

Optimising excipient properties to prevent aggregation in biopharmaceutical formulations

Toby E. King,[†] James R. Humphrey,[‡] Charles A. Laughton,[†] Neil R. Thomas,[¶]
and Jonathan D. Hirst^{*,§}

[†]*Biodiscovery Institute, School of Pharmacy, University Park, Nottingham, NG7 2RD, UK*

[‡]*Croda Europe Ltd, Cowick Hall, Snaith, DN14 9AA, UK*

[¶]*Biodiscovery Institute, School of Chemistry, University Park, Nottingham, NG7 2RD, UK*

[§]*School of Chemistry, University Park, Nottingham, NG7 2RD, UK*

E-mail: jonathan.hirst@nottingham.ac.uk

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 structure and anti-aggregation activity of excipients by utilising coarse-grained molecular dynamics modelling of protein-excipient interaction. We have studied human serum albumin as a model protein, and 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. Physicochemical descriptors generated by Mordred are

employed to probe the structure-activity 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. Generally, longer molecules with a large number of alcohol-terminated PEG units tended to interact more, with qualitatively different protein interaction, wrapping around the protein, and would have a lesser protective effect on stability. 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 five to ten 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 and quantitative modelling is an important first step in an all-purpose protein-independent model for the computer-aided design of stabilising excipients.

Introduction

Protein-based biotherapeutics are a growing market, with significantly more treatment options based on biologics under development and a multi-billion 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 delivery⁹ and cancer vaccines.¹⁰ Generally, therapeutic proteins are produced in bioreactors using recombinant cell lines,¹¹ and are often lyophilised 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 resuspending.¹ These aggregates have reduced function¹³ and an increased specific immune response when administered;¹⁴ indeed, the association constant of 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 non-physiological pH, ionic strength, extremes of temperature, interactions with impurities, 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 structure may expose hydrophobic residues, which form patches on the surface of the protein.¹⁷ The energy landscape changes; it becomes more favourable to bury the hydrophobic patches by interacting 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 modifica-

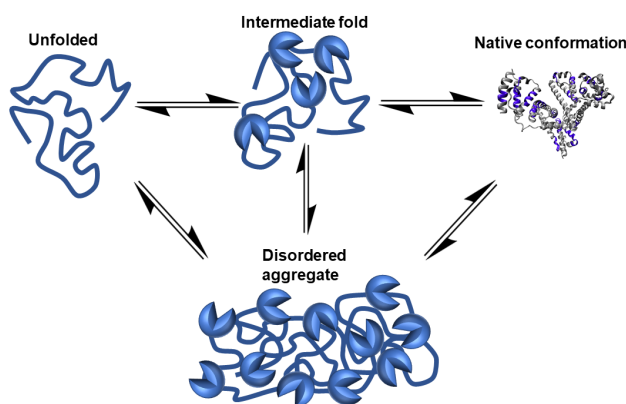


Figure 1: Folding and misfolding pathway of a protein. An unfolded protein assumes more stable intermediate folding conformations until arriving at the native configuration. If subjected to non-native conditions, the equilibrium position can change to favour the formation of a stable, disordered conformation,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

tion 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 poloxamers²⁷ 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 favourable and acting as a chaperone to facilitate native folding.³³

Differences in protein structure complicate the understanding of aggregation prevention; as proteins' structures differ, so too will their aggregation propensity, as well as their interaction with anti-aggregation 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 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 SAP³⁴ and Aggrescan-3D.³⁵ Generally, excipients are not chosen in light of efficacy as an anti-aggregation agent, but due to their well-established safety profiles from other uses;³⁶ for example, polysorbates are popular emulsifiers, particularly in cosmetics³⁷ and in the food industry.³⁸ Therefore, there is chemical space to explore to optimise anti-aggregation excipients.

