model predictions with $\psi = 0$ to the corresponding experimental value of $B_{22}/B_{22,ST}$ at the highest TIS value available, such that the Debye length is small and electrostatic interactions are essentially fully screened. With that fixed ϵ_{SR} value, ψ was sampled until an optimal value was found when comparing to the full experimental $B_{22}/B_{22,ST}$ vs. TIS profile.

2.3 Results and discussion

Illustrative experimental B_{22} values as a function of TIS are shown in Figure 2.2 for selected cases that show qualitatively different behavior. In Case (i), values of $B_{22}/B_{22,ST}$ are large and positive at low TIS and decline monotonically with

increasing ionic strength until a plateau near zero $B_{22}/B_{22,ST}$ at high TIS values. In Case (ii), values of $B_{22}/B_{22,ST}$ are positive or near zero at very low TIS, then decrease and become negative at intermediate TIS, and increase monotonically with increasing ionic strength until a plateau near zero $B_{22}/B_{22,ST}$ at high TIS values. In Case (iii), values of $B_{22}/B_{22,ST}$ are large and negative at low TIS and increase monotonically with increasing ionic strength until a plateau near zero $B_{22}/B_{22,ST}$ at high TIS values. Experimental data for all five antibodies as a function of pH and TIS is shown in Figure A.1 (Appendix A).

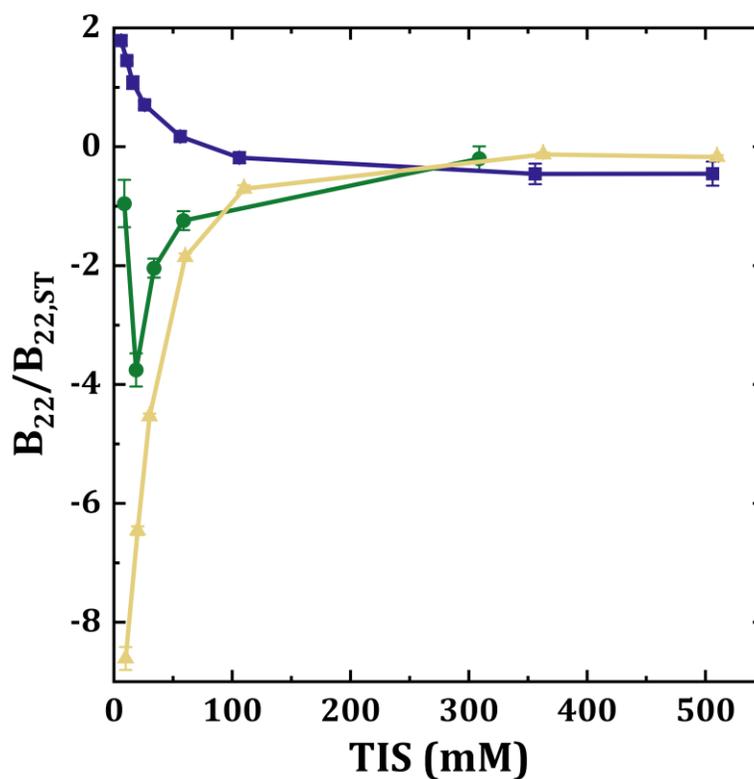


Figure 2.2: Experimental $B_{22}/B_{22,ST}$ values as a function of TIS for MAb 1 at pH 5 (indigo squares), MAb B at pH 5 (green circles), and MAb 2 at pH 6.5 (yellow triangles) to illustrate characteristic behavior for Cases (i), (ii), and (iii), respectively.^{42,49}

Case (i) is an example of classic colloid-like behavior that would be expected for proteins with high net surface charge leading to self-interactions that are more repulsive than steric-only interactions (i.e., $B_{22}/B_{22,ST}$ values greater than 1) due to long-ranged electrostatic repulsions at low ionic strength, while electrostatic interactions become screened at high ionic strength and $B_{22}/B_{22,ST}$ values plateau at values close to zero. Case (ii) is expected for proteins with both positive and negative charges but there are sufficiently attractive interactions between opposite charges on neighboring proteins that there are net attractions at low to intermediate ionic strength, but at very low ionic strength the Debye screening length becomes so large that the high net charge on the proteins causes the net interactions to resemble more of a classic colloidal model. Case (iii) is expected if the net charge on the protein is near zero and the distribution of charges on the surface is such that enough charges of similar value are “clustered” spatially to create charge “patches” that can lead to very strong attractive interactions between groups of oppositely charged amino acids, although Case (iii) behavior can also occur when there is significant net charge on the protein. For the CG models considered in this chapter and elsewhere^{28,50,63,71,73,192–194}, the underlying assumptions of a screened-Coulomb implicit solvent and implicit ion model are questionable at very low ionic strength for typical protein concentrations for measuring B_{22} (e.g., significantly below 10 mM TIS where the mean-field approximation likely breaks down).^{184,185} The comparisons between models and experimental data below will focus on TIS values above approximately 10 mM.

The interaction potentials for the CG models that were used require the specification of bead diameters (sets of σ_i values, i denoting different bead types that are summarized in Appendix A, see Table A3 and Table A6), the maximum strength

of short-ranged interactions (ϵ_{SR}) and the adjustment factor for effective charges (ψ). Sets of σ_i and ϵ_{SR} values will be discussed in what follows and optimizing ψ will be discussed in the next section. Bead diameters appear in all the interaction potentials and are the key parameter(s) that influences steric interactions once one sets the location of each bead. $B_{22,ST}$ accounts for the steric or excluded volume contributions to B_{22} . The HEXA and DODECA bead diameters were chosen as 4.4 nm and 3.5 nm respectively in prior work to match all-atom steric behavior ($B_{22,ST}$) of a model MAb, with all six (twelve) domain beads having the same diameter in the HEXA (DODECA) model.^{28,189} A similar procedure was used to determine the diameter of each of the twelve domain beads in the 1bC/D model, except that the value of $B_{22,ST}$ from an atomistic model for each MAb was used to match $B_{22,ST}$ for the 1bC/D model. The value of $B_{22,ST}$ was 11.4 mL/g for MAb A, 9.7 mL/g for MAb B, 12.0 mL/g for MAb C, 9.9 mL/g for MAb 1 and 10.6 mL/g for MAb 2.^{42,49,50} $B_{22,ST}$ was calculated for the 1bC/D model for each MAb as a function of domain bead diameter, shown in Figure A5 (Appendix A), with all charged sites represented explicitly and contributing to steric interactions. The domain bead diameters for the 1bC/D model to match with the $B_{22,ST}$ values from the all-atom calculations were 2.7 nm for MAb A, 3.15 nm for MAb B, 2.9 nm for MAb C, 3.1 nm for MAb 1 and 3.35 nm for MAb 2. Protein shape plays a role in excluded volume, for example MAb A and MAb C have larger values of $B_{22,ST}$ because the homology model structure is somewhat more extended in the hinge region than the homology models for the other antibodies considered here.

At high TIS (e.g., ~300 mM) the Debye screening length is small, causing electrostatic interactions to be heavily screened and solution behavior to be dominated by the influence of steric repulsions and short-ranged non-electrostatic attractions. As

steric interactions are already determined by the choice of bead diameters, the maximum strength of non-electrostatic short-ranged attractions (ϵ_{SR}) can be chosen by matching MSOS simulation results to experimental self-interaction measurements at high TIS. Figure 2.3 shows $B_{22}/B_{22,ST}$ as a function of ϵ_{SR} for the 1bAA model for each MAb under conditions where the electrostatic potential is not included (equivalently, with $\psi = 0$). Analogous results for the other CG models are shown in Figure A.6 (Appendix A). For the 1bC/D and 1bAA models, ϵ_{SR} is chosen to match the experimental $B_{22}/B_{22,ST}$ for each MAb at TIS greater than 300 mM (independent of pH). The HEXA and DODECA models have fixed structures regardless of the choice of MAb, so the curves in Figure A.6A are the same for any given MAb. Values of ϵ_{SR} for each MAb and choice of CG model and pH are listed in Table 2.1. At sufficiently high TIS, $B_{22}/B_{22,ST}$ values should not depend on pH. For some of the molecules compared here, sufficiently high TIS experimental data were available to show that expected behavior. However, for some of the molecules/pH conditions, the highest TIS values did show somewhat different values of $B_{22}/B_{22,ST}$, and therefore the ϵ_{SR} values were not necessarily the same within the available data. As HEXA has fewer beads, a higher value of ϵ_{SR} is needed to achieve the same magnitude of $B_{22}/B_{22,ST}$ at high TIS when compared to DODECA.

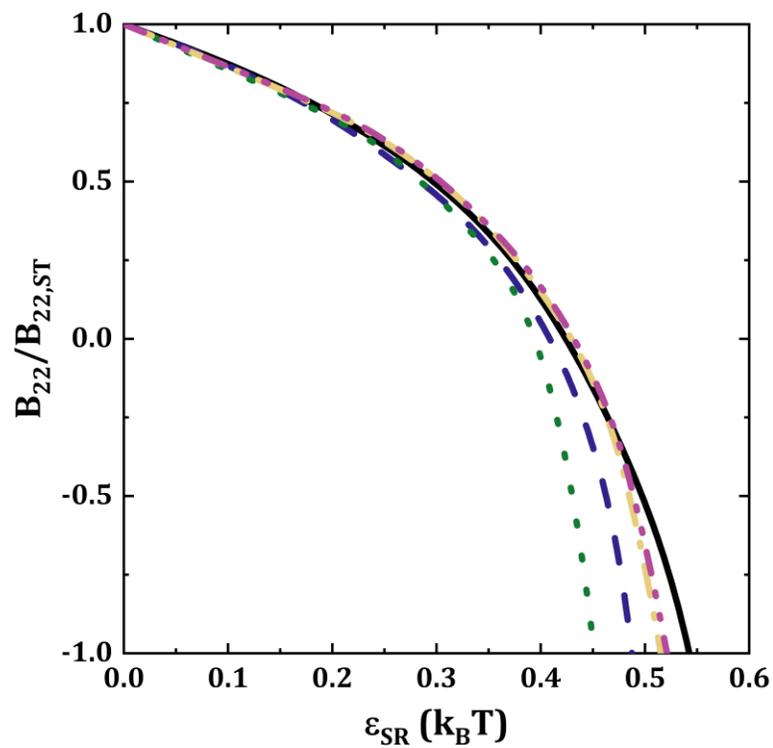


Figure 2.3: $B_{22}/B_{22,ST}$ as a function of ε_{SR} in the limit where electrostatic interactions are not included (i.e., $\psi = 0$) for the 1bAA model for MABs 1 (solid black), 2 (dashed indigo), A (dotted green), B (dash-dotted yellow), and C (dash-double dotted purple).

Table 2.1: Optimized values of ϵ_{SR} (units of $k_B T$) and ψ (dimensionless) for each MAb, pH and CG model and the resulting percent ARD values.^a

MAb (pH)	HEXA			DODECA			1bC/D			1bAA		
	ϵ_{SR}	ψ	ARD									
Case (i)												
MAb 1 (5)	1.40	0.42	25	0.72	0.68	22	1.07	0.9	32	0.50	0.67	16
MAb 1 (6.5)	1.40	0.80	16	0.68	1.04	14	1.1	0.80	517	0.46	0.95	280
MAb 2 (5)	1.30	0.44	18	0.64	0.60	12	0.75	0.98	38	0.44	0.75	24
MAb A (5)	1.24	0.29	35	0.58	0.33	38	2.46	0.66	48	0.42	0.60	75
MAb A (8)	1.32	0.23	19	0.66	0.30	14	2.54	0.6	13	0.42	0.20	25
Case (ii)												
MAb B (5)	1.50	2.60	31	0.70	3.35	40	0.84	1.24	6	0.45	1.05	17
MAb B (8)	1.10	0.56	33	0.55	1.03	32	0.50	1.3	7	0.31	1.20	29
Case (iii)												
MAb 2 (6.5)	1.30	2.90	35	0.63	3.54	38	0.82	1.58	16	0.44	1.40	39
MAb C (5)	1.55	0.00	72	0.80	0.00	72	1.90	1.48	20	0.52	1.35	21
MAb C (8)	1.35	2.20	8	0.73	2.33	21	1.68	1.62	19	0.52	1.65	37

^aARD values are calculated using the listed ϵ_{SR} and ψ values for all TIS values for a given experimental data set. Cells shaded green, gray, and yellow correspond to good, fair, and poor qualitative agreement with the experimental data set, respectively.

The full set of experimental $B_{22}/B_{22,ST}$ values as a function of pH and TIS are shown in Figure A.1 (Appendix A). As noted above, the ϵ_{SR} value for a given model was set to match the value of $B_{22}/B_{22,ST}$ at the highest TIS value for a given experimental data set. The optimal ψ values were selected based on minimizing the ARD between the simulated and experimental $B_{22}/B_{22,ST}$ values as a function of TIS, for a given choice of protein, pH, and CG model. Figure 2.4A shows an example of

Case (i) behavior and the comparison of the experimental data (symbols) with each optimized CG model (curves). Figure 2.4B shows an example of how ARD depends on ψ for each model with the experimental conditions in Figure 2.4A, illustrating how ψ values were optimized by selecting ψ to minimize ARD. Figure 2.5 shows examples of Case (ii) (Figure 2.5A) and Case (iii) (Figure 2.5B and 2.5C), analogous to Figure 2.4A for Case (i). The other examples for each choice of MAb and pH across all the CG models are provided in Figure A.7 (Appendix A).

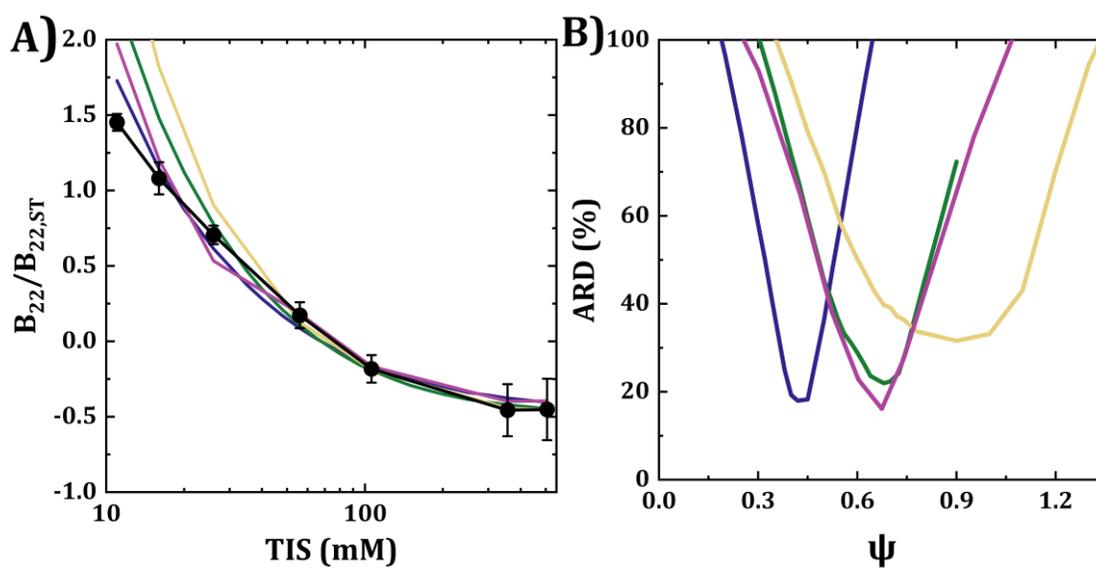


Figure 2.4: (A) Experimental $B_{22}/B_{22,ST}$ values (black symbols and curve as guide to the eye) and simulated $B_{22}/B_{22,ST}$ values using ψ values that minimized ARD for different models for MAb 1 at pH 5: HEXA (indigo curve), DODECA (green curve), 1bC/D (yellow curve) and 1bAA (purple curve); (B) ARD vs. ψ for the experimental data in panel A for HEXA (indigo curve), DODECA (green curve), 1bC/D (yellow curve) and 1bAA (purple curve), illustrating how the ψ value for a given model was selected for a given experimental data set.

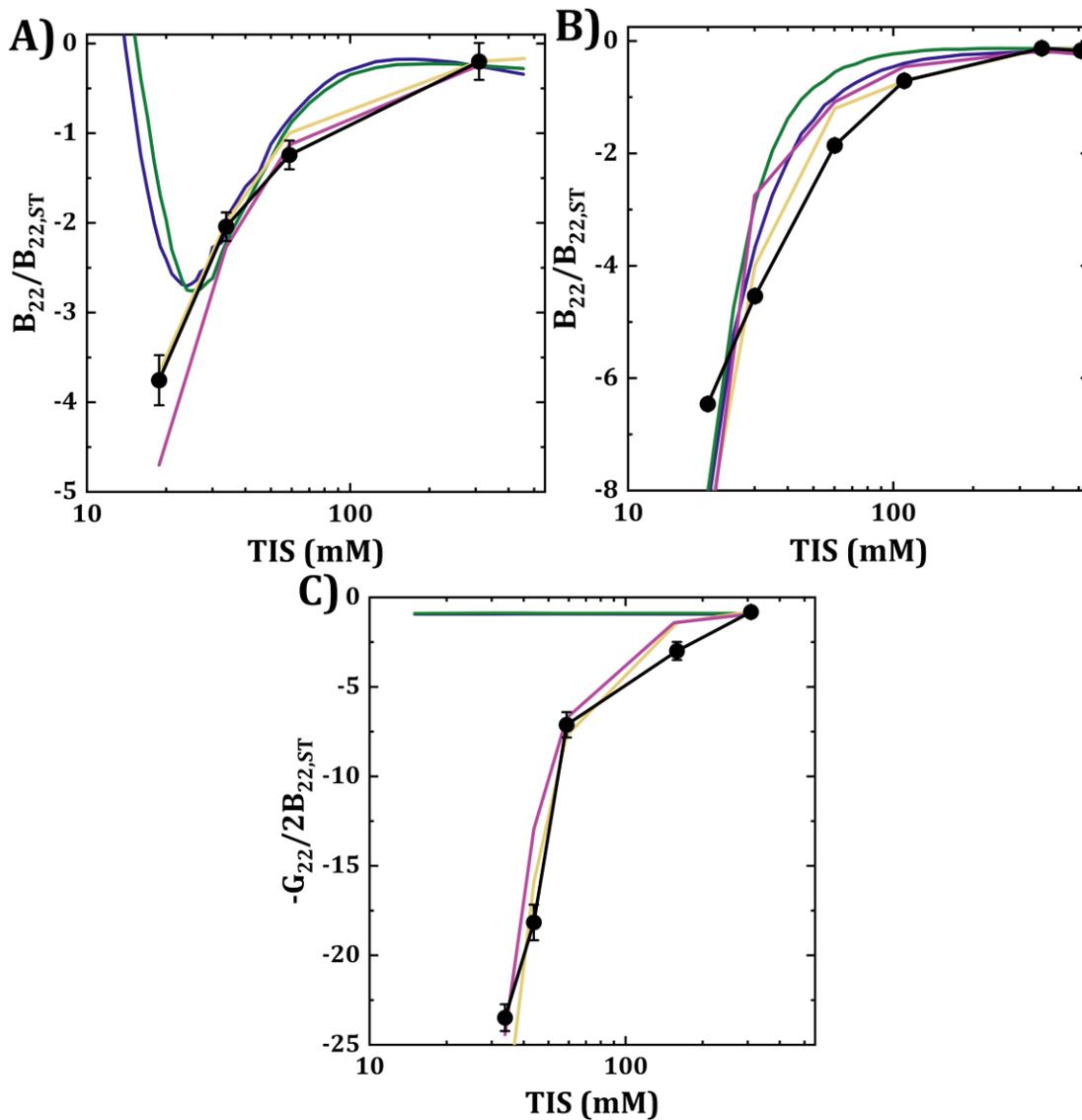


Figure 2.5: Comparison of experimental $B_{22}/B_{22,ST}$ values (black symbols and line) and simulated predictions with ψ values that minimized ARD for HEXA (indigo), DODECA (green), 1bC/D (yellow) and 1bAA (purple) models for: (A) MAb B at pH 5; (B) MAb 2 at pH 6.5; (C) MAb C at pH 5. Values in panel (C) are designated as $-G_{22}/2B_{22,ST}$ instead of $B_{22}/B_{22,ST}$ because some values are so large that they indicate multi-body interactions at the experimental protein concentrations, rather than two-body interactions that are captured by B_{22} . The corresponding ARD plots are shown in Figure A.7 (Appendix A).

Figure 2.5A shows that it is possible for all of the models to capture some or all of the experimental behavior over most of the TIS conditions, although there are inconsistencies at the lowest TIS values for the domain-level models (HEXA and DODECA). Figure 2.5B shows an example where all the CG models are able to capture the experimental behavior if the model parameters are sufficiently adjusted. Figure 2.5C shows a case where the lower resolution (domain-level) models cannot even qualitatively capture the experimental behavior.

Table 2.1 provides a summary for each CG model across all the experimental data sets (figures are provided for each data set and model in Figure A.7 in Appendix A). The entries are color coded based on whether the model is able to reasonably capture the experiment behavior. If one does not have experimental data to parameterize the CG models, inspection of Table 2.1 for the cases where the model(s) capture the experimental data suggests some starting ranges for the adjustable parameters. The ε_{SR} values for the HEXA, DODECA, and 1bAA models fall in a reasonably narrow range (units of $k_B T$): 1.2-1.5, 0.55-0.7, and 0.4-0.5, respectively. For the 1bC/D model, ε_{SR} values range more widely, in part because the domain bead diameter (and therefore the range of the short-ranged attractions) is adjusted on a case-by-case basis in the 1bC/D model to capture the steric-only B_{22} value for a given all-atom protein structure/homology model.

ψ values for Case (i) with the available experimental data were 0.2-0.5, 0.3-0.7, 0.6-1, and 0.6-0.75 for the HEXA, DODECA, 1bC/D, and 1bAA models, respectively. It can be seen that it is possible for each of the models to quantitatively or semi-quantitatively capture the experimental data for proteins and solution conditions that fall in the category of Case (i). This is not surprising, given that the

electrostatic contributions to the net self-interactions are predominantly repulsive, and therefore “lumping” charges into the domains does not cause significant attractive charge-charge interactions between oppositely charged amino acids to be missed in the domain-level CG potentials, compared to the higher resolution models.

Because the models were restricted to data above 10 mM TIS, the “upturn” in the value of $B_{22}/B_{22,ST}$ at low TIS for MAb B at pH 5 was not considered when comparing the models. That notwithstanding, there are cases where the different CG models were or were not able to capture the experimental profiles that were in the Case (ii) category. The lower-resolution, domain-level CG models tended to be less effective at the lowest TIS values. They overpredicted increases in repulsive interactions at the lowest TIS conditions, as illustrated in Figure 2.5A with the upturns in B_{22} values versus TIS that the models predicted at higher ionic strength than experiments showed. This is perhaps not surprising since those models “lump” positive and negative charges into single domains with net charges, and therefore the strength of charge-charge repulsions between the domains would be overestimated at low TIS values. This is due, in part, because of the longer charge screening lengths since the interaction length-scales scale with the bead diameter. It is also because the charge-charge repulsions scale as the square of the domain charge for those models instead of the sum over the squared amino-acid/amino-acid charge pairs between antibody molecules.

ψ values for Cases (ii) and (iii) with the available experimental data were 1-1.5 for the 1bAA model. For the examples of Case (iii) behavior it is sometimes possible for the lower-resolution CG models to provide a reasonable fit to the $B_{22}/B_{22,ST}$ vs. TIS profiles, but it requires one to allow the model parameter(s) to reach extreme

values (e.g., $\psi \sim 3.5$ for MAb 2 at pH 6.5, DODECA). In another case (e.g., MAb C at pH 5, Figure 2.5C), there were no combinations of model parameters for the domain-level models that could even qualitatively reproduce the experimental results. Based on the results and analysis from higher resolution models such as 1bAA and 1bC/D for that case⁶⁹, it is not surprising that domain-level models will not capture the strongly attractive interactions because those are due to a relatively small number oppositely charged amino acids that interact strongly between neighboring proteins. In addition, the domains in which those amino acids are located do not carry opposite charges to each other. As such, “lumped” domain-domain interactions will give qualitatively different behavior than interactions with explicit charges on the surface. This is a common issue of CG models where amino acids are grouped or lumped, whether by domains, sub-domains, or other rationales for grouping amino acids.^{28,62,73,106,193–198}

ψ values for the 1bC/D model are more nuanced, as they depend on the value of the putative hydration layer (Δ_H). The relationship between Δ_H and the profile of ARD as a function of ψ for each MAb and pH is illustrated in Figure A.8 (Appendix A). For Case (i) behavior, there were broad ranges of ψ that reproduced the experimental trends accurately, the putative hydration layer (Δ_H) did not play a significant role in the optimal value of ψ , and the results in Table 2.1 for Case (i) were not sensitive to Δ_H . This is reasonable, since Case (i) corresponds to net colloiddally repulsive interactions with primarily repulsions between charged sites, and close contacts between charged groups are less likely.

However, for Case (ii) and (iii) behaviors (i.e., those that are strongly influenced by attractive electrostatic interactions), there were narrow minima for the optimal values of ψ (denoted ψ_{min}) and there was a direct correlation between Δ_H and

ψ_{min} . The corresponding minimum ARD values and agreement between model and experimental results were similarly good for any of the Δ_H values, indicating that the model is still able to capture the experimental data. The ψ_{min} value decreased with decreasing Δ_H , as expected since the charged sites can approach more closely as Δ_H decreases. The results in Appendix A use the values of the domain diameter (σ_i) optimized using $\Delta_H = 3 \text{ \AA}$ and $\psi = 0$, and those σ_i values will change for different Δ_H values since the charged sites also contribute to the net $B_{22,ST}$ values. Fully optimizing Δ_H and ψ would therefore require a convoluted multi-parameter search of Δ_H , ψ , and of σ_i , and could also be expanded to re-parameterizing the values of $\sigma_{c,i}$ from Cornell et al.¹⁸³ That notwithstanding, these examples illustrate the balance of effects that need to be considered in designing CG models such as those portrayed in Figure 2.1.

2.3.1 Higher order virial coefficients and higher protein concentrations

This section illustrates extending the use of CG models to higher protein concentrations, as this becomes a major challenge for higher resolution models and is currently not tenable with all-atom models for most proteins of interest.^{28,49,50,68,76} This section will focus on higher-order virial coefficients that capture multi-protein interactions. As a starting point, steric interactions are fundamental for any CG model, as they determine the excluded volume of the protein and impact packing in concentrated systems.^{28,70,76} Virial coefficients for just the steric protein-protein interactions ($A_{i,ST}$) up to the fifth virial coefficient ($i=2,3,4,5$) for MAb B are shown in Figure 2.6 as an example for each CG model, normalized by the value calculated using the all-atom model. The structures for HEXA and DODECA are independent of MAb identity, so the results are general beyond MAb B. However, the bead sizes for the HEXA and DODECA models were not determined by the structure for MAb B, and

therefore it is not unexpected that the $B_{22,ST}$ is not the same as for the all-atom model of MAb B.^{28,189} The domain bead diameters in the 1bC/D model are MAb-specific and chosen to match $B_{22,ST}$, so the results match those for all-atom.

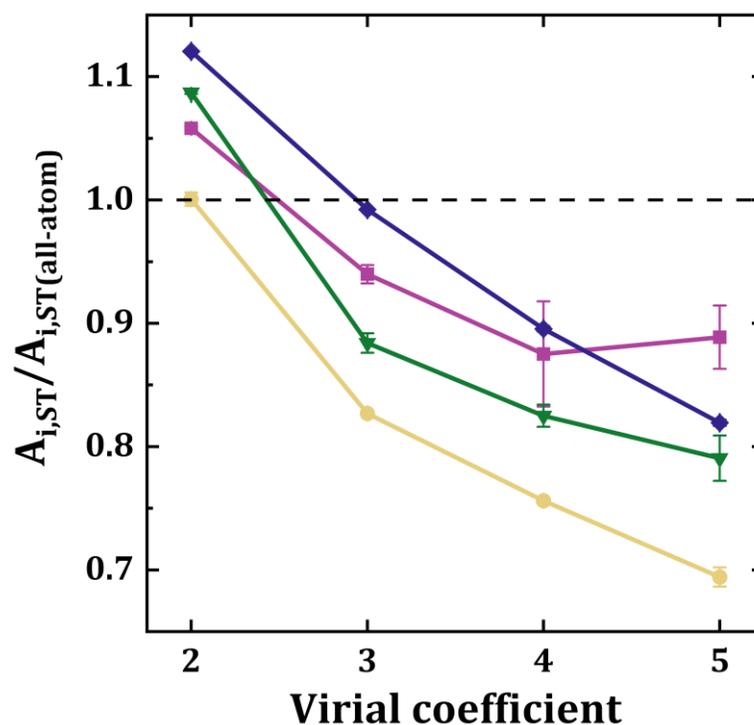


Figure 2.6: Illustrative steric-only virial coefficient values ($A_{i,ST}$, $i = 2$ to 5) for MAb B, normalized by the value of $A_{i,ST}$ for the all-atom model (for a given virial coefficient, $i=2,3,4,5$), for HEXA (indigo), DODECA (green), 1bC/D (yellow), and 1bAA (purple). Steric-only virial coefficients $i = 2$ to 5 for the DODECA model were reported in prior work.⁵⁰ The values of $A_{i,ST}$ for $i = 2$ to 5 for the all-atom model were $(9.46 \pm 0.07) \times 10^{-3}$ L/g, $(5.45 \pm 0.08) \times 10^{-5}$ (L/g)², $(2.06 \pm 0.02) \times 10^{-7}$ (L/g)³, and $(6.29 \pm 0.02) \times 10^{-9}$ (L/g)⁴, respectively. Error was calculated as the 95% confidence interval from three independent simulations. The terms $B_{22,ST}$ and $A_{2,ST}$ are used interchangeably.

Generally, as one considers higher-order virial coefficients, the models deviate from all-atom behavior to a greater extent. The 1bC/D and DODECA models have similar profiles in Figure 2.6 as the models deal with steric interactions in a similar way, although the charged sites in the 1bC/D model also contribute to a smaller extent to the net steric repulsions. The 1bAA model is the most accurate from among the current set of CG models for reproducing the all-atom steric interactions (within ~ 10% for all coefficients). The results illustrate the importance of considering steric interactions when choosing a CG model and illustrate a procedure for determining how well a given CG model reproduces the steric interactions of an all-atom representations as one increasing protein concentration.

Considering beyond steric-only interactions, higher order virial coefficients were calculated using the HEXA and DODECA models to predict self-interactions from low to high concentration at low and intermediate ionic strength values for MAb 1 and 2 at pH 5 and 6.5 and shown to be tractable in prior work.⁴⁹ The 1bAA model is overly computationally burdensome once one includes more than just steric interactions.²⁸ The 1bC/D model is tractable for multi-body simulations, and provides the opportunity to study the effect of charge anisotropy at increased concentrations (see Chapter 4). As an illustration, Figure 2.7 shows up to the fifth virial coefficient for MAb B at pH 5 using the 1bC/D model as a function of TIS, with the virial coefficient values normalized by their respective steric-only value (i.e., from Figure 2.6). The black horizontal line at an $A_i/A_{i,ST}$ value of 1 represents when a given virial coefficient transitions from net repulsive interactions (greater than 1) to net attractive interactions (less than 1) with respect to steric-only interactions. Based on Figure 2.7, the second virial coefficient (black symbols and curve) is always negative, suggesting

that in the regime of two body interactions, MAb B experiences net attractive self-interactions. The attractions are strong at low TIS and decrease with increasing ionic strength (Case iii in the nomenclature above). By adding a third MAb B molecule, the interactions in low TIS change from strong attractions to strong repulsions (indigo symbols and curve). The strength of interactions becomes weaker as the salt concentration increases toward the region where short-ranged non-electrostatic interactions overcome electrostatic interactions. Adding a fourth molecule (green symbols and curve) changes the net interactions again from net repulsive to strongly attractive interactions in low TIS. The results for the fifth virial coefficient (yellow symbols and curve) are only shown for high TIS (125 mM and higher) because the magnitude of attractive interactions was so large that the simulations for the fifth virial coefficient did not converge for lower TIS values.

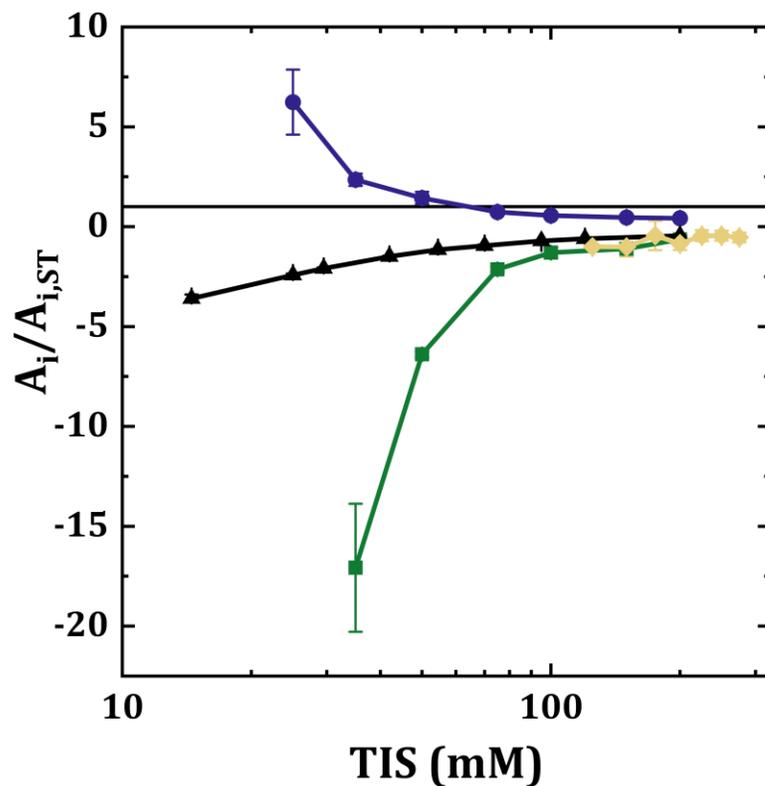


Figure 2.7: Virial coefficients with respect to steric-only behavior ($A_i/A_{i,ST}$) illustrated for MAb B at pH 5 as a function of TIS using the 1bC/D model: $B_{22}/B_{22,ST}$ (black triangles and line), $A_3/A_{3,ST}$ (indigo circles and line), $A_4/A_{4,ST}$ (green squares and line) and $A_5/A_{5,ST}$ (yellow diamonds and line) coefficients. The black horizontal line represents net interactions equivalent to steric-only interactions for a given virial coefficient ($i=2, 3, 4, 5$).

2.3.2 Comparison of computational burden

Computational burden is important to consider when selecting a CG model. While higher resolution models might provide more accurate results, they require more computational resources and can limit the practical application of a given model. The CG models were compared based on their relative computational burdens. Table 2.2 shows the relative CPU time for 10^7 MSOS steps for each CG model, relative to

that for the HEXA model, as a function of the virial coefficient being calculated. The CPU times scale roughly with the square of the number of beads in the simulation, as expected. The footnote to Table 2.2 also indicates the typical order of magnitude for the CPU time for the HEXA model for each virial coefficient to allow the results for any of the models to be translated to real time, assuming a comparable processor. The 1bC/D and 1bAA models do not have a fixed number of beads across all monoclonal antibodies, as the number of charged or titratable amino acids in the 1bC/D model will depend on the antibody sequence, and the number of total amino acids in the 1bAA model can also differ between different antibodies. For the example antibodies used here, the 1bC/D model showed computational burdens that were between 25 and 50 times lower than the 1bAA model, while still providing a similar or improved resolution for the location of surface charges. CPU cost per configuration is not the only factor in the net computational burden, as different CG models may require a different number of configurations to converge for a given property (e.g., virial coefficient). Typically, solution conditions that result in strongly attractive net interactions (large and negative virial coefficient) require significantly larger numbers of configurations to converge the simulations; this is as expected based on the choice of reference state as steric-only interactions.^{178,180,199}

Table 2.2: Order-of-magnitude relative CPU time for calculating virial coefficients for each CG model, normalized by the CPU time using the HEXA model^a

Virial coefficient	Normalized CPU time		
	DODECA	1bC/D	1bAA
2 nd	30	10 ³	5 x 10 ⁴
3 rd	20	7 x 10 ³	2 x 10 ⁵
4 th	10	7 x 10 ³	4 x 10 ⁵
5 th	30	9 x 10 ³	5 x 10 ⁵

^aThe CPU time for the HEXA model was 29, 48, 96, and 132 seconds for the second, third, fourth, and fifth virial coefficients, respectively, for 10⁷ MSOS steps in this particular example.

2.3.3 Considerations in selecting among different CG models

Ideally, one would be able to know *a priori* what level of CG model is needed to predict or capture the MAb self-interactions qualitatively or quantitatively, but results in the literature^{28,49,50,68} and illustrated here indicate that some experimental data is required so that regression or refinement of model parameters are needed on a case-by-case basis, both in terms of the protein in question and the solution conditions (e.g., pH and ionic strength, as well as the salt type^{28,49,50,200–202}, and co-solute concentrations such as sucrose^{49,50}). In principle, one should use the lowest resolution CG model needed to capture the relevant physics and contributions to the self-interactions, so as to minimize the computational burden. It is clear that a major issue with low resolution models (e.g., domain-only models such as HEXA and DODECA in the present report) is that they do not accurately account for the location of individual charges. One can consider “lumping” charges in smaller groups^{73,198}, or

trying to assign charge “patches” in an analogous way to assigning hydrophobic “patches”.^{75,192,203–207} However, it is questionable whether one can know *a priori* how to assign such “patches”. A hybrid approach such as the 1bC/D model offers one approach that balances the advantages of domain-level models (e.g., HEXA and DODECA) with explicit amino acid models (e.g., 1bAA) to explicitly account for the impact of the surface charge distribution while lumping the steric repulsions and non-electrostatic short-ranged attractions. This does not account for geometrically highly specific hydrophobic or van der Waals “lock and key” configurations/interactions.^{208,209}

In addition, while a lower resolution model may be able to “tune” the model parameters to force the model to recapitulate the experimental values (e.g., B_{22}) at low protein concentrations, the model may be (greatly) inaccurate at higher concentrations where multi-protein interactions become important.⁴⁹ This is particularly a concern for conditions with strongly net attractive electrostatic interactions. One may still be able to fit or refine model parameters for such simplified models against the high concentration data directly^{38,155,195,210,211}, but at that point the models are only recapitulating the known data rather than predicting the experimental behavior without already having the experimental results for comparison.⁴⁹

2.4 Summary and conclusions

A series of CG models for proteins that spanned from domain-level descriptions to amino-acid-level descriptions were compared based on their ability to quantitatively and qualitatively capture experimental self-interactions at low protein concentrations as a function of solution ionic strength for five published monoclonal antibodies across multiple pH-buffer systems. The models were compared based on

their ability to capture qualitatively different experimental profiles, relative computational burdens, and extension to high protein-concentration conditions. Comparisons were made with emphasis on the ability of each model to accurately represent steric repulsions, electrostatic repulsions and attractions, and to a lesser extent on non-electrostatic attractions for net behavior that ranged from “colloid-like” electrostatic repulsions to the opposite extreme where particular oppositely charged amino acids lead to strong electrostatic attractions. The focus was on the impact of different levels of coarse-graining for the distribution of charged amino acids on the protein surface, rather than the distribution of hydrophobic amino acids. Domain-level models predicted net repulsive and mildly attractive net self-interactions with reasonable accuracy and much lower computational burden compared to higher resolution models but were inherently limited in the context of protein-solution conditions when attractive electrostatic interactions between oppositely charged amino acid residues dominated. In the latter case, explicit sites were needed for each charged amino acid or charged “site” on the protein surface. This is expected to be general across CG models beyond those considered here, and examples here illustrate that this will be exacerbated at higher protein concentrations. A hybrid model (1bC/D) was introduced that helps to balance each of these considerations for future applications to predicting the behavior of challenging MAb systems at high concentrations.

Chapter 3

ELECTROSTATICALLY MEDIATED ATTRACTIVE SELF-INTERACTIONS AND REVERSIBLE SELF-ASSOCIATION OF FC-FUSION PROTEINS

3.1 Introduction

As discussed in section 1.6.2, Fc-fusion proteins are a growing class of therapeutic proteins that are composed of an Fc domain from a MAb linked to a fusion partner (FP) protein with a peptide linker. The diversity in the FPs and linkers of Fc-fusion proteins presents a greater challenge in their biophysical characterization in comparison with more established and structurally similar proteins such as MAbs.^{163,212,213} Biophysical behaviors that can negatively impact drug product development such as aggregation, elevated viscosity, low solubility, and phase separation are not uncommon for Fc-fusion proteins.^{118,153,214–218} Biophysical behaviors are challenging to predict *a priori* for Fc-fusion proteins (and therapeutic proteins as a whole) as they are dependent on not only the chosen Fc, fusion partner, and linker, but also on sequence optimization and formulation conditions (see also section 1.5).^{11,118,215,216,219} Given the significant biophysical challenges in the development of Fc-fusion proteins, there is an outstanding need for methods and tools to screen prospective drug candidates and formulation conditions.¹⁶³

This chapter is focused on a systematic biophysical characterization of two Fc-fusion proteins (monovalent and bivalent) and the corresponding fusion partner protein, depicted in Figure 3.1. The fusion partner protein is a 15 kDa globular protein,

while the monovalent and bivalent Fc-fusion proteins have one and two fusion partner proteins, respectively, joined to the Fc region of a MAb (roughly 50 kDa) with a flexible peptide linker. SLS and DLS measurements were performed for the three proteins at two pH values (5 and 6.5) as a function of solution ionic strength (from 10 mM to at least 300 mM) to characterize net self-interactions and reversible self-association. In pH 6.5 solutions, the proteins displayed net attractive self-interactions and a propensity to form reversible oligomers, while for pH 5 solutions, the proteins typically displayed “simple-colloidal” electrostatic repulsions and a reduced propensity for reversible self-association. CG molecular simulations with a 1-bead-per-amino-acid (1bAA) model (visualized in Figure 3.1) were used to model the self-interactions and probe the pairwise electrostatic interactions between charged amino acids that were implicated in the net attractive experimental self-interactions and reversible self-association. The content in this chapter has been submitted to a peer-reviewed journal.

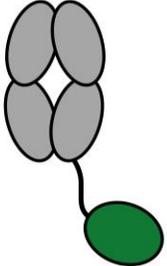
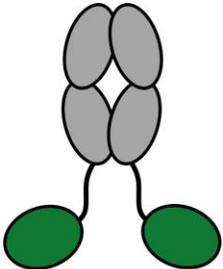
Protein	Diagram	1bAA model
Fusion partner		
Monovalent Fc-fusion		
Bivalent Fc-fusion		

Figure 3.1: Schematic diagrams (left) and 1bAA models (right), shown roughly to scale, compared to the 1bAA structures, of the three proteins in this chapter.

3.2 Materials and methods

3.2.1 Sample preparation

Sodium acetate solutions were prepared at a concentration of 15.6 mM and at pH 5.0 ± 0.05 with glacial acetic acid (Thermo Fisher Scientific, Waltham, MA). Histidine hydrochloride solutions were prepared at a concentration of 10 mM and at pH 6.5 ± 0.05 with L-histidine monochloride monohydrate (Sigma-Aldrich, St. Louis, MO). Buffer concentrations were chosen such that the contribution to the total ionic strength of the solution was 10 mM. All solutions were prepared with deionized water (Elga LabWater, Veolia, Woodridge, IL), titrated with a 5 M sodium hydroxide

solution (Thermo Fisher Scientific), and filtered with 0.45 μm membrane filters (MilliporeSigma, Burlington, MA).

Protein stock solutions were provided by Merck & Co., Inc. Three 24-h buffer exchanges against the buffer of interest were performed at 4 $^{\circ}\text{C}$ with either 3.5 kDa molecular weight cutoff Slide-A-Lyzer dialysis cassettes (Thermo Fisher Scientific) for the fusion partner protein, or 10 kDa molecular weight cutoff dialysis tubing (Repligen, Waltham, MA) for the two Fc-fusion proteins. The protein solutions were then filtered with 0.22 μm low protein-binding filters (Celltreat, Pepperell, MA), and protein concentration was determined by UV absorbance at 280 nm (DeNovix DS-11, Wilmington, DE). Samples at varying protein concentrations and NaCl concentrations were prepared by gravimetric dilution with solutions of only buffer, and buffer with approximately 2 M NaCl (Thermo Fisher Scientific). Samples were centrifuged at 20,000 rcf for 10 minutes immediately prior to light scattering measurements.

3.2.2 Static light scattering

SLS measurements were performed at 20 $^{\circ}\text{C}$ with a DynaPro NanoStar instrument (Wyatt Technology, Santa Barbara, CA) at a laser wavelength of 662 nm. Excess Rayleigh scattering (R_{90}^{ex}) was determined from the mean scattering intensity at a 90 $^{\circ}$ scattering angle. R_{90}^{ex} was divided by the optical constant K , to calculate the excess Rayleigh ratio (R_{90}^{ex}/K). K is $\frac{(2\pi)^2 n^2 \left(\frac{dn}{dc_2}\right)^2}{N_A \lambda^4}$, where n is solvent refractive index, $\left(\frac{dn}{dc_2}\right)$ is the refractive index increment with respect to protein concentration, N_A is Avogadro's number, and λ is the laser wavelength. For a given protein, pH, and NaCl concentration, measurements of R_{90}^{ex}/K as a function of protein concentration (c_2) were regressed using Equation 3.1 along with the true protein molecular weight (M_w)

to calculate the protein-protein Kirkwood-Buff integral (G_{22}) and apparent molecular weight ($M_{w,app}$).³⁵

$$R_{90}^{ex}/K = M_{w,app}c_2 + M_w G_{22}c_2^2 \quad 3.1$$

$M_{w,app}$ is the weight average of the molecular weights of all species in the solution, with some contribution also from protein-solvent interactions.³⁵ G_{22} quantifies the net self-interactions between proteins in solution. In cases where interactions are sufficiently weak (typically, $|c_2 G_{22}| \ll 1$), G_{22} can be related to the second osmotic virial coefficient (B_{22}) via $B_{22} = -\frac{1}{2} \lim_{c_2 \rightarrow 0} G_{22}$.³⁵ B_{22} is the concentration-independent analogue to G_{22} that quantifies net two-body self-interactions. A value of B_{22} greater (less) than zero denotes net repulsions (attractions) with respect to an ideal gas mixture. Due to strong self-interactions at multiple conditions in this work, all experimental SLS measurements of net self-interactions were presented as $-G_{22}/2$ values so that the scale of the data were the same as that for B_{22} and the measurements could be directly compared to simulated B_{22} values.

3.2.3 Dynamic light scattering

DLS was measured simultaneously with SLS with a Wyatt DynaPro NanoStar instrument. Time-dependent fluctuations in scattered light intensity at a 90° scattering angle were processed by an autocorrelator to calculate the intensity autocorrelation function $g_2(\tau)$. For solutions with a sufficiently monodisperse particle size distribution (i.e., $g_2(\tau)$ with a single exponential decay), $g_2(\tau)$ was analyzed using the method of cumulants^{45,220}

$$g_2(\tau) = B + \beta e^{-2\tau\Gamma} \left(1 + \frac{\mu_2}{2} \tau^2\right)^2 \quad 3.2$$

where B is the average baseline (~ 1), β is the coherence factor (dependent on instrumental configuration), τ is delay time, Γ is the first cumulant (average decay rate of the distribution), and μ_2 is the second cumulant (variance in the decay rate). The method of cumulants quantifies the collective diffusion coefficient (D_C) via $D_C = \Gamma/q^2$ and the dimensionless solution polydispersity (p_2) via $p_2 = \mu_2/(D_C q^2)^2$, where q is the scattering wave vector $q = \frac{4\pi n}{\lambda} \sin(\theta/2)$. Measurements of D_C as a function of protein concentration (c_2) were used to calculate the interaction parameter k_D via $D_C = D_0(1 + k_D c_2)$, where D_0 is the infinite-dilution or self-diffusion coefficient. k_D quantifies net self-interactions, but in contrast to B_{22} which is a purely thermodynamic quantity, k_D also contains contributions from hydrodynamic interactions.^{184,221} D_0 is related to hydrodynamic radius (r_H) via the Stokes-Einstein relation, $r_H = k_B T / 6\pi\eta D_0$, where k_B is Boltzmann's constant, T is absolute temperature, and η is the solvent viscosity.³⁶

3.2.4 1bAA CG molecular simulations and parameter optimization

The 1bAA model, MSOS simulation algorithm, and interaction potentials used in this chapter are the same as those described in section 2.2.2. At least 10^6 configurations were generated for each simulation, and simulations were performed in triplicate. In some cases, up to 10^8 configurations were needed to converge the integral in Equation 2.1, typically for systems with strong net attractions, similar to observations in previous work.^{48,49,179} The MSOS algorithm calculates B_{22} with respect to $B_{22,ST}$, the steric contribution to B_{22} , such that the output of a simulation is $B_{22}/B_{22,ST}$. The experimental $-G_{22}/2$ values were also normalized by $B_{22,ST}$, which was calculated by a separate all-atom MSOS simulation.²⁸ The values of $B_{22,ST}$ were 4.6, 9.1, and 9.5 mL/g for the FP, monovalent Fc-fusion, and bivalent Fc-fusion,

respectively. These values are in line with results for similar proteins in prior work (other globular proteins and MAbs).^{68,70,179}

As discussed in section 2.2.2 and section 2.2.4, the 1bAA model has two adjustable parameters that are optimized by experimental measurement of net self-interactions ($-G_{22}/2$ in this chapter). The ϵ_{SR} parameter represents attractive interactions other than those from electrostatics and was chosen to reproduce the experimental data at high ionic strength (> 300 mM) where electrostatic interactions are minimized due to Debye screening. For each protein and pH, MSOS simulations were performed to calculate the profile of $B_{22}/B_{22,ST}$ as a function of ionic strength over a range of ψ values. The ψ parameter adjusts the strength of electrostatic interactions to account for solution nonidealities, and was optimized by minimizing the error between the simulated predictions, $B_{22}^{predicted,i}$, and the experimental data, $(-G_{22}/2)^{experimental,i}$.

The error function used in this chapter (and Chapter 4) was different from ARD, which was used in Chapter 2 (Equation 2.7), to be more robust when considering experimental measurements near zero. Error was calculated as root-mean-square deviation (RMSD), shown in Equation 3.3, where n is the number of experimental datapoints for that protein and pH.

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{(-G_{22}/2)^{experimental,i} - B_{22}^{predicted,i}}{B_{22,ST}} \right)^2} \quad 3.3$$

Uncertainty in the optimal value of ψ was determined as the range of ψ values that resulted in RMSD values within 20% of the minimum RMSD value, similar to prior work.⁴⁸

3.2.5 Mayer-weighted average pairwise electrostatic energies

Pairwise electrostatic interaction energies in each configuration (k) in the 1bAA simulations were calculated by averaging the electrostatic energy values (Equation 2.4) for each interacting pair of charged beads. The electrostatic energy between charged bead i in protein a and charged bead j in protein b , and vice versa were averaged via Equation 3.4.

$$\phi_{ij}^{(k)} = \frac{u_{EL,ij}^{ab(k)} + u_{EL,ji}^{ab(k)}}{2} \quad 3.4$$

The Mayer-weighted average ($\tilde{\phi}_{ij}$) of each i - j electrostatic pairwise interaction was determined by averaging $\phi_{ij}^{(k)}$ values over the total number of configurations generated in the simulation (N_C), weighted by the Mayer function for the given k th configuration ($e^{-W_{22}/k_B T} - 1$) _{k} via

$$\tilde{\phi}_{ij} = \frac{1}{N_C} \frac{\sum_{k=1}^{N_C} \left(e^{-\frac{W_{22}}{k_B T}} - 1 \right)_k \times \phi_{ij}^{(k)}}{\sum_{k=1}^{N_C} \left(e^{-\frac{W_{22}}{k_B T}} - 1 \right)_k} \quad 3.5$$

where W_{22} is the potential of mean force for configuration k .^{44,48,69} ($e^{-W_{22}/k_B T} - 1$) is the same as the integrand for B_{22} (Equation 2.1), so $\tilde{\phi}_{ij}$ provides a reasonable measure of the average contribution of the interactions of a given pair of interacting charged beads to the computed $B_{22}/B_{22,ST}$.

3.3 Results and discussion

3.3.1 SLS and DLS measurements of self-interactions and reversible self-association

SLS and DLS measurements were performed at a range of pH and ionic strength conditions designed to map a broad range of possible electrostatic contributions to the net self-interactions, while remaining within typical formulation

conditions for pH, ionic strength, and choice of buffer of commercially available therapeutic proteins.³ The two chosen pH values, 5 and 6.5, modulated the charge states of the ionizable residues in the proteins (primarily His at these pH values), and the elevated ionic strength conditions reduced the strength of electrostatic interactions via Debye screening. An illustrative subset of profiles of excess Rayleigh ratio values (R_{90}^{ex}/K) as a function of protein concentration (c_2) are shown in Figure 3.2A and 3.2B for the FP protein and the monovalent Fc-fusion, respectively. Data are shown for buffer-only conditions at both pH values, and at both pH values with added 300 mM NaCl, to illustrate the impact of the different pH values and elevated ionic strength conditions and for subsequent comparison with the data presented below for the bivalent Fc-fusion. Excess Rayleigh ratio profiles for the other ionic strength conditions can be found in Figure B.1 (Appendix B). The initial slope of the profile corresponds to the apparent molecular weight ($M_{w,app}$), and the curvature is due to net self-interactions that are quantified by G_{22} , where downward curvature is due to net repulsive self-interactions and upward curvature is due to net attractive self-interactions.

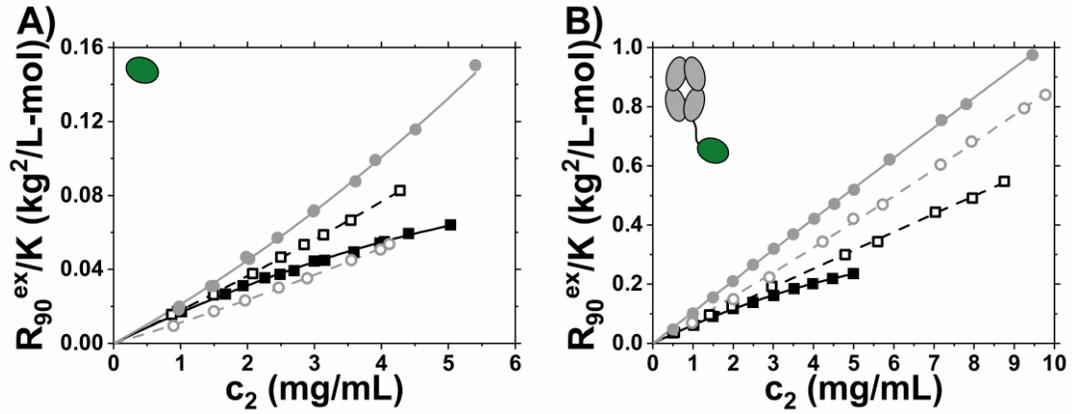


Figure 3.2: Excess Rayleigh ratio values (R_{90}^{ex}/K) as a function of protein concentration (c_2) for the fusion partner protein (A) and the monovalent Fc-fusion (B). Data are shown for a subset of the experimental conditions: pH 5 buffer-only (solid black squares), pH 5 + 300 mM NaCl (open black squares), pH 6.5 buffer-only (solid gray circles), and pH 6.5 + 300 mM NaCl (open gray circles). The curves are fits to Equation 3.1 used to calculate G_{22} and $M_{w,app}$. Illustrations in the corners are included to orient the reader when comparing between the different proteins (see Figure 3.1).

$-G_{22}/2B_{22,ST}$ values calculated from SLS for the FP protein and the monovalent Fc-fusion protein are shown in Figure 3.3A and 3.3B, respectively. $-G_{22}/2B_{22,ST}$ values above (below) 0 indicate repulsions (attractions) with respect to an ideal gas mixture, and values above (below) 1 indicate repulsions (attractions) with respect to a reference state with only steric repulsions. For both the FP and the monovalent Fc-fusion, electrostatic interactions were net repulsive at pH 5 and displayed classical “simple-colloidal” behavior where the value of $-G_{22}/2B_{22,ST}$ was positive at low ionic strength and decreased monotonically with increased ionic strength (see Figure 3.3A and Figure 3.3B). In contrast, electrostatic interactions were strongly attractive for the FP protein at pH 6.5, as $-G_{22}/2B_{22,ST}$ values were negative and large in magnitude at low ionic strength and they increased with increasing ionic

strength (see Figure 3.3A). The monovalent Fc-fusion displayed a nonmonotonic relationship between $-G_{22}/2B_{22,ST}$ and ionic strength at pH 6.5 (see Figure 3.3B). At intermediate ionic strength conditions (30 - 110 mM), electrostatic interactions were attractive as $-G_{22}/2B_{22,ST}$ values were negative and less than those at the highest ionic strength conditions (> 300 mM), where electrostatic interactions are minimized due to Debye screening. At low ionic strength conditions (10 - 20 mM), electrostatic interactions were approximately net neutral as values of $-G_{22}/2B_{22,ST}$ were increased compared to those at intermediate ionic strength, and nearly equivalent to $-G_{22}/2B_{22,ST}$ values at high ionic strength. The attractive electrostatic interactions between specific residues at the intermediate ionic strength conditions were offset by longer-ranged (“simple-colloidal”) repulsions driven by the net charge of the protein at low ionic strength.¹⁷⁹ This behavior has been observed in other proteins such as MAbs in a manner consistent with the results shown here.^{42,48,171,205,222,223}

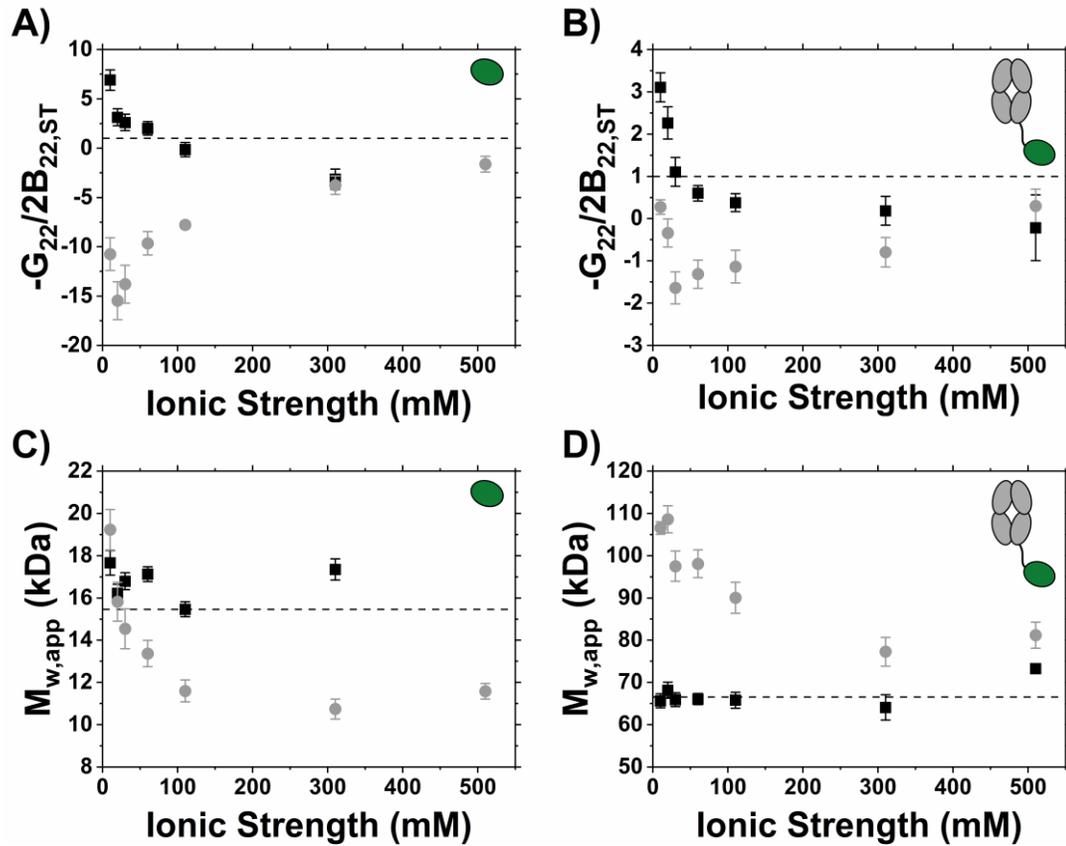


Figure 3.3: $-G_{22}/2B_{22,ST}$ as a function of ionic strength for the fusion partner protein (A), and the monovalent Fc-fusion (B) at pH 5 (black squares) and pH 6.5 (gray circles). The horizontal dashed line at $-G_{22}/2B_{22,ST} = 1$ corresponds to a steric-only reference state. Apparent molecular weight ($M_{w,app}$) as a function of ionic strength for the fusion partner protein (C), monovalent Fc-fusion (D), following the same color scheme. The horizontal dashed line corresponds to the true molecular weight. Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1).

DLS measurements that were collected simultaneously with the SLS measurements broadly corroborate the conclusions from SLS outlined above. DLS correlograms for 2 mg/mL solutions of the fusion partner protein and the monovalent Fc-fusion are shown in Figure 3.4 for the same pH and ionic strength conditions as in Figure 3.2. The correlograms show a primarily monodisperse size distribution (i.e., with a single exponential decay of $g_2(\tau)$ without the presence of a significant “shoulder” and low polydispersity values), except for the fusion partner protein at pH 5 conditions which showed a transient shoulder indicative of a relatively small number of multimers in solution. The experimental SLS and DLS data for the fusion partner protein at pH 5 should thus be considered qualitatively, but that does not impact the conclusions or simulations below.

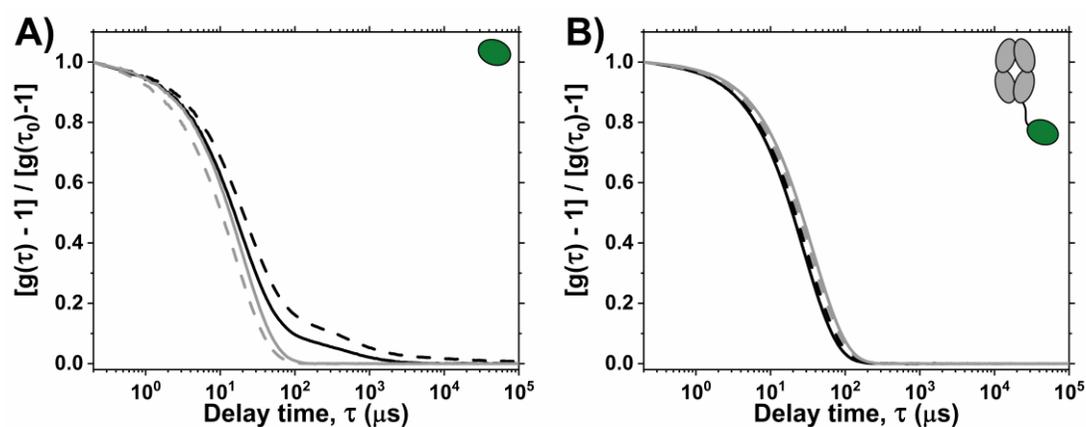


Figure 3.4: DLS correlograms for 2 mg/mL solutions of the FP protein (A) and the monovalent Fc-fusion (B). The correlograms are normalized by the value of the autocorrelation function at the shortest delay time ($g(\tau_0) - 1$) for ease of comparison between conditions. Data are shown for a subset of the experimental conditions: pH 5 buffer-only (solid black lines), pH 5 + 300 mM NaCl (dashed black lines), pH 6.5 buffer-only (solid gray lines), and pH 6.5 + 300 mM NaCl (dashed gray lines). Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1).

Because the correlograms were largely indicative of a near-monodisperse size distribution, DLS data were analyzed with a standard cumulants analysis. Net self-interactions, quantified by k_D , were qualitatively similar to B_{22} values as a function of pH and ionic strength, as shown in Figure 3.5A and Figure 3.5B for the FP protein and the monovalent Fc-fusion, respectively. The full set of DLS measurements (D_C as a function of and the corresponding polydispersity index values) are shown in Figure B.2 and B.3 (Appendix B) for the FP protein and the monovalent Fc-fusion, respectively.

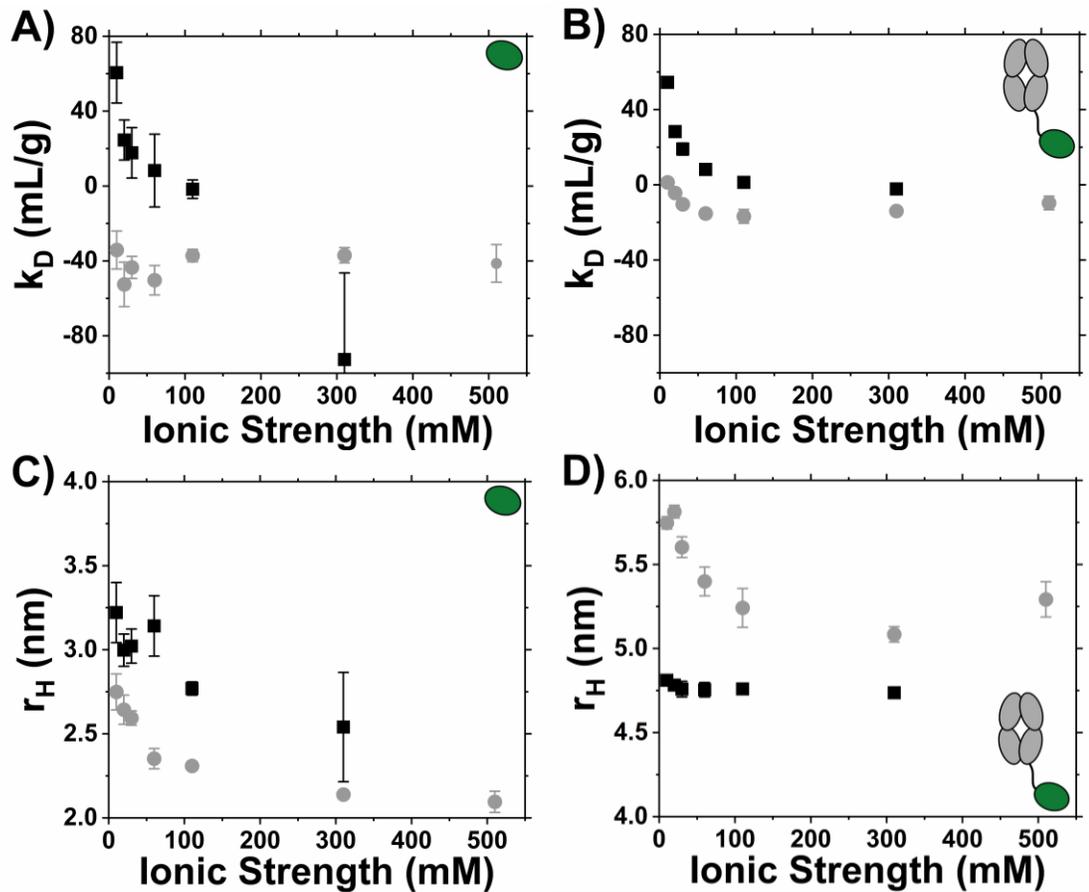


Figure 3.5: k_D as a function of ionic strength for the fusion partner protein (A), and the monovalent Fc-fusion (B) at pH 5 (black squares) and pH 6.5 (gray circles). Hydrodynamic radius (r_H) as a function of ionic strength for the fusion partner protein (C), monovalent Fc-fusion (D), following the same color scheme. Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1).

Apparent molecular weight ($M_{w,app}$) values from SLS for the FP protein and the monovalent Fc-fusion are shown in Figure 3.3C and 3.3D, respectively. $M_{w,app}$ values at pH 5 for both proteins were close to the true molecular weight across the ionic strength conditions, suggesting no significant formation of reversible oligomers. As mentioned above, the DLS measurements of the fusion partner protein at pH 5 suggested the presence of a small number of multimers, which could also have caused the inconsistent $M_{w,app}$ values in Figure 3.3C, and therefore the results for the fusion partner protein at pH 5 should be considered qualitatively. In contrast, at pH 6.5 both proteins had $M_{w,app}$ values greater than the true molecular weight at low ionic strength, and as ionic strength was increased, the value of $M_{w,app}$ decreased to near or less than the true molecular weight. Trends in hydrodynamic radius (r_H) values from DLS (Figure 3.5A and 3.5B) as a function of ionic strength were qualitatively consistent with $M_{w,app}$ values. The $M_{w,app}$ and r_H values at pH 6.5 conditions suggest the presence of reversible oligomers at low ionic strength that were mediated by electrostatic interactions and disrupted at higher ionic strength conditions when electrostatics were screened. Given the qualitative similarities between the fusion partner protein and the monovalent Fc-fusion in their net self-interactions and reversible self-association behavior, electrostatic interactions involving residues on the surface of the fusion partner protein are anticipated to play an important role.

Although the $-G_{22}/2B_{22,ST}$ values for the FP and monovalent Fc-fusion in cases at pH 6.5 conditions where the $M_{w,app}$ value was larger than the monomeric molecular weight may have been somewhat biased by interactions with reversible oligomers, the extent of self-association was relatively low (in comparison with, e.g., the bivalent Fc-fusion shown below) and the analysis utilizing G_{22} and k_D values were

used as a reasonable estimate of interactions involving monomeric proteins. The low polydispersity index values and primarily monodisperse correlograms from DLS also support that approximation. $M_{w,app}$ values significantly less than the true molecular weight at elevated ionic strength conditions for the fusion partner protein can be attributed to nonideal interactions between the protein and NaCl.³⁵ Shifts in $M_{w,app}$ values due to nonideal interactions between a given protein and solution components (e.g., NaCl) are not uncommon for other proteins.^{41,48}

SLS and DLS measurements of the bivalent Fc-fusion were performed at pH 5 and pH 6.5 for a subset of the ionic strength conditions, 10 mM (buffer-only) and 310 mM, the same conditions as those shown in Figure 3.2 for the FP protein and monovalent Fc-fusion. Excess Rayleigh ratio values (R_{90}^{ex}/K) as a function of protein concentration (c_2) are shown in Figure 3.6A (analogous to Figure 3.2). $-G_{22}/2B_{22,ST}$ values are not presented as the net self-interactions were strongly biased by interactions with a large population of higher molecular weight species present in each formulation (some of which were reversible oligomers) as evidenced by R_{90}^{ex}/K values that were an order of magnitude larger at the lowest protein concentrations (c_2) compared to those for the monovalent Fc-fusion (see Figure 3.2B). $M_{w,app}$ values (Figure 3.6B) calculated from the R_{90}^{ex}/K profiles were 4-10 times larger than the monomeric molecular weight, indicating the presence of a significant number of higher molecular weight species that greatly skewed the $M_{w,app}$ values. The curvature in each profile in Figure 3.6A is potentially the result of a convolution of monomer-monomer, monomer-dimer, dimer-dimer, etc. interactions.

