



This is a repository copy of *Optimizing excipient properties to prevent aggregation in biopharmaceutical formulations*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/211443/>

Version: Published Version

Article:

King, T.E. orcid.org/0009-0004-2020-311X, Humphrey, J.R., Laughton, C.A. orcid.org/0000-0003-4090-3960 et al. (2 more authors) (2024) Optimizing excipient properties to prevent aggregation in biopharmaceutical formulations. *Journal of Chemical Information and Modeling*, 64 (1). pp. 265-275. ISSN 1549-9596

<https://doi.org/10.1021/acs.jcim.3c01898>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Optimizing Excipient Properties to Prevent Aggregation in Biopharmaceutical Formulations

Toby E. King, James R. Humphrey, Charles A. Laughton, Neil R. Thomas, and Jonathan D. Hirst*



Cite This: *J. Chem. Inf. Model.* 2024, 64, 265–275



Read Online

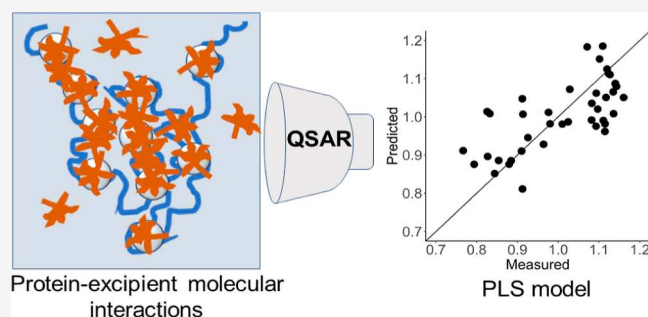
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: Excipients are included within protein biotherapeutic solution formulations to improve colloidal and conformational stability but are generally not designed for the specific purpose of preventing aggregation and improving cryoprotection in solution. In this work, we have explored the relationship between the structure and antiaggregation activity of excipients by utilizing coarse-grained molecular dynamics modeling of protein–excipient interaction. We have studied human serum albumin as a model protein, and we report the interaction of 41 excipients (polysorbates, fatty alcohol ethoxylates, fatty acid ethoxylates, phospholipids, glucosides, amino acids, and others) in terms of the reduction of solvent accessible surface area of aggregation-prone regions, proposed as a mechanism of aggregation prevention. Polyoxyethylene sorbitan had the greatest degree of interaction with aggregation-prone regions, decreasing the solvent accessible surface area of APRs by 20.7 nm² (40.1%). Physicochemical descriptors generated by Mordred are employed to probe the structure–property relationship using partial least-squares regression. A leave-one-out cross-validated model had a root-mean-square error of prediction of 4.1 nm² and a mean relative error of prediction of 0.077. Generally, longer molecules with a large number of alcohol-terminated PEG units tended to interact more, with qualitatively different protein interactions, wrapping around the protein. Shorter or less ethoxylated compounds tend to form hemimicellar clusters at the protein surface. We propose that an improved design would feature many short chains of 5 to 10 PEG units in many distinct branches and at least some hydrophobic content in the form of medium-length or greater aliphatic chains (i.e., six or more carbon atoms). The combination of molecular dynamics simulation and quantitative modeling is an important first step in an all-purpose protein-independent model for the computer-aided design of stabilizing excipients.



INTRODUCTION

Protein-based biotherapeutics are a growing market, with significantly more treatment options based on biologics under development and a multibillion dollar industry revolving around their research and manufacture; in 2021, 28% of all FDA-approved drugs were biologics.¹ The majority of biotherapeutics include hormones,² plasma proteins,³ enzymes,⁴ coagulation factors,⁵ vaccines,⁶ and monoclonal antibodies (mAb) and their fragments.⁷ mAbs are the largest fraction⁸ and are used primarily as immunotherapeutics, for targeted delivery,⁹ and cancer vaccines.¹⁰ Generally, therapeutic proteins are produced in bioreactors using recombinant cell lines¹¹ and are often lyophilized or frozen for storage. One of the key challenges facing protein biotherapeutics is their conformational and colloidal stability as formulation and storage conditions can induce aggregation and agglomeration¹² during both freezing and rethawing or resuspension.¹ These aggregates have reduced function¹³ and an increased specific immune response when administered;¹⁴ indeed, the association constant of human serum albumin (HSA) to ketuprofen decreased by 42% after the formation of fibrillar aggregates by HSA.¹⁵

As folding occurs, the tertiary structure of a protein changes as hydrophobic residues are buried within the 3D structure. The folding protein assumes transient intermediate structures of increasing stability and reaches a thermodynamic global minimum at the native conformation, sometimes guided by molecular chaperone proteins.^{4,13,16} During manufacture and storage, proteins are exposed to non-native conditions, such as nonphysiological pH, ionic strength, extremes of temperature, interactions with impurities, and hydrophobic interactions at interfaces with synthetic surfaces or air, which may induce partial unfolding or misfolding and can lead to noncovalent aggregation (Figure 1). The change in the structure may expose hydrophobic residues, which form patches on the surface of the protein.¹⁷ The energy landscape changes; it

Received: November 26, 2023

Revised: December 4, 2023

Accepted: December 8, 2023

Published: December 19, 2023



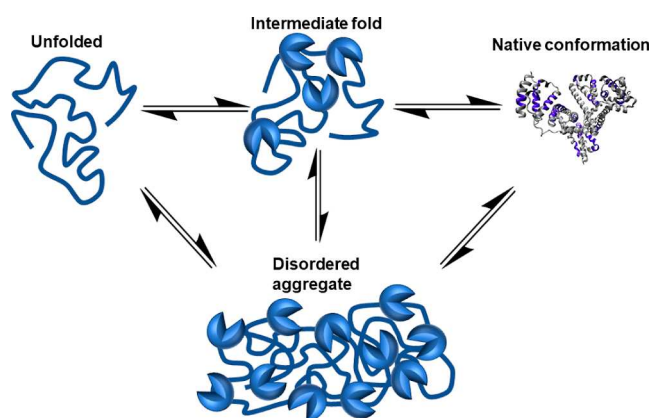


Figure 1. Folding and misfolding pathways of a protein. An unfolded protein assumes more stable intermediate folding conformations until arriving at the native configuration. If subjected to non-native conditions, the equilibrium position can change to favor the formation of a stable, disordered conformation, which can form an aggregation nucleus while residing within a thermodynamic energy minimum.

becomes more favorable to bury the hydrophobic patches by interaction with hydrophobic surfaces, such as similar patches on other protein molecules. This process is driven primarily by hydrophobic interaction, but electrostatics and hydrogen bonding also contribute.¹³ Solvent is preferentially excluded from the protein surface as the protein molecules interact with one another, and more molecules are recruited into the aggregation nucleus in an irreversible process.¹⁸

The tendency of protein biotherapeutics to aggregate can be mitigated by the modification of conditions, such as pH¹⁹ and ionic strength,²⁰ as well as the inclusion of excipients into biotherapeutic formulation.^{12,21} Excipients used to improve biotherapeutic stability include histidine,²² arginine,²³ sugars,²⁴ fatty alcohol ethoxylates,²⁵ alkylsaccharides,²⁶ poloxamers,²⁷ and polysorbates.²⁸ The mechanism by which aggregation is prevented is not fully understood. One proposal is the formation of protein–excipient complexes, which could shield aggregation-prone regions (APRs) of the protein from solvent or other hydrophobic surfaces.^{29,30} Competitive adsorption at surface interfaces, particularly by surfactants, may prevent aggregation by reducing the exposure of the protein to another hydrophobic surface, thereby reducing partial unfolding and aggregation nucleation.^{31,32} Excipients also modify the energetics of native intermediates and increase stability, by making disordered intermediates less favorable and acting as a chaperone to facilitate native folding.³³

Differences in the protein structure complicate the understanding of aggregation prevention; as proteins' structures differ, so too will their aggregation propensity, as well as their interaction with antiaggregation agents. Hydrophobic patches of proteins are exposed to solvent to different extents, and APRs will not have the same topology and charge distribution across different proteins.¹⁷ There are multiple approaches to predicting APRs using differing levels of the protein structure. Some, such as Aggrescan,¹⁷ work solely from the primary structure and determine APRs by comparing the amino acid sequence against an experimentally determined aggregation propensity. Others account for the 3D structure and, thus, the solvent-accessible surface area (SASA). Examples of this approach include SAP¹⁷ and Aggrescan-3D.³⁴ Generally, excipients are chosen not in light of efficacy as an antiaggregation agent, but due to their well-established safety

profiles from other uses;³⁵ for example, polysorbates are popular emulsifiers, particularly in cosmetics³⁶ and in the food industry.³⁷ Therefore, there is chemical space to explore to optimize antiaggregation excipients.

