ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Structure-Based Computational Scanning of Chemical Modification Sites in Biologics

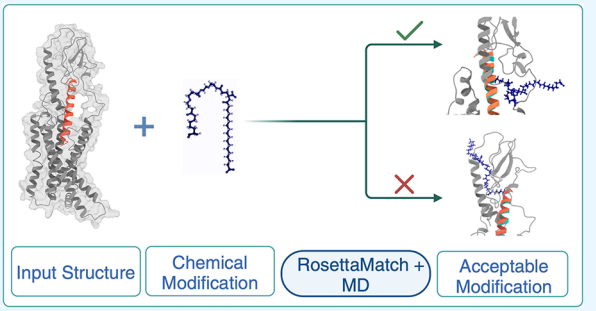
Nidhin Thomas,[‡] Tanmoy Sanyal,[‡] Per Greisen, Jr., and Kristine Deibler*

Cite This: *ACS Omega* 2024, 9, 36787−36794

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: To address the challenges of short half-life, immunogenicity, and nonspecific distribution, chemical modifications of peptide and protein-based drugs have emerged as a versatile strategy for improving their therapeutic efficacy. One such modification involves the derivatization of peptides and proteins with fatty acids, which can protract their half-life, modify their biodistribution, and potentially enable targeted delivery to specific tissues or disease sites of interest. However, the present strategies for the synthesis of such synthetically modified biologics require numerous rounds of experimental testing and often yield unstable, inactive, or heterogeneous products. To address the inefficiencies in designing modified biologics, we developed a hybrid computational workflow that integrates RosettaMatch from the Rosetta suite of protein modeling tools with molecular dynamics (MD) simulations. This approach not only reduces the number of amino acid positions that need to be experimentally tested by targeting only the most promising candidates for modification but also expedites the design of chemically modified biologics with the desired properties, ensuring a rapid and cost-effective development cycle. Although we demonstrate the utility of our method on a peptide therapeutic, GLP-1, with different fatty acid derivatizations, this straightforward approach has the potential to streamline the design process of a diverse range of chemically modified therapeutics, enabling tailored enhancements to their pharmacokinetic properties.



INTRODUCTION

Peptide and protein drugs have proven successful in the treatment of a wide range of medical conditions.[1] Whereas some classes of therapeutic proteins, including antibodies, have long inherent half-lives through their constant region, many native molecules such as peptide hormones are often susceptible to enzymatic degradation, renal clearance, and rapid receptor-mediated clearance, which lead to a short plasma elimination half-life.[2] Therefore, considerable research effort has gone into the development of a diverse range of principles and technologies for prolonging the half-life of peptides and proteins including N/O-linked glycosylation, PEGylation, Fc-fusion, albumin-fusion, and reversible albumin binding techniques.[2−4] In particular, fatty acid derivatization technology has proven extremely useful in many cases, influencing both the rate of absorption after subcutaneous administration and the rate of elimination by various clearance pathways.[5−9]

Fatty acid derivatization (known as protraction) entails the chemical coupling of a fatty acid side chain onto a peptide or a protein to confer reversible binding to human serum albumin (HSA), which is a multifunctional and abundant protein that plays a key role in transporting endogenous fatty acids and other substances as well as small-molecule drug compounds.[10] Early in the development of "protracted" biologics, it became apparent that more than just albumin binding was at play and that fatty acid derivatization could offer several mechanistic opportunities for tailoring both the pharmacokinetic half-life and distribution parameters to specific needs for a given peptide drug.[10] Half-life extension is a critical aspect of developing effective therapeutic agents, particularly in the context of biologics. Numerous studies have identified specific derivatives and dimers that exhibit strong HSA-binding affinity and can effectively extend half-life.[3,11,12]

Despite the promise of therapeutic biologics, their optimization remains a significant challenge.[2,13,14] Traditional development is a manual, iterative process, requiring extensive experimental testing that often yields unstable, inactive, or heterogeneous products.[14−18] Our computational solution addresses this bottleneck by rapidly identifying optimal chemical modification sites, aiming to streamline the development process and reduce reliance on costly and time-consuming experimental iterations that not only slow the development cycle but also increase the cost and labor