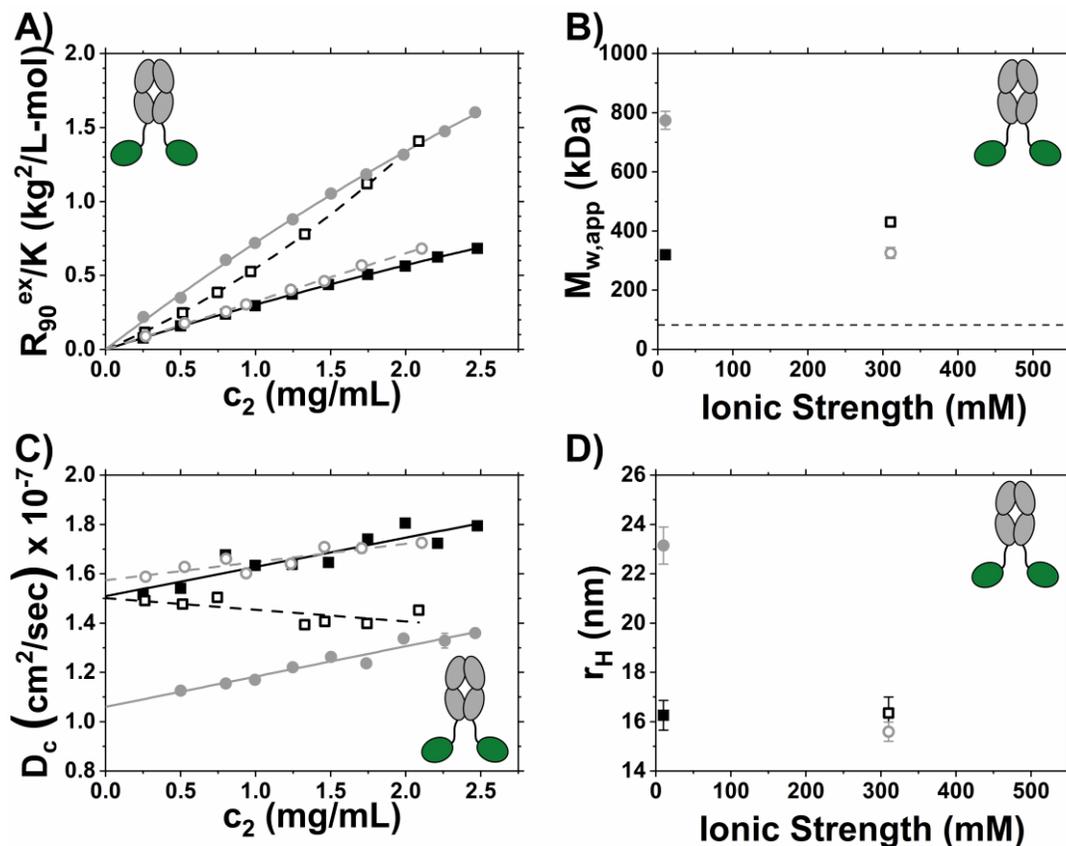


Figure 3.6: SLS and DLS measurements for the bivalent Fc-fusion at pH 5 buffer-only (solid black squares), pH 5 + 300 mM NaCl (open black squares), pH 6.5 buffer-only (solid gray circles), and pH 6.5 + 300 mM NaCl (open gray circles). A) Excess Rayleigh ratio values (R_{90}^{ex}/K) as a function of protein concentration (c_2). The curves are fits to Equation 3.1 used to calculate $M_{w,app}$ values. B) Apparent molecular weight, $M_{w,app}$, as a function of ionic strength, where the horizontal dashed line corresponds to the true molecular weight. C) Collective diffusion coefficient values, D_c , as a function of protein concentration (c_2). The lines are linear fits, and D_0 (the y-intercept of a given linear fit) was used to calculate hydrodynamic radius, r_H , shown in panel D. Illustrations in the corners are included to orient the reader when comparing between the different proteins (see Figure 3.1).

$M_{w,app}$ values contain uncertainty due to strong interactions, particularly at the pH 5 + 300 mM NaCl condition ($-G_{22}/2B_{22,ST} = -76 \pm 6$), so the simultaneous DLS measurements were used as an orthogonal measure to support qualitative conclusions about the populations of higher molecular weight species and the dependence of reversible self-association on solution conditions. Collective diffusion coefficient values (D_c) as a function of protein concentration (c_2) are shown in Figure 3.6C, as an analogue to the R_{90}^{ex}/K profiles in Figure 3.6A. Hydrodynamic radius (r_H) values calculated by the Stokes-Einstein relation are shown in Figure 3.6D, to compare against $M_{w,app}$ values from Figure 3.6B. DLS correlograms for 2 mg/mL solutions and polydispersity index as a function of c_2 are shown in Figure 3.7 and Figure B.4 (Appendix B), respectively. The correlograms show a broad, single exponential decay at much longer delay times than what were observed for the monovalent Fc-fusion, and polydispersity index values were fairly high (between 0.2 and 0.5), suggesting that there were a range of oligomer species present. Collective diffusion coefficient values (D_c) and hydrodynamic radius (r_H) values should be interpreted as representative of the distribution of species and not indicative of a monomeric solution and are shown as a method to quantify the relative differences between each formulation. DLS data were analyzed with a standard cumulants analysis as more complex analyses (e.g., regularization methods) yielded broad size distributions that did not indicate an obvious key oligomeric species. Like the FP protein and the monovalent Fc-fusion, $M_{w,app}$ and r_H values were similar as a function of ionic strength at pH 5 conditions. At pH 6.5, $M_{w,app}$ and r_H values decreased when ionic strength was increased, again indicative of electrostatically driven reversible self-association. There was also a “baseline” of higher molecular weight species for the bivalent Fc-fusion that was

invariant with the different solution conditions in this study and led to a $M_{w,app}$ value of approximately 300 kDa.

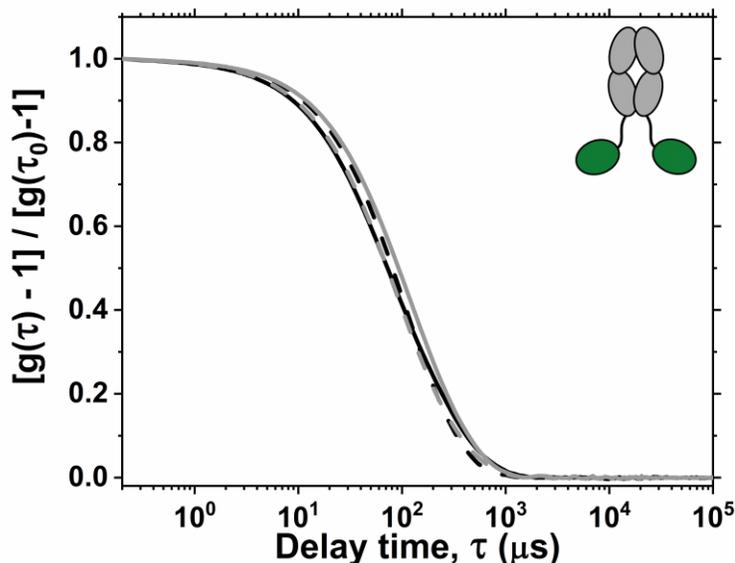


Figure 3.7: DLS correlograms for 2 mg/mL solutions of the bivalent Fc-fusion. The correlograms are normalized by the value of the autocorrelation function at the shortest delay time ($g(\tau_0) - 1$) for ease of comparison between conditions. Data are shown for a subset of the experimental conditions: pH 5 buffer-only (solid black lines), pH 5 + 300 mM NaCl (dashed black lines), pH 6.5 buffer-only (solid gray lines), and pH 6.5 + 300 mM NaCl (dashed gray lines). Illustrations are included in the corners to orient the reader when comparing between the different proteins (see Figure 3.1).

3.3.2 CG molecular simulations to model self-interactions

The FP protein, monovalent Fc-fusion, and bivalent Fc-fusion all displayed electrostatically mediated reversible self-association at pH 6.5 conditions, where $M_{w,app}$ values indicated the presence of a sub-population of reversible oligomers that dissociated in solutions with elevated ionic strength. Also, the net electrostatic self-interactions for the fusion partner protein and monovalent Fc-fusion were attractive at

pH 6.5 (i.e., negative values of $-G_{22}/2B_{22,ST}$ that were less negative at high ionic strength). This finding was unsurprising, as attractive self-interactions and reversible self-association are phenomenologically linked.^{24,101–105} However, measurements of net self-interactions do not always directly correlate to reversible self-association and the formation of reversible oligomers.¹⁰² Even in this work, $-G_{22}/2$ values for the monovalent Fc-fusion at pH 6.5 were not consistently correlated with $M_{w,app}$ values, as $M_{w,app}$ values monotonically decreased with increased ionic strength, and $-G_{22}/2$ values were nonmonotonic with ionic strength. To address the link between self-interactions and reversible self-association for these proteins in more detail than what can be ascertained from experimental measurements of net self-interactions (i.e., G_{22} or k_D values), 1bAA molecular simulations were used that were tuned to reproduce the experimental net self-interactions. The 1bAA simulations were applied to: 1) evaluate whether the simulations could qualitatively or semi-quantitatively capture the experimental trends of net self-interactions for the fusion partner protein and monovalent Fc-fusion; 2) make qualitative predictions about net electrostatic self-interactions for the bivalent Fc-fusion; and 3) gain insight into potential key charged amino acids involved in attractive electrostatic self-interactions at pH 6.5 that could also be implicated in electrostatically mediated reversible self-association.

Experimental values of $-G_{22}/2B_{22,ST}$ were used to optimize two parameters in the 1bAA model, ϵ_{SR} and ψ , such that the simulated values of $B_{22}/B_{22,ST}$ reproduced the experimental data. The ϵ_{SR} parameter scales the strength of short-ranged non-electrostatic attractions and was chosen to match the data at high ionic strength (> 300 mM) where electrostatic interactions were effectively fully screened. Then, the ψ parameter was chosen to minimize the prediction error (in the form of RMSD,

Equation 3.3) of $B_{22}/B_{22,ST}$ as a function of ionic strength for a given pH. Optimal values of ϵ_{SR} and ψ are shown in Table 3.1, and plots of RMSD vs. ψ are available in Figure B.5 (Appendix B). Uncertainty in ψ was estimated as the range of ψ values that fell within 20% of the minimum RMSD.

Table 3.1: Parameter values for 1bAA simulations

Protein	pH	ϵ_{SR} ($k_B T$)	ψ^β
Fusion partner	5	0.40	1.00 ± 0.20
Fusion partner	6.5	0.50	1.40 ± 0.10
Monovalent Fc-fusion	5	0.38	0.72 ± 0.06
Monovalent Fc-fusion	6.5	0.50	1.25 ± 0.05
Bivalent Fc-fusion ^a	5	0.38	0.72 ± 0.06
Bivalent Fc-fusion ^a	6.5	0.50	1.25 ± 0.05

^aParameter values for the bivalent Fc-fusion were chosen to match those for the monovalent Fc-fusion because the presence of large amounts of higher molecular weight species precluded reliable experimental measurement of G_{22} values for the bivalent Fc-fusion.

^βUncertainty in the ψ parameter was defined as the range that resulted in RMSD values within 20% of the minimum.

Predictions of $B_{22}/B_{22,ST}$ as a function of ionic strength and pH are shown in Figure 3.8A for the FP protein, and Figure 3.8B for the monovalent Fc-fusion. The 1bAA simulations reproduced the qualitative behavior of net self-interactions for the FP protein at both pH values, although they could not reach the magnitude of strong repulsions present at pH 5. Electrostatic repulsions for the monovalent Fc-fusion at pH 5 were captured quantitatively by the 1bAA simulations. The simulations accurately predicted the presence and magnitude of strong electrostatic attractions at pH 6.5, but

not the nonmonotonic trend in the experimental data, as the simulations did not capture that the net self-interactions became less attractive at the lowest ionic strength values. This deficiency may stem from representing charged amino acids in the 1bAA model as having static, rather than fluctuating, charges. Quantitative accuracy for complex systems such as the monovalent Fc-fusion at pH 6.5 might be improved by implementation of charge fluctuations, but that was out of scope for this chapter (see Chapter 4) and could prove to be computationally burdensome with the 1bAA model.⁸³

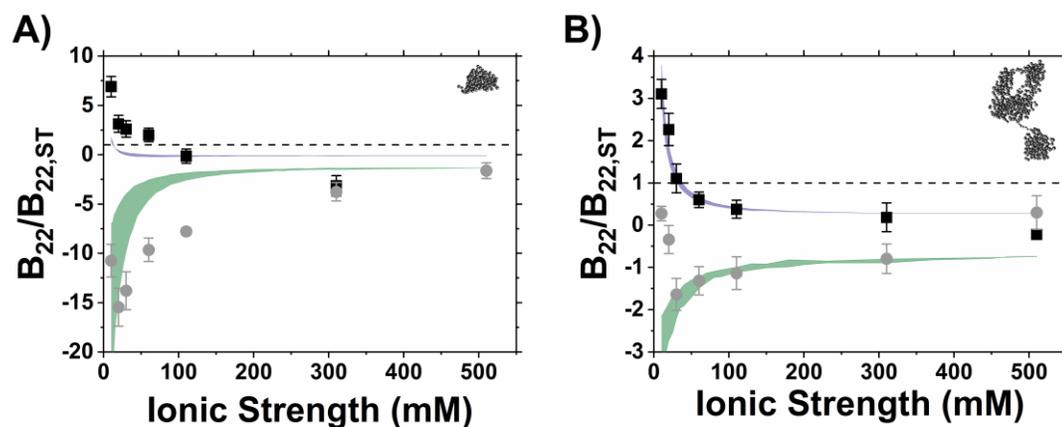


Figure 3.8: $B_{22}/B_{22,ST}$ as a function of ionic strength calculated by 1bAA simulations for the FP protein (A) and the monovalent Fc-fusion (B). Experimental measurements are reproduced from Figure 3.3 where data at pH 5 is shown as black squares and at pH 6.5 as gray circles. The shaded regions correspond to predictions from the 1bAA simulations at pH 5 (indigo) and pH 6.5 (green) based on uncertainty in the ψ parameter. 1bAA diagrams in the corners are included to orient the reader when comparing between the different proteins (see Figure 1).

The experimental data for the bivalent Fc-fusion was convoluted by the presence of many higher molecular weight species and therefore were not used to

calculate $-G_{22}/2$ values. The 1bAA simulations were leveraged as an alternative approach to make qualitative predictions of the net electrostatic self-interactions via contour plots. These plots indicated how the simulated $B_{22}/B_{22,ST}$ was predicted to change for the bivalent Fc-fusion as a function of ionic strength and ψ for a given pH, shown in Figure 3.9A (pH 5) and Figure 3.9B (pH 6.5). These simulations did not require experimental data from SLS or DLS and were based only on the amino acid sequence, homology model structure, and a fixed value of ϵ_{SR} , while the ionic strength and ψ values were sampled systematically. Analogous contour plots for the other two proteins are available in Figure B.6 (Appendix B). This method is useful to qualitatively screen potential candidates or formulation conditions (pH and ionic strength) based on the potential net electrostatic self-interactions. The simulations predicted similar qualitative behavior to the monovalent Fc-fusion and the FP protein: net electrostatic repulsions at pH 5 and net electrostatic attractions at pH 6.5. With the optimal ϵ_{SR} and ψ values for the monovalent Fc-fusion, the 1bAA simulations predicted stronger repulsions at pH 5 and stronger attractions at pH 6.5 for the bivalent Fc-fusion.

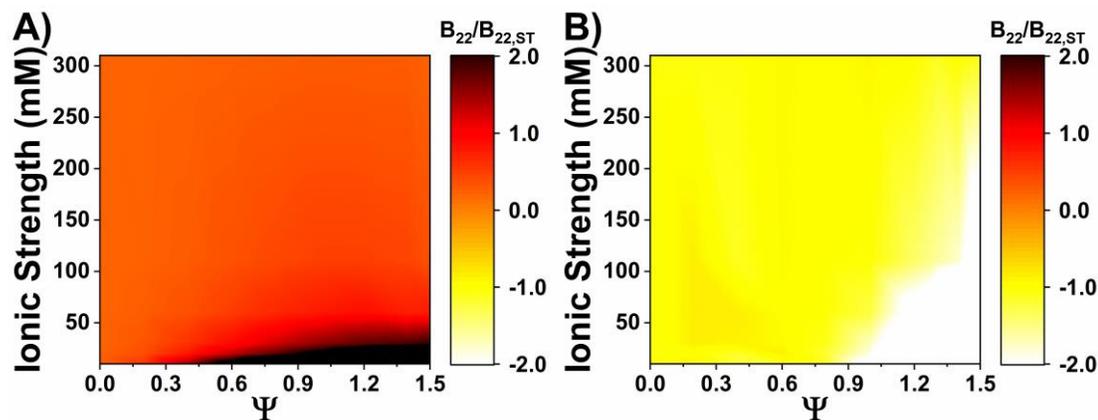


Figure 3.9: Simulated contour plots of $B_{22}/B_{22,ST}$ as a function of ψ and ionic strength for the bivalent Fc-fusion at pH 5 (A) and pH 6.5 (B).

3.3.3 Identification of specific electrostatic interactions involved in attractive self-interactions

The presence of strongly attractive electrostatic self-interactions and reversible self-association in pH 6.5 conditions for all three proteins that could be reduced by elevated ionic strength conditions was noteworthy, but the results above did not directly address the question of which charged amino acids or pairs of charged amino acids interacting between adjacent proteins were more influential in mediating this behavior. To help address this question, interaction heatmaps of average Mayer-weighted pairwise electrostatic energy values ($\tilde{\phi}_{ij}$ from Equation 3.5) were generated from the 1bAA simulations and are shown in Figure 3.10 for each protein at pH 6.5 and 10 mM ionic strength conditions. The corresponding simulated $B_{22}/B_{22,ST}$ values for the FP protein (panel A), monovalent Fc-fusion (panel B), and bivalent Fc-fusion (panel C) at these conditions were -7.9 ± 0.2 , -2.7 ± 0.1 , and -11.1 ± 0.4 , respectively. The heatmaps show the $\tilde{\phi}_{ij}$ value for each pairwise interaction between two charged amino acids in the simulations, where attractive $\tilde{\phi}_{ij}$ values are shown in red, repulsive

$\tilde{\phi}_{ij}$ values are shown in blue, and $\tilde{\phi}_{ij}$ values smaller in magnitude transition to white. The x-axis and y-axis represent the charged sites in the protein in sequence order, and the color in the heatmap for charged amino acid i interacting with charged amino acid j on the adjacent protein is the $\tilde{\phi}_{ij}$ value. Heatmaps for the two Fc-fusion proteins (panels B and C) were split by whether the charged amino acid was in the Fc portion or FP domain of the molecule.

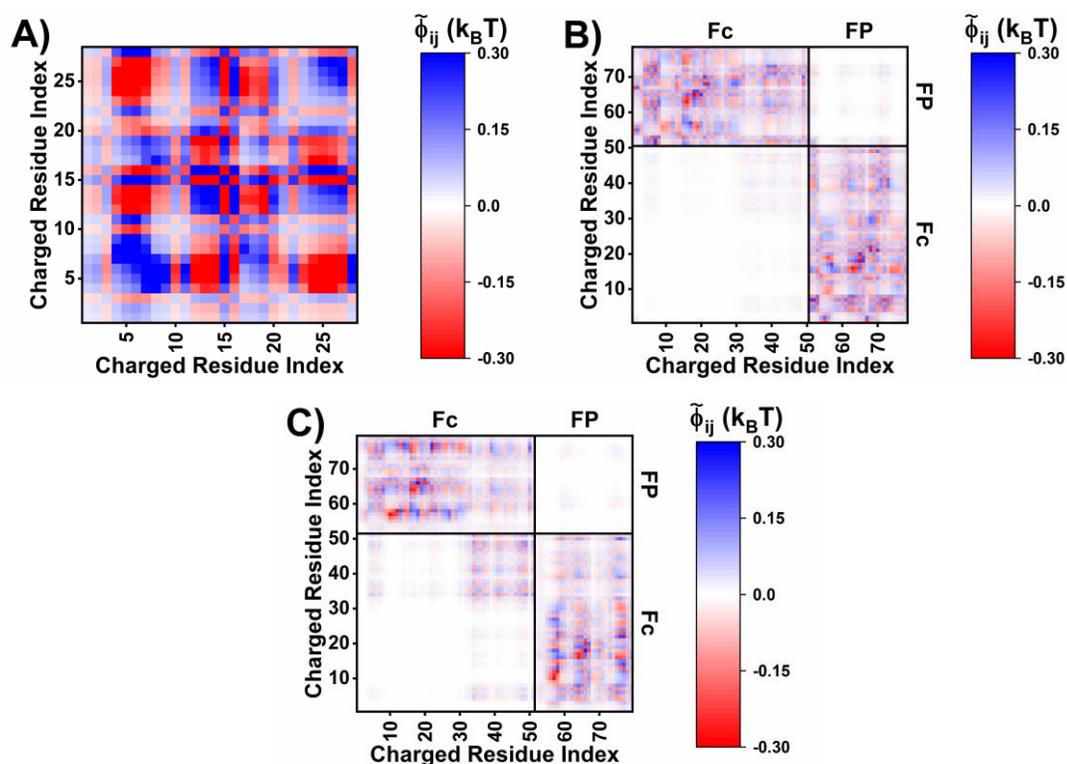


Figure 3.10: Mayer-weighted average pairwise electrostatic energy values ($\tilde{\phi}_{ij}$) for charged sites in the FP protein (A), the monovalent Fc-fusion (B), and the bivalent Fc-fusion (C) at pH 6.5 and 10 mM ionic strength. Pairwise repulsions (positive $\tilde{\phi}_{ij}$ values) are shown in blue, and attractions (negative $\tilde{\phi}_{ij}$ values) are shown in red. The heatmaps for the Fc-fusion proteins are ordered based on the structural domains.

Results at pH 5 and 10 mM ionic strength are shown in Figure 3.11 as a contrast for solution conditions where the simulated B_{22} values correspond to net repulsions. The corresponding simulated $B_{22}/B_{22,ST}$ values for the FP (panel A), monovalent Fc-fusion (panel B), and bivalent Fc-fusion (panel C) at these conditions were 1.69 ± 0.0 , 3.2 ± 0.0 , and 5.3 ± 0.0 respectively.

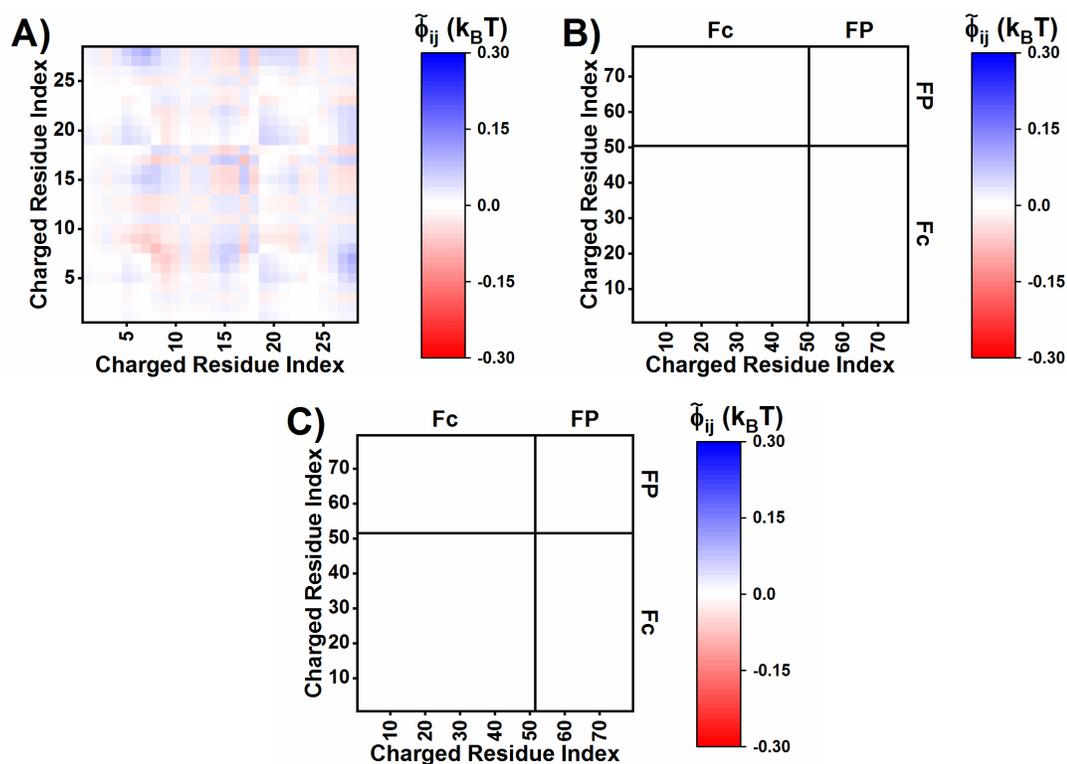


Figure 3.11: Mayer-weighted average pairwise electrostatic energy values ($\tilde{\phi}_{ij}$) for charged sites in the FP protein (A), the monovalent Fc-fusion (B), and the bivalent Fc-fusion (C) at pH 5 and 10 mM ionic strength. Pairwise repulsions (positive $\tilde{\phi}_{ij}$ values) are shown in blue, and attractions (negative $\tilde{\phi}_{ij}$ values) are shown in red. The heatmaps for the Fc-fusion proteins are ordered based on the structural domains.

Simulations were performed at 10 mM ionic strength, where electrostatic interactions are only mildly screened. Because the 1bAA simulations did not capture the nonmonotonic experimental behavior of $-G_{22}/2B_{22,ST}$ as a function of ionic strength for the monovalent Fc-fusion at pH 6.5, the heatmaps were also generated at 30 mM ionic strength for the two Fc-fusions where the model could quantitatively capture the experimental value of $B_{22}/B_{22,ST}$ for the monovalent Fc-fusion, shown in Figure 3.12. The results were qualitatively similar, with average energy values smaller in magnitude than they were at 10 mM ionic strength, as expected due to charge screening. The corresponding simulated $B_{22}/B_{22,ST}$ values for the monovalent Fc-fusion (panel A), and bivalent Fc-fusion (panel B) at these conditions were -1.8 ± 0.1 , and -3.4 ± 0.3 respectively.

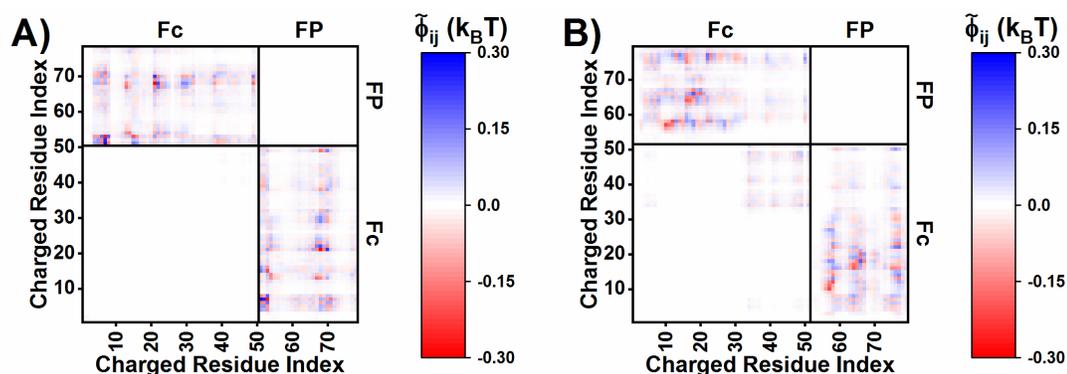


Figure 3.12: Mayer-weighted average pairwise electrostatic energies ($\tilde{\phi}_{ij}$) for charged sites in the monovalent Fc-fusion (A) and bivalent Fc-fusion (B) at pH 6.5 and 30 mM ionic strength. Repulsions are shown in blue, and attractions are shown in red. The heatmaps are ordered based on the structural domains.

The FP protein by itself had both relatively large positive (repulsive) and negative (attractive) values of $\tilde{\phi}_{ij}$ at pH 6.5 (see Figure 3.10A) that on balance lead to net electrostatic attractions, consistent with findings from prior work.^{44,69} In comparison, $\tilde{\phi}_{ij}$ values were significantly smaller at pH 5 (by approximately 1 order of magnitude). $\tilde{\phi}_{ij}$ values (both positive and negative) were close to zero for all pairwise interactions at pH 5 for the two Fc-fusion proteins (Figure 3.11B and Figure 3.11C), which is consistent with “simple colloidal” models for electrostatic net repulsions of this magnitude.^{44,69} While the Fc-Fc and FP-FP interaction regions in the heatmaps for the monovalent Fc-fusion and bivalent Fc-fusion at pH 6.5 (Figure 3.10B and Figure 3.10C, respectively) primarily contained $\tilde{\phi}_{ij}$ values near zero, the Fc-FP cross-domain interaction regions were saturated with positive (repulsive) and negative (attractive) values of $\tilde{\phi}_{ij}$. CH3-CH3 interactions (the top right quadrant in the Fc-Fc region) contained some significant $\tilde{\phi}_{ij}$ values for the bivalent Fc-fusion as well. Those results suggest that the net attractive electrostatic self-interactions observed for the monovalent Fc-fusion and the bivalent Fc-fusion at pH 6.5 were principally due to many of the pairwise cross-domain interactions between the fusion partner and the Fc domains. Given the phenomenological links between attractive self-interactions and reversible self-association as well as the apparent relationship between net attractive electrostatic self-interactions and reversible self-association in this work, that finding also suggests that Fc-FP cross-domain interactions were relevant to the reversible self-association that was observed for the two Fc-fusion proteins.

Similar to the Fc-fusion proteins, the fusion partner protein by itself displayed attractive electrostatic self-interactions and reversible self-association at pH 6.5. However, there was no Fc domain present, so that behavior could not have been due to

cross-domain interactions between the fusion partner and the Fc domains. There were several amino acid pairings that had strongly attractive electrostatic self-interactions (negative $\tilde{\phi}_{ij}$ values in Figure 3.10A), but those same pairings were not particularly notable in terms of their $\tilde{\phi}_{ij}$ values for the Fc-fusion proteins (i.e., the FP-FP region in Figure 3.10B and Figure 3.10C). Many of the individual amino acids that were involved in strongly attractive pairwise electrostatic self-interactions for the FP protein were also involved in different strongly attractive pairwise electrostatic self-interactions with amino acids in the Fc region of either Fc-fusion protein. For example, charged residues 12-14 (each of which are negatively charged) in the fusion partner protein had several strongly attractive values of $\tilde{\phi}_{ij}$ with charged residues 4-7 (each of which are positively charged), as shown in Figure 3.10A. As a part of an Fc-fusion protein, those same negatively charged amino acids in the FP (charged residues 63-65 for the Fc-fusion proteins) had strongly attractive pairwise electrostatic interactions with charged residues 15-17 (each of which are positively charged) in the Fc region (see Figure 3.10B and Figure 3.10C). That suggests that there are particular charged amino acids on the surface of the fusion partner protein that are more likely to be involved in strongly attractive electrostatic interactions. Those charged amino acid subgroups and others that may appear to be groups by inspection of the heatmaps are not necessarily spatially proximal to each other on the protein surface (i.e., they do not form a visually obvious “charge patch”), but rather were adjacent in the protein sequence when uncharged residues were removed. As a precursor to future experimental work, preliminary simulations showed that the value of $B_{22}/B_{22,ST}$ and the heatmaps of average Mayer-weighted pairwise electrostatic energy values were sensitive to charge swap mutations of key charged amino acids based on the results

from Figure 3.10 (see Appendix B.1), and those may be candidate mutations for mitigating the reversible self-association behavior of these proteins.

It is important to note that the simulations in this chapter do not explicitly model the intermittent intermolecular interactions that link reversible oligomers, but rather sample spatial configurations between two identical proteins that could include interactions that are relevant to reversible self-association. The configuration space that is sampled in the simulations is limited by the rigid structure of the 1bAA model that does not capture unfolding or flexibility of the peptide linker used to connect the fusion partner protein(s) to the Fc. Addressing these limitations was beyond the scope of this chapter but could also be the subject of future work to refine the simulation methods to more accurately model reversible self-association and the amino acids or regions (e.g., APRs) in the protein that are prone to reversible self-association. Applying experimental techniques to isolate and characterize oligomer species, validate the regions active in reversible self-association^{111,224}, or screen excipients that disrupt electrostatic attractions and/or inhibit reversible self-association are also promising areas for future studies.

3.4 Summary and conclusions

Net self-interactions and reversible self-association were characterized experimentally for a range of pH (5 and 6.5) and ionic strength conditions (10 mM to at least 300 mM) via SLS and DLS for a FP protein and its corresponding monovalent Fc-fusion and bivalent Fc-fusion proteins. All three proteins displayed electrostatically mediated attractions and reversible self-association at pH 6.5 conditions, and this was mitigated upon increased solution ionic strength. The FP protein and the corresponding monovalent Fc-fusion had similar experimental self-interaction

behaviors: “simple colloidal” net electrostatic repulsions at pH 5, but notable net electrostatic attractions at pH 6.5. 1bAA molecular simulations reasonably captured those net self-interactions, except for the nonmonotonic dependence of the self-interactions with respect to ionic strength for the monovalent Fc-fusion at pH 6.5. The bivalent Fc-fusion results indicated reversible higher molecular weight species, across all solution conditions which precluded the use of standard G_{22} or k_D analyses of the SLS and DLS data. As an analog to the results for the monovalent fusion protein and the partner protein, 1bAA simulations for the bivalent Fc-fusion were used to qualitatively compare the results at pH 5 (net repulsive interactions) and pH 6.5 (net attractive interactions). Analysis of Mayer-weighted average pairwise electrostatic energy values across the different proteins and solution conditions in those simulations highlighted the potential importance of cross-domain interactions between the Fc and fusion partner domains as a primary source of electrostatic attractions. Future work could explore the space of possible charge variants, potential isolation of oligomer species, determination of regions involved in reversible self-association, or screening of excipients that could reduce net attractive electrostatic self-interactions and/or inhibit the formation of reversible oligomers.

Chapter 4

SIMULATION OF HIGH-CONCENTRATION SELF-INTERACTIONS FOR MONOCLONAL ANTIBODIES FROM WELL-BEHAVED TO POORLY-BEHAVED SYSTEMS

4.1 Introduction

As discussed in Chapter 1, prediction and mitigation of potential poor high-concentration behaviors of therapeutic proteins is of great interest in industrial drug development. Many of those poor behaviors (e.g., aggregation, reversible self-association, opalescence, and phase separation) are related to self-interactions and can be key bottlenecks in the development process, in part because material availability may limit experimental evaluation of high-concentration samples of drug candidates at early stages of development.^{16,23,25,125} Moreover, these behaviors are dependent on the formulation conditions, e.g., protein concentration, pH, cosolute concentrations, and ionic strength. This presents a complex, multiparameter optimization problem when determining the most promising candidate(s) and formulation(s).⁹¹

Self-interactions can be challenging to predict at high-concentration via extrapolation of low-concentration measures as the balance between electrostatic interactions, short-ranged non-electrostatic attractions, and steric repulsions can shift as intermolecular distances decrease and, on average, proteins experience significant interactions with multiple neighboring proteins.⁴⁹ To help address this, prior work used measurements of B_{22} via SLS to parameterize relatively simple domain-level CG models for use in MC molecular simulations to quantitatively predict self-interactions

at high protein concentration that could be directly compared to experimental SLS data.^{49,50,68} The predictions were found to be accurate for “simple-colloidal” net repulsions and mildly net attractive conditions. However, strongly net attractive conditions due to electrostatic interactions were not well-captured by the domain-level CG models. The highest resolution CG model was the DODECA model, which contains 12 beads, one for each MAb domain.^{28,49,50}

In Chapter 2, it was noted that the 1-bead-per-charge-and-domain (1bC/D) model was potentially tractable for use in high-concentration simulations, while providing improved accuracy in representing strong electrostatic attractions in MSOS simulations of $B_{22}/B_{22,ST}$. The key advantage of the 1bC/D model over the DODECA model is the increased resolution of electrostatic attractions and repulsions, where instead of charged interactions being “lumped” into domain beads (that represent ~100 amino acids per domain for a typical MAb), the model has explicit charged sites for each charged atom in the protein. Schematic illustrations of the DODECA and 1bC/D models are shown in Figure 4.1.

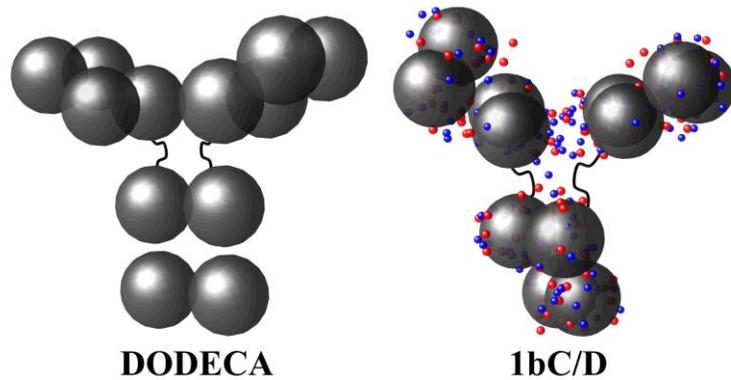


Figure 4.1: Schematic diagrams of the DODECA and 1bC/D models shown to scale. Linkers shown between the Fc and Fab are for illustrative purposes only. Charged sites in the 1bC/D model are blue (red) for positively (negatively) charged atoms. The DODECA model is reproduced from Calero-Rubio et al.²⁸

This chapter improves upon previously developed methodologies to couple sample-sparing experimental measurements of B_{22} with low- to high-concentration 1bC/D CG molecular simulations to accurately predict a broad range of high-concentration net self-interactions. Predictions of the zero- q limit of the static structure factor, $S_{q=0}$, as a function of protein concentration (up to 160 mg/mL) were compared between the DODECA and 1bC/D models for six MAbs at two pH values (5 and 6.5), and at two ionic strengths (~ 10 mM and ~ 110 mM). The accuracy of those predictions were determined by comparison with experimental SLS measurements at the same conditions. The six MAbs are MAb 1, MAb 2, MAb 3, MAb 4, MAb 5, and MAb 6. $B_{22}/B_{22,ST}$ values for MAb 1 and MAb 2 were also used as test cases in Chapter 2, and data for the other four MAbs are presented here for the first time. The data span a broad range of net self-interactions at elevated protein concentrations, from strongly repulsive to strongly attractive. The results indicate that the 1bC/D model provides improved predictions of electrostatic attractions at high protein

concentration in comparison with the DODECA model. Some challenges with respect to the representation of charge equilibria were identified and methods to model the effects of charge equilibria more accurately were evaluated, with partial success. The two models provided equivalent predictions for cases where there were strong net repulsions between MAbs and/or weak net repulsions or attractions. The 1bC/D model was tractable at high protein concentrations with an expected increase in computational burden compared to dilute protein solutions, in agreement with Chapter 2.¹⁷⁹ Parts of this chapter were published in a peer-reviewed journal and are reproduced here with permission (Appendix E).⁴⁸

4.2 Materials and methods

4.2.1 Sample preparation

The solution conditions for low-concentration SLS measurements are the same as described in section 3.2.1. Stock solutions of MAb 3, MAb 4, MAb 5, and MAb 6 were provided by Bristol-Myers Squibb (New Brunswick, NJ). Low-concentration MAb stock solutions (below approximately 15 mg/mL) were prepared by three 24 hour buffer exchanges at 4 °C against buffer solutions using 10 kDa molecular weight cutoff dialysis tubing (Repligen).

High-concentration MAb stock solutions (for measurements above ~10 mg/mL) were prepared at roughly 150 mg/mL by membrane centrifugation at 3×10^3 rcf using 10 kDa molecular weight cutoff Amicon Ultra centrifugal filtration tubes (MilliporeSigma). An iterative procedure of concentration and reconstituting the concentrate with buffer four times was used to buffer exchange the MAb stocks. Buffer solutions used in this process had reduced pH values to account for shifts in pH

at high protein concentration due to the Donnan equilibrium effect. The initial pH values before centrifugal concentration steps for MAb 3 and MAb 4 were approximately pH 4.3 for low ionic strength pH 5 solutions, pH 4.8 for intermediate ionic strength pH 5 solutions, and pH 6.3 for pH 6.5 solutions (low and intermediate ionic strength). pH values were verified to be within 0.1 of the target after the centrifugation procedure to concentrate and buffer exchange MAb stock solutions.

MAb stock concentrations were determined by UV absorbance at 280 nm (Agilent 8453 or DeNovix DS-11). Samples at low protein concentrations were prepared by gravimetric dilution to the desired protein concentration and desired NaCl concentration with an appropriate ratio of buffer solution and buffer with 1 M NaCl (Fisher Scientific). Buffer solutions and MAb stock solutions for samples at high protein concentrations were prepared with the desired NaCl concentration. Solution density was used to adjust calculated protein concentrations based on density measurements made with a DDM 2911 Plus (Rudolph Scientific) at 22.00 ± 0.02 °C, shown in Figure 4.2. All buffer and low-concentration MAb stock solutions were filtered using 0.22 μm low protein binding filters (Celltreat) before use. High-concentration MAb stock solutions were filtered prior to centrifugation.

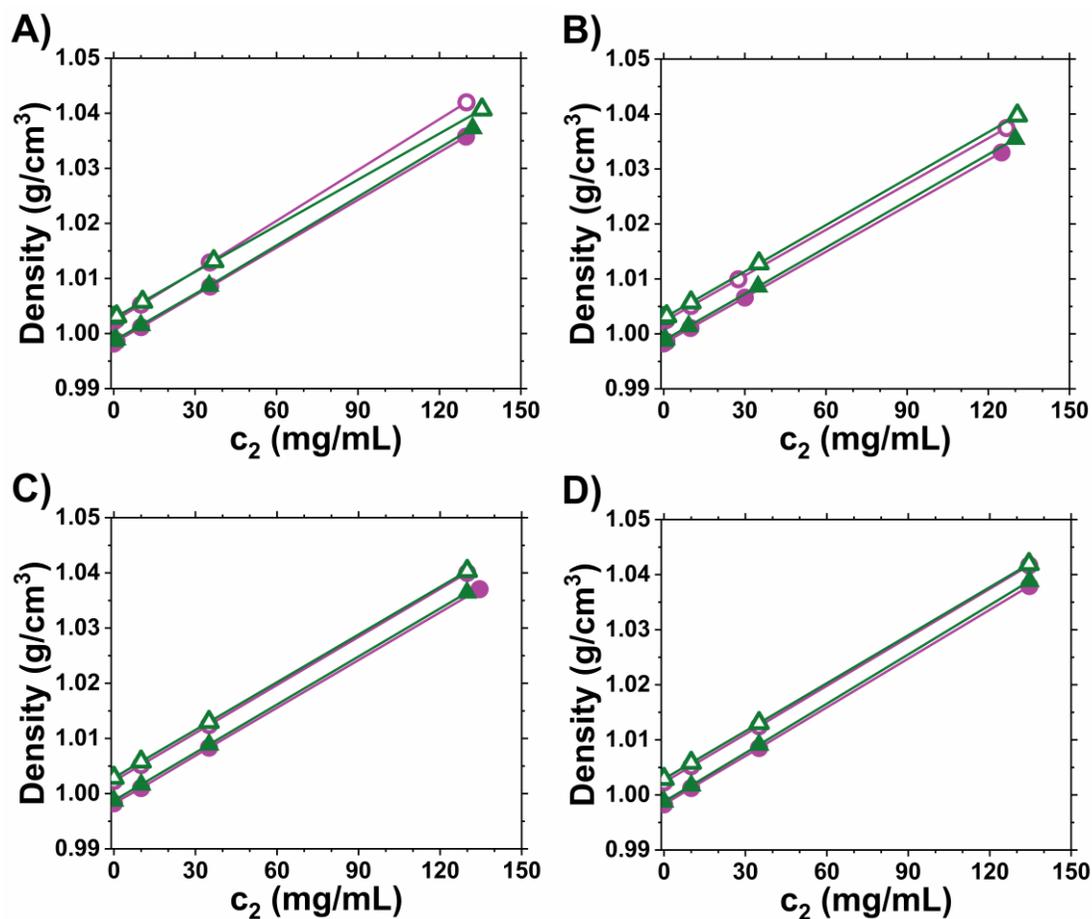


Figure 4.2: Solution density as a function of MAb concentration (c_2) for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D) in the following buffer solutions: pH 5 low IS (filled purple circles), pH 5 high IS (open purple circles), pH 6.5 low IS (filled green triangles) and pH 6.5 high IS (open green triangles).

4.2.2 Static and dynamic light scattering

SLS and DLS measurements were conducted using a Wyatt DynaPro NanoStar with a laser wavelength of 662 nm at 25.0 °C with the same methodologies as described in sections 3.2.2 and 3.2.3, respectively. High- c_2 data are not presented in terms of a single parameter (e.g., B_{22} or k_D) because multi-body interactions lead to concentration-dependent self-interactions. Instead, high- c_2 SLS data are presented and

discussed in terms of the concentration-dependent profile of the zero- q limit of the static structure factor, $S_{q=0}$, where $S_{q=0} = 1 + G_{22}c_2$. $S_{q=0}$ values above (below) 1 correspond to net attractions (repulsions).²²⁵ The corresponding high- c_2 DLS data are presented as collective diffusion coefficients (D_C values). Only DLS measurements with a polydispersity value below 0.1 are presented to ensure that measurements were not influenced significantly by the presence of aggregates or other unfiltered particles. D_C is related to the zero- q limit of the static structure factor ($S_{q=0}$) and the hydrodynamic factor ($H_{q=0}$) by $D_C = \frac{D_0 H_{q=0}}{S_{q=0}}$, where D_0 is the infinite-dilution or self-diffusion coefficient.²²¹

4.2.3 CG models, interaction potentials, B_{22} simulations and CG model parameterization

The CG models, interaction potentials, and MSOS simulations of $B_{22}/B_{22,ST}$ are largely the same as what was described in section 2.2. The error function for simulated and experimental $B_{22}/B_{22,ST}$ values was RMSD, described in section 3.2.4 and Equation 3.3. A summary of some of the most pertinent details, a description of one change in the 1bC/D model, and some MAb-specific details are provided below.

The DODECA and 1bC/D models use the same mathematical form for the interaction potentials for short-ranged non-electrostatic attractions (e.g., hydrophobic interactions and van der Waals forces) as well as electrostatic attractions and repulsions, but with different parameter values that are optimized by minimizing RMSD. The DODECA model and the 1bC/D model both contain 12 beads that each correspond to one MAb domain composed of a roughly equivalent number of amino acids. In the DODECA model those 12 beads are used to model both short-ranged non-electrostatic attractions and electrostatic attractions and repulsions. Domain bead

valences in DODECA models for a given MAb and pH are shown in Table C.1 (Appendix C). However, in addition to the 12 domain beads, the 1bC/D model contains explicit charged beads at the location of each charged atom in the MAb. The domain beads in the 1bC/D model participate in short-ranged non-electrostatic interactions, and the charged sites contribute to electrostatic attractions and repulsions. Domain beads in the 1bC/D model correspond to the same amino acid sequences in the DODECA model. Each domain bead has the same diameter and value of ϵ_{sr} . However, in contrast to the DODECA model, domain beads are placed at the geometric center of each domain based on a homology model for a given MAb, and therefore their locations can be somewhat different for each MAb or homology model. Domain beads in the 1bC/D model have hard-sphere steric repulsions and short-ranged non-electrostatic attractions with domain beads in other MAbs, but do not contribute to electrostatic attractions and repulsions.

ϵ_{sr} was chosen for each MAb (independent of pH) by matching the predicted $B_{22}/B_{22,ST}$ values where no electrostatic interactions are present (i.e., $\psi = 0$) to experimental data at high ionic strength, where electrostatic interactions are sufficiently screened such that and there is no longer a change in B_{22} with increased ionic strength and experimental data are equivalent at both pH values. After ϵ_{sr} was selected, simulations were performed sampling across a range of values for ψ to determine the optimal value (i.e., minimum RMSD value) that best reproduced the full experimental profile of $B_{22}/B_{22,ST}$ as a function of ionic strength. The uncertainty of ψ was estimated as values of ψ for which RMSD fell within 20% of the minimum RMSD. If the minimum RMSD was less than 0.05, the uncertainty of ψ was estimated as values of ψ for which the RMSD fell within 0.05 of the minimum RMSD.

The treatment of steric repulsions was different for the 1bC/D model here as compared to Chapter 2 to better account for packing in crowded systems (see the following subsection on high-concentration simulations). In the 1bC/D model, domain beads only have hard-sphere steric repulsions with other domain beads and charged beads only have hard-sphere steric repulsions with other charged beads. This approach was chosen to ensure that the charged beads were not buried, but remained surface exposed as they would be in the real protein structure. Additionally, the charged beads can impede domain beads from other MAb molecules from approaching each other when compared to a domain-level model such as DODECA.¹⁷⁹ This results in larger values of ϵ_{ST} compared to those for a purely domain-level model such as DODECA, in order to provide equivalent model results at high ionic strength where non-electrostatic attractions and steric repulsions dominate the behavior.¹⁷⁹

As described in section 2.2.2, domain bead diameters for the 1bC/D model are MAb-specific and chosen to reproduce the second osmotic virial coefficient due to only steric repulsions, $B_{22,ST}$, as calculated by an implicit solvent, all-atom MC simulation for each MAb.¹⁷⁹ The values of $B_{22,ST}$ were 9.9 mL/g for MAb 1, 9.9 mL/g MAb 2, 10.0 mL/g for MAb 3, 10.3 mL/g for MAb 4, 9.9 mL/g for MAb 5, and 10.2 mL/g for MAb 6. The corresponding domain bead diameters (σ_D) were 3.47 nm for MAb 1, 3.36 nm MAb 2, 3.35 nm for MAb 3, 3.34 nm for MAb 4, 3.30 for MAb 5, and 3.39 for MAb 6.

4.2.4 Simulations at high protein concentrations

High- c_2 (≥ 10 mg/ml) simulations were performed in the grand canonical ensemble (fixed temperature, system volume, and protein chemical potential)²²⁶ at 25 °C with a box length of 60 nm with periodic boundary conditions and a cubic box

geometry. Simulations were run with the optimal ψ value, and the lower and upper bounds of ψ based on its uncertainty. Protein chemical potential values were determined by preliminary simulations that mapped protein chemical potential to c_2 . For each MC attempt, one of the following moves was attempted: a translation, a rigid-body rotation around the MAb center of mass, a MAb insertion, or a MAb deletion. 30% of the moves were translations or rotations, and the remaining 70% were insertions or deletions. A pre-equilibration period was used to determine translation distance and rotation angle. The probability distribution of N_2 (the number of MAbs in the simulation box) was used to compute G_{22} with Equation 4.1 where V is the box volume, $\langle N_2 \rangle$ is the ensemble average value of N_2 , and $\langle N_2^2 \rangle - \langle N_2 \rangle^2$ is the variance of N_2 . To calculate $S_{q=0}$, the simplifying assumption that $M_{w,app} = M_w$ was used and Equation 3.1 was inverted to yield $S_{q=0} = 1 + G_{22}c_2$.⁵⁰

$$G_{22} = v \left(\frac{\langle N_2^2 \rangle - \langle N_2 \rangle^2}{\langle N_2 \rangle^2} - \frac{1}{\langle N_2 \rangle} \right) \quad 4.1$$

High- c_2 simulations with the DODECA model were performed using transition matrix Monte Carlo (TMMC) using the methodology described in prior work.⁵⁰ With TMMC, the simulation is biased towards sampling less likely states (values of N_2), leading to roughly even sampling in the specified bounds of N_2 (so-called “flat-histogram” sampling).²²⁷ After each cycle (5000 MC attempts in the present work), the probability distribution of N_2 was recalculated and used to rationally bias the simulation. Simulations at varying bounds of N_2 and different protein chemical potential values were performed, at least in triplicate for at least 10^4 cycles, until the combined probability distribution for the full desired protein concentration range had converged. The probability distribution of N_2 used in Equation 4.1 was reweighted

using histogram reweighting techniques to compute G_{22} as a continuous function of c_2 .^{28,227}

1bC/D simulations were performed using grand canonical MC (GCMC), the unbiased analog to TMMC. This reduced the computational burden associated with simulating a higher resolution CG model and resulted in predictions of self-interactions at a specific protein concentration that corresponded with the specified protein chemical potential. Each simulation was run in triplicate, where the number of MC attempts required to reach convergence was a function of both the protein concentration being simulated and how attractive the self-interactions were. For simulations at ~ 130 mg/mL, at least 10^7 attempts were needed, with more than 10^8 for strongly attractive systems.

4.2.5 Explicit charge equilibria algorithm

Simulations with an algorithm for explicit representation of charge fluctuations were performed in this chapter for pH 6.5 conditions where electrostatic interactions were not always properly captured in 1bC/D simulations. That algorithm was introduced and explained in further detail in Shahfar et al.⁸³ For each generated configuration, each of the His residues had its valence (either 0 or +1) randomized based on the charge probability at the simulated pH via the Henderson-Hasselbalch equation. In the context of this chapter where the pH condition of interest is 6.5, the probability that a given His residue ($pK_a = 6.04$) was charged was approximately 26%. To be specific, for each His residue, a random number between 0 and 1 was generated. The His charged site had +1 valence if the random number was less than ~ 0.26 and it had 0 valence otherwise. For GCMC simulations, His charge states were

fluctuated only for the MAb associated with the MC attempt (a translation, rotation, or insertion).

4.2.6 Mayer-averaged electrostatic energies

Similar to Chapter 3, averaged electrostatic energies for charged sites in the 1bC/D model were calculated from the MSOS (i.e., B_{22}) simulations to quantify the specific contributions to electrostatic attractions at the amino acid level and rank their relative importance. Some details are different here than in Chapter 3, so the full method is provided here for clarity.

The electrostatic energy ($\phi_i^{(k)}$) for a given charged site (i) on one protein molecule, with contributions from all other charged sites (j) on the adjacent protein, was calculated for a given configuration (k) as the sum of the electrostatic interactions (u_{ij}^{EL} in Equation 2.4) via

$$\phi_i^{(k)} = \sum_{j=1}^{N_{CS}} \frac{u_{EL,ij}^{ab(k)} + u_{EL,ji}^{ab(k)}}{2} \quad 4.2$$

where a and b refer to the two interacting MAbs, and j is the index of summation for all charged sites, N_{CS} .⁴² The Mayer-weighted average of the electrostatic energy for a given charged site ($\tilde{\phi}_i$) was defined as

$$\tilde{\phi}_i = \frac{1}{N_C} \frac{\sum_k^{N_C} (e^{-W_{22}/k_B T} - 1) \times \phi_i^{(k)}}{\sum_k^{N_C} (e^{-W_{22}/k_B T} - 1)} \quad 4.3$$

Where k is the index of summation for the number of simulated configurations N_C and W_{22} is the potential of mean force for configuration k (see also Equation 2.1).

4.3 Results and discussion

4.3.1 Experimental SLS and DLS measurements

SLS and DLS measurements were performed at dilute protein concentrations (c_2) to determine the second osmotic virial coefficient (B_{22}) and interaction parameter (k_D) for a given pH and ionic strength. SLS and DLS measurements at elevated protein concentrations (greater than 10 mg/mL) quantified the concentration-dependent self-interactions by the zero- q limit of the static structure factor ($S_{q=0}$; see section 4.2.2) and by the collective diffusion coefficient (D_c), respectively. SLS measurements were used to parameterize and validate the results of the CG molecular simulations, and DLS measurements corroborated the findings from SLS and provided information about the polydispersity of the samples. A value of B_{22} above 0, or a value of $S_{q=0}$ below 1, indicates net repulsions relative to an ideal gas mixture reference state.²²⁵ This will be further discussed below in the context of the experimental data and computational predictions.

At low- c_2 , MAb solutions were formulated at pH 5 and pH 6.5 for a range of ionic strength values that were adjusted by the addition of NaCl. The two selected pH values modulated the charge states of ionizable residues in the MAbs (primarily histidine residues at these pH values). Increasing ionic strength was used to decrease the impact of electrostatic interactions via charge screening.²⁹ At high- c_2 , measurements were performed at pH values of 5 and 6.5 and at two ionic strength conditions: (1) low ionic strength (6-10 mM) where electrostatic interactions are relatively strong, and (2) intermediate ionic strength (106-110 mM) where electrostatic interactions are weaker but might not be fully screened. c_2 values ranged from 10

mg/mL to at least 130 mg/mL. Table 4.1 lists the solution conditions for samples at high- c_2 .

Table 4.1: Summary of solution conditions for samples at high protein concentration

Formulation	Ionic Strength
pH 5, low ionic strength	6-10 mM
pH 5, intermediate ionic strength	106 mM
pH 6.5, low ionic strength	10 mM
pH 6.5, intermediate ionic strength	110 mM

SLS measurements for MAb 1 and MAb 2 were reported in prior work^{49,50}, while SLS and DLS measurements for MAb 3, MAb 4, MAb 5, and MAb 6 have not been reported previously. $B_{22}/B_{22,ST}$ values for each MAb are shown in Figure 4.3. B_{22} values were normalized by the value of B_{22} due to steric repulsions ($B_{22,ST}$) for each MAb that were calculated by an all-atom MSOS simulation.¹⁷⁹ Each of the $B_{22,ST}$ values for the four MAbs were roughly 10 mL/g (see section 4.2.3).

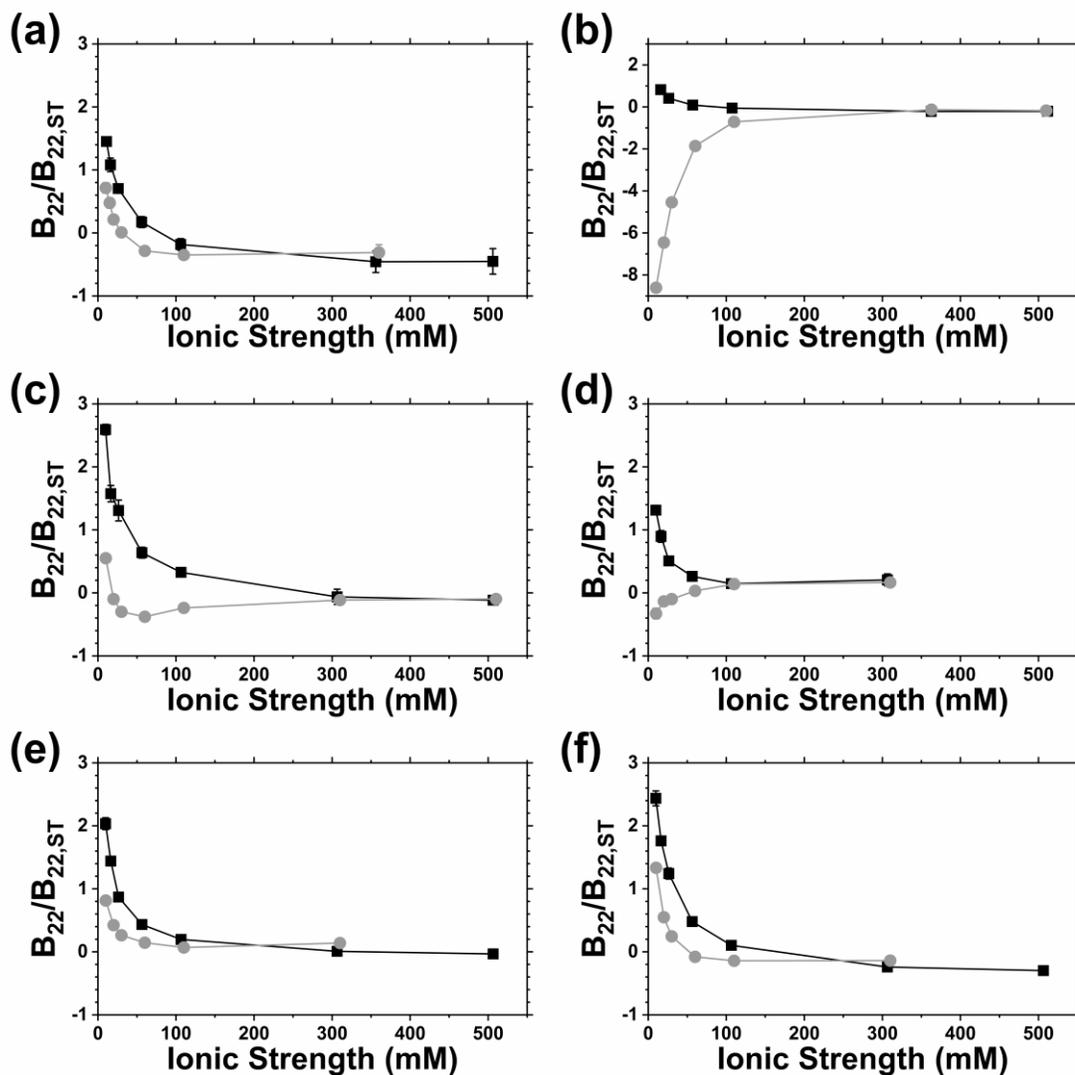


Figure 4.3: $B_{22}/B_{22,ST}$ as a function of ionic strength for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) at pH 5 (black squares) and pH 6.5 (gray circles), with lines to guide the eye.^{49,50} The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

All of the MAbs in this chapter showed strong electrostatic repulsions ($B_{22}/B_{22,ST} > 1$) at pH 5 and low ionic strength, and those repulsions decreased monotonically with increased ionic strength until the net interactions were effectively independent of pH or ionic strength (i.e., not dependent on electrostatics) at high ionic strength. This was consistent with classic “simple-colloidal” behavior for proteins with a high net surface charge due to the solution environment being far above or far below the protein’s pI.¹⁷⁹ The experimental pI values were 8.6 for MAb 1, 7.8 for MAb 2, 8.4 for MAb 3, 7.5 for MAb 4, 8.5 for MAb 5, and 9.2 for MAb 6. Using the Henderson-Hasselbalch equation with nominal pK_a values for each ionizable amino acid, the computed pI values were 9.0 for MAb 1, 8.4 for MAb 2, 9.0 for MAb 3, 8.8 for MAb 4, 8.9 for MAb 5, and 9.3 for MAb 6. At pH 6.5, the results were dependent on the nature of electrostatic interactions as a function of ionic strength. If the value of $B_{22}/B_{22,ST}$ at a given ionic strength was above (below) the value at high ionic strength (i.e., at the plateau in Figure 4.3), then electrostatic interactions were net repulsive (net attractive) for that condition. For MAb 1, MAb 5, and MAb 6 (Figure 4.3a, 4.3e, and 4.3f, respectively), electrostatic interactions were net repulsive in all cases, but less so at pH 6.5 than at pH 5. Electrostatic attractions at low ionic strength were strong for MAb 2 (Figure 4.3b) and weak for MAb 4 (Figure 4.3d). MAb 3 (Figure 4.3c) showed electrostatic repulsions at the lowest ionic strength values tested (i.e., 10 mM). However, there was a nonmonotonic dependence of B_{22} as a function of ionic strength, where electrostatic interactions were net attractive at intermediate ionic strengths (i.e., 30 to 110 mM). This suggests that MAb 3 has considerable charge anisotropy at pH 6.5 that results in attractive interactions at intermediate ionic strengths, while the net charge dominates at low ionic strength where the Debye

screening length becomes large enough to shift the self-interactions to “simple-colloidal” net repulsions.¹⁷⁹ A nonmonotonic dependence of B_{22} as a function of ionic strength has been observed in other MAbs and proteins (including the monovalent Fc-fusion in Chapter 3), consistent with that interpretation of the results for MAb 3.^{42,171,205,222}

Simultaneous low- c_2 DLS measurements of k_D for MAb 3, MAb 4, MAb 5, and MAb 6, shown in Figure 4.4, broadly corroborate the conclusions about two-body net self-interactions from the SLS measurements of $B_{22}/B_{22,ST}$ outlined above. Although DLS is influenced by hydrodynamic interactions in addition to thermodynamic interactions^{184,221}, the value of k_D is well-correlated with B_{22} for MAbs and other proteins^{41,42,54}, and the data in this work is consistent with those results (see also Figure 1.1). The qualitative conclusions from molecular simulations below would be equivalent if experimental B_{22} values had been inferred from the k_D measurements. In practice, one might prefer to use DLS to refine model parameters if it can be performed with higher throughput.

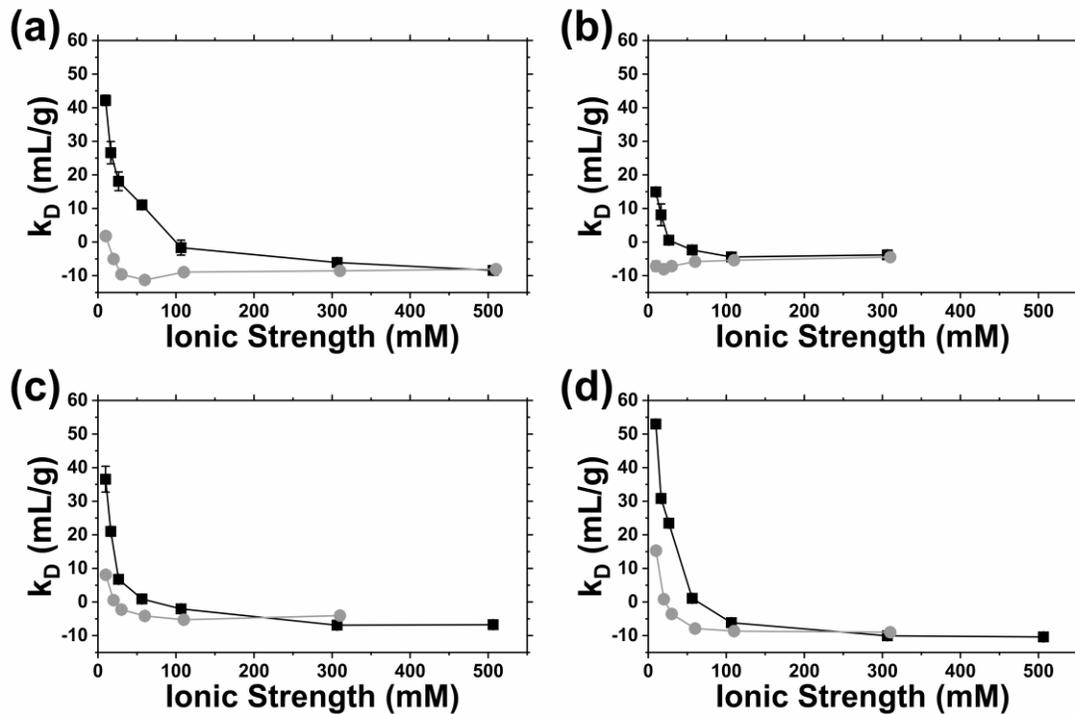


Figure 4.4: k_D as a function of ionic strength for MAb 3 (a), MAb 4 (b), MAb 5 (c), and MAb 6 (d) at pH 5 (black squares) and pH 6.5 (gray circles), with lines to guide the eye.

The corresponding apparent molecular weight ($M_{w,app}$) values from low- c_2 SLS and DLS measurements are shown in Figure 4.5. $M_{w,app}$ values at low IS in most cases were near or below the protein molecular weight (the black dashed lines in Figure 4.5), indicating a largely monodisperse solution. c_2 -dependent values of polydispersity index (PDI) from DLS support that conclusion for low- to high-concentration samples for MAb 3, MAb 4, MAb 5, and MAb 6, shown in Figure 4.6. $M_{w,app}$ generally reduced as ionic strength was increased, which can be explained by nonideal interactions between the MAbs and NaCl ions.^{35,41} MAb 2 at pH 6.5 and MAb 3 at pH 5 have low IS $M_{w,app}$ values significantly above the protein molecular

weight, indicative of a relatively small subpopulation of reversible oligomers or aggregates. MAb 2 at pH 6.5 had strongly attractive electrostatic self-interactions (i.e., $B_{22}/B_{22,ST} \ll 0$) and had a tendency for aggregation and phase separation; one or both of those factors likely accounts for the observed increase in $M_{w,app}$. Preliminary experiments for MAb 3 at pH 5 showed that the increased $M_{w,app}$ values were due to aggregates formed during sample preparation and that the values of $B_{22}/B_{22,ST}$ and k_D were approximately the same when the sample preparation procedure was improved such that the aggregates were no longer present. The high- c_2 SLS and DLS experiments discussed below for MAb 3 were performed with 3 independently prepared protein stocks that did not contain that subpopulation of aggregates.

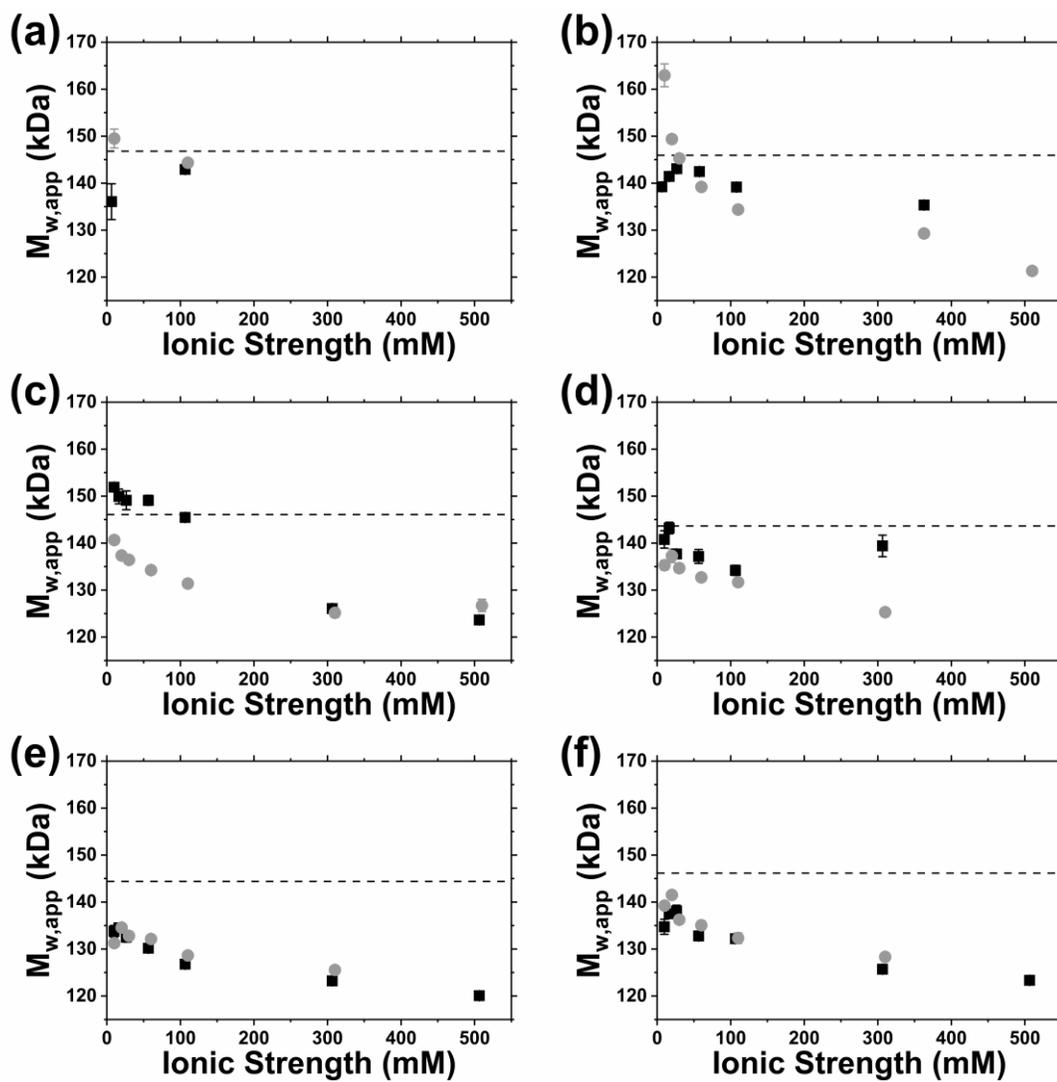


Figure 4.5: $M_{w,app}$ as a function of ionic strength for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) at pH 5 (black squares) and pH 6.5 (gray circles).^{49,50} $M_{w,app}$ values at other ionic strength conditions for MAb 1 were not reported in prior work.