Computational techniques can provide mechanistic insights into length and time scales that are inaccessible to conventional wet lab methods.³⁸ Molecular dynamics (MD) simulations have been applied in the study of surfactant behavior in solution;^{39,40} protein–surfactant interaction,²⁹ including stability modulation⁴¹ and binding;⁴² protein aggregation⁴³ and folding;⁴⁴ and the modulation of protein stability by excipients such as histidine.⁴⁵ Atomistic or pseudoatomistic MD models often have a prohibitively high computational expense to be applied in large numbers of simulations that examine microsecond-time scale events, such as many aspects of protein dynamics.⁴⁶

There are few investigations of the nonspecific interaction between excipients and APRs as a mechanism of aggregation prevention that considers all areas of the protein. No quantitative structure–property relationship model has been derived that probes the relationship between the excipient structure and antiaggregation activity. In this work, we present an MD model that investigates APR–excipient interaction to determine the stabilizing effect on protein biotherapeutics, coupled with a quantitative model which uses physicochemical descriptors in statistical analysis to reveal the impact of the key features on antiaggregation activity. In doing so, we investigate the model of the shielding of APRs from solvent as a mechanism of aggregation prevention, hypothesizing that a smaller SASA of APRs leads to greater stability. To produce sufficient data for a quantitative model, a coarse-grained (CG) force field was selected, as they allow access to microsecond simulation time scales at reasonable computational expense and without the need for enhanced sampling methods. CG force fields decrease the computational cost at the expense of resolution by representing multiple atoms as a single interaction site; doing so can facilitate the large-scale simulation at microsecond time scales, as there are fewer degrees of freedom to consider.

MARTINI^{47,48} is a prominent CG force field which maps atoms to beads at an approximately 4:1 ratio in a building-block approach. It has been applied to many different biomolecular systems, such as membrane studies, protein–ligand binding, phase behavior, carbohydrates, and nucleic acids. MARTINI has also been applied specifically in the context of improving protein stability by including excipients that reduce antibody self-association; Lui et al. utilized a docking approach to screen excipients by binding with the most significant APR. The Docking Assay For Transmembrane components (DAFT) method for the high-throughput study of dimer/trimer association⁴⁹ was applied in order to sample sufficient initial relative poses of antibody fragments, resulting in a CG-MD model of antibody self-association and the effect of excipients on aggregation kinetics.⁵⁰ Similarly, insulin self-association and its non-Arrhenius behavior were investigated in a study of aggregation nucleation kinetics in MARTINI,⁵¹ finding that the insulin unfolding equilibration constant is the single most important kinetic parameter in nucleation time.

Excipients were selected based on their prevalence in the industry as solution state stability enhancers, their prior parametrization by the MARTINI development team, or their utility to a quantitative model. PEG alkyl amides (PAAs) consist of a PEG chain, amide linker, and alkyl chain. Fatty acid

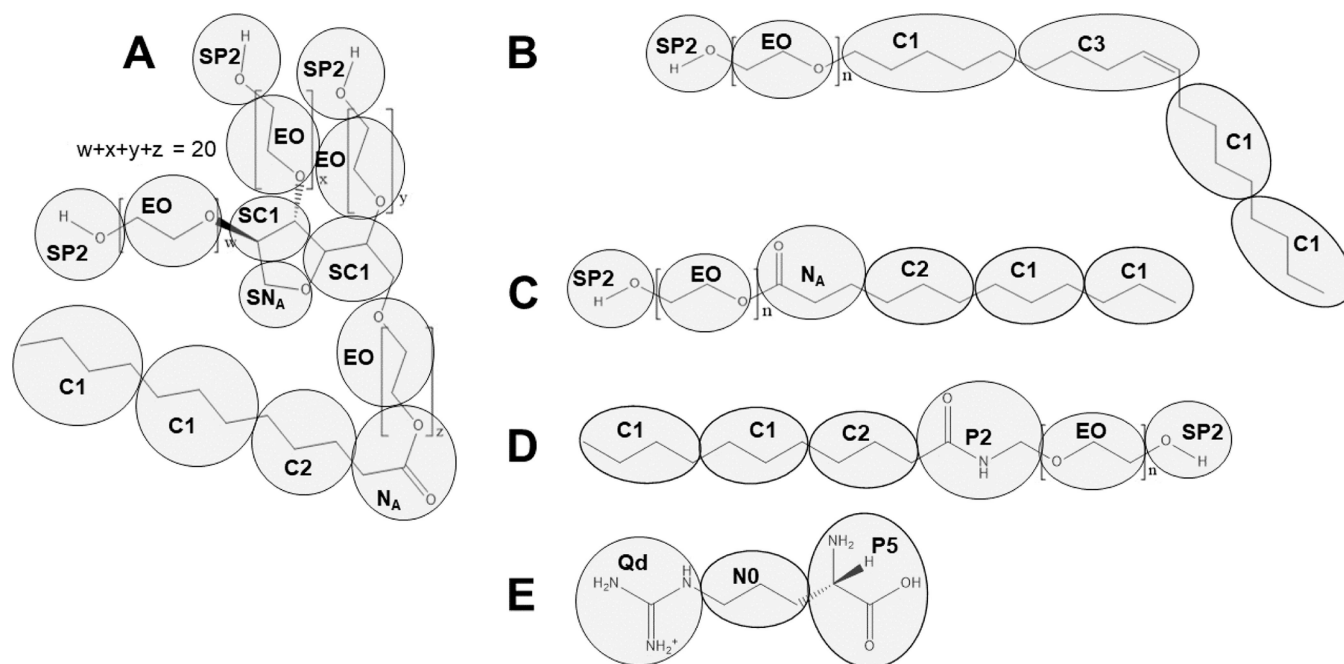


Figure 2. Chemical structures of studied excipients with their MARTINI mapping superimposed. (A) Polysorbate 20, (B) fatty alcohol ethoxylate (Brij) On , (C) FAE Ln , (D) PEG alkyl amide Ln , and (E) L-arginine

ethoxylates (FAEs) and fatty alcohol ethoxylates compounds are similar but have an ester bond or an ether bond in place of the amide linker, respectively. Polysorbates are fatty acid esters of polyoxyethylene sorbitan (PSBN). Spans are similar to polysorbates but are not ethoxylated.⁵² Other compounds include cholesteryl glucopyranoside, a range of phospholipids, fatty acids, arginine, and *n*-octyl glucoside. This range of chemically diverse compounds facilitates the extraction of useful information for quantitative modeling and allows data-driven decisions to be made in the design of antiaggregation excipients. The application of these data could improve biotherapeutic formulation design by lowering costs, improving therapeutic outcomes, and elucidating structure–property relationships.

HSA was chosen as a model protein due to its use in biotherapeutic formulations, both as an active pharmaceutical ingredient⁵³ and an excipient,⁵⁴ its loss of function after aggregation,¹⁵ and its manageable size of 585 residues. Some evidence indicates that the binding between HSA and excipients (specifically polysorbates) occurs within endogenous binding sites⁵⁵ and thus could pose difficulties in extrapolating the model to other therapeutically relevant proteins, particularly as the same study indicated negligible interaction between polysorbates and IgG. However, there is also evidence of polysorbates interacting with pharmaceutically relevant proteins, including human growth hormone,⁵⁶ an IgG mAb,⁵⁷ filgrastim,^{58,59} lysozyme, RN29S, and recombinant factor VIII,⁶⁰ imparting improvements to their physical stability, in conjunction with surfactant and interfacial stabilizing interactions. Thus, the interaction between HSA and the excipients selected for this study could feasibly be applied to different proteins to elucidate excipient and protein interactions and their potential roles in preventing aggregation.

MATERIALS AND METHODS

The initial structure of HSA was obtained from the RCSB Protein Data Bank (code 4L8U)⁶¹ and processed into the

MARTINI force field via the `martinize2` script, from the `vermouth` package.⁶² Its APRs were highlighted using the Aggrescan web server¹⁷ and its FASTA sequence; the APRs consisted of 25.4% of the sequence in 18 different patches.

Parameterization. To parametrize excipients that are not available from MARTINI, initial united-atom coordinates and topologies were generated using the Automated Topology Builder⁶³ in the GROMOS 54a7 force field⁶⁴ and converted into a MARTINI model. The MARTINI mapping was based on existing MARTINI beads and their use in the literature, as well as the preservation and representation of functional groups (Figure 2). Molecule parameters reported in the previous work by the MARTINI group and used here include phospholipids, ceramides, and glycerols,⁶⁵ as well as sugars,⁶⁶ fatty acids,⁶⁷ and sterol groups.⁶⁸

The initial united-atom structure is simulated for 10 ns in water at pH 7.0 and indexed so that each index group of atoms corresponds to a MARTINI bead. The angles and distances between these beads are measured and used as the bonded parameters in the MARTINI topology, a frame is extracted and used as the initial structure for a MARTINI simulation, and the bond lengths and angles are measured. These values and their force constants are modified in an iterative process until their distributions throughout both the MARTINI and indexed simulations are approximately matched. Polyply⁶⁹ was also used to generate initial MARTINI topologies for some compounds.