Computational techniques can provide mechanistic insight at length- and timescales that are inaccessible to conventional wet lab methods.³⁹ Molecular dynamics (MD) simulations have been applied in the study of surfactant behaviour in solution;^{40,41} protein-surfactant interac-

tion,²⁹ including stability modulation⁴² and binding;⁴³ protein aggregation⁴⁴ and folding;⁴⁵ and the modulation of protein stability by excipients such as histidine.⁴⁶ Atomistic or pseudoatomistic MD models often have a prohibitively high computational expense to be applied in large numbers of simulations that examine microsecond-timescale events, such as many aspects of protein dynamics.⁴⁷

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-activity relationship model has been derived that probes the relationship between excipient structure and anti-aggregation activity. In this work, we present an MD model that investigates APR-excipient interaction to determine the stabilising effect on protein biotherapeutics, coupled with a quantitative model which uses physicochemical descriptors in statistical analysis to reveal the impact of key features on anti-aggregation activity. In doing so, we investigate the model of the shielding of APRs from solvent as a mechanism of aggregation prevention, hypothesising that a smaller SASA of APRs leads to greater stability. To produce sufficient data for a quantitative model, a CG force field was selected, as they allow access to microsecond simulation timescales at reasonable computational expense and without the need for enhanced sampling methods. Coarse-grained (CG) force fields decrease 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 timescales, as there are fewer degrees of freedom to consider.

MARTINI⁴⁸ 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 behaviour, 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.* utilised 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 behaviour 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 industry as solution state stability enhancers, their prior parameterisation 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 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 modelling and allows data-driven decisions to be made in the design of anti-aggregation excipients. The application of these data could improve biotherapeutic formulation design by lowering costs, improving therapeutic outcomes, and elucidating structure-activity relationships.

Materials and Methods

Human serum albumin (HSA) was selected as a model protein. The initial structure was obtained from the RCSB Protein Data Bank (code 4L8U)⁵³ and processed into the MARTINI force field via the martinize2 script, from the vermouth package.⁵⁴ 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.

Parameterisation

To parameterise excipients that are not available from MARTINI, initial united-atom coordinates and topologies were generated using the Automated Topology Builder⁵⁵ in the GRO-MOS 54a7 force field,⁵⁶ and converted into a MARTINI model. The MARTINI mapping was based on existing MARTINI beads, their use in the literature, as well as the preservation and representation of functional groups (Figure 2). Molecule parameters reported in previous work by the MARTINI group and used here include phospholipids, ceramides, and glycerols,⁵⁷ as well as sugars,⁵⁸ fatty acids,⁵⁹ and sterol groups.⁶⁰