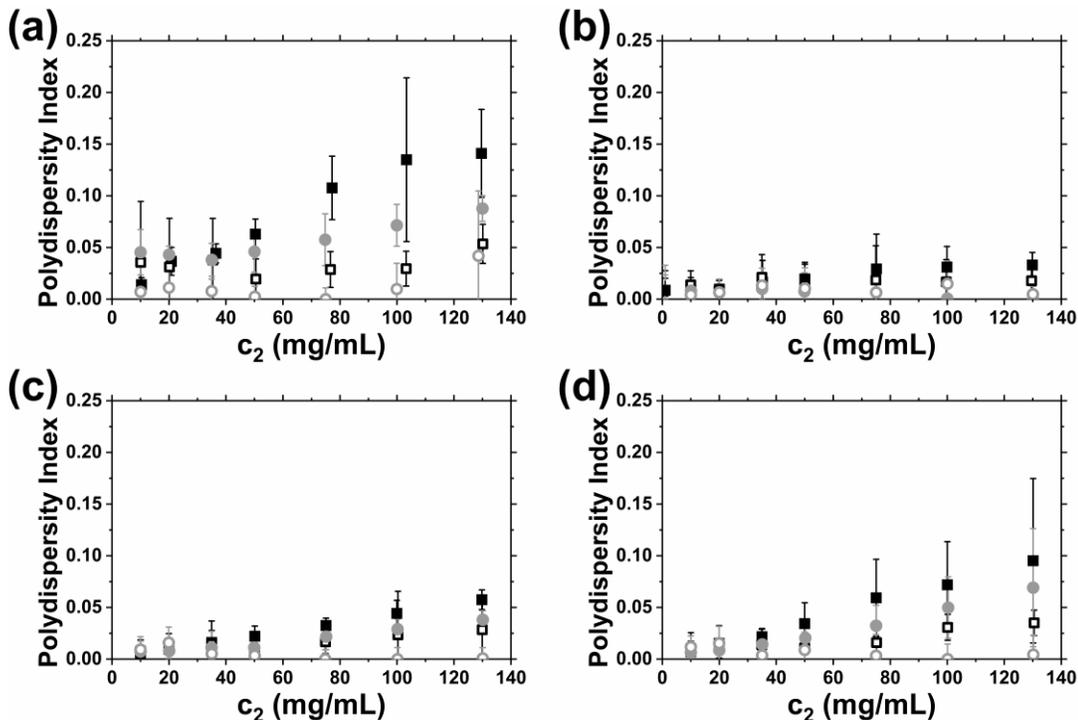


Figure 4.6: Polydispersity index (PDI) as a function of c_2 for MAb 3 (a), MAb 4 (b), MAb 5 (c), and MAb 6 (d) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles).

High- c_2 SLS measurements of $S_{q=0}$ are shown in Figure 4.7, and the G_{22} values used to calculate $S_{q=0}$ are shown in Figure C.1 (Appendix C). A reference state for steric repulsions is also included in Figure 4.7 based on simulation results from prior work.⁵⁰ Measurements at intermediate ionic strength at both pH values were attractive with respect to steric repulsions for all six MAbs. Intermediate ionic strength data also gave insight into electrostatic interactions. When $S_{q=0}$ values for low ionic strength conditions were below (above) the higher ionic strength measurements, electrostatic interactions were net repulsive (attractive). High- c_2 SLS measurements at

pH 5 and low ionic strength for all six MAbs had $S_{q=0}$ values far below one in accordance with the strong repulsive self-interactions observed at low- c_2 . The same qualitative low ionic strength electrostatic contributions were consistent at high- c_2 at pH 6.5 and low ionic strength: MAb 1, MAb 3, MAb 5, and MAb 6 had electrostatic repulsions, MAb 4 had weak electrostatic attractions, and MAb 2 had strong electrostatic attractions. $S_{q=0}$ decreased at high- c_2 values due at least in part to steric repulsions from crowding. Despite the qualitative agreement with low- c_2 data, one could not extrapolate these results with just a measurement of B_{22} or k_D as multi-body interactions and steric repulsions from crowding at high- c_2 lead to nonlinearity in $S_{q=0}$ profiles.

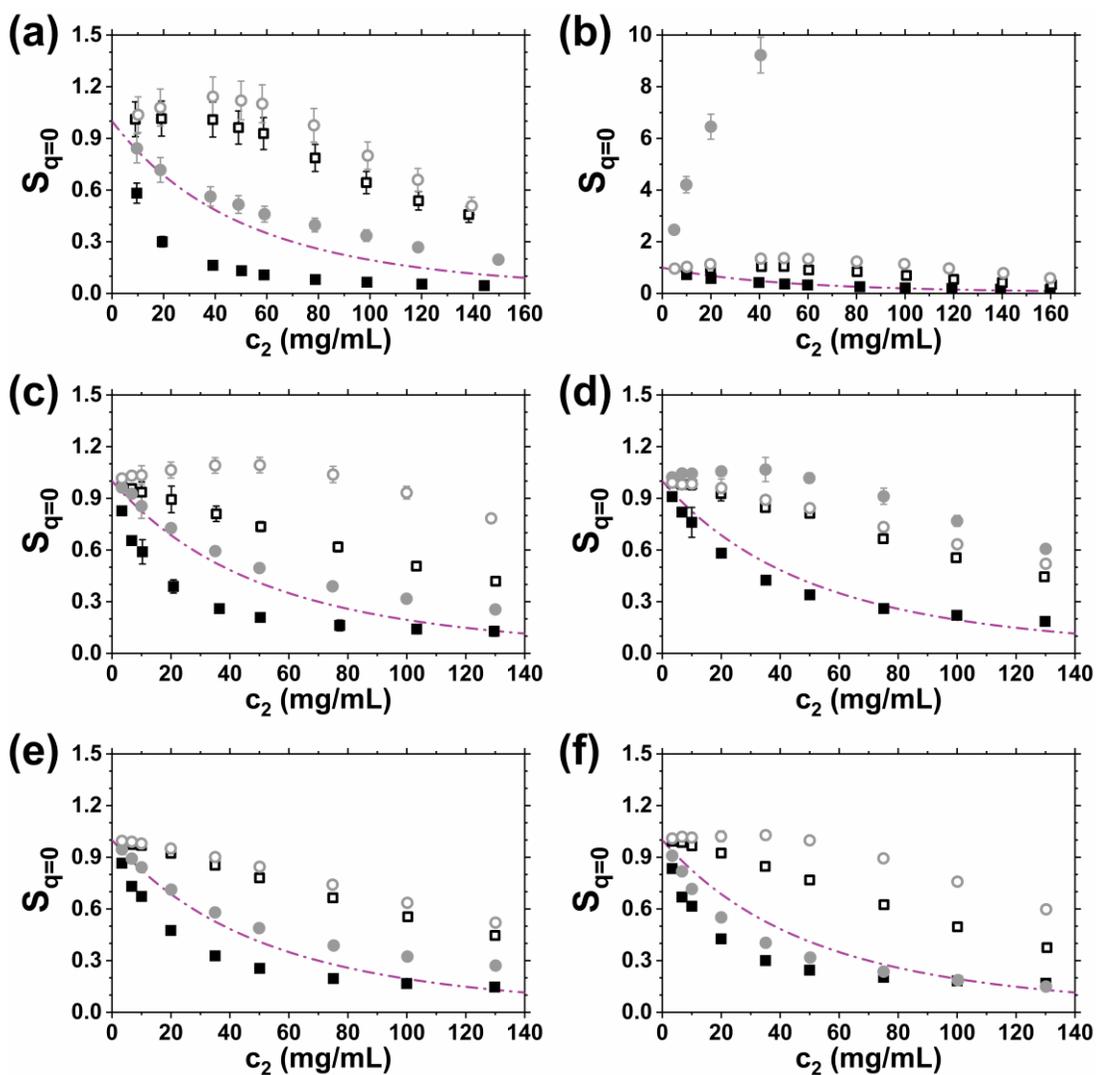


Figure 4.7: $S_{q=0}$ as a function of c_2 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles).^{49,50} The purple dot-dashed line is a steric repulsion equation of state.⁵⁰ The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

Simultaneous DLS measurements of D_C are shown in Figure 4.8 and are consistent with the conclusions from SLS outlined below, although perhaps more convoluted due to the less straightforward interpretation of D_C with respect to e.g., a steric reference state or consideration of hydrodynamic contributions. There are ongoing efforts to predict hydrodynamic contributions to D_C for MABs in the context of viscosity^{74,194,228}, however this is out of the scope of this work.

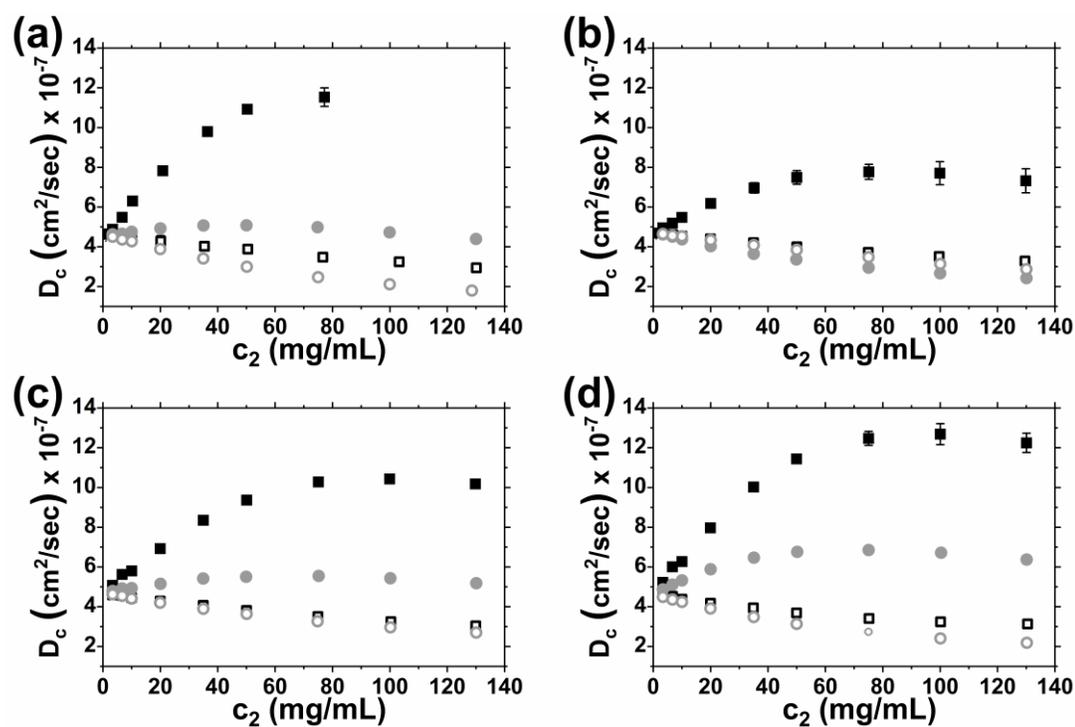


Figure 4.8: D_C as a function of c_2 for MAb 3 (a), MAb 4 (b), MAb 5 (c), and MAb 6 (d) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles). Data for MAb 3 at pH 5 at low ionic strength above 75 mg/mL are omitted due to polydispersity values significantly above 0.1.

4.3.2 High-concentration predictions of self-interactions from CG simulations

As described in section 4.2.3, MSOS simulations were performed to optimize parameters that scale the relative strength of short-ranged non-electrostatic and electrostatic interactions, ϵ_{SR} and ψ , respectively. ϵ_{SR} was set for each MAb and CG model and was chosen to match $B_{22}/B_{22,ST}$ at high ionic strength independent of pH, illustrated in Figure 4.9.

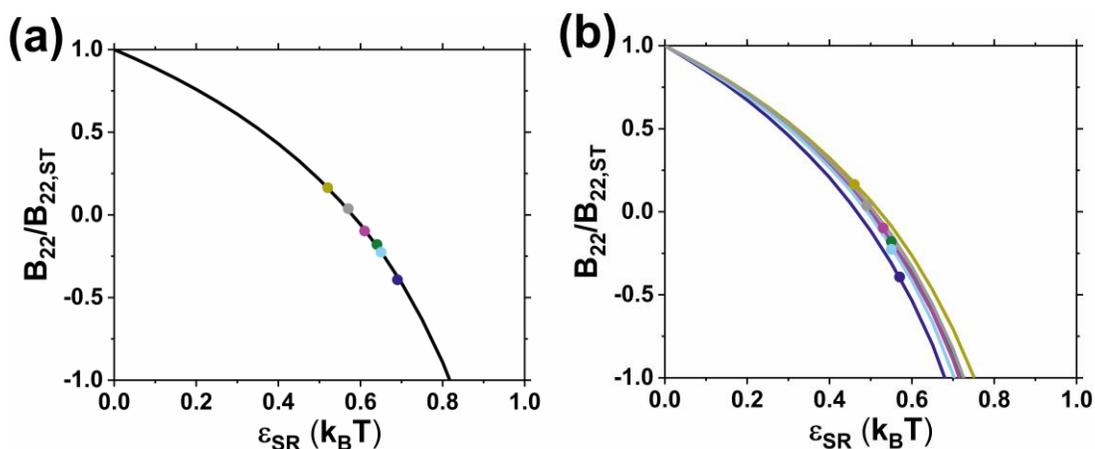


Figure 4.9: $B_{22}/B_{22,ST}$ as a function of ϵ_{SR} where no electrostatic interactions are present (i.e., $\psi = 0$) for the DODECA (a) and 1bC/D (b) models. The DODECA model is not MAb-specific, so one curve is shown in black. The results for the 1bC/D model are specific to the MAb structure and are shown in indigo for MAb 1, green for MAb 2, purple for MAb 3, yellow for MAb 4, light blue for MAb 5, and gray for MAb 6. Points in both figures correspond to the chosen ϵ_{SR} value for each MAb and match the previously listed colors.

The values of ψ were determined for each MAb, pH, and CG model to minimize the RMSD (see Equation 3.3) between the simulated and experimental profile of $B_{22}/B_{22,ST}$ as a function of ionic strength. Optimized values of ϵ_{SR} and ψ are shown in Table 4.2 for the 1bC/D and DODECA models, separated by whether the

net electrostatic self-interactions were attractive or repulsive. The corresponding figures of RMSD as a function of ψ are shown in Figure C.2 and C.3 (Appendix C) for pH 5 and pH 6.5 formulations, respectively.

Table 4.2: Optimized values of ϵ_{sr} and ψ for each MAb, pH and CG model and qualitative agreement with high- c_2 experimental data

MAb (pH)	DODECA			1bC/D		
	Pred. ^α	ϵ_{sr} ($k_B T$)	ψ ^β	Pred. ^α	ϵ_{sr} ($k_B T$)	ψ ^β
Electrostatic repulsions						
MAb 1 (5)	✓	0.69	0.45 (0.41-0.48)	~	0.57	0.54 (0.50-0.58)
MAb 1 (6.5)	✓	0.69	0.78 (0.72-0.88)	~	0.57	1.05 (0.90-1.20)
MAb 2 (5)	✓	0.64	0.54 (0.49-0.59)	✓	0.55	0.78 (0.74-0.80)
MAb 3 (5)	✓	0.61	0.66 (0.63-0.73)	✓	0.53	0.70 (0.66-0.75)
MAb 3 (6.5)	~	0.61	0.42 (0.24-0.69)	~	0.53	1.05 (0.90-1.10)
MAb 4 (5)	✓	0.52	0.41 (0.37-0.44)	✓	0.46	0.50 (0.46-0.54)
MAb 5 (5)	✓	0.57	0.58 (0.55-0.60)	✓	0.49	0.68 (0.64-0.72)
MAb 5 (6.5)	✓	0.57	0.91 (0.88-0.95)	X	0.49	0.90 (0.50-1.10)
MAb 6 (5)	✓	0.65	0.54 (0.50-0.58)	✓	0.55	0.60 (0.54-0.64)
MAb 6 (6.5)	✓	0.65	0.71 (0.66-0.76)	✓	0.55	1.02 (0.95-1.10)
Electrostatic attractions						
MAb 2 (6.5)	X	0.64	1.73 (1.66-1.78)	✓	0.55	1.125 (1.10-1.15)
MAb 4 (6.5)	X	0.52	1.48 (1.45-1.52)	✓	0.46	0.92 (0.88-0.96)

^αCheck marks (✓), tildes (~), and Xs correspond to good, fair, and poor qualitative agreement with high- c_2 experimental data, respectively. Results are shown for the 1bC/D model with nominal pK_a values.

^βThe first value listed is the optimal value, and in parentheses are the uncertainty bounds that shown as error bars in the high- c_2 predictions.

Simulated profiles of $B_{22}/B_{22,ST}$ as a function of ionic strength are shown in Figure 4.10 for the six MAbs at pH 5 conditions, where experimental results indicated “simple colloidal” behavior. These simulations were used to optimize the ψ parameter and what is shown in Figure 4.10 are the results using the optimal ψ values and their estimated uncertainties (see Table 4.2). Both the DODECA and 1bC/D models capture the “simple colloidal” electrostatic repulsions well, quantitatively or semi-quantitatively in all cases. High- c_2 simulations with the DODECA and 1bC/D models were performed with the optimal ϵ_{sr} value and three ψ values: the optimal value, and the lower and upper bounds of the estimated uncertainty (see Table 4.2 and section 4.2.3). The simulations were performed for the same formulations as the experimental measurements (see Table 4.1) and were used to predict self-interactions in the form of $S_{q=0}$ profiles as a function of c_2 , shown in Figure 4.11. Predictions from the DODECA model are presented as curves based on TMMC simulations, while predictions from the 1bC/D model were from direct GCMC simulations at particular values of c_2 to reduce computational burden (see section 4.2.3). High- c_2 predictions were also accurate at a quantitative or semi-quantitative level for all MAbs with both CG models, with the exception of the 1bC/D model for MAb 1 where the predicted electrostatic repulsions were not as strong as the experimental measurements. At pH 5, each of the MAbs was far below its pI, thus “simple-colloidal” net repulsions were expected to be dominant at low ionic strength and driven by the net charge of the protein, while at intermediate ionic strength, electrostatic repulsions were weakened by Debye screening. This behavior is captured by the DODECA model, corroborating prior work (including in Chapter 2) that has shown that domain-level models can

capture repulsive and weakly attractive (non-electrostatic) self-interactions at high- c_2 conditions.^{49,50,68}

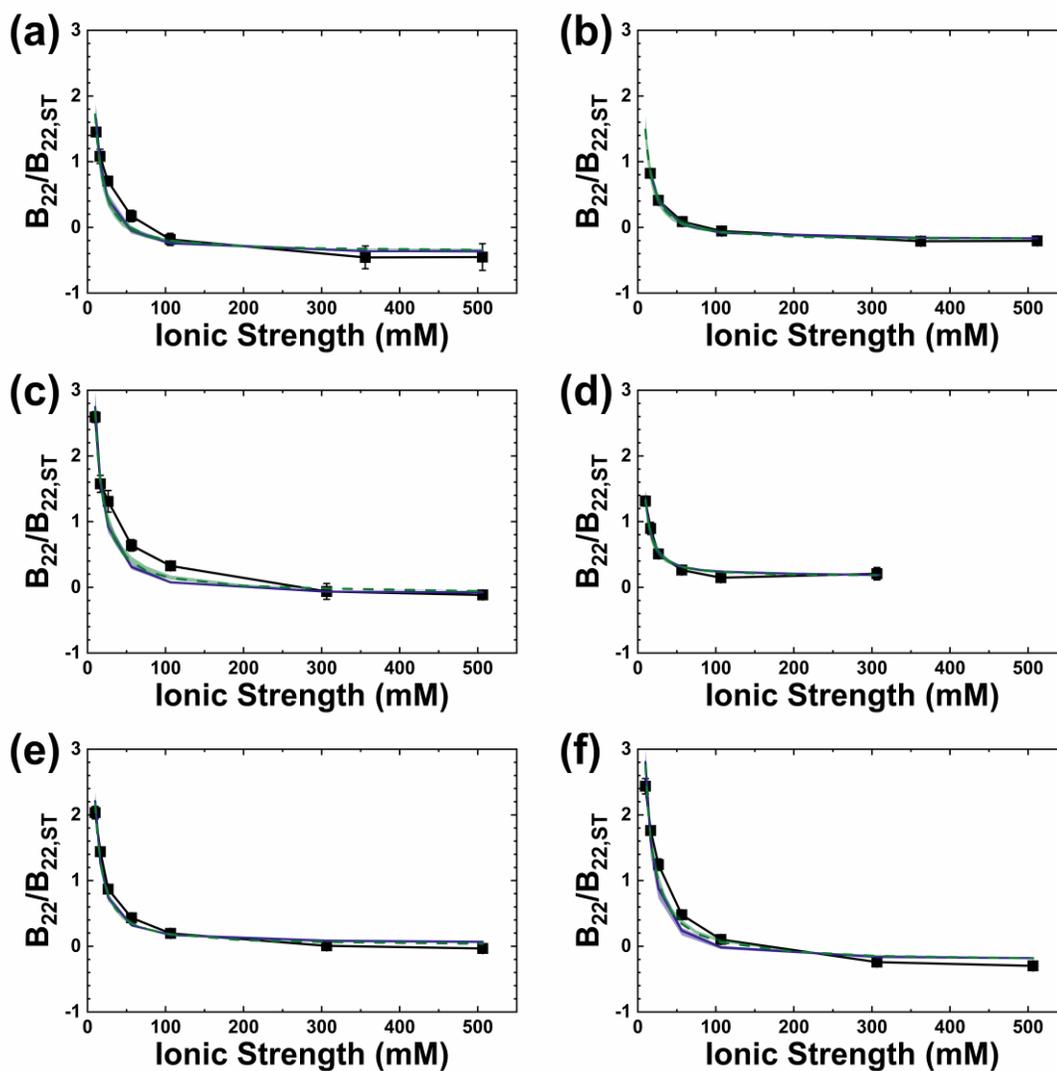


Figure 4.10: MSOS simulations of $B_{22}/B_{22,ST}$ as a function of ionic strength at pH 5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 (black squares), DODECA results are dashed green lines, and 1bC/D results are solid indigo lines. Shaded regions correspond to uncertainty in the ψ parameter from Table 4.2.

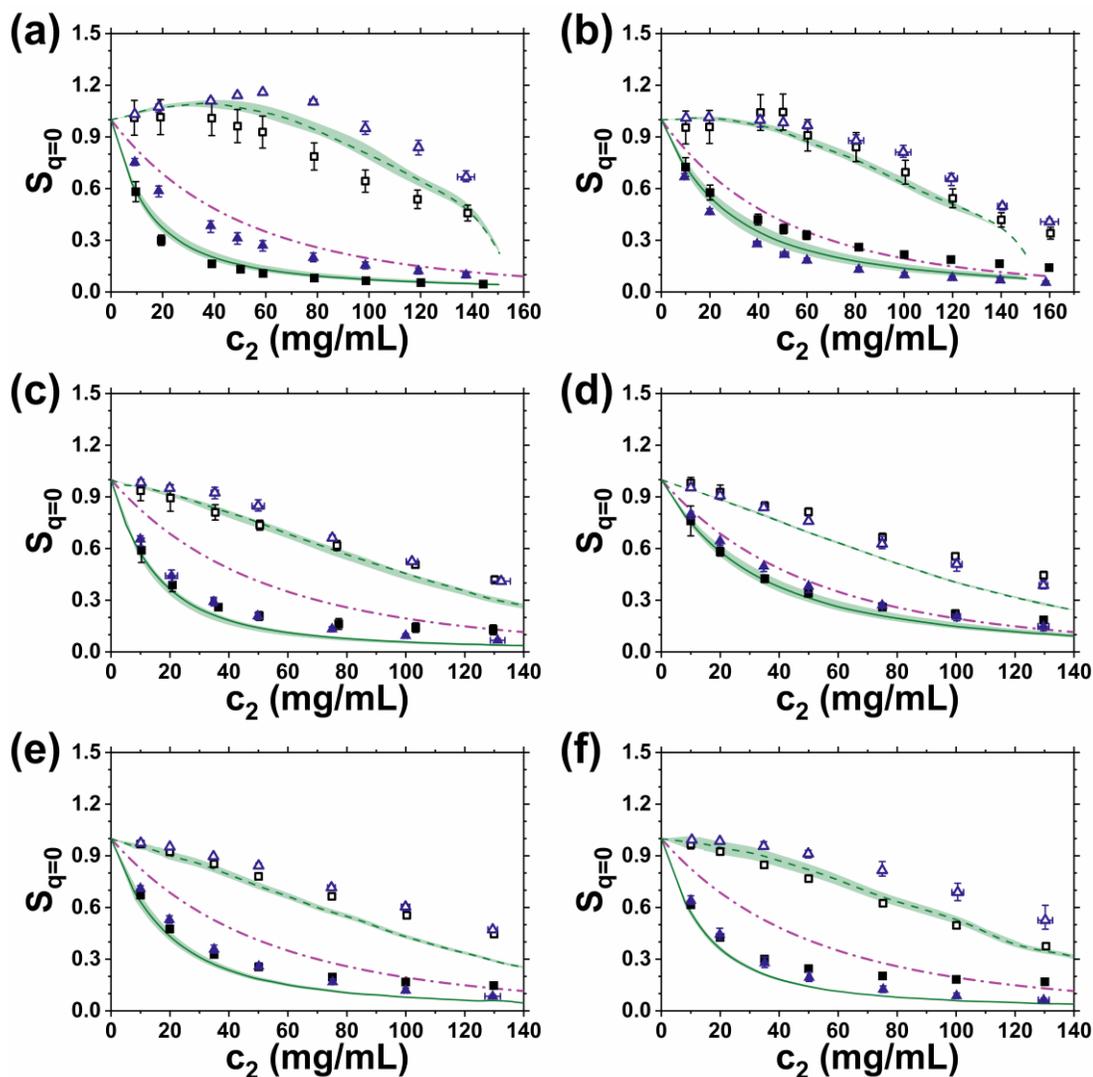


Figure 4.11: High- c_2 predictions of $S_{q=0}$ as a function of MAb concentration (c_2) at pH 5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 as filled (open) black squares for low (intermediate) ionic strength conditions. Predictions with the optimal ψ value by the 1bC/D model are filled (open) indigo triangles for low (intermediate) ionic strength. Predictions with the optimal ψ value by the DODECA model are solid (dashed) green lines for low (intermediate) ionic strength. Shaded regions for DODECA and error bars for 1bC/D correspond to uncertainty of the ψ parameter (see Table 4.2). The purple dash-dotted line is a steric-only equation of state.⁵⁰

Similar to the results presented above for pH 5 formulations, low- c_2 simulations of $B_{22}/B_{22,ST}$ as a function of ionic strength and the predicted profiles of $S_{q=0}$ as a function of c_2 for pH 6.5 conditions are shown in Figure 4.12 and 4.13, respectively. Most notably, the 1bC/D model correctly predicts significant electrostatic attractions at high- c_2 for MAb 2 (Figure 4.13b) and MAb 4 (Figure 4.13d), while the DODECA model underestimates electrostatic attractions (MAb 2) or predicts electrostatic repulsions (MAb 4). The consolidation of all charged sites in a MAb into domains (in this case, 12 domain beads) with “lumped” net charges in the DODECA model is clearly an inadequate estimation when attempting to capture electrostatic interactions that are not well described as “simple-colloidal” net repulsions. This is because the DODECA model does not have the proper structural resolution to account for molecular configurations with oppositely charged amino acids that contribute significantly to, or even dominate, electrostatic attractions. It is clear from these results that the surface charge locations that are more accurately depicted by the 1bC/D model were key to accurately model electrostatic attractions at high- c_2 . DODECA results for MAb 2 at high- c_2 are a useful example of where the domain-level model is misleading, as the 1bC/D results clearly indicate poor solution behavior due to strong attractions while the DODECA model is unable to capture that behavior. The experimental behavior corroborates this, as MAb 2 in the pH 6.5, low ionic strength condition displays opalescence and ultimately liquid-liquid phase separation on longer time scales at protein concentrations above ~40 mg/mL. It is important for a predictive tool to avoid “false negatives”, in this case predictions that do not accurately reproduce strong attractions at high- c_2 like those for MAb 2.

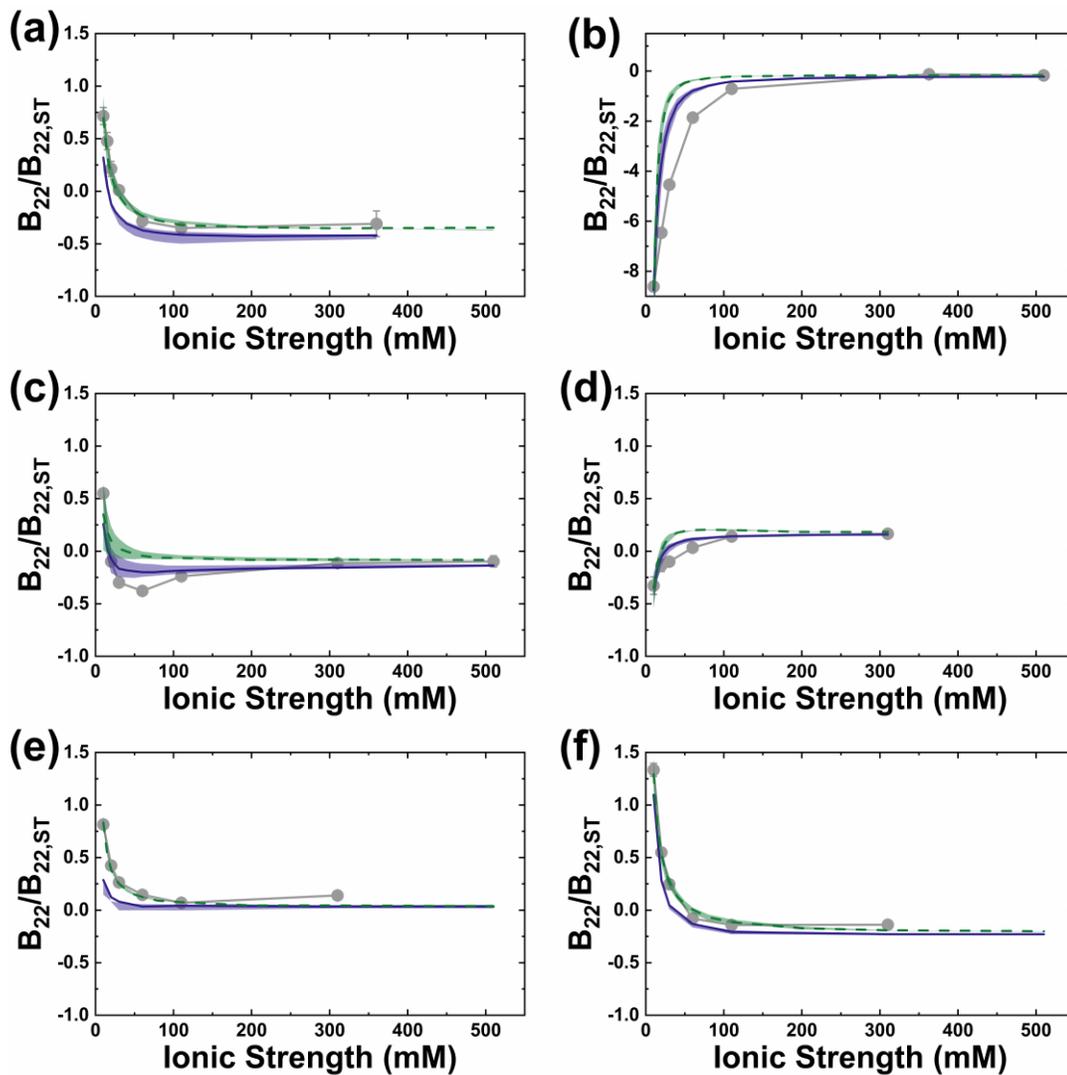


Figure 4.12: MSOS simulations of $B_{22}/B_{22,ST}$ as a function of ionic strength at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 (gray circles), DODECA results are dashed green lines, and 1bC/D results are solid indigo lines. Shaded regions correspond to uncertainty in the ψ parameter from Table 4.2. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

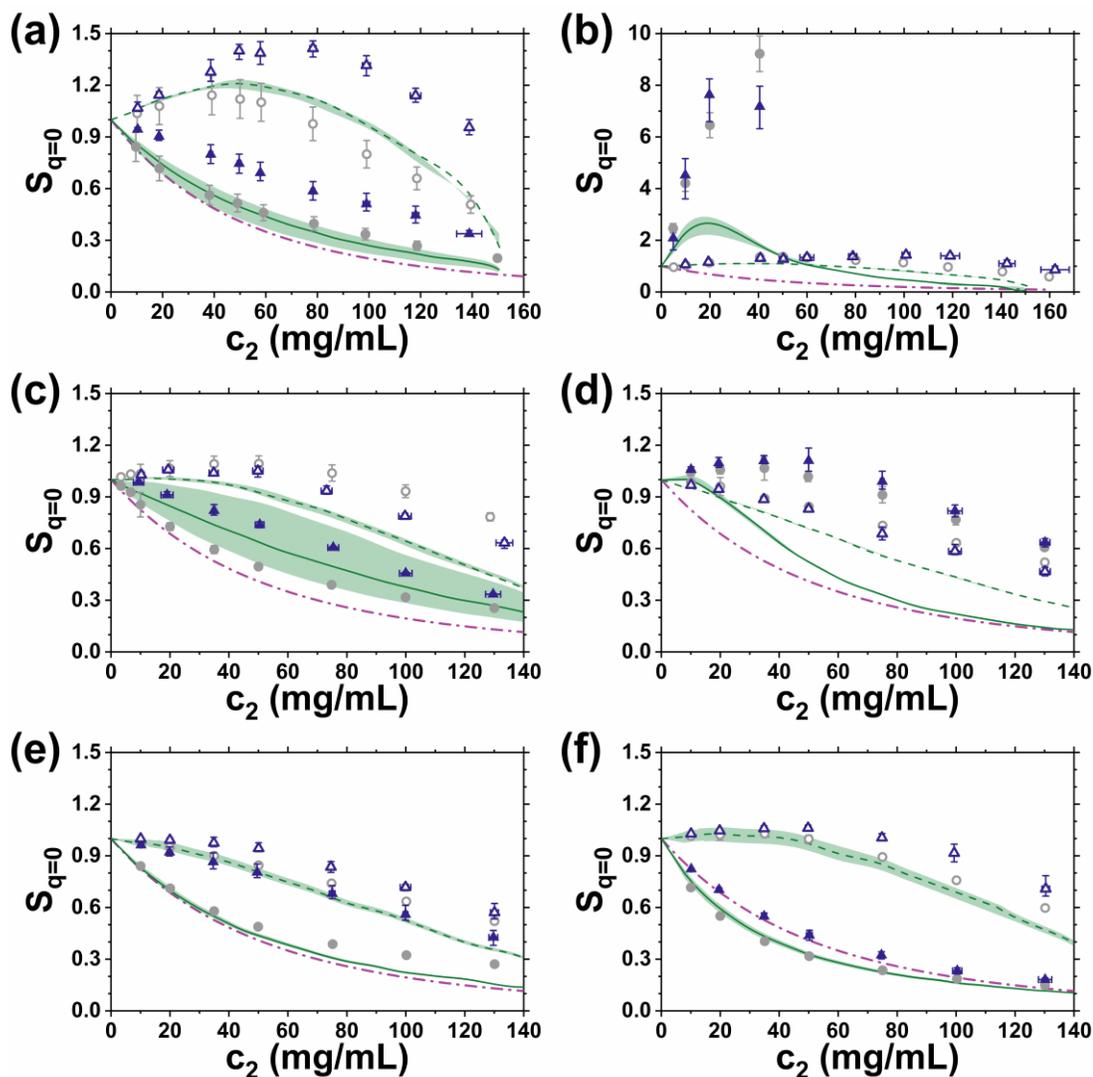


Figure 4.13: High- c_2 predictions of $S_{q=0}$ as a function of MAb concentration (c_2) at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are reproduced from Figure 4.3 as filled (open) gray circles for low (intermediate) ionic strength conditions. Predictions with the optimal ψ value by the 1bC/D model are filled (open) indigo triangles for low (intermediate) ionic strength. Predictions with the optimal ψ value by the DODECA model are solid (dashed) green lines for low (intermediate) ionic strength. Shaded regions for DODECA and error bars for 1bC/D correspond to uncertainty of the ψ parameter (see Table 4.2). The purple dash-dotted line is a steric-only equation of state.⁵⁰ The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

Qualitative accuracy of the high- c_2 predictions (i.e., either “good”, “fair”, or “poor”) are summarized in Table 4.2. While the 1bC/D model was clearly preferable to the DODECA model in cases with net attractive electrostatic interactions (i.e., MAb 2 and MAb 4), the DODECA model was equal to or preferable to the 1bC/D model for systems with net repulsions (i.e., MAb 1, MAb 3, MAb 5, and MAb 6). High- c_2 predictions at pH 6.5 from the DODECA model were nearly quantitative for MAb 1, MAb 5, and MAb 6 (Figure 4.13a, 4.13e, and 4.13f, respectively) in comparison with the 1bC/D model which underestimated electrostatic repulsions at low and intermediate ionic strength values. 1bC/D predictions were poor for MAb 5 where the model predicted quite mild electrostatic repulsions that were particularly inconsistent with the experimental data. The results are similar between the two CG models for MAb 3 at pH 6.5 (Figure 4.13c), although there was large uncertainty in the ψ parameter for the DODECA model. The broad range of predicted profiles would make the DODECA model less useful from a quantitative rather than qualitative perspective. This was in part due to the choice of matching the entire B_{22} vs. ionic strength profile when determining ψ rather than only a particular ionic strength of interest for high- c_2 predictions as neither model could quantitatively capture the nonmonotonic behavior of B_{22} . For MAb 3, the ψ value that matched the low ionic strength $B_{22}/B_{22,ST}$ value for DODECA yielded the best high- c_2 prediction (the lower bound of the $S_{q=0}$ prediction).

4.3.3 Methods to improve representation of charge equilibria

As noted above, many of the high- c_2 predictions from the 1bC/D model were lacking at pH 6.5 formulations (see Figure 4.13). For MAb 1, MAb 3, MAb 5, and MAb 6, there was no value of ψ that could capture the strength of electrostatic

repulsions at low ionic strength and low- c_2 (i.e., the $B_{22}/B_{22,ST}$ value at 10 mM ionic strength). The nonmonotonic $B_{22}/B_{22,ST}$ profile as a function of ionic strength for MAb 3 at pH 6.5 was not captured by either model. Although the 1bC/D model could properly capture the low ionic strength $B_{22}/B_{22,ST}$ values for MAb 2 and MAb 4, which both had net attractive electrostatic interactions, the ionic strength dependence could be improved and the high- c_2 predictions for MAb 2 had high error and required an exceptional amount of computational resources compared to the other conditions.

These challenges could be attributed to the treatment of charge states in the model; His residues were uncharged at pH 6.5 in the 1bC/D model, and the corresponding charged sites were excluded, while each His residue in the DODECA model was assigned a partial charge based on the Henderson-Hasselbalch equation and contributed to the overall domain bead charge.²⁸ This is of particular practical relevance as the nominal pK_a value of His is approximately 6.0, and is within one pH unit of the pH of most commercially-available high- c_2 antibody therapies.^{3,229} Two different approaches to more realistically represent charge equilibria with the 1bC/D model were evaluated: 1) assigning pK_a values of each ionizable residue individually, rather than assuming nominal pK_a values for all residues of a given amino acid type, and 2) explicitly modeling dynamic charge equilibria in the simulations.

Unique pK_a values for each ionizable residue were estimated with PROPKA¹⁸⁶, one of many algorithms available for this purpose.²³⁰⁻²³² The PROPKA algorithm accounts for the impact of the local environment in the protein (i.e., the location and identity of proximal residues) on the pK_a of a given sidechain. This technique is frequently used in molecular dynamics simulations of proteins.^{106,233,234} At pH 6.5, PROPKA estimated that a subset of the His residues had pK_a values greater

than 6.5, which led to the addition of roughly 20 positive charged sites for each MAb, compared to ~250 total charged sites. The algorithm for explicit charge fluctuations was developed in prior work and is described in section 4.2.5.⁸³ The valence of the His charged sites could fluctuate between +1 and 0 per configuration which was also implicitly dependent on the local environment in a given configuration. At pH 6.5, no other ionizable residues have nominal pK_a values close enough to 6.5 for the potential fluctuations to be significant. While applying PROPKA pK_a values did not increase the computational burden of the simulations, explicit fluctuating charges increased the computational burden by approximately 2 times as the sampling space was larger. More configurations were needed to converge the value of $B_{22}/B_{22,ST}$ or $S_{q=0}$ because the simulations also had to sample the distribution of valence microstates for all of the His residues.⁸³ The three different methods will be referred to hereafter as 1) “static nominal”, 2) “static PROPKA”, and 3) “fluctuating nominal”. Preliminary simulations where the charge states fluctuated with pK_a values calculated by the PROPKA algorithm (which would be termed “fluctuating PROPKA”) suggested that the two methods combined did not provide any additional improvements.

A comparison of the results from 1bC/D MSOS simulations between the different methods is shown in Figure 4.14. ψ values and a summary of qualitative agreement with the low- c_2 experimental data are shown in Table 4.3. RMSD as a function of ψ is shown in Figure C.4 (Appendix C). In cases where the static nominal simulations could not reproduce the strength of electrostatic repulsions at low ionic strength (i.e., MAb 1, MAb 3, MAb 5, and MAb 6), both the static PROPKA and fluctuating nominal methods were able to capture the strength of those repulsions quantitatively. Both the static PROPKA and fluctuating nominal methods increased

the average net valence of the MAbs by adding positive charges to the proteins, thus “simple-colloidal” net repulsions would be expected to increase in strength. The nonmonotonic profile of $B_{22}/B_{22,ST}$ as a function of ionic strength for MAb 3 was captured by the fluctuating nominal method, although there was quite high sensitivity with respect to the ψ parameter which required many simulations to optimize. $B_{22}/B_{22,ST}$ values for MAb 2 were not considerably different between the three methods and some of the apparent differences in Figure 4.14 are due to sensitivity to the exact value of the ψ parameter. For MAb 4, the static PROPKA method erroneously produced a nonmonotonic profile, while the fluctuating nominal simulations matched the ionic strength dependence more accurately than the static nominal simulations. Overall, the static PROPKA simulations performed better than the static nominal simulations only for net repulsive electrostatic self-interactions, while the fluctuating nominal simulations were also equal or better for net attractive electrostatic self-interactions, but with significantly increased computational burden.

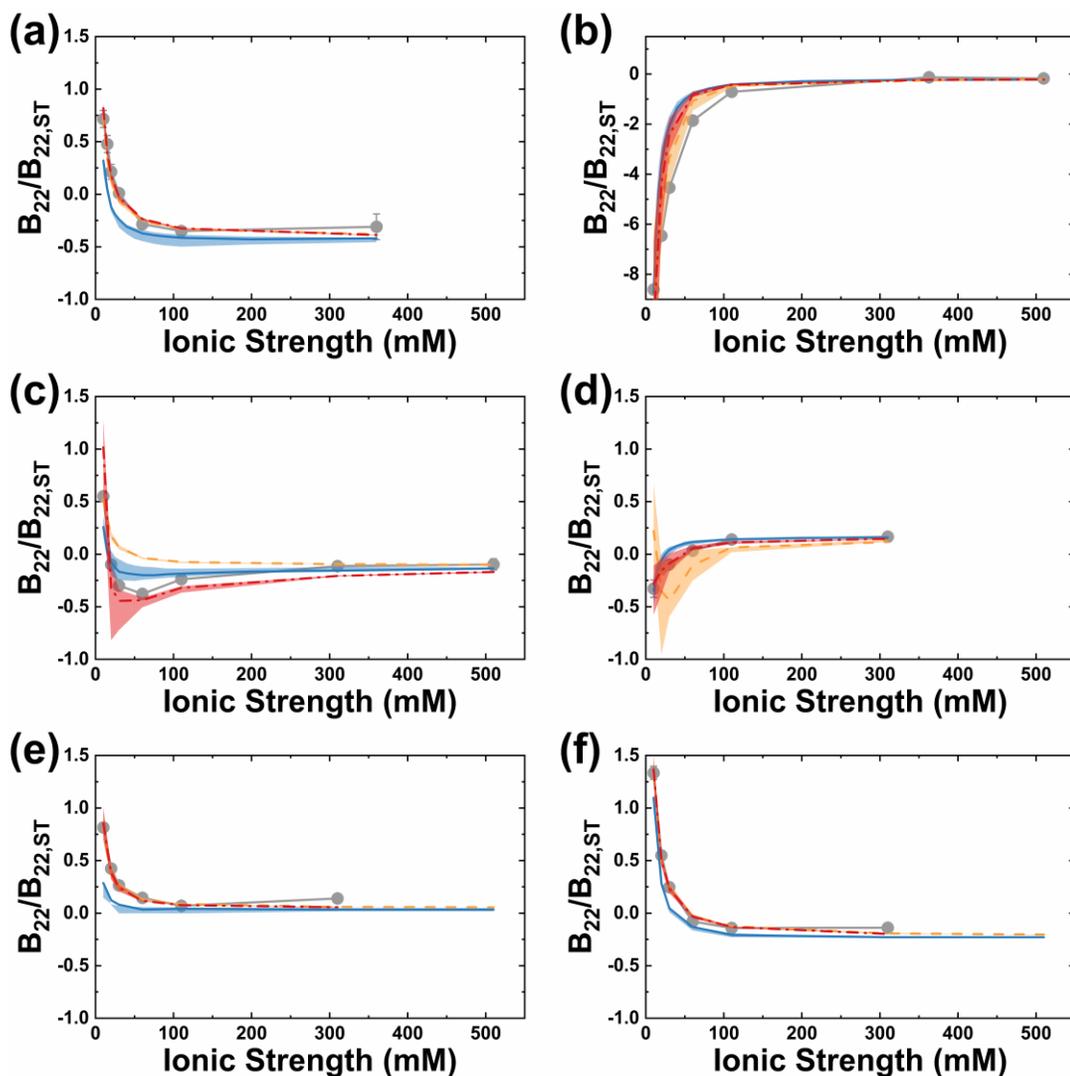


Figure 4.14: MSOS simulations of $B_{22}/B_{22,ST}$ as a function of ionic strength at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are gray circles (reproduced from Figure 4.3), static nominal results are solid blue lines (reproduced from Figure 4.12), static PROPKA results are dashed orange lines, and fluctuating nominal results are dash-dotted red lines. Shaded regions correspond to uncertainty in the ψ parameter from Table 4.3. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

Table 4.3: Optimized values of ψ for each MAb and charge equilibria model and qualitative agreement with low- c_2 experimental data

MAb	Static nominal		Static PROPKA		Fluctuating nominal	
	Pred. ^α	ψ ^β	Pred. ^α	ψ ^β	Pred. ^α	ψ ^β
Electrostatic repulsions						
MAb 1	~	1.05 (0.90-1.20)	✓	0.475 (0.45-0.50)	✓	0.74 (0.70-0.76)
MAb 3	~	1.05 (0.90-1.10)	~	0.40 (0.38-0.43)	✓	1.31 (1.30-1.33)
MAb 5	X	0.90 (0.50-1.10)	✓	0.56 (0.50-0.62)	✓	0.80 (0.70-0.90)
MAb 6	~	1.02 (0.95-1.10)	✓	0.52 (0.50-0.56)	✓	0.68 (0.64-0.72)
Electrostatic attractions						
MAb 2	✓	1.125 (1.10-1.15)	✓	1.225 (1.20-1.25)	✓	1.155 (1.13-1.18)
MAb 4	✓	0.92 (0.88-0.96)	X	1.26 (1.22-1.28)	✓	1.125 (1.10-1.15)

^αCheck marks (✓), tildes (~), and Xs correspond to good, fair, and poor qualitative agreement with low- c_2 experimental data, respectively.

^βThe first value listed is the optimal value, and in parentheses are the uncertainty bounds that shown as error bars in the low- c_2 predictions.

An analogous comparison of the high- c_2 predictions of $S_{q=0}$ between the three charge representation methods for the 1bC/D model is shown in Figure 4.15. The analysis was limited to low ionic strength conditions, where the differences in the treatment of electrostatic interactions were the most impactful, and where the error in high- c_2 predictions was generally the largest. Qualitative agreement with the experimental data for each method is shown in Table 4.4. Predictions for MAb 1, MAb 5, and MAb 6 were improved by both the static PROPKA and the fluctuating nominal methods with statistically indistinguishable results for all three MAbs.

Predictions were improved for MAb 3 as well, but with significantly stronger repulsions than even the experimental data with the fluctuating nominal method. Overall, predictions of high- c_2 self-interactions where electrostatic interactions were net repulsive were improved by both alternate methods of considering charge equilibria.

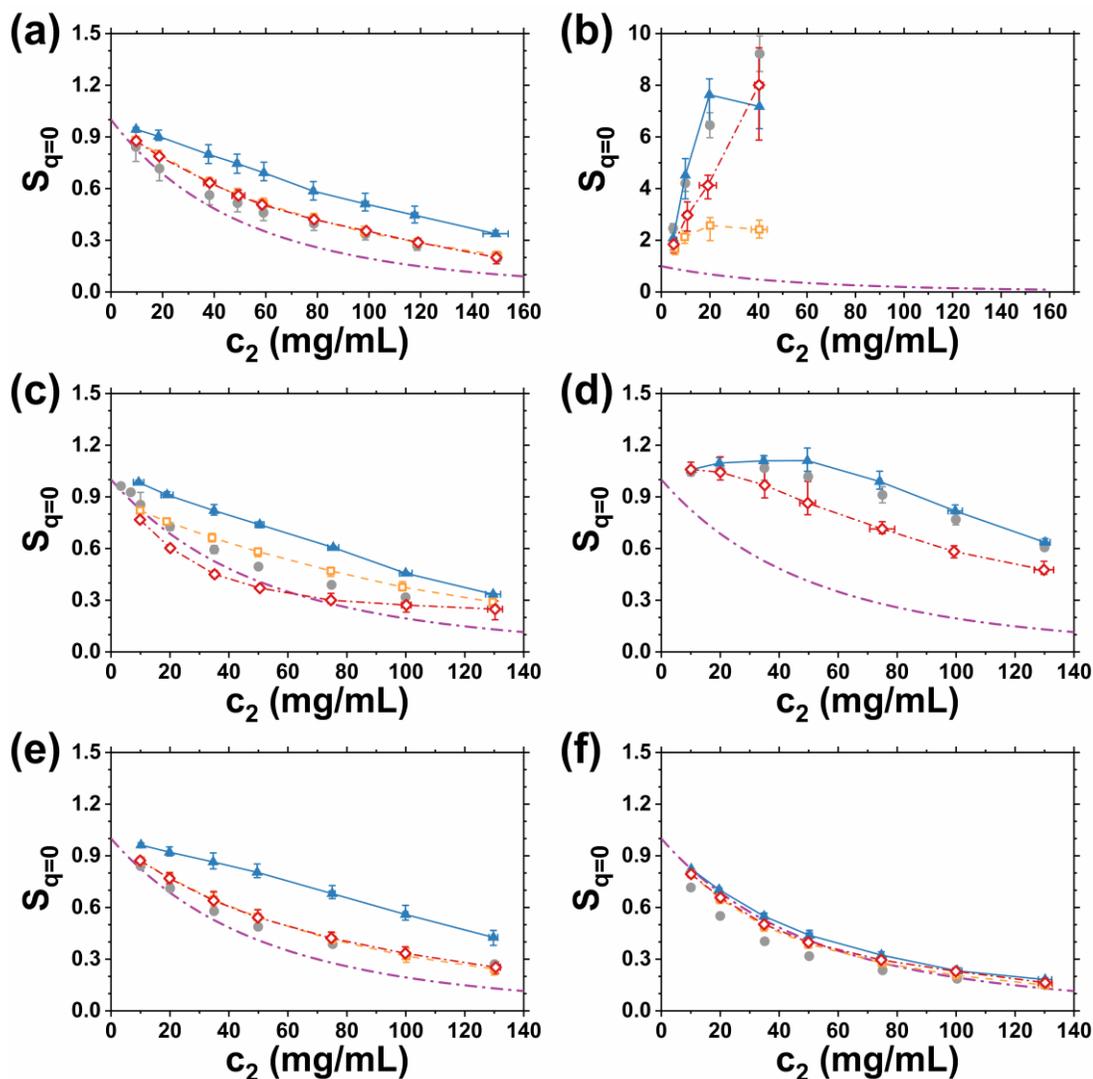


Figure 4.15: High- c_2 predictions of $S_{q=0}$ as a function of MAb concentration (c_2) at the pH 6.5 low ionic strength formulation for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f). The experimental data are filled gray circles (reproduced from Figure 4.3) and predictions with the static nominal method are filled blue triangles (reproduced from Figure 4.13). Predictions from the static PROPKA and fluctuating nominal methods are shown as open orange squares and open red diamonds, respectively. Lines connecting the predictions are guides to the eye and match the formatting from Figure 4.14. Error bars correspond to uncertainty of the ψ parameter (see Table 4.3). The purple dash-dotted line is a steric-only equation of state.⁵⁰ The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

Table 4.4: Qualitative agreement with high- c_2 experimental data for each MAb and charge equilibria method

MAb	Static nominal	Static PROPKA	Fluctuating nominal
	Pred. ^α	Pred. ^α	Pred. ^α
Electrostatic repulsions			
MAb 1	~	✓	✓
MAb 3	~	✓	~
MAb 5	X	✓	✓
MAb 6	✓	✓	✓
Electrostatic attractions			
MAb 2	✓	X	✓
MAb 4	✓	N/A	~

^αCheck marks (✓), tildes (~), and Xs correspond to good, fair, and poor qualitative agreement with high- c_2 experimental data, respectively.

Predictions of high- c_2 self-interactions where electrostatic interactions were net attractive were not particularly improved when compared to the static nominal method that was used in the previous section. For MAb 2, the static PROPKA method predicted much weaker electrostatic attractions and the results were not representative of the experimental data. The predictions with the fluctuating nominal method were qualitatively similar to the static nominal method in that they predicted strong electrostatic attractions, but the computational burden was much higher (approximately an order of magnitude more) compared to the static nominal method. For MAb 4, the nominal PROPKA method was not attempted as the simulated low- c_2 $B_{22}/B_{22,ST}$ profiles were inconsistent with the experimental data. The high- c_2 predictions with the nominal fluctuating method were significantly worse than those

with the static nominal method. The predicted high- c_2 net self-interactions were less attractive than the results from the static nominal simulations or experimental measurements, and the net electrostatic interactions were approximately neutral such that the $S_{q=0}$ values did not change significantly at higher ionic strength (see Figure C.5 in Appendix C).

4.3.4 Context-driven CG model and charge equilibria method selection

In summary, the ideal CG model and for the 1bC/D model, the ideal method used to represent charge equilibria, was context dependent with no clear universal setup to capture low- to high- c_2 net self-interactions regardless of their magnitude or sign (i.e., repulsive or attractive). In cases with “simple-colloidal” electrostatic net repulsions, one can reliably apply the DODECA model, or for a significant increase in computational burden, the 1bC/D model with the static PROPKA or nominal fluctuating charge representation methods. However, those methods all had limitations in cases with net attractive electrostatic self-interactions. The static nominal method with the 1bC/D model works well for cases with net attractive electrostatic self-interactions but performs very poorly for cases with “simple-colloidal” electrostatic net repulsions at pH 6.5. With experimental $B_{22}/B_{22,ST}$ values, a conditional approach to predicting high- c_2 net self-interactions would be successful, at least with the dataset in this Chapter.

Without experimental data (e.g., if used as a screening tool in early-stage drug development), the DODECA model would be prone to false negatives as it erroneously predicted electrostatic repulsions at high- c_2 and could not capture the nonmonotonic value of $B_{22}/B_{22,ST}$ as a function of ionic strength at low- c_2 for MAb 3 at pH 6.5. There were other cases in Chapter 2 where the DODECA model could not

capture electrostatic attractions at low- c_2 as well. The static PROPKA method could not capture the nonmonotonic low- c_2 $B_{22}/B_{22,ST}$ values for MAb 3 either and struggled generally in cases with net attractive electrostatic interactions. The static nominal method would be prone to false positives as it could not consistently capture the strength of electrostatic repulsions, particularly in the case of MAb 5 that might be flagged as potentially problematic. The fluctuating nominal method was the only universally successful simulation setup at low- c_2 , but the high- c_2 predictions for MAb 3 and in cases with net attractive electrostatic interactions were equal to or worse than those for the static nominal method. Further improvements would be needed for consistently reliable qualitative low- to high- c_2 predictions.

4.3.5 Ranking average electrostatic energies of charged sites for strong electrostatic attractions

In this chapter, the net self-interactions for MAb 2 at pH 6.5 were by far the most problematic. The net electrostatic interactions were strongly attractive, which is associated with many poor biophysical behaviors, some of which were observed experimentally (e.g., phase separation and opalescence).⁴⁹ To further investigate the electrostatic origins of those attractions, the Mayer-weighted average electrostatic energy for each charged site ($\tilde{\phi}_i$) in 1bC/D MSOS simulations (with the static nominal method) of MAb 2 was calculated and is shown in Figure 4.16a for pH 5 at low ionic strength and Figure 4.16b for pH 6.5 at low ionic strength. His residues are shown for pH 5 but omitted for pH 6.5 as they were uncharged in the model. In recent work, a similar procedure was used with the 1bC/D model to predict MAb variants that would reduce strongly attractive electrostatic self-interactions, which were confirmed by SLS and DLS measurements.⁴⁴ The results shown in this section are similar in concept to

those in section 3.3.3, but on the individual amino acid level rather than based on specific charge-charge pairwise interactions.

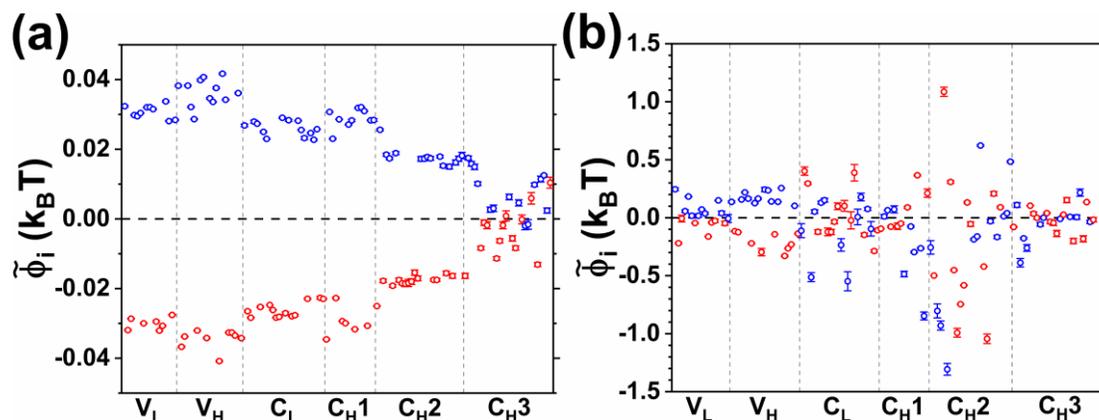


Figure 4.16: Mayer-weighted average electrostatic energy ($\tilde{\phi}_i$) for charged sites in MAb 2 at pH 5 at low ionic strength (a) and pH 6.5 at low ionic strength (b). Positively (negatively) charged atoms are shown in blue (red). Domain boundaries are vertical dashed grey lines with labels on the x-axis. The black dashed line at $\tilde{\phi}_i = 0$ denotes zero net energy. Error bars are 95% confidence intervals from 3 independent simulations.

The results for MAb 2 at pH 5 (Figure 4.16a) show that all Mayer-weighted energies were small with respect to $k_B T$, thus specific, strongly attractive electrostatic interactions were not necessarily favored. This was expected given the repulsive B_{22} value at pH 5, which suggested that more general “simple-colloidal” net repulsions based on overall net surface charge were dominant. Results for other MABs at pH 5 and pH 6.5 were analogous (i.e., small values of $\tilde{\phi}_i$ with respect to $k_B T$), consistent with B_{22} values that were repulsive or at most were weakly attractive and thus were not as dependent on specific molecular configurations with strong electrostatic attractions between particular residues.

However, for MAb 2 at pH 6.5, where B_{22} indicated strongly attractive self-interactions, there were several residues that had large, negative values of $\tilde{\phi}_i$ (note the difference in scale between panels a and b). Charged sites in the C_{H2} domain played the largest role in the strong electrostatic attractions compared to the other domains, although some charged sites in other domains had notably negative values of $\tilde{\phi}_i$ as well. Table 4.5 lists the amino acids with the most negative average electrostatic energies at pH 6.5. This analysis gives insight into the amino acids that are most relevant to the specific configurations that lead to electrostatic attractions and could help identify promising protein variants.

Table 4.5: Ranking of the 10 amino acids with the most negative $\tilde{\phi}_i$ values for 1bC/D MSOS simulations of MAb 2 at pH 6.5

Charged residue	$\tilde{\phi}_i$ (k _B T)
Arg474	-1.31
Glu512	-1.04
Glu488	-0.99
Lys467	-0.93
Arg436	-0.85
Lys465	-0.80
Asp489	-0.75
Glu491	-0.58
Lys183	-0.55
Lys126	-0.51

4.4 Summary and conclusions

Self-interactions for six MABs in a range of industrially relevant solution conditions were experimentally quantified with SLS at low and high- c_2 conditions. For four of the MABs, DLS was used as an orthogonal measure of self-interactions that corroborated findings from SLS measurements. A domain-level CG molecular model, DODECA, and a higher resolution hybrid model, 1bC/D, were compared based on their accuracy in predicting high- c_2 net self-interactions in the form of zero- q static structure factor ($S_{q=0}$) profiles from SLS. When low- c_2 behavior was “simple-colloidal” (i.e., electrostatic repulsions at low ionic strength that monotonically decay to weak non-electrostatic attractions at high ionic strength), the domain-level model (DODECA) predicted high- c_2 behavior well, consistent with previously reported work and findings in Chapter 2.^{49,50,68,179} The 1bC/D model was able to correctly predict the nature of electrostatic interactions (i.e., attractive or repulsive) in all cases, even for strong electrostatic attractions, while the DODECA model gave poor predictions for systems with significant electrostatic attractions. Those poor predictions would be misleading if used when determining formulation conditions or selecting prospective MAb candidate(s) based on their anticipated developability challenges. With only low- c_2 SLS measurements of the second osmotic virial coefficient B_{22} (or k_D from DLS), the 1bC/D model can predict the balance between steric repulsions due to crowding, short-ranged non-electrostatic interactions such as hydrogen bonding or hydrophobic interactions, and electrostatic attractions and repulsions at high- c_2 . This method is not exclusive to MABs and is easily transferable to other proteins provided that their sequences and homology models are available.

Challenges that stemmed from the representation of charge equilibria were addressed by two different methods: one that individually assigned pK_a values for

each ionizable residue and another that explicitly represented dynamic charge fluctuations. For cases with net repulsive electrostatic interactions, low- to high- c_2 simulations were more accurate, particularly with the latter method. Analysis of average interaction energies from charged sites for the MAb with the strongest electrostatic attractions (MAb 2) suggested that specific interactions involving amino acids in the C_{H2} domain were prominent in driving those strong attractions. Future work with the 1bC/D model could include the incorporation of hinge flexibility to sample a more complex configuration space, or algorithms such as configurational bias to improve acceptance rates at high- c_2 .^{76,226} Future work could also focus on the effects of cosolutes such as ions, surfactants, or ionic liquids on electrostatic self-interactions by explicitly representing them in the simulations.

Chapter 5

DECONVOLUTING THE BIOPHYSICAL PHENOMENA THAT INFLUENCE LONG-TERM AGGREGATION RATES OF HIGH-CONCENTRATION MONOCLONAL ANTIBODY FORMULATIONS

5.1 Introduction

As discussed in section 1.3.3, irreversible protein aggregation (hereafter referred to as aggregation) is a problematic and ubiquitous challenge within drug development of therapeutic proteins. Aggregation can create challenges in many processes within development and manufacturing, limit product shelf-life, and trigger unwanted immunogenic responses when introduced to a patient.^{114,125–128} This chapter is focused on the long-term stability of high-concentration (i.e., > 100 mg/mL) MAb solutions, such as those intended for subcutaneous injection. In this context, long-term stability refers to the rate of aggregation over timescales on the order of months to years. Formulations that minimize aggregation can be challenging to develop as solution conditions like pH, excipient concentrations, and protein concentration can have complex relationships with the aggregation process that can differ between different proteins.^{125,147,235–237} With a large selection of solution conditions to choose from, and limited available protein material in earlier stages of development, determining stable solution conditions can be a difficult process to do quickly (see also section 1.5).^{91,229,238}

From an industrial / practical perspective, there is a need for tools and methods to predict long-term, high-concentration aggregation rates as early as possible in the

product discovery and development lifecycles as a part of developability assessment, or even to expand the space of “developable” proteins by accurate prediction of suitable solution conditions. The range of possible solution conditions is typically reduced by high-throughput screening experiments (i.e., sample sparing and short time-scale) that measure quantities that could be considered to be a proxy for poor long-term stability.^{11,53} Measurements of conformational stability (e.g., via differential scanning calorimetry (DSC) or isothermal chemical denaturation) are common (see also section 1.3.1), but only directly apply to a subcase of conditions where the measured conformational change (e.g., domain-level unfolding or misfolding) is quantified and determinate of the aggregation rate.^{92,139,239–245} Measurements of self-interactions are also commonplace (e.g., via SLS and DLS) and similarly, there is evidence to support that these quantities are related to aggregation rates in certain subcases, such as when electrostatic repulsions reduce aggregation rates.^{43,85,137–140} Computational methods are available, with varied degrees of success, that attempt to predict fundamental quantities related to conformational stability, self-interactions, or amino acid sequences that could be considered aggregation “hot spots”.^{11,25,91,151} Experimental and computational screening may be able to partially reduce the space of possible solution conditions (and protein candidates), but have not been broadly successful in quantitatively predicting long-term aggregation rates, particularly with respect to changes in solution conditions.^{53,56–61} There has been much recent interest in using machine learning (ML) methods to parse the features relevant to protein aggregation and make predictions of long-term physical stability.^{40,85,148,170,233,246–249} Significant progress has been made in other related areas, particularly in prediction of elevated viscosity.^{99,233,250,251} The inherently confidential nature of drug development

is a challenge as many pharmaceutical companies generate data that could be useful in effectively parsing the important features to protein aggregation, but that data is not publicly available. The long-term stability of commercially available therapeutics is inherently biased towards stable formulations, and information about the formulations that were less stable, and thus the potential sources of their instability, is not typically disclosed.

In this chapter, aggregation rates for four MAbs (introduced in Chapter 4 as MAb 3, MAb 4, MAb 5, and MAb 6) were measured at a range of solution conditions (pH 5 and pH 6.5; 10 mM and ~110 mM ionic strength), MAb concentrations (10, 35, and 130 mg/mL) and incubation temperatures (4 °C, 30 °C, and 45-50 °C) representative of industrial stability studies and/or commercial drug product storage conditions.^{3,15,91,252} The measured aggregation rates spanned several orders of magnitude, from hours to years. The dataset in this chapter was designed to systematically test specific formulation features (i.e., pH, ionic strength, and protein concentration) with respect to incubation temperature to better understand the global features of MAb aggregation. Aggregation rates in the initial-rate regime were quantified with size-exclusion chromatography (SEC), and subvisible particle formation was monitored with backgrounded membrane imaging (BMI). Each MAb was also characterized based on its conformational stability (via DSC), and self-interactions (via SLS and DLS, data shown in Chapter 4) across the solution conditions from the stability studies. MAb 1 from Chapters 2 and 4 is also referenced, but was not studied in the context of long-term stability in this work.^{43,48-50,179} Interpretable statistical models based on ML methods were applied to parse and quantify the features relevant to the measured aggregation rates at high-concentration

(i.e., 130 mg/mL). The models provided a more rigorous framework for determining statistical significance that could distinguish between somewhat collinear features. Aggregation rates of high-concentration protein solutions at realistic long-term storage conditions such as what are presented here have been limited to a relatively small number of reports in the literature, and have been confined to smaller sets of formulations and/or proteins.^{56,129,145,151,160,169,214,253} The result indicate some guidance for rational stability study design, as well as toward the development of robust ML models for prediction of long-term aggregation rates. The formulation conditions in this work are broadly similar to those from Chapter 4 and are shown in Table 5.1.