Received: June 24, 2024
Revised: July 30, 2024
Accepted: August 6, 2024
Published: August 14, 2024

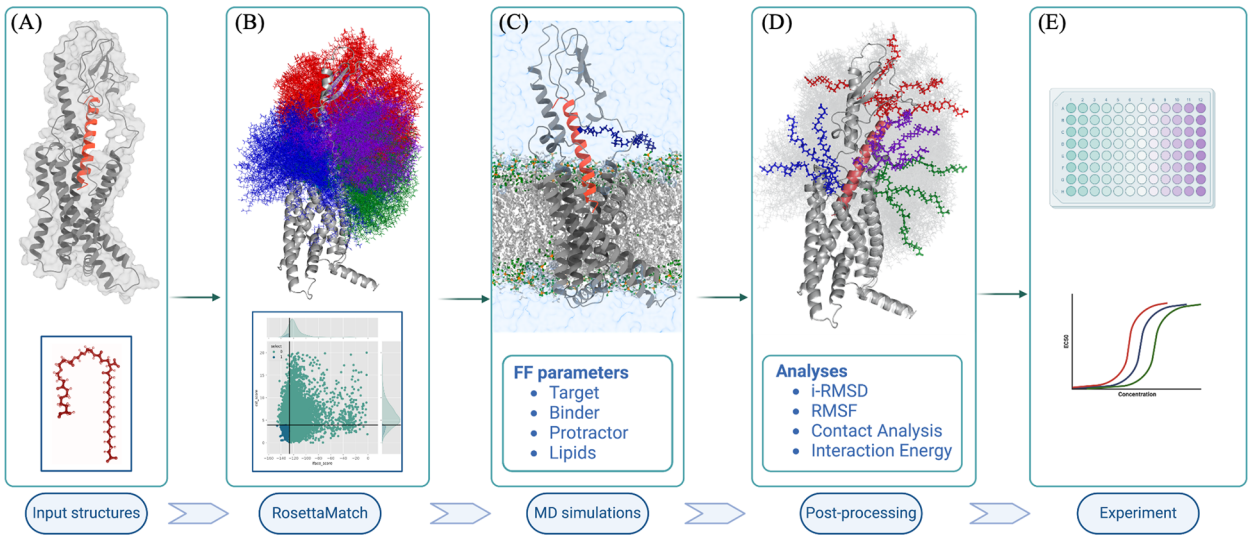


Figure 1. Protractor scan workflow demonstrated on the GLP1-GLP1R system with a fatty-diacid-based protractor. (A) Input the GLP-1/GLP-1R co-complex structure, and process the extended protractor conformations (generated from SMILES strings) first through (B) high-throughput sampling using RosettaMatch and then through (C) a low-throughput MD simulation stage to enumerate sterically feasible protractor conformations at allowed residues on GLP-1. (D) Structural parameters and interaction energies calculated from the MD simulation help to filter choices for protraction sites which can (E) ultimately guide the experiment.

involved. Our innovative computational solution has the potential to advance this paradigm by augmenting and automating the selection process. By rapidly identifying optimal chemical modification sites, our method reduces the need for experimental testing and accelerates the development of stable and effective biologic therapeutics. Our workflow does not in any way seek to replace experimental protraction assays with an in silico decision. Rather, we generate a set of hypotheses about protraction sites and supplement them as an estimation of the structural and energetic effects of placing a protractor at these sites, which can narrow the search space of protraction experiments.