Molecular Dynamics Simulation. All MD simulations were carried out using GROMACS 2019 and 2021.2 in the Martini 2.3P force field, and five independent simulations were performed for each system. A truncated octahedral box was built around a single molecule of HSA, with the distance between opposing hexagonal sides equaling 34.24 nm, leading in practice to a volume of 30841.5 nm³. Each simulation contained a single molecule of HSA and approximately 233,000 MARTINI water molecules, for a protein concentration of 0.0538 mM or 3.61 mg/mL; therapeutic HSA

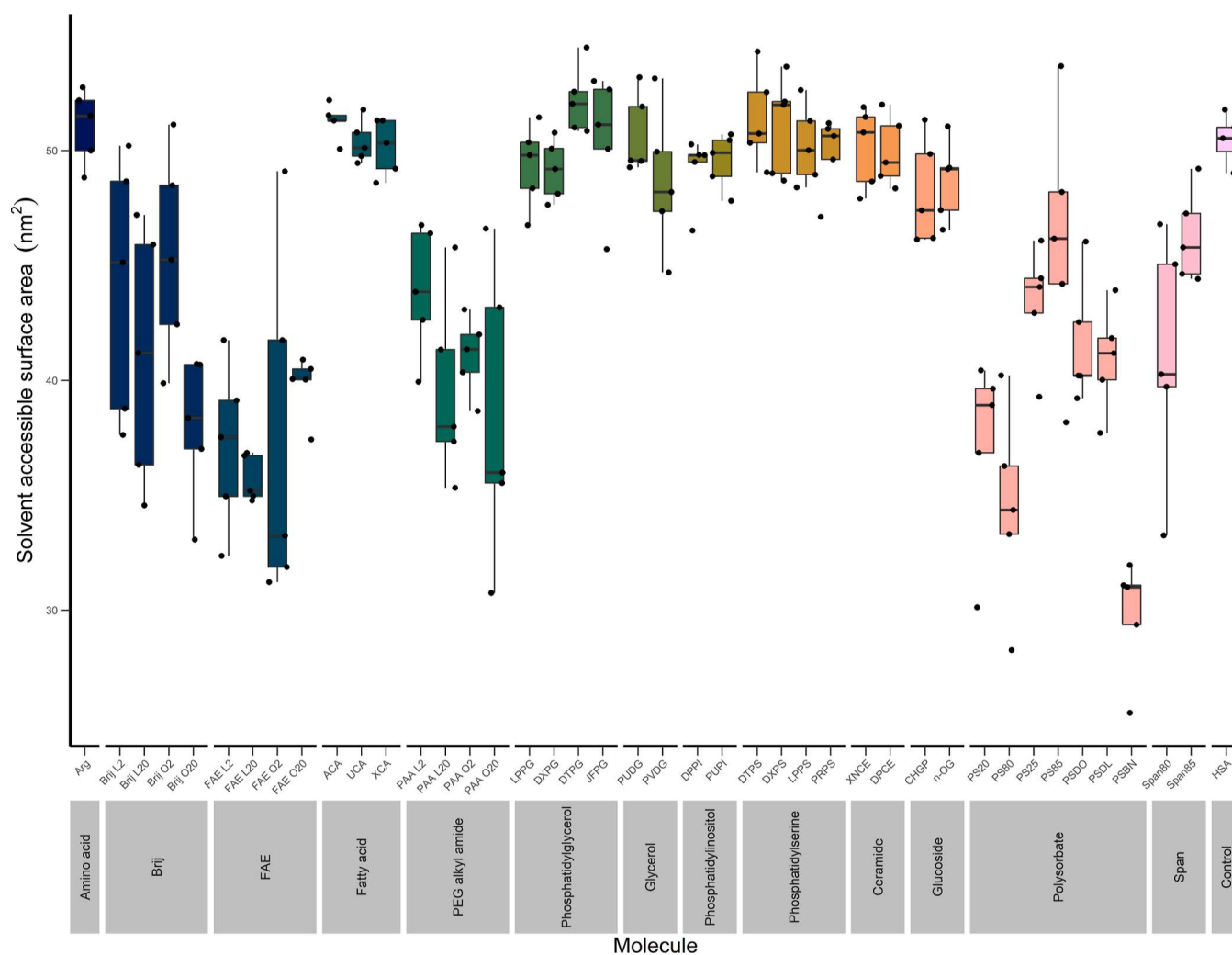


Figure 3. Average SASA of APRs, averaged for each trajectory. Polysorbates have the greatest effect on the SASA of APRs. Of the linear, ethoxylated surfactants, there is little significant difference between them, across all classes, but they are all significantly different from the control, with the exceptions of Brij O2 and L2. Arginine, phospholipids, fatty acids, and glucosides had an insignificant effect on the SASA of the APRs.

formulations are typically between 5 and 25% w/v.⁷⁰ This size was a compromise between having a sufficiently large system to model a comparatively low excipient concentration, with enough excipient molecules for the system to be thermodynamically realistic, and the prohibitive computational expense that would result from larger systems to model concentrations approaching those found in HSA therapeutic formulations. Sufficient excipient molecules were added to bring their concentration to 0.1% w/w, an industrially relevant concentration^{71,72} via *gmx insert-molecules* inserting into vacuum. In practice, this leads to a variable molar concentration, proportional to the molecular weight of the excipient. This is not an issue as it maintains the quantity of Martini “beads” across all simulations and makes comparisons between them more straightforward. The vacuum system was minimized for 1000 steps using the steepest descent algorithm and solvated using the MARTINI polarizable water model.⁷³ Sodium ions were added to neutralize the system by replacing water molecules at random, and the system was minimized again for 1000 steps. The system is relaxed in the *NPT* ensemble, with a 5 fs time step, *V*-scale thermostat at 300 K, and isotropic pressure coupling at 1.0 bar using the Berendsen barostat.⁷⁴ This relaxation phase consisted of 100 ps. Production MD was

performed in the same ensemble, with the same thermo- and barostats, a time step of 20 fs, and a total time of 1 μ s. For some compounds, particularly those with ring structures, a time step of 10 fs was necessary to run stable MD; the overall time remained 1 μ s. Coulombic and Lennard-Jones cutoffs were 1.1 nm and used the reaction field and potential shift Verlet modifiers, respectively, in the Verlet cutoff scheme. All trajectories were found to be equilibrated and converged, which in detail can be found in the [Supporting Information](#). Full parameter files can be found in the Github repository (see [Supporting Information](#)).

The SASA of the APRs was calculated using *gmx sasa* within Gromacs, indexed to calculate the SASA of APRs alone, using lone HSA as a control. Bartlett’s test⁷⁵ was utilized to indicate homoscedasticity between distributions for each excipient–protein simulation, and the results directed whether the Kruskal–Wallis⁷⁶ (homoscedastic) or Welch’s⁷⁷ (heteroscedastic) analyses of variance were employed to determine statistical significance. All analysis scripts can be found in the Github repository ([Supporting Information](#)).

Structure–Property Relationship. To probe the structure–property relationship of antiaggregation activity, partial least-squares (PLS)⁷⁸ regression was performed, using a set of