Table 5.1: Formulation conditions for stability studies in this chapter^a

Formulation	NaCl concentration (mM)	Total ionic strength (mM)	Symbol types ^b
pH 5 low IS	4	10	Purple filled circle
pH 5 high IS	100	106	Purple open circle
pH 6.5 low IS	0	10	Green filled triangle
pH 6.5 high IS	110	110	Green open triangle

^a All studies were performed for 10, 35, and 130 mg/mL MAb concentrations

^b Symbol types that correspond to results in Figures in this chapter

5.2 Materials and methods

5.2.1 Backgrounded membrane imaging

Subvisible particle formation was quantified with a Halo Labs HORIZON instrument (Halo Labs, Burlingame, CA). Samples were diluted to 1 mg/mL with matching buffer, and 3 50 μ L aliquots of a given sample were pipetted onto 3 different

wells of a 96-well membrane plate with 0.4 μm pore size polycarbonate filters. A 350 mbar vacuum was applied to remove the liquid and immobilize the subvisible particles, and wicking paper was used to remove liquid that might have adhered to the bottom of the membranes. Images were taken of the membrane and image analysis was performed with HORIZON VUE software. Further details about the technique and comparisons with similar instruments (e.g., MFI, FlowCam, and HIAC) have been reported elsewhere.^{88,254–257}

The histogram of particle counts for a given sample by equivalent circular diameter, minus any particles from a buffer-only measurement, were converted to particle area. The cumulative subvisible particles of all sizes present in the solution were quantified as total particle area per 100 mg protein, to normalize for different starting concentrations. Although particle counts, rather than areas, are the typical quantity monitored by regulatory agencies, particle areas were used here to connect more directly to the mass of protein being incorporated into the subvisible particles.

In this chapter, BMI results were separated into 2 binary categories based on whether significant subvisible particle formation occurred over the span of the stability study. A stability study was considered prone to subvisible particle formation if the total subvisible particle area consistently increased over time, or if the following threshold was ever exceeded: no more than 5,000 particles greater than 10 μm in diameter, and no more than 500 particles greater than 50 μm in diameter (based on USP <787>, and <788> for products with dosage volumes less than 100 mL).^{135,136} That threshold was converted roughly to particle area by assuming 500 particles of 75 μm diameter, and 4,500 particles of 30 μm diameter, which have a total particle area per 100 mg protein of $5.4 \times 10^6 \mu\text{m}^2$. This additional threshold was considered to take

into account formulations that were prone to subvisible particle formation such that the unstressed, “T0” sample already contained many subvisible particles that in some cases would dissociate when heated.

5.2.2 Differential scanning calorimetry

Differential scanning calorimetry was performed with a MicroCal VP-DSC instrument (Malvern Panalytical, Malvern, UK) for 1 mg/mL solutions of all 4 MAbs at each of the formulations listed in Table 5.1. Thermal scans were performed from 25 °C to 90 °C at a rate of 60 °C per hour. At least 4 buffer/buffer baseline scans were averaged to determine the instrument background that was subtracted from the subsequent protein scan. The thermograms were transformed to absolute, partial specific heat capacity ($c_{p,abs}$) via Equation 5.1²⁵⁸

$$c_{p,abs} = c_{p,w} \frac{V_p}{V_w} - \frac{\Delta c_{p,app}}{m_p} \quad 5.1$$

$c_{p,w}$ is the heat capacity of water (approximately 1 cal/g-K), V_p is the partial specific volume of the protein (estimated as 0.72 mL/g), V_w is the specific volume of water (approximately 1 mL/g), $\Delta c_{p,app}$ is the buffer subtracted DSC signal scaled by the scan rate, and m_p is the total protein mass in the cell (roughly 0.5 mg in this work). Midpoint unfolding temperatures ($T_{m,app1}$ and $T_{m,app2}$) were the temperatures where a local maximum occurred in the $c_{p,abs}$ profile, and the onset unfolding temperature ($T_{m,onset}$) was the temperature where $c_{p,abs}$ increased 10% from the baseline value. In all cases, either the $c_{p,abs}$ profiles had a sharp exotherm before reaching 90 °C, or the MAb solutions were visibly hazy after cooling to room temperature, indicating significant irreversible aggregation and precipitation. As unfolding was likely not reversible in some cases, the results were considered as nonequilibrium and additional

thermodynamic quantities (e.g., enthalpy and free energy of unfolding) were not calculated.^{43,92,258,259}

5.2.3 Size-exclusion chromatography

An Agilent 1100 high-performance liquid chromatography (HPLC) instrument (Agilent Technologies) with a Tosoh TSKgel G3000SWxl (Tosoh Bioscience, Montgomeryville, PA) size-exclusion column was used for size-exclusion chromatography (SEC) for MAb samples from stability studies. Integrated peak areas for aggregates, monomeric protein, and low molecular weight (LMW) MAb fragments were determined with Agilent ChemStation (Agilent Technologies) software from chromatograms of absorbance at a 280 nm detection wavelength and a 1 mL/min flowrate. Aggregate, monomer, and LMW fragment fractions were determined by integrated peak areas calibrated to standard samples that were also used to normalize for variability in the laser and changes in separation performance of the SEC column over time. Samples were diluted to 10 mg/mL with matching buffer, centrifuged at 5000 rcf for 1 minute to remove impurities or insoluble aggregates, and the supernatant was used for subsequent SEC analysis. Integrated peak areas for 3 nonconsecutive 100 µg injections were averaged for a given sample.

A preliminary screen of mobile phase conditions was performed for each MAb to choose conditions where aggregates did not precipitate or dissociate, and that provided clearly resolved peaks for aggregates, monomeric proteins, and LMW fragments.^{43,260} All mobile phase solutions were prepared with deionized water and 0.5% volume fraction ortho-phosphoric acid (Fisher Scientific). For MAb 3 and MAb 4, the mobile phase was at pH 5 with 100 mM NaCl. For MAb 5 and MAb 6, the mobile phase was at pH 3.5 with 50 mM NaCl.

5.2.4 Isothermal stability studies

Samples were prepared for stability studies at MAb concentrations of 10, 35, or 130 mg/mL. Deactivated clear glass screw neck autosampler vials (Waters Corporation, Milford, MA) were filled with minimal headspace and then hermetically sealed with PTFE screw caps (Waters Corporation). Quiescent, isothermal incubations were performed on the stability samples with an incubator set to the desired incubation temperature (T_{inc}) and with the vials stored upright in an opaque box. The samples for refrigerated studies were stored in a dedicated laboratory refrigerator with more strict temperature control than domestic refrigerators (approximately ± 1 °C). At a given timepoint, a sample was removed from the incubator and cooled via immersion in an ice-water bath for at least 5 minutes. Stability studies were performed for the four MAbs in the formulations shown in Table 5.1, at refrigerated, 30 °C, and accelerated conditions, and at 3 MAb concentrations: 10, 35, and 130 mg/mL.

All refrigerated studies proceeded for at least 12 months with at least 3 samples analyzed at varying timepoints. 30 °C studies proceeded for at least 6 months with at least 5 samples analyzed at varying timepoints. At least 6 samples were analyzed for accelerated studies, where the studies proceeded such that the final sample had at least 10% monomer loss. The T_{inc} values for accelerated studies were chosen for a given MAb and pH (i.e., the same T_{inc} for both low and high IS) by preliminary stability studies such that a 10 mg/mL solution at the low IS formulation had a monomer loss rate of about 1% per week. The accelerated T_{inc} was also confirmed to be at least 10 °C below the first midpoint unfolding temperature ($T_{m,app1}$) from DSC measurements (see also section 5.2.2 above).

5.2.5 Aggregation rate calculations

Analysis was restricted to stability samples within the initial-rate regime, where aggregate concentration as a function of time was linear and thus the dominant aggregation pathway was likely the same throughout.⁹² The earliest extents/stages of aggregation are also the most directly applicable to the problem of long-term stability for drug products.⁹² The observed aggregation rate coefficient (k_{obs}) was determined by linear regression of aggregate fraction (a) as a function of incubation time (t) via Equation 5.2.

$$\frac{da}{dt} = k_{obs} \quad 5.2$$

Reported uncertainty in k_{obs} were 95% confidence intervals from the linear regression. The minimum aggregation rate that could be quantified in this work was $\log_{10}[k_{obs}(\text{days}^{-1})] = -4.86$ (0.5% aggregate per year, on a mass basis).

Aggregate concentration was determined for each stability sample and an unheated, “T0” sample with SEC (see section 5.2.3 above). All T0 samples had less than 1% aggregate fraction. There were two different procedures for calculating aggregate fraction, based on the potential for bias from the formation of fragments: case 1, where negligible fragmentation occurred (i.e., little to no change in LMW fragment peak area in SEC chromatograms), and case 2, where significant fragmentation was observed. In case 1, aggregate concentration was determined as the decrease in monomer fraction (i.e., the change in monomeric protein peak area in SEC chromatograms). In case 2, some fragments (e.g., ~100 kDa) eluted at similar retention volumes as monomer and could not be reliably deconvoluted, so aggregate fraction was determined directly from the peak area(s) for aggregates. The advantage of case 1 is that any insoluble aggregates or aggregates that might travel through the column but

have absorbance at 280 nm disproportionate to their mass do not impact the calculated aggregate concentration.²⁶⁰ In either case, it was confirmed that the aggregate fractions were reasonably linear with time, satisfying the initial-rate assumption. A table of aggregate concentration case by formulation is shown in Table D.1 (Appendix D), and a parity plot comparing aggregation rates calculated by case 1 vs. case 2 is shown in Figure D.1 (Appendix D).

LMW fragment fraction was also monitored for linearity, in case of “cross-interactions” where fragments could be incorporated into aggregates, or where aggregates could lose mass due to fragmentation. There were no cases in this work that suggested that fragments were significantly impacting the measured aggregation rates.

5.2.6 Elastic net regression models

Possible input features for a given high-concentration (i.e., 130 mg/mL) stability study condition (i.e., a specific MAb, pH, NaCl concentration, and T_{inc}) were collected from information about the formulation (e.g., pH and ionic strength), experimental measurements (e.g., from SLS and DSC), and results from stability studies at other conditions (e.g., the same formulation but at accelerated T_{inc}). The output or predictor variable was the measured aggregation rate at that condition, which was converted to $\log_{10}(k_{obs})$. The features were screened for groups of features that were intended to describe the same phenomenon and were highly correlated (via the Pearson correlation coefficient, R) to avoid overfitting of the statistical models. The full initial set of features, those that were highly correlated ($|R| > \sim 0.7$), and the features that remained after screening are shown in Appendix D.1. Inputs from stability studies were also chosen such that no overlapping information was used (e.g.,

if an Arrhenius extrapolated k_{obs} value was selected, the k_{obs} values used in that extrapolation were not selected). The input features were standardized such that the mean of each feature was 0, and the standard deviation was 1. This put all of the inputs on a similar scale to avoid bias and so that the coefficients for each feature could be quantitatively interpreted in terms of which features were more or less important.

Elastic net regression (ENR)²⁶¹ was performed to relate the screened feature sets (described above) to aggregation rates for all 130 mg/mL formulations at a given stability condition (refrigerated, 30 °C, or accelerated conditions). ENR is an ML method that adds two penalties to the typical least-squares residuals from multilinear regression, the L1 and the L2 norms (also known as lasso and ridge regression, respectively). Two hyperparameters in the elastic net models, α and λ_{EL} , scale the relative contributions of the L1 and L2 norms, and the strength of the penalties. ENR was used to reduce overfitting and help in feature selection by shrinking coefficients towards 0.²⁴⁷ Linear ENR was performed for all three T_{inc} conditions, and for the refrigerated condition, logistic ENR was also performed. Logistic regression predicts the log-odds (logit) of the probability of a binary outcome occurring. The binary criterion was whether the value of $\log_{10}(k_{obs})$ was greater than -4.56 (1% aggregation per year).

ENR models were trained iteratively with recursive feature elimination (RFE) to generate a model with the lowest number of features while maintaining that each feature was statistically significant in reducing the model error. After each feature selection “round”, the feature with the lowest coefficient magnitude (or all features with a coefficient of 0) was removed, and the model was trained again. That process repeated until one feature remained, and the optimal model was chosen based on the

95% confidence interval overlap test. “Leave-one-protein-out” cross-validation (LOPO CV) was used to train the ENR models where the dataset was split into “folds” or subsets based on the MAb identity.²⁴⁸ The model with α and λ_{EL} values that minimized the mean absolute error (MAE; linear ENR) or deviance (logistic ENR) between the model predictions and test sets was chosen.

The observed aggregation rates in some cases contained significant uncertainties, which were accounted for with Monte Carlo (MC) sampling. For a given feature selection round, each k_{obs} value that the model tested against for a given condition was sampled 50 times from the Gaussian distribution defined by the k_{obs} value that was measured and its variance. Reported uncertainties for feature coefficients and model error values are based on the 95% confidence interval of results with different MC sampled training sets.

5.3 Results and discussion

5.3.1 Stability study design

Stability studies for the four MAbs were designed to parse some fundamental features of typical solution conditions for commercial MAb therapies, in ambient conditions characteristic of industrial accelerated and long-term stability studies used to screen MAb candidates and formulations, and demonstrate satisfactory long-term shelf-life.^{144,262} pH and ionic strength were varied in a 2-level full factorial design (see Table 5.1) intended to sample a range of electrostatic contributions to the protein self-interactions and conformational stability that can be fundamental to protein aggregation. The two pH values (5 and 6.5) altered the charge state of ionizable residues in the proteins (primary His at these pH values). The two ionic strength (IS)

values (10 mM and ~110 mM, referred to as low IS and high IS hereafter; note that “high” was referred to as “intermediate” in the previous chapter) scaled the strength of electrostatic interactions from relatively strong to weakened but likely not fully screened via the Debye screening effect. For each combination of pH and ionic strength (i.e., each row in Table 5.1), stability studies were performed for formulations with MAb concentrations (c_2 values) of 10, 35, and 130 mg/mL. Those concentrations are roughly log-spaced, to scale with phenomenological models of aggregation that propose that some rate limiting processes for aggregation (e.g., monomer addition for aggregate growth) are first order or pseudo first order with protein concentration.^{105,117,263} 130 mg/mL is representative of high-concentration MAb formulations intended for subcutaneous injection^{3,15}, but it would be sample-sparing to understand when low-concentration aggregation rates correlate with those at high-concentration, and phenomenologically insightful to understand how aggregation rates change over a range of concentrations. MAb subtype, experimental pI values, and accelerated T_{inc} values for each MAb are shown in Table 5.2.

Table 5.2: Subtype, pI, and incubation temperature (T_{inc}) for accelerated condition of each MAb

MAb	Subtype	pI ^α	Accelerated T_{inc} (°C)
MAb 3	IgG1	8.4	50 (pH 5); 45 (pH 6.5)
MAb 4	IgG4	7.5	45
MAb 5	IgG1	8.5	50
MAb 6	IgG1	9.2	50

^α pI values are certain to approximately ± 0.1

Each formulation (MAb, pH, IS, and c_2) was tested at refrigerated (4 °C), 30 °C, and accelerated conditions (45 °C or 50 °C, see Table 5.2). Preliminary studies were used to determine the T_{inc} value for the accelerated conditions in an attempt to keep accelerated studies fairly rapid (i.e., useful in the industrial context) while avoiding even faster rates where the dominant aggregation mechanism might be more likely to be different than that for 30 °C or refrigerated conditions (see section 5.2.4).²⁶³ The accelerated temperatures chosen from those stability studies are a rough estimate of the rankings of relative rates for 10 mg/mL low IS formulations, and should not necessarily be considered as indicative of stability at other conditions. All midpoint unfolding temperatures ($T_{m,app1}$ values) from DSC were at least 10 °C higher than the accelerated T_{inc} that was chosen.

High-concentration MAb formulations are commonly intended for refrigerated storage¹⁶⁰, where 30 °C might be considered an “accelerated” condition as well as a conservative estimate of room temperature conditions.^{252,262} 45 °C to 50 °C are denoted as accelerated conditions here to represent the more preliminary studies performed in early-stage development for formulation and candidate screening.^{91,144} Based on the length of the long-term studies and inherent uncertainty in aggregate concentrations for samples with a relatively low number of aggregates, the minimum aggregation rate that could be reliably quantified in this chapter was $\log_{10}[k_{obs}(days^{-1})] = -4.86$ (0.5% aggregation per year, on a mass basis.).

5.3.2 Results from stability studies

Aggregation rates at accelerated conditions (45 °C or 50 °C; see Table 5.2) are shown in Figure 5.1 in the form of observed rate coefficient values (k_{obs} ; see Equation 5.2). Generally, the pH 6.5 formulations had slower aggregation rates than pH 5

formulations, although that finding was not clear for MAb 3 as the T_{inc} was 50 °C for pH 5 conditions, and 45 °C for pH 6.5 conditions. The pH 5 high IS condition had the fastest aggregation rate of all four formulations for a given MAb and c_2 , except for MAb 4 at 130 mg/mL where the aggregation rate at the pH 5 low IS condition was similar. Aggregation rates were broadly similar between the low IS and high IS formulations at pH 6.5 for a given MAb and c_2 . MAb 5 and MAb 6 were typically more stable than MAb 3 and MAb 4 for a given formulation and c_2 , particularly considering that the T_{inc} was lower for MAb 3 and MAb 4 in most cases.

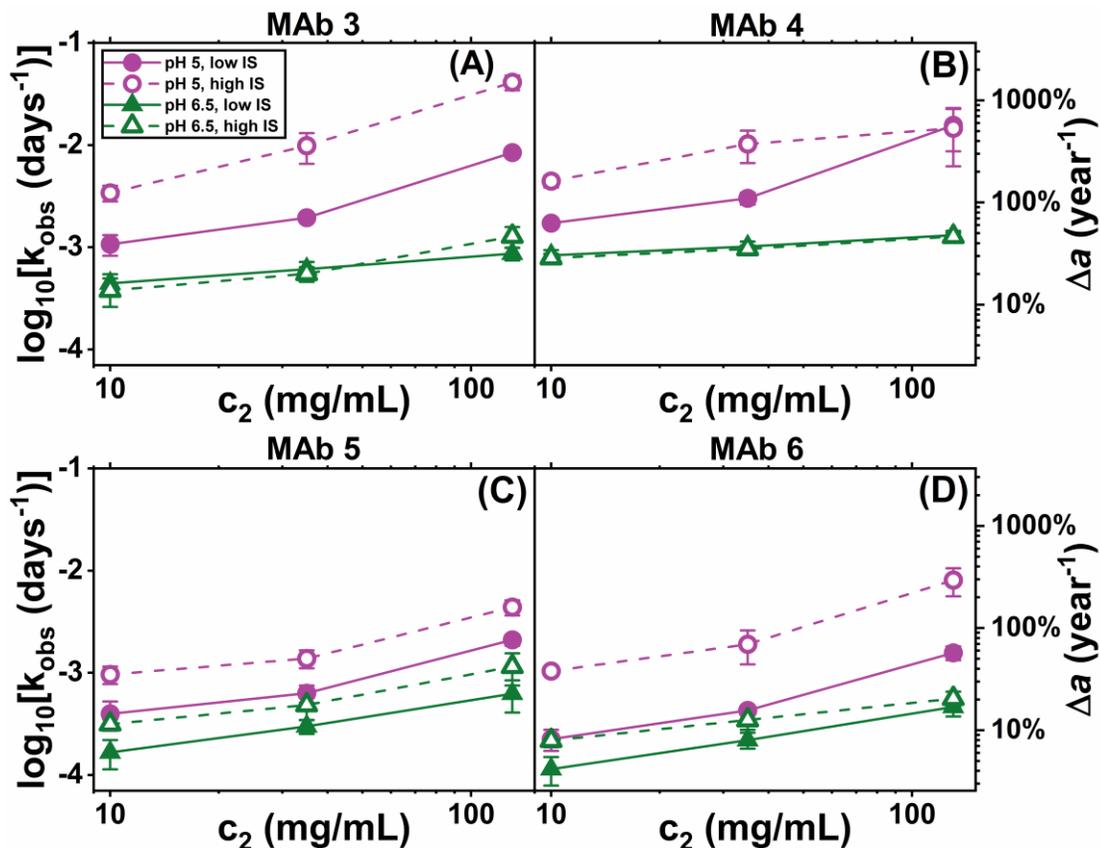


Figure 5.1: Observed aggregation rate coefficient (k_{obs}) values as a function of MAb concentration (c_2) at accelerated conditions (see Table 5.2) for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D), with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. Solid (dashed) lines connecting the data points correspond to low (high) IS and are guides to the eye.

Aggregation rates (k_{obs} values) at 30 °C conditions are shown in Figure 5.2. Similar to the accelerated studies (see Figure 5.1), MAb 5 and MAb 6 had slower aggregation rates than MAb 3 and MAb 4 for a given formulation and c_2 value. The pH 5 high IS condition still typically had the fastest aggregation rate, but only at 10

mg/mL. In contrast with aggregation rates for accelerated conditions, the relative stability rankings of each formulation for a given MAb and c_2 at 30 °C were not as systematically consistent and often statistically indistinguishable. Notably, there was a “crossover” for MAb 4 where the pH 6.5 formulations were more stable than the pH 5 formulations at 10 and 35 mg/mL but had faster aggregation rates than pH 5 formulations at 130 mg/mL.

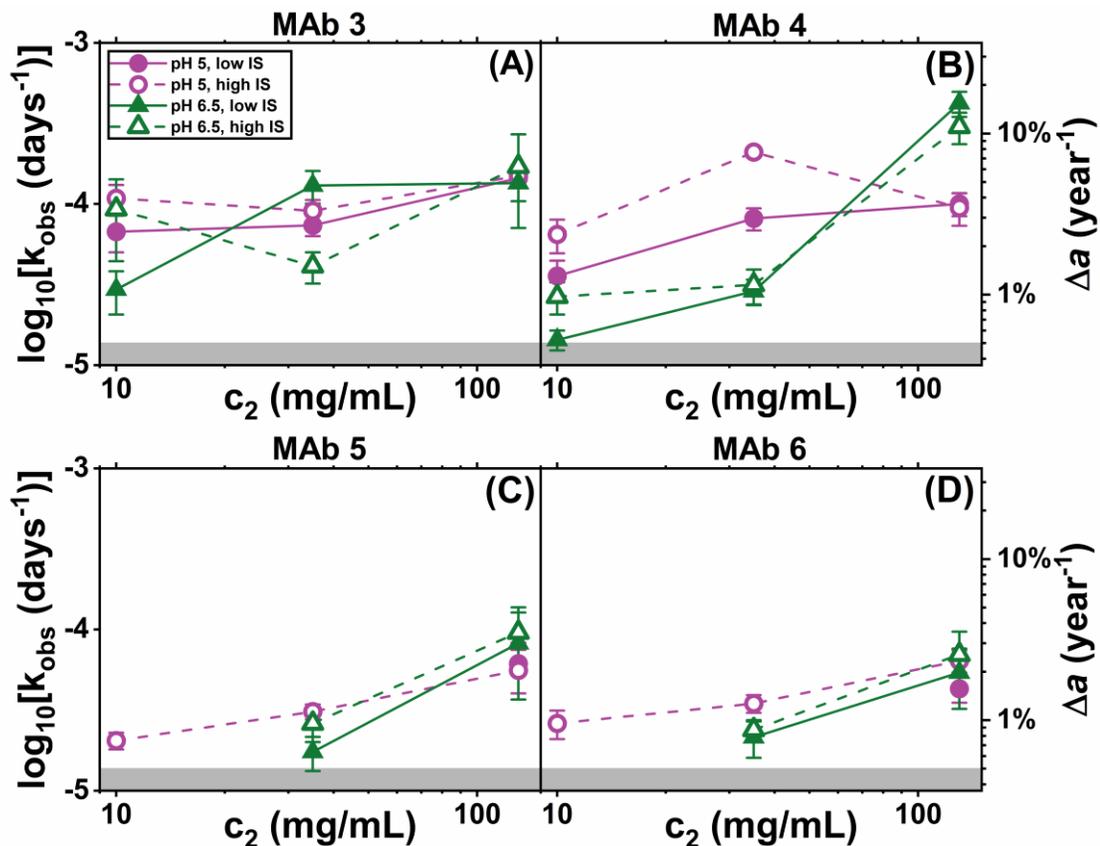


Figure 5.2: Observed aggregation rate coefficient (k_{obs}) values as a function of MAb concentration (c_2) at 30 °C for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D), with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. Solid (dashed) lines connecting the data points correspond to low (high) IS and are guides to the eye. The gray shaded region for $\log_{10}[k_{obs} \text{ (days}^{-1}\text{)}] < -4.86$ represents the measurement noise floor and any formulations with data not shown were below the measurement noise floor.

Aggregation rates (k_{obs} values) at refrigerated conditions for 130 mg/mL formulations are shown in Figure 5.3. All of the 10 and 35 mg/mL formulations had aggregation rates below the measurement floor, without exception. For all four MAbs,

the pH 5 formulations had significantly slower aggregation rates than the pH 6.5 formulations. MAb 3 was the most stable (had the slowest aggregation rate) in every formulation, although at the pH 5 low IS and pH 6.5 high IS formulations, aggregation rates were not significantly different between MAb 3 and MAb 6.

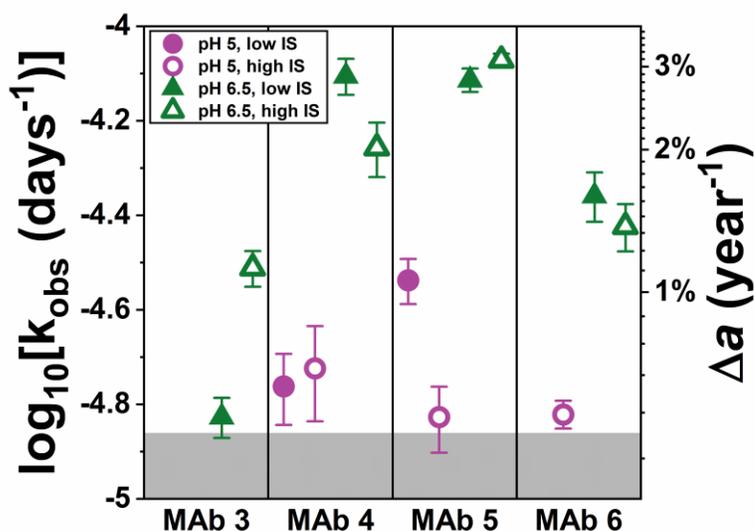


Figure 5.3: Observed aggregation rate coefficient (k_{obs}) values at refrigerated conditions for 130 mg/mL solutions, with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. Symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals. MAb identities are listed on the x-axis. Aggregation rates for MAb 3 at pH 5 low IS, MAb 3 at pH 5 high IS, and MAb 6 at pH 5 low IS were below the measurement floor (the shaded region for $\log_{10}[k_{obs}(days^{-1})] < -4.86$) after at least 1 year. Error bars are 95% confidence intervals.

Aggregation rates were slower as T_{inc} was decreased, which was generally expected for the T_{inc} values in this work.^{24,264} The only exceptions were 130 mg/mL, pH 6.5 formulations of MAb 5 and MAb 6, which had similar aggregation rates

between 30 °C and refrigerated conditions. However, the relative rankings of each formulation at 130 mg/mL for a given MAb were inconsistent between different T_{inc} conditions, as depicted in Figure 5.4. Studies at accelerated conditions suggested that pH 6.5 formulations were the most stable, there was not a clear trend with pH in 30 °C studies, and at refrigerated conditions, pH 6.5 formulations were the least stable. The aggregation rates at accelerated and refrigerated conditions were fairly anti-correlated ($R = -0.58$). The only case where relative formulation rankings for 130 mg/mL formulations at refrigerated conditions were consistent with a different T_{inc} was for MAb 4 at 30 °C, where there was a “crossover” between 35 and 130 mg/mL.

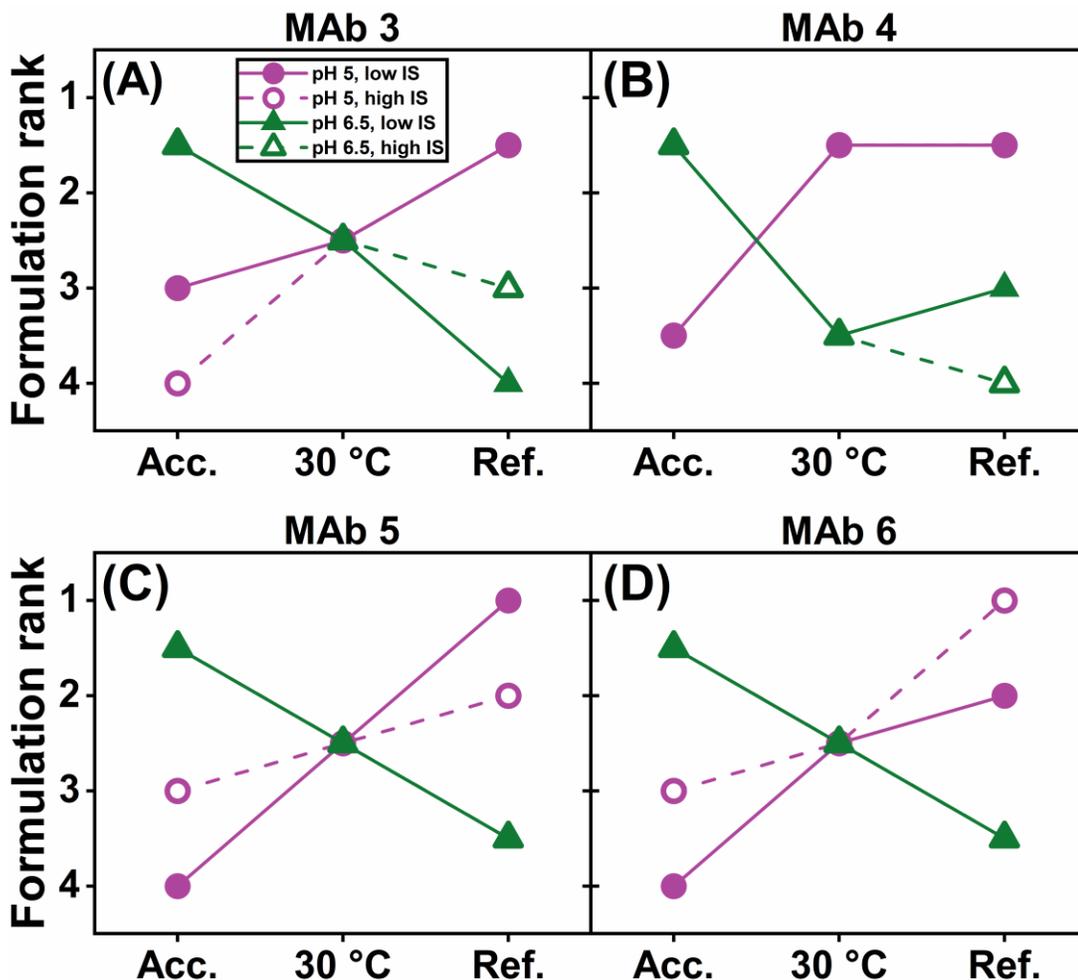


Figure 5.4: Relative stability rankings, where a higher ranking was a slower k_{obs} value, of each 130 mg/mL formulation as a function of T_{inc} (nonlinear in x-axis; “Acc.” is short for accelerated and “Ref.” is short for refrigerated) for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D). The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and solid (dashed) lines connecting the data points correspond to low (high) IS. When the k_{obs} values of multiple formulations were statistically indistinguishable, the rankings were averaged.

The results from the accelerated and 10-35 mg/mL 30 °C studies suggested that low IS formulations were most stable, but at 130 mg/mL 30 °C and refrigerated

conditions, there was not as significant of a trend with respect to IS values. Parity plots comparing aggregation rates between pH 5 and pH 6.5 formulations and low and high IS formulations are shown in Figure 5.5.

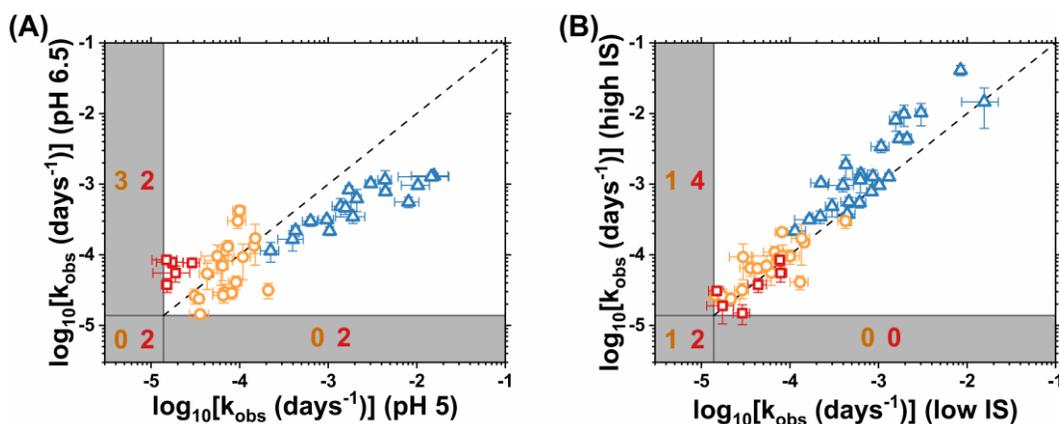


Figure 5.5: Parity plots comparing aggregation rates for accelerated (blue triangles), 30 °C (orange circles), and refrigerated (red squares, 130 mg/mL only) conditions between: (A) pH 5 and pH 6.5 formulations, and (B) low and high ionic strength formulations. Aggregation rates for MAb 3 at accelerated temperature conditions are omitted from panel A as the incubation temperatures were different between the two pH conditions (see Table 5.2). The black dashed line is the parity line, and the gray region represents the measurement floor of the study ($\log_{10}[k_{obs}(days^{-1})] = -4.86$). The numbers shown in the measurement floor regions correspond to formulations where one or both of the aggregation rates were less than the measurement floor at 30 °C or refrigerated conditions.

When comparing aggregation rates for a given formulation across the different T_{inc} conditions, it is clear that the aggregation rates were non-Arrhenius, as shown in Figure 5.6 which compares measured aggregation rate at 130 mg/mL and refrigerated conditions with the aggregation rate predicted by applying Arrhenius extrapolation to

the measured 130 mg/mL aggregation rates at accelerated and 30 °C conditions. The Arrhenius extrapolated k_{obs} values were all underestimates of the measured k_{obs} values, especially for the pH 5 formulations where extrapolated k_{obs} values were several orders of magnitude lower than the measured values. That result is indicative of either a change in mechanism or an inherently non-Arrhenius mechanism (e.g., one dependent on a Gibbs free energy of unfolding), and has been reported several times for MAbs and other proteins.^{92,129,239,263,265–268} Despite clear non-Arrhenius behavior, the Arrhenius extrapolated k_{obs} values were fairly correlated with the measured values at refrigerated conditions ($R = 0.70$). Many other analytical models could be used as well, but broadly those models are out of scope for this study as aggregation rates were only measured at 3 temperatures.¹⁴⁵

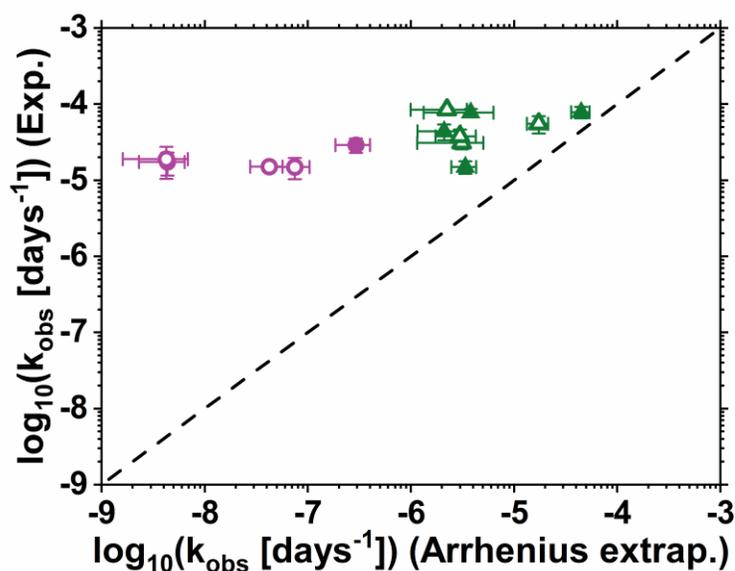


Figure 5.6: Parity plot comparing k_{obs} values at refrigerated conditions for 130 mg/mL formulations that were measured experimentally (y-axis) and k_{obs} values predicted by Arrhenius extrapolation (x-axis) via rates at 30 °C (Figure 5.2) and accelerated conditions (Figure 5.1). The black dashed line is the parity line. Symbols for each formulation correspond to Table 1 and error bars are 95% confidence intervals.

As mentioned above, the dependence of aggregation rates on c_2 was different between the accelerated and 30 °C conditions. At accelerated conditions, the k_{obs} values scaled consistently with c_2 (linear in the context of the log-log plot in Figure 5.1), with the possible exception of MAb 4 at the pH 5 low IS condition. That result would suggest that the rate limiting step in the dominant aggregation mechanism might have been constant as a function of c_2 . The slope of c_2 with respect to k_{obs} was generally larger at pH 5 conditions than pH 6.5 conditions, suggesting that there were different driving forces for aggregation at those conditions. In contrast, the dependence of k_{obs} values on c_2 was less apparent at 30 °C conditions, which suggests that the dominant aggregation mechanism might have been dependent on c_2 .

The fractional monomer loss based aggregation rates for all T_{inc} conditions generally increased as c_2 was increased, which rules out zeroth order mechanisms such as those where unfolding is rate limiting or where aggregation is surface mediated.^{92,141}

Aggregate size as measured by SEC was also highly dependent on T_{inc} . For samples from refrigerated studies, there were no formulations where a significant high molecular weight (HMW) aggregate population (i.e., oligomers or multimers that eluted at the void volume in SEC) was detected. Only 13 out of 40 formulations from 30 °C studies that showed significant aggregation (i.e., $\log_{10}[k_{obs}(days^{-1})] > -4.86$) had HMW aggregates in SEC chromatograms: 9 out of 12 for MAb 3 and 4 out of 12 for MAb 4. Conversely, 39 out of 48 formulations in the accelerated studies had HMW aggregates. Illustrative chromatograms and a summary table of HMW formation are shown in Figure 5.7 and Table 5.3, respectively. The aggregates that were formed at accelerated conditions were much more likely to grow into HMW aggregates than those at 30 °C or refrigerated conditions, which is consistent with other reports of MAb physical stability.^{129,239,269} The observed differences in aggregate sizes were consistent with shifting aggregation mechanisms between the different T_{inc} conditions in this study, although that does not indicate whether mechanistically aggregate growth was rate limiting with respect to monomer loss. Formation of HMW aggregates did not necessarily translate into higher propensity to form subvisible particles (discussed in more detail in section 5.3.5) for a given MAb and formulation, but it was generally true that subvisible particles were more likely to form at higher T_{inc} conditions.

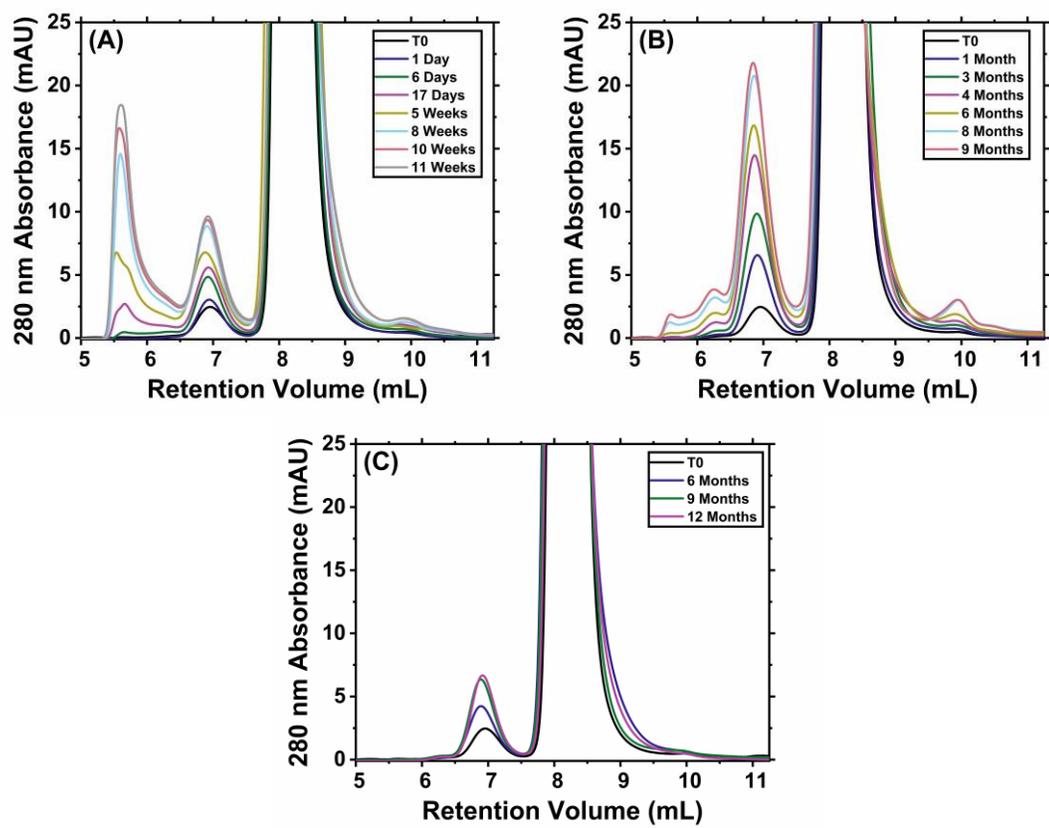


Figure 5.7: SEC chromatograms that illustrate differences in HMW formation as a function of T_{inc} . The 130 mg/mL MAb 4 at pH 6.5 low IS formulation was used as the example case. Each panel is a time course of samples from (A) accelerated, (B) 30 °C, and (C) refrigerated studies. Incubation times are shown in the legend, where T0 represents the unstressed sample.

Table 5.3: High molecular weight aggregate formation case by formulation^α

MAB	Form.	Accelerated			30 °C			Refrigerated		
		10 mg/mL	35 mg/mL	130 mg/mL	10 mg/mL	35 mg/mL	130 mg/mL	10 mg/mL	35 mg/mL	130 mg/mL
MAB 3	pH 5 low IS	+	+	+	+	+	+	N/A	N/A	N/A
	pH 5 high IS	+	+	+	+	+	+	N/A	N/A	N/A
	pH 6.5 low IS	+	+	+	-	+	-	N/A	N/A	-
	pH 6.5 high IS	+	+	+	+	-	+	N/A	N/A	-
MAB 4	pH 5 low IS	-	+	+	-	-	-	N/A	N/A	-
	pH 5 high IS	+	+	+	-	+	+	N/A	N/A	-
	pH 6.5 low IS	+	+	+	-	-	+	N/A	N/A	-
	pH 6.5 high IS	+	+	+	-	-	+	N/A	N/A	-
MAB 5	pH 5 low IS	-	-	-	N/A	N/A	-	N/A	N/A	-
	pH 5 high IS	+	+	+	-	-	-	N/A	N/A	-
	pH 6.5 low IS	+	+	+	N/A	-	-	N/A	N/A	-
	pH 6.5 high IS	+	+	+	N/A	-	-	N/A	N/A	-
MAB 6	pH 5 low IS	-	-	-	N/A	N/A	-	N/A	N/A	-
	pH 5 high IS	+	+	+	-	-	-	N/A	N/A	-
	pH 6.5 low IS	+	+	+	N/A	-	-	N/A	N/A	-
	pH 6.5 high IS	+	+	+	N/A	-	-	N/A	N/A	-

^α Formulations that were prone to HMW aggregate formation are shown in purple and with + symbols, those that were not prone to HMW aggregate formation are shown in green and with - symbols, and cases with aggregation rates below the measurement noise floor are shown in gray and with “N/A”.

At refrigerated conditions, fragmentation rates were slow and typically a significant increase in LMW fragments (~50 kDa, putatively a single MAb fragment) was not observed. Significant fragmentation was observed at 30 °C and accelerated conditions, which was expected based on the temperature dependence of peptide bond cleavage.⁸⁹ Fragmentation rates were fastest for pH 5 high IS, then pH 5 low IS formulations, followed by both pH 6.5 formulations. At a given T_{inc} , the three IgG1 MAbs (MAb 3, MAb 5, and MAb 6) had nearly identical fragmentation rates (with respect to fractional concentration) at a given pH and IS that were independent of c_2 , which is consistent with prior reports of IgG1 fragmentation.^{89,265,270} MAb 4 had notably the slowest fragmentation rates, which was expected as the IgG4 subtype is known to be more resistant to fragmentation.¹⁶⁰ This analysis assumes that the fragmentation mechanism is a single step where a monomeric MAb fragments into one ~50 kDa LMW fragment (that is resolved by SEC) and one ~100 kDa HMW fragment (Fc+Fab or Fab+Fab; convoluted with the monomer peak in SEC). If there are other steps that split the HMW fragment into two ~50 kDa fragments, or if any fragments were involved in the formation of or fragmentation of aggregates, then the fragmentation rate of monomeric proteins may not be the same as what would be inferred from analysis of the LMW fragment concentration measured by SEC. The nearly identical fragmentation rates between the three IgG1 MAbs for a given pH and IS would suggest that the same fragmentation pathway was dominant at a given condition and was not significantly impacted by sequence differences or the presence of aggregates in solution. LMW fragment concentrations were linear with time and not dependent on c_2 , suggesting that fragmentation did not significantly impact the measured aggregation rates.

5.3.3 Impacts of conformational stability and self-interactions on aggregation rates

The impact of the formulation conditions (i.e., c_2 , pH, and IS) on relative aggregation rates was not consistent for different T_{inc} conditions, or between the different MAbs, which is agreement with prior reports^{129,160,265} and with the current paradigm of industrial formulation screening.^{24,91,156,271} Formulation development relies in part on high-throughput experimental assays that probe how the formulation conditions impact conformational stability and self-interactions, which are known to be some of the most important factors in mediating protein aggregation.^{24,53,91} Solution conditions have an indirect influence on aggregation rates by mediating changes in more fundamental phenomena that can be specific to the protein in question. For example, pH is not expected to be directly relevant to aggregation in a quantitative sense, but rather how the pH impacts the spatial distribution of charged residues and the resultant intermolecular and intramolecular electrostatic interactions.

In this chapter, the difference between the pI and the pH was fairly correlated with aggregation rates at 130 mg/mL for a given T_{inc} , as shown in Figure 5.8. In this chapter the quantity is described as the absolute value of the difference between the pI and the pH ($|\text{pH}-\text{pI}|$), despite all pI values being greater than 6.5 (see Table 5.2), to clarify that the magnitude of the net valence was likely the relevant fundamental feature, rather than the magnitude and sign of the net valence. $|\text{pH}-\text{pI}|$ was an accurate proxy for net valence as evidenced by Figure D.1.1.1 (Appendix D.1) where $|\text{pH}-\text{pI}|$ was highly correlated with theoretical net valences calculated by the Henderson–Hasselbalch equation with either nominal pK_a values or pK_a values computed by the PROPKA algorithm ($R = 0.93$ and $R = 0.97$, respectively).¹⁸⁶ In practice, one may prefer to use a theoretical calculation if experimental measures are not available, but

here the experimental pI was used because it did not make any assumptions about the pK_a values of ionizable residues.

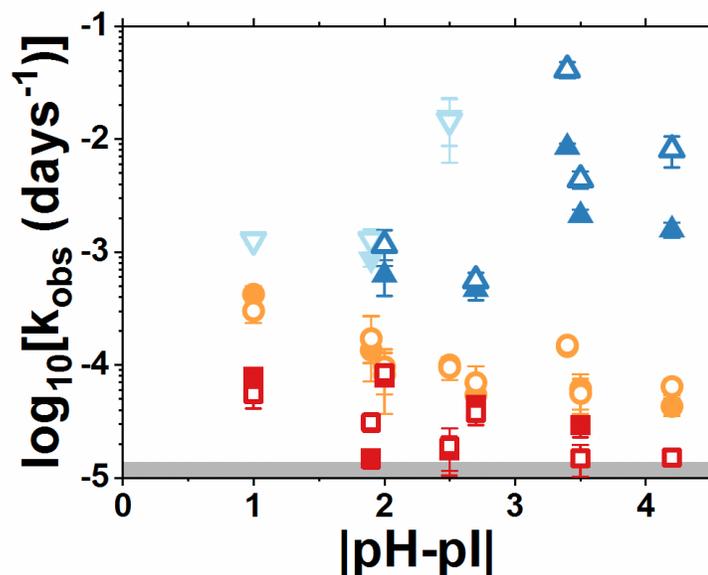


Figure 5.8: Observed aggregation rate coefficient values (k_{obs}) as a function of the difference between the pH and pI ($|\text{pH-pI}|$) for 130 mg/mL formulations at refrigerated (red squares), 30 °C (orange circles), and accelerated (blue triangles) conditions. For the accelerated conditions, 50 °C rates are shown as blue upwards triangles, and 45 °C rates are shown as light blue downwards triangles. Filled and open symbols correspond to low IS and high IS conditions, respectively. The gray region represents the measurement floor of the study ($\log_{10}[k_{obs} (days^{-1})] = -4.86$).

The relationship between $|\text{pH-pI}|$ and aggregation rate was differentiated in terms of the value of the incubation temperature: at accelerated T_{inc} conditions, lower $|\text{pH-pI}|$ conditions had slower aggregation rates ($R = 0.42$, but note that there were two different T_{inc} values), while at 30 °C and refrigerated conditions, lower $|\text{pH-pI}|$ conditions had faster aggregation rates ($R = -0.75$ and $R = -0.62$, respectively). At

refrigerated conditions, the three formulations with aggregation rates below the measurement floor had $|\text{pH-pI}|$ values of 3.4 and 4.2. Analogous plots to Figure 5.8 for 10 and 35 mg/mL formulations are shown in Figure 5.9. Qualitatively, the results were not as clear for accelerated conditions but were consistent for 30 °C conditions when accounting for the many formulations with aggregation rates below the measurement floor. There have been several reported cases of electrostatic repulsions reducing aggregation rates^{43,85,137–140}, which could also explain the negative correlation between $|\text{pH-pI}|$ and aggregation rates for 30 °C and refrigerated conditions in this work, as high net valence will typically lead to strong net repulsive electrostatic self-interactions.^{179,272} Proteins are more likely to undergo thermally induced unfolding at higher incubation temperature, and high net valence can reduce conformational stability as well.^{273–277} If the conformational stability was a primary factor in the dominant aggregation pathway at accelerated conditions, which has been shown in other reports^{92,139,239–245}, that would be consistent with the positive correlation between $|\text{pH-pI}|$ and aggregation rates at accelerated conditions.^{273–277} The competing impacts of conformational stability and self-interactions (typically electrostatic repulsions) with respect to aggregation rates as a function of T_{inc} has been reported previously for MAbs and other proteins and is explored in more detail for this dataset below.^{24,129,141,243,265}

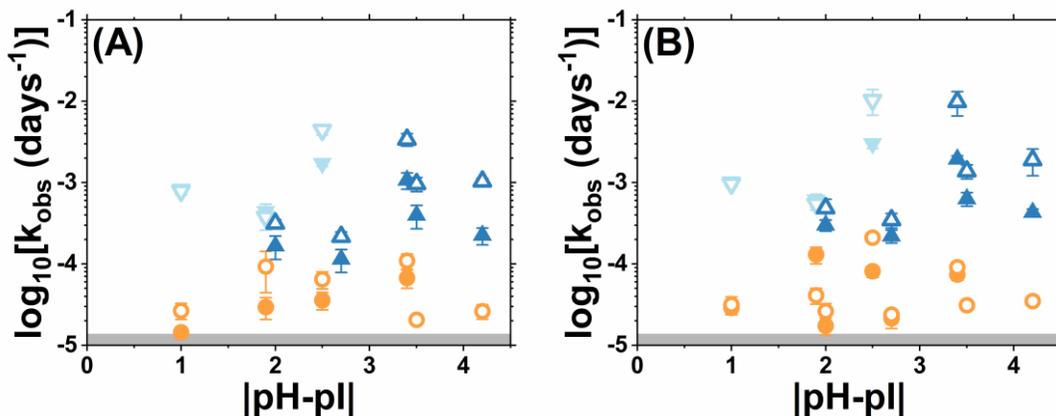


Figure 5.9: Observed aggregation rate coefficient values (k_{obs}) as a function of the difference between the pH and pI ($|\text{pH-pI}|$) for (A) 10 mg/mL and (B) 35 mg/mL formulations at 30 °C (orange circles) and accelerated (blue triangles) conditions. For the accelerated conditions, 50 °C rates are shown as blue upwards triangles, and 45 °C rates are shown as light blue downwards triangles. Filled and open symbols correspond to low IS and high IS conditions, respectively. The gray region represents the measurement floor of the study ($\log_{10}[k_{obs}(\text{days}^{-1})] = -4.86$).

The impacts of each formulation on conformational stability were quantified with DSC, where unfolding events were described based on midpoint unfolding temperatures ($T_{m,app}$ values) of DSC thermograms shown in Figure 5.10. MAbs typically have 3 independent unfolding transitions that correspond to the C_{H2} , C_{H3} , and Fab regions. The first small peak corresponds to the C_{H2} domain, the second small peak corresponds to the C_{H3} domain, and the large peak corresponds to the Fab fragment, any of which might overlap in a given thermogram.^{89,259} $T_{m,app}$ values were used as a proxy for more specific thermodynamic quantities that describe conformational stability such as the Gibbs free energy of unfolding for a given unfolding transition that were not accessible in this work as unfolding was likely irreversible in most cases (see section 5.2.2).^{43,92} The DSC results from this work

(Figure 5.10 and Table 5.4) confirmed that for a given MAb, the pH 5 conditions had lower T_{onset} and $T_{m,app1}$ values than the pH 6.5 conditions, except for MAb 3 where the low IS formulations had approximately the same T_{onset} and $T_{m,app1}$ values.

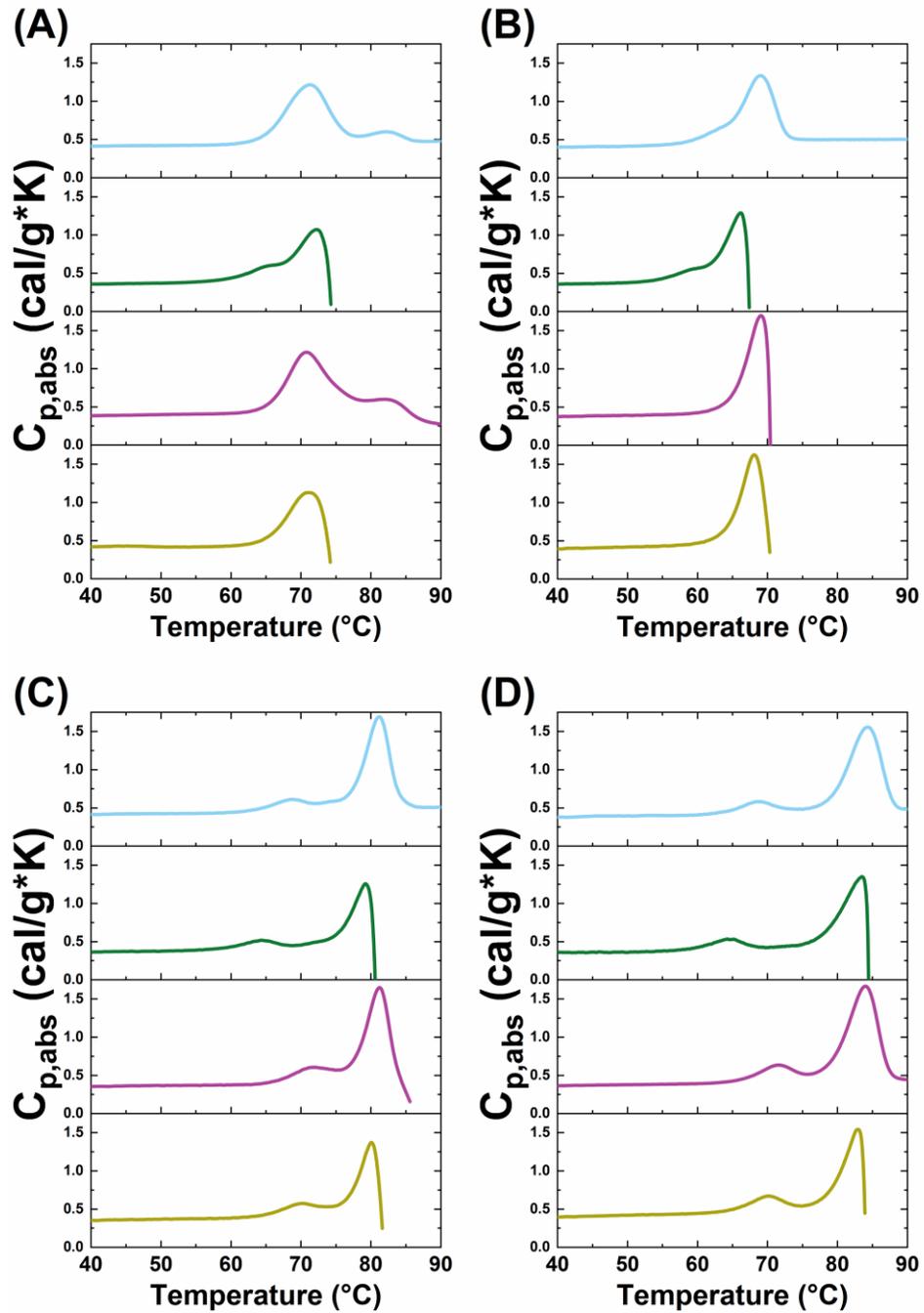


Figure 5.10: DSC thermograms for MAb 3 (A), MAb 4 (B), MAb 5 (C), and MAb 6 (D) in the following buffer solutions: pH 5 low IS (blue), pH 5 high IS (green), pH 6.5 low IS (purple) and pH 6.5 high IS (yellow).

Table 5.4: Onset unfolding temperature and midpoint unfolding temperatures for each formulation via DSC

MAB	Formulation	$T_{m,onset}$ (°C) ^α	$T_{m,app1}$ (°C)	$T_{m,app2}$ (°C) ^β
MAB 3	pH 5 low IS	62.8	71.2	82.2
	pH 5 high IS	58.4	65.0	72.3
	pH 6.5 low IS	63.2	70.8	82.1
	pH 6.5 high IS	63.2	71.1	N/A
MAB 4	pH 5 low IS	59.3	63.0	69.0
	pH 5 high IS	53.4	59.0	66.2
	pH 6.5 low IS	61.4	69.0	N/A
	pH 6.5 high IS	60.4	68.2	N/A
MAB 5	pH 5 low IS	63.2	68.7	81.1
	pH 5 high IS	59.1	64.4	79.3
	pH 6.5 low IS	65.7	72.0	81.2
	pH 6.5 high IS	64.4	70.4	80.1
MAB 6	pH 5 low IS	63.1	68.7	84.4
	pH 5 high IS	58.1	64.3	83.5
	pH 6.5 low IS	65.3	71.6	84.0
	pH 6.5 high IS	64.1	70.3	82.9

^α $T_{m,onset}$ was calculated as the temperature where $c_{p,abs}$ had increased 10% from the baseline value.

^β N/A is listed for conditions where a sharp exotherm occurred before a second peak.

The difference between the first midpoint unfolding temperature ($T_{m,app1}$) and the T_{inc} was highly correlated to the aggregation rate at accelerated conditions for a given MAb and c_2 , as shown in Figure 5.11. The linear fits all had high correlation coefficients ($|R| > 0.95$ for all, except MAB 4 at 130 mg/mL, where $|R| = 0.89$). In Figure 6, the lowest value of $T_{m,app1} - T_{inc}$ (~14 °C for all 4 MAbs) corresponded to

the pH 5 high IS formulation, the second lowest value to the pH 5 low IS formulation, and the two highest values to the two pH 6.5 formulations.

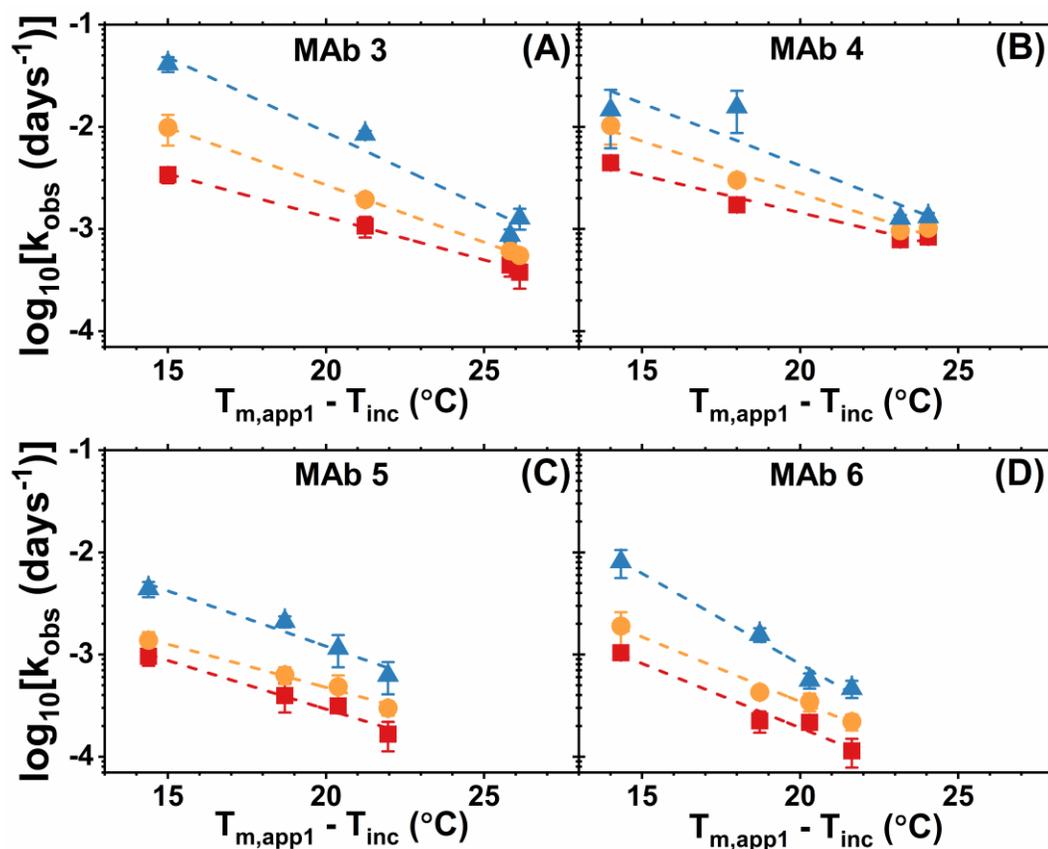


Figure 5.11: Aggregation rates (k_{obs} values) at accelerated conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the T_{inc} (45 °C or 50 °C, see Table 5.2) and the first midpoint unfolding temperature ($T_{m,app1}$). Rates at c_2 values of 10, 35, and 130 mg/mL are shown as red squares, orange circles, and blue triangles, respectively. Dashed lines are linear fits for a given c_2 and error bars are 95% confidence intervals.

These results suggest that aggregation rates at accelerated conditions for a given MAb were mediated by conformational stability. Mechanistically, that is consistent with some degree of unfolding being involved in the formation of key intermediate species that impacts the overall aggregation rate.²⁶³ However, as mentioned previously, the unfolding step itself was not rate limiting because the aggregation rates (based on monomer fraction) were dependent on c_2 (i.e., the rate-limiting step was not unimolecular). Corresponding plots for 30 °C and refrigerated conditions are shown in Figure 5.12 and Figure 5.13, respectively. Those results did not reveal the same relationship and statistical uncertainties were much higher.

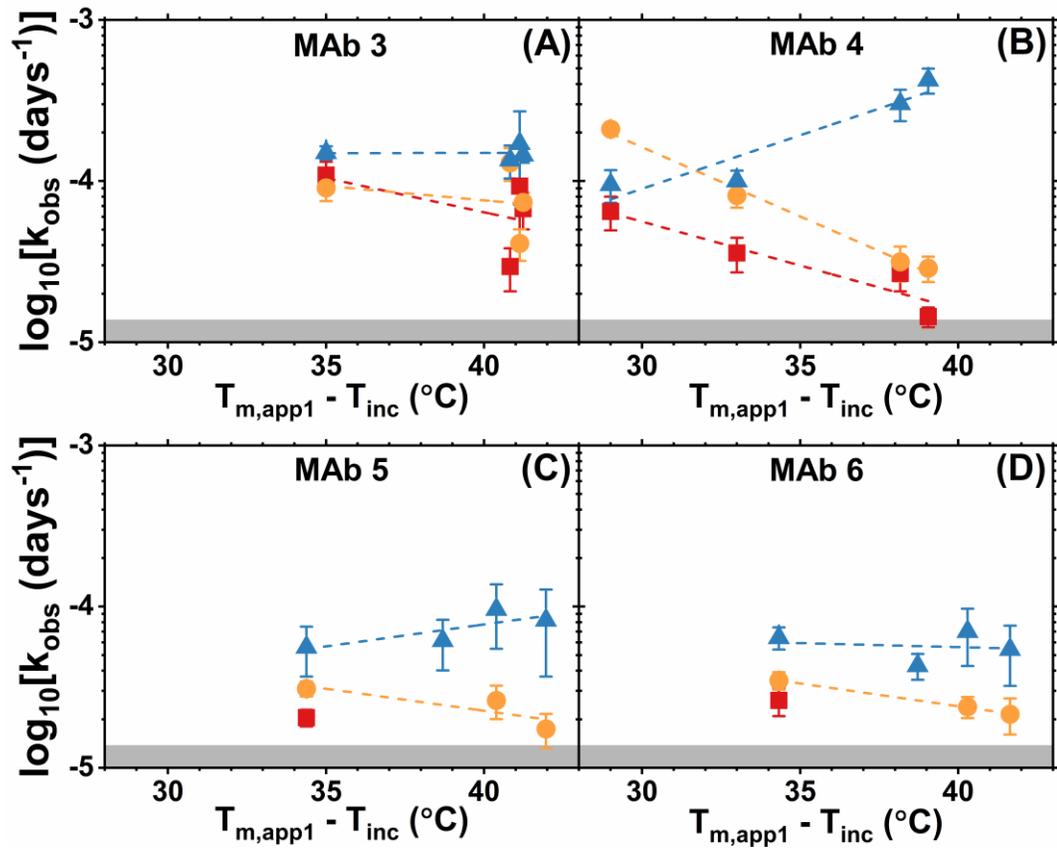


Figure 5.12: Aggregation rates (k_{obs} values) at 30 °C conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the T_{inc} (30 °C) and the first midpoint unfolding temperature ($T_{m,app1}$). Rates at c_2 values of 10, 35, and 130 mg/mL are shown as red squares, orange circles, and blue triangles, respectively. Dashed lines are linear fits for a given c_2 and error bars are 95% confidence intervals. The gray region represents the measurement floor of the study ($\log_{10}[k_{obs} \text{ (days}^{-1}\text{)}] = -4.86$).

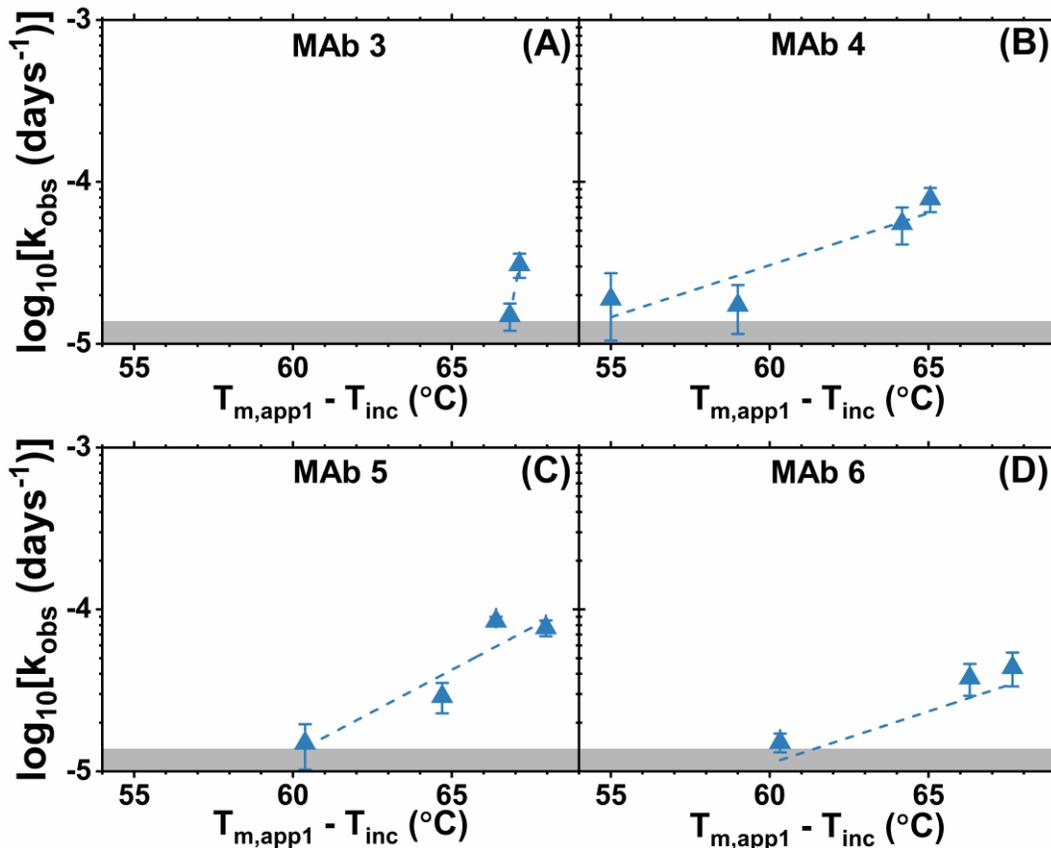


Figure 5.13: Aggregation rates (k_{obs} values) at refrigerated conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the T_{inc} (4 °C) and the first midpoint unfolding temperature ($T_{m,app1}$). Rates at 130 mg/mL are shown as blue triangles. Dashed lines are linear fits for a given c_2 and error bars are 95% confidence intervals. The gray region represents the measurement floor of the study ($\log_{10}[k_{obs} \text{ (days}^{-1}\text{)}] = -4.86$).

For MAb 5 and MAb 6, $T_{m,app1}$ appears to correspond to unfolding of the C_H2 domain for all the formulations, which was likely relevant to the dominant aggregation pathway because the correlation with $T_{m,app2} - T_{inc}$ was significantly worse than that for $T_{m,app1} - T_{inc}$ ($|R| \sim 0.7$ for MAb 5 and $|R| \sim 0.0$ for MAb 6; see Figure 5.14).