In this paper, we demonstrate our structure-based computational scanning protocol on a candidate therapeutic, glucagon-like-peptide-1 (GLP-1). The quantitative effect of different fatty acid chemistries and their acylation locations on GLP-1 on peptide half-life has been studied in detail previously.[9,15,19] Here, we use a C-18 diacid linked to GLP-1 with a γGlu-2xOEG linker that provides a demonstrably superior half-life extension over other fatty-acid-based modifications.[15,19] We refer to the fatty acid and linker combination as the "protractor" for the remainder of this paper. Fatty acid modification of therapeutic peptides, such as GLP-1, can significantly enhance their pharmacokinetic properties. However, identifying optimal sites for attaching these modifications (which we term "protractors") remains a challenge. In this study, we present a computational protocol that combines RosettaMatch[16,17] for initial structural screening with molecular dynamics (MD) simulations to assess the dynamic behavior and interactions of protracted GLP-1 systems. This approach enables the rapid identification of optimal protractor combinations, offering the potential to substantially advance the design of fatty-acid-modified biologics for enhanced therapeutic efficacy.

METHODS

Protractor Conformer Generation. As demonstrated in Figure 1, our proposed workflow requires only a SMILES string for the protractor molecule and an input structure of the biologic molecule of interest. We used the Chem.CanonSmiles() method in RDKit to canonicalize the SMILES string and used the Experimental-Torsion-(basic) Knowledge Distance Geometry (ETKDG) approach from RDKi[20,21] and enforced chirality to generate an initial ensemble of conformers that were further pruned to remove members within 1 Å heavy atom RMSD of each other. The details are provided in Appendix 1 in the Supporting Information. Initially, 100 conformations for the γGlu-2xOEG-C18-diacid protractor were generated from the following SMILES string:

O=CCOCCOCCNC(=O)COCCOCCNC(=O)CC[C@@H(NC(=O)CCCCCCCCCCCCCCCCC(=O)[O−])C(=O)[O−]

Preparation of the GLP-1/GLP-1R Complex and Energy Minimization. For the benchmark system in this paper, GLP-1:7−36 in complex with full length GLP-1-receptor (GLP-1R) (residues 29−423) was taken from the 2.10 Å cryo-EM structure 6X18.[22] Unresolved loop regions of the GLP-1R structure (PDB: 6X18) were added using the looping model protocols in Rosetta.[23−27] Specifically, this experimental structure was missing loops, and we chose to use the remodel executable to model potential conformations. Remodel allows us to place missing atoms and utilize cyclic coordinate descent (CCD) for building the missing loops. We generated a blueprint file by running the getBluePrintFromCoords.pl script (in Rosetta/demos/public/design_w_flex_loops_using_RosettaRemodel/scripts), and that was modified to contain the missing residues. In addition, we added a fast relaxation step to the refinement stage in the flags (see code in GitHub).

RosettaMatch Protocol for Sampling Sterically Feasible Protraction Sites. RosettaMatch[16,21] was employed to identify sterically feasible conformations and acylation sites for the γGlu-2xOEG-C18-diacid protractor within the GLP-1 peptide scaffold, ensuring compatibility with the desired binding interface geometry. To achieve this, a parameter file