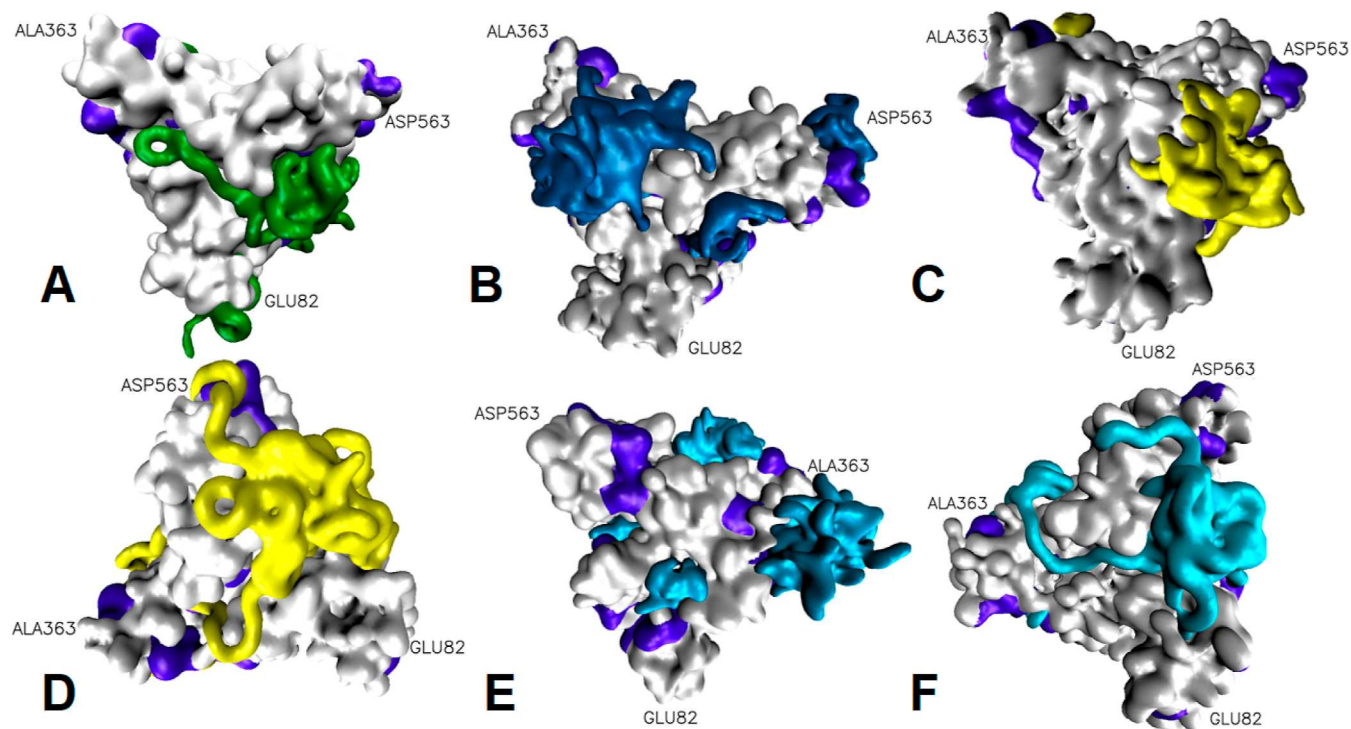


Figure 4. Select snapshots of trajectories at the end point of simulation, with the water removed for clarity. HSA is in gray with its APRs colored in violet. Glu82, Ala363, and Asp563 are labeled for orientation. Longer molecules wrap around the protein, while smaller molecules form clusters at the protein surface. (A) FAE O20. (B) Polysorbate 20. (C) PAA L2. (D) PAA O20. (E) Fatty alcohol ethoxylate (Brij) L2. (F) Fatty alcohol ethoxylate (Brij) O20. Gray, protein; violet, APR; green, FAE; dark blue, polysorbate; yellow, PAA; and light blue, fatty alcohol ethoxylate.

physicochemical descriptors as input. Molecular descriptors were generated using the Mordred package⁷⁹ in Python and filtered based on their utility in the context of chemical intuition, leaving a total of 106 descriptors. PLS regression was performed on the entire data set, employing leave-one-out cross validation⁸⁰ (LOO-CV) to find the optimal number of components to include in the model. This is achieved by using a number of components that cause the root mean squared error of prediction to be at a minimum, while also taking into account the principle of parsimony and avoiding overfitting. Four components were used in the final PLS model. To measure the robustness and efficacy of the model in predicting data, the data set was split into a partition of 0.8/0.2 training data/test data. LOO-CV was performed on the training data set, the model was applied to predict the test data set, and the Q^2 was recorded as a measure of predictive accuracy. This was repeated 1000 times; the Q^2 reported hereafter is the median of these repetitions.

RESULTS

Protein–Excipient Interaction. The shielding of APRs from solvent by excipient molecules is a key mechanism in the prevention of aggregation and increase in stability of biotherapeutic protein formulations; this can be quantified in an MD model by the extent to which the SASA of APRs reduces. HSA without any excipients was found to have a SASA of 271.7 nm²; within that, its APRs have an average SASA of 50.5 nm² polysorbate compounds have the greatest impact on the SASA of APRs (Figure 3) and are all statistically significant from the HSA-only control, according to Kruskal–Wallis⁷⁶ and Dunn tests. PSBN, the strongest performer, is significantly different from Brij L2 ($p < 0.05$), Brij O2 ($p <$

0.05), and PS85 ($p < 0.01$). PS80 is significantly different from Brij O2 ($p < 0.05$), which is somewhat surprising, given that they contain the same aliphatic chain content (a single oleate). Linear ethoxylated compounds were not significantly different from one another, with the exception of Brij O2, which was different from every other linear ethoxylated compound. ($p < 0.05$). The only ethoxylated compounds to not be significantly different from the control were Brij O2, Brij L2, and Span 85. Span 80 was significantly different from the control, but no difference was found between it and any polysorbate compound. None of the other compounds under study were found to have an impact on the SASA of the APRs of HSA that was significantly different from the control.

There is a significant degree of heterogeneity in performance, within both a single class and repetitions of the same excipient. This could be indicative of the nonspecific nature of binding; the interaction within each individual repetition and each individual molecule could be between many different residues in a heterogeneous manner, and a weak interaction might not guarantee the formation of an HSA–excipient complex within the simulation time. In each trajectory, protein–excipient contacts remained dynamic to some degree, fluctuating above and below the average. Each trajectory appeared to be at equilibrium in this way. This is indicative of the interaction being somewhat reversible, although the deviation from the average throughout a given trajectory is not large.

The significant α -helical content of HSA will have an effect, as the configuration in space will affect both the accessibility of specific residues and the local environment in which they reside. This is represented in MARTINI as a change in the polarity of the backbone bead of all residues present in a helix

as well as the side chains of glycine, alanine, and protein, all represented as significantly less polar beads.⁸¹ Therefore, an alanine residue within an α -helix will have significantly less hydrophobic character than an alanine residue outside a helix. As 87.5% of the APRs are found within α -helices and 39.1% of the residues within the helices are APRs, it is likely that the interaction between helices and excipient or between helices and solvent is significant in aggregation prevention. Indeed, α -helices have been shown to induce the formation of protein aggregates.^{82,83}

Visual inspection of the trajectories can also reveal the characteristics of the excipient–protein interaction. Qualitatively, compounds with a high PEG content, such as polysorbates or linear compounds with 20 PEG units, have a tendency to wrap around the protein while shorter ethoxylated compounds form localized, hemicellar clusters around a small number of residues (Figure 4). Unsurprisingly, of the simulations that showed little to no contact (such as phospholipids), little information can be gleaned from the nature of their interaction from the inspection of the arrangement in space. However, in the trajectories containing free arginine as an excipient, there is little evidence of continued sustained interaction, supporting the notion that its interaction is transient.

Structure–Property Relationship. The final PLS model of two components, validated with LOOCV, has an R^2 value of 0.398 and a mean relative error of prediction of 0.077. To gain an understanding of the robustness of the data set and its validity in regression, the data set of excipient simulations was split 0.8/0.2 training data set/testing data set, and the Q^2 was 0.344, with median root-mean-square errors of 4.10 and 4.37 nm² for the training and test sets, respectively. These distributions of a measure of goodness of fit gives confidence that there is sufficient variation within the data set for its utility in a quantitative structure–property relationship application. Independently, a new model was constructed trained on all 41 instances to determine the importance of descriptors (and not to assess the predictive accuracy). There is a distinct divide between heavy molecules containing a relatively large amount of PEG that performed well in shielding APRs and thus improving stability and both smaller ethoxylated molecules and larger ones without any PEG (Figure 5). The PLS results show a clear demarcation between strongly interacting molecules and weak or noninteracting molecules and reveals physicochemical and structural differences between the two groups. Broadly, highly branched molecules and those with a high PEG content are within the well-performing cluster (cluster 1), while linear molecules and compounds with little to no PEG content are found within the other, broader cluster of poorly performing antiaggregation agents, with small ethoxylated compounds forming their own grouping along with fatty acids and arginine (cluster 2). The other poorly performing and/or PEG-lacking compounds make up a broadly dispersed cluster (cluster 3). There appears to be a moderate negative correlation between component 1 and the SASA of APRs.

The coordinates of Mordred variables in latent space, and their relation to compounds' coordinates in the same space, can indicate the physicochemical forces involved in APR shielding. There are broadly similar but decidedly more scattered clusters within the variable space. Descriptors with a positive score in the second component and a negative score in the first component include those related to the number of oxygen atoms and the nature of their bonds, the number of

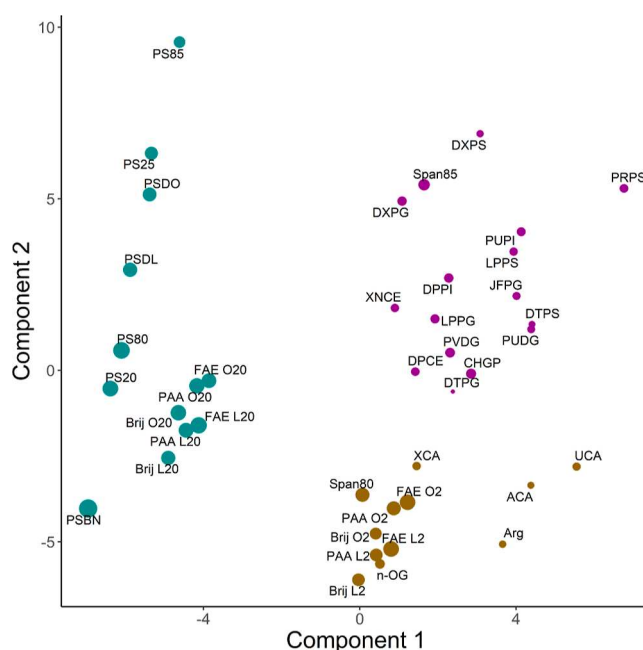


Figure 5. Distribution of the data in the latent variable space using Mordred descriptors as independent variables. Compounds are sized in proportion to the percentage decrease of APR SASA relative to the control and are colored according to their cluster. Cluster 1, teal; Cluster 2, gold; and Cluster 3, magenta.