Many DSC thermograms for MAb 3 and MAb 4 contained sharp exotherms or

convoluted peaks that were less straightforward to interpret, although the $T_{m,app1}$ for the pH 5 high IS condition also appeared to correspond to the C_{H2} domain for both MAbs. It should be noted that the “true” first midpoint unfolding temperature in a thermodynamic sense may be somewhat higher than what is reported here for conditions where irreversible aggregation and precipitation occurred simultaneously with the first unfolding transition.⁴³ In practice, that limitation could be partially overcome by the use of a capillary DSC instrument rather than the instrument used in this work.²⁶⁵ The DSC measurements in this work also do not account for any potential c_2 dependence in conformational stability, which is challenging to probe experimentally but has been theorized to have an impact on the Gibbs free energy of unfolding and has been shown experimentally to impact the structure or stability of the folded state.^{43,95–98}

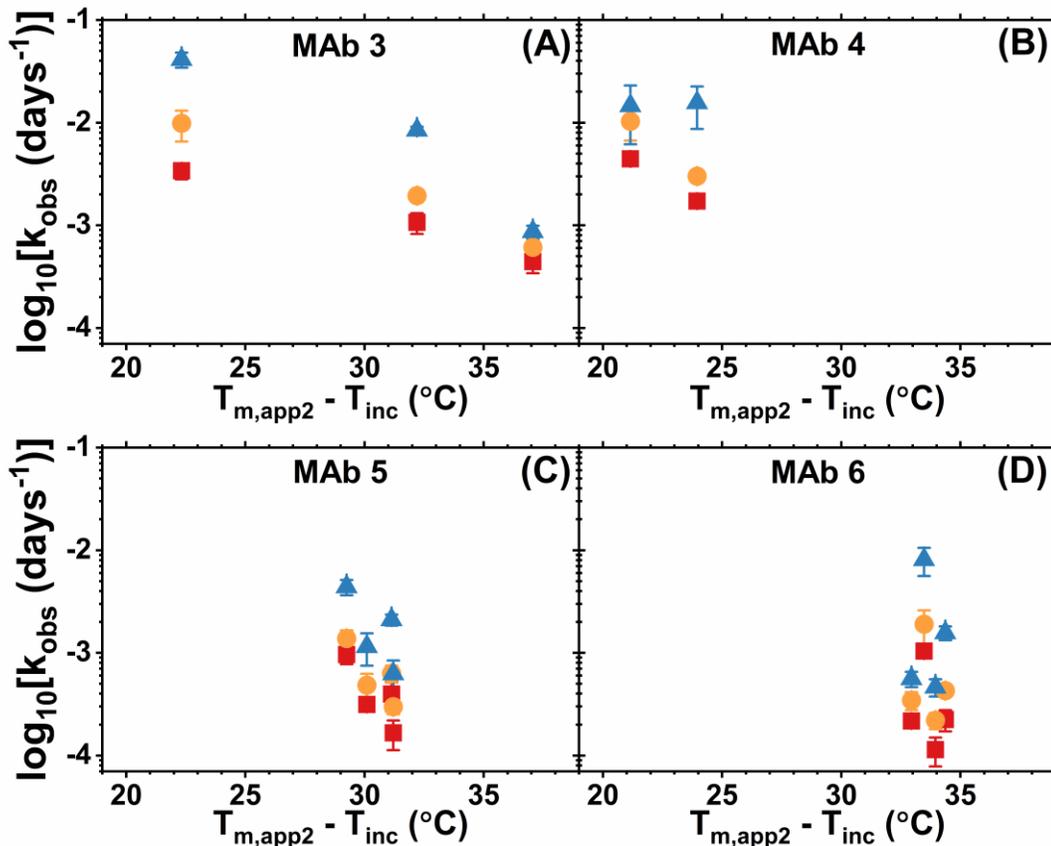


Figure 5.14: Aggregation rates (k_{obs} values) at accelerated conditions for MAb 3 (A), MAb 4 (B), MAb 5 (C) and MAb 6 (D) plotted against the difference between the T_{inc} (45 °C or 50 °C, see Table 5.2) and the second midpoint unfolding temperature ($T_{m,app2}$). Rates at c_2 values of 10, 35, and 130 mg/mL are shown as red squares, orange circles, and blue triangles, respectively. Error bars are 95% confidence intervals.

A parallel dataset (similar pH, IS, and c_2 values) of aggregation rates only at 50 °C were reported for a different MAb in prior work (referred to as MAb 1, the same MAb as in Chapters 2 and 4).^{43,48–50,179} Additional studies at 30 °C or refrigerated conditions for MAb 1 were not performed due to material constraints. MAb 1 had a pI of 8.6, and the relative high-concentration aggregation rates between pH 5 and pH 6.5 conditions were in qualitative agreement with the results with respect to $|\text{pH-pI}|$ in

Figure 5.9.⁴⁸ Aggregation rates for MAb 1, reproduced in Figure 5.15, were in many cases orders of magnitude larger than those for the four MAbs in this work. MAb 1, compared to the MAbs studied in this work, had less spread in $T_{m,app1} - T_{inc}$ (approximately 4 °C), but the analogous plot to Figure 5.11 for MAb 1 (Figure 5.16) suggests a similar relationship between conformational stability (i.e., $T_{m,app1} - T_{inc}$) and aggregation rate at a given c_2 ($|R| > 0.80$).

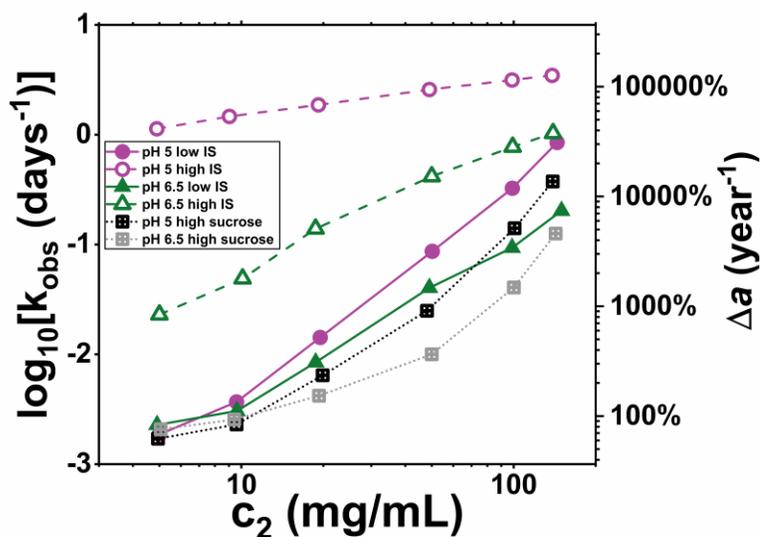


Figure 5.15: Observed aggregation rate coefficient (k_{obs}) values as a function of MAb concentration (c_2) at 50 °C for MAb 1, with the corresponding increase in aggregate fraction (Δa) per year on the right y-axis. Symbols correspond to Table 5.1, with two other solution conditions Lines connecting the data points are guides to the eye and error bars are 95% confidence intervals. Adapted from Ghosh et al.⁴³

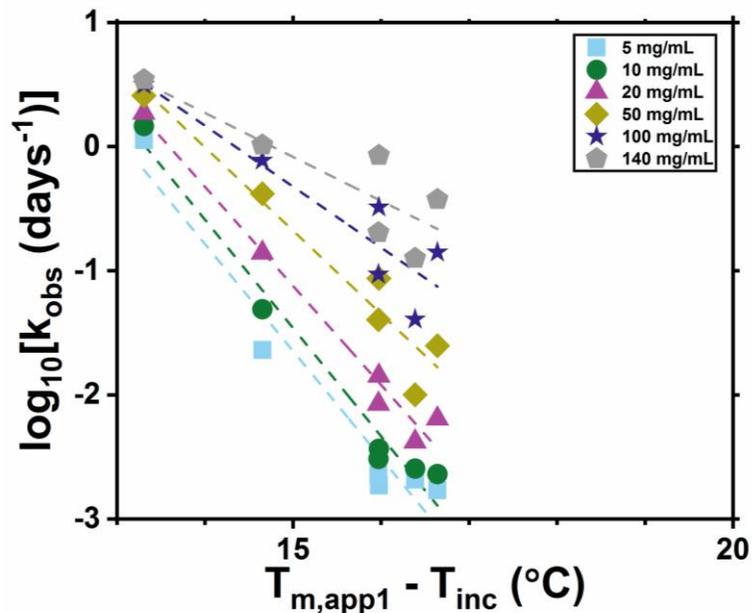


Figure 5.16: Aggregation rates at accelerated temperature conditions for MAb 1 plotted against the difference between the T_{inc} (50 °C) and the first midpoint unfolding temperature ($T_{m,app1}$). Dashed lines are linear fits for a given MAb concentration and error bars are 95% confidence intervals. Note that the x-axis is narrowed compared to Figure 5.11 in the main text. Data was adapted from Ghosh et al.⁴³

SLS and DLS measurements of net self-interactions (i.e., values of B_{22} and G_{22} from SLS, and values of k_D and D_c from DLS, shown in Chapter 4) showed that for all conditions except MAb 4 at pH 6.5 (which had the lowest $|\text{pH-pI}|$ at 1.0), electrostatic interactions were net repulsive, and typically displayed “simple colloidal” behavior where the strength of electrostatic repulsions were reduced at high IS due to Debye screening. These results were not surprising due to the expected high net valence for the $|\text{pH-pI}|$ conditions in this chapter.^{179,272} The mild impact of IS on 130 mg/mL aggregation rates at 30 °C and refrigerated conditions could be explained by the crowded environment at high c_2 where proteins are forced to be fairly close to

each other and electrostatic repulsions are therefore not as successful at keeping proteins apart and inhibiting aggregation.

In prior work focused on MAb 1, a semi-quantitative correlation between the change in aggregation rate at 50 °C and reduction in net repulsive electrostatic self-interactions due to increased IS was reported.⁴³ The net self-interactions were quantified by the c_2 -dependent quantity G_{22} via SLS. The semi-quantitative correlation reported in prior work was compared to the results in this work at accelerated, 30 °C, and refrigerated conditions (shown in Figure 5.17). The correlation did not apply to the MAbs in this work at any of the T_{inc} conditions either broadly or for each MAb individually. However, the relationship was qualitatively consistent at accelerated conditions, where high IS formulations typically had faster aggregation rates than their low IS counterparts (see also Figure 5.5B).

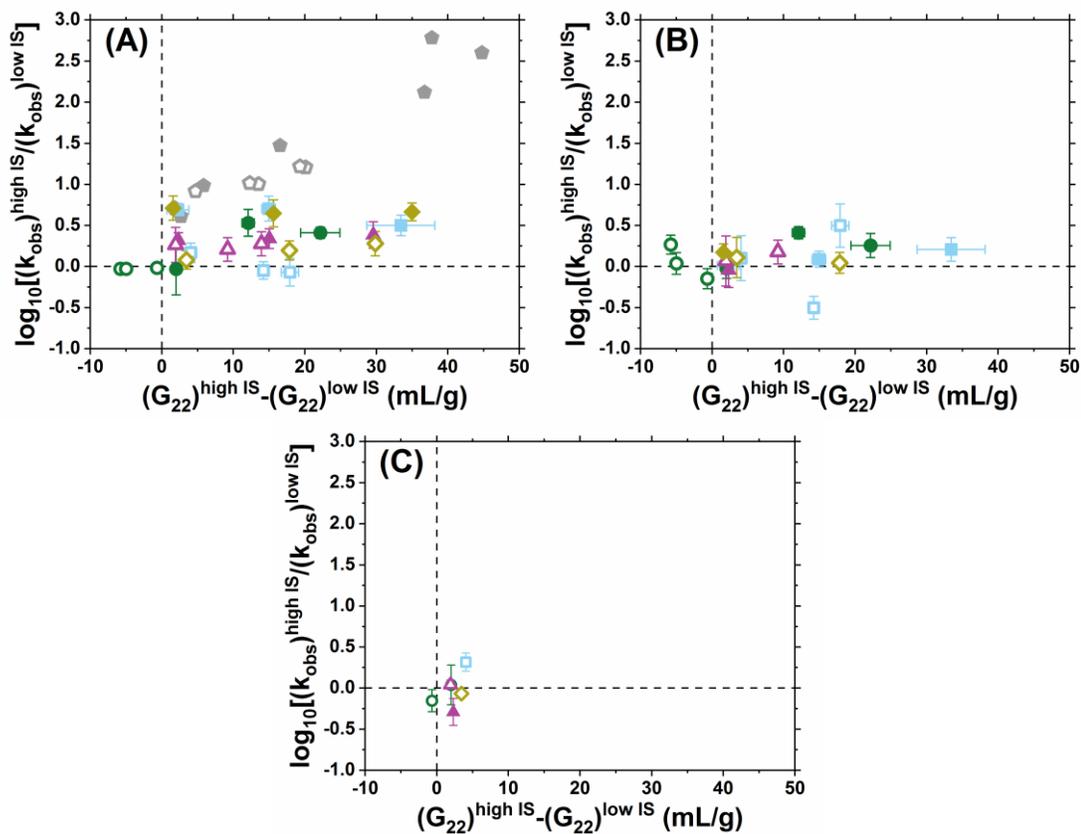


Figure 5.17: Comparison of the change in G_{22} and the change in the aggregation rate (k_{obs} values) at (A) accelerated, (B) 30 °C, and (C) refrigerated conditions due to addition of 100 mM NaCl (referred to as high IS in this chapter). Data for MAb 3, MAb 4, MAb 5, and MAb 6 are shown as light blue squares, green circles, purple triangles, and yellow diamonds, respectively. pH 5 conditions are filled symbols, and pH 6.5 conditions are open symbols. Data for MAb 1 (gray pentagons, panel A only) at 50 °C are reproduced from Ghosh et al.⁴³

Several other quantities that could be used to represent high-concentration net self-interactions were considered in this analysis and for use in the ML models described in the following section. Simultaneous DLS measurement of the 130 mg/mL samples yielded the collective diffusion coefficient (D_c), which is analogous to G_{22} but also includes contributions from hydrodynamic interactions.^{221,278} Low-

concentration SLS and DLS measurements were used to calculate B_{22}^* and k_D , respectively. Those quantities are more experimentally accessible, but only describe two-body net self-interactions which do not include self-interactions specific to high-concentration systems such as increased steric repulsions or multibody interactions. Because the impact of self-interactions on the aggregation rates was hypothesized to be due to electrostatic repulsions, the difference between the $B_{22}/B_{22,ST}$ (referred to hereafter as B_{22}^*) value at a given formulation and the B_{22}^* value at > 300 mM IS (referred to as $B_{22,NE}^*$) was also considered (referred to as ΔB_{22}^*). At > 300 mM IS conditions, electrostatic interactions were effectively fully screened as the value of B_{22}^* was not dependent on pH or IS. The net valence (calculated via the Henderson–Hasselbalch equation with nominal pK_a values or pK_a values calculated by the PROPKA algorithm¹⁸⁶) was multiplied by the Debye length (λ_{DE}) to also represent screened electrostatic repulsions, similar to ΔB_{22}^* . All of the aforementioned quantities were highly correlated with one another, as shown in the correlation matrix in Figure 5.18. For any pair of quantities in Figure 5.18, the magnitude of the correlation coefficient ($|R|$) was at least 0.80. G_{22} was expected to be inversely correlated as B_{22} is calculated via $-\frac{1}{2} \lim_{c_2 \rightarrow 0} G_{22}$. These results suggest that for a given c_2 , explicit measurements of net self-interactions (i.e., G_{22} or D_c values) can be semi-quantitatively estimated by low- c_2 measurements or even theoretical calculations for at least the range net self-interactions in this chapter (i.e., strong net repulsions to weak net attractions). Chapter 4 showed that molecular simulations could also be used for this task with quantitative accuracy in many cases, typically parameterized by measurements of B_{22}^* .^{48,50}

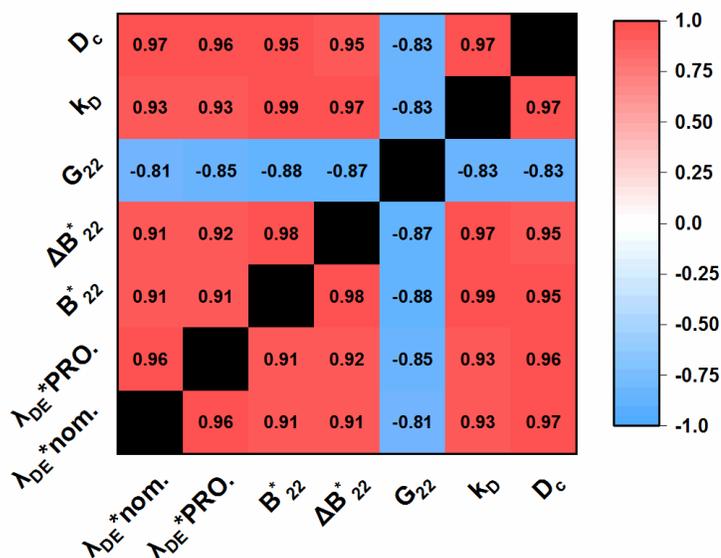


Figure 5.18: Correlation coefficients (R values) for quantities that were considered to represent net self-interactions at 130 mg/mL. D_c and G_{22} were measured at 130 mg/mL and all others listed on the axes were not c_2 dependent.

The measurements of conformational stability and self-interactions in this chapter were consistent with MAbs and formulations that would not be expected to be particularly problematic with respect to aggregation or other related physical instabilities such as low solubility or liquid-liquid phase separation.^{16,109} Midpoint unfolding temperatures from DSC were fairly high, and net self-interactions ranged from strong repulsions to weak attractions, which is consistent with prior reports of approved MAb products.^{54,279} Still, many of these formulations had significant aggregation rates at high-concentration, even at refrigerated conditions. The dataset in this chapter is thus best interpreted in the context of optimizing formulation conditions for protein candidates that have successfully passed through early-stage screening. The dataset here is likely not representative of the broader range of MAbs and formulations

that are tested at early stages of candidate selection where there are typically more poorly behaved conditions.

5.3.4 ML models for feature selection and prediction of high-concentration aggregation rates

The dataset in this chapter was designed to systematically map the impact of formulation conditions and incubation temperatures (T_{inc}), within typical scope of commercial products and industrial stability studies.^{3,91,252,262} The impacts of those conditions on phenomena relevant to aggregation (e.g., self-interactions and conformational stability) were quantified by several theoretical or experimental methods but as evidenced with respect to net self-interactions in Figure 5.18, there were many measures that were highly collinear. To reduce the dimensionality of the dataset and avoid overfitting in the statistical models outlined below, it was necessary to use a systematic approach to reduce the initial large feature set to a smaller set that were not highly correlated and represented distinct properties or behaviors. This procedure is described in detail in Appendix D.1. The final feature set, shown in Table 5.5, included features that describe the solution conditions, net self-interactions, conformational stability, and results from stability studies at higher T_{inc} values or lower c_2 values.

Table 5.5: Feature coefficients and model performance for elastic net regression models of 130 mg/mL aggregation rates

Feature ^a	Refrigerated (linear)	Refrigerated (logistic)	30 °C (linear)	Accelerated (linear)
pH – pI	-0.111 ± 0.007 (-0.188 ± 0.008)	0 (-0.773 ± 0.107)	-0.171 ± 0.016 (-0.190 ± 0.017)	0.106 ± 0.014 (0)
ln(IS)	0 (0)	0 (0)	0 (0)	0 (0)
ΔB_{22}^*	0 (0)	0 (0)	0 (0)	0 (0)
$B_{22,NE}^*$	0 (0)	0 (0)	0 (0)	0 (0)
$T_{m,app1}$ – T_{inc}	0 (0)	0 (0.753 ± 0.089)	0 (0)	0 (-0.222 ± 0.013)
y-intercept	-4.59 (-4.58)	0.442 (0.466)	-4.00 (-4.00)	-2.59 (-2.60)
Features derived from other stability studies				
$k_{obs,Arr\ ex}$	0.110 ± 0.011	1.97 ± 0.20	N/A	N/A
$k_{obs,10}$	N/A	N/A	0.095 ± 0.020	N/A
$k_{obs,acc}$	N/A	N/A	0	N/A
$k_{obs,c_2\ ex}$	N/A	N/A	N/A	0.521 ± 0.021
Model performance				
R ^β	0.61 (0.52)	0.66 (0.58)	0.73 (0.56)	0.93 (0.49)
Error ^γ	0.213 ± 0.004 (0.237 ± 0.003)	3.58 ± 0.23 (4.36 ± 0.21)	0.167 ± 0.005 (0.208 ± 0.007)	0.155 ± 0.005 (0.418 ± 0.006)
Null error ^δ	0.308	5.91	0.284	0.516

^a Values in parenthesis are for models without features derived from other stability studies and uncertainties are 95% confidence intervals of the coefficient values.

^β Correlation coefficient between the predicted and measured aggregation rates

^γ Error was MAE for linear models and deviance for logistic models

^δ Null error is the intercept-only error, where all coefficients besides the constant term are set to 0 and all predicted k_{obs} values are the same

The quantity $\ln(\text{IS})$ captures the Debye screening effect and was linearized via the natural log because IS has an exponential relationship with Debye length.²⁹ $|\text{pH}-\text{pI}|$ essentially captures differences in the net valence at a given pH (see also Figure 5.8). $T_{m,app1} - T_{inc}$ is a surrogate for changes in the conformational stability at a given T_{inc} , pH, and IS (see also Figure 5.11). At 30 °C and refrigerated conditions, T_{inc} is invariant, so $T_{m,app1} - T_{inc}$ is the same as $T_{m,app1}$ because the data were standardized for this analysis. $B_{22,NE}^*$ was measured at > 300 mM IS where electrostatic interactions were largely screened and describes net non-electrostatic self-interactions, such as those from hydrophobic interactions and hydrogen bonding. ΔB_{22}^* , the difference between $B_{22,NE}^*$ and B_{22}^* at a given IS, quantifies the electrostatic contribution to B_{22}^* . The correlation coefficient for each of those features with 130 mg/mL aggregation rates as a function of T_{inc} is shown in Figure 5.19. The results in Figure 5.19 reiterate that for this dataset, the phenomena relevant to aggregation can change in their relative importance and directional impact (i.e., promoting or inhibiting aggregation) depending on the choice of incubation temperature. The features themselves are somewhat correlated, however (see Appendix D.1.7), and more sophisticated analysis (i.e., the ML models developed below) was required to deconvolute the variance in aggregation rates that was captured by each feature.

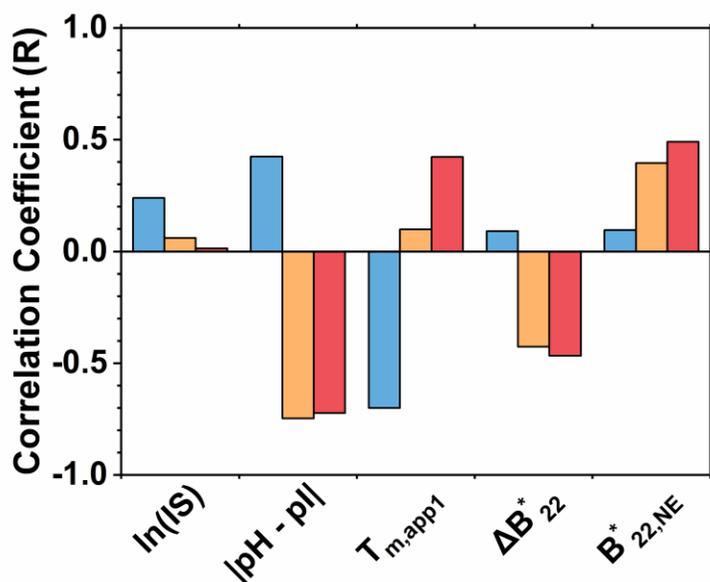


Figure 5.19: Correlation coefficients (R values) between input features and 130 mg/mL aggregation rates (k_{obs} values) at accelerated (blue), 30 °C (orange), and refrigerated (red) conditions. $T_{m,app1} - T_{inc}$ is shortened to $T_{m,app1}$ for readability.

A diagram of the ML statistical framework described below is shown in Figure 5.20. The gray boxes represent the full set of inputs to the model that broadly fall into two categories: information intrinsic to a given MAb and stability study, and results from experimental measurements. The green boxes are inputs into the statistical model. The green box labeled “Feature screening” refers to the process described in Appendix D.1 wherein the large set of input features were filtered to a subset that describe distinct phenomena and are not highly collinear. The other green box, labeled “ k_{obs} database” is the measured aggregation rates at the condition(s) of interest that are used as training data. The purple boxes and arrows are all part of the ML algorithm that is described below, and the light blue box is the output: a numerical model of the most important or relevant features and their relative weights that can be applied to

predict aggregation rates for each of the conditions in the training data and perhaps for therapeutic proteins outside the dataset.

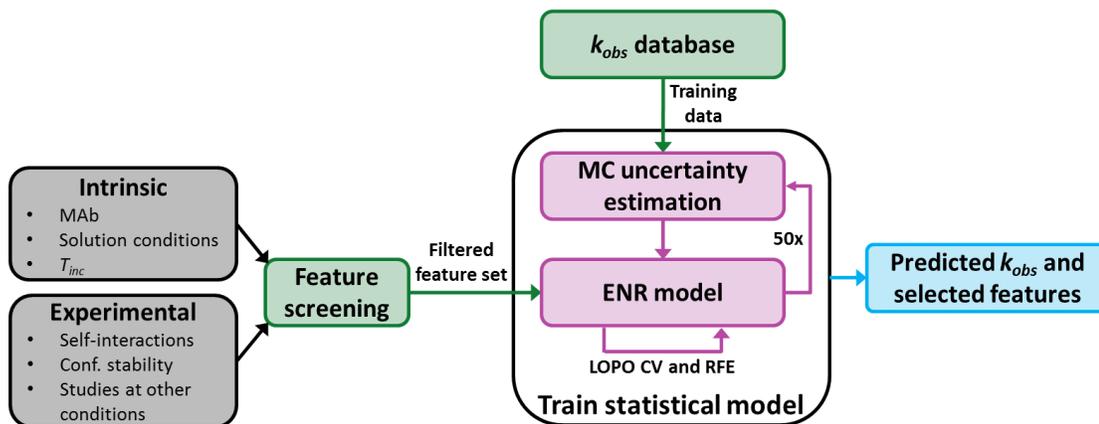


Figure 5.20: Statistical framework applied in this section. Green boxes are the inputs and the light blue box is the output.

ENR models of 130 mg/mL aggregation rates for a given T_{inc} value were built with the features shown in Table 5.5 (y-intercepts are included for completeness). ENR was used to build linear combination models (i.e., $\log_{10}(k_{obs}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$), where x_i is a given feature in Table 5.5. For refrigerated conditions, logistic regression models were also considered, where the aggregation rates were transformed into binary data via the threshold $\log_{10}[k_{obs}(\text{days}^{-1})] > -4.56$ (1% aggregation per year). Logistic models predict the log-odds (logit) of the probability of the binary data being true (i.e., $P(\log_{10}[k_{obs}(\text{days}^{-1})] > -4.56) = (1 + \exp[-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots)])^{-1}$). The logistic models explicitly consider the aggregation rates that were below the measurement threshold and are perhaps better suited in cases like the dataset in this chapter where many of the measured rates had

relatively significant uncertainties. Feature coefficients from logistic models vary in their magnitudes from the linear models and should not be directly compared. A few elements were added to the ENR models to improve their ability to be applied not only to predict aggregation rates, but also to deconvolute the impacts of each feature and more rigorously identify which features were statistically significant. The models were iteratively generated with RFE to remove the least important features such that the final models (shown in Table 5.5) only contained features that reduced the error of the predicted k_{obs} values in a statistically significant manner ($p < 0.05$). LOPO CV²⁴⁸ was used so that the models explicitly considered MAb identity and would be more robust for predictions of aggregation rates of MAbs outside of this dataset. MC sampling was used for uncertainty estimation to account for the significant uncertainties in some of the k_{obs} values. Feature coefficients and model error values had uncertainties due to MC sampling, which were useful to reduce overfitting and improve how robust the feature selection procedure was.

ENR models with and without features derived from stability studies at other conditions (i.e., higher T_{inc} or lower c_2 values) were considered. For accelerated conditions, the predicted aggregation rate via extrapolation from 10 and 35 mg/mL studies was used to capture differences in the dependence of k_{obs} on c_2 of the different formulations ($R = 0.95$; referred to as $k_{obs,c_2 ex}$). For 30 °C conditions, the aggregation rate at 10 mg/mL ($R = 0.41$; referred to as $k_{obs,10}$), and the aggregation rate at accelerated conditions ($R = 0.12$; referred to as $k_{obs,acc}$) were both considered. The aggregation rates at accelerated conditions were incongruous because of the two different T_{inc} conditions chosen (see Table 5.2), so the feature was adjusted by extrapolating the k_{obs} values at 45 °C to 50 °C via Arrhenius extrapolation with the

k_{obs} values at 30 °C (see Figure D.1.5.2). For refrigerated conditions, the Arrhenius extrapolated rate from 30 °C and accelerated conditions (see also Figure 5.6) was used as an input ($R = 0.70$; referred to as $k_{obs,Arr ex}$). A review of the highly correlated features from stability studies at other conditions that were also considered is shown in Appendix D.1. Predicted k_{obs} values where features derived from stability studies at other conditions were considered are shown in Figure 5.21, and the analogous predictions without features from other stability studies are shown in Figure 5.22. Feature coefficients for each of the models are shown in Table 5.5. Because the features were standardized for ENR, the coefficients are directly comparable (with the exception of the logistic ENR models) where a larger magnitude indicates more significant relationship.

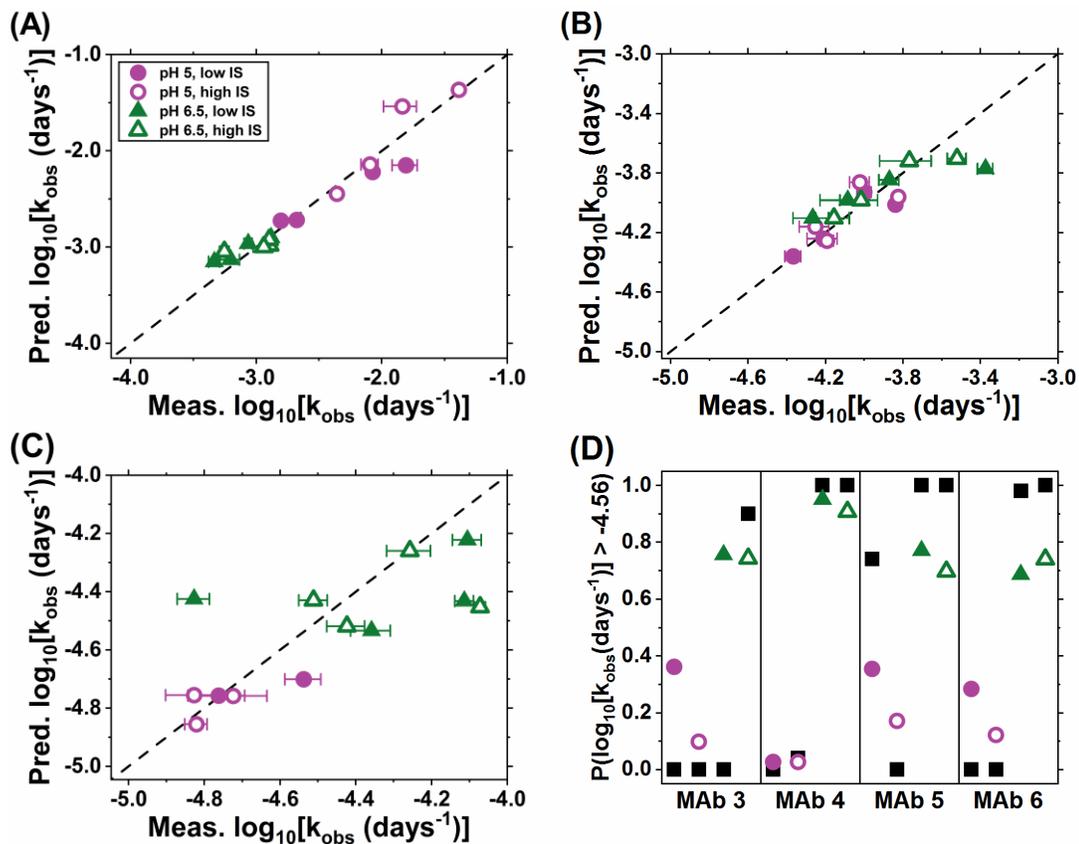


Figure 5.21: ENR model predictions for 130 mg/mL aggregation rates for models with feature(s) derived from stability studies at other conditions at: (A) accelerated conditions, (B) 30 °C, and (C and D) refrigerated conditions. Panels A, B, and C are parity plots of measured aggregation rates (x-axis) and aggregation rates predicted by a linear ENR model (y-axis). Panel D shows predicted aggregation rates by the logistic ENR model with $(\log_{10}[k_{obs}(\text{days}^{-1})] > -4.56)$ as the boundary criterion, where observed aggregation rates are shown as filled black squares. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals.

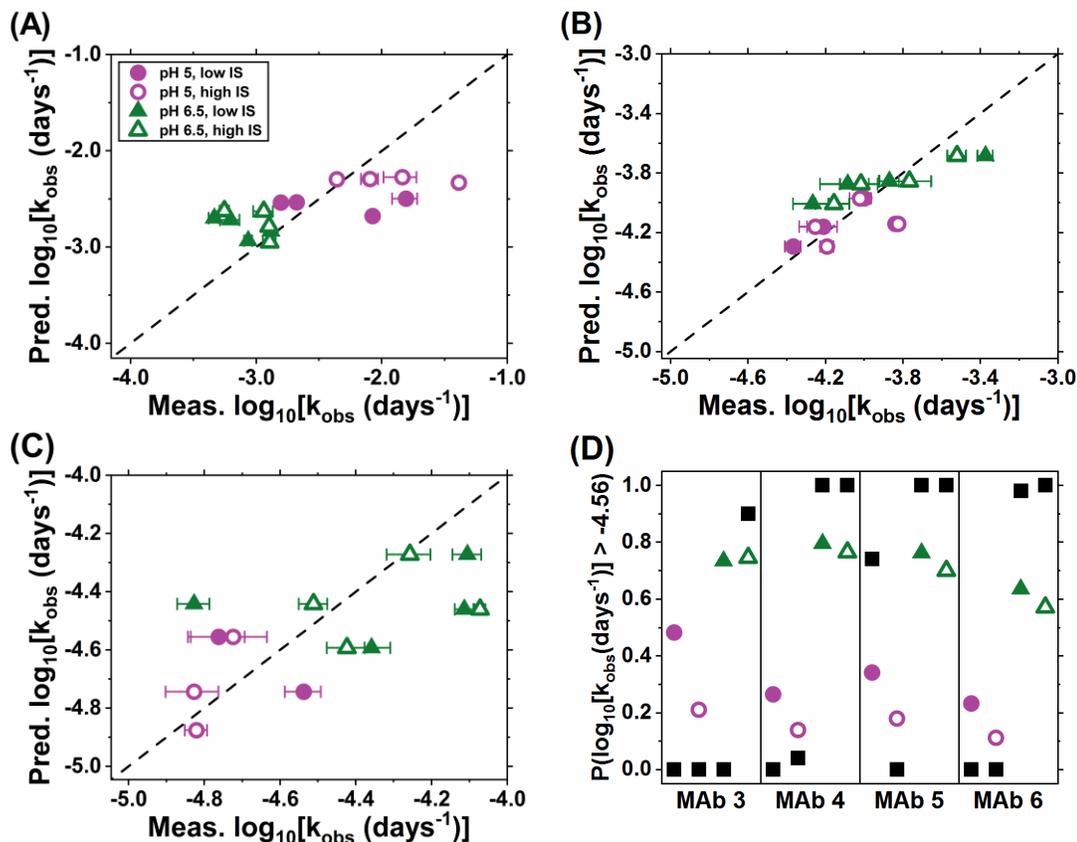


Figure 5.22: ENR model predictions for 130 mg/mL aggregation rates for models without feature(s) derived from stability studies at other conditions at: (A) accelerated conditions, (B) 30 °C, and (C and D) refrigerated conditions. Panels A, B, and C are parity plots of measured aggregation rates (x-axis) and aggregation rates predicted by a linear ENR model (y-axis). Panel D shows predicted aggregation rates by the logistic ENR model with $(\log_{10}[k_{obs} \text{ (days}^{-1}\text{)}] > -4.56)$ as the boundary criterion, where observed aggregation rates are shown as filled black squares. The legend is the same for all panels where symbols for each formulation correspond to Table 5.1 and error bars are 95% confidence intervals.

The ENR model for accelerated incubation temperatures selected $k_{obs,c_2 ex}$ as well as $|\text{pH-pI}|$. $k_{obs,c_2 ex}$ was expected to be selected due to the high correlation of $R = 0.95$. A model with only $k_{obs,c_2 ex}$ had a 65% reduction from the null error ($R =$

0.90), which mildly improved to 69% when $|\text{pH-pI}|$ was included ($R = 0.93$). When $k_{obs,c_2\ ex}$ was excluded, the model selected only $T_{m,app1} - T_{inc}$, but with only a ~20% reduction from the null error ($R = 0.49$). $T_{m,app1} - T_{inc}$ was only mildly useful as a quantitative predictor of aggregation rates when considering multiple MAbs, but as noted previously it was useful in understanding the differences in aggregation rates between formulations for a specific MAb and c_2 , which suggests that there were other MAb-specific traits not captured by the features considered here that impacted the aggregation rates.

Similarly, the linear models for 30 °C and refrigerated conditions were improved by results from other stability studies ($k_{obs,10}$ and $k_{obs,Arr\ ex}$, respectively), and $|\text{pH-pI}|$ was the other selected feature that helped to account for additional variance in the measured aggregation rates. When the features derived from other stability studies were excluded, $|\text{pH-pI}|$ was the only selected feature and the coefficient increased in magnitude (i.e., it was more strongly weighted). The linear model for refrigerated aggregation rates had a 31% reduction from the null error ($R = 0.61$) with $k_{obs,Arr\ ex}$ and $|\text{pH-pI}|$, which dropped to 23% ($R = 0.52$) with only $|\text{pH-pI}|$. The model for 30 °C aggregation rates had a 41% reduction from the null error ($R = 0.73$) with $k_{obs,10}$ and $|\text{pH-pI}|$, which dropped to 27% ($R = 0.56$) with only $|\text{pH-pI}|$. The results from accelerated conditions were not at all useful for predicting aggregation rates at 30 °C as the ML models did not select the $k_{obs,acc}$ feature, and the correlation with the k_{obs} at 30 °C was weak ($R = 0.12$). However, the 10 mg/mL aggregation rate at 30 °C ($k_{obs,10}$; $R = 0.41$) was selected by the ML model, despite changes in intermolecular interactions due to greatly reduced crowding, and the

significant number of conditions where the $k_{obs,10}$ value was below the measurement floor (6 out of 16).

By inspection of Figure 5.3, it is clear that the aggregation rates for a given MAb at refrigerated conditions were primarily distinguished by the pH of the formulation. Both the linear and logistic models selected $k_{obs,Arr ex}$ as an important feature, which also clearly separated the formulations by pH (see Figure 5.6). This was likely incidental; conformational stability clearly mediated aggregation rates at accelerated conditions, which resulted in faster aggregation rates for pH 5 conditions, but results were generally more mixed at 30 °C. That lead to a reversal in the relative formulation rankings for $k_{obs,Arr ex}$. Without $k_{obs,Arr ex}$, both models selected $|pH-pI|$ (as discussed above), but the logistic model also selected $T_{m,app1} - T_{inc}$. This was likely also incidental as $T_{m,app1}$ values from DSC were also separated by pH. Also, the positive coefficient would suggest that formulations with higher conformational stability would have faster aggregation rates, counter to fundamental intuition. Neither the linear or logistic model selected aggregation rates from 30 °C stability studies when they were used in replacement of $k_{obs,Arr ex}$. Aggregation rates from accelerated stability studies were selected when used in replacement of $k_{obs,Arr ex}$, but with larger model error and feature coefficients that would suggest that conditions with faster aggregation rates at accelerated conditions would have slower aggregation rates at refrigerated conditions.

No model, regardless of T_{inc} , selected $\ln(IS)$, ΔB_{22}^* , or $B_{22,NE}^*$ as an important feature. The impact of Debye screening on conformational stability and self-interactions was captured by other features, and features derived from stability studies at other conditions also implicitly considered IS. ΔB_{22}^* differs from $|pH-pI|$ in its

consideration of screened electrostatics, but is also fairly correlated with $|\text{pH-pI}|$ ($R = 0.57$) and even more so when considering only low IS formulations ($R = 0.85$). As mentioned previously, the mild impact of IS on high-concentration aggregation rates might suggest that long-ranged electrostatic interactions were less important than shorter-ranged interactions between specific residues. $B_{22,NE}^*$ happened to also be fairly correlated with $|\text{pH-pI}|$ ($R=0.60$) but was quite sparse given only 4 datapoints. It is therefore unclear from this chapter whether net self-interactions due to short-ranged attractions can be a useful predictor of high-concentration aggregation rates.

As it was observed that aggregation rates were non-Arrhenius with respect to T_{inc} (see Figure 5.6), and the influence of relevant phenomena were different at different T_{inc} values (see Figure 5.19), a “unified” model with aggregation rates at all T_{inc} conditions was not considered. Similarly, models with c_2 from 10 to 130 mg/mL were not created as it was observed that the c_2 dependence of aggregation rates was not the same between different formulations.

5.3.5 Subvisible particle formation

Subvisible particle formation was also monitored with BMI for the high-concentration stability studies. Results for MAb 3 at the pH 5 low IS condition were excluded because the instrument was not available when those studies began, so only partial, inconclusive results were available. The protein mass incorporated into subvisible particles is relatively small (e.g., tenths or hundredths of a percent of the total protein mass) and is challenging to ascertain without ambiguity in how particle counts or areas are interpreted.¹³⁴ Instead, the results were quantified by binary categories based on whether the formulation was prone to the formation of subvisible particles, including a threshold based on regulatory standards, on the basis of particle

area normalized by the initial c_2 (see section 5.2.1 for further detail). The binary results are shown in Table 5.6.

Table 5.6: Results from BMI measurements of subvisible particle formation^α

MAB	Formulation	Accelerated	30 °C	Refrigerated
MAb 3	pH 5 low IS ^β	ND	ND	ND
	pH 5 high IS	+	-	-
	pH 6.5 low IS	+	+	-
	pH 6.5 high IS	+	+	-
MAb 4	pH 5 low IS	-	-	-
	pH 5 high IS	-	+	+
	pH 6.5 low IS	+	+	-
	pH 6.5 high IS	+	+	-
MAb 5	pH 5 low IS	+	-	-
	pH 5 high IS	+	+	+
	pH 6.5 low IS	+	+	+
	pH 6.5 high IS	+	-	-
MAb 6	pH 5 low IS	-	-	-
	pH 5 high IS	+	+	+
	pH 6.5 low IS	-	-	-
	pH 6.5 high IS	-	+	-

^α Formulations that were prone to subvisible particle formation are shown in purple and with + symbols, and those that were not prone to subvisible particle formation are shown in green and with - symbols.

^β Results for MAb 3 at pH 5 low IS is shown as “ND” because only partial BMI data were collected.

Out of the 15 formulations at 130 mg/mL, 10, 9, and 4 were prone to forming subvisible particles at accelerated, 30 °C, and refrigerated conditions, respectively. In

part, this was convoluted by the different aggregation rates (and resultant aggregate fractions) of these conditions as the formation of HMW aggregates is a necessary precursor to the formation of subvisible particles that are orders of magnitude larger in terms of mass.²⁵⁷ Logistic regression models per T_{inc} were developed with the features in Table 5.5, excluding features derived from other stability studies. The logistic models also considered the aggregation rate (k_{obs} value) at that formulation, a binary category of whether SEC showed the formation of HMW aggregates (see Table 5.3), and the binary subvisible particle formation results from studies at higher T_{inc} , where applicable. Predictions and feature coefficients from those logistic models are shown in Figure 5.23 and Table 5.7, respectively.

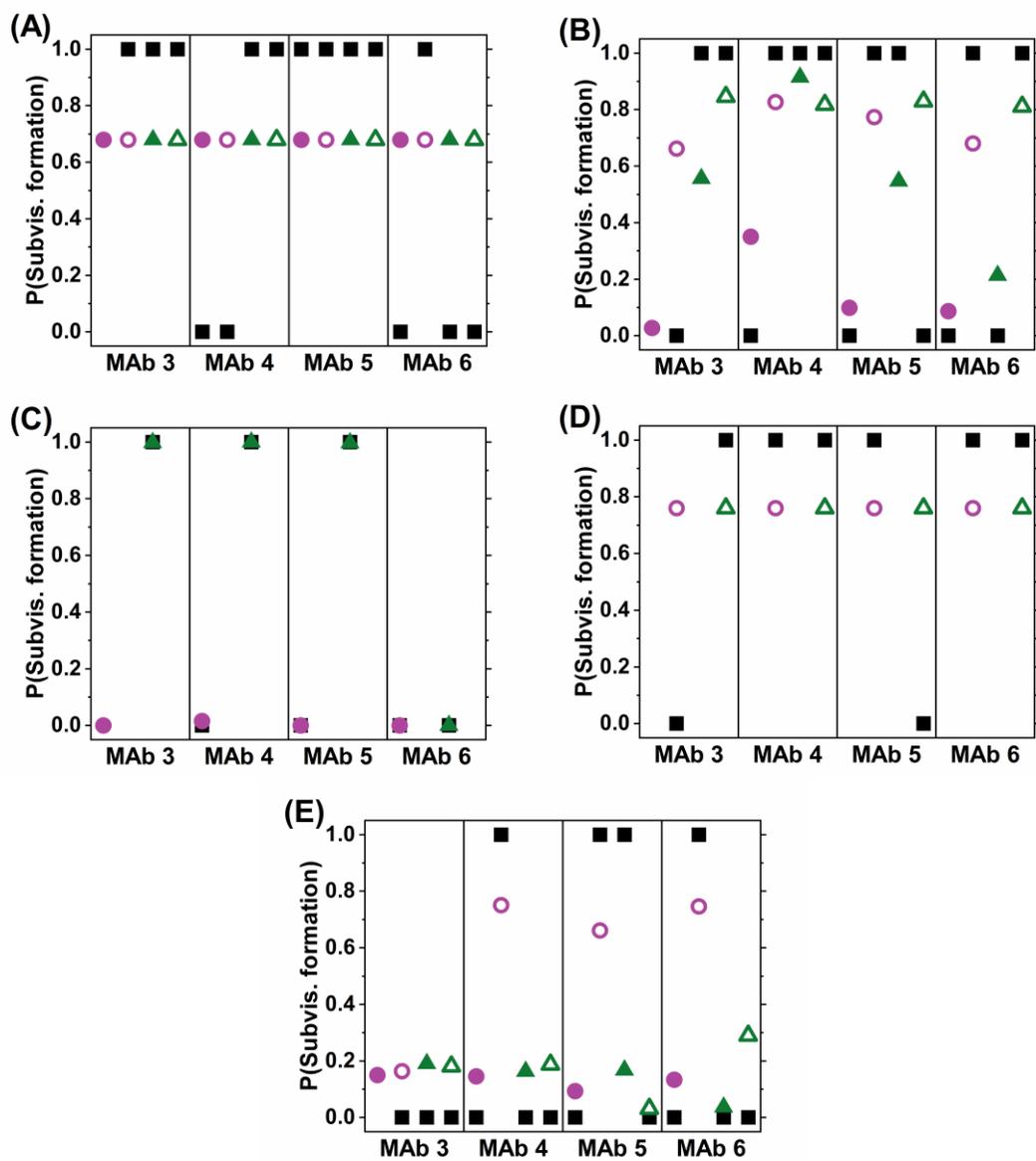


Figure 5.23: Logistic models for 130 mg/mL subvisible particle formation for the following conditions: (A) accelerated, (B) 30 °C, (C) 30 °C, low IS only, (D) 30 °C, high IS only, and (E) refrigerated. The black squares are the experimental results (see Table 5.6 above) and symbols for predicted values correspond to Table 5.1.

Table 5.7: Feature coefficients and model performance for elastic net logistic regression models of 130 mg/mL subvisible particle formation

Coefficient	Refrigerated	30 °C	30 °C (low IS)	30 °C (high IS)	Accelerated
y-intercept	0.205	0.550	0.000	0.760	0.679
pH – pI	0.569	0	0	0	0
ln(IS)	0	0	0	0	0
ΔB_{22}^*	0	-1.73	-22.9	0	0
$B_{22,NE}^*$	0	0	0	0	0
$T_{m,app1} - T_{inc}$	-0.703	0	0	0	0
k_{obs}	0	0	0	0	0
HMW agg	N/A	0	0	0	0
Subvis. (acc.)	0	0	0	0	N/A
Subvis. (30 °C)	1.04	N/A	N/A	N/A	N/A
Model performance					
R^a	0.66	0.46	1.00	N/A	N/A
Error (deviance)	3.46	3.52	0.29	2.79	6.30
Null error ^b	4.77	5.21	2.99	2.79	6.30

^a Correlation coefficient between the predicted and measured aggregation rates

^b Null error is the intercept-only error, where all coefficients besides the constant term are set to 0 and all predicted values are the same

At accelerated conditions, MAb 5 and to some extent MAb 3 were more prone to subvisible particle formation than MAb 4 and MAb 6. pH 6.5 formulations of MAb 3 also displayed complex behavior at accelerated conditions where subvisible particles that were present initially (likely created during sample preparation) dissociated and then new subvisible particles formed at later timepoints (see Figure 5.24). The logistic model did not select any of the features for accelerated conditions, so MAb identity

appeared to play the most important role. At 30 °C conditions, the only feature selected by the logistic model was ΔB_{22}^* . By splitting the dataset by IS, it was clear that this result was specific to low IS formulations, and subvisible particle formation was inhibited by conditions with strong electrostatic repulsions. The three conditions with $\Delta B_{22}^* < 0.7$ were all prone to subvisible particle formation, while the four conditions with $\Delta B_{22}^* > 1.1$ were all not prone to subvisible particle formation. At high IS, 6 out of 8 formulations were prone to subvisible particle formation, with no clear relationship to any features besides IS (i.e., the logistic model did not select any features). These results are consistent with other reports where electrostatic repulsions were found to inhibit the formation of subvisible or insoluble particles.^{57,243,275,280} Subvisible particle formation at refrigerated conditions was fairly correlated with 30 °C conditions ($R = 0.49$), and the logistic model returned that as the most important feature, followed by $T_{m,app1} - T_{inc}$ and $|\text{pH-pI}|$. With only 4 out of 15 formulations displaying subvisible particle formation, and relatively slow aggregation rates, there is significant uncertainty that precludes deeper analysis or interpretation of those results.

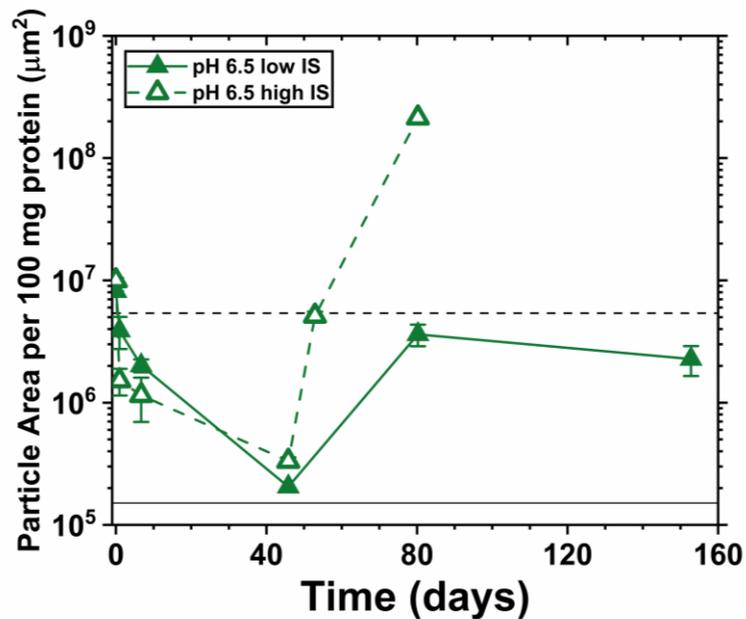


Figure 5.24: Subvisible particle areas for 130 mg/mL, pH 6.5 MAb 3 formulations at accelerated (45 °C) conditions. The low and high IS conditions are shown as filled and open green triangles, respectively (to match Table 5.1). Solid and dashed lines connecting the data points correspond to low and high IS, respectively, and are guides to the eye.

5.3.6 Outlook and industrial applications

The relative rankings of aggregation rates by MAb or by formulation were systematically inconsistent between accelerated and refrigerated conditions, which contributes to a growing body of knowledge that the industrial practice of using stability studies at elevated temperature conditions to extrapolate to refrigerated conditions can be inaccurate or misleading, at least for MAb products.^{129,145,263} The MAb aggregation rates were non-Arrhenius at 130 mg/mL in all cases, indicating that there was likely a change in the dominant aggregation mechanism as a function of T_{inc} or an intrinsically non-Arrhenius dominant aggregation mechanism.²⁶³ Despite that, aggregation rates at refrigerated conditions that were predicted via Arrhenius

extrapolation from the accelerated and 30 °C conditions were somewhat correlated with the measured aggregation rates at refrigerated conditions ($R = 0.67$; see Figure 5.6). This result is consistent with some other reports^{214,243}, but it is unlikely that Arrhenius extrapolation was truly phenomenologically relevant or quantitatively useful given that the predicted rates were several orders of magnitude smaller than the measured rates.

Conformational stability was a primary factor in mediating aggregation rates at accelerated conditions, despite performing those studies well below the $T_{m,app1}$ values ($T_{m,app1} - T_{inc} > 10$ °C) as measured by DSC. At 30 °C and refrigerated conditions, $T_{m,app1} - T_{inc}$ was not found to be related to high-concentration aggregation rates. Instead, the ML models showed that the net valence was the most informative feature for predicting aggregation rates or explaining the variance in aggregation rates between different solution conditions at refrigerated storage conditions. For refrigerated storage conditions, it was always better to choose a formulation with a pH further from the pI (pH 5 for the four MAbs in this work). It was also notable that at high-concentration, ionic strength had a limited impact on aggregation rates, which suggested that the Debye screening effect had a diminished influence when proteins were more crowded and intermolecular distances were necessarily smaller. Also, the impact of charged residues from the proteins in solution on the total solution ionic strength was not considered.

It is important to note that any statistical model is limited by its training dataset, which in this case consisted of MAbs that were not particularly problematic and likely would not be screened out during high-throughput screening activities (e.g., measurements of conformational stability and self-interactions) in early-stage

development. In part, this chapter provides guidance for the development of more robust ML models with expanded datasets that may not be publicly available but exist within companies that have performed numerous stability studies across a large number of candidates and formulations. ENR was selected over other statistical and/or ML methods that were also considered (e.g., neural networks, support vector machines, principal component analysis (PCA), and partial least squares (PLS) regression) because the ENR results were directly interpretable in terms of most important features and less likely to create overfit models given the size of this dataset. For a significantly larger dataset, those methods could provide more robust predictions with the features presented here. Preliminary analysis with PCA and PLS yielded principal components and latent variables that were not insightful or phenomenon-specific.

Many molecular modeling and simulation techniques have been developed with the goal of predicting protein aggregation, although most methods predict a proxy for quantitative aggregation rates (e.g., conformational stability, or identification of potentially aggregation prone regions) rather than being directly trained against physical stability data at intended storage conditions and representative formulations.^{100,150} Many of these models rely on data from proteins that are quite distinct from MAbs, like much smaller peptides or amyloid proteins, and to date it is not clear how translatable those models are to MAbs and other related therapeutic proteins.^{100,151} Regardless, it could be worthwhile to incorporate computational methods into the ML models developed in this work, especially because they do not require physical material and are typically higher throughput than experiments. That was out of the scope of this chapter, but has been somewhat successful in other reports

for aggregation rates of therapeutic MABs, although the datasets were typically smaller, not at high-concentration, and/or did not consider differences in solution conditions.^{40,99,150,169,233,281}

5.4 Summary and conclusions

Stability studies for four MABs (MAB 3, MAB 4, MAB 5, and MAB 6) were performed at incubation temperatures (T_{inc} values) representative of industrial studies: refrigerated (4 °C), 30 °C, and accelerated (45-50 °C) conditions. Studies were performed for formulations at every combination of two pH values (5 and 6.5), two ionic strengths (10 and ~110 mM), and three MAB concentrations (10, 35, and 130 mg/mL). Aggregation rates were calculated by analyzing the stability samples with size-exclusion chromatography, and subvisible particles were monitored with backgrounded membrane imaging. The results were analyzed from the perspective of how the formulation conditions impacted fundamental phenomena such as self-interactions (quantified by static and dynamic light scattering experiments) and conformational stability (quantified by differential scanning calorimetry experiments). Interpretable machine learning models (i.e., elastic net regression models) were developed to rigorously select the features that were relevant to the measured aggregation rates, deconvolute the impacts of correlated features, and make predictions of high-concentration aggregation rates.

A key goal of this chapter was to provide insights that would help in predicting which proteins or solution conditions would be more stable as a commercial MAB product for subcutaneous administration. Those products are formulated at high protein concentrations (represented by 130 mg/mL in this work) and often stored at refrigerated conditions. Stability studies at higher T_{inc} conditions (30 °C to 50 °C)

were misleading with respect to which MAb or solution condition (i.e., pH and ionic strength values) was more or less stable, and aggregation rates were unsurprisingly non-Arrhenius. Aggregation rates at the accelerated condition (45-50 °C) were clearly mediated by conformational stability, as the difference between the first midpoint unfolding temperature from DSC ($T_{m,app1}$) and the T_{inc} (i.e., $T_{m,app1} - T_{inc}$) was highly correlated with the aggregation rate for a given MAb and MAb concentration. The net valence of the MAb at a given pH, represented by the difference between the pH and pI (i.e., |pH-pI|), was reasonably predictive of high-concentration aggregation rates at 30 °C and refrigerated conditions. That was theorized to be due to the effect of repulsive electrostatic self-interactions, but the ML models found that net valence (quantified by |pH-pI|) was a more significant feature than the net strength of electrostatic repulsions that were quantified by SLS and DLS. The ML models also suggested that long-ranged electrostatic repulsions were relevant in inhibiting subvisible particle formation at 30 °C and perhaps refrigerated conditions.

The dataset of aggregation rates in this chapter is unique in its breadth compared to other published studies as it includes multiple MAbs, multiple solution conditions, high MAb concentrations and realistic refrigerated storage conditions. Measurements of conformational stability and net self-interactions for each MAb were consistent with relatively well-behaved MAbs that might not be screened out during early-stage development, thus the results in this work should be interpreted with respect to optimizing solution conditions for MAbs in later stages of development. Also, pharmaceutical companies that develop therapeutic proteins likely have even larger datasets that are not publicly available. The ML methods presented in this work were developed with that context in mind and included elements to more rigorously

estimate uncertainty, identify relevant features, and be specifically suited for predicting aggregation rates for a protein outside of the training data. The ML models might be improved by the incorporation of molecular modeling and molecular simulations, or other experimental methods such as intrinsic fluorescence and circular dichroism.

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

This thesis was focused on addressing challenges in industrial drug development of therapeutic proteins through the lens of characterizing and predicting self-interactions and improving understanding of how self-interactions are related to reversible self-association and aggregation. Emphasis was placed on specific practical applications such as predicting high-concentration net self-interactions and predicting how different solution conditions will impact MAb aggregation rates. Experimental datasets were generated for several proteins (four MAbs, two Fc-fusion proteins, and the corresponding FP protein) at varying solution conditions that systematically parsed fundamental features of commercial protein formulations (i.e., pH, ionic strength, and protein concentration). The proteins were provided by industrial collaborators and were directly representative of commercial therapeutic protein products. Data from prior studies of five other MAbs were also used in assessing and validating the molecular simulation methods. Overall, the scope of the studies in this thesis were quite broad compared to other self-consistent and publicly-available datasets, and therefore the results were more certain in a statistical sense and more robust with respect to being potentially applicable to other proteins (especially other MAbs).

A collection of CG models for molecular simulations were curated in Chapter 2 that ranged from domain-level models (e.g., DODECA), to relatively high-resolution models (e.g., 1bAA). Guidance for application-specific model selection was provided

based on detailed analyses of the tradeoffs between computational burden and structural resolution and/or predictive accuracy. A dataset from prior work of net self-interactions (i.e., B_{22} values) for five MAbs at many different solution conditions (with respect to pH and ionic strength) were used as an illustrative example for quantitative comparisons. The dataset of net self-interactions ranged from “simple-colloidal” electrostatic repulsions to strong net attractions driven by electrostatic interactions between particular amino acids. Domain-level models can be applied in a high-throughput manner and are tractable at high-concentrations but lack the structural resolution to model interactions between specific amino acids such those that can facilitate strong electrostatic attractions. Higher-resolution models sacrificed computational efficiency by orders of magnitude but were significantly more accurate in their structural resolutions and representation of attractive electrostatic self-interactions. The improved resolution of steric and electrostatic behavior for high-resolution models would be appealing for use in high-concentration simulations, but models like 1bAA were not tractable at those concentrations to run at a practical throughput. The 1bC/D model was introduced which retained amino acid resolution of the charged sites in the protein, while greatly reducing the total number of interaction sites or beads in the model by “lumping” the noncharged amino acids into domain beads akin to the DODECA domain-level model. The 1bC/D model was able to capture the range of net self-interactions with equal or improved accuracy compared to the 1bAA model, with computational burdens that were 25-50 times lower. Some of these models were used in the two subsequent chapters: 1bAA in Chapter 3 for low-concentration simulations of two Fc-fusion proteins and their FP protein, and both

DODECA and 1bC/D in Chapter 4 for low- to high-concentration simulations for six MAbs.

SLS and DLS experiments and 1bAA molecular simulations were applied to investigate and model attractive electrostatic self-interactions and their potential relationship with reversible self-association for two Fc-fusion proteins (monovalent and bivalent) and the corresponding FP protein in Chapter 3. At pH 6.5 conditions, the three proteins formed oligomers at low ionic strength conditions that fully or partially reversed when the ionic strength in the solution was increased. For the monovalent Fc-fusion and FP protein, that reversible self-association appeared to be related to attractive electrostatic self-interactions. For the bivalent Fc-fusion, a “baseline” of higher molecular weight species at all conditions complicated experimental quantification of net self-interactions, but 1bAA simulations were used to qualitatively predict that at pH 6.5, electrostatic interactions would also be net attractive. Analysis of specific pairwise electrostatic interactions between charged beads in the 1bAA simulations suggested that cross-domain electrostatic interactions between the Fc and FP domains were responsible for attractive electrostatic self-interactions and thus possibly also reversible self-association. Fc-fusion proteins can be particularly challenging to develop as they can be prone to many challenging biophysical behaviors which can be difficult to predict as well as time- and material-intensive to screen for experimentally. Computational methods such as those presented in Chapter 3 showed substantial potential for screening Fc-fusion drug candidates and designing charge variants that can mitigate poor behaviors like reversible self-association.

Typically, the preferred administration route for commercial protein therapeutics is via subcutaneous injection, which requires relatively small volumes (<

~2 mL). To achieve the needed dosage in a single injection, liquid solutions with relatively high protein concentrations (i.e., on the order of 100 mg/mL) are needed. Chapters 4 and 5 focused on problematic behaviors at high-concentrations, how they are related to low-concentration behaviors, efforts to better understand the underlying phenomena, and predictions of net self-interactions and aggregation rates at high-concentration conditions. In Chapter 4, experimental measurements of net self-interactions (primarily via SLS) from low- to high-concentration for six MAbs as a function of pH and ionic strength were used as a test case where similar to Chapter 2, the net self-interactions spanned from strong repulsions to strong attractions. The variance in net self-interactions was primarily electrostatically driven via differences in the surface distribution of charged residues due to the unique amino acid sequences of the six MAbs and changes in pH and ionic strength. In Chapter 5, that dataset was expanded for four of the MAbs to include measurements of conformational stability via DSC and short- to long-term stability studies to quantify aggregation rates as a function of pH, ionic strength, protein concentration, and incubation temperature. The dataset of aggregation rates was unique in its size and scope compared to other published works.

A previously developed method for combining low-concentration experimental SLS measurements and low- to high-concentration CG molecular simulations to predict high-concentration net self-interactions was improved in Chapter 4 by the implementation of the hybrid 1bC/D model that was introduced in Chapter 2. The high-concentration predictions were compared between the 1bC/D model and the DODECA model, and the results showed that while the DODECA model consistently gave poor predictions for systems with net attractive electrostatic self-interactions, the

1bC/D model was able to properly capture net attractive electrostatic interactions. However, the 1bC/D model sometimes struggled to quantitatively capture net repulsive electrostatic interactions at pH 6.5 conditions. Those deficiencies were addressed by considering two alternative methods of representing of charge states of ionizable residues in the simulations. Like Chapter 3, specific electrostatic interactions between charged residues were analyzed and problematic amino acids were identified for a MAb with strongly attractive electrostatic self-interactions and a propensity for phase separation.

The stability study conditions in Chapter 5 were intended to mimic commercial MAb products (i.e., in terms of protein concentration and formulation components) and stability studies that are commonly performed in an industrial setting (i.e., in terms of incubation temperatures). The MAbs in this chapter were relatively “well-behaved” by industrial standards, at least with respect to the experimental measurements of net self-interactions and conformational stability. Thus, the results are perhaps best interpreted in the context of formulation development where a given MAb candidate was not disqualified by early-stage high-throughput screening measurements. Short- to medium-term stability studies are performed at “accelerated” temperatures to screen drug candidates and formulations and were represented by studies at incubation temperatures of 30-50 °C. Long-term stability studies are used to demonstrate adequate shelf life at the intended storage condition and were represented by studies at refrigerated conditions (4 °C). The most important goal of these studies was to relate measurements of net self-interactions, conformational stability, and aggregation rates at accelerated conditions to high-concentration (i.e., 130 mg/ml) aggregation rates at refrigerated storage conditions. The results from those studies

were first analyzed in terms of straightforward statistical relationships between the high-concentration aggregation rates and measured properties such as $T_{m,app1}$ from DSC and B_{22} from SLS. Conformational stability mediated aggregation rates at the highest incubation temperatures (i.e., 45 °C and 50 °C) for a given MAb but was not very predictive of aggregation rates when data for all four MAbs were combined, indicating that there were other MAb-specific features that impacted the aggregation rates. At 30 °C and refrigerated conditions, conformational stability was not particularly correlated with aggregation rates. However, there was a clear trend with the net valence at a particular pH, which was proposed to be due to repulsive electrostatic self-interactions that inhibited aggregation. It was notable that at high-concentration, ionic strength had a limited impact on aggregation rates, which suggested that the Debye screening effect had a diminished influence when proteins were more crowded and intermolecular distances were necessarily smaller.

The relative rankings of aggregation rates by MAb or by formulation were systematically inconsistent between accelerated and refrigerated conditions, which contributes to a growing body of knowledge that the industrial practice of using stability studies at elevated temperature conditions to extrapolate to refrigerated conditions can be inaccurate or misleading.^{129,145,263} Interpretable ML methods were developed in part to provide a platform for more robust predictions of aggregation rates that focused on deconvoluting related features and considered uncertainty in the measured and predicted aggregation rates. The models were also used to evaluate when studies at more accessible conditions (i.e., lower concentration or higher incubation temperature) could be useful to predict the high-concentration aggregation rates. The models were useful in hypothesis testing (e.g., is conformational stability

related to aggregation rates at refrigerated conditions?), and clearly suggested that net valence was the most significant predictor of aggregation rates at realistic storage conditions. Although the dataset of aggregation rates in Chapter 5 was large compared to other similar published datasets, it is certainly much smaller than the data that has been collected at pharmaceutical companies over the last few decades, so the ML methods could be of particular utility when trained on those much larger datasets. In the context of the dataset in Chapter 5, the ML models were most useful to parse the contributions of fundamental phenomena on aggregation rates with higher statistical certainty.

6.2 Future work

6.2.1 Characterization of oligomers and aggregates

The methods used in this thesis to characterize reversible oligomers and aggregates were constrained by several practical considerations such as instrument availability and whether the technique might perturb the aggregates or reversible oligomers. Also, relatively high throughput methods were needed for characterization of aggregates formed during stability studies to maximize the number of conditions that could be tested. More detailed insights into structure and size were compromised in favor of more precise quantification in terms of aggregate mass concentration that could be performed at higher throughput. The experimental methods described below could be used as orthogonal measures or to investigate other relevant phenomena.

It was key that the reversible self-association be characterized *in situ* as it is inherently transient and was highly dependent on the solution conditions. SLS and DLS were used in Chapter 3, which satisfy that requirement but are at best semi-

quantitative techniques for characterizing reversible oligomers. SLS was used to calculate the apparent molecular weight of the solution ($M_{w,app}$), which is a single value that quantifies the weight average of all scattering species in solution which could arise from many different molecular weight distributions. DLS correlograms were a convolution of autocorrelation functions from a likely broad and “continuous” distribution of oligomers (i.e., monomers, dimers, trimers, tetramers, etc.). The correlograms qualitatively described the size distribution of oligomers and were useful for qualitative comparisons between proteins and formulations. However, the exact size distribution of oligomers was difficult or impossible to reliably deconvolute in this context, even with advanced analysis techniques like the regularization method. Nondestructive techniques that can separate species by size such as sedimentation velocity AUC or field flow fractionation (FFF) could provide more quantitative results if the timescale of reversible oligomerization is slow enough to isolate the oligomers. Preliminary AUC measurements had a similar limitation to DLS where the exact size of oligomers was challenging to ascertain.

It would be of interest to understand the mechanism of the reversible self-association in more detail, especially which region(s) are active in reversible self-association and how strong the association is. Binding affinity measurements via isothermal titration calorimetry and surface plasmon resonance were applied by Mieczkowski et al. to quantify the strength of reversible self-association for a MAb.²²⁴ Protein-protein docking simulations and homology modeling were also used in that work to identify sites where point mutations could reduce self-association affinity.²²⁴ The molecular simulations in Chapter 3 suggested that reversible self-association proceeded through cross-domain FP-Fc interactions, which could be corroborated by

SLS, DLS, or binding affinity measurements using the “unfused” proteins (e.g., measurements where Fc and FP concentrations are varied and compared to measurements of the Fc and FP on their own). Another promising methodology was described in Dobson et al. where the residues relevant to self-association of a MAb were mapped via hydrogen/deuterium exchange and cross-linking mass spectroscopy and that information was used to rationally design point mutations.¹¹¹

The MAb aggregates that were generated from stability studies had many of the same characterization limitations. SEC could clearly separate LMW fragments and dimers, but higher molecular weight aggregates (i.e., trimers, tetramers, and so on) had overlapping leaks and their concentrations could not be deconvoluted. The stability samples were also diluted in mobile phase (which was significantly different in composition compared to the MAb formulations) when injected into the HPLC and could have interacted with the SEC column, both of which could have perturbed the aggregates. Many of the stability samples were also measured with SLS and DLS, but those measurements had the same limitations as outlined above and in general were uncertain due to a number of factors including: 1) relatively low concentrations of aggregates that did not contribute significantly to the scattering intensity, 2) the presence of large aggregates that would intermittently interact with the laser and drastically increase the scattering intensity, 3) insoluble particles that could not be analyzed with SLS or DLS, and 4) the presence of fragments which would decrease the scattering signal. Analyzing the eluate from the outlet of the SEC column with an in-line MALS detector (termed SEC-MALS) is commonly used and might have been beneficial to better understand the size distribution of higher molecular weight aggregates that were not disrupted by SEC. Some aggregates also undergo significant

conformational changes, which could be probed by many spectroscopic techniques.²⁸² Preliminary measurements of higher order structure via intrinsic fluorescence and circular dichroism did not reveal significant changes in secondary or tertiary structure for a given MAb as a function of pH, ionic strength, or extent of aggregation. Sedimentation velocity AUC and FFF might also be more informative in deconvoluting the size distribution of aggregates larger than dimers and would avoid any column or mobile phase interactions.