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,

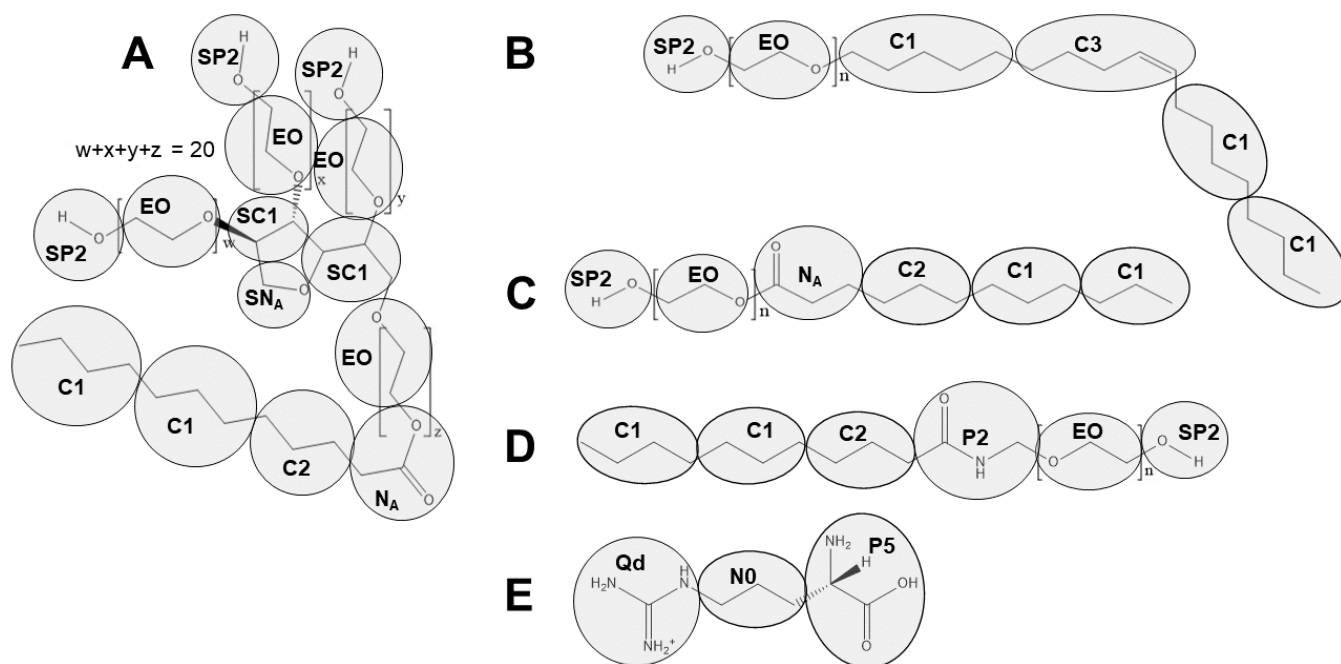


Figure 2: Chemical structures of studied excipients, with their MARTINI mapping superimposed. A: polysorbate 20. B: fatty alcohol ethoxylate (BrijTM On). C: fatty acid ethoxylate Ln. D: PEG alkyl amide Ln. E: L-arginine.

and the bond lengths and angles 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. Polyply⁶¹ was also used to generate initial MARTINI topologies for some compounds.

Molecular dynamics

All MD simulations were carried out using GROMACS 2019 and 2021.2 in the Martini 2.3P forcefield 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 equalling 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 formulations are typically between 5% and 25% w/v.⁶² 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,^{63,64} 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 minimised for 1000 steps using the steepest descent algorithm and solvated using the MARTINI polarisable water model.⁶⁵ Sodium ions were added to neutralise the system by replacing water molecules at random and the system was minimised again for 1000 steps. The system is relaxed in the NPT ensemble, with a 5 fs timestep, V-scale thermostat at 300 K, and isotropic

pressure coupling at 1.0 bar using the Berendsen barostat.⁶⁶ This relaxation phase consisted of 100 ps. Production MD was performed in the same ensemble, with the same thermo- and barostats, a timestep of 20 fs and a total time of 1 μ s. For some compounds, particularly those with ring structures, a timestep of 10 fs was necessary to run stable MD; the overall time remained 1 μ s. Coulombic and Lennard-Jones cut-offs were 1.1 nm and used the reaction field and potential shift Verlet modifiers respectively in the Verlet cut-off scheme. Full parameter files can be found at the Github repository (see Supplementary Information).

The SASA of the APRs were calculated using `gmx sasa` within Gromacs, indexed to calculate the SASA of APRs alone, using lone HSA as a control. Bartlett's test⁶⁷ was utilised to indicate homoscedasticity between distributions for each excipient-protein simulation, and the results directed whether the Kruskal-Wallis⁶⁸ (homoscedastic) or Welch's⁶⁹ (heteroscedastic) analyses of variance were employed to determine statistical significance. All analysis scripts can be found in the GitHub repository (see Supplementary Information).

Machine learning

To probe the structure-activity relationship of anti-aggregation activity, partial least squares (PLS)⁷⁰ regression was performed, using a set of physicochemical descriptors as input. Molecular descriptors were generated using the Mordred package⁷¹ 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 dataset, employing leave-one-out cross validation⁷² (LOO-CV) to find the optimal number of components to include in the model. This is achieved by using a number of components that causes 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 dataset was split into a partition of 0.8/0.2 training data/ test data. LOO-CV was performed on the training dataset, the model was

applied to predict the test dataset 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 average 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 that 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-Wallis and Dunn tests. PSBN, the strongest performer, is significantly different to Brij L2 ($p < 0.05$), Brij O2 ($p < 0.05$), and PS85 ($p < 0.01$). PS80 is significantly different to 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 to 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 to 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, both within 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