heteroatoms, 5-membered rings, bond and atom polarizability, topological polar surface area, radius, and complexity, among others. Many of these descriptors have a clear relationship between them, such as the number of oxygen atoms and polar surface area. This specific example could indicate that there is a significant polar component that drives the shielding of the APRs from the solvent. The presence of the Bertz complexity score, a measure of molecular complexity and the distribution of heteroatoms, along with sp^3 carbons bound to a single additional carbon, which in this context is either a terminal carbon or one within a furan ring, implies that greater APR shielding (and therefore enhanced stability) would be achieved by a branched compound with short aliphatic chains, a high degree of complexity and a broad distribution in space of a large number of heteroatoms. This is further supported by the lower impact on decreasing the SASA of APRs of compounds with a high alkyl chain content: phospholipids, Span 85 (three oleates), and glycerols all had little impact on the SASA of the APRs of HSA.

The poor performance of PS85 and Span 85 in particular could indicate an “activity cliff” relationship between APR shielding and aliphatic content, reflected in the positive coordinates in the latent space of descriptors concerning hydrocarbon content for both components in the region of cluster 2. Atom and bond polarizability are both influential in the formation of cluster 1 (negative value for component 1 and positive value for component 2), but mean polarizability is within cluster 2. This apparent discrepancy can be explained by the more highly mixed content of the well-performing ethoxylated compounds, having high molecular weights and their structures comprising polarizable and nonpolarizable bonds and atoms interspersed; conversely, those with a higher mean polarizability and less polarizable bonds and atoms have more chemically compact head groups, consisting of a small number of atoms with a high polarizability, and smaller

hydrophobic tails, leading to a higher ratio of polarizability to molecular weight. This indicates that the shielding of APRs by excipients is better achieved by structures that have larger, but less extreme, polar characters like that which can be achieved by repeating ethoxylate units. Additionally, supporting this hypothesis is the position of the E-state descriptors that describe double-bonded oxygen (SdO and NdO) and the number of acid and base groups (nAcid and nBase). The mean van der Waals volume can also be found in this region, which could also be explained by the presence of bulky head groups in phospholipids, which are also found in this area as opposed to the lower occupational volume of PEG chains. Further evidence supporting this hypothesis is the positions and relative importance of the topological radius, topological polar surface area, and the number of rotatable bonds. Qualitatively, many of these qualities can be found in compounds with high PEG content, and the data reflect the preferential interaction to APRs of polysorbate species and linear compounds with high PEG content.

The impact of PEG content on increasing protein–excipient interaction could be due to entropic effects; longer ethoxylated compounds would have a greater loss of entropy upon burial, as the hydrophobic tail is more readily buried within the hydrophilic head groups. This is indicated by the SASA differential observed between linear PEGylated compounds of differing PEG length. Compounds with 20 PEG units typically have a greater effect on the reduction of the SASA than those with 2 PEG units, despite these simulations having approximately the same quantity of EO beads but different numbers of molecules. These compounds also appear to more readily form intermolecular clusters, independent of the protein, that are reminiscent of micelles. These behaviors are also exhibited by the polysorbate compounds. Together, these behaviors indicate a strong influence of entropy on the interaction between protein APRs and stabilizing excipients. Qualitatively, long PEG chains have a greater tendency to occupy channels on the surface of the protein; these valleys are lined with polar residues, but nonpolar residues typically make up the “floor”. Thus, the larger PEG chains are able to make a large number of polar–polar contacts to reduce the SASA, and their intermediate polarity as Martini beads allows them to occupy these surface channels without a prohibitive degree of repulsion.

DISCUSSION

MD simulations have been employed to investigate the efficacy of excipients as antiaggregation agents and probe the importance of APR interaction as a mechanism for the prevention of biotherapeutic aggregation. The APRs of HSA have been identified using an experimentally derived aggregation propensity score via the Aggrescan web server, and the propensity of an excipient molecule to interact with both the APR and the entirety of HSA has been utilized as an effective demonstration of the APR-shielding mechanism of the arrest of aggregation. Generally, molecules with a high degree of PEG content reduced the SASA of APRs, with little impact from any differences in hydrophobic content within ethoxylated compounds and almost no change between HSA and compounds with high hydrophobic content that lack PEG. This suggests that the interaction between the protein and the polar PEG chains that constitute the headgroup is driving the overall increase in interaction, a finding that is supported by the literature.^{84–86} As protein aggregation is driven primarily

by hydrophobic interactions with contributions from polar interactions,¹³ this could indicate that the increase in polar interaction is contributing to the overall stability of the protein by tipping the scales in the direction of polar interaction and making the hydrophobic destabilizing interactions less significant overall. This notion is further supported by the near total lack of interaction between HSA and the naturally occurring phospholipids under study; compounds with the largest hydrophobic tails and comparatively small head groups have little interaction. Similarly, Spans (in essence, polysorbates lacking PEG) and polysorbates with more than one fatty acid ester, such as PS85 and PS25, perform worse in terms of APR SASA shielding than PSBN, a branched compound with little nonpolar content and a high proportion of PEG content. This also implies that an increase in molecular weight is not sufficient to increase antiaggregation activity, further supported by the absence of impact of molecular weight as a descriptor or as a factor within a descriptor within PLS. Together with the observation that larger molecules have a tendency to wrap around HSA, this could imply that the headgroup initiates the interaction before recruiting the tail in wrapping around more hydrophobic areas of the protein. It can also be seen that longer interacting compounds are making end-to-end contact with each other within a shallow channel on the protein surface (Figure 4A,F). This is reminiscent of binding behavior observed in crystallographic binding studies with short- and medium-chain fatty acids.⁸⁷ Polysorbate 20 and 80 specifically have also been found to interact with HSA, albeit weakly,⁵⁵ which has also been reproduced in this study. The use of Aggrescan, which calculates the average aggregation propensity of sequences based on experimentally derived values for each amino acid in the context of the formation of amyloid plaques,⁸⁸ as the sole indication of APRs could be improved by the inclusion of other methods in a comparative way. One such method would be spatial aggregation propensity (SAP),¹⁷ which considers whether residues are either exposed to the solvent or buried. Using additional methods to flag APRs would ensure a comprehensive approach in finding areas of the protein that are significant in the aggregation process and therefore improve the robustness of the model.

The lack of interaction between HSA and every phospholipid under study is surprising, given HSA's role in transporting fatty acids⁸⁹ and cholesterol⁹⁰ in circulation and studies of its interaction with phospholipid membranes.^{91,92} However, the concentrations of lipid used in the membrane studies are typically significantly greater than those of excipients in the present study; typically, these are millimolar as opposed to 0.1% w/w, which results in concentrations in the range of 0.10–0.18 mM. For all phospholipids with at least 12 carbons in their fat chains, this concentration range is above the critical micellar concentration (CMC);⁹³ the lack of differentiation along the CMC of the compounds under study implies that it is not of critical importance in this context; heavy phospholipids above the CMC perform equally poorly to lighter phospholipids below it, and so, other factors are more significant in determining the extent of interaction. This concentration of 0.1% w/w was chosen to emulate industrial conditions for primarily surfactant excipients used in biotherapeutic stabilization formulations; for other excipients such as those that include sugar residues and arginine, their working concentrations are typically higher.

One limitation of this study is the modeling of polysorbates as homogeneous additives, when in reality, they are typically a

heterogeneous mixture that contains byproducts with ranges of differences in aliphatic and PEG chain lengths and number.^{94,95} This is particularly of note as the heterogeneity of polysorbate commercial products impacts their ability to prevent aggregation; polysorbate fractions vary in their performance in this context.⁹⁶ Therefore, it could be prudent to model polysorbate as a heterogeneous mixture; to maintain concentrations that are industrially relevant, this would likely require the modeling of extremely large systems.

Validation could also be provided in the characterization of excipient effects on protein stability, by monitoring changes in aggregate size, protein secondary and tertiary structures, and biological activity assays. However, the stability of HSA and its own use as an antiaggregation agent⁹⁷ would make reliably inducing (and measurably arresting or preventing) aggregation challenging. This points to a need for a protein-independent model, which would be most easily developed by modeling one or more different therapeutically relevant proteins, ideally with their own stability issues, such as insulin or the binding fragment of an antibody. Additionally, validation of the PLS model can be increased by the introduction of more simulation data, which can be either included in the predictive model or excluded from it and used as a validation test set.