6.2.2 Advanced algorithms for high-concentration molecular simulations

The high-concentration molecular simulations in Chapter 4 were made tractable for the 1bC/D model by using straightforward GCMC, rather than GCMC with the TMMC algorithm. TMMC provides a continuous prediction of net self-interactions as a function of concentration by applying histogram reweighting techniques and the biased sampling can be useful to push the system to different configurations thus increasing acceptance rates. Sampling with GCMC was sometimes inefficient because acceptance rates were quite low in particularly attractive systems (i.e., the simulation would get stuck in potential energy “wells”). Higher concentration simulations were also inherently prone to low acceptance rates due to the tight intermolecular distances between molecules. The computational efficiency of the simulations could be improved if optimized sampling methods were applied such as configurational-bias, domain decomposition, or aggregation-volume bias.^{283–285} Configurational bias is particularly promising as it has been successfully implemented previously in similar simulations for proteins.^{207,285,286}

The high-concentration simulations in Chapter 4 that could accurately reproduce net self-interactions for a broad range of conditions contained a

considerable amount of potentially useful information about high-concentration MAb solutions that were not analyzed in detail in this thesis. The spatial arrangements and specific interactions between amino acids (particularly the charged sites) could be investigated to further understand the solution microstructure. Preliminary work to begin to collect and analyze this information suggested that unique spatial correlations between specific pairs of MAb domains were present and dependent on the MAb and formulation (i.e., pH, ionic strength, and MAb concentration). Examples of this finding in the form of domain-domain radial distribution functions (also known as $g(r)$ functions) are shown in Figure 6.1 for 130 mg/mL simulations of MAb 4 with the 1bC/D model (static charge states with nominal pK_a values) at each of the formulations that were studied in Chapters 4 and 5 (see Table 4.1 and Table 5.1). The radial distribution functions for each pair of domains (e.g., V_H and C_{H2}) are overlaid and Figure 6.1 shows that not only are specific pairs of domains more likely to be close to one another, but that is mediated by electrostatic interactions as the results changed as a function of pH and ionic strength.

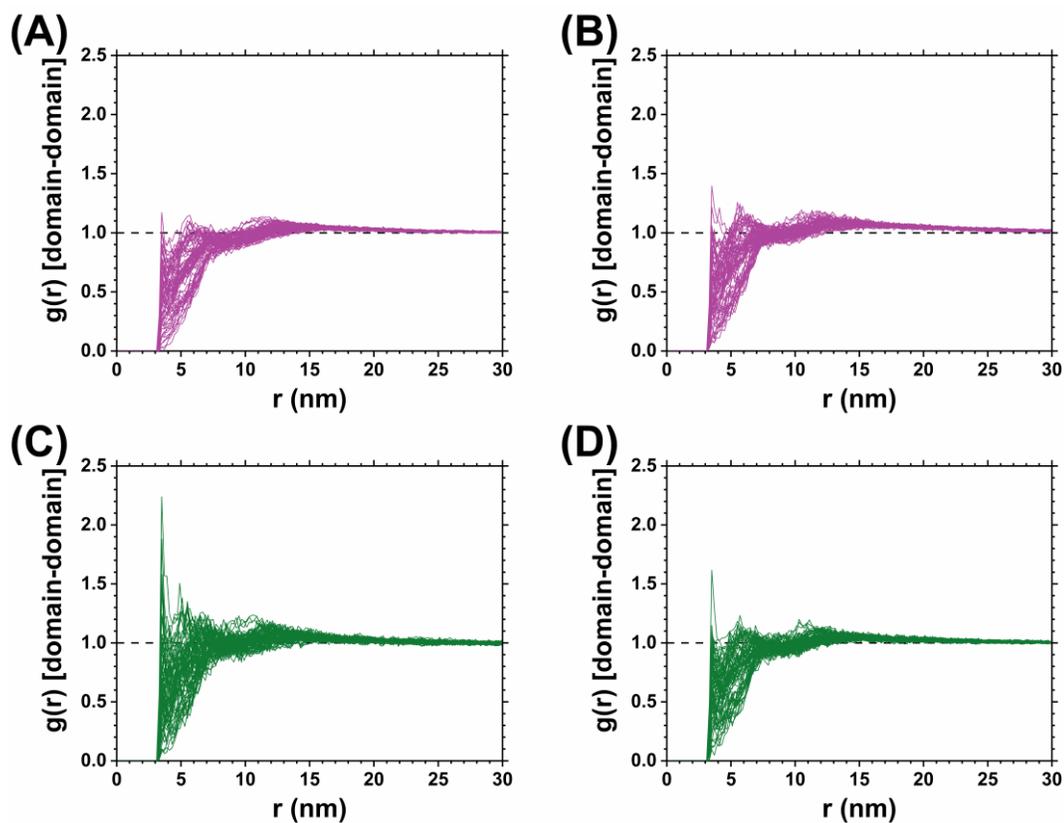


Figure 6.1: Domain-domain radial distribution functions between 1bC/D domain beads in 130 mg/mL simulations of Mab 4 at the following formulations: (A) pH 5 low ionic strength, (B) pH 5 high ionic strength, (C) pH 6.5 low ionic strength, and (D) pH 6.5 high ionic strength. pH 5 results are shown in purple and pH 6.5 results are shown in green to match the color scheme in Chapter 5.

Other reports have used CG models similar to the DODECA model to predict and model high-concentration solution viscosities for therapeutic proteins (particularly MAbs), often when coupled with SAXS experiments.^{34,287,288} The simulations in this thesis are also well-suited for that application especially with respect to the electrostatic contributions to viscosity. The formation of “clusters”, or persistent groups of proteins that are spatially proximate but have not formed oligomers, has

been putatively related to poor biophysical behaviors (especially elevated viscosity) and could be explored further by applying one of the many previously developed algorithms for tracking clustering with the high-concentration molecular simulations from Chapter 4.^{71,77,155,194} Preliminary work in this area yielded quantities that were all highly correlated with net self-interactions, which is perhaps unsurprising given the many previously reported correlations between self-interactions and elevated viscosity.^{54,121,289} The high-concentration MAb solutions in this thesis did not have particularly elevated solution viscosities, except for MAb 2 at the pH 6.5 low ionic strength formulation at concentrations that were also prone to phase separation. A dataset with other proteins would be necessary to better assess whether the simulations could be more insightful in predicting viscosity than e.g., k_D values from DLS.

Because the high-concentration simulations are MC and not MD, they could also be used to probe phenomena that occur on longer timescales than what has been typically accessible with MD simulations, such as specific steps in the mechanisms of reversible self-association and/or aggregation (e.g., reversible dimerization). Results from those simulations could be incorporated into phenomenological models of aggregation to enable more robust predictions of aggregation rates. An ambitious target would be to combine the phenomenological models with information from the simulations to identify what step was rate limiting and the regions in the protein that were responsible. The reactive MC algorithm is a promising method for probing associations between proteins as it specifically samples the phase space of a given reaction.^{290,291} Reactive MC and MD simulations have been applied to proteins in other areas such as bond scission reactions, calculating pK_a values of ionizable residues, and conformational stability.^{66,292,293}

6.2.3 Experimental and simulation methods to screen excipients

Another improvement to the simulations used in this thesis could be to explicitly represent cosolutes. Currently, cosolutes play at most an implicit role in the simulations if they are charged and thus impact the Debye screening length, which does not at all account for the identity of the cosolute species except for its net valence. Although uncharged cosolutes can impact the solution permittivity which will change the Bjerrum length variable, that will only serve to mildly shift the optimal value of the ψ parameter as the two are multiplied together in the electrostatic interaction potential (see Equation 2.4). If explicitly represented in the simulations, preferential interactions could be modeled for a broad range of excipients commonly used in commercial formulations (e.g., salts, sugars, surfactants, buffers, and amino acids) or less established “next-generation” excipients such as ionic liquids and glycopolymers.²⁹⁴ For more structurally complex molecules such as surfactants, including flexibility in both those molecules and/or the proteins in the simulation might be necessary. Explicit representation of preferential accumulation of ions on the protein surface (e.g., that form a Stern layer) is of particular interest and relevance as it mediates electrostatic interactions and can play an important role in packing and diffusion.²²² The simulations could be validated with experimental data such measurements of net self-interactions or partial molar volumes as a function of cosolute concentration and/or protein concentration. In the context of the results in Chapters 3 and 5, these methods could be used to screen cosolutes that could impact reversible self-association and/or aggregation. There are some reports similar to what is suggested here in terms of explicit representation of cosolutes in simulations, but direct practical applications are still limited and the simulation methods used in this

thesis could be well-suited to address this topic, especially at high protein concentrations.^{81,204,206,247,295–297}

REFERENCES

- (1) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J. Health Econ.* **2016**, *47*, 20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>.
- (2) Brown, A. Top Product Forecasts for 2023. *Nat. Rev. Drug Discov.* **2023**, *22* (1), 8. <https://doi.org/10.1038/d41573-022-00193-0>.
- (3) Strickley, R. G.; Lambert, W. J. A Review of Formulations of Commercially Available Antibodies. *J. Pharm. Sci.* **2021**, *110* (7), 2590-2608.e56. <https://doi.org/10.1016/j.xphs.2021.03.017>.
- (4) Lu, R. M.; Hwang, Y. C.; Liu, I. J.; Lee, C. C.; Tsai, H. Z.; Li, H. J.; Wu, H. C. Development of Therapeutic Antibodies for the Treatment of Diseases. *J. Biomed. Sci.* **2020**, *27* (1), 1–30. <https://doi.org/10.1186/s12929-019-0592-z>.
- (5) Pecetta, S.; Finco, O.; Seubert, A. Quantum Leap of Monoclonal Antibody (MAb) Discovery and Development in the COVID-19 Era. *Semin. Immunol.* **2020**, *50*, 101427. <https://doi.org/10.1016/j.smim.2020.101427>.
- (6) Cui, Y.; Cui, P.; Chen, B.; Li, S.; Guan, H. Monoclonal Antibodies: Formulations of Marketed Products and Recent Advances in Novel Delivery System. *Drug Dev. Ind. Pharm.* **2017**, *43* (4), 519–530. <https://doi.org/10.1080/03639045.2017.1278768>.
- (7) Grilo, A. L.; Mantalaris, A. The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.* **2019**, *37* (1), 9–16. <https://doi.org/10.1016/j.tibtech.2018.05.014>.
- (8) Wang, S. S.; Yan, Y.; Ho, K. US FDA-Approved Therapeutic Antibodies with High-Concentration Formulation: Summaries and Perspectives. *Antib. Ther.* **2021**, *4* (4), 262–273. <https://doi.org/10.1093/abt/tbab027>.
- (9) *The Antibody Society. Therapeutic monoclonal antibodies approved or in regulatory review.* www.antibodysociety.org/antibody-therapeutics-product-data (accessed 2023-09-07).
- (10) Fogel, D. B. Factors Associated with Clinical Trials That Fail and Opportunities for Improving the Likelihood of Success: A Review. *Contemp. Clin. Trials Commun.* **2018**, *11* (August), 156–164. <https://doi.org/10.1016/j.conctc.2018.08.001>.
- (11) Bailly, M.; Mieczkowski, C.; Juan, V.; Metwally, E.; Tomazela, D.; Baker, J.; Uchida, M.; Kofman, E.; Raoufi, F.; Motlagh, S.; Yu, Y.; Park, J.; Raghava, S.; Welsh, J.; Rauscher, M.; Raghunathan, G.; Hsieh, M.; Chen, Y. L.; Nguyen, H. T.; Nguyen, N.; Cipriano, D.; Fayadat-Dilman, L. Predicting Antibody Developability Profiles Through Early Stage Discovery Screening. *MAbs* **2020**,

- 12 (1), 1–28. <https://doi.org/10.1080/19420862.2020.1743053>.
- (12) Jain, T.; Sun, T.; Durand, S.; Hall, A.; Houston, N. R.; Nett, J. H.; Sharkey, B.; Bobrowicz, B.; Caffry, I.; Yu, Y.; Cao, Y.; Lynaugh, H.; Brown, M.; Baruah, H.; Gray, L. T.; Krauland, E. M.; Xu, Y.; Vásquez, M.; Wittrup, K. D. Biophysical Properties of the Clinical-Stage Antibody Landscape. *Proc. Natl. Acad. Sci.* **2017**, *114* (5), 944–949. <https://doi.org/10.1073/PNAS.1616408114>.
- (13) Wei, J. Y.; Bou-Assaf, G. M.; Houde, D.; Weiskopf, A. Technical Decision-Making with Higher Order Structure Data: Detecting Reversible Concentration-Dependent Self-Association in a Monoclonal Antibody and a Preliminary Investigation to Eliminate It. *J. Pharm. Sci.* **2015**, *104* (11), 3984–3989. <https://doi.org/10.1002/jps.24616>.
- (14) Shukla, A. A.; Wolfe, L. S.; Mostafa, S. S.; Norman, C. Evolving Trends in MAb Production Processes. *Bioeng. Transl. Med.* **2017**, *2* (1), 58–69. <https://doi.org/10.1002/btm2.10061>.
- (15) Ghosh, I.; Gutka, H.; Krause, M. E.; Clemens, R.; Kashi, R. S. A Systematic Review of Commercial High Concentration Antibody Drug Products Approved in the US: Formulation Composition, Dosage Form Design and Primary Packaging Considerations. *MAbs* **2023**, *15* (1), 1–20. <https://doi.org/10.1080/19420862.2023.2205540>.
- (16) Starr, C. G.; Tessier, P. M. Selecting and Engineering Monoclonal Antibodies with Drug-like Specificity. *Curr. Opin. Biotechnol.* **2019**, *60*, 119–127. <https://doi.org/10.1016/j.copbio.2019.01.008>.
- (17) Fernández-Quintero, M. L.; Ljungars, A.; Waibl, F.; Greiff, V.; Andersen, J. T.; Gjølberg, T. T.; Jenkins, T. P.; Voldborg, B. G.; Grav, L. M.; Kumar, S.; Georges, G.; Kettenberger, H.; Liedl, K. R.; Tessier, P. M.; McCafferty, J.; Laustsen, A. H. Assessing Developability Early in the Discovery Process for Novel Biologics. *MAbs* **2023**, *15* (1). <https://doi.org/10.1080/19420862.2023.2171248>.
- (18) Jain, T.; Sun, T.; Durand, S.; Hall, A.; Houston, N. R.; Nett, J. H.; Sharkey, B.; Bobrowicz, B.; Caffry, I.; Yu, Y.; Cao, Y.; Lynaugh, H.; Brown, M.; Baruah, H.; Gray, L. T.; Krauland, E. M.; Xu, Y.; Vásquez, M.; Wittrup, K. D. Biophysical Properties of the Clinical-Stage Antibody Landscape. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (5), 944–949. <https://doi.org/10.1073/pnas.1616408114>.
- (19) Bittner, B.; Richter, W.; Schmidt, J. Subcutaneous Administration of Biotherapeutics: An Overview of Current Challenges and Opportunities. *BioDrugs* **2018**, *32* (5), 425–440. <https://doi.org/10.1007/s40259-018-0295-0>.
- (20) Badkar, A.; Wolf, A.; Bohack, L.; Kolhe, P. Development of Biotechnology Products in Pre-Filled Syringes: Technical Considerations and Approaches. *AAPS PharmSciTech* **2011**, *12* (2), 564–572. <https://doi.org/10.1208/s12249-011-9617-y>.
- (21) Blanco, M. A. Computational Models for Studying Physical Instabilities in High Concentration Biotherapeutic Formulations. *MAbs* **2022**, *14* (1).

- <https://doi.org/10.1080/19420862.2022.2044744>.
- (22) Barros, M.; Zhang, X.; Kenrick, S.; Valente, J. J. Opalescence Measurements: Improvements in Fundamental Knowledge, Identifying Sources of Analytical Biases, and Advanced Applications for the Development of Therapeutic Proteins. *J. Pharm. Sci.* **2021**, *110* (11), 3550–3557. <https://doi.org/10.1016/j.xphs.2021.06.013>.
- (23) Jarasch, A.; Koll, H.; Regula, J. T.; Bader, M.; Papadimitriou, A.; Kettenberger, H. Developability Assessment During the Selection of Novel Therapeutic Antibodies. *J. Pharm. Sci.* **2015**, *104* (6), 1885–1898. <https://doi.org/10.1002/jps.24430>.
- (24) Wang, W.; Roberts, C. J. Protein Aggregation – Mechanisms, Detection, and Control. *Int. J. Pharm.* **2018**, *550* (1–2), 251–268. <https://doi.org/10.1016/j.ijpharm.2018.08.043>.
- (25) Krause, M. E.; Sahin, E. Chemical and Physical Instabilities in Manufacturing and Storage of Therapeutic Proteins. *Curr. Opin. Biotechnol.* **2019**, *60*, 159–167. <https://doi.org/10.1016/j.copbio.2019.01.014>.
- (26) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic Cluster Formation Determines Viscosity and Diffusion in Dense Protein Solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (20), 9843–9852. <https://doi.org/10.1073/pnas.1817564116>.
- (27) Blanco, M. A.; Sahin, E.; Robinson, A. S.; Roberts, C. J. Coarse-Grained Model for Colloidal Protein Interactions, B22, and Protein Cluster Formation. *J. Phys. Chem. B* **2013**, *117* (50), 16013–16028. <https://doi.org/10.1021/jp409300j>.
- (28) Calero-Rubio, C.; Saluja, A.; Roberts, C. J. Coarse-Grained Antibody Models for “Weak” Protein-Protein Interactions from Low to High Concentrations. *J. Phys. Chem. B* **2016**, *120* (27), 6592–6605. <https://doi.org/10.1021/acs.jpcc.6b04907>.
- (29) Israelachvili, J. N. Van Der Waals Forces between Particles and Surfaces. In *Intermolecular and Surface Forces*; Academic Press, 2011; pp 253–289. <https://doi.org/10.1016/b978-0-12-375182-9.10013-2>.
- (30) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. Molecular Origins of Osmotic Second Virial Coefficients of Proteins. *Biophys. J.* **1998**, *75* (5), 2469–2477. [https://doi.org/https://doi.org/10.1016/S0006-3495\(98\)77691-X](https://doi.org/https://doi.org/10.1016/S0006-3495(98)77691-X).
- (31) Khaniya, U.; Mao, J.; Wei, R. J.; Gunner, M. R. Characterizing Protein Protonation Microstates Using Monte Carlo Sampling. *J. Phys. Chem. B* **2022**, *126* (13), 2476–2485. <https://doi.org/10.1021/acs.jpcc.2c00139>.
- (32) Bashford, D.; Karplus, M. PKa’s of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry* **1990**, *29* (44), 10219–10225. <https://doi.org/10.1021/bi00496a010>.
- (33) Wang, W.; Roberts, C. J. *Aggregation of Therapeutic Proteins*, 1st ed.; John Wiley & Sons: Hoboken, NJ, 2010. <https://doi.org/10.1002/9780470769829.ch1>.

- (34) Chowdhury, A. A.; Manohar, N.; Witek, M. A.; Woldeyes, M. A.; Majumdar, R.; Qian, K. K.; Kimball, W. D.; Xu, S.; Lanzaro, A.; Truskett, T. M.; Johnston, K. P. Subclass Effects on Self-Association and Viscosity of Monoclonal Antibodies at High Concentrations. *Mol. Pharm.* **2023**, *20* (6), 2991–3008. <https://doi.org/10.1021/acs.molpharmaceut.3c00023>.
- (35) Blanco, M. A.; Sahin, E.; Li, Y.; Roberts, C. J. Reexamining Protein-Protein and Protein-Solvent Interactions from Kirkwood-Buff Analysis of Light Scattering in Multi-Component Solutions. *J. Chem. Phys.* **2011**, *134* (22), 1–12. <https://doi.org/10.1063/1.3596726>.
- (36) Minton, A. P. Recent Applications of Light Scattering Measurement in the Biological and Biopharmaceutical Sciences. *Anal. Biochem.* **2016**, *501* (February 2016), 4–22. <https://doi.org/10.1016/j.ab.2016.02.007>.
- (37) Cole, J. L.; Lary, J. W.; P. Moody, T.; Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **2008**, *84* (07), 143–179. [https://doi.org/10.1016/S0091-679X\(07\)84006-4](https://doi.org/10.1016/S0091-679X(07)84006-4).
- (38) Yearley, E. J.; Zarraga, I. E.; Shire, S. J.; Scherer, T. M.; Gokarn, Y.; Wagner, N. J.; Liu, Y. Small-Angle Neutron Scattering Characterization of Monoclonal Antibody Conformations and Interactions at High Concentrations. *Biophys. J.* **2013**, *105* (3), 720–731. <https://doi.org/10.1016/j.bpj.2013.06.043>.
- (39) Tessier, P. M.; Lenhoff, A. M.; Sandler, S. I. Rapid Measurement of Protein Osmotic Second Virial Coefficients by Self-Interaction Chromatography. *Biophys. J.* **2002**, *82* (3), 1620–1631. [https://doi.org/10.1016/S0006-3495\(02\)75513-6](https://doi.org/10.1016/S0006-3495(02)75513-6).
- (40) Lai, P. K.; Fernando, A.; Cloutier, T. K.; Kingsbury, J. S.; Gokarn, Y.; Halloran, K. T.; Calero-Rubio, C.; Trout, B. L. Machine Learning Feature Selection for Predicting High Concentration Therapeutic Antibody Aggregation. *J. Pharm. Sci.* **2021**, *110* (4), 1583–1591. <https://doi.org/10.1016/j.xphs.2020.12.014>.
- (41) Holloway, L.; Roche, A.; Marzouk, S.; Uddin, S.; Ke, P.; Ekizoglou, S.; Curtis, R. Determination of Protein-Protein Interactions at High Co-Solvent Concentrations Using Static and Dynamic Light Scattering. *J. Pharm. Sci.* **2020**, *109* (9), 2699–2709. <https://doi.org/10.1016/j.xphs.2020.05.023>.
- (42) Ferreira, G. M.; Calero-Rubio, C.; Sathish, H. A.; Remmele, R. L.; Roberts, C. J. Electrostatically Mediated Protein-Protein Interactions for Monoclonal Antibodies: A Combined Experimental and Coarse-Grained Molecular Modeling Approach. *J. Pharm. Sci.* **2019**, *108* (1), 120–132. <https://doi.org/10.1016/j.xphs.2018.11.004>.
- (43) Ghosh, R.; Calero-Rubio, C.; Saluja, A.; Roberts, C. J. Relating Protein-Protein Interactions and Aggregation Rates from Low to High Concentrations. *J. Pharm. Sci.* **2016**, *105* (3), 1086–1096. <https://doi.org/10.1016/j.xphs.2016.01.004>.
- (44) Shahfar, H.; Du, Q.; Parupudi, A.; Shan, L.; Esfandiary, R.; Roberts, C. J.

- Electrostatically Driven Protein–Protein Interactions: Quantitative Prediction of Second Osmotic Virial Coefficients to Aid Antibody Design. *J. Phys. Chem. Lett.* **2022**, *13*, 1366–1372. <https://doi.org/10.1021/acs.jpcclett.1c03669>.
- (45) Woldeyes, M. A.; Qi, W.; Razinkov, V. I.; Furst, E. M.; Roberts, C. J. How Well Do Low- and High-Concentration Protein Interactions Predict Solution Viscosities of Monoclonal Antibodies? *J. Pharm. Sci.* **2018**, *108* (1), 1–13. <https://doi.org/10.1016/j.xphs.2018.07.007>.
- (46) Woldeyes, M. A.; Josephson, L. L.; Leiske, D. L.; Galush, W. J.; Roberts, C. J.; Furst, E. M. Viscosities and Protein Interactions of Bispecific Antibodies and Their Monospecific Mixtures. *Mol. Pharm.* **2018**, *15*, 4745–4755. <https://doi.org/10.1021/acs.molpharmaceut.8b00706>.
- (47) Woldeyes, M. A.; Qi, W.; Razinkov, V. I.; Furst, E. M.; Roberts, C. J. Temperature Dependence of Protein Solution Viscosity and Protein–Protein Interactions: Insights into the Origins of High-Viscosity Protein Solutions. *Mol. Pharm.* **2020**. <https://doi.org/10.1021/acs.molpharmaceut.0c00552>.
- (48) Forder, J. K.; Ilott, A. J.; Sahin, E.; Roberts, C. J. Simulation of High-Concentration Self-Interactions for Monoclonal Antibodies from Well-Behaved to Poorly-Behaved Systems. *AIChE J.* **2023**, *69* (4). <https://doi.org/https://doi.org/10.1002/aic.17965>.
- (49) Calero-Rubio, C.; Saluja, A.; Sahin, E.; Roberts, C. J. Predicting High-Concentration Interactions of Monoclonal Antibody Solutions: Comparison of Theoretical Approaches for Strongly Attractive Versus Repulsive Conditions. *J. Phys. Chem. B* **2019**, *123* (27), 5709–5720. <https://doi.org/10.1021/acs.jpccb.9b03779>.
- (50) Calero-Rubio, C.; Ghosh, R.; Saluja, A.; Roberts, C. J. Predicting Protein–Protein Interactions of Concentrated Antibody Solutions Using Dilute Solution Data and Coarse-Grained Molecular Models. *J. Pharm. Sci.* **2018**, *107* (5), 1269–1281. <https://doi.org/10.1016/j.xphs.2017.12.015>.
- (51) Xu, A. Y.; Clark, N. J.; Pollastrini, J.; Espinoza, M.; Kim, H.; Kanapuram, S.; Kerwin, B.; Treuheit, M. J.; Krueger, S.; Mcauley, A.; Curtis, J. E. Effects of Monovalent Salt on Protein–Protein Interactions of Dilute and Concentrated Monoclonal Antibody Formulations. *antibodies* **2022**, *11* (24), 1–19.
- (52) Kirkwood, J. G.; Buff, F. P. On the Statistical Mechanical Theory of Solutions. *J. Chem. Phys.* **1958**, *19* (6), 774–777. <https://doi.org/10.1063/1.1744666>.
- (53) Hofmann, M.; Gieseler, H. Predictive Screening Tools Used in High-Concentration Protein Formulation Development. *J. Pharm. Sci.* **2018**, *107* (3), 772–777. <https://doi.org/10.1016/j.xphs.2017.10.036>.
- (54) Kingsbury, J. S.; Saini, A.; Auclair, S. M.; Fu, L.; Lantz, M. M.; Halloran, K. T.; Calero-Rubio, C.; Schwenger, W.; Airiau, C. Y.; Zhang, J.; Gokarn, Y. R. A Single Molecular Descriptor to Predict Solution Behavior of Therapeutic Antibodies. *Sci. Adv.* **2020**, *6* (32). <https://doi.org/10.1126/sciadv.abb0372>.
- (55) Sibanda, N.; Shanmugam, R. K.; Curtis, R. The Relationship between Protein–Protein Interactions and Liquid–Liquid Phase Separation for Monoclonal

- Antibodies. *Mol. Pharm.* **2023**, *20*, 2662–2674.
<https://doi.org/10.1021/acs.molpharmaceut.3c00090>.
- (56) Thiagarajan, G.; Semple, A.; James, J. K.; Cheung, J. K.; Shameem, M. A Comparison of Biophysical Characterization Techniques in Predicting Monoclonal Antibody Stability. *MAbs* **2016**, *8* (6), 1088–1097.
<https://doi.org/10.1080/19420862.2016.1189048>.
- (57) Svilenov, H. L.; Winter, G. Formulations That Suppress Aggregation During Long-Term Storage of a Bispecific Antibody Are Characterized by High Refoldability and Colloidal Stability. *J. Pharm. Sci.* **2020**, *109* (6), 2048–2058.
<https://doi.org/10.1016/j.xphs.2020.03.011>.
- (58) Meric, G.; Naik, S.; Hunter, A. K.; Robinson, A. S.; Roberts, C. J. Challenges for Design of Aggregation-Resistant Variants of Granulocyte Colony-Stimulating Factor. *Biophys. Chem.* **2021**, *277*, 106630.
<https://doi.org/10.1016/j.bpc.2021.106630>.
- (59) Fang, J. A Critical Review of Five Machine Learning-Based Algorithms for Predicting Protein Stability Changes upon Mutation. *Brief. Bioinform.* **2019**, *21* (4), 1285–1292. <https://doi.org/10.1093/bib/bbz071>.
- (60) O'Brien, C. J.; Blanco, M. A.; Costanzo, J. A.; Enterline, M.; Fernandez, E. J.; Robinson, A. S.; Roberts, C. J. Modulating Non-Native Aggregation and Electrostatic Protein-Protein Interactions with Computationally Designed Single-Point Mutations. *Protein Eng. Des. Sel.* **2016**, *29* (6), 231–243.
<https://doi.org/10.1093/protein/gzw010>.
- (61) Zhang, C.; Samad, M.; Yu, H.; Chakroun, N.; Hilton, D.; Dalby, P. A. Computational Design to Reduce Conformational Flexibility and Aggregation Rates of an Antibody Fab Fragment. *Mol. Pharm.* **2018**, *15* (8), 3079–3092.
<https://doi.org/10.1021/acs.molpharmaceut.8b00186>.
- (62) Kastelic, M.; Vlachy, V. Theory for the Liquid-Liquid Phase Separation in Aqueous Antibody Solutions. *J. Phys. Chem. B* **2018**, *122* (21), 5400–5408.
<https://doi.org/10.1021/acs.jpcc.7b11458>.
- (63) Rosch, T. W.; Errington, J. R. Investigation of the Phase Behavior of an Embedded Charge Protein Model through Molecular Simulation. *J. Phys. Chem. B* **2007**, *111* (43), 12591–12598. <https://doi.org/10.1021/jp075455q>.
- (64) Shen, V. K.; Cheung, J. K.; Errington, J. R.; Truskett, T. M. Coarse-Grained Strategy for Modeling Protein Stability in Concentrated Solutions. II: Phase Behavior. *Biophys. J.* **2006**, *90* (6), 1949–1960.
<https://doi.org/10.1529/biophysj.105.076497>.
- (65) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936. <https://doi.org/10.1021/acs.chemrev.6b00163>.
- (66) Cheung, J. K.; Truskett, T. M. Coarse-Grained Strategy for Modeling Protein Stability in Concentrated Solutions. *Biophys. J.* **2005**, *89* (4), 2372–2384.
<https://doi.org/10.1529/biophysj.105.062067>.
- (67) Brien, C. J. O.; Calero-rubio, C.; Razinkov, V. I.; Robinson, A. S.; Roberts, C.

- J.; O'Brien, C. J.; Calero-rubio, C.; Razinkov, V. I.; Robinson, A. S.; Roberts, C. J. Biophysical Characterization and Molecular Simulation of Electrostatically Driven Self-Association of a Single-Chain Antibody. *Protein Sci.* **2018**, *27* (7), 1275–1285. <https://doi.org/10.1002/pro.3415>.
- (68) Woldeyes, M. A.; Calero-Rubio, C.; Furst, E. M.; Roberts, C. J. Predicting Protein Interactions of Concentrated Globular Protein Solutions Using Colloidal Models. *J. Phys. Chem. B* **2017**, *121* (18), 4756–4767. <https://doi.org/10.1021/acs.jpcc.7b02183>.
- (69) Ferreira, G. M.; Shahfar, H.; Sathish, H. A.; Remmele, R. L.; Roberts, C. J. Identifying Key Residues That Drive Strong Electrostatic Attractions between Therapeutic Antibodies. *J. Phys Chem B* **2019**, *123*, 10642–10653. <https://doi.org/10.1021/acs.jpcc.9b08355>.
- (70) Grünberger, A.; Lai, P. K.; Blanco, M. A.; Roberts, C. J. Coarse-Grained Modeling of Protein Second Osmotic Virial Coefficients: Sterics and Short-Ranged Attractions. *J. Phys. Chem. B* **2013**, *117* (3), 763–770. <https://doi.org/10.1021/jp308234j>.
- (71) Blanco, M. A.; Sahin, E.; Robinson, A. S.; Roberts, C. J. Coarse-Grained Model for Colloidal Protein Interactions, B22, and Protein Cluster Formation. *J. Phys. Chem. B* **2013**, *117* (50), 16013–16028. <https://doi.org/10.1021/jp409300j>.
- (72) Bereau, T.; Deserno, M. Generic Coarse-Grained Model for Protein Folding and Aggregation. *J. Chem. Phys.* **2009**, *130*, 1–15. <https://doi.org/10.1063/1.3152842>.
- (73) Chaudhri, A.; Zarraga, I. E.; Kamerzell, T. J.; Brandt, J. P.; Patapoff, T. W.; Shire, S. J.; Voth, G. A. Coarse-Grained Modeling of the Self-Association of Therapeutic Monoclonal Antibodies. *J. Phys. Chem. B* **2012**, *116* (28), 8045–8057. <https://doi.org/10.1021/jp301140u>.
- (74) Godfrin, P. D.; Zarraga, I. E.; Zarzar, J.; Porcar, L.; Falus, P.; Wagner, N. J.; 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. <https://doi.org/10.1021/acs.jpcc.5b07260>.
- (75) Skar-Gislinge, N.; Ronti, M.; Garting, T.; Rischel, C.; Schurtenberger, P.; Zaccarelli, E.; Stradner, A. A Colloid Approach to Self-Assembling Antibodies. *Mol. Pharm.* **2019**, *16* (6), 2394–2404. <https://doi.org/10.1021/acs.molpharmaceut.9b00019>.
- (76) Blanco, M. A.; Hatch, H. W.; Curtis, J. E.; Shen, V. K. Evaluating the Effects of Hinge Flexibility on the Solution Structure of Antibodies at Concentrated Conditions. *J. Pharm. Sci.* **2019**, *108* (5), 1663–1674. <https://doi.org/10.1016/j.xphs.2018.12.013>.
- (77) Chaudhri, A.; Zarraga, I. E.; Yadav, S.; Patapoff, T. W.; Shire, S. J.; Voth, G. A. The Role of Amino Acid Sequence in the Self-Association of Therapeutic Monoclonal Antibodies: Insights from Coarse-Grained Modeling. *J. Phys.*

- Chem. B* **2013**, *117* (5), 1269–1279. <https://doi.org/10.1021/jp3108396>.
- (78) Dear, B. J.; Bollinger, J. A.; Chowdhury, A.; Hung, J. J.; Wilks, L. R.; Karouta, C. A.; Ramachandran, K.; Shay, T. Y.; Nieto, M. P.; Sharma, A.; Cheung, J. K.; Nykypanchuk, D.; Godfrin, P. D.; Johnston, K. P.; Truskett, T. M. X-Ray Scattering and Coarse-Grained Simulations for Clustering and Interactions of Monoclonal Antibodies at High Concentrations. *J. Phys. Chem. B* **2019**, *123*, acs.jp cb.9b04478. <https://doi.org/10.1021/acs.jp cb.9b04478>.
- (79) Sun, G.; Wang, Y.; Lomakin, A.; Benedek, G. B.; Stanley, H. E.; Xu, L.; Buldyrev, S. V. The Phase Behavior Study of Human Antibody Solution Using Multi-Scale Modeling. *J. Chem. Phys.* **2016**, *145* (19). <https://doi.org/10.1063/1.4966972>.
- (80) Hage, K. E. L.; Hédin, F.; Gupta, P. K.; Meuwly, M.; Karplus, M. Valid Molecular Dynamics Simulations of Human Hemoglobin Require a Surprisingly Large Box Size. *Elife* **2018**, *7*, 1–16. <https://doi.org/10.7554/eLife.45318.001>.
- (81) Asthagiri, D.; Tomar, D. S. System Size Dependence of Hydration-Shell Occupancy and Its Implications for Assessing the Hydrophobic and Hydrophilic Contributions to Hydration. *J. Phys. Chem. B* **2020**, *124* (5), 798–806. <https://doi.org/10.1021/acs.jp cb.9b11200>.
- (82) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. Challenges in Protein Folding Simulations: Timescale, Representation, and Analysis. *Nat Phys.* **2010**, *6* (10), 751–758. <https://doi.org/10.1038/nphys1713.Challenges>.
- (83) Shahfar, H.; O'Brien, C. J.; Budyak, I. L.; Roberts, C. J. Predicting Experimental B22 Values and the Effects of Histidine Charge States for Monoclonal Antibodies Using Coarse-Grained Molecular Simulations. *Mol. Pharm.* **2022**, *19* (11), 3820–3830. <https://doi.org/10.1021/acs.molpharmaceut.2c00337>.
- (84) Woldeyes, M. A.; Calero-Rubio, C.; Furst, E. M.; Roberts, C. J. Predicting Protein Interactions of Concentrated Globular Protein Solutions Using Colloidal Models. *J. Phys. Chem. B* **2017**, *121* (18), 4756–4767. <https://doi.org/10.1021/acs.jp cb.7b02183>.
- (85) Kulakova, A.; Augustijn, D.; Bialy, I. El; Gentiluomo, L.; Greco, M. L.; Hervøhansen, S.; Indrakumar, S.; Mahapatra, S.; Morales, M. M.; Pohl, C.; Polimeni, M.; Roche, A.; Svilenov, H. L.; Tosstorff, A.; Zalar, M.; Curtis, R.; Derrick, J. P.; Frieß, W.; Golovanov, A. P.; Lund, M.; Nørgaard, A.; Khan, T. A.; Peters, G. H. J.; Pluen, A.; Roessner, D.; Streicher, W. W.; Walle, C. F. Van Der; Warwicker, J.; Uddin, S.; Winter, G.; Bukrinski, J. T.; Rinnan, Å.; Harris, P. Chemometrics in Protein Formulation: Stability Governed by Repulsion and Protein Unfolding. *Mol. Pharm.* **2023**, *20*, 2951–2965. <https://doi.org/10.1021/acs.molpharmaceut.3c00013>.
- (86) Wang, W.; Roberts, C. J. Protein Aggregation – Mechanisms, Detection, and Control. *Int. J. Pharm.* **2018**, *550* (1), 251–268. <https://doi.org/https://doi.org/10.1016/j.ijpharm.2018.08.043>.

- (87) Kalonia, C.; Toprani, V.; Toth, R.; Wahome, N.; Gabel, I.; Middaugh, C. R.; Volkin, D. B. Effects of Protein Conformation, Apparent Solubility, and Protein-Protein Interactions on the Rates and Mechanisms of Aggregation for an IgG1 Monoclonal Antibody. *J. Phys. Chem. B* **2016**, *120* (29), 7062–7075. <https://doi.org/10.1021/acs.jpcc.6b03878>.
- (88) Wood, C. V.; Razinkov, V. I.; Qi, W.; Furst, E. M.; Roberts, C. J. A Rapid, Small-Volume Approach to Evaluate Protein Aggregation at Air-Water Interfaces. *J. Pharm. Sci.* **2021**, *110* (3), 1083–1092. <https://doi.org/10.1016/j.xphs.2020.11.024>.
- (89) Vlasak, J.; Ionescu, R. Fragmentation of Monoclonal Antibodies. *MABs* **2011**, *3* (3), 253–263. <https://doi.org/10.4161/mabs.3.3.15608>.
- (90) Roberts, C. J. Protein Aggregation and Its Impact on Product Quality. *Curr. Opin. Biotechnol.* **2014**, *30*, 211–217. <https://doi.org/10.1016/j.copbio.2014.08.001>.
- (91) Wang, W.; Ohtake, S. Science and Art of Protein Formulation Development. *Int. J. Pharm.* **2019**, *568*, 118505. <https://doi.org/10.1016/j.ijpharm.2019.118505>.
- (92) Weiss IV, W. F.; Young, T. M.; Roberts, C. J. Principles, Approaches, and Challenges for Predicting Protein Aggregation Rates and Shelf Life. *J. Pharm. Sci.* **2009**, *98* (4), 1246–1277. <https://doi.org/10.1002/jps>.
- (93) Seelig, J.; Seelig, A. Protein Unfolding—Thermodynamic Perspectives and Unfolding Models. *Int. J. Mol. Sci.* **2023**, *24* (6). <https://doi.org/10.3390/ijms24065457>.
- (94) Roberts, C. J. Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends Biotechnol.* **2014**, *32* (7), 372–380. <https://doi.org/10.1016/j.tibtech.2014.05.005>.
- (95) Shen, V. K.; Cheung, J. K.; Errington, J. R.; Truskett, T. M. Insights into Crowding Effects on Protein Stability from a Coarse-Grained Model. *J. Biomech. Eng.* **2009**, *131* (7), 1–7. <https://doi.org/10.1115/1.3127259>.
- (96) Harn, N.; Allan, C.; Oliver, C.; Middaugh, C. R. Highly Concentrated Monoclonal Antibody Solutions: Direct Analysis of Physical Structure and Thermal Stability. *J. Pharm. Sci.* **2007**, *96* (3), 532–546. <https://doi.org/10.1002/jps.20753>.
- (97) Sarangapani, P. S.; Hudson, S. D.; Migler, K. B.; Pathak, J. A. The Limitations of an Exclusively Colloidal View of Protein Solution Hydrodynamics and Rheology. *Biophys. J.* **2013**, *105* (10), 2418–2426. <https://doi.org/https://doi.org/10.1016/j.bpj.2013.10.012>.
- (98) Kamerzell, T. J.; Kanai, S.; Liu, J.; Shire, S. J.; Wang, Y. J. Increasing IgG Concentration Modulates the Conformational Heterogeneity and Bonding Network That Influence Solution Properties. *J. Phys. Chem. B* **2009**, *113*, 6109–6118.
- (99) Kuroda, D.; Tsumoto, K. Engineering Stability, Viscosity, and Immunogenicity of Antibodies by Computational Design. *Journal of Pharmaceutical Sciences*.

- American Pharmacists Association 2020, pp 1631–1651.
<https://doi.org/10.1016/j.xphs.2020.01.011>.
- (100) Meric, G.; Robinson, A. S.; Roberts, C. J. Driving Forces for Nonnative Protein Aggregation and Approaches to Predict Aggregation-Prone Regions. *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8* (1), 139–159. <https://doi.org/10.1146/annurev-chembioeng-060816-101404>.
- (101) Arora, J.; Hu, Y.; Esfandiary, R.; Sathish, H. A.; Bishop, S. M.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. Charge-Mediated Fab-Fc Interactions in an IgG1 Antibody Induce Reversible Self-Association, Cluster Formation, and Elevated Viscosity. *MAbs* **2016**, *8* (8), 1561–1574. <https://doi.org/10.1080/19420862.2016.1222342>.
- (102) Parupudi, A.; Chaturvedi, S. K.; Adão, R.; Harkness, R. W.; Dragulin-Otto, S.; Kay, L. E.; Esfandiary, R.; Zhao, H.; Schuck, P. Global Multi-Method Analysis of Interaction Parameters for Reversibly Self-Associating Macromolecules at High Concentrations. *Sci. Rep.* **2021**, *11* (1), 1–15. <https://doi.org/10.1038/s41598-021-84946-8>.
- (103) Alford, J. R.; Kendrick, B. S.; Carpenter, J. F.; Randolph, T. W. Measurement of the Second Osmotic Virial Coefficient for Protein Solutions Exhibiting Monomer–Dimer Equilibrium. *Anal. Biochem.* **2008**, *377* (2), 128–133. <https://doi.org/https://doi.org/10.1016/j.ab.2008.03.032>.
- (104) Pathak, J. A.; Nugent, S.; Bender, M.; Roberts, C. J.; Curtis, R. J.; Douglas, J. F. Comparison of Huggins Coefficients and Osmotic Second Virial Coefficients of Buffered Solutions of Monoclonal Antibodies. *Polymers (Basel)*. **2021**, *13* (4), 1–21. <https://doi.org/10.3390/polym13040601>.
- (105) Shire, S. J.; Shahrokh, Z.; Liu, J. Challenges in the Development of High Protein Concentration Formulations. *J. Pharm. Sci.* **2004**, *93* (6), 1390–1402. <https://doi.org/10.1002/jps.20079>.
- (106) Izadi, S.; Patapoff, T. W.; Walters, B. T. Multiscale Coarse-Grained Approach to Investigate Self-Association of Antibodies. *Biophys. J.* **2020**, *118* (11), 2741–2754. <https://doi.org/10.1016/j.bpj.2020.04.022>.
- (107) Yearley, E. J.; Godfrin, P. D.; Perevozchikova, T.; Zhang, H.; Falus, P.; Porcar, L.; Nagao, M.; Curtis, J. E.; Gawande, P.; Taing, R.; Zarraga, I. E.; Wagner, N. J.; 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. <https://doi.org/10.1016/j.bpj.2014.02.036>.
- (108) Lilyestrom, W. G.; Yadav, S.; Shire, S. J.; Scherer, T. M. Monoclonal Antibody Self-Association, Cluster Formation, and Rheology at High Concentrations. *J. Phys. Chem. B* **2013**, *117* (21), 6373–6384. <https://doi.org/10.1021/jp4008152>.
- (109) Sibanda, N.; Shanmugam, R. K.; Curtis, R. The Relationship between Protein-Protein Interactions and Liquid-Liquid Phase Separation for Monoclonal Antibodies. *Mol. Pharm.* **2023**, *20* (5), 2662–2674. <https://doi.org/10.1021/acs.molpharmaceut.3c00090>.
- (110) Clodfelter, D. K.; Pekar, A. H.; Rebhun, D. M.; Destrampe, K. A.; Havel, H.

- A.; Myers, S. R.; Brader, M. L. Effects of Non-Covalent Self-Association on the Subcutaneous Absorption of a Therapeutic Peptide. *Pharmaceutical Research*. 1998, pp 254–262. <https://doi.org/10.1023/A:1011918719017>.
- (111) Dobson, C. L.; Devine, P. W. A.; Phillips, J. J.; Higazi, D. R.; Lloyd, C.; Popovic, B.; Arnold, J.; Buchanan, A.; Lewis, A.; Goodman, J.; Van Der Walle, C. F.; Thornton, P.; Vinall, L.; Lowne, D.; Aagaard, A.; Olsson, L. L.; Wollberg, A. R.; Welsh, F.; Karamanos, T. K.; Pashley, C. L.; Iadanza, M. G.; Ranson, N. A.; Ashcroft, A. E.; Kippen, A. D.; Vaughan, T. J.; Radford, S. E.; Lowe, D. C. Engineering the Surface Properties of a Human Monoclonal Antibody Prevents Self-Association and Rapid Clearance in Vivo. *Sci. Rep.* **2016**, *6* (June), 1–14. <https://doi.org/10.1038/srep38644>.
- (112) Kelly, R. L.; Sun, T.; Jain, T.; Caffry, I.; Yu, Y.; Cao, Y.; Lynaugh, H.; Brown, M.; Vásquez, M.; Dane Witttrup, K.; Xu, Y. High Throughput Cross-Interaction Measures for Human IgG1 Antibodies Correlate with Clearance Rates in Mice. *MAbs* **2015**, *7* (4), 770–777. <https://doi.org/10.1080/19420862.2015.1043503>.
- (113) Avery, L. B.; Wade, J.; Wang, M.; Tam, A.; King, A.; Piche-Nicholas, N.; Kavosi, M. S.; Penn, S.; Cirelli, D.; Kurz, J. C.; Zhang, M.; Cunningham, O.; Jones, R.; Fennell, B. J.; McDonnell, B.; Sakorafas, P.; Apgar, J.; Finlay, W. J.; Lin, L.; Bloom, L.; O’Hara, D. M. Establishing in Vitro in Vivo Correlations to Screen Monoclonal Antibodies for Physicochemical Properties Related to Favorable Human Pharmacokinetics. *MAbs* **2018**, *10* (2), 244–255. <https://doi.org/10.1080/19420862.2017.1417718>.
- (114) Moussa, E. M.; Panchal, J. P.; Moorthy, B. S.; Blum, J. S.; Joubert, M. K.; Narhi, L. O.; Topp, E. M. Immunogenicity of Therapeutic Protein Aggregates. *J. Pharm. Sci.* **2016**, *105* (2), 417–430. <https://doi.org/10.1016/j.xphs.2015.11.002>.
- (115) Le Basle, Y.; Chennell, P.; Tokhadze, N.; Astier, A.; Sautou, V. Physicochemical Stability of Monoclonal Antibodies: A Review. *J. Pharm. Sci.* **2020**, *109* (1), 169–190. <https://doi.org/10.1016/j.xphs.2019.08.009>.
- (116) Fathallah, A. M.; Chiang, M.; Mishra, A.; Kumar, S.; Xue, L.; Middaugh, R.; Balu-Iyer, S. V. The Effect of Small Oligomeric Protein Aggregates on the Immunogenicity of Intravenous and Subcutaneous Administered Antibodies. *J. Pharm. Sci.* **2015**, *104* (11), 3691–3702. <https://doi.org/10.1002/jps.24592>.
- (117) Andrews, J. M.; Roberts, C. J. A Lumry-Eyring Nucleated Polymerization Model of Protein Aggregation Kinetics: 1. Aggregation with Pre-Equilibrated Unfolding. *J. Phys. Chem. B* **2007**, *111* (27), 7897–7913. <https://doi.org/10.1021/jp070212j>.
- (118) Gokarn, Y. R.; Fesinmeyer, R. M.; Saluja, A.; Cao, S.; Dankberg, J.; Goetze, A.; Remmele, R. L.; Narhi, L. O.; Brems, D. N. Ion-Specific Modulation of Protein Interactions: Anion-Induced, Reversible Oligomerization of a Fusion Protein. *Protein Sci.* **2009**, *18* (1), 169–179. <https://doi.org/10.1002/pro.20>.
- (119) O’Brien, C. J.; Calero-Rubio, C.; Razinkov, V. I.; Robinson, A. S.; Roberts, C. J. Biophysical Characterization and Molecular Simulation of Electrostatically

- Driven Self-Association of a Single-Chain Antibody. *Protein Sci.* **2018**, *27* (7), 1275–1285. <https://doi.org/10.1002/pro.3415>.
- (120) Xu, A. Y.; Castellanos, M. M.; Mattison, K.; Krueger, S.; Curtis, J. E. Studying Excipient Modulated Physical Stability and Viscosity of Monoclonal Antibody Formulations Using Small-Angle Scattering. *Mol. Pharm.* **2019**, *16* (10), 4319–4338. <https://doi.org/10.1021/acs.molpharmaceut.9b00687>.
- (121) Saito, S.; Hasegawa, J.; Kobayashi, N.; Kishi, N.; Uchiyama, S.; Fukui, K. Behavior of Monoclonal Antibodies: Relation between the Second Virial Coefficient (B₂) at Low Concentrations and Aggregation Propensity and Viscosity at High Concentrations. *Pharm. Res.* **2012**, *29* (2), 397–410. <https://doi.org/10.1007/s11095-011-0563-x>.
- (122) Gomes, D.; Kalman, R. K.; Pagels, R. K.; Rodrigues, M. A.; Roberts, C. J. Parallel Chromatography and in Situ Scattering to Interrogate Competing Protein Aggregation Pathways. *Protein Sci.* **2018**, *27* (7), 1325–1333. <https://doi.org/10.1002/pro.3435>.
- (123) Shi, H.; Singh, N.; Esselborn, F.; Blobel, G. Structure of a Myosin•adaptor Complex and Pairing by Cargo. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (12). <https://doi.org/10.1073/pnas.1401428111>.
- (124) Philo, J. S. Is Any Measurement Method Optimal for All Aggregate Sizes and Types? *AAPS J.* **2006**, *8* (3), 564–571. <https://doi.org/10.1208/aapsj080365>.
- (125) Wang, W.; Roberts, C. J. Protein Aggregation – Mechanisms, Detection, and Control. *Int. J. Pharm.* **2018**, *550* (1–2), 251–268. <https://doi.org/10.1016/j.ijpharm.2018.08.043>.
- (126) Ratanji, K. D.; Derrick, J. P.; Dearman, R. J.; Kimber, I. Immunogenicity of Therapeutic Proteins: Influence of Aggregation. *J. Immunotoxicol.* **2014**, *11* (2), 99–109. <https://doi.org/10.3109/1547691X.2013.821564>.
- (127) Lundahl, M. L. E.; Fogli, S.; Colavita, P. E.; Scanlan, E. M. Aggregation of Protein Therapeutics Enhances Their Immunogenicity: Causes and Mitigation Strategies. *RSC Chem. Biol.* **2021**, *2* (4), 1004–1020. <https://doi.org/10.1039/d1cb00067e>.
- (128) Hung, J. J.; Borwankar, A. U.; Dear, B. J.; Truskett, T. M.; Johnston, K. P. High Concentration Tangential Flow Ultrafiltration of Stable Monoclonal Antibody Solutions with Low Viscosities. *J. Memb. Sci.* **2016**, *508*, 113–126. <https://doi.org/10.1016/j.memsci.2016.02.031>.
- (129) Wälchli, R.; Vermeire, P. J.; Massant, J.; Arosio, P. Accelerated Aggregation Studies of Monoclonal Antibodies: Considerations for Storage Stability. *J. Pharm. Sci.* **2020**, *109* (1), 595–602. <https://doi.org/10.1016/j.xphs.2019.10.048>.
- (130) Li, Y.; Roberts, C. J. Lumry-Eyring Nucleated-Polymerization Model of Protein Aggregation Kinetics. 2. Competing Growth via Condensation and Chain Polymerization. *J. Phys. Chem. B* **2009**, *113* (19), 7020–7032. <https://doi.org/10.1021/jp8083088>.
- (131) Kijanka, G.; Bee, J. S.; Korman, S. A.; Wu, Y.; Roskos, L. K.; Schenerman, M.

- A.; Slütter, B.; Jiskoot, W. Submicron Size Particles of a Murine Monoclonal Antibody Are More Immunogenic Than Soluble Oligomers or Micron Size Particles Upon Subcutaneous Administration in Mice. *J. Pharm. Sci.* **2018**, *107* (11), 2847–2859. <https://doi.org/10.1016/j.xphs.2018.06.029>.
- (132) Joubert, M. K.; Hokom, M.; Eakin, C.; Zhou, L.; Deshpande, M.; Baker, M. P.; Goletz, T. J.; Kerwin, B. A.; Chirmule, N.; Narhi, L. O.; Jawa, V. Highly Aggregated Antibody Therapeutics Can Enhance the in Vitro Innate and Late-Stage T-Cell Immune Responses. *J. Biol. Chem.* **2012**, *287* (30), 25266–25279. <https://doi.org/10.1074/jbc.M111.330902>.
- (133) Telikepalli, S. N.; Kumru, O. S.; Kalonia, C.; Esfandiary, R.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B. Structural Characterization of IgG1 MAb Aggregates and Particles Generated under Various Stress Conditions. *J. Pharm. Sci.* **2014**, *103* (3), 796–809. <https://doi.org/10.1002/jps.23839>.
- (134) Barnard, J. G.; Singh, S.; Randolph, T. W.; Carpenter, J. F. Subvisible Particle Counting Provides a Sensitive Method of Detecting and Quantifying Aggregation of Monoclonal Antibody Caused by Freeze-Thawing: Insights Into the Roles of Particles in the Protein Aggregation Pathway. *J. Pharm. Sci.* **2011**, *100* (2), 492–503. <https://doi.org/10.1002/jps>.
- (135) *USP 787 Subvisible Particulate Matter in Therapeutic Protein Injections*. The United States Pharmacopeia-National Formulary. <https://www.usp.org> (accessed 2023-12-05).
- (136) *USP 788 Particulate Matter in Injections*. The United States Pharmacopeia-National Formulary. <https://www.usp.org> (accessed 2023-12-05).
- (137) Saluja, A.; Kalonia, D. S. Nature and Consequences of Protein-Protein Interactions in High Protein Concentration Solutions. *Int. J. Pharm.* **2008**, *358* (1–2), 1–15. <https://doi.org/10.1016/j.ijpharm.2008.03.041>.
- (138) Kumar, V.; Dixit, N.; Zhou, L.; Fraunhofer, W. Impact of Short Range Hydrophobic Interactions and Long Range Electrostatic Forces on the Aggregation Kinetics of a Monoclonal Antibody and a Dual-Variable Domain Immunoglobulin at Low and High Concentrations. *Int. J. Pharm.* **2011**, *421* (1), 82–93. <https://doi.org/10.1016/j.ijpharm.2011.09.017>.
- (139) Chou, D. K.; Krishnamurthy, R.; Manning, M. C.; Randolph, T. W.; Carpenter, J. F. Physical Stability of Albinterferon-A2b In Aqueous Solution: Effects of Conformational Stability and Colloidal Stability on Aggregation. *J. Pharm. Sci.* **2012**, *101* (8), 2702–2719. <https://doi.org/10.1002/jps.23215>.
- (140) Laber, J. R.; Dear, B. J.; Martins, M. L.; Jackson, D. E.; Divenere, A.; Gollihar, J. D.; Ellington, A. D.; Truskett, T. M.; Johnston, K. P.; Maynard, J. A. Charge Shielding Prevents Aggregation of Supercharged GFP Variants at High Protein Concentration. *Mol. Pharm.* **2017**, *14* (10), 3269–3280. <https://doi.org/10.1021/acs.molpharmaceut.7b00322>.
- (141) Wood, C. V.; Mcevoy, S.; Razinkov, V. I.; Qi, W.; Furst, E. M.; Roberts, C. J. Kinetics and Competing Mechanisms of Antibody Aggregation via Bulk and Surface-Mediated Pathways. *J. Pharm. Sci.* **2020**, *109*, 1449–11459.

- <https://doi.org/10.1016/j.xphs.2020.01.005>.
- (142) Lazar, K. L.; Patapoff, T. W.; Sharma, V. K. Cold Denaturation of Monoclonal Antibodies. *MAbs* **2010**, *2* (1), 42–52. <https://doi.org/10.4161/mabs.2.1.10787>.
- (143) Du, C.; Barnett, G.; Borwankar, A.; Lewandowski, A.; Singh, N.; Ghose, S.; Borys, M.; Li, Z. J. Protection of Therapeutic Antibodies from Visible Light Induced Degradation: Use Safe Light in Manufacturing and Storage. *Eur. J. Pharm. Biopharm.* **2018**, *127* (February), 37–43. <https://doi.org/10.1016/j.ejpb.2018.02.007>.
- (144) Halley, J.; Chou, Y. R.; Cicchino, C.; Huang, M.; Sharma, V.; Tan, N. C.; Thakkar, S.; Zhou, L. L.; Al-Azzam, W.; Cornen, S.; Gauden, M.; Gu, Z.; Kar, S.; Lazar, A. C.; Mehndiratta, P.; Smith, J.; Sosic, Z.; Weisbach, P.; Stokes, E. S. E. An Industry Perspective on Forced Degradation Studies of Biopharmaceuticals: Survey Outcome and Recommendations. *J. Pharm. Sci.* **2020**, *109* (1), 6–21. <https://doi.org/10.1016/j.xphs.2019.09.018>.
- (145) Kayser, V.; Chennamsetty, N.; Voynov, V.; Helk, B.; Forrer, K.; Trout, B. L. Evaluation of a Non-Arrhenius Model for Therapeutic Monoclonal Antibody Aggregation. *J. Pharm. Sci.* **2011**, *100* (7), 2526–2542. <https://doi.org/10.1002/jps.22493>.
- (146) Li, Y. I.; Ogunnaike, B. A.; Roberts, C. J. Multi-Variate Approach to Global Protein Aggregation Behavior and Kinetics: Effects of PH, NaCl, and Temperature for α -Chymotrypsinogen A. *J. Pharm. Sci.* **2010**, *99* (2), 645–662. <https://doi.org/10.1002/jps.21869>.
- (147) Barnett, G. V.; Qi, W.; Amin, S.; Neil Lewis, E.; Roberts, C. J. Aggregate Structure, Morphology and the Effect of Aggregation Mechanisms on Viscosity at Elevated Protein Concentrations. *Biophys. Chem.* **2015**, *207*, 21–29. <https://doi.org/10.1016/j.bpc.2015.07.002>.
- (148) Gentiluomo, L.; Svilenov, H. L.; Augustijn, D.; El Bialy, I.; Greco, M. L.; Kulakova, A.; Indrakumar, S.; Mahapatra, S.; Morales, M. M.; Pohl, C.; Roche, A.; Tosstorff, A.; Curtis, R.; Derrick, J. P.; Nørgaard, A.; Khan, T. A.; Peters, G. H. J.; Pluen, A.; Rinnan, Å.; Streicher, W. W.; Van Der Walle, C. F.; Uddin, S.; Winter, G.; Roessner, D.; Harris, P.; Frieß, W. Advancing Therapeutic Protein Discovery and Development through Comprehensive Computational and Biophysical Characterization. *Mol. Pharm.* **2020**, *17* (2), 426–440. <https://doi.org/10.1021/acs.molpharmaceut.9b00852>.
- (149) van der Kant, R.; Karow-Zwick, A. R.; Van Durme, J.; Blech, M.; Gallardo, R.; Seeliger, D.; Abfal, K.; Baatsen, P.; Compennolle, G.; Gils, A.; Studts, J. M.; Schulz, P.; Garidel, P.; Schymkowitz, J.; Rousseau, F. Prediction and Reduction of the Aggregation of Monoclonal Antibodies. *J. Mol. Biol.* **2017**, *429* (8), 1244–1261. <https://doi.org/10.1016/j.jmb.2017.03.014>.
- (150) Fang, Y.; Gao, S.; Tai, D.; Middaugh, C. R.; Fang, J. Identification of Properties Important to Protein Aggregation Using Feature Selection. *BMC Bioinformatics* **2013**, *14* (1). <https://doi.org/10.1186/1471-2105-14-314>.
- (151) Prabakaran, R.; Rawat, P.; Thangakani, A. M.; Kumar, S.; Gromiha, M. M.

- Protein Aggregation: In Silico Algorithms and Applications. *Biophys. Rev.* **2021**, *13* (1), 71–89. <https://doi.org/10.1007/s12551-021-00778-w>.
- (152) Tomar, D. S.; Kumar, S.; Singh, S. K.; 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. <https://doi.org/10.1080/19420862.2015.1128606>.
- (153) Krause, M. E.; Narang, A. S.; Barker, G.; Herzer, S.; Deshmukh, S.; Lan, W.; Fichana, D.; Wasyluk, J. M.; Demirdirek, B.; Zhang, L.; Fiske, J.; McGann, M.; Adams, M. L.; Gandhi, R. B. Buffer Exchange Path Influences the Stability and Viscosity upon Storage of a High Concentration Protein. *Eur. J. Pharm. Biopharm.* **2018**, *131*, 60–69. <https://doi.org/10.1016/j.ejpb.2018.07.014>.
- (154) Duerkop, M.; Berger, E.; Dürauer, A.; Jungbauer, A. Impact of Cavitation, High Shear Stress and Air/Liquid Interfaces on Protein Aggregation. *Biotechnol. J.* **2018**, *13* (7), 1–9. <https://doi.org/10.1002/biot.201800062>.
- (155) Chowdhury, A.; Bollinger, J. A.; Dear, B. J.; Cheung, J. K.; Johnston, K. P.; Truskett, T. M. 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.* **2020**, *17* (5), 1748–1756. <https://doi.org/10.1021/acs.molpharmaceut.9b00960>.
- (156) Whitaker, N.; Xiong, J.; Pace, S. E.; Kumar, V.; Middaugh, C. R.; Joshi, S. B.; Volkin, D. B. A Formulation Development Approach to Identify and Select Stable Ultra-High-Concentration Monoclonal Antibody Formulations With Reduced Viscosities. *J. Pharm. Sci.* **2017**, *106* (11), 3230–3241. <https://doi.org/10.1016/j.xphs.2017.06.017>.
- (157) Kelley, B. Industrialization of MAb Production Technology. *MABs* **2009**, *1* (5), 443–452.
- (158) Elgert, K. D. Antibody Structure and Function. In *Immunology: Understanding the Immune System*; Wiley-Blackwell, 2009; pp 58–78.
- (159) Sha, S.; Agarabi, C.; Brorson, K.; Lee, D. Y.; Yoon, S. N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies. *Trends Biotechnol.* **2016**, *34* (10), 835–846. <https://doi.org/10.1016/j.tibtech.2016.02.013>.
- (160) Cain, P.; Huang, L.; Tang, Y.; Anguiano, V.; Feng, Y. Impact of IgG Subclass on Monoclonal Antibody Developability. *MABs* **2023**, *15* (1), 1–12. <https://doi.org/10.1080/19420862.2023.2191302>.
- (161) Schmidt, S. R. Fusion Proteins: Applications and Challenges. *Fusion Protein Technol. Biopharm. Appl. Challenges* **2009**, *12* (2), 1–24. <https://doi.org/10.1002/9781118354599.ch1>.
- (162) Czajkowsky, D. M.; Hu, J.; Shao, Z.; Pleass, R. J. Fc-Fusion Proteins: New Developments and Future Perspectives. *EMBO Mol. Med.* **2012**, *4* (10), 1015–1028. <https://doi.org/https://doi.org/10.1002/emmm.201201379>.
- (163) Duivelshof, B. L.; Murisier, A.; Camperi, J.; Fekete, S.; Beck, A.; Guillarme, D.; D’Atri, V. Therapeutic Fc-Fusion Proteins: Current Analytical Strategies. *J. Sep. Sci.* **2021**, *44* (1), 35–62.

- <https://doi.org/https://doi.org/10.1002/jssc.202000765>.
- (164) Chen, X.; Zaro, J. L.; Shen, W. C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **2013**, *65* (10), 1357–1369. <https://doi.org/10.1016/j.addr.2012.09.039>.
- (165) Ducharme, E.; Weinberg, J. M. Etanercept. *Expert Opin. Biol. Ther.* **2008**, *8* (4), 491–502. <https://doi.org/10.1136/bmj.3.5611.176-a>.
- (166) Larsen, C. P.; Pearson, T. C.; Adams, A. B.; Tso, P.; Shirasugi, N.; Strobert, E.; Anderson, D.; Cowan, S.; Price, K.; Naemura, J.; Emswiler, J.; Greene, J.; Turk, L. A.; Bajorath, J.; Townsend, R.; Hagerty, D.; Linsley, P. S.; Peach, R. J. Rational Development of LEA29Y (Belatacept), a High-Affinity Variant of CTLA4-Ig with Potent Immunosuppressive Properties. *Am. J. Transplant.* **2005**, *5* (3), 443–453. <https://doi.org/https://doi.org/10.1111/j.1600-6143.2005.00749.x>.
- (167) Cavaco, M.; Castanho, M. A. R. B.; Neves, V. Peptibodies: An Elegant Solution for a Long-Standing Problem. *Pept. Sci.* **2018**, *110* (1), e23095. <https://doi.org/https://doi.org/10.1002/bip.23095>.
- (168) Dumont, J. A.; Low, S. C.; Peters, R. T.; Bitonti, A. J. Monomeric Fc Fusions. *BioDrugs* **2006**, *20* (3), 151–160. <https://doi.org/10.2165/00063030-200620030-00002>.
- (169) Lauer, T. M.; Agrawal, N. J.; Chennamsetty, N.; Egodage, K.; Helk, B.; Trout, B. L. Developability Index: A Rapid in Silico Tool for the Screening of Antibody Aggregation Propensity. *J. Pharm. Sci.* **2012**, *101* (1), 102–115. <https://doi.org/10.1002/jps.22758>.
- (170) Obrezanova, O.; Arnell, A.; De La Cuesta, R. G.; Berthelot, M. E.; Gallagher, T. R. A.; Zurdo, J.; Stallwood, Y. Aggregation Risk Prediction for Antibodies and Its Application to Biotherapeutic Development. *MAbs* **2015**, *7* (2), 352–363. <https://doi.org/10.1080/19420862.2015.1007828>.
- (171) Bye, J. W.; Curtis, R. A. Controlling Phase Separation of Lysozyme with Polyvalent Anions. *J. Phys. Chem. B* **2019**, *123* (3), 593–605. <https://doi.org/10.1021/acs.jpcc.8b10868>.
- (172) Roberts, D.; Keeling, R.; Tracka, M.; Van Der Walle, C. F.; Uddin, S.; Warwick, J.; Curtis, R. Specific Ion and Buffer Effects on Protein-Protein Interactions of a Monoclonal Antibody. *Mol. Pharm.* **2015**, *12* (1), 179–193. <https://doi.org/10.1021/mp500533c>.
- (173) Connolly, B. D.; Petry, C.; Yadav, S.; Demeule, B.; Ciaccio, N.; Moore, J. M. R.; Shire, S. J.; Gokarn, Y. R. Weak Interactions Govern the Viscosity of Concentrated Antibody Solutions: High-Throughput Analysis Using the Diffusion Interaction Parameter. *Biophys. J.* **2012**, *103* (1), 69–78. <https://doi.org/10.1016/j.bpj.2012.04.047>.
- (174) Lehermayr, C.; Mahler, H. C.; Mäder, K.; Fischer, S. Assessment of Net Charge and Protein-Protein Interactions of Different Monoclonal Antibodies. *J. Pharm. Sci.* **2011**, *100* (7), 2551–2562. <https://doi.org/10.1002/jps.22506>.
- (175) Fossat, M. J.; Pappu, R. V. Q-Canonical Monte Carlo Sampling for Modeling

- the Linkage between Charge Regulation and Conformational Equilibria of Peptides. *J. Phys. Chem. B* **2019**, *123* (32), 6952–6967.
<https://doi.org/10.1021/acs.jpcc.9b05206>.
- (176) Ben-Naim, A. Y. *Statistical Thermodynamics for Chemists and Biochemists*, 1st ed.; Plenum Press, New York, 1992.
- (177) Schultz, A. J.; Kofke, D. A. Virial Coefficients of Model Alkanes. *J. Chem. Phys.* **2010**, *133* (10). <https://doi.org/10.1063/1.3486085>.
- (178) Singh, J. K.; Kofke, D. A. Mayer Sampling: Calculation of Cluster Integrals Using Free-Energy Perturbation Methods. *Phys. Rev. Lett.* **2004**, *92* (22), 22–25. <https://doi.org/10.1103/physrevlett.92.220601>.
- (179) Shahfar, H.; Forder, J. K.; Roberts, C. J. Toward a Suite of Coarse-Grained Models for Molecular Simulation of Monoclonal Antibodies and Therapeutic Proteins. *J. Phys. Chem. B* **2021**, *125* (14), 3574–3588.
<https://doi.org/10.1021/acs.jpcc.1c01903>.
- (180) Nandou, L.; A., K. D.; B., W. T. Improving the Efficiency and Reliability of Free Energy Perturbation Calculations Using Overlap Sampling Methods. *J. Comput. Chem.* **2003**, *25* (1), 28–40. <https://doi.org/10.1002/jcc.10369>.
- (181) Schultz, A. J.; Kofke, D. A. Sixth, Seventh and Eighth Virial Coefficients of the Lennard-Jones Model. *Mol. Phys.* **2009**, *107* (21), 2309–2318.
<https://doi.org/10.1080/00268970903267053>.
- (182) Schultz, A. J.; Kofke, D. A. Fifth to Eleventh Virial Coefficients of Hard Spheres. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2014**, *90* (2), 1–14.
<https://doi.org/10.1103/PhysRevE.90.023301>.
- (183) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (184) Sorret, L. L.; DeWinter, M. A.; Schwartz, D. K.; Randolph, T. W. Challenges in Predicting Protein-Protein Interactions from Measurements of Molecular Diffusivity. *Biophys. J.* **2016**, *111* (9), 1831–1842.
<https://doi.org/10.1016/j.bpj.2016.09.018>.
- (185) McQuarrie, D. A. *Statistical Mechanics*, 2nd ed.; University Science Books, 2000.
- (186) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.
<https://doi.org/10.1021/ct100578z>.
- (187) Miyazawa, S.; Jernigan, R. L. Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *J Mol Biol* **1996**, 623–644.
- (188) Pacios, L. F. Distinct Molecular Surfaces and Hydrophobicity of Amino Acid Residues in Proteins. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3–6), 1427–1435.
<https://doi.org/10.1021/ci010369n>.