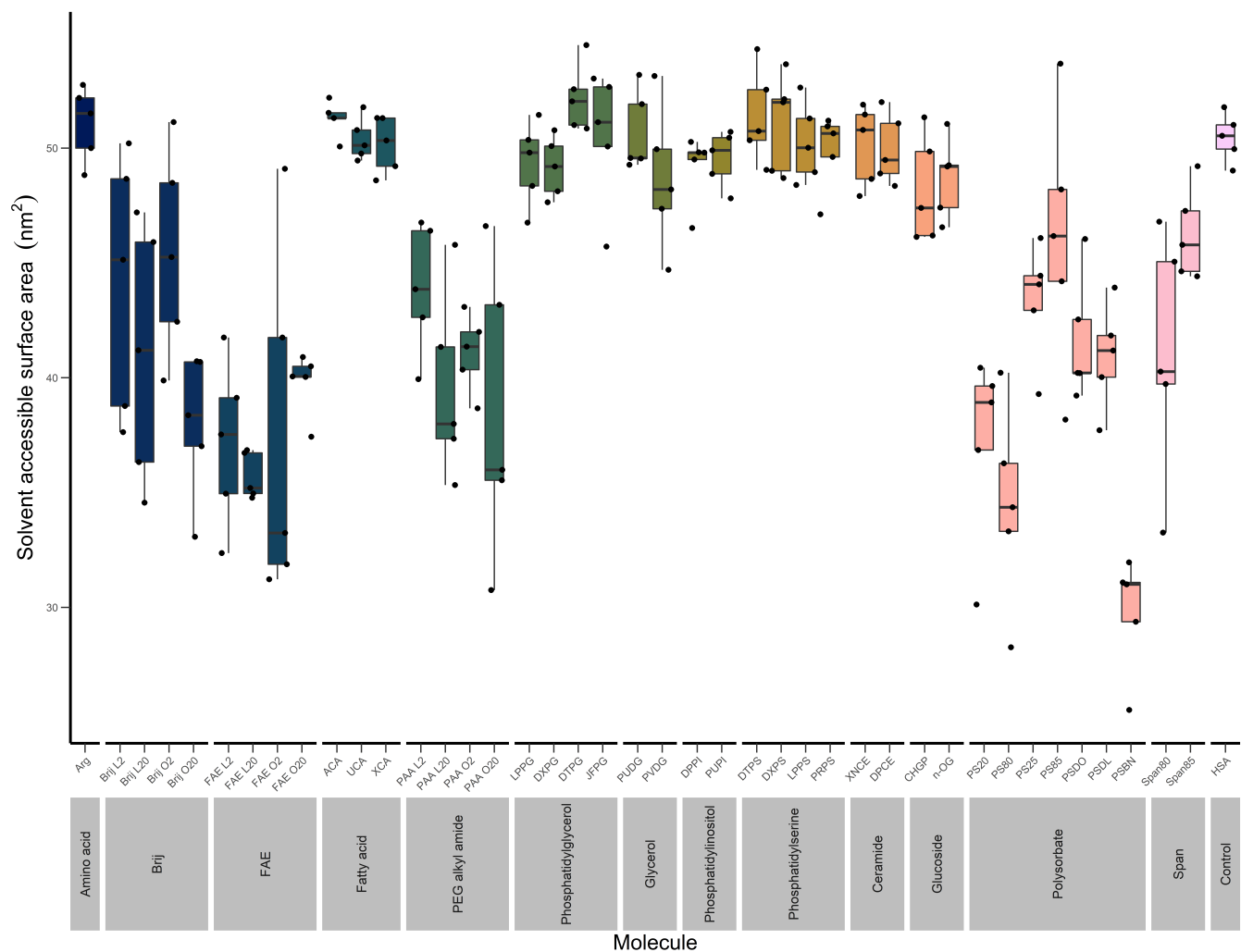


Figure 3: The average solvent accessible surface area (SASA) of aggregation-prone regions (APRs), averaged for each trajectory. Polysorbates have the greatest impact on the SASA of APRs. Of the linear, ethoxylated surfactants, there is little significant difference between them, across all classes, but they are all significantly different to 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.

might not guarantee the formation of a HSA- excipient complex within the simulation time.

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.⁷³ Therefore, an alanine residue within an α -helix will have significantly less hydrophobic character than one outside of 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 helices and solvent is significant in aggregation prevention. Indeed, α -helices have been shown to induce the formation of protein aggregates.^{74,75}

Visual inspection of the trajectories can also reveal characteristics of 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 localised, 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 inspecting 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.