By using techniques to explore latent variable space and probe the physicochemical properties of each excipient and how they correlate with antiaggregation activity, hypotheses on the design of novel excipients with greater APR SASA shielding, and therefore improved performance as antiaggregation agents, can be postulated. Particularly, variable importance in projection (VIP) plots are used in feature selection in drug design⁹⁸ and are a useful tool in investigating the structure–property relationship within a PLS model by indicating the critical descriptors that explain the maximal variance in both dependent and independent variables. An optimized excipient would be a large, branched compound which is highly polar (i.e., with several oxygen atoms) and also of some hydrophobic character. Practically, this could be achieved by the incorporation of multiple PEG chains into the excipient design around a central scaffold and at least one aliphatic chain. This is broadly descriptive of a polysorbate compound, and this is perhaps unsurprising considering their performance, but it also indicates that there is chemical space that is underutilized by the current antiaggregation excipient design paradigm. It implies that the exact degree of hydrophobic content is not significant, provided that there is some present in a localized area in order to provide amphiphilic character to the excipient. The findings suggest that perhaps a compound with a lower molecular weight and a higher number of short branches would be more effective in APR shielding than heavier compounds with a small number of large chains. Such a compound might be achieved by the utilization of an oligopeptide or dendrimer central scaffold, functionalized by multiple short-chain ethoxylation and fatty acid groups on termini and side chains.

CONCLUSIONS

The coarse-grained modeling of HSA with a series of cosolutes has revealed structural and physicochemical features that are highly influential to the prevention of aggregation via APR shielding. Broadly, ethoxylated compounds had the greatest performance as APR-shielding antiaggregation agents, and polysorbate species specifically were the highest performing class. Branched compounds tended to make greater contact

with APRs, particularly those with PEG chains, while phospholipids and fatty acids performed very poorly in shielding APRs from solvent and thereby preventing aggregation. The use of dimensionality reduction coupled with physicochemical descriptors has revealed structural features that are key to optimizing protein–excipient interaction. The overall weight of aliphatic chains does not appear to influence the performance of antiaggregation agents, provided that some is present. The significance of polarity, polarizability, and polar heteroatom content in predicting HSA interaction also suggests that the interaction between APRs and excipients is driven by polar interactions to a significant degree. The quantitative model would be well-supported by future endeavors that elucidate free energy differences, provide validation via wet-lab work or atomistic MD, and move away from a singular protein to develop a more widely applicable, predictive model to aid in computational excipient design and improve the stability of biotherapeutic formulations.

ASSOCIATED CONTENT

Data Availability Statement

SASA data can be found at the GitHub repository (<https://github.com/TobyEdwardKing/Excipient-Optimisation>), as can the descriptor data and the compounds' SMILES. Gromacs is a freely available software package for molecular dynamics, and details on its installation can be found on their Web site: www.gromacs.org. The following packages in R were used in the extraction of data, development of the model, and generation of figures: Peptides, scico, tidyverse, ggpubr, pls, webchem, rcdk, and vip. All are freely available from the CRAN repository. Mordred, a Python package, was used to extract quantitative structure–property activity information from SMILES structures, in conjunction with rdkit, numpy, and pandas, and all can be retrieved freely. Some molecular dynamics graphics were created with VMD, freely available from <http://www.ks.uiuc.edu/Research/vmd>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.3c01898>.

Bash script for molecular dynamics simulations and their subsequent analysis, parameter files for excipient compounds, structural files for excipient compounds, structural depictions of excipients under study, R script detailing PLS model, raw data in an archive, Mordred data input for the PLS model, and data table of excipient SMILES and associated data (PDF)

AUTHOR INFORMATION

Corresponding Author

Jonathan D. Hirst – School of Chemistry, Nottingham NG7 2RD, U.K.; orcid.org/0000-0002-2726-0983;
Email: jonathan.hirst@nottingham.ac.uk

Authors

Toby E. King – Biodiscovery Institute, School of Pharmacy, Nottingham NG7 2RD, U.K.; orcid.org/0009-0004-2020-311X

James R. Humphrey – Croda Europe Ltd, Snaith DN14 9AA, U.K.

Charles A. Loughton – Biodiscovery Institute, School of Pharmacy, Nottingham NG7 2RD, U.K.; orcid.org/0000-0003-4090-3960

Neil R. Thomas – *Biodiscovery Institute, School of Chemistry, Nottingham NG7 2RD, U.K.*; orcid.org/0000-0002-9260-5423

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.3c01898>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Calculations were performed using the Sulis Tier 2 HPC platform hosted by the Scientific Computing Research Technology Platform at the University of Warwick. Sulis is funded by EPSRC Grant EP/T022108/1 and the HPC Midlands + consortium. We are also grateful for access to the University of Nottingham Augusta high performance computing (HPC) services as well as the School of Pharmacy HPC services. J.D.H. is supported by the Department of Science, Innovation and Technology (DSIT) and the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme. We thank the EPSRC for funding the Centre for Doctoral Training in Transformative Pharmaceutical Technologies (EP/S023054/1).