- (189) Padlan, E. A. Anatomy of the Antibody Molecule. *Mol. Immunol.* **1994**, *31* (3), 169–217. [https://doi.org/10.1016/0161-5890\(94\)90001-9](https://doi.org/10.1016/0161-5890(94)90001-9).
- (190) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (191) Blanco, M. A.; Sahin, E.; Li, Y.; Roberts, C. J. Reexamining Protein-Protein and Protein-Solvent Interactions from Kirkwood-Buff Analysis of Light Scattering in Multi-Component Solutions. *J. Chem. Phys.* **2011**, *134* (22), 1–12. <https://doi.org/10.1063/1.3596726>.
- (192) Blanco, M. A.; Shen, V. K. Effect of the Surface Charge Distribution on the Fluid Phase Behavior of Charged Colloids and Proteins. *J. Chem. Phys.* **2016**, *145* (15). <https://doi.org/10.1063/1.4964613>.
- (193) Buck, P. M.; Chaudhri, A.; Kumar, S.; Singh, S. K. Highly Viscous Antibody Solutions Are a Consequence of Network Formation Caused by Domain-Domain Electrostatic Complementarities: Insights from Coarse-Grained Simulations. *Mol. Pharm.* **2015**, *12* (1), 127–139. <https://doi.org/10.1021/mp500485w>.
- (194) Wang, G.; Varga, Z.; Hofmann, J.; Zarraga, I. E.; Swan, J. W. Structure and Relaxation in Solutions of Monoclonal Antibodies. *J. Phys. Chem. B* **2018**, *122* (11), 2867–2880. <https://doi.org/10.1021/acs.jpccb.7b11053>.
- (195) Corbett, D.; Hebditch, M.; Keeling, R.; Ke, P.; Ekizoglou, S.; Sarangapani, P.; Pathak, J.; Van Der Walle, C. F.; Uddin, S.; Baldock, C.; Avendanlo, C.; Curtis, R. A. Coarse-Grained Modeling of Antibodies from Small-Angle Scattering Profiles. *J. Phys. Chem. B* **2017**, *121* (35), 8276–8290. <https://doi.org/10.1021/acs.jpccb.7b04621>.
- (196) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The Multiscale Coarse-Graining Method. II. Numerical Implementation for Coarse-Grained Molecular Models. *J. Chem. Phys.* **2008**, *128* (24). <https://doi.org/10.1063/1.2938857>.
- (197) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-Grained Model of the SARS-CoV-2 Virion. *Biophys. J.* **2020**, 1–8. <https://doi.org/10.1016/j.bpj.2020.10.048>.
- (198) Cao, Z.; Voth, G. A. The Multiscale Coarse-Graining Method. XI. Accurate Interactions Based on the Centers of Charge of Coarse-Grained Sites. *J. Chem. Phys.* **2015**, *143* (24). <https://doi.org/10.1063/1.4933249>.
- (199) Shaul, K. R. S.; Schultz, A. J.; Perera, A.; Kofke, D. A.; Schultz, A. J.; Perera, A.; Integral-equation, D. A. K. Integral-Equation Theories and Mayer-Sampling Monte Carlo: A Tandem Approach for Computing Virial Coefficients of Simple Fluids. *Mol. Phys.* **2011**, *8976*, 2395–2406. <https://doi.org/10.1080/00268976.2011.615764>.
- (200) Curtis, R. A.; Lue, L. A Molecular Approach to Bioseparations: Protein-Protein and Protein-Salt Interactions. *Chem. Eng. Sci.* **2006**, *61* (3), 907–923.

- <https://doi.org/10.1016/j.ces.2005.04.007>.
- (201) Pasquier, C.; Vazdar, M.; Forsman, J.; Jungwirth, P.; Lund, M. Anomalous Protein-Protein Interactions in Multivalent Salt Solution. *J. Phys. Chem. B* **2017**, *121* (14), 3000–3006. <https://doi.org/10.1021/acs.jpcc.7b01051>.
- (202) Hladílková, J.; Callisen, T. H.; Lund, M. Lateral Protein-Protein Interactions at Hydrophobic and Charged Surfaces as a Function of PH and Salt Concentration. *J. Phys. Chem. B* **2016**, *120* (13), 3303–3310. <https://doi.org/10.1021/acs.jpcc.5b12225>.
- (203) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. Design of Therapeutic Proteins with Enhanced Stability. *Pnas* **2009**, *106* (29), 11937–11942.
- (204) Lund, M.; Jungwirth, P. Patchy Proteins, Anions and the Hofmeister Series. *J. Phys. Condens. Matter* **2008**, *20* (49). <https://doi.org/10.1088/0953-8984/20/49/494218>.
- (205) Li, W.; Persson, B. A.; Morin, M.; Behrens, M. A.; Lund, M.; Zackrisson Oskolkova, M. Charge-Induced Patchy Attractions between Proteins. *J. Phys. Chem. B* **2015**, *119* (2), 503–508. <https://doi.org/10.1021/jp512027j>.
- (206) Yigit, C.; Heyda, J.; Dzubiella, J. Charged Patchy Particle Models in Explicit Salt: Ion Distributions, Electrostatic Potentials, and Effective Interactions. *J. Chem. Phys.* **2015**, *143* (6). <https://doi.org/10.1063/1.4928077>.
- (207) Cheung, J. K.; Shen, V. K.; Errington, J. R.; Truskett, T. M. Coarse-Grained Strategy for Modeling Protein Stability in Concentrated Solutions. III: Directional Protein Interactions. *Biophys. J.* **2007**, *92* (12), 4316–4324. <https://doi.org/10.1529/biophysj.106.099085>.
- (208) Quang, L. J.; Sandler, S. I.; Lenhoff, A. M. Anisotropic Contributions to Protein-Protein Interactions. *J. Chem. Theory Comput.* **2014**, *10* (2), 835–845. <https://doi.org/10.1021/ct4006695>.
- (209) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. Molecular Origins of Osmotic Second Virial Coefficients of Proteins. *Biophys. J.* **1998**, *75* (5), 2469–2477. [https://doi.org/10.1016/S0006-3495\(98\)77691-X](https://doi.org/10.1016/S0006-3495(98)77691-X).
- (210) Kaieda, S.; Lund, M.; Plivelic, T. S.; Halle, B. Weak Self-Interactions of Globular Proteins Studied by Small-Angle X-Ray Scattering and Structure-Based Modeling. *J. Phys. Chem. B* **2014**, *118* (34), 10111–10119. <https://doi.org/10.1021/jp505809v>.
- (211) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42* (1), 73–93. <https://doi.org/10.1146/annurev-biophys-083012-130348>.
- (212) Camperi, J.; Dahotre, S.; Guillarme, D.; Stella, C. Monitoring Multiple Quality Attributes of a Complex Fc-Fusion Protein during Cell Culture Production Processes by MD-LC-MS Peptide Mapping. *Talanta* **2022**, *246*, 123519. <https://doi.org/https://doi.org/10.1016/j.talanta.2022.123519>.
- (213) Yu, Z.; Reid, J. C.; Yang, Y. P. Utilizing Dynamic Light Scattering as a Process Analytical Technology for Protein Folding and Aggregation Monitoring in

- Vaccine Manufacturing. *J. Pharm. Sci.* **2013**, *102* (12), 4284–4290.
<https://doi.org/10.1002/jps.23746>.
- (214) Kuzman, D.; Bunc, M.; Ravnik, M.; Reiter, F.; Žagar, L.; Bončina, M. Long-Term Stability Predictions of Therapeutic Monoclonal Antibodies in Solution Using Arrhenius-Based Kinetics. *Sci. Rep.* **2021**, *11* (1), 1–15.
<https://doi.org/10.1038/s41598-021-99875-9>.
- (215) Kim, N. A.; An, I. B.; Lim, D. G.; Lim, J. Y.; Lee, S. Y.; Shim, W. S.; Kang, N. G.; Jeong, S. H. Effects of PH and Buffer Concentration on the Thermal Stability of Etanercept Using DSC and DLS. *Biol. Pharm. Bull.* **2014**, *37* (5), 808–816. <https://doi.org/10.1248/bpb.b13-00926>.
- (216) Patel, T. R.; Meier, M.; Li, J.; Morris, G.; Rowe, A. J.; Stetefeld, J. T-Shaped Arrangement of the Recombinant Agrin G3 - IgG Fc Protein. *Protein Sci.* **2011**, *20* (6), 931–940. <https://doi.org/10.1002/pro.628>.
- (217) Chen, Z.; Huang, C.; Chennamsetty, N.; Xu, X.; Li, Z. J. Insights in Understanding Aggregate Formation and Dissociation in Cation Exchange Chromatography for a Structurally Unstable Fc-Fusion Protein. *J. Chromatogr. A* **2016**, *1460* (2016), 110–122. <https://doi.org/10.1016/j.chroma.2016.07.023>.
- (218) Wang, T.; Fodor, S.; Hapuarachchi, S.; Jiang, X. G.; Chen, K.; Apostol, I.; Huang, G. Analysis and Characterization of Aggregation of a Therapeutic Fc-Fusion Protein. *J. Pharm. Biomed. Anal.* **2013**, *72*, 59–64.
<https://doi.org/10.1016/j.jpba.2012.09.010>.
- (219) Fast, J. L.; Cordes, A. A.; Carpenter, J. F.; Randolph, T. W. Physical Instability of a Therapeutic Fc Fusion Protein: Domain Contributions to Conformational and Colloidal Stability. *Biochemistry* **2009**, *48* (49), 11724–11736.
<https://doi.org/10.1021/bi900853v>.
- (220) Frisken, B. J. Revisiting the Method of Cumulants for Analysis of Dynamic Light Scattering Data. In *Optics InfoBase Conference Papers; 2001; Vol. 40*, pp 4087–4091. <https://doi.org/10.1364/ao.40.004087>.
- (221) Nägele, G. On the Dynamics and Structure of Charge-Stabilized Suspensions. *Phys. Rep.* **1996**, *272* (5–6), 215–372. [https://doi.org/10.1016/0370-1573\(95\)00078-X](https://doi.org/10.1016/0370-1573(95)00078-X).
- (222) Roberts, D.; Keeling, R.; Tracka, M.; Van Der Walle, C. F.; 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.
<https://doi.org/10.1021/mp5002334>.
- (223) Allahyarov, E.; Löwen, H.; Hansen, J. P.; Louis, A. A. Nonmonotonic Variation with Salt Concentration of the Second Virial Coefficient in Protein Solutions. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **2003**, *67* (5), 13. <https://doi.org/10.1103/PhysRevE.67.051404>.
- (224) Mieczkowski, C.; Cheng, A.; Fischmann, T.; Hsieh, M.; Baker, J.; Uchida, M.; Raghunathan, G.; Strickland, C.; Fayadat-Dilman, L. Characterization and Modeling of Reversible Antibody Self-Association Provide Insights into Behavior, Prediction, and Correction. *Antibodies* **2021**, *10* (1).

- <https://doi.org/10.3390/antib10010008>.
- (225) Blanco, M. A.; Perevozchikova, T.; Martorana, V.; Manno, M.; Roberts, C. J. Protein-Protein Interactions in Dilute to Concentrated Solutions: α -Chymotrypsinogen in Acidic Conditions. *J. Phys. Chem. B* **2014**, *118* (22), 5817–5831. <https://doi.org/10.1021/jp412301h>.
- (226) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press: London, UK, 2002.
- (227) Errington, J. R. Direct Calculation of Liquid-Vapor Phase Equilibria from Transition Matrix Monte Carlo Simulation. *J. Chem. Phys.* **2003**, *118* (22), 9915–9925. <https://doi.org/10.1063/1.1572463>.
- (228) Lai, P. K.; Swan, J. W.; Trout, B. L. Calculation of Therapeutic Antibody Viscosity with Coarse-Grained Models, Hydrodynamic Calculations and Machine Learning-Based Parameters. *MAbs* **2021**, *13* (1). <https://doi.org/10.1080/19420862.2021.1907882>.
- (229) Uchiyama, S. Liquid Formulation for Antibody Drugs. *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844* (11), 2041–2052. <https://doi.org/10.1016/j.bbapap.2014.07.016>.
- (230) Reis, P. B. P. S.; Vila-Vicosa, D.; Rocchia, W.; Machuqueiro, M. PypKA: A Flexible Python Module for Poisson-Boltzmann-Based PKa Calculations. *J. Chem. Inf. Model.* **2020**, *60* (10), 4442–4448. <https://doi.org/10.1021/acs.jcim.0c00718>.
- (231) Song, Y.; Mao, J.; Gunner, M. R. MCCE2: Improving Protein PKa Calculations with Extensive Side Chain Rotamer Sampling. *J. Comput. Chem.* **2009**, *32*. <https://doi.org/10.1002/jcc.21222>.
- (232) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), 537–541. <https://doi.org/10.1093/nar/gks375>.
- (233) Lai, P.-K.; Gallegos, A.; Mody, N.; Sathish, H. A.; Trout, B. L. Machine Learning Prediction of Antibody Aggregation and Viscosity for High Concentration Formulation Development of Protein Therapeutics. *MAbs* **2022**, *14* (1). <https://doi.org/10.1080/19420862.2022.2026208>.
- (234) Yadav, S.; Laue, T. M.; Kalonia, D. S.; Singh, S. N.; Shire, S. J. The Influence of Charge Distribution on Self-Association and Viscosity Behavior of Monoclonal Antibody Solutions. *Mol. Pharm.* **2012**, *9* (4), 791–802. <https://doi.org/10.1021/mp200566k>.
- (235) Saito, S.; Hasegawa, J.; Kobayashi, N.; Tomitsuka, T.; Uchiyama, S.; Fukui, K. Effects of Ionic Strength and Sugars on the Aggregation Propensity of Monoclonal Antibodies: Influence of Colloidal and Conformational Stabilities. *Pharm. Res.* **2013**, *30* (5), 1263–1280. <https://doi.org/10.1007/s11095-012-0965-4>.
- (236) Svilenov, H. L.; Kulakova, A.; Zalar, M.; Golovanov, A. P.; Winter, G. Orthogonal Techniques to Study the Effect of PH, Sucrose and Arginine Salts

- on Monoclonal Antibody Physical Stability and Aggregation during Long-Term Storage. *J. Pharm. Sci.* **2019**, *109* (1), 584–594. <https://doi.org/10.1016/j.xphs.2019.10.065>.
- (237) Kim, N.; Remmele, R. L.; Liu, D.; Razinkov, V. I.; Fernandez, E. J.; Roberts, C. J. 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. <https://doi.org/10.1016/j.bpc.2012.12.004>.
- (238) Svilenov, H.; Winter, G. Rapid Sample-Saving Biophysical Characterisation and Long-Term Storage Stability of Liquid Interferon Alpha2a Formulations: Is There a Correlation? *Int. J. Pharm.* **2019**, *562* (December 2018), 42–50. <https://doi.org/10.1016/j.ijpharm.2019.03.025>.
- (239) Perico, N.; Purtell, J.; Dillon, T. M.; Ricci, M. S. Conformational Implications of an Inversed PH-Dependent Antibody Aggregation. *J. Pharm. Sci.* **2009**, *98* (9), 3031–3042. <https://doi.org/10.1002/jps.21539>.
- (240) Mahler, H. C.; Friess, W.; Grauschopf, U.; Kiese, S. Protein Aggregation: Pathways, Induction Factors and Analysis. *J. Pharm. Sci.* **2009**, *98* (9), 2909–2934. <https://doi.org/10.1002/jps.21566>.
- (241) Nicoud, L.; Jagielski, J.; Pfister, D.; Lazzari, S.; Massant, J.; Lattuada, M.; Morbidelli, M. Kinetics of Monoclonal Antibody Aggregation from Dilute toward Concentrated Conditions. *J. Phys. Chem. B* **2016**, *120* (13), 3267–3280. <https://doi.org/10.1021/acs.jpcc.5b11791>.
- (242) Costanzo, J. A.; O'Brien, C. J.; Tiller, K.; Tamargo, E.; Robinson, A. S.; Roberts, C. J.; Fernandez, E. J. Conformational Stability as a Design Target to Control Protein Aggregation. *Protein Eng. Des. Sel.* **2014**, *27* (5), 157–167. <https://doi.org/10.1093/protein/gzu008>.
- (243) Roberts, C. J.; Das, T. K.; Sahin, E. Predicting Solution Aggregation Rates for Therapeutic Proteins: Approaches and Challenges. *Int. J. Pharm.* **2011**, *418* (2), 318–333. <https://doi.org/10.1016/j.ijpharm.2011.03.064>.
- (244) Banks, D. D.; Latypov, R. F.; Ketchem, R. R.; Woodard, J.; Scavezze, J. L.; Siska, C. C.; Razinkov, V. I. Native-State Solubility and Transfer Free Energy as Predictive Tools for Selecting Excipients to Include in Protein Formulation Development Studies. *J. Pharm. Sci.* **2012**, *101* (8), 2720–2732. <https://doi.org/10.1002/jps.23219>.
- (245) Kayser, V.; Chennamsetty, N.; Voynov, V.; Helk, B.; Trout, B. L. Conformational Stability and Aggregation of Therapeutic Monoclonal Antibodies Studied with ANS and Thioflavin T Binding. *MAbs* **2011**, *3* (4), 408–411. <https://doi.org/10.4161/mabs.3.4.15677>.
- (246) King, A. C.; Woods, M.; Liu, W.; Lu, Z.; Gill, D.; Krebs, M. R. H. High-Throughput Measurement, Correlation Analysis, and Machine-Learning Predictions for PH and Thermal Stabilities of Pfizer-Generated Antibodies. *Protein Sci.* **2011**, *20* (9), 1546–1557. <https://doi.org/10.1002/pro.680>.
- (247) Cloutier, T. K.; Sudrik, C.; Mody, N.; Sathish, H. A.; Trout, B. L. Machine

- Learning Models of Antibody-Excipient Preferential Interactions for Use in Computational Formulation Design. *Mol. Pharm.* **2020**, *17* (9), 3589–3599. <https://doi.org/10.1021/acs.molpharmaceut.0c00629>.
- (248) Gentiluomo, L.; Roessner, D.; Frieß, W. Application of Machine Learning to Predict Monomer Retention of Therapeutic Proteins after Long Term Storage. *Int. J. Pharm.* **2020**, *577* (January), 119039. <https://doi.org/10.1016/j.ijpharm.2020.119039>.
- (249) Narayanan, H.; Dingfelder, F.; Butté, A.; Lorenzen, N.; Sokolov, M.; Arosio, P. Machine Learning for Biologics: Opportunities for Protein Engineering, Developability, and Formulation. *Trends Pharmacol. Sci.* **2021**, *42* (3), 151–165. <https://doi.org/10.1016/j.tips.2020.12.004>.
- (250) Lai, P.-K.; Swan, J. W.; Trout, B. L. Calculation of Therapeutic Antibody Viscosity with Coarse-Grained Models, Hydrodynamic Calculations and Machine Learning-Based Parameters. *MABs* **2021**, *13* (1), 1–10. <https://doi.org/10.1080/19420862.2021.1907882>.
- (251) Lai, P. K.; Fernando, A.; Cloutier, T. K.; Gokarn, Y.; Zhang, J.; Schwenger, W.; Chari, R.; Calero-Rubio, C.; Trout, B. L. Machine Learning Applied to Determine the Molecular Descriptors Responsible for the Viscosity Behavior of Concentrated Therapeutic Antibodies. *Mol. Pharm.* **2021**, *18* (3), 1167–1175. <https://doi.org/10.1021/acs.molpharmaceut.0c01073>.
- (252) Kaur, H. Stability Testing in Monoclonal Antibodies. *Crit. Rev. Biotechnol.* **2021**, *41* (5), 692–714. <https://doi.org/10.1080/07388551.2021.1874281>.
- (253) Wälchli, R.; Fanizzi, F.; Massant, J.; Arosio, P. Relationship of PEG-Induced Precipitation with Protein-Protein Interactions and Aggregation Rates of High Concentration MAb Formulations at 5 °C. *Eur. J. Pharm. Biopharm.* **2020**, *151* (March), 53–60. <https://doi.org/10.1016/j.ejpb.2020.03.011>.
- (254) Helbig, C.; Ammann, G.; Menzen, T.; Friess, W.; Wuchner, K.; Hawe, A. Backgrounded Membrane Imaging (BMI) for High-Throughput Characterization of Subvisible Particles During Biopharmaceutical Drug Product Development. *J. Pharm. Sci.* **2020**, *109* (1), 264–276. <https://doi.org/10.1016/j.xphs.2019.03.024>.
- (255) Grabarek, A. D.; Bozic, U.; Rousel, J.; Menzen, T.; Kranz, W.; Wuchner, K.; Jiskoot, W.; Hawe, A. What Makes Polysorbate Functional? Impact of Polysorbate 80 Grade and Quality on IgG Stability During Mechanical Stress. *J. Pharm. Sci.* **2020**, *109* (1), 871–880. <https://doi.org/10.1016/j.xphs.2019.10.015>.
- (256) Murphy, M. I.; Bruque, M.; Hanford, A.; Trayton, I.; Handali, M.; Leissa, J.; Hasige, S.; Day, K.; Patel, S. Qualitative High-Throughput Analysis of Subvisible Particles in Biological Formulations Using Backgrounded Membrane Imaging. *J. Pharm. Sci.* **2022**, *111* (6), 1605–1613. <https://doi.org/10.1016/j.xphs.2022.03.010>.
- (257) Vargas, S. K.; Eskafi, A.; Carter, E.; Ciaccio, N. A Comparison of Background Membrane Imaging versus Flow Technologies for Subvisible Particle Analysis

- of Biologics. *Int. J. Pharm.* **2020**, *578* (August 2019), 1–9.
<https://doi.org/10.1016/j.ijpharm.2020.119072>.
- (258) Andrews, J. M.; Roberts, C. J. Non-Native Aggregation of α -Chymotrypsinogen Occurs through Nucleation and Growth with Competing Nucleus Sizes and Negative Activation Energies. *Biochemistry* **2007**, *46* (25), 7558–7571. <https://doi.org/10.1021/bi700296f>.
- (259) Garber, E.; Demarest, S. J. A Broad Range of Fab Stabilities within a Host of Therapeutic IgGs. *Biochem. Biophys. Res. Commun.* **2007**, *355* (3), 751–757. <https://doi.org/10.1016/j.bbrc.2007.02.042>.
- (260) Sahin, E.; Roberts, C. J. Size-Exclusion Chromatography with Multi-Angle Light Scattering for Elucidating Protein Aggregation Mechanisms. In *Therapeutic Proteins: Methods and Protocols*; Voynov, V., Caravella, J. A., Eds.; Springer International Publishing, 2012; Vol. 899, pp 403–423. <https://doi.org/10.1007/978-1-61779-921-1>.
- (261) Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67* (2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- (262) González-González, O.; Ramirez, I. O.; Ramirez, B. I.; O’Connell, P.; Ballesteros, M. P.; Torrado, J. J.; Serrano, D. R. Drug Stability: ICH versus Accelerated Predictive Stability Studies. *Pharmaceutics* **2022**, *14* (11). <https://doi.org/10.3390/pharmaceutics14112324>.
- (263) Wang, W.; Roberts, C. J. Non-Arrhenius Protein Aggregation. *AAPS J.* **2013**, *15* (3), 840–851. <https://doi.org/10.1208/s12248-013-9485-3>.
- (264) Rosa, M.; Roberts, C. J.; Rodrigues, M. A. Connecting High-Temperature and Low-Temperature Protein Stability and Aggregation. *PLoS One* **2017**, *12* (5). <https://doi.org/10.1371/journal.pone.0176748>.
- (265) Roberts, C. J.; Nesta, D. P.; Kim, N. Effects of Temperature and Osmolytes on Competing Degradation Routes for an IgG1 Antibody. *J. Pharm. Sci.* **2013**, *102* (10), 3556–3566. <https://doi.org/10.1002/jps.23668>.
- (266) Andrews, J. M.; Weiss IV, W. F.; Roberts, C. J. Nucleation, Growth, and Activation Energies for Seeded and Unseeded Aggregation of α -Chymotrypsinogen A. *Biochemistry* **2008**, *47* (8), 2397–2403. <https://doi.org/10.1021/bi7019244>.
- (267) Brummitt, R. K.; Nesta, D. P.; Roberts, C. J. Predicting Accelerated Aggregation Rates for Monoclonal Antibody Formulations, and Challenges for Low-Temperature Predictions. *J. Pharm. Sci.* **2011**, *100* (10), 4234–4243. <https://doi.org/10.1002/jps.22633>.
- (268) Roberts, C. J.; Darrington, R. T.; Whitley, M. B. Irreversible Aggregation of Recombinant Bovine Granulocyte-Colony Stimulating Factor (BG-CSF) and Implications for Predicting Protein Shelf Life. *J. Pharm. Sci.* **2003**, *92* (5), 1095–1111. <https://doi.org/10.1002/jps.10377>.
- (269) Bunc, M.; Hadži, S.; Graf, C.; Bončina, M.; Lah, J. Aggregation Time Machine: A Platform for the Prediction and Optimization of Long-Term Antibody

- Stability Using Short-Term Kinetic Analysis. *J. Med. Chem.* **2022**, *65* (3), 2623–2632. <https://doi.org/10.1021/acs.jmedchem.1c02010>.
- (270) Kamerzell, T. J.; Li, M.; Arora, S.; Ji, J. A.; Wang, Y. J. The Relative Rate of Immunoglobulin Gamma 1 Fragmentation. *J. Pharm. Sci.* **2011**, *100* (4), 1341–1349. <https://doi.org/10.1002/jps.22389>.
- (271) Tian, X.; Langkilde, A. E.; Thorolfsson, M.; Rasmussen, H. B.; Vestergaard, B. Small-Angle X-Ray Scattering Screening Complements Conventional Biophysical Analysis: Comparative Structural and Biophysical Analysis of Monoclonal Antibodies IgG1, IgG2, and IgG4. *J. Pharm. Sci.* **2014**, *103* (6), 1701–1710. <https://doi.org/10.1002/jps.23964>.
- (272) Yadav, S.; Shire, S. J.; Kalonia, D. S. Viscosity Behavior of High-Concentration Monoclonal Antibody Solutions: Correlation with Interaction Parameter and Electroviscous Effects. *J. Pharm. Sci.* **2012**, *101* (3), 998–1011. <https://doi.org/10.1002/jps.22831>.
- (273) Wu, H.; Kroe-Barrett, R.; Singh, S.; Robinson, A. S.; Roberts, C. J. Competing Aggregation Pathways for Monoclonal Antibodies. *FEBS Lett.* **2014**, *588* (6), 936–941. <https://doi.org/10.1016/j.febslet.2014.01.051>.
- (274) Ionescu, R. M.; Vlasak, J.; Price, C.; Kirchmeier, M. Contribution of Variable Domains to the Stability of Humanized IgG1 Monoclonal Antibodies. *J. Pharm. Sci.* **2008**, *97* (4), 1414–1426. <https://doi.org/10.1002/jps.21104>.
- (275) Sahin, E.; Grillo, A. O.; Perkins, M. D.; Roberts, C. J. Comparative Effects of PH and Ionic Strength on Protein-Protein Interactions, Unfolding, and Aggregation for IgG1 Antibodies. *J. Pharm. Sci.* **2010**, *99* (12), 4830–4848. <https://doi.org/10.1002/jps.22198>.
- (276) Fukada, H.; Tsumoto, K.; Arakawa, T.; Ejima, D. Long-Term Stability and Reversible Thermal Unfolding of Antibody Structure at Low PH: Case Study. *J. Pharm. Sci.* **2018**, *107* (11), 2965–2967. <https://doi.org/10.1016/j.xphs.2018.07.001>.
- (277) He, F.; Woods, C. E.; Becker, G. W.; Narhi, L. O.; Razinkov, V. I. High-Throughput Assessment of Thermal and Colloidal Stability Parameters for Monoclonal Antibody Formulations. *J. Pharm. Sci.* **2011**, *100* (12), 5126–5141. <https://doi.org/10.1002/jps.22712>.
- (278) Sorret, L. L.; DeWinter, M. A.; Schwartz, D. K.; Randolph, T. W. Challenges in Predicting Protein-Protein Interactions from Measurements of Molecular Diffusivity. *Biophys. J.* **2016**, *111* (9), 1831–1842. <https://doi.org/10.1016/J.BPJ.2016.09.018>.
- (279) Brader, M. L.; Estey, T.; Bai, S.; Alston, R. W.; Lucas, K. K.; Lantz, S.; Landsman, P.; Maloney, K. M. Examination of Thermal Unfolding and Aggregation Profiles of a Series of Developable Therapeutic Monoclonal Antibodies. *Mol. Pharm.* **2015**, *12* (4), 1005–1017. <https://doi.org/10.1021/mp400666b>.
- (280) Li, Y.; Ogunnaike, B. A.; Roberts, C. J. Multi-Variate Approach to Global Protein Aggregation Behavior and Kinetics: Effects of PH, NaCl, and

- Temperature for α -Chymotrypsinogen A. *J. Pharm. Sci.* **2010**, *99* (2), 645–662. <https://doi.org/10.1002/jps.21869>.
- (281) Trikeriotis, M.; Akbulatov, S.; Esposito, U.; Anastasiou, A.; Leszczyszyn, O. I. Analytical Workflows to Unlock Predictive Power in Biotherapeutic Developability. *Pharm. Res.* **2023**, *40* (2), 487–500. <https://doi.org/10.1007/s11095-022-03448-y>.
- (282) Barnett, G. V.; Balakrishnan, G.; Chennamsetty, N.; Hoffman, L.; Bongers, J.; Tao, L.; Huang, Y.; Slaney, T.; Das, T. K.; Leone, A.; Kar, S. R. Probing the Tryptophan Environment in Therapeutic Proteins: Implications for Higher Order Structure on Tryptophan Oxidation. *J. Pharm. Sci.* **2019**, *108* (6), 1944–1952. <https://doi.org/10.1016/j.xphs.2018.12.027>.
- (283) Shen, V. K.; Siderius, D. W. Elucidating the Effects of Adsorbent Flexibility on Fluid Adsorption Using Simple Models and Flat-Histogram Sampling Methods. *J. Chem. Phys.* **2014**, *140* (24). <https://doi.org/10.1063/1.4884124>.
- (284) Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I.; Krishna, R. Improving the Efficiency of the Configurational-Bias Monte Carlo Algorithm. *Mol. Phys.* **1998**, *94* (4), 727–733. <https://doi.org/10.1080/002689798167881>.
- (285) Hatch, H. W.; Siderius, D. W.; Errington, J. R.; Shen, V. K. Efficiency Comparison of Single- and Multiple-Macrostate Grand Canonical Ensemble Transition-Matrix Monte Carlo Simulations. *J. Phys Chem B* **2023**, *127*, 3041–3051. <https://doi.org/10.1021/acs.jpcc.3c00613>.
- (286) Blanco, M. A.; Hatch, H. W.; Curtis, J. E.; Shen, V. K. A Methodology to Calculate Small-Angle Scattering Profiles of Macromolecular Solutions from Molecular Simulations in the Grand-Canonical Ensemble. *J. Chem. Phys.* **2018**, *149* (8). <https://doi.org/10.1063/1.5029274>.
- (287) Hung, J. J.; Dear, B. J.; Karouta, C. A.; Chowdhury, A. A.; Godfrin, P. D.; Bollinger, J. A.; Nieto, M. P.; Wilks, L. R.; Shay, T. Y.; Ramachandran, K.; Sharma, A.; Cheung, J. K.; Truskett, T. M.; Johnston, K. P. Protein-Protein Interactions of Highly Concentrated Monoclonal Antibody Solutions via Static Light Scattering and Influence on the Viscosity. *J. Phys. Chem. B* **2019**, *123* (4), 739–755. <https://doi.org/10.1021/acs.jpcc.8b09527>.
- (288) Chowdhury, A.; Bollinger, J. A.; Dear, B. J.; Cheung, J. K.; Johnston, K. P.; Truskett, T. M. 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.* **2020**, *17* (5), 1748–1756. <https://doi.org/10.1021/acs.molpharmaceut.9b00960>.
- (289) Roche, A.; Gentiluomo, L.; Sibanda, N.; Roessner, D.; Friess, W.; Trainoff, S. P.; Curtis, R. Towards an Improved Prediction of Concentrated Antibody Solution Viscosity Using the Huggins Coefficient. *J. Colloid Interface Sci.* **2022**, *607*, 1813–1824. <https://doi.org/10.1016/j.jcis.2021.08.191>.
- (290) Meuwly, M. Reactive Molecular Dynamics: From Small Molecules to Proteins. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9* (1), 1–22. <https://doi.org/10.1002/wcms.1386>.

- (291) Johnson, J. K. Reactive Canonical Monte Carlo. In *Advances in Chemical Physics*; 1999; Vol. 105, pp 461–481.
<https://doi.org/10.1002/9780470141649.ch15>.
- (292) Zuchniarz, J.; Liu, Y.; Li, C.; Voth, G. A. Accurate pKa Calculations in Proteins with Reactive Molecular Dynamics Provide Physical Insight into the Electrostatic Origins of Their Values. *J. Phys. Chem. B* **2022**, *126* (38), 7321–7330. <https://doi.org/10.1021/acs.jpcc.2c04899>.
- (293) Rennekamp, B.; Kutzki, F.; Obarska-Kosinska, A.; Zapp, C.; Gräter, F. Hybrid Kinetic Monte Carlo/Molecular Dynamics Simulations of Bond Scissions in Proteins. *J. Chem. Theory Comput.* **2020**, *16* (1), 553–563.
<https://doi.org/10.1021/acs.jctc.9b00786>.
- (294) Wang, W. Advanced Protein Formulations. *Protein Sci.* **2015**, *24* (7), 1031–1039. <https://doi.org/10.1002/pro.2684>.
- (295) Abriata, L. A.; Spiga, E.; Peraro, M. D. Molecular Effects of Concentrated Solutes on Protein Hydration, Dynamics, and Electrostatics. *Biophys. J.* **2016**, *111* (4), 743–755. <https://doi.org/10.1016/j.bpj.2016.07.011>.
- (296) Pandey, R. B.; Jacobs, D. J.; Farmer, B. L. Preferential Binding Effects on Protein Structure and Dynamics Revealed by Coarse-Grained Monte Carlo Simulation. *J. Chem. Phys.* **2017**, *146* (19), 1–9.
<https://doi.org/10.1063/1.4983222>.
- (297) Sudrik, C. M.; Cloutier, T.; Mody, N.; Sathish, H. A.; Trout, B. L. Understanding the Role of Preferential Exclusion of Sugars and Polyols from Native State IgG1 Monoclonal Antibodies and Its Effect on Aggregation and Reversible Self-Association. *Pharm. Res.* **2019**, *36* (8), 1–12.
<https://doi.org/10.1007/s11095-019-2642-3>.
- (298) Calero-Rubio, C. Protein Interactions, Unfolding and Aggregation from Low to High Protein Concentrations via Coarse-Grained Molecular Modeling and Experimental Characterization, University of Delaware, 2017.

Appendix A

SUPPORTING INFORMATION FOR CHAPTER 2

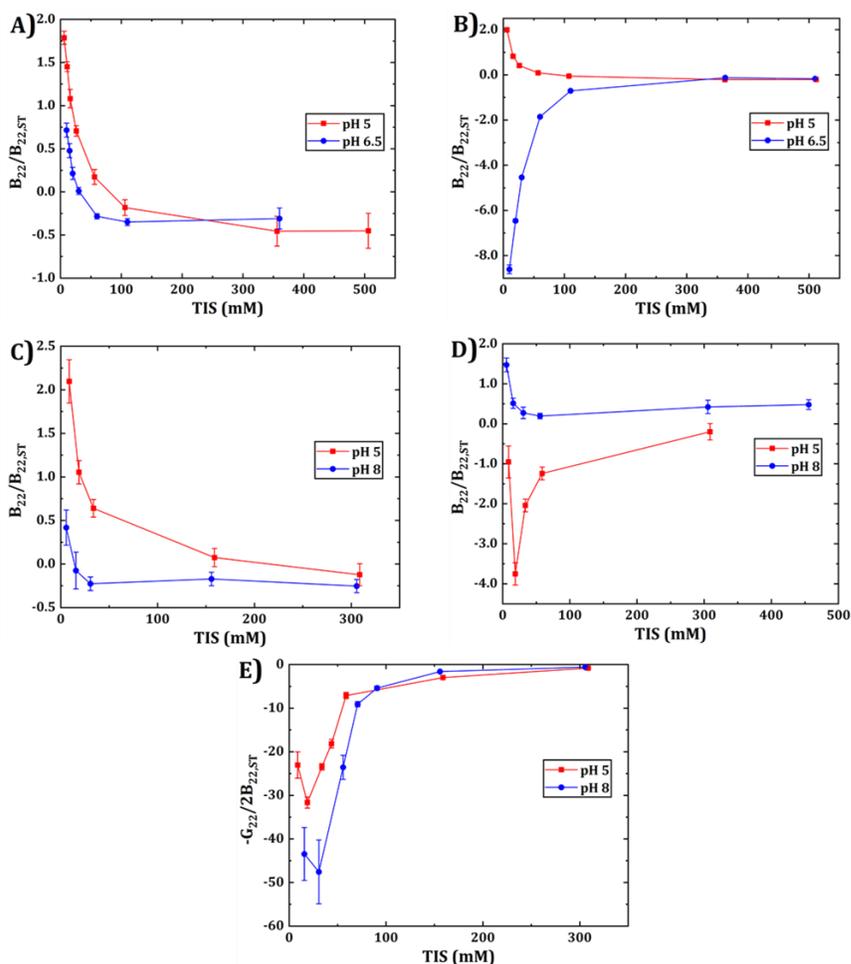


Figure A.1: Experimental $B_{22}/B_{22,ST}$ values from static light scattering for MAB 1 (A), MAB 2 (B), MAB A (C), MAB B (D) and MAB C (E).^{42,49,50} Values in panel (E) are designated as $-G_{22}/2B_{22,ST}$ instead of $B_{22}/B_{22,ST}$ because some values are so large that they indicate multi-body interactions at the experimental protein concentrations, rather than two-body interactions that are captured by B_{22} .¹⁹¹

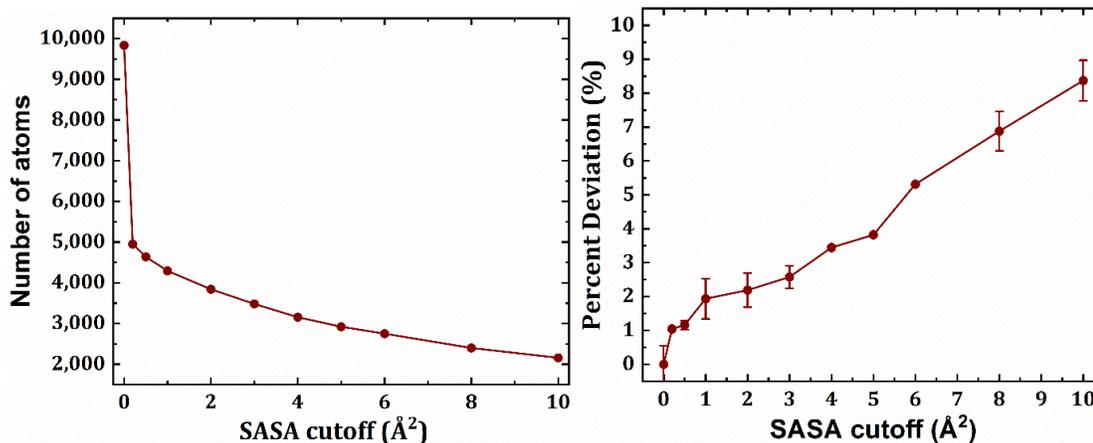


Figure A.2: Left: The number of atoms in MAb B that have SASA above a given cutoff value. 50.3% of the atoms have SASA > 0.2 Å². Right: Percent deviation of all-atom steric-only simulations of the second virial coefficient as a function of SASA. Percent deviation was defined as

$$100 * \frac{B_{22,ST}^{all-atom} - B_{22,ST}^{cutoff}}{B_{22,ST}^{all-atom}}.$$

Table A.1: Valence values for HEXA beads for all conditions. The HEXA domains are F_v (V_L + V_H), C1 (C_L + C_{H1}), C2 (C_{H2} + C_{H2}) and C3 (C_{H3} + C_{H3}), where there are two domains for both F_v and C1.^{49,50}

MAb (pH)	C3	C2	C1	F _v	Total
MAb 1 (5)	5.56	10.14	9.21	2.61	39.3
MAb 1 (6.5)	-0.40	3.64	5.08	0.31	14.0
MAb 2 (5)	5.78	2.30	4.21	5.40	27.3
MAb 2 (6.5)	-0.38	-2.88	0.08	4.04	5.0
MAb A (5)	1.56	15.5	8.01	10.2	53.5
MAb A (8)	-5.95	8.03	2.03	6.88	19.9
MAb B (5)	-2.23	6.29	7.09	1.68	21.6
MAb B (8)	-9.95	0.01	2.02	-2.12	-10.1
MAb C (5)	1.56	15.5	9.90	0.82	38.5
MAb C (8)	-5.95	8.03	4.03	-2.13	5.9

Table A.2: Valence values for DODECA beads for all conditions. The last column denotes the total net charge of the MAb at the given pH value.^{49,50}

MAB (pH)	V _H	C _{H1}	C _{H2}	C _{H3}	V _L	C _L	Total
MAB 1 (5)	1.89	6.27	5.07	2.78	0.72	2.94	39.3
MAB 1 (6.5)	0.29	4.53	1.82	-0.20	0.02	0.55	14.0
MAB 2 (5)	2.79	2.27	1.15	2.89	2.61	1.94	27.3
MAB 2 (6.5)	2.02	0.53	-1.44	-0.19	2.02	-0.45	5.0
MAB A (5)	3.79	7.18	7.75	0.78	6.42	0.83	53.5
MAB A (8)	1.94	4.02	4.01	-2.98	4.94	-1.99	19.9
MAB B (5)	-0.84	5.37	3.15	-1.11	2.53	1.72	21.6
MAB B (8)	-3.06	3.01	0.01	-4.97	0.94	-0.99	-10.1
MAB C (5)	0.25	8.18	7.75	0.78	0.57	1.72	38.5
MAB C (8)	-1.07	5.02	4.01	-2.98	-1.06	-0.99	5.87

Table A.3: Relative hydrophobicity values (ϵ_i , dimensionless) and bead diameters (σ_i , angstroms) for the 1bAA model, where i denotes the chemical identity of the amino acid. Reproduced from Blanco et al.⁷¹

residue	σ_i (Å)	ϵ_i	residue	σ_i (Å)	ϵ_i
Lys	7.03	0.00	His	6.29	0.25
Glu	6.40	0.05	Ala	5.02	0.26
Asp	5.83	0.06	Tyr	7.11	0.49
Asn	5.95	0.10	Cys	4.92	0.54
Ser	5.28	0.11	Trp	6.70	0.64
Arg	7.32	0.13	Val	6.05	0.65
Gln	6.35	0.13	Met	6.32	0.67
Pro	5.62	0.14	Ile	6.36	0.84
Thr	5.81	0.16	Phe	6.95	0.91
Gly	4.31	0.17	Leu	6.55	1.00

Table A.4: Residue indices for each domain in the DODECA and 1bC/D models. Indices start at the N-terminal amino acid for the heavy or light chain.^{69,298}

Domain	Residue Index
MAb 1 & 2 V _L	1-107
MAb 1 & 2 C _L	108-214
MAb 1 & 2 V _H	1-118
MAb 1 & 2 C _{H1}	119-234
MAb 1 & 2 C _{H2}	244-357
MAb 1 & 2 C _{H3}	358-474
MAb A & C V _L	1-111
MAb A & C C _L	112-216
MAb A & C V _H	1-121
MAb A & C C _{H1}	122-232
MAb A & C C _{H2}	233-344
MAb A & C C _{H3}	345-447
MAb B V _L	1-113
MAb B C _L	114-217
MAb B V _H	1-120
MAb B C _{H1}	121-230
MAb B C _{H2}	231-340
MAb B C _{H3}	341-444

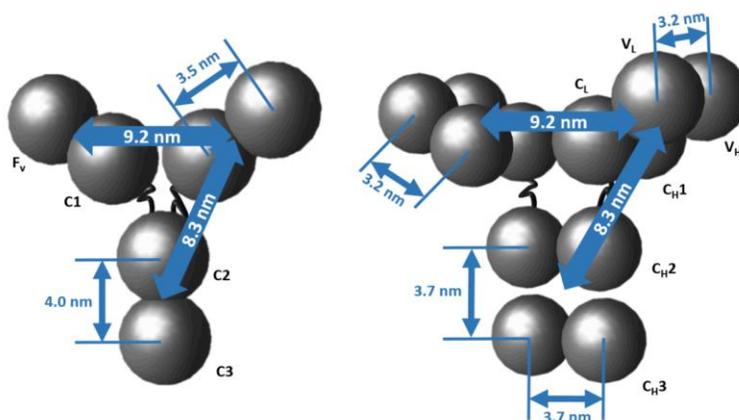


Figure A.3: Geometric features for the HEXA and DODECA models. Reproduced from Calero-Rubio et al.⁵⁰

Table A.5: Charged site diameters ($\sigma_{c,i}$, angstroms) for charged sites in a given amino acid for the 1bC/D model, where i denotes the chemical identity of the amino acid.¹⁸³ For data shown in the main text, Δ_H was 3 Å (e.g., the final bead diameter for the charged oxygen site in Glu was 6.8 Å for $\sigma_{c,i} + \Delta_H$).

Amino Acid	$\sigma_{c,i}$ (Å)
Lys	3.7
Arg	3.4
His	3.4
Glu	3.8
Asp	3.8

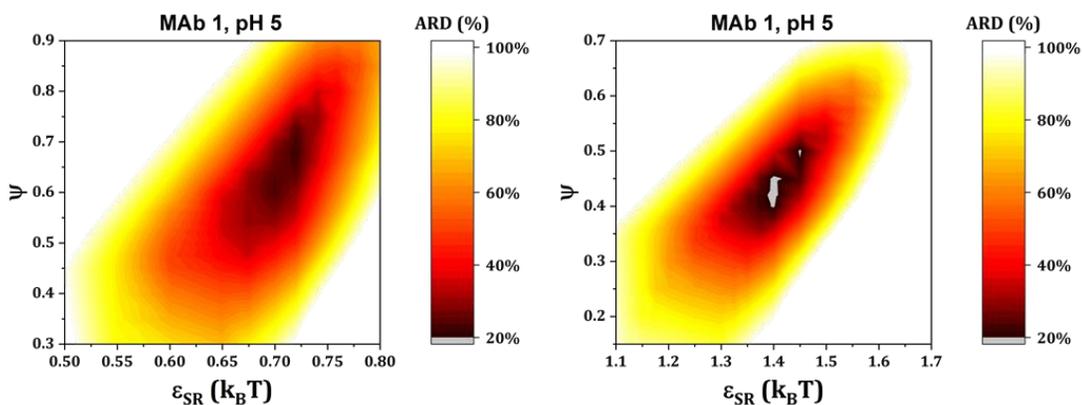


Figure A.4: DODECA (left) and HEXA (right) surface response of ARD as a function of ϵ_{SR} and ψ for MAb 1 at pH 5 as an example. Because HEXA and DODECA are much more computationally efficient, a global optimization was performed for both ϵ_{SR} and ψ which takes into account the contribution from short-ranged non-electrostatic interactions at intermediate TIS as well (e.g., 100 mM). For monotonic profiles of $B_{22}/B_{22,ST}$ vs. TIS, the difference in the resulting ϵ_{SR} values between the two procedures was negligible.

Table A.6: Domain bead diameters (σ_i , nm) for the HEXA and DODECA models (independent of MAb identity) and the 1bC/D model for each MAb.^{28,50}

Model	σ_i (nm)
HEXA	4.4
DODECA	3.5
MAb 1 1bC/D	3.1
MAb 2 1bC/D	3.35
MAb A 1bC/D	2.7
MAb B 1bC/D	2.9
MAb C 1bC/D	3.1

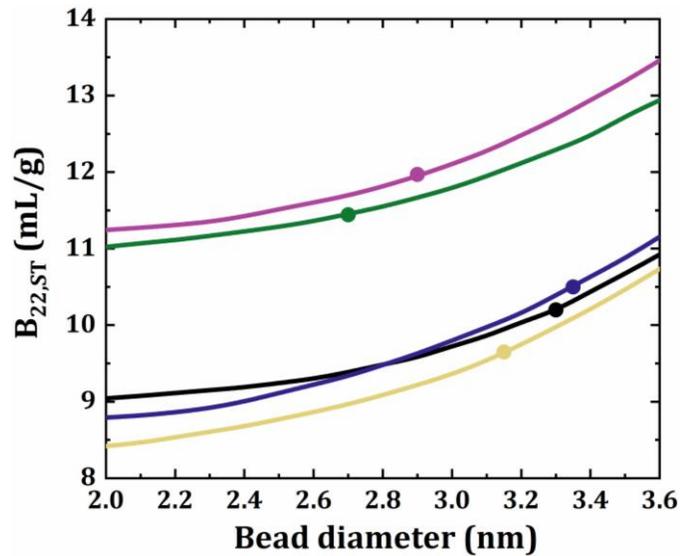


Figure A.5: $B_{22,ST}$ as a function of the domain bead diameter for MAb 1 (black), 2 (indigo), A (green), B (yellow) and C (purple) for the 1bC/D model. Circles on each curve show the bead diameter that results in the correct value of $B_{22,ST}$. The curves for MAb 1 and MAb 2 overlap due to domain beads covering more charged sites as bead diameter is increased. The effect of charged beads and non-planar structure of domain beads can also be seen in the lower value of bead diameter for all MAbs compared to 3.5 nm for DODECA.

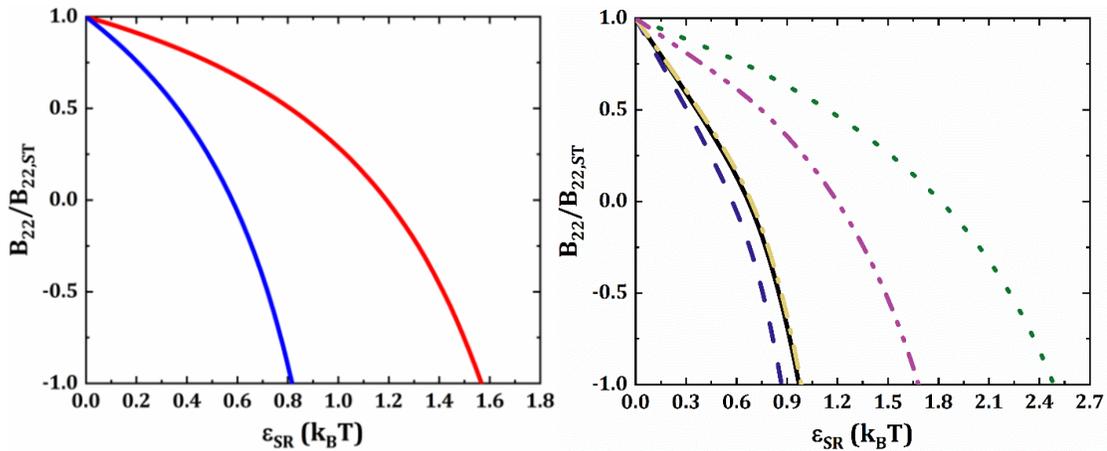
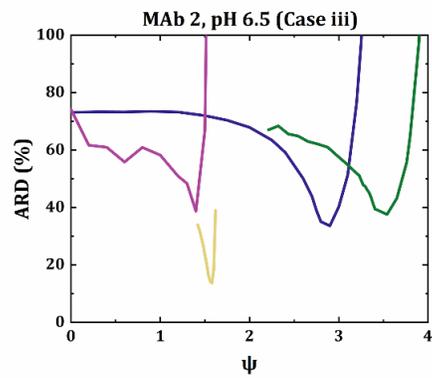
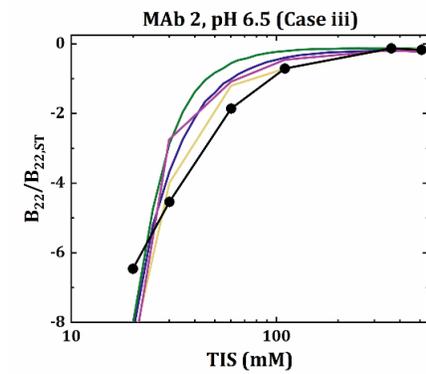
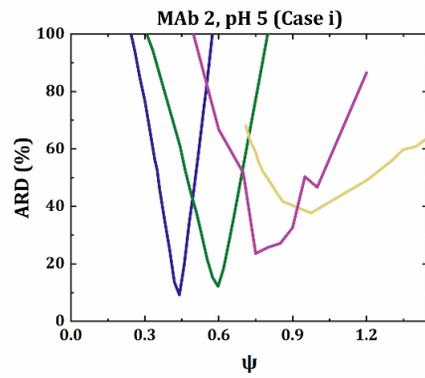
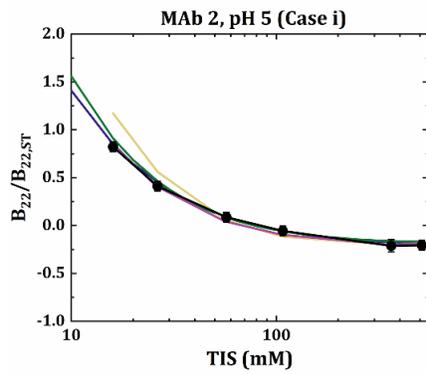
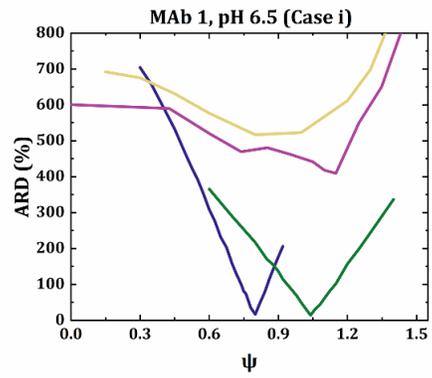
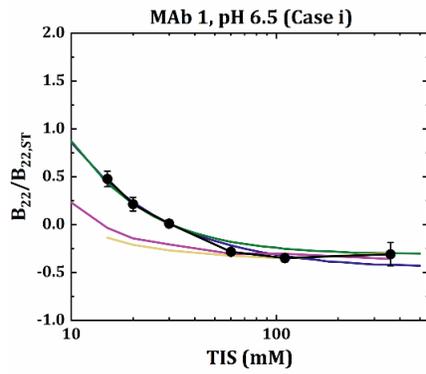
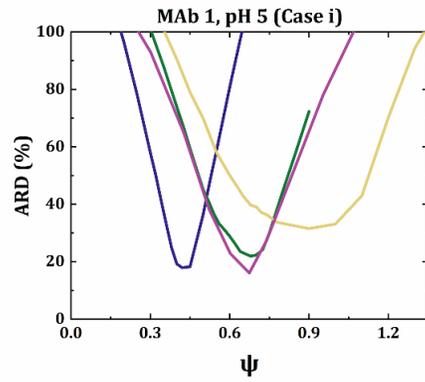
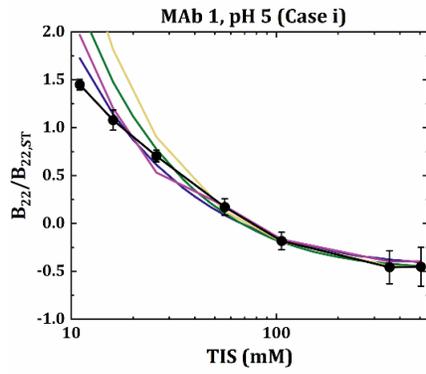
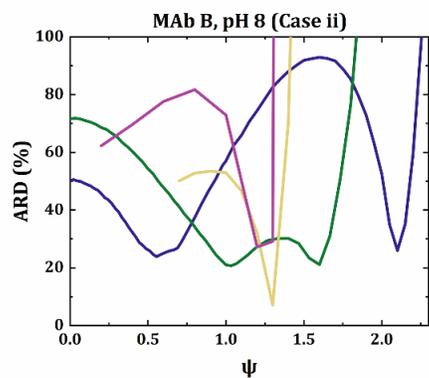
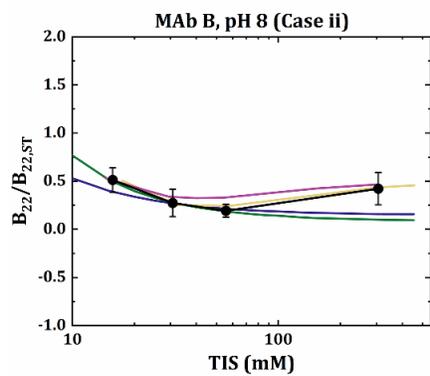
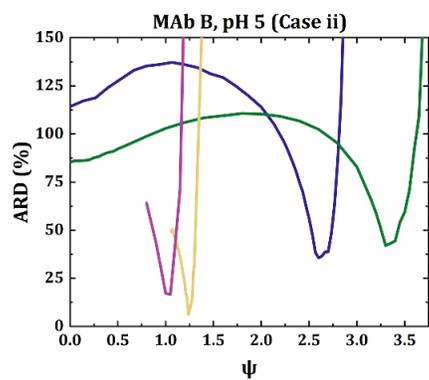
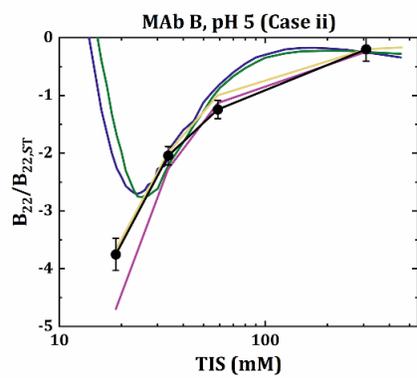
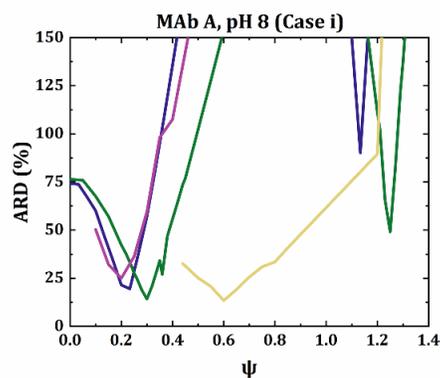
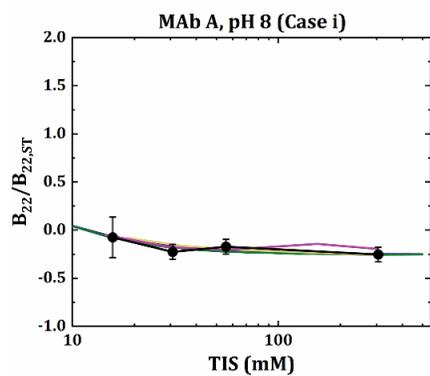
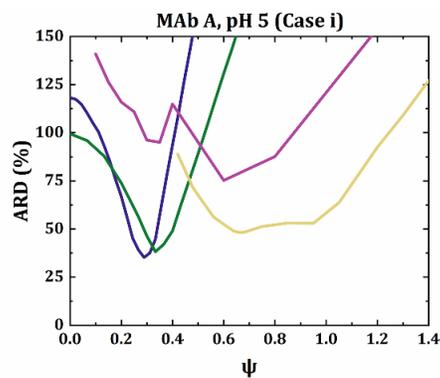
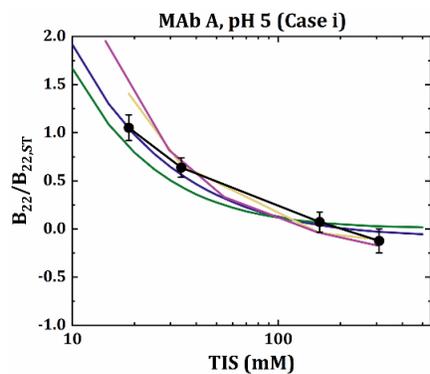


Figure A.6: $B_{22}/B_{22,ST}$ as a function of ϵ_{SR} , where electrostatic interactions are ignored (i.e., $\psi = 0$) for (left) HEXA (red) and DODECA (blue), and (right) 1bC/D for MAbs 1 (solid black), 2 (dashed indigo), A (dotted green), B (dash-dotted yellow), and C (dash-dot-dotted purple).





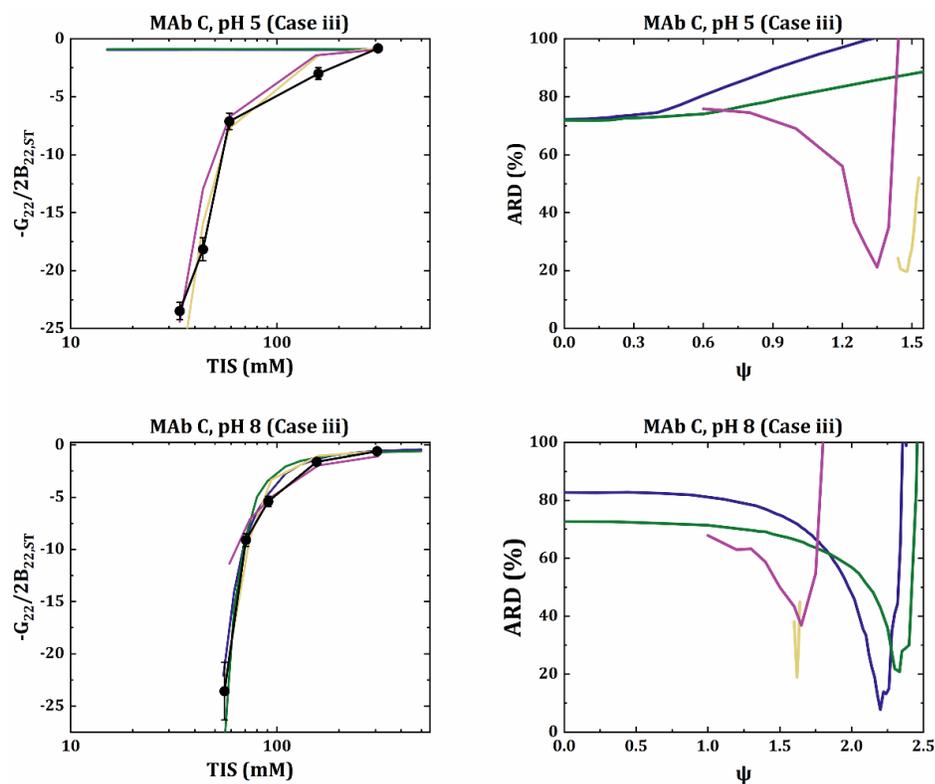
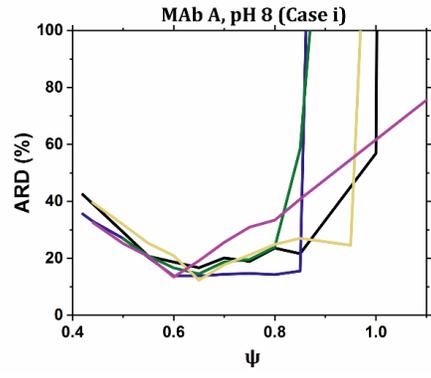
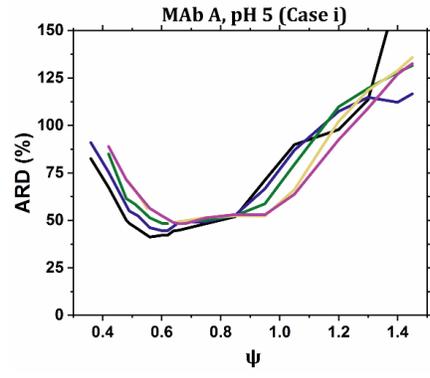
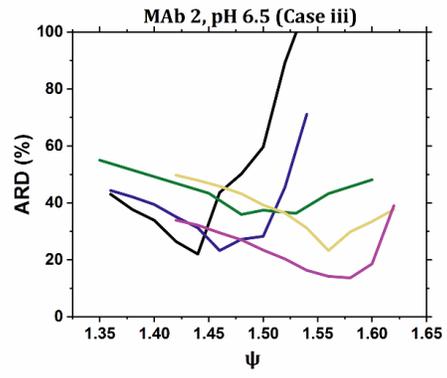
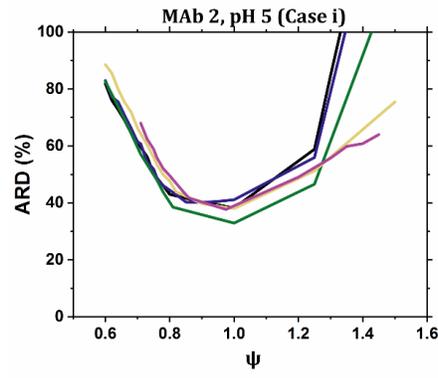
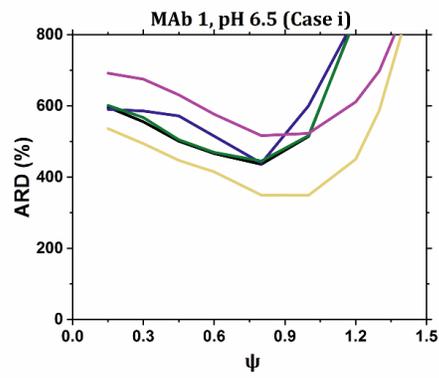
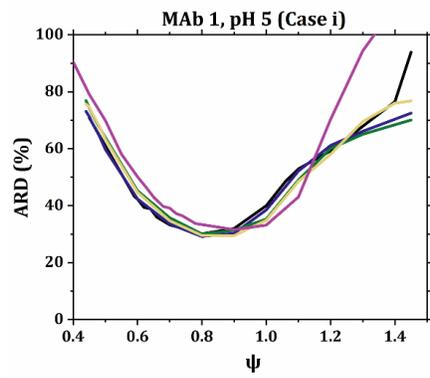


Figure A.7: Left: Experimental $B_{22}/B_{22,ST}$ measurements (black) and simulated predictions with ψ that minimized ARD for HEXA (indigo), DODECA (green), 1bC/D (yellow) and 1bAA (purple). Right: ARD for HEXA, DODECA, 1bC/D and 1bAA as a function of ψ . The qualitative experimental self-interaction behavior (i.e., Case i, ii or iii) is listed in the headers of each panel.



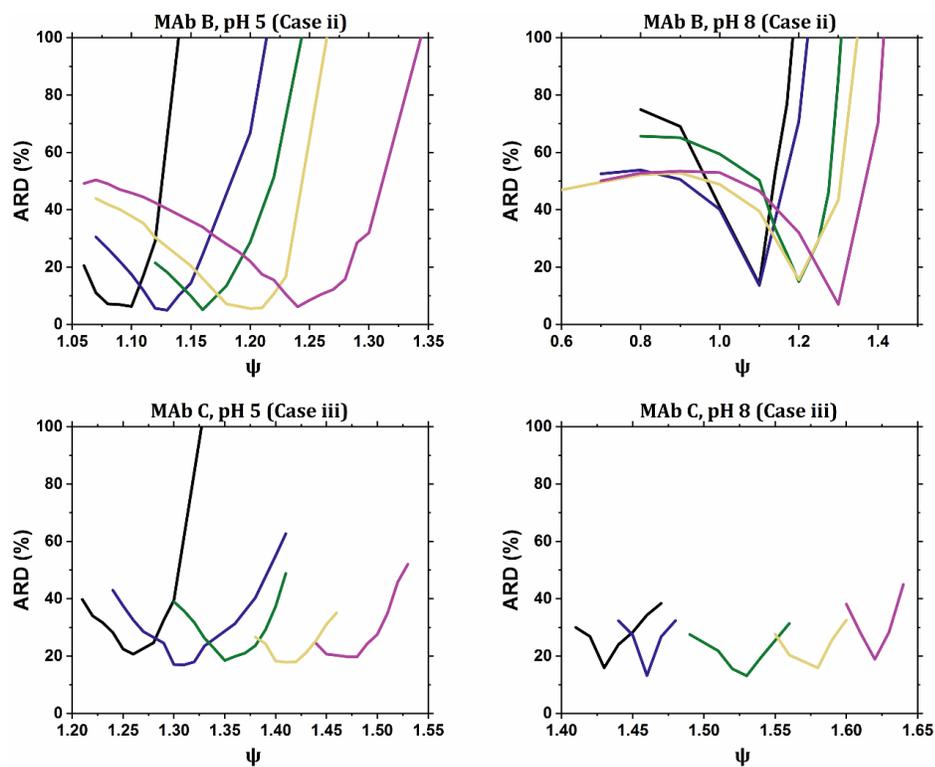


Figure A.8: ARD as a function of ψ for different values of Δ_H for the 1bC/D model, with Δ_H values: 0 Å (black), 0.75 Å (indigo), 1.5 Å (green), 2.25 Å (yellow), and 3 Å (purple). The qualitative experimental self-interaction behavior (i.e., Case i, ii or iii) is listed in the headers of each panel.