https://doi.org/10.1021/acsomega.4c05857
ACS Omega 2024, 9, 36787−36794

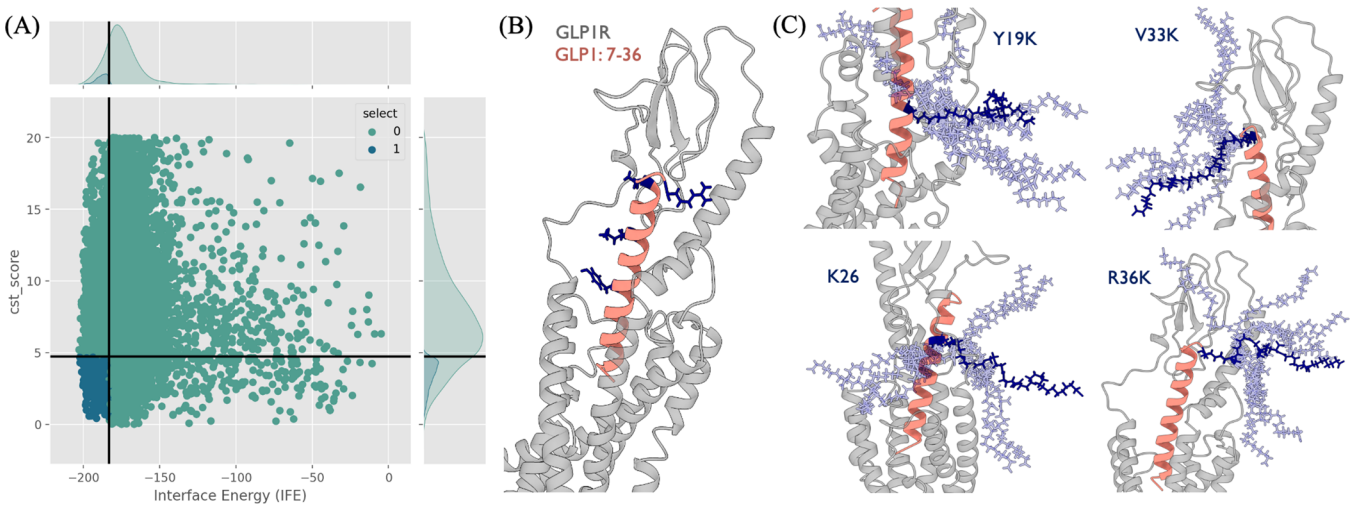


Figure 2. High-throughput scan of protraction sites on GLP-1(7−36) (in complex with GLP-1R) using the RosettaMatch protocol. (A) Clustering of the landscape of match constraint score ("cst_score") and GLP-1/GLP-1R interface energy ("IFE") into a binary decision variable where 1 (blue) corresponds to linker conformers that have both low interface energy and low match constraint score and can be selected for acylation, whereas 0 (green) corresponds to all other rejected conformers. (B) Some representative acylation residues taken from the viable region in (A) are mapped to the energy minimized GLP-1/GLP-1R co-complex. (C) Sample protractor conformations at residues that cover the length of the GLP-1 fragment: Y19, K26, V33, and R36. Protractor conformations are energy minimized random samples from an RMSD-based clustering of solutions returned by the RosettaMatch protocol for a given residue, and the conformation in dark blue represents the cluster center.

for the protractor was generated using the molfile_to_params.py script, assigning an overall charge of −2. A position file was then generated to define the residues on the GLP-1 peptide to be sampled. Because of the relatively small size of the peptide, all positions were screened to comprehensively identify potential protraction sites.

The geometry of the chemical bond linking the fatty acid protractor to the amino acid side chain (lysine in this case) was specified using six degrees of freedom (one bond distance, two angles, three dihedrals), defining the amide bond linkage. RosettaMatch hits, representing potential protractor attachment points, were grouped into bins based on a 3 Å RMSD similarity criterion.

To select the most productive attachment points, hits were subjected to energy minimization and side-chain repacking using the ref2015_cst scoring function in Rosetta.[28] Minimization consisted of 1000 iterations, with side-chain repacking limited to residues within a 10 Å radius of the GLP-1/GLP-1R binding site. The resulting relaxed conformations were further filtered based on the peptide-receptor interface energy (IFE) and RosettaMatch's constraint score, selecting candidates within the top 25th percentile of both features (IFE < −183.086 Rosetta energy units (REU) and constraint score < 4.719) to form the final ensemble of valid protractor conformations. The minimization and repacking were conducted using RosettaScripts,[29] as shown within run_matcher.py.

MD Simulation Setup for Protracted GLP-1/GLP-1R Systems. For each identified protraction site on GLP-1, a random subset of up to 10 protractor conformations, selected from the previously filtered RosettaMatch results, was used to initiate individual molecular dynamics (MD) simulations. Further, because the MD stage serves to reassess hypotheses generated previously with RosettaMatch, we used a more biological representation of the system, which includes a lipid bilayer separating the extracellular and transmembrane domains of GLP-1R.