Machine learning and structure-activity relationship

The final PLS model of two components, validated with LOOCV, has an R^2 value of 0.398 and mean relative error of prediction of 0.077. To gain an understanding of the robustness of the dataset and its validity in regression, the dataset of excipient simulations was split 0.8/0.2 training dataset/ testing dataset, and the Q^2 was 0.344, with median root mean square errors of 4.10 nm² and 4.37 nm² for the training and test sets respectively. These distributions of measure of goodness of fit gives confidence that there is sufficient variation within the dataset for its utility in a quantitative structure-activity 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 relatively large amount of PEG that performed

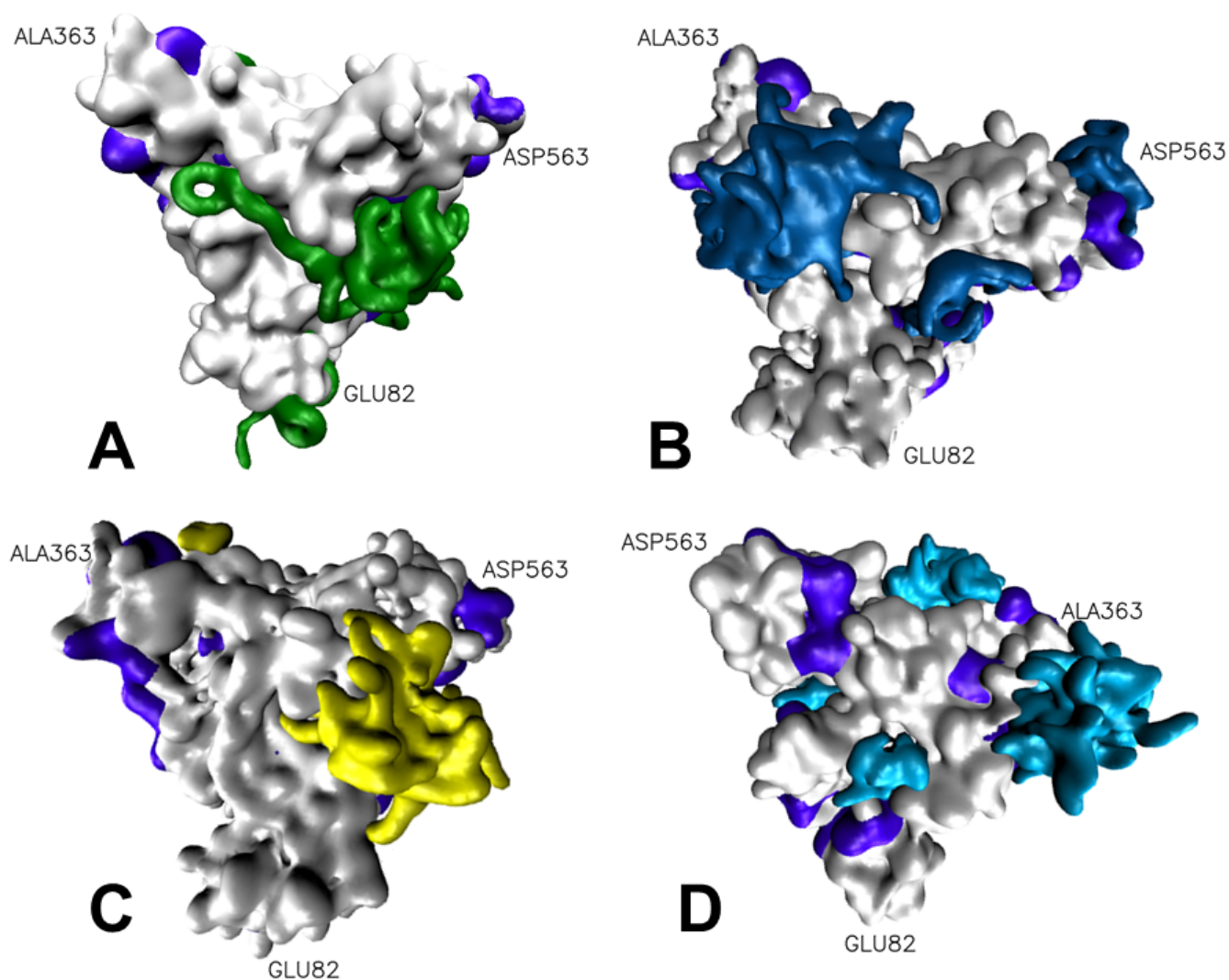


Figure 4: Select snapshots of trajectories at the end point of simulation, with the water removed for clarity. HSA is in grey with its APRs coloured in violet. Glu82, Ala363 and Asp563 are labelled for orientation. Longer molecules wrap around the protein while smaller molecules form clusters at the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