REFERENCES

- (1) de la Torre, B. G.; Albericio, F. The Pharmaceutical Industry in 2021. An Analysis of FDA Drug Approvals from the Perspective of Molecules. *Molecules* **2022**, *27* (3), 1075.
- (2) Perrin, C.; Ewen, M.; Beran, D. The role of biosimilar manufacturers in improving access to insulin globally. *Lancet Diabetes Endocrinol.* **2017**, *5*, 578.
- (3) Davies, J. R.; Micucci, V.; Baker, D.; Turner, P. J. Large scale manufacture and properties of chromatographically purified albumin for therapeutic use. *Australas. Biotechnol.* **1996**, *6*, 100–102.
- (4) Clarkson, B. R.; Schon, A.; Freire, E. Conformational stability and self-association equilibrium in biologics. *Drug Discovery Today* **2016**, *21*, 342–347.
- (5) Casademunt, E.; Martinelle, K.; Jernberg, M.; Winge, S.; Tiemeyer, M.; Biesert, L.; Knaub, S.; Walter, O.; Schroder, C. The first recombinant human coagulation factor VIII of human origin: human cell line and manufacturing characteristics. *Eur. J. Haematol.* **2012**, *89*, 165–176.
- (6) Pardi, N.; Hogan, M. J.; Porter, F. W.; Weissman, D. mRNA vaccines - a new era in vaccinology. *Nat. Rev. Drug Discovery* **2018**, *17*, 261–279.
- (7) dos Santos, M. L.; Quintilio, W.; Manieri, T. M.; Tsuruta, L. R.; Moro, A. M. Advances and challenges in therapeutic monoclonal antibodies drug development. *Braz. J. Pharm. Sci.* **2018**, *54*, 15.
- (8) Lagassé, H. D.; Alexaki, A.; Simhadri, V. L.; Katagiri, N. H.; Jankowski, W.; Sauna, Z. E.; Kimchi-Sarfaty, C. Recent advances in (therapeutic protein) drug development. *F1000Res.* **2017**, *6*, 113.
- (9) Zahavi, D.; Weiner, L. Monoclonal Antibodies in Cancer Therapy. *Antibodies (Basel)* **2020**, *9*, 34.
- (10) Saxena, M.; van der Burg, S. H.; Melief, C. J. M.; Bhardwaj, N. Therapeutic cancer vaccines. *Nat. Rev. Cancer* **2021**, *21*, 360–378.
- (11) Butreddy, A.; Janga, K. Y.; Ajjarapu, S.; Sarabu, S.; Dudhipala, N. Instability of therapeutic proteins — An overview of stresses, stabilization mechanisms and analytical techniques involved in lyophilized proteins. *Int. J. Biol. Macromol.* **2021**, *167*, 309–325.
- (12) Vazquez-Rey, M.; Lang, D. A. Aggregates in Monoclonal Antibody Manufacturing Processes. *Biotechnol. Bioeng.* **2011**, *108*, 1494–1508.
- (13) Dobson, C. M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **2004**, *15*, 3–16.
- (14) Ratanji, K. D.; Derrick, J. P.; Dearman, R. J.; Kimber, I. Immunogenicity of therapeutic proteins: Influence of aggregation. *J. Immunotoxicol.* **2014**, *11*, 99–109.
- (15) Maciążek-Jurczyk, M.; Janas, K.; Pożycka, J.; Szkudlarek, A.; Rogóż, W.; Owczarzy, A.; Kulig, K. Human Serum Albumin Aggregation/Fibrillation and its Abilities to Drugs Binding. *Molecules* **2020**, *25*, 618.
- (16) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (17) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. Prediction of Aggregation Prone Regions of Therapeutic Proteins. *J. Phys. Chem. B* **2010**, *114*, 6614–6624.
- (18) Roberts, C. J. Non-native protein aggregation kinetics. *Biotechnol. Bioeng.* **2007**, *98*, 927–938.
- (19) Tedeschi, G.; Mangiagalli, M.; Chmielewska, S.; Lotti, M.; Natalello, A.; Brocca, S. Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue. *Biochim. Biophys. Acta. Gen. Subj.* **2017**, *1861*, 2543–2550.
- (20) 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*, 4830–4848.
- (21) Vasilescu, M.; Angelescu, D.; Almgren, M.; Valstar, A. Interactions of globular proteins with surfactants studied with fluorescence probe methods. *Langmuir* **1999**, *15*, 2635–2643.
- (22) Al-hussein, A.; Gieseler, H. Investigation of histidine stabilizing effects on LDH during freeze-drying. *J. Pharm. Sci.* **2013**, *102*, 813–826.
- (23) Shukla, D.; Trout, B. L. Interaction of Arginine with Proteins and the Mechanism by Which It Inhibits Aggregation. *J. Phys. Chem. B* **2010**, *114*, 13426–13438.
- (24) Leblanc, Y.; Bihoreau, N.; Jube, M.; Andre, M.-H.; Tellier, Z.; Chevreux, G. Glycation of polyclonal IgGs: Effect of sugar excipients during stability studies. *Eur. J. Pharm. Biopharm.* **2016**, *102*, 185–190.
- (25) Yue, L. Y.; Yan, Z.; Li, H.; Liu, X.; Sun, P. Y. Brij-58, a potential injectable protein-stabilizer used in therapeutic protein formulation. *Eur. J. Pharm. Biopharm.* **2020**, *146*, 73–83.
- (26) Kale, S. S.; Akamanchi, K. G. Trehalose Monooleate: A Potential Antiaggregation Agent for Stabilization of Proteins. *Mol. Pharm.* **2016**, *13*, 4082–4093.
- (27) Kannan, A.; Shieh, I. C.; Fuller, G. G. Linking aggregation and interfacial properties in monoclonal antibody-surfactant formulations. *J. Colloid Interface Sci.* **2019**, *550*, 128–138.
- (28) Kerwin, B. A. Polysorbates 20 and 80 used in the formulation of protein biotherapeutics: Structure and degradation pathways. *J. Pharm. Sci.* **2008**, *97*, 2924–2935.
- (29) Arsiccio, A.; Pisano, R. Surfactants as stabilizers for biopharmaceuticals: An insight into the molecular mechanisms for inhibition of protein aggregation. *Eur. J. Pharm. Biopharm.* **2018**, *128*, 98–106.
- (30) Lee, H. J.; McAuley, A.; Schilke, K. F.; McGuire, J. Molecular origins of surfactant-mediated stabilization of protein drugs. *Adv. Drug Delivery Rev.* **2011**, *63*, 1160–1171.
- (31) Yano, Y. F. Kinetics of protein unfolding at interfaces. *J. Phys.: Condens. Matter* **2012**, *24*, S03101.
- (32) Ghazvini, S.; Kalonia, C.; Volkin, D. B.; Dhar, P. Evaluating the Role of the Air-Solution Interface on the Mechanism of Subvisible Particle Formation Caused by Mechanical Agitation for an IgG1 mAb. *J. Pharm. Sci.* **2016**, *105*, 1643–1656.
- (33) Saha, D.; Ray, D.; Kohlbrecher, J.; Aswal, V. K. Unfolding and Refolding of Protein by a Combination of Ionic and Nonionic Surfactants. *ACS Omega* **2018**, *3*, 8260–8270.
- (34) Kuriata, A.; Iglesias, V.; Pujols, J.; Kurcinski, M.; Kmiecik, S.; Ventura, S. Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. *Nucleic Acids Res.* **2019**, *47*, W300–W307.
- (35) Lanigan, R. S. Final report on the safety assessment of PEG-20 sorbitan cocoate; PEG-40 sorbitan diisostearate; PEG-2,-5, and-20 sorbitan isostearate; PEG-40 and-75 sorbitan lanolate; PEG-10,-40,-

- 44,-75, and-80 sorbitan laurate; PEG-3, and-6 sorbitan oleate; PEG-80 sorbitan palmitate; PEG-40 sorbitan perisostearate; PEG-40 sorbitan peroleate; PEG-3,-6,-40, and-60 sorbitan stearate; PEG-20,-30,-40, and-60 sorbitan tetraoleate; PEG-60 sorbitan tetrastearate; PEG-20 and-160 sorbitantriostearate; PEG-18 sorbitan trioleate; PEG-40 and-50 sorbitol hexaoleate; PEG-30 sorbitol tetraoleate laurate; and PEG-60 sorbitol tetrastearate - Addendum to the final report on the safety assessment of polysorbates. *Int. J. Toxicol.* **2000**, *19*, 43–89.
- (36) Fratter, A.; Semenzato, A. New association of surfactants for the production of food and cosmetic nanoemulsions: preliminary development and characterization. *Int. J. Cosmet. Sci.* **2011**, *33*, 443–449.
- (37) Goff, H. D. Colloidal aspects of ice cream - A review. *Int. Dairy J.* **1997**, *7*, 363–373.
- (38) Gelpi, J.; Hospital, A.; Goñi, R.; Orozco, M. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.
- (39) Hossain, M. S.; Berg, S.; Bergstrom, C. A. S.; Larsson, P. Aggregation Behavior of Medium Chain Fatty Acids Studied by Coarse-Grained Molecular Dynamics Simulation. *AAPS PharmSciTech* **2019**, *20*, 61.
- (40) Morrow, B. H.; Koenig, P. H.; Shen, J. K. Self-Assembly and Bilayer-Micelle Transition of Fatty Acids Studied by Replica-Exchange Constant pH Molecular Dynamics. *Langmuir* **2013**, *29*, 14823–14830.
- (41) Li, J.; Chu, M. K.; Lu, B.; Mirzaie, S.; Chen, K.; Gordijo, C. R.; Plettenburg, O.; Giacca, A.; Wu, X. Y. Enhancing thermal stability of a highly concentrated insulin formulation with Pluronic F-127 for long-term use in microfabricated implantable devices. *Drug Delivery Transl. Res.* **2017**, *7*, 529–543.
- (42) Valoerdi, F. M.; Farasat, A.; Shariatifar, H.; Gheibi, N. Study of HSA interactions with arachidonic acid using spectroscopic methods revealing molecular dynamics of HSA-AA interactions. *Biomed. Rep.* **2020**, *12*, 125–133.
- (43) Qi, R. X.; Wei, G. H.; Ma, B. Y.; Nussinov, R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Methods Mol. Biol.* **2018**, *1777*, 101–119.
- (44) Meli, M.; Colombo, G. A Hamiltonian Replica Exchange Molecular Dynamics (MD) Method for the Study of Folding, Based on the Analysis of the Stabilization Determinants of Proteins. *Int. J. Mol. Sci.* **2013**, *14*, 12157–12169.
- (45) Saurabh, S.; Kalonia, C.; Li, Z.; Hollowell, P.; Waigh, T.; Li, P.; Webster, J.; Seddon, J. M.; Lu, J. R.; Bresme, F. Understanding the Stabilizing Effect of Histidine on mAb Aggregation: A Molecular Dynamics Study. *Mol. Pharm.* **2022**, *19*, 3288–3303.
- (46) Otsu, T.; Ishii, K.; Tahara, T. Microsecond protein dynamics observed at the single-molecule level. *Nat. Commun.* **2015**, *6*, 7685.
- (47) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (48) Grunewald, F.; Rossi, G.; de Vries, A. H.; Marrink, S. J.; Monticelli, L. Transferable MARTINI Model of Poly(ethylene Oxide). *J. Phys. Chem. B* **2018**, *122*, 7436–7449.
- (49) Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2278–2291.
- (50) Lui, L. H.; van der Walle, C. F.; Brocchini, S.; Velayudhan, A. Discovering Novel Small Molecule Compound for Prevention of Monoclonal Antibody Self-Association. *Antibodies* **2022**, *11*, 40.
- (51) Mishra, R. P.; Goel, G. Multiscale Model for Quantitative Prediction of Insulin Aggregation Nucleation Kinetics. *J. Chem. Theory Comput.* **2021**, *17*, 7886–7898.
- (52) González-Rodríguez, M. V.; Dopico-García, M. S.; Nogueroles-Cal, R.; Carballeira-Amarelo, T.; López-Vilariño, J. M.; Fernández-Martínez, G. Application of liquid chromatography in polymer non-ionic antistatic additives analysis. *J. Sep. Sci.* **2010**, *33*, 3595–3603.
- (53) Roberts, I.; Blackhall, K.; Alderson, P.; Bunn, F.; Schierhout, G. Human albumin solution for resuscitation and volume expansion in critically ill patients. *Cochrane Database Sys. Rev.* **2011**, *2011*, CD001208.
- (54) Strickley, R. G.; Lambert, W. J. A review of formulations of commercially available antibodies. *J. Pharm. Sci.* **2021**, *110*, 2590–2608.e56.
- (55) Garidel, P.; Hoffmann, C.; Blume, A. A thermodynamic analysis of the binding interaction between polysorbate 20 and 80 with human serum albumins and immunoglobulins: A contribution to understand colloidal protein stabilisation. *Biophys. Chem.* **2009**, *143*, 70–78.
- (56) Bam, N. B.; Cleland, J. L.; Yang, J.; Manning, M. C.; Carpenter, J. F.; Kelley, R. F.; Randolph, T. W. Tween protects recombinant human growth hormone against agitation-induced damage via hydrophobic interactions. *J. Pharm. Sci.* **1998**, *87*, 1554–1559.
- (57) Singh, S. M.; Bandi, S.; Jones, D. N. M.; Mallela, K. M. G. Effect of Polysorbate 20 and Polysorbate 80 on the Higher-Order Structure of a Monoclonal Antibody and Its Fab and Fc Fragments Probed Using 2D Nuclear Magnetic Resonance Spectroscopy. *J. Pharm. Sci.* **2017**, *106*, 3486–3498.
- (58) Chang, P. K.; Prestidge, C. A.; Barnes, T. J.; Bremmell, K. E. Impact of PEGylation and non-ionic surfactants on the physical stability of the therapeutic protein filgrastim (G-CSF). *RSC Adv.* **2016**, *6*, 78970–78978.
- (59) Kim, H. L.; McAuley, A.; McGuire, J. Protein Effects on Surfactant Adsorption Suggest the Dominant Mode of Surfactant-Mediated Stabilization of Protein. *J. Pharm. Sci.* **2014**, *103*, 1337–1345.
- (60) Joshi, O.; Chu, L. P.; McGuire, J.; Wang, D. Q. Adsorption and Function of Recombinant Factor VIII at the Air-Water Interface in the Presence of Tween 80. *J. Pharm. Sci.* **2009**, *98*, 3099–3107.
- (61) Wang, Z. M.; Ho, J. X.; Ruble, J. R.; Rose, J.; Ruker, F.; Ellenburg, M.; Murphy, R.; Click, J.; Soistman, E.; Wilkerson, L.; Carter, D. C. Structural studies of several clinically important oncology drugs in complex with human serum albumin. *Biochim. Biophys. Acta Gen. Subj.* **2013**, *1830*, 5356–5374.
- (62) Kroon, P. C.; Grünewald, F.; Barnoud, J.; van Tilburg, M.; Souza, P. C. T.; Wassenaar, T. A.; Marrink, S.-J. Martinize2 and Vermouth: Unified Framework for Topology Generation. *arXiv* **2022**, arXiv:2212.01191. <https://arxiv.org/abs/2212.01191>.
- (63) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037.
- (64) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions S4A7 and S4B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.
- (65) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144–2155.
- (66) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- (67) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (68) Melo, M. N.; Ingólfsson, H. I.; Marrink, S. J. Parameters for Martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **2015**, *143*, 243152.
- (69) Grünewald, F.; Alessandri, R.; Kroon, P. C.; Monticelli, L.; Souza, P. C. T.; Marrink, S. J. PolyPy; a python suite for facilitating simulations of macromolecules and nanomaterials. *Nat. Commun.* **2022**, *13*, 68.