Protracted co-complexes from the RosettaMatch stage were aligned (using PyMol 2.2.5[30]) to a pre-equilibrated GLP-1/GLP-1R complex inserted into a DPPC lipid-bilayer sourced from MemProtMD.[31] The alignment ensured the correct orientation of the protein complex with respect to the lipid membrane. We used the AmberFF03[32,33] and CHARMM36[34] force field for protein and lipid, respectively. Atomic charges of protracted lysine residue were computed using the RESP-A1 (HF/6-31G*) charge model and the GAMESS-US[35] quantum mechanics program given in the Q4MD Force Field Tools Web server.[36−38] The output AMBERFF03 parameters were converted to GROMACS readable.itp files using the Parmed package.[39] The system was solvated using TIP3P[40,41] water molecules, and a buffer of 150 mM NaCl ions was added to neutralize the system.

MD simulations were performed using GROMACS 2021.4.[42−44] The simulations began with energy minimization using the steepest descent algorithm until the maximum force was smaller than 700.0 kJ mol^{-1} nm^{-1}. This was followed by a 3.25 ns equilibration period, performed in three distinct steps, maintaining the system's temperature at 310.15 K using velocity-rescale temperature coupling with a time constant of 1.0 ps. During this phase, positional restraints were placed on the protein complex and lipid phosphorus atoms and were successively reduced to ensure physiological conditions for the proteins and prevent phase transitions for the DPPC lipids. Specifically, even though the phase transition temperature for DPPC lipids is around 315 K,[45,46] equilibrium simulations at 310.15 K for ∼100 ns ensured that they remained in a liquid-disordered state. Subsequently, the system underwent 15 ns of NPT equilibration at a temperature of 310.15 K and a pressure of 1 atm, employing Berendsen semi-isotropic pressure coupling,[47] with a time-constant of 5.0 ps and compressibility value of 4.5×10^{-5} bar^{-1}. During this period, positional restraints on the protein complex, lipid phosphorus atoms, and protein side chains were gradually removed. In the final production phase that ran for 100 ns, all positional restraints were lifted, and Parrinello−Rahman semi-isotropic pressure

