SUPPORTING INFORMATION

Predicting High-Concentration Interactions of Monoclonal Antibody Solutions: Comparison of Theoretical Approaches for Strongly Attractive vs. Repulsive Conditions

Cesar Calero-Rubio\textsuperscript{1,a}, Atul Saluja\textsuperscript{2,a}, Erinc Sahin\textsuperscript{2}, Christopher J. Roberts\textsuperscript{1,*}

\textsuperscript{1}Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, USA
\textsuperscript{2}Drug Product Science and Technology, Bristol-Myers Squibb, New Brunswick, New Jersey 08901, USA
\textsuperscript{a}Present address: Biologics, Drug Product Development and Manufacturing, Sanofi, Framingham, Massachusetts 01701, USA
\textsuperscript{*Corresponding author: tel. 302-831-1048; email: cjr@udel.edu

1. Simplification to the MSOS algorithm for \textit{B}_2 simulations

The original MSOS algorithm developed by Kofke and co-workers\textsuperscript{1-3} is composed of a double importance sampling (one for the computed or simulated system and one for the reference system). Consequently, its speed (for the same number of molecules and MC attempts) is half than that of the regular MS algorithm.\textsuperscript{2-6} However, previous work by the current authors has shown that this decrease in speed is justifiable as the MSOS algorithm can converge up to 100 times faster (e.g., \textit{10}^\textit{6} MC attempts on the MSOS would require up to \textit{10}^\textit{8} MC attempts with the regular MS algorithm), resulting in an overall speed-up of around 50 times. Here, an optimization to further speed up the MSOS algorithm is offered by analytically solving the importance sampling of the reference system, reducing the number of total importance samplings to only one. Thus, this is expected to further increase computational speed by up to a factor of two.

The original MSOS algorithm states that:

\begin{equation}
\Gamma_{\text{SYS}}(T, X) = \Gamma_{\text{REF}}(T) \frac{\langle y/\pi \rangle_{\pi}/\langle y_{\text{OS}}/\pi \rangle_{\pi}}{\langle y_o/\pi_o \rangle_{\pi_o}/\langle y_{\text{OS}}/\pi_o \rangle_{\pi_o}} \quad (A1)
\end{equation}
\[ \gamma_{OS} = \frac{|\gamma_o| |\gamma|}{\alpha |\gamma_o| + |\gamma|} \]  

where \( \Gamma_{SYS} \) is the computed virial coefficient (which is a function of the system temperature \( T \) and a set of model parameters \( X \)), \( \Gamma_{REF} \) is the virial coefficient of the reference system, \( \gamma \) is the Mayer function (for the second virial coefficient, \( B_{22} \)) or integrand (for higher order virial coefficients) of the computed system, \( \gamma_o \) is the Mayer function or integrand of the reference system, \( \gamma_{OS} \) is the overlap of \( \gamma \) and \( \gamma_o \), and \( \alpha \) is the overlap parameter.\(^{1,3} \) \( \pi \) and \( \pi_o \) indicate the distribution under which the importance sampling is performed.\(^{1,3} \) For simplicity, one can assume that \( \pi \) is equal to \( |\gamma| \) and \( \pi_o \) is equal to \( |\gamma_o| \). This leads to two independent simulations (one in \( \pi \) space and another one in \( \pi_o \) space) to solve equation A1 and obtain \( \Gamma_{SYS} \). It is worth to highlight that the value of \( \alpha \) in equation A2 must not be too small (\( i.e. \), close to 0) as this would reduce the MSOS algorithm into the MS version, deleting any gained advantage. Kofke and coworkers stated that \( \alpha \) needs to be tuned for convergence.\(^1 \) However, the authors in this work have not noticed any dependence on \( \alpha \) for values between 0.5 and 2.

In what follows, one needs to focus on the denominator of equation A1. If the reference system is fully embedded and overlapped with the computed system, any sampling on \( \pi_o \) space would always lead to \( \gamma = \gamma_o \) for \( \pi_o = |\gamma_o| \). For instance, this applies when the steric-only behavior is used as a reference state. For any steric overlap between the molecules, \( \gamma \) and \( \gamma_o \) would result in a value of -1 during \( B_{22} \) simulations (or different than 0 for higher order virial coefficients). Conversely, a non-overlap will result in a value of \( \gamma_o = 0 \) and \( \gamma \neq 0 \) (if the computed system includes interactions beyond steric). However, these non-overlapped conditions are never accepted in \( \pi_o \) space,\(^1 \) so \( \gamma \) will always be equal to \( \gamma_o \) for this evaluated example. Consequently, equation A2 in \( \pi_o \) space can be simplified as:

\[ \gamma_{OS} = \frac{|\gamma_o| |\gamma|}{\alpha |\gamma_o| + |\gamma|} = \frac{|\gamma_o|}{\alpha + 1}, \quad \text{in } \pi_o \text{ space and } \pi_o = |\gamma_o| \]  

Combining equations A1 and A3, and using the simplification \( \pi_o = |\gamma_o| \), one would obtain:
\[ \Gamma_{SYS}(T,X) = \Gamma_{REF}(T) \frac{\langle y/\pi \rangle_\pi / \langle y_{OS}/\pi \rangle_\pi}{\langle y_o/|y_o| \rangle_{\pi_o} / (1/(\alpha + 1))_{\pi_o}} \] (A4)