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 non-interacting 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 anti-aggregation 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 heteroatoms, 5-membered rings, bond and atom polarisability, topological polar surface area, radius, and complexity, among others. Many of these descriptors have a clear relationship between one another, 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 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 coordi-

nates in latent space of descriptors concerning hydrocarbon content for both components, in the region of cluster 2. Atom and bond polarisability are both influential in the formation of cluster 1 (negative value for component 1 and positive value for component 2, but mean polarisability 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 polarisable and nonpolarisable bonds and atoms interspersed; conversely, those with a higher mean polarisability and less polarisable bonds and atoms have more chemically compact head groups, consisting of a small number of atoms with a high polarisability, and smaller hydrophobic tails, leading to a higher ratio of polarisability to molecular weight. This indicates that the shielding of APRs by excipients is better achieved by structures which have larger, but less extreme, polar character, like that which can be achieved by repeating ethoxylate units. Additionally supporting this hypothesis is the position of 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 which supports this hypothesis are 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.

Discussion

MD simulations have been employed to investigate the efficacy of excipients as anti-aggregation 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-

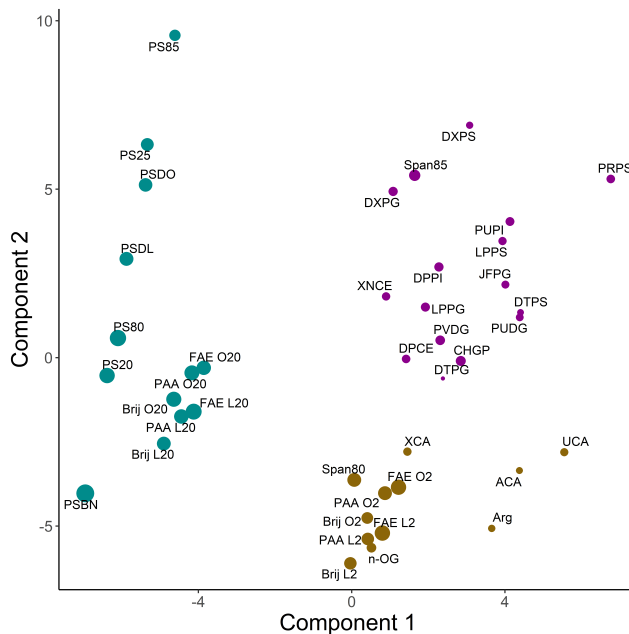


Figure 5: Distribution of the data in the latent variable space, using Mordred descriptors as independent variables. Compounds are sized in proportion to the percentage decrease of APR SASA relative to the control, and are coloured according to their cluster. Cluster 1, teal. Cluster 2, gold. Cluster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

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 utilised 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 head group is driving the overall increase in interaction, a finding that is supported by the literature.^{76–78} 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 destabilising 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 anti-aggregation 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 head group 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 4E). This is reminiscent of binding behaviour observed in crystallographic binding studies with short- and medium-chain fatty acids.⁷⁹ Polysorbate 20 and 80 specifically have also been found to interact with HSA, albeit weakly,⁸⁰ 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,⁸¹ 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),³⁴ 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 acids⁸² and cholesterol⁸³ in circulation, and studies of its interaction with phospholipid membranes.^{84,85} 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 to 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);⁸⁶ 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 stabilisation 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 modelling 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.^{87,88} 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.⁸⁹ Therefore, it could be prudent to model polysorbate as a heterogeneous mixture; to maintain concentrations that are industrially relevant, this would likely require the modelling of extremely large systems.

Validation could also be provided in the characterisation of excipient effects on protein stability, by monitoring changes in aggregate size, protein secondary and tertiary structure and biological activity assays. However, the stability of HSA and its own use as an anti-aggregation agent⁹⁰ 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 modelling 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 either be 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 anti-aggregation activity, hypotheses on the design of novel excipients with greater APR SASA shielding, and therefore improved performance as anti-aggregation agents, can be postulated. Particularly, variable importance in projection (VIP) plots are used in feature selection in drug design⁹¹ and are a useful tool in investigating the structure-activity relationship within a PLS model by indicating the critical descriptors that explain the maximal variance in both dependent and independent variables. An optimised excipient would be a large, branched compound which is highly polar (i.e., with several oxygen atoms) and also 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 is also indicative that there is chemical space that is underutilised by the current anti-aggregation excipient design paradigm. It implies that the exact degree of hydrophobic content is not significant, provided that there is some present in a localised area in order to provide amphiphilic character to the excipient. The findings suggest that perhaps a lighter compound with a higher number of short branches would be more effective in APR shielding than heavier compounds with a small number of large chains.