https://doi.org/10.1021/acsomega.4c05857
ACS Omega 2024, 9, 36787−36794

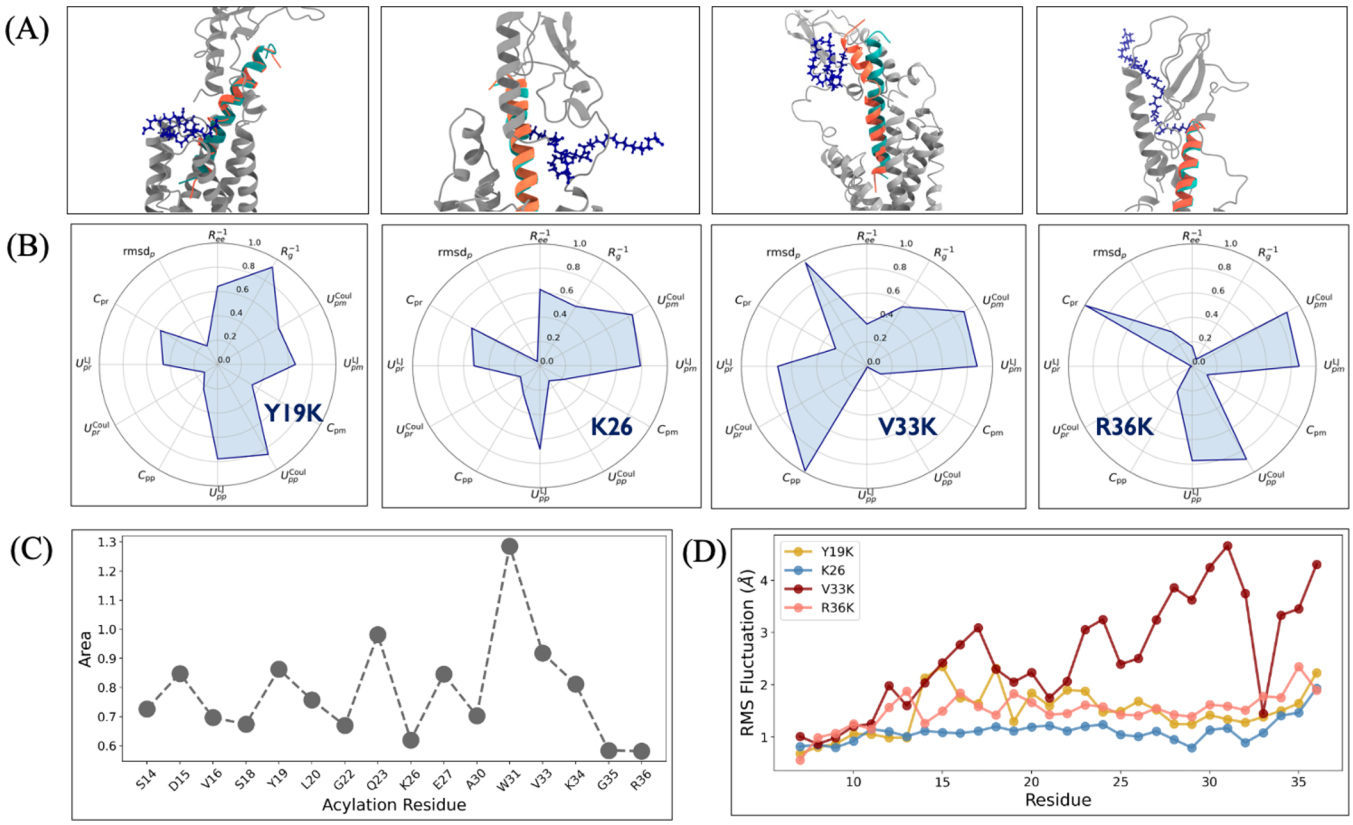


Figure 3. Comparative analysis of protraction site efficacy through structural and energetic parameters. (A) Snapshots from MD simulations of the protracted GLP-1(7−36)/GLP-1R complex. The lipid bilayer and solvent were removed to enhance the clarity. GLP-1R, GLP-1, and protractor side chain are shown in gray, salmon, and blue, respectively. The reference conformation of the GLP-1 peptide from RosettaMatch is shown in cyan. (B) Radar plots for structural parameters and interaction energies calculated from the MD simulations. Results are presented for a γGlu-2xOEG-C18-diacid protractor acylated at residues Y19, K26, V33, and R36 (mutating to K when required) and initial conformations for Y19 and V33 protracted peptide shown in green. Each property in panel B is dimensionless and has been normalized to (0,1) using a sigmoid transformation, and lower values of each property are desirable. Properties included are the inverse of protractor radius of gyration (R_g^{-1}), inverse of protractor end-to-end distance (R_{ee}^{-1}), RMSD of protracted peptide relative to unprotracted state ($rmsd_p$), number of contacts ($C_{\alpha\beta}$), electrostatic interaction energy ($U_{\alpha\beta}^{Coul}$) and van-der Waals interaction energy ($U_{\alpha\beta}^{LJ}$) between protractor−receptor ($\alpha\beta$ = pr), protractor−peptide ($\alpha\beta$ = pp), and protractor−membrane ($\alpha\beta$ = pm), respectively. (C) Area under the radar plot calculated for all acylation residues enumerated by RosettaMatch. (D) RMS fluctuations of GLP-1 residues, measured relative to unprotracted GLP-1, for the four acylated residues.

coupling[48] with a time constant of 5.0 ps was used. All simulations adopted a 0.002 ps time step, and statistics were collected from the last 80 ns of each production run trajectory. We verified that the lipid bilayer remained in a liquid-disorder phase by calculating time-resolved profiles of the bilayer thickness of the lipid bilayer (Figure S1 in the Supporting Information).