Finally, if the interactions of the reference system are always repulsive (or attractive), the term \( \langle y_o/|y_o| \rangle_{\pi_o} \) simplifies into \(-1^j\) in the case of \( \Gamma_{SYS} = B_{22} \), where \( j = 1 \) for always repulsive conditions (e.g., for cases where sterics are used as the reference system) and \( j = 0 \) for always attractive conditions. This arises from the fact that \( y_o/|y_o| \) is always equal to 1 for attractions and to -1 for repulsions, regardless of the interaction potential model. Finally, equations A1 and A4 can be written as:

\[ \Gamma_{SYS}(T,X) = (-1)^j \frac{\Gamma_{REF}(T)}{\alpha + 1} \cdot \frac{\langle y/\pi \rangle_\pi}{\langle y_{OS}/\pi \rangle_\pi} \] (A5)

As it can be observed, equation A5 only requires a single importance sampling in \( \pi \) space, improving computational times by up to a factor of two in comparison to equation A1. Additionally, A5 is the simplified version of A1 only for \( B_{22} \) simulations as a pre-factor to \((-1)^j\) must be added for higher order virial coefficients. This was outside of the scope of this work and is left open for additional contributions within the field.

The reader must be aware that the simplified version presented in equation A5 would require the reference system to: (i) always be either attractive or repulsive (never attractive and repulsive), (ii) be fully embedded into the main system and (iii) be fully overlapped with the computed system (i.e., \( \gamma = \gamma_o \) if \( \gamma_o \neq 0 \) in \( \pi_o \) space). This leads to a significant limitation in the use of the simplified form of the MSOS algorithm: the magnitude of \( \Gamma_{SYS}(T,X) \) must be close to that of \( \Gamma_{REF}(T) \). For values with significantly different magnitudes (e.g., for a parameter set of \( X \) that leads to strongly attractive inter-molecular interactions while using the steric-only behavior as the reference system and stronger than those exhibit by MAb2 at pH 6.5 and very low TIS within the main document), the MSOS algorithm might not fully converge. Thus, the increase in computational speed provided by equation A5 is offset by a lack of convergence for considerably different interaction conditions between the computed and the reference systems. Nevertheless, for the experimental conditions studied in this and some previous works by the same authors, this was not a relevant limitation.
2. Charge distribution of MAb1

**Figure S1.** Theoretical charge distribution for the DODECA model at pH 5 and 6.5 for MAb1 as reported in Ref 2.
3. HEXA simulations for MAb2 with the TMMC and MSOS algorithms

Figure S2. **Panels A-B:** Comparison of $B_{22}/B_{22,ST}$ as a function of TIS between experimental (symbols) and simulated values (shaded areas) using the HEXA model at pH 5 for buffer (A) and 5% w/w added sucrose (B) conditions. The insets correspond to surface response of ARD values as a function of $\varepsilon_{SR}$ and $\psi$. **Panels C-D:** High-$c_2$ predictions of $R^{ex}/K$ and $S_{q=0}$ from low-$c_2$ parameters with the HEXA model shown in Figure 4, and for buffer-only (black squares), 5% w/w sucrose (red triangles) and 100 mM NaCl (gray circles). The blue dashed-line corresponds to the steric-only behavior.
Figure S3. Comparison of the MSOS and TMMC approaches for case studies using the HEXA model for net-repulsive (panel A) and net-attractive (panel B) conditions. Black solid lines represent the TMMC results, while red dashed, blue dotted, green dash-dotted and gray solid lines represent the MSOS results with up to the 2nd, 3rd, 4th and 5th virial coefficient, respectively. Insets correspond to the relative deviation as a function of \( c_2 \) for each model using the TMMC results as the reference.
4. Additional results from 1bAA simulations for MAb1 and MAb2

![Graph showing B22 as a function of εSR using the 1bAA CG model for MAb1 (gray) and MAb2 (black) molecules. Inset shows the same data as the main panel, but on a scaled focused to where experimental B22/B22,ST values lie for MAb1 and MAb2 in this work.](image)

**Figure S4.** $B_{22}$ as a function of $\varepsilon_{SR}$ using the 1bAA CG model for MAb1 (gray) and MAb2 (black) molecules. Inset shows the same data as the main panel, but on a scaled focused to where experimental $B_{22}/B_{22,ST}$ values lie for MAb1 and MAb2 in this work.

![Graph showing experimental B22/B22,ST vs TIS with best fit parameters using the 1bAA model for MAb1 (panel A, $\varepsilon_{SR} = 0.5$ k_BT, $\psi = 0.4$ at pH 5 and $\psi = 0.74$ at pH 6.5) and MAb 2 (panel B, $\varepsilon_{SR} = 0.44$ k_BT, $\psi = 0.7$ at pH 5 and $\psi = 1.4$ at pH 6.5). Black lines and symbols correspond to pH 5 and gray lines and symbols to pH 6.5. Symbols correspond to the experimental data in Figure 3 and in Ref. 2, while lines correspond to simulated values.](image)

**Figure S5.** Experimental $B_{22}/B_{22,ST}$ vs TIS with best fit parameters using the 1bAA model for MAb1 (panel A, $\varepsilon_{SR} = 0.5$ k_BT, $\psi = 0.4$ at pH 5 and $\psi = 0.74$ at pH 6.5) and MAb 2 (panel B, $\varepsilon_{SR} = 0.44$ k_BT, $\psi = 0.7$ at pH 5 and $\psi = 1.4$ at pH 6.5). Black lines and symbols correspond to pH 5 and gray lines and symbols to pH 6.5. Symbols correspond to the experimental data in Figure 3 and in Ref. 2, while lines correspond to simulated values.
References