- (70) Haynes, G. R.; Bassiri, K. Hyper-oncotic vs. Hypo-oncotic Albumin Solutions: a Systematic Review of Clinical Efficacy and Safety. *SN Compr. Clin. Med.* **2021**, *3*, 1137–1147.
- (71) Ohtake, S.; Kita, Y.; Arakawa, T. Interactions of formulation excipients with proteins in solution and in the dried state. *Adv. Drug Delivery Rev.* **2011**, *63*, 1053–1073.
- (72) Castañeda Ruiz, A. J.; Shetab Boushehri, M. A.; Phan, T.; Carle, S.; Garidel, P.; Buske, J.; Lamprecht, A. Alternative Excipients for Protein Stabilization in Protein Therapeutics: Overcoming the Limitations of Polysorbates. *Pharmaceutics* **2022**, *14*, 2575.
- (73) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.
- (74) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (75) Chao, M.-T.; Glaser, R. E. The Exact Distribution of Bartlett's Test Statistic for Homogeneity of Variances with Unequal Sample Sizes. *J. Am. Stat. Assoc.* **1978**, *73*, 422–426.
- (76) Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621.
- (77) Jan, S. L.; Shieh, G. Sample size determinations for Welch's test in one-way heteroscedastic ANOVA. *Br J. Math Stat Psychol* **2014**, *67*, 72–93.
- (78) Stanton, D. T. QSAR and QSPR Model Interpretation Using Partial Least Squares (PLS) Analysis. *Curr. Comput.-Aided Drug Des.* **2012**, *8*, 107–127.
- (79) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminform.* **2018**, *10*, 4.
- (80) Ramírez-Galicia, G.; Garduño-Juárez, R.; Hemmateenejad, B.; Deeb, O.; Deciga-Campos, M.; Moctezuma-Eugenio, J. C. QSAR Study on the Antinociceptive Activity of Some Morphinans. *Chem. Biol. Drug Des.* **2007**, *70*, 53–64.
- (81) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (82) Lin, Z.; Zhou, B.; Wu, W.; Xing, L.; Zhao, Q. Self-assembling amphipathic alpha-helical peptides induce the formation of active protein aggregates in vivo. *Faraday Discuss.* **2013**, *166*, 243–256.
- (83) Kunjithapatham, R.; Oliva, F. Y.; Doshi, U.; Perez, M.; Avila, J.; Munoz, V. Role for the alpha-helix in aberrant protein aggregation. *Biochem* **2005**, *44*, 149–156.
- (84) Pongprayoon, P.; Gleeson, M. P. Probing the binding site characteristics of HSA: a combined molecular dynamics and cheminformatics investigation. *J. Mol. Graph. Model.* **2014**, *54*, 164–173.
- (85) Guizado, T. R. Analysis of the structure and dynamics of human serum albumin. *J. Mol. Model.* **2014**, *20*, 2450.
- (86) Paris, G.; Ramseyer, C.; Enescu, M. A Principal Component Analysis of the Dynamics of Subdomains and Binding Sites in Human Serum Albumin. *Biopolymers* **2014**, *101*, 561–572.
- (87) Bhattacharya, A. A.; Grüne, T.; Curry, S. Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin 1 I Edited by R. Huber. *J. Mol. Biol.* **2000**, *303*, 721–732.
- (88) Conchillo-Sole, O.; de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Daura, X.; Ventura, S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinf.* **2007**, *8*, 65.
- (89) van der Vusse, G. J. Albumin as fatty acid transporter. *Drug Metab. Pharmacokinet.* **2009**, *24*, 300–307.
- (90) Sankaranarayanan, S.; de la Llera-Moya, M.; Drazul-Schrader, D.; Phillips, M. C.; Kellner-Weibel, G.; Rothblat, G. H. Serum albumin acts as a shuttle to enhance cholesterol efflux from cells. *J. Lipid Res.* **2013**, *54*, 671–676.
- (91) Thakur, R.; Das, A.; Chakraborty, A. Interaction of human serum albumin with liposomes of saturated and unsaturated lipids with different phase transition temperatures: a spectroscopic investigation by membrane probe PRODAN. *RSC Adv.* **2014**, *4*, 14335–14347.
- (92) van de Wouw, J.; Joles, J. A. Albumin is an interface between blood plasma and cell membrane, and not just a sponge. *Clin. Kidney J.* **2022**, *15*, 624–634.
- (93) Marsh, D. *Handbook of Lipid Bilayers*, array ed.; CRC Press: Boca Raton, FL, 2013; p 1145.
- (94) Tomlinson, A.; Zarraga, I. E.; Demeule, B. Characterization of Polysorbate Ester Fractions and Implications in Protein Drug Product Stability. *Mol. Pharm.* **2020**, *17*, 2345–2353.
- (95) Konya, Y.; Ochiai, R.; Fujiwara, S.; Tsujino, K.; Okumura, T. Profiling polysorbate 80 components using comprehensive liquid chromatography–tandem mass spectrometry analysis. *Rapid Commun. Mass Spectrom.* **2023**, *37*, No. e9438.
- (96) Diederichs, T.; Mittag, J. J.; Humphrey, J.; Voss, S.; Carle, S.; Buske, J.; Garidel, P. Existence of a superior polysorbate fraction in respect to protein stabilization and particle formation? *Int. J. Pharm.* **2023**, *635*, 122660.
- (97) Kamerzell, T. J.; Esfandiary, R.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B. Protein–excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. *Adv. Drug Delivery Rev.* **2011**, *63*, 1118–1159.
- (98) Mehmood, T.; Sæbø, S.; Liland, K. H. Comparison of variable selection methods in partial least squares regression. *J. Chemom.* **2020**, *34*, No. e3226.