We derived several structural parameters and interaction energies from MD simulations. These include the protractor's radius of gyration (R_g), the protractor's end-to-end distance (R_{ee}), and root-mean-square deviation (RMSD) of the protracted peptide in comparison to its unprotracted state ($rmsd_p$), along with the number of contacts ($C_{\alpha\beta}$). Additionally, we computed the electrostatic interaction energy ($U_{\alpha\beta}^{Coul}$) and the van der Waals interaction energy ($U_{\alpha\beta}^{LJ}$) for the protractor−receptor ($\alpha\beta$ = pr), protractor−peptide ($\alpha\beta$ = pp), and protractor−membrane ($\alpha\beta$ = pm) using the "gmx energy" command. The index groups used in "gmx energy" for the receptor, membrane, and protractor residue were created using "gmx make_ndx".

For each protraction site, up to 10 independent trajectories were simulated, each initiated from a different RosettaMatch-derived protractor conformation. These trajectories were divided into four contiguous blocks, and the aforementioned properties were averaged to enhance the statistical accuracy and reduce the noise. Subsequently, the structural parameters and interaction energies were transformed to a (0,1) domain using a sigmoid function whose slope was estimated from the data. Thus, for a property x, with minimum, maximum, and average values across all MD simulations for different acylation residues given by x_{min}, x_{max}, and x_0, the corresponding transformed value is given by $y = \frac{1}{1+\exp[-k(x-x_0)]}$. The slope k was calculated as $k = -\frac{1}{x_{max}-x_{min}}\log\left(\frac{1-y_{max}}{1-y_{min}}\right)$, where we set the numerical minimum and maximum values of the transformed property as y_{min} = 0.001 and y_{max} = 0.999, respectively.

RESULTS AND DISCUSSION

Analysis and Selection of Protractor Conformations. To identify optimal protractor conformations from the RosettaMatch output, a rigorous filtering strategy was implemented. Figure 2A illustrates the distribution of feasible GLP-1/GLP-1R complex conformations returned by RosettaMatch plotted according to their interface energy (IFE) and

https://doi.org/10.1021/acsomega.4c05857
ACS Omega 2024, 9, 36787−36794

constraint score (cst_score). The IFE, representing the cumulative Rosetta energy score at the GLP-1/GLP-1R interface, serves as a proxy for energetic favorability. The cst_score, derived from the EnzScore filter,[16] quantifies the geometric compatibility of the protractor attachment in terms of the degrees of freedom around the amide bond.

Optimal solutions, as exemplified in Table S1, are characterized by low values for both scores, indicating both steric compatibility and minimal disruption of the GLP-1/GLP-1R interface. To identify these optimal solutions, a threshold was applied, selecting the top 25th percentile of conformations in the lower left quadrant of Figure 2A. Notably, the choice of the 25th percentile threshold is user-defined and can be adjusted based on the desired number of solutions.

Within this optimal ensemble, multiple protractor conformations may correspond to the same acylation residue on GLP-1. Therefore, the "protractibility" of a residue was qualitatively defined as the presence of at least one optimal protractor conformation (blue shaded area in Figure 2A). Although this measure is approximate, it serves as a valuable tool for narrowing down the potential protraction sites in larger biologics like mini-proteins or antibodies, where the sheer number of possibilities can be overwhelming.

Figure 2B highlights a subset of optimal solutions identified by RosettaMatch for four representative acylation residues: Y19, K26, V33, and R36. These residues exemplify the diverse protraction scenarios along the GLP-1:7−36 peptide: Y19 is proximal to the GLP-1R binding pocket; V33 and R36 are near the GLP-1R extracellular region, and K26 occupies an intermediate region.

As shown in Table S1, N-terminal residues[7−14] buried within the GLP-1R binding pocket are unsuitable for protraction, as evidenced by the absence of valid RosettaMatch solutions. Similarly, C-terminal residues[32−35] are likely incompatible due to steric clashes with the GLP-1R extracellular domain (ECD).[40] This observation is consistent with literature reports of undesirably high EC50 values for GLP-1 analogs protracted at residues 36 and 37,[11] suggesting the existence of an optimal location for protractor placement along the peptide axis.

This notion is further supported by the wide range of protractor conformations observed when attached to K26, which are predominantly orthogonal to the peptide axis. Interestingly, K26 has been reported to yield maximal half-life extension for the protractor used in this study.[9,11] However, the quantity of optimal RosettaMatch solutions can vary depending on the diversity of protractor conformations considered; thus, these results may not definitively determine the feasibility of protractor attachment at a specific residue.

Molecular Dynamics Simulations of Protracted GLP-1 Systems. To complement the RosettaMatch findings and gain a deeper insight into the structural dynamics of the protracted GLP-1 systems, we conducted MD simulations. In Figure 3A, we show representative snapshots from MD simulations of the four (Y19, K26, V33, and R36) protracted GLP-1 systems. Figure 3B shows metrics calculated from the MD simulations in the form of radar charts. The chosen metrics, including the radius of gyration (R_g) and end-to-end distance (R_{ee}), are essential for quantifying the conformation and relative size of the protractor chain.[49,50] Additionally, the backbone RMSD of protracted GLP-1 relative to the bare peptide provides insight into how the protractor influences the structural dynamics of the peptide. We also report energies of the protractor's interaction with GLP-1, GLP-1R, and the lipid bilayer as measured by the number of residue−residue contacts (with a 5 Å CA−CA cutoff), van der Waals, and electrostatic components of energies. Protraction of biologics with fatty acids impacts receptor binding through different mechanisms including reduction of the free concentration of drug available for receptor binding (e.g., only the unbound drug is recognized by the receptor) or reduction of receptor affinity (e.g., the biologic retains some receptor binding even when bound to albumin).[2] An ideal protraction site should provide enhanced binding with serum albumin while shielding the peptide−receptor interface from steric interactions with the protractor. Because albumin is not included in the simulation, high R_g and R_{ee} and low protractor−membrane interaction signal extended protractor conformations that may bind more easily with albumin. Given that albumin is not incorporated in the simulation, the observed high values of R_g and R_{ee}, coupled with the low protractor−membrane interaction, suggest extended protractor conformations. These extended conformations are less likely to interact with the rest of the protein, thereby reducing potential steric hindrances and unwanted conformational changes.[2,51−53] As a result, these conformations may facilitate more efficient and specific binding with albumin once it is introduced into the system. On the other hand, lower interaction of the protractor with the peptide and receptor reflects preservation of the peptide−receptor binding interface. For consistency, we present R_g^{-1} and R_{ee}^{-1} in Figure 3B such that for all in silico observables, lower values are better. We tried different data transforms that could normalize the data to a (0,1) domain such as a linear scaling using the maximum range or Z scores. However, we observed that only a sigmoid type transform (detailed in the Methods section) prevented artificial inflation of differences in data points that are very close to each other. Therefore, all properties are normalized to a fraction between (0,1) using a sigmoid transformation. Thus, for the radar plots in Figure 3B, a lower total shaded area is approximately correlated to an optimal protraction strategy. We calculated the shaded area for radar plots for all GLP-1 residues deemed feasible by RosettaMatch (see Figure. S2 in the Supporting Information) and present the area value as a function of the acylation residue in Figure 3C. Finally, Figure 3D shows the root-mean-square fluctuation (RMSF) of individual residues of the four protracted GLP-1(7−36) fragments relative to the unprotracted peptide.

In Figure 3B, protraction at K26 has the smallest footprint among all others on the radar chart, among the four residues presented here, whereas protraction at V33 produces the largest total shaded area, ∼ 46% more than that at K26. The MD simulation results recapitulate the optimality of K26 as an attractive candidate for protraction. But more importantly, they highlight the fact that V33 is problematic because a protractor placed at this residue interacts undesirably with GLP-1, nearly "twisting" the peptide axis (Figure 3A, column 3, orange) and producing a very high backbone RMSD relative to the bare peptide (Figure 3A, column 3, green) and high protractor−peptide contacts (Figure 3B), and significantly disrupts the interface with GLP1R as measured by very high RMSF of ∼4 Å (Figure 3D). Protraction at R36 avoids interaction