Conclusion

The coarse-grained modelling of HSA with a series of co-solutes 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 anti-aggregation agents, and polysorbate species specifically were the highest performing

class. Branched compounds tended to make greater contact to 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 optimising protein-exciipient interaction. The overall weight of aliphatic chains does not appear to influence the performance of anti-aggregation agents, provided that some is present. The significance of polarity, polarisability 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 endeavours 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.

Data and software availability

SASA data can be found at the GitHub repository, 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 website: 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 QSAR information from SMILES structures, in conjunction with rdkit, numpy and pandas, and can all be retrieved freely. Some molecular dynamics graphics were created with VMD, freely available from <http://www.ks.uiuc.edu/Research/vmd>.

Acknowledgement

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 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).

Supporting Information Available

The following files are available free of charge and can be found at the GitHub repository (<https://github.com/TobyEdwardKing/Excipient-Optimisation>).

- Bash script for molecular dynamics simulations and their subsequent analysis (.sh)
- Parameter files for excipient compounds (.itp)
- Structural files for excipient compounds (.gro)
- Structural depictions of excipients under study (.pdf)
- R script detailing PLS model (.r)
- Raw data in an archive (.rar)
- Mordred data input for the PLS model (.csv)
- Data table of excipient SMILES and associated data (.csv)

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*, 578–578.
- (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. A.; 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*, 3.
- (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*, 503101.
- (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) 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.
- (35) 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.
- (36) 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 sorbitantriisostearate; 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.
- (37) 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.
- (38) Goff, H. D. Colloidal aspects of ice cream - A review. *Int. Dairy J.* **1997**, *7*, 363–373.
- (39) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.
- (40) 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.
- (41) 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.
- (42) 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 Deliv. Transl. Res.* **2017**, *7*, 529–543.
- (43) Valojerdi, 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.
- (44) 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.

- (45) 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.
- (46) 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.
- (47) Otsu, T.; Ishii, K.; Tahara, T. Microsecond protein dynamics observed at the single-molecule level. *Nat. Commun.* **2015**, *6*, 7685.
- (48) 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.
- (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-Amarello, 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) 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.
- (54) 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. 2022; <https://arxiv.org/abs/2212.01191>.
- (55) 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.
- (56) 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 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.
- (57) 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.
- (58) 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.
- (59) 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.
- (60) 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.

- (61) Grünewald, F.; Alessandri, R.; Kroon, P. C.; Monticelli, L.; Souza, P. C. T.; Marrink, S. J. Polyply; a python suite for facilitating simulations of macromolecules and nanomaterials. *Nat. Commun.* **2022**, *13*, 68.
- (62) 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.
- (63) 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.
- (64) 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.
- (65) 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*, 1–17.
- (66) 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.
- (67) 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.
- (68) Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621.

- (69) 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.
- (70) T. Stanton, D. QSAR and QSPR Model Interpretation Using Partial Least Squares (PLS) Analysis. *Curr. Comput.-Aided Drug Des.* **2012**, *8*, 107–127.
- (71) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminform.* **2018**, *10*, 4.
- (72) 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.
- (73) Monticelli, L.; Kandasamy, S. K.; Periolo, 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.
- (74) 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.
- (75) 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.
- (76) 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.
- (77) Guizado, T. R. Analysis of the structure and dynamics of human serum albumin. *J. Mol. Model.* **2014**, *20*, 2450.
- (78) 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.
- (79) 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. *J. Mol. Biol.* **2000**, *303*, 721–732.
- (80) 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.
- (81) 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 Bioinform.* **2007**, *8*, 65.
- (82) van der Vusse, G. J. Albumin as fatty acid transporter. *Drug Metab. Pharmacokinet.* **2009**, *24*, 300–307.
- (83) 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.
- (84) 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.
- (85) 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.* **2021**, *15*, 624–634.

- (86) Marsh, D. *Handbook of lipid bilayers*, array ed.; CRC Press: Boca Raton, FL, 2013; p 1145.
- (87) 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.
- (88) 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*, e9438.
- (89) 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.
- (90) 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.
- (91) Mehmood, T.; Sæbø, S.; Liland, K. H. Comparison of variable selection methods in partial least squares regression. *J. Chemom.* **2020**, *34*, e3226.

TOC Graphic