Appendix B

SUPPORTING INFORMATION FOR CHAPTER 3

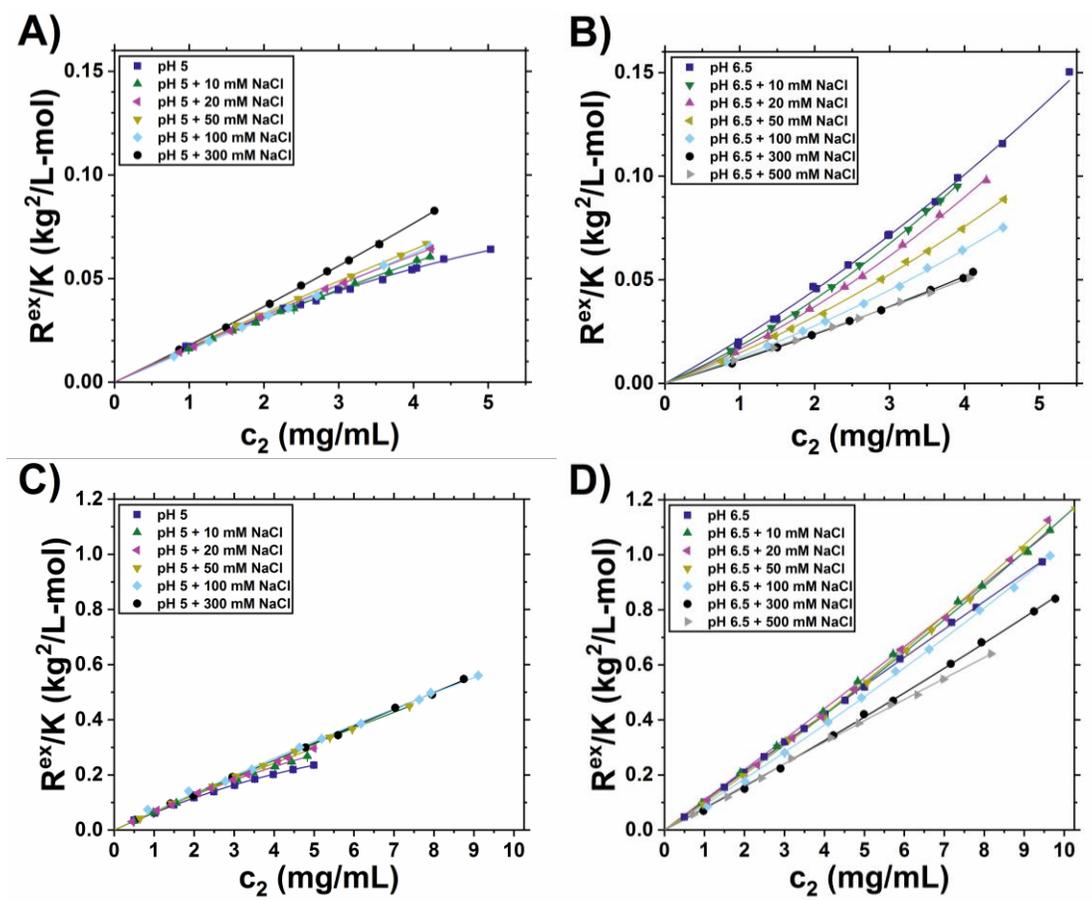


Figure B.1: Excess Rayleigh ratio (R_{90}^{ex}/K) as a function of protein concentration (c_2) for the FP protein at pH 5 (A) and pH 6.5 (B) and the monovalent Fc-fusion at pH 5 (C) and pH 6.5 (D). The curves are fits to Equation 3.1 used to calculate $M_{w,app}$.

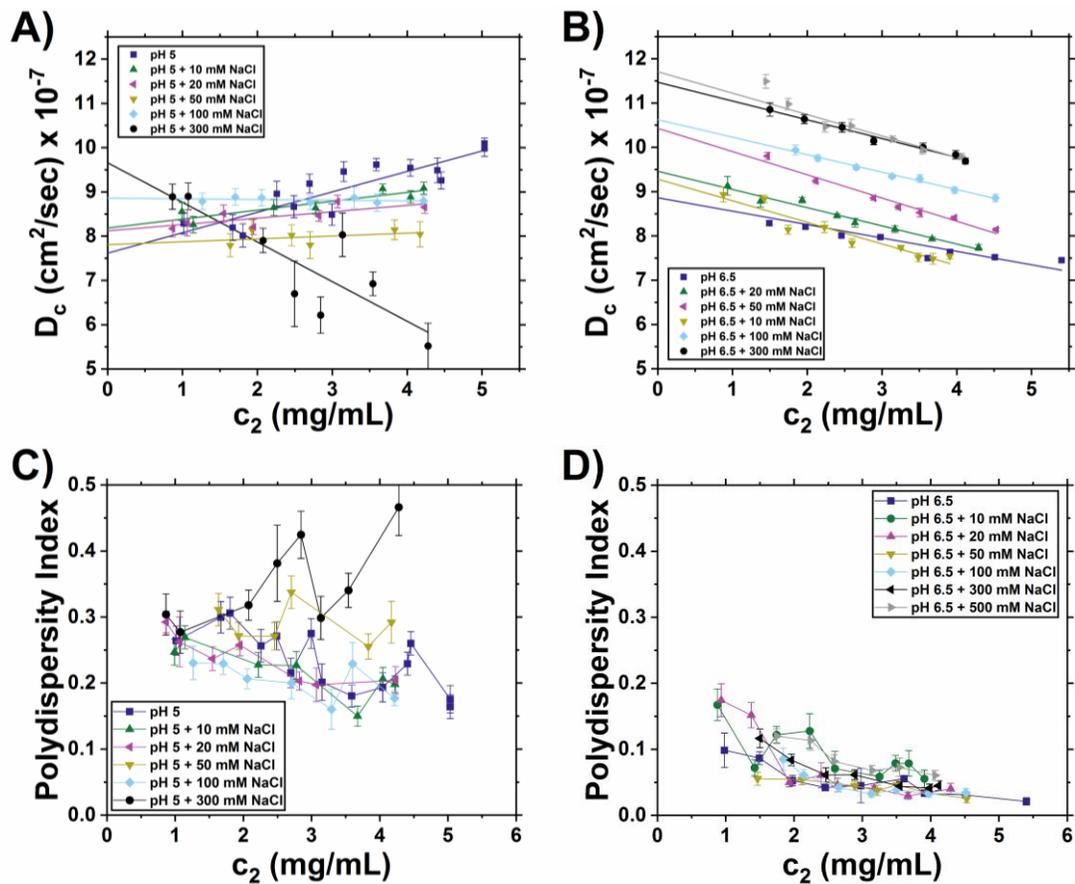


Figure B.2: DLS results for the FP protein. Collective diffusion coefficient, D_c at pH 5 (A) and pH 6.5 (B), and polydispersity index at pH 5 (C) and pH 6.5 (D) as a function of protein concentration (c_2).

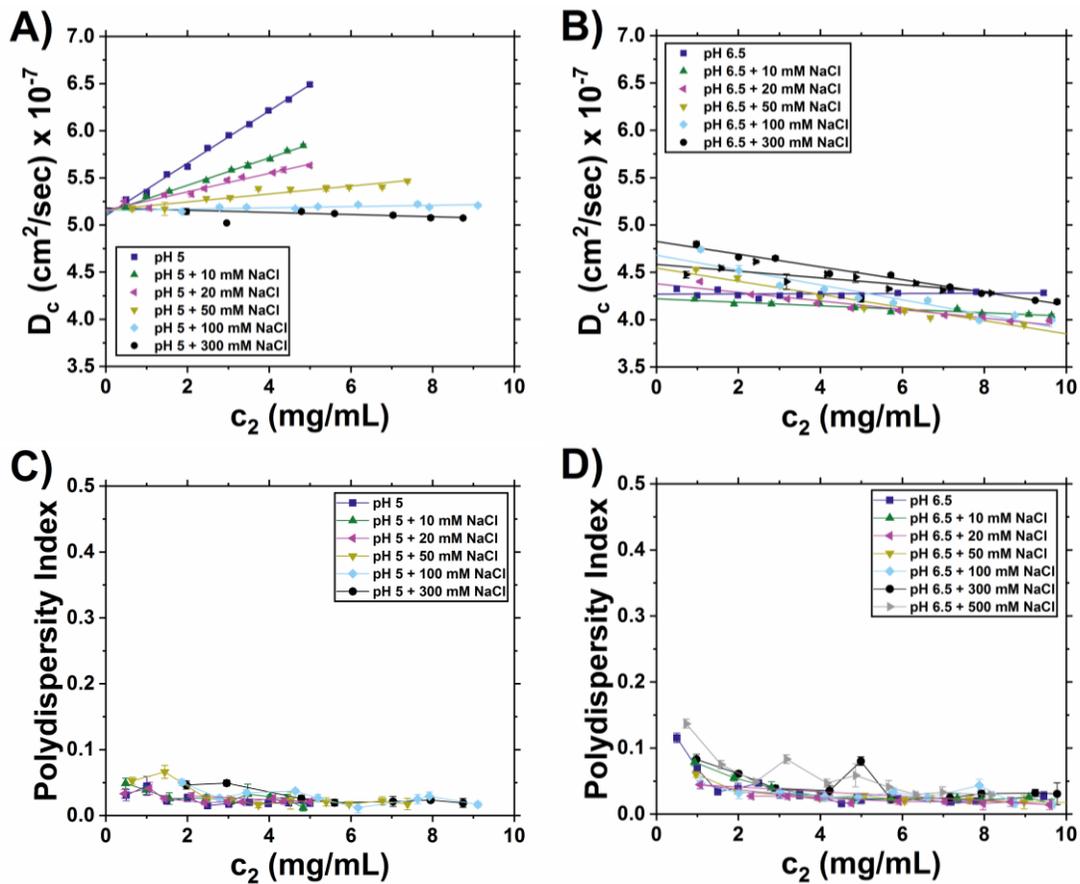


Figure B.3: DLS results for the monovalent Fc-fusion protein. Collective diffusion coefficient, D_c at pH 5 (A) and pH 6.5 (B), and polydispersity index at pH 5 (C) and pH 6.5 (D) as a function of protein concentration (c_2).

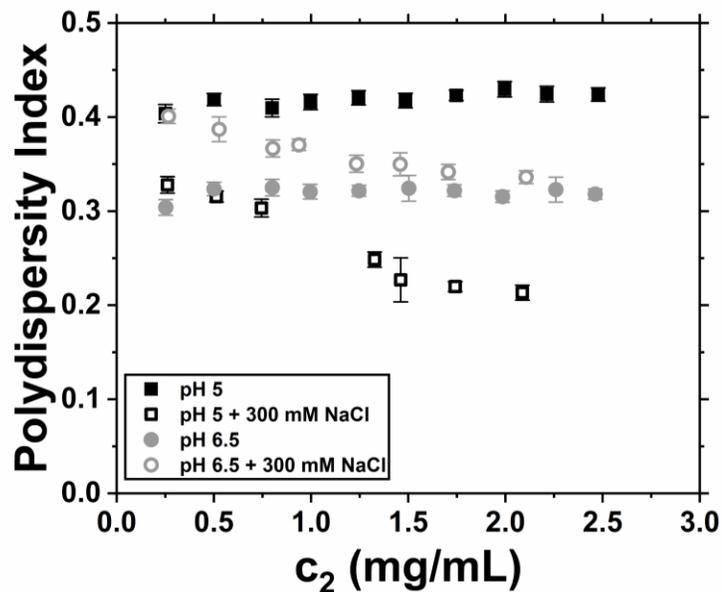


Figure B.4: Polydispersity as a function of protein concentration (c_2) for the bivalent Fc-fusion protein at pH 5 buffer-only (solid black squares), pH 5 + 300 mM NaCl (open black squares), pH 6.5 buffer-only (solid gray circles), and pH 6.5 + 300 mM NaCl (open gray circles).

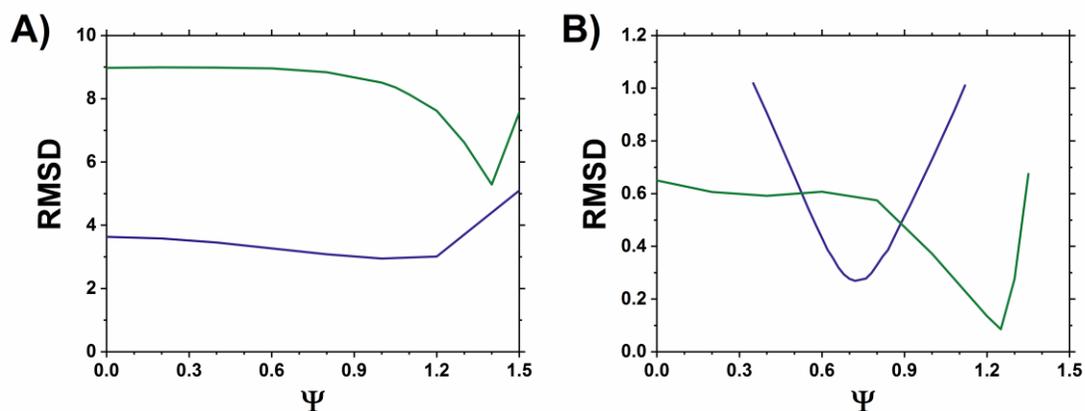


Figure B.5: RMSD as a function of ψ for the FP protein (A) and the monovalent Fc-fusion (B) at pH 5 (indigo) and pH 6.5 (green).

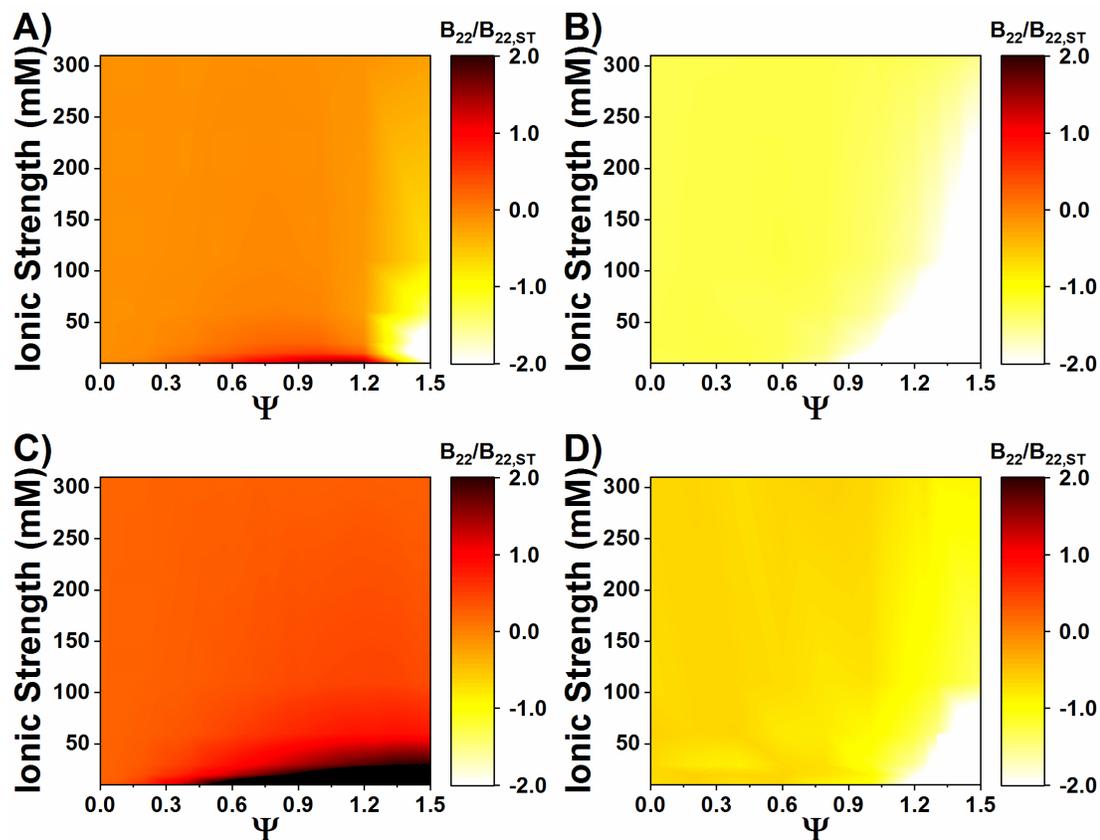


Figure B.6: Simulated contour plots of $B_{22}/B_{22,ST}$ as a function of ψ and ionic strength for the FP protein at pH 5 (A) and pH 6.5 (B), and the monovalent Fc-fusion at pH 5 (C) and pH 6.5 (D).

B.1 Preliminary computational charge swap mutations

To address the potential utility of the 1bAA simulations to predict point mutations of charged amino acids that could disrupt the pattern of electrostatic attractions present in Fc-FP cross-domain interactions (Figure 3.10B and Figure 3.10C in Chapter 3), preliminary trials of computational charge swap mutations were performed for the monovalent Fc-fusion and bivalent Fc-fusion. Charge swap mutations were performed independently for each amino acid that participated in the strongest attractive electrostatic pairwise interaction (most negative $\tilde{\phi}_{ij}$ value in Figure 3.10B or Figure 3.10C). A positively charged amino acid was replaced with Asp, and a negatively charged amino acid was replaced with Arg. For both Fc-fusion proteins, the amino acid that was swapped from positive to negative charge was located in the Fc domain, and the amino acid that was swapped from negative to positive charge was located in the fusion partner domain. For amino acids in the Fc domain, or the fusion partner domain in the bivalent Fc-fusion, both identical residues were swapped simultaneously. The same analysis as what was performed for the wild-type proteins in the main text (Figure 3.10) was performed for each charge swap mutant and results are shown in Figure B.7.

In summary, performing a charge swap mutation for the amino acids that participated in the strongest attractive electrostatic pairwise interaction reduced electrostatic attractions in some cases (i.e., less negative value of $B_{22}/B_{22,ST}$), and dramatically increased FP-FP electrostatic attractions in another (the FP negative-to-positive swap for the monovalent Fc-fusion) such that the value of $B_{22}/B_{22,ST}$ was much more negative. As there was no straightforward criterion for identifying point mutations to reduce electrostatic attractions for the Fc-fusion proteins in this work, it

is instead recommended to screen all possible charge swap or charge-to-neutral mutations, similar to recent work by Shahfar et al.⁴⁴ In that work, several MAb charge variants that were predicted to strongly reduce electrostatic attractions via similar molecular simulation methods were confirmed via experimental SLS and DLS measurements.

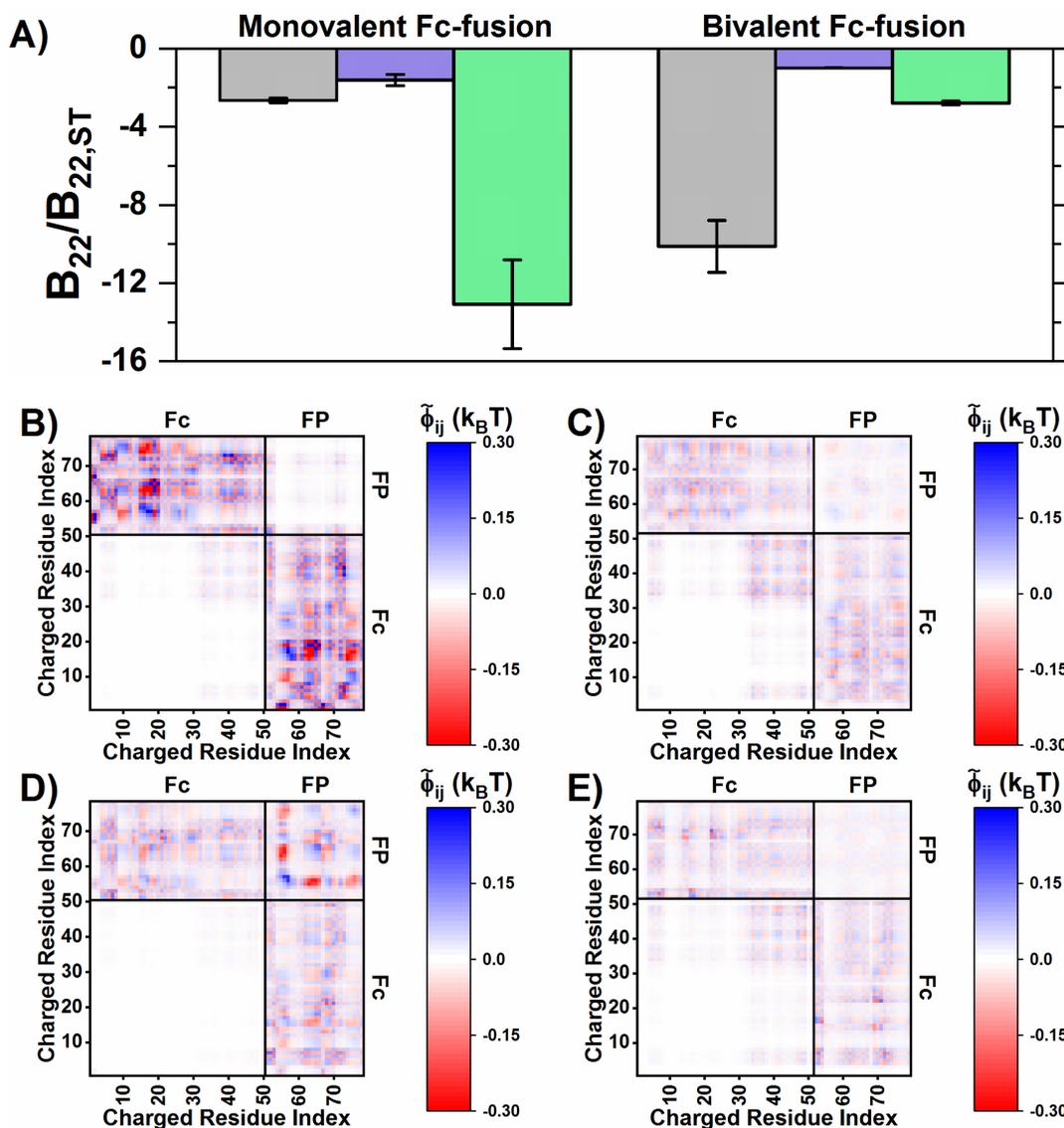


Figure B.7: A) Simulated $B_{22}/B_{22,ST}$ from 1bAA simulations of the wild-type protein (gray), the Fc positive-to-negative charge swap variant (indigo), and the fusion partner negative-to-positive charge swap variant (green) for the monovalent Fc-fusion and bivalent Fc-fusion. B-E) Corresponding Mayer-weighted average pairwise electrostatic energy ($\tilde{\phi}_{ij}$) heatmaps for the monovalent Fc-fusion (B: Fc swap, D: fusion partner swap) and bivalent Fc-fusion (C: Fc swap, E: fusion partners swap). The heatmaps are ordered based on the structural domains with fusion partner protein abbreviated as FP.

Appendix C

SUPPORTING INFORMATION FOR CHAPTER 4

Table C.1: Valence values for DODECA models.

MAb (pH)	V_H	C_{H1}	C_{H2}	C_{H3}	V_L	C_L	Net valence
MAb 1 (5)	2.88	6.19	3.05	4.78	1.71	1.94	41.1
MAb 1 (6.5)	1.29	4.53	-0.92	1.81	1.02	-0.45	14.5
MAb 2 (5)	3.79	3.44	2.04	0.99	3.61	0.94	29.6
MAb 2 (6.5)	3.02	1.54	-0.44	-2.19	3.02	-1.45	7.0
MAb 3 (5)	2.04	5.26	4.98	3.78	2.61	0.94	39.2
MAb 3 (6.5)	1.03	3.53	1.07	0.81	2.02	-1.45	14.0
MAb 4 (5)	4.75	5.37	1.15	1.89	2.75	0.94	33.7
MAb 4 (6.5)	3.28	3.53	-1.44	-1.19	2.02	-1.45	9.5
MAb 5 (5)	0.86	7.18	3.06	3.78	2.75	0.94	37.1
MAb 5 (6.5)	0.03	4.79	-0.18	0.81	2.02	-1.45	12.0
MAb 6 (5)	5.70	6.19	3.05	4.78	2.61	0.94	46.6
MAb 6 (6.5)	4.28	4.53	-0.92	1.81	2.02	-1.45	20.5

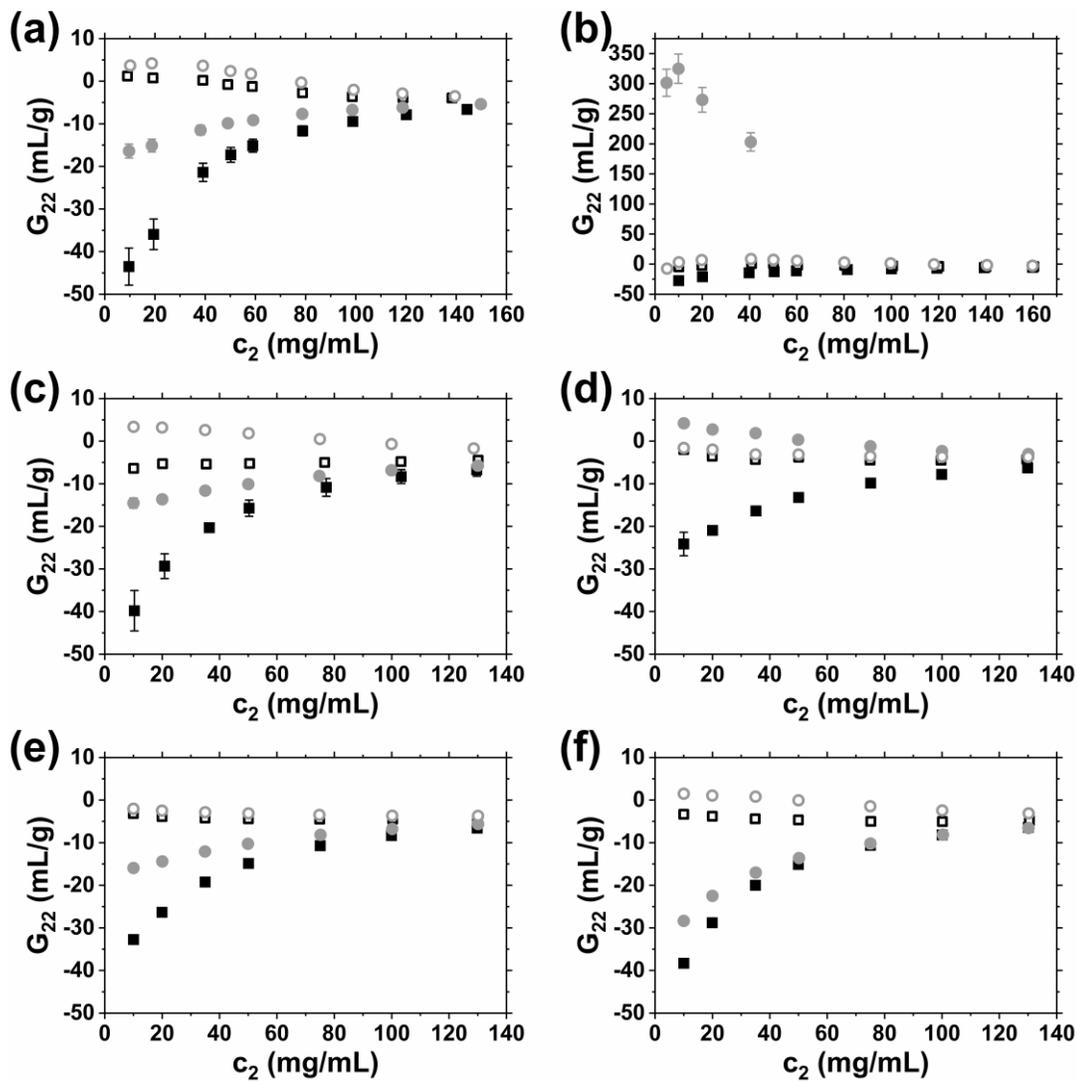


Figure C.1: G_{22} as a function of c_2 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (f) for the formulations shown in Table 4.1: pH 5 at low ionic strength (filled black squares), pH 5 at intermediate ionic strength (open black squares), pH 6.5 at low ionic strength (filled gray circles), and pH 6.5 at intermediate ionic strength (open gray circles).^{49,50} The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

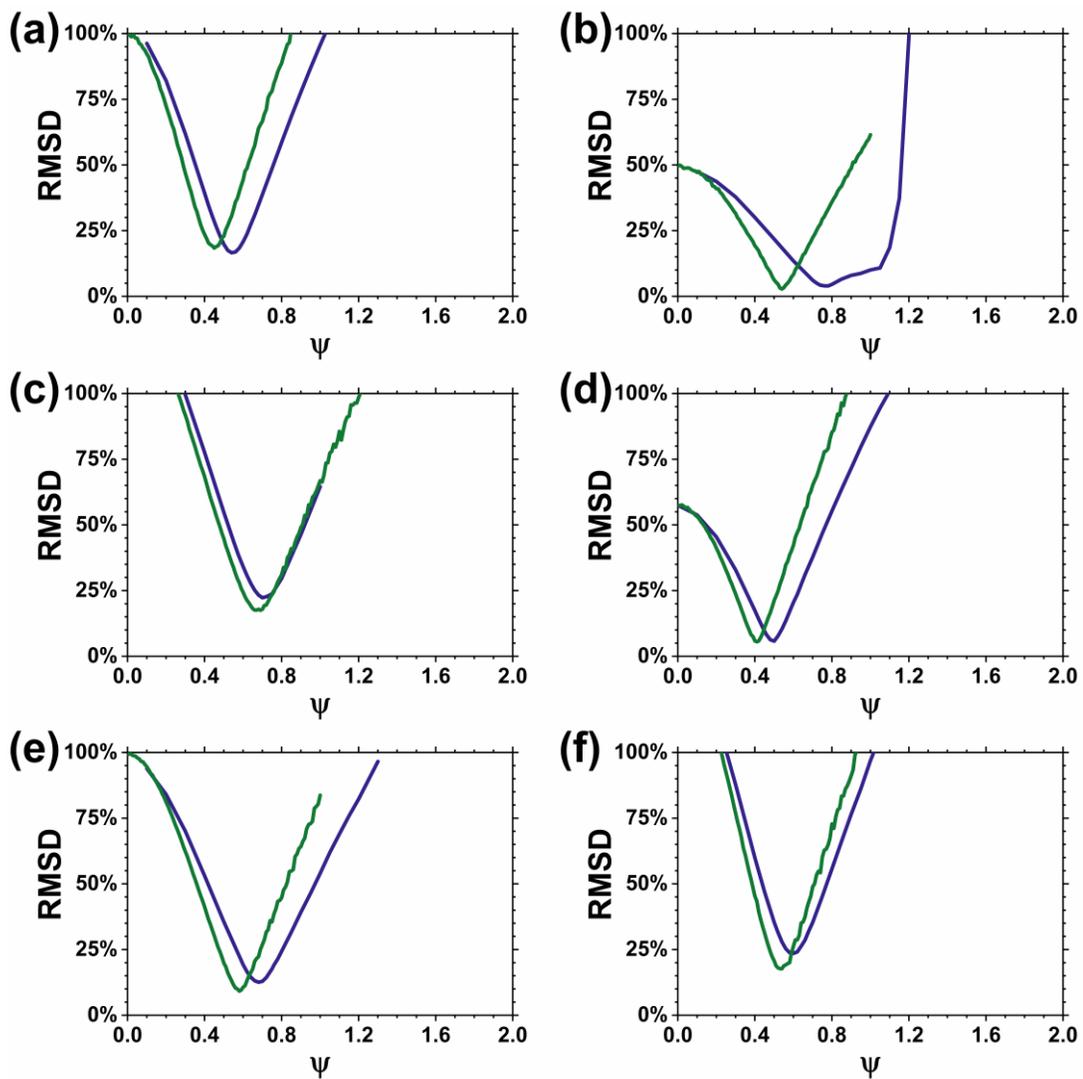


Figure C.2: RMSD as a function of ψ from MSOS simulations of $B_{22}/B_{22,ST}$ at pH 5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (e). DODECA results are green lines, and 1bC/D results are indigo lines.

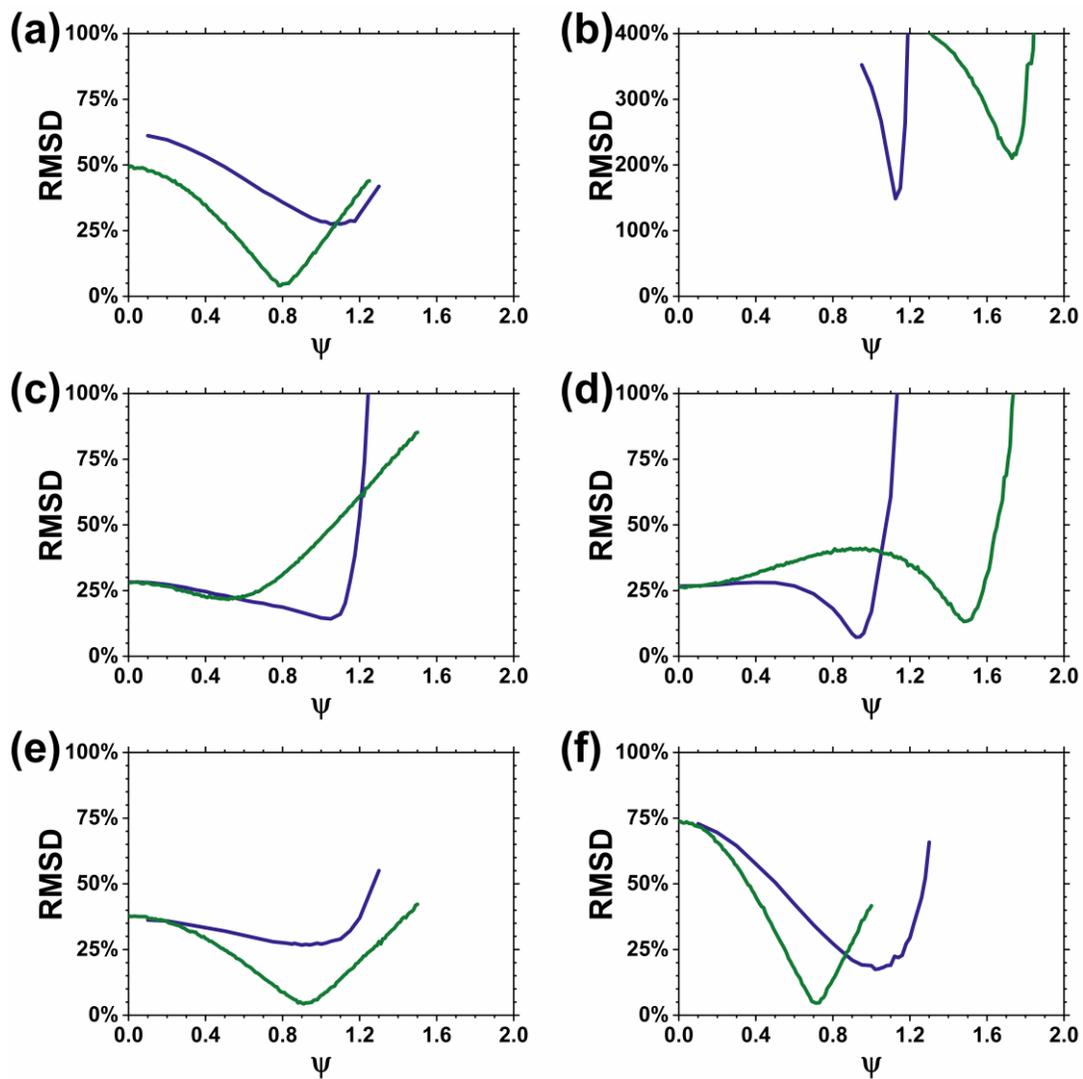


Figure C.3: RMSD as a function of ψ from MSOS simulations of $B_{22}/B_{22,ST}$ at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (e). DODECA results are green lines, and 1bC/D results are indigo lines. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

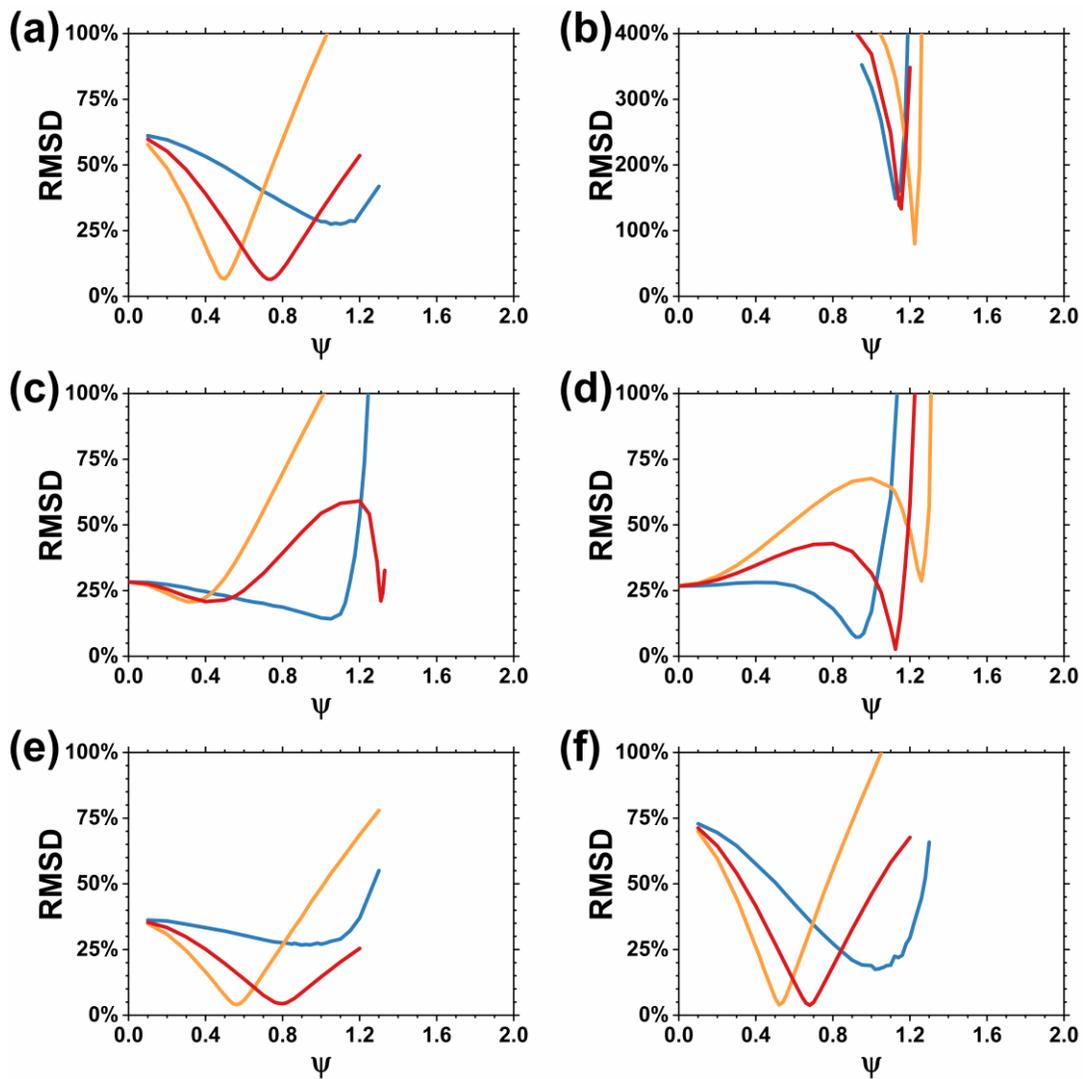


Figure C.4: RMSD as a function of ψ from MSOS simulations of $B_{22}/B_{22,ST}$ at pH 6.5 for MAb 1 (a), MAb 2 (b), MAb 3 (c), MAb 4 (d), MAb 5 (e), and MAb 6 (e). Static nominal results are blue lines, static PROPKA results are orange lines, and fluctuating nominal results are red lines. The y-axis scale for MAb 2 (panel b) is extended to fit all of the data.

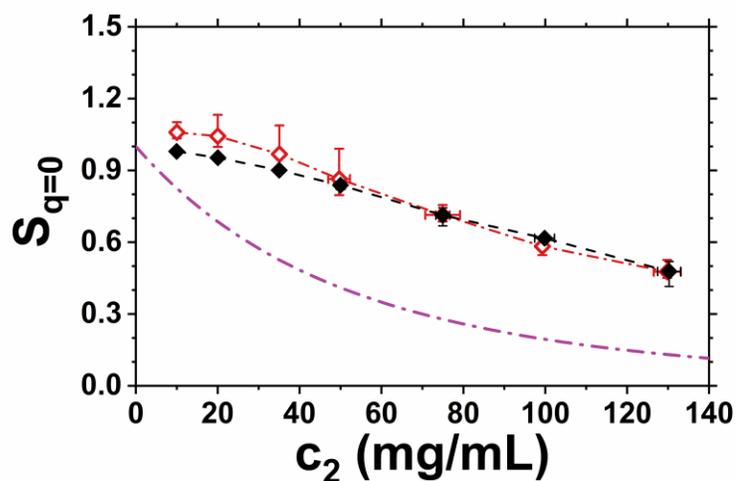


Figure C.5: High- c_2 predictions of $S_{q=0}$ as a function of MAb concentration (c_2) at pH 6.5 for MAb 4 with the fluctuating nominal charge representation method. Low ionic strength results (reproduced from Figure 4.14) are shown as open red diamonds with dash-dotted lines and intermediate ionic strength results are shown as filled black diamonds with dashed lines. Lines connecting the predictions are guides to the eye and match the formatting from Figure 4.14. Error bars correspond to uncertainty of the ψ parameter (see Table 4.3). The purple dash-dotted line is a steric-only equation of state.⁵⁰

Appendix D

SUPPORTING INFORMATION FOR CHAPTER 5

Table D.1 Aggregation rate calculation case by formulation^α

MAb	Form.	Accelerated			30 °C			Refrigerated		
		10 mg/mL	35 mg/mL	130 mg/mL	10 mg/mL	35 mg/mL	130 mg/mL	10 mg/mL	35 mg/mL	130 mg/mL
MAb 3	pH 5 low IS	2	2	1	2	2	2	N/A	N/A	N/A
	pH 5 high IS	2	1	1	2	2	2	N/A	N/A	N/A
	pH 6.5 low IS	2	2	2	2	2	2	N/A	N/A	2
	pH 6.5 high IS	2	2	2	2	2	2	N/A	N/A	2
MAb 4	pH 5 low IS	1	1	1	2	2	1	N/A	N/A	2
	pH 5 high IS	1	1	1	2	1	1	N/A	N/A	2
	pH 6.5 low IS	1	1	1	2	2	1	N/A	N/A	1
	pH 6.5 high IS	1	1	1	2	2	1	N/A	N/A	1
MAb 5	pH 5 low IS	2	2	2	N/A	N/A	2	N/A	N/A	2
	pH 5 high IS	2	2	2	2	2	2	N/A	N/A	2
	pH 6.5 low IS	2	2	2	N/A	2	2	N/A	N/A	1
	pH 6.5 high IS	2	2	2	N/A	2	2	N/A	N/A	1
MAb 6	pH 5 low IS	2	2	2	N/A	N/A	2	N/A	N/A	2
	pH 5 high IS	2	2	2	2	2	2	N/A	N/A	2
	pH 6.5 low IS	2	2	2	N/A	2	2	N/A	N/A	2
	pH 6.5 high IS	2	2	2	N/A	2	2	N/A	N/A	2

^α Case 1 (via monomer fraction) is shown in green and with the number 1, case 2 (via aggregate fraction) is shown in purple and with the number 2, and cases with aggregation rates below the measurement noise floor are shown in gray and with “N/A”.

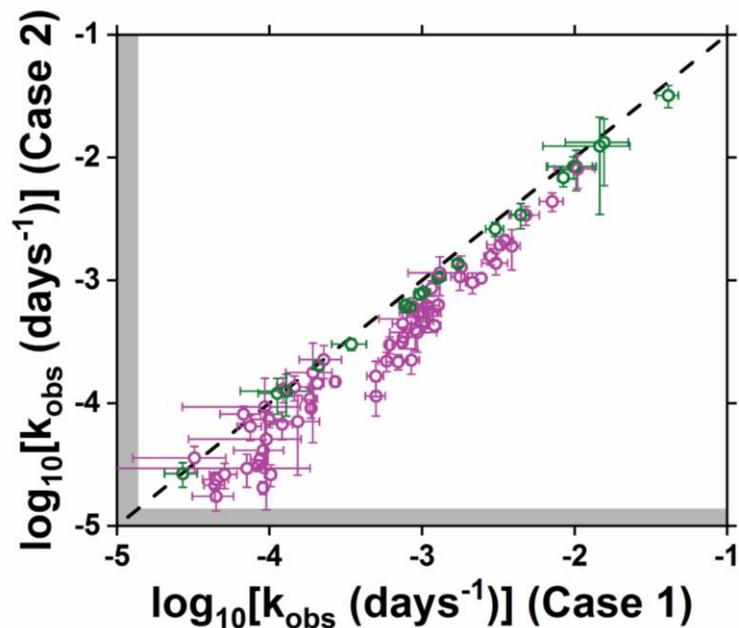


Figure D.1: Parity plot of k_{obs} calculated by change in monomer concentration (case 1; x-axis) and change in aggregate peak area (case 2; y-axis). Formulations where case 1 was chosen are shown in green, and formulations where case 2 was chosen are shown in purple, matching the color scheme in Table D.1. The gray shaded region for $\log_{10}[k_{obs}(\text{days}^{-1})] < -4.86$ represents the measurement noise floor.

D.1 Feature screening

A larger set of features (see Table D.1.1) were filtered to the set that are considered in the ML models in Chapter 5. The total solution ionic strength (IS) has an exponential relationship with the Debye screening effect²⁹, so the input feature was $\ln(\text{IS})$. The features were sorted into subgroups based on the phenomenon that they describe, and for each subgroup, correlation matrices (R values for each pair of features) were calculated with the standardized variables. Features that were highly correlated ($|\text{R}| > 0.7$) were condensed based on which feature most directly quantifies the relevant phenomenon. For features from stability studies, there was an additional stipulation that features could not contain overlapping information (e.g., if an Arrhenius extrapolated k_{obs} value was selected, the k_{obs} values used in that extrapolation were not used). The correlation matrices and filtered features are shown by subgroup below, followed by the final set and related correlation matrices.

Table D.1.1: Initial set of input features that were considered for ENR models

Feature	Source	Phenomenon
$\ln(\text{IS})$	Formulation	Debye screening effect
$ \text{pH} - \text{pI} $	Formulation and cIEF	Net valence at a given pH
Net valence (nominal pKa values)	Formulation and sequence	Net valence at a given pH
Net valence (PROPKA pKa values)	Formulation, sequence, and computational	Net valence at a given pH
Debye length (λ_{DE}) x nominal valence	Formulation and sequence	Net charge at a given pH (i.e., screened electrostatic self-interactions)
Debye length (λ_{DE}) x PROPKA valence	Formulation, sequence, and computational	Net charge at a given pH (i.e., screened electrostatic self-interactions)
$T_{m,onset} - T_{inc}$	DSC	Conformational stability
$T_{m,app1} - T_{inc}$	DSC	Conformational stability
$T_{m,app2} - T_{inc}$	DSC	Conformational stability
B_{22}^*	SLS	Net self-interactions (2-body)
$B_{22,NE}^*$	SLS	Non-electrostatic net self- interactions (2-body)
ΔB_{22}^* (= $B_{22}^* - B_{22,NE}^*$)	SLS	Electrostatic net self- interactions (2-body)
G_{22}	SLS	Net self-interactions (at c_2)
k_D	DLS	Net self-interactions (2-body)
D_c	DLS	Net self-interactions (at c_2)

Features derived from other stability studies (accelerated)		
$k_{obs,10}$ (accelerated; 10 mg/mL)	Accelerated stability studies	Aggregation rate at lower c_2
$k_{obs,35}$ (accelerated; 35 mg/mL)	Accelerated stability studies	Aggregation rate at lower c_2
$k_{obs,c_2 ex}$ (130 mg/mL; extrapolated from 10 and 35 mg/mL)	Accelerated stability studies	c_2 dependence of aggregation rate
Features derived from other stability studies (30 °C)		
$k_{obs,acc}$ (accelerated; 130 mg/mL) ^α	Accelerated stability studies	Aggregation rate at higher T_{inc}
$k_{obs,10}$ (30 °C; 10 mg/mL)	30 °C stability studies	Aggregation rate at lower c_2
$k_{obs,35}$ (30 °C; 35 mg/mL)	30 °C stability studies	Aggregation rate at lower c_2
Features derived from other stability studies (refrigerated)		
$k_{obs,acc}$ (accelerated; 130 mg/mL) ^α	Accelerated stability studies	Aggregation rate at higher T_{inc}
$k_{obs,30 °C}$ (30 °C; 130 mg/mL)	30 °C stability studies	Aggregation rate at higher T_{inc}
$k_{obs,Arr ex}$ (130 mg/mL; Arrhenius extrapolated) ^β	Accelerated and 30 °C stability studies	Aggregation rate based on assumed Arrhenius behavior

^α Aggregation rates at 45 °C were extrapolated to 50 °C using Arrhenius extrapolation with the rate at 30 °C

^β Aggregation rate at refrigerated conditions was estimated by Arrhenius extrapolation with aggregation rates at the 30 °C and accelerated T_{inc} conditions

D.1.1 Net valence at a given pH

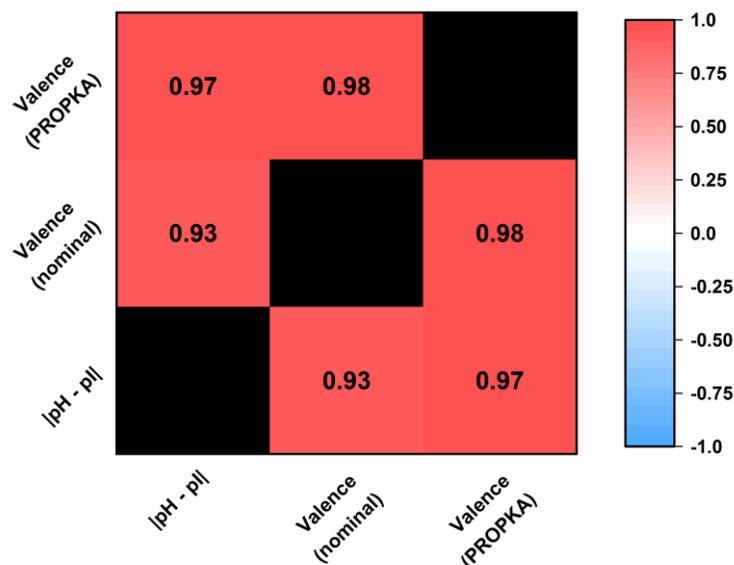


Figure D.1.1.1: Correlation coefficients for features in the net valence at a given pH subgroup.

Table D.1.1.1: Features in the net valence at a given pH subgroup and the feature(s) that were selected for use in ENR models

Feature	Source	Selected feature(s)	Justification
pH – pI	Formulation and cIEF	pH – pI	Explicit experimental measurement; no assumptions about individual pK _a values
Net valence (nominal pK _a values)	Formulation and sequence		
Net valence (PROPKA pK _a values)	Formulation, sequence, and computational		

D.1.2 Net self-interactions

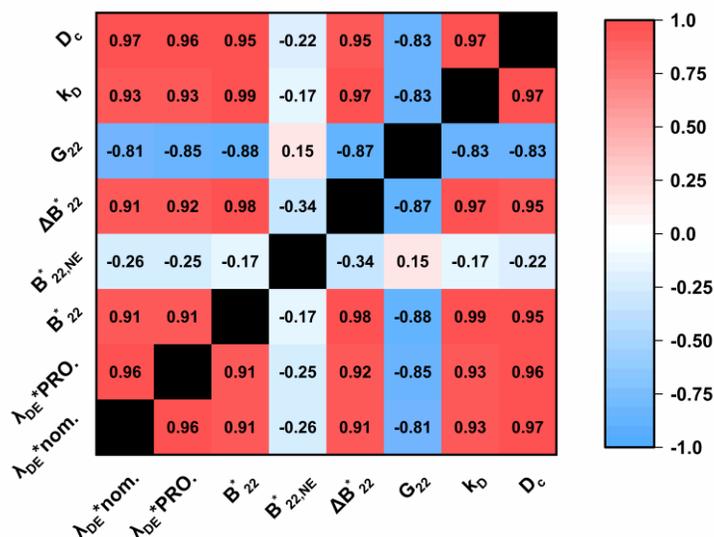


Figure D.1.2.1: Correlation coefficients for features in the net self-interactions subgroup. This figure is largely the same as what is shown in Figure 5.18 but is included here for completeness.

Table D.1.2.1: Features in the net self-interactions subgroup and the feature(s) that were selected for use in ENR models

Feature	Source	Selected feature(s)	Justification
Debye length (λ_{DE}) x nominal valence	Formulation and sequence	$B_{22,NE}^*$	Describes non-electrostatic self-interactions Not highly correlated with the other features
Debye length (λ_{DE}) x PROPKA valence	Formulation, sequence, and computational		
B_{22}^*	SLS		
$B_{22,NE}^*$	SLS		

ΔB_{22}^* ($= B_{22}^* - B_{22,NE}^*$)	SLS	ΔB_{22}^* ($= B_{22}^* - B_{22,NE}^*$)	Describes electrostatic self-interactions Directly comparable to $B_{22,NE}^*$
G_{22}	SLS		
k_D	DLS		
D_c	DLS		

D.1.3 Conformational stability

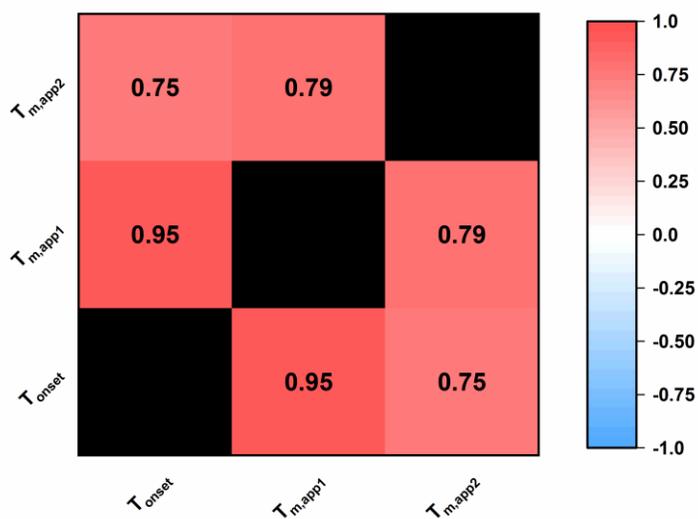


Figure D.1.3.1: Correlation coefficients for features in the conformational stability subgroup.

Table D.1.3.1: Features in the conformational stability subgroup and the feature(s) that were selected for use in ENR models

Feature ^α	Source	Selected feature(s)	Justification
$T_{m,onset} - T_{inc}$	DSC	$T_{m,app1} - T_{inc}$	$T_{m,app2}$ is sparse (some formulations only had one $T_{m,app}$) $T_{m,onset}$ is ambiguous based on how it is defined
$T_{m,app1} - T_{inc}$	DSC		
$T_{m,app2} - T_{inc}$	DSC		

^α These results are equivalent to if the features were not subtracted by T_{inc}

D.1.4 Features derived from other stability studies (accelerated)

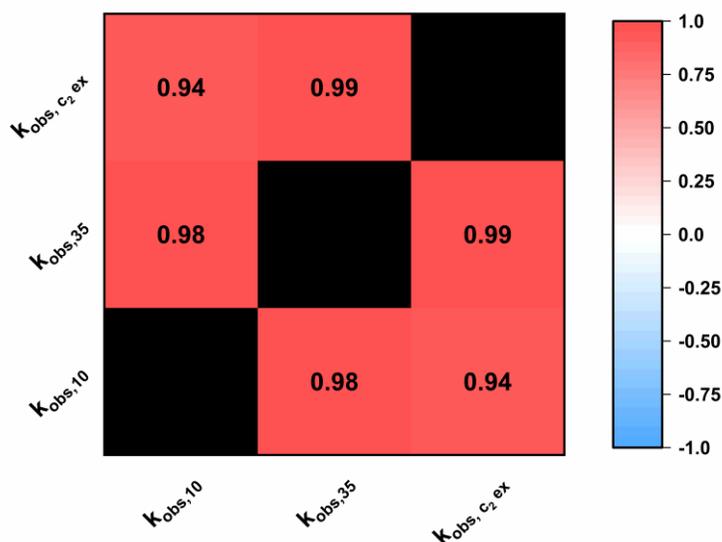


Figure D.1.4.1: Correlation coefficients for features in the stability studies (accelerated) subgroup.

Table D.1.4.1: Features in the stability studies (accelerated) subgroup and the feature(s) that were selected for use in ENR models

Feature	Selected feature(s)	Justification
$k_{obs,10}$ (accelerated; 10 mg/mL)	$k_{obs,c_2 ex}$ (130 mg/mL; extrapolated from 10 and 35 mg/mL)	$\log_{10}(k_{obs})$ scales differently with c_2 for different formulations; the other two inputs do not account for that
$k_{obs,35}$ (accelerated; 35 mg/mL)		
$k_{obs,c_2 ex}$ (130 mg/mL; extrapolated from 10 and 35 mg/mL)		

D.1.5 Features derived from other stability studies (30 °C)

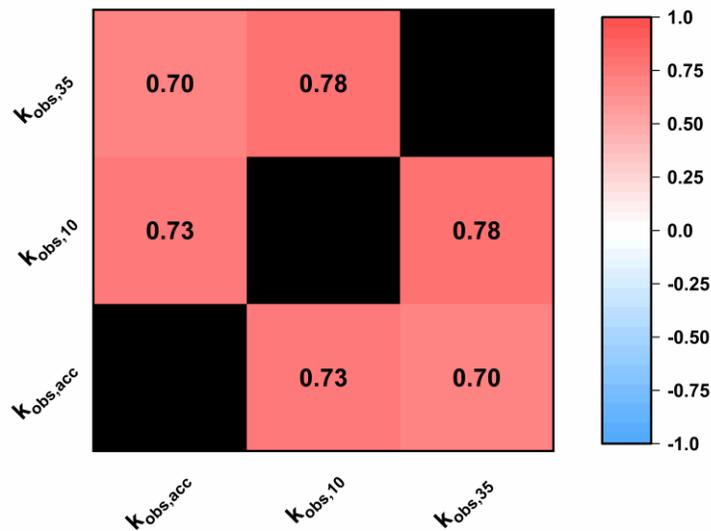


Figure D.1.5.1: Correlation coefficients for features in the stability studies (30 °C) subgroup.

Table D.1.5.1: Features in the stability studies (30 °C) subgroup and the feature(s) that were selected for use in statistical models

Feature	Selected feature(s)	Justification
$k_{obs,acc}$ (accelerated; 130 mg/mL) ^α	$k_{obs,acc}$ (accelerated; 130 mg/mL) ^α	The most experimentally accessible
$k_{obs,10}$ (30 °C; 10 mg/mL)		
$k_{obs,35}$ (30 °C; 35 mg/mL)	$k_{obs,10}$ (30 °C; 10 mg/mL)	

^α Aggregation rates at 45 °C were extrapolated to 50 °C using Arrhenius extrapolation with the rate at 30 °C

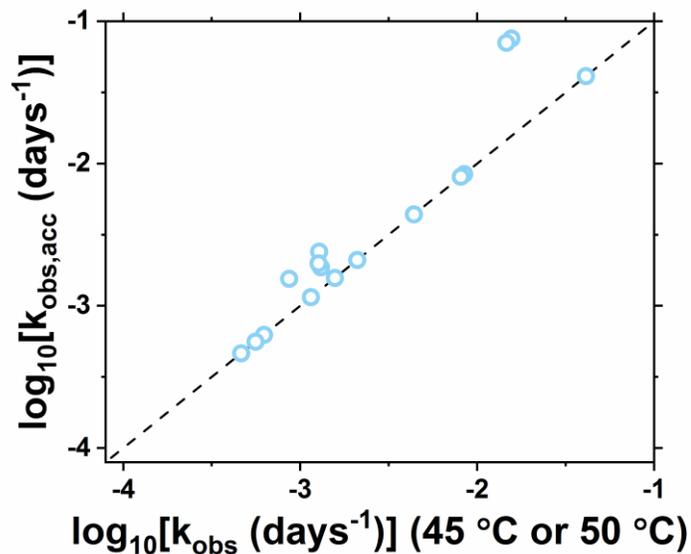


Figure D.1.5.2: Comparison of k_{obs} values from 130 mg/mL accelerated studies (x-axis; either 45 °C or 50 °C), and $k_{obs,acc}$, where rates at 45 °C were converted to estimated rates at 50 °C via Arrhenius extrapolation (y-axis). The black dashed line is the parity line.

D.1.6 Features derived from other stability studies (refrigerated)

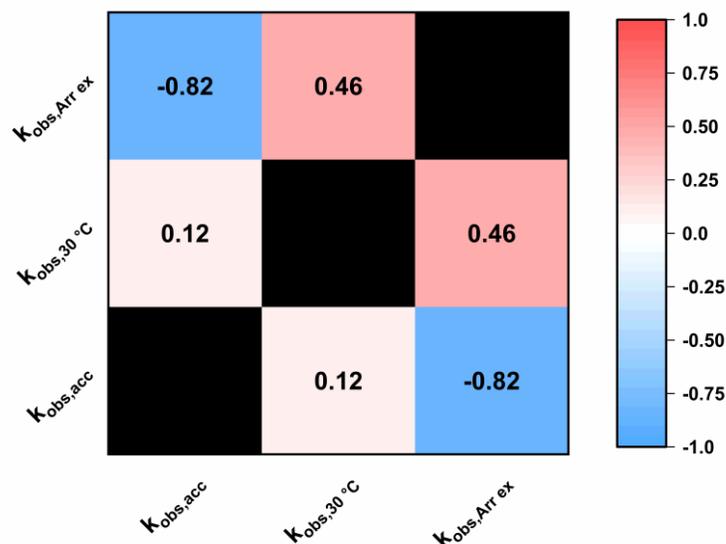


Figure D.1.6.1: Correlation coefficients for features in the stability studies (refrigerated) subgroup.

Table D.1.6.1: Features in the stability studies (refrigerated) subgroup and the feature(s) that were selected for use in ENR models

Feature	Selected feature(s)	Justification
$k_{obs,acc}$ (accelerated; 130 mg/mL) ^α	$k_{obs,Arr\ ex}$ (130 mg/mL; Arrhenius extrapolated)	$k_{obs,Arr\ ex}$ uses the other two features and is more correlated with k_{obs} values at refrigerated conditions The other two features were tested as well and were not selected by the ENR models
$k_{obs,30^{\circ}C}$ (30 °C; 130 mg/mL)		
$k_{obs,Arr\ ex}$ (130 mg/mL; Arrhenius extrapolated) ^β		

^α Aggregation rates at 45 °C were extrapolated to 50 °C using Arrhenius extrapolation with the rate at 30 °C

^β Aggregation rate at refrigerated conditions was estimated by Arrhenius extrapolation with aggregation rates at the 30 °C and accelerated T_{inc} conditions

D.1.7 Final feature set

Table D.1.7.1: Final set of input features that were considered for statistical modeling

Feature	Source	Phenomenon
$\ln(\text{IS})$	Formulation	Debye screening
$ \text{pH} - \text{pI} $	Formulation and cIEF	Net valence at a given pH
$T_{m,app1} - T_{inc}$	DSC	Conformational stability
ΔB_{22}^* ($= B_{22}^* - B_{22,NE}^*$)	SLS	Electrostatic net self-interactions (2-body)
$B_{22,NE}^*$	SLS	Non-electrostatic net self-interactions (2-body)
Features derived from other stability studies (accelerated)		
$k_{obs,c_2 ex}$ (130 mg/mL; extrapolated from 10 and 35 mg/mL)	Accelerated stability studies	c_2 dependence of aggregation rate
Features derived from other stability studies (30 °C)		
$k_{obs,acc}$ (accelerated; 130 mg/mL) ^α	Accelerated stability studies	Aggregation rate at higher T_{inc}
$k_{obs,10}$ (30 °C; 10 mg/mL)	30 °C stability studies	Aggregation rate at lower c_2
Features derived from other stability studies (refrigerated)		
$k_{obs,Arr ex}$ (130 mg/mL; Arrhenius extrapolated) ^β	Accelerated and 30 °C stability studies	Aggregation rate based on assumed Arrhenius behavior

^α Aggregation rates at 45 °C were extrapolated to 50 °C using Arrhenius extrapolation with the rate at 30 °C

^β Aggregation rate at refrigerated conditions was estimated by Arrhenius extrapolation with aggregation rates at 30 °C and the accelerated temperature condition

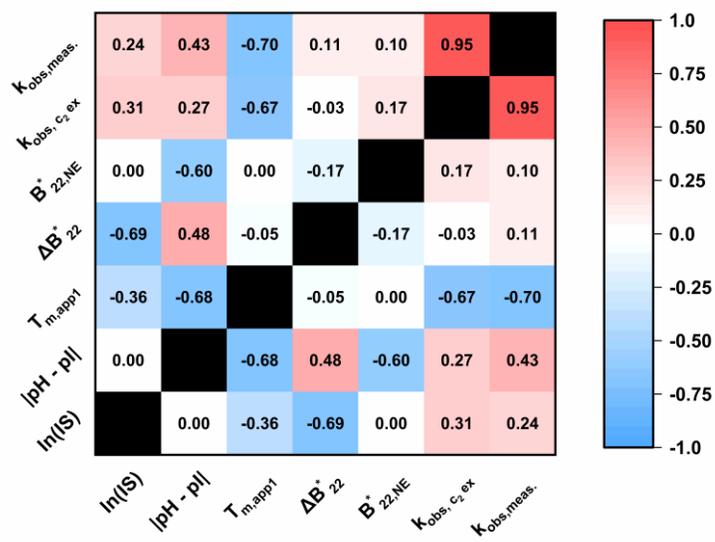


Figure D.1.7.1: Correlation matrix for input features for models of accelerated conditions. The last entry is the $\log_{10}(k_{obs})$ value for 130 mg/mL accelerated conditions. $T_{m,app1} - T_{inc}$ is shortened to $T_{m,app1}$ for readability.

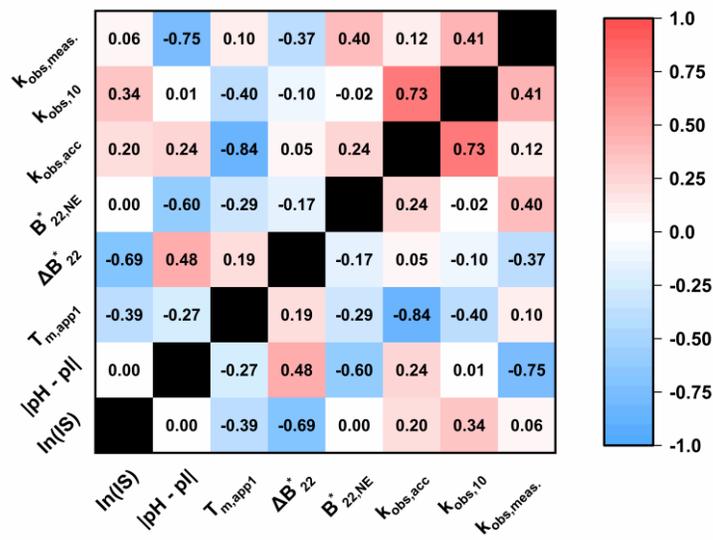


Figure D.1.7.2: Correlation matrix for input features for models of 30 °C conditions. The last entry is the $\log_{10}(k_{obs})$ value for 130 mg/mL 30 °C conditions.

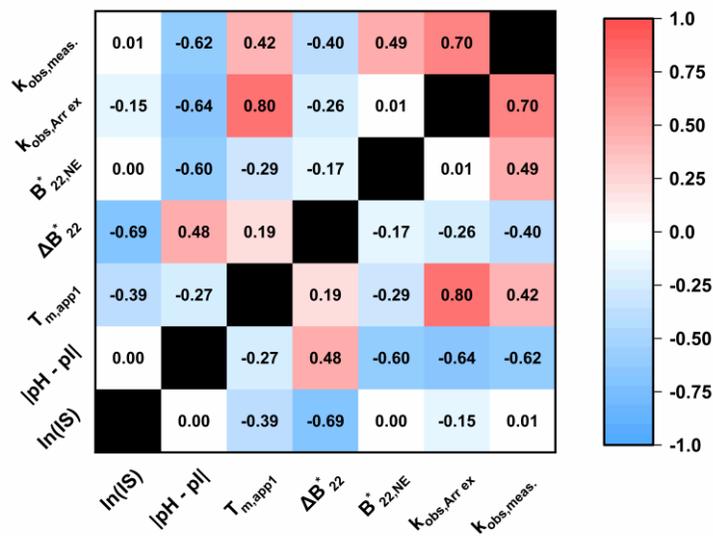


Figure D.1.7.3: Correlation matrix for input features for models of refrigerated conditions. The last entry is the $\log_{10}(k_{obs})$ value for 130 mg/mL refrigerated conditions.

Appendix E

REPRINT PERMISSION LETTERS



Home



Help ▾



Email Support



Sign in



Create Account



Toward a Suite of Coarse-Grained Models for Molecular Simulation of Monoclonal Antibodies and Therapeutic Proteins

Author: Hassan Shahfar, James K. Forder, Christopher J. Roberts

Publication: The Journal of Physical Chemistry B

Publisher: American Chemical Society

Date: Apr 1, 2021

Copyright © 2021, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

This is a License Agreement between James Forder ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	02-Oct-2023	Type of Use	Republish in a thesis/dissertation
Order License ID	1402451-1	Publisher	JOHN WILEY & SONS, INC.
ISSN	1547-5905	Portion	Chapter/article

LICENSED CONTENT

Publication Title	AIChE journal	Publication Type	e-journal
Article Title	Simulation of high-concentration self-interactions for monoclonal antibodies from well-behaved to poorly-behaved systems	Start Page	n/a
		End Page	n/a
		Issue	4
Author/Editor	American Institute of Chemical Engineers.	Volume	69
		URL	http://www3.interscience.wiley.com/cg...
Date	01/01/1955		
Language	English		
Country	United States of America		
Rightsholder	John Wiley & Sons - Books		

REQUEST DETAILS

Portion Type	Chapter/article	Rights Requested	Main product
Page Range(s)	1-13	Distribution	Worldwide
Total Number of Pages	13	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the Disabled?	Yes
Who Will Republish the Content?	Academic institution	Minor Editing Privileges?	Yes
Duration of Use	Life of current edition	Incidental Promotional Use?	No
Lifetime Unit Quantity	Up to 499	Currency	USD

NEW WORK DETAILS

Title	Graduate Student	Institution Name	University of Delaware
Instructor Name	Christopher J. Roberts	Expected Presentation Date	2024-01-01

ADDITIONAL DETAILS

The Requesting Person/Organization to Appear on the License: James Forder

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Simulation of high-concentration self-interactions for monoclonal antibodies from well-behaved to poorly-behaved systems	Title of the Article/Chapter the Portion Is From	Simulation of high-concentration self-interactions for monoclonal antibodies from well-behaved to poorly-behaved systems
Editor of Portion(s)	Forder, James K.; Ilott, Andrew J.; Sahin, Erinc; Roberts, Christopher J.	Author of Portion(s)	Forder, James K.; Ilott, Andrew J.; Sahin, Erinc; Roberts, Christopher J.
Volume / Edition	69	Publication Date of Portion	2023-04-01
Page or Page Range of Portion	n/a-n/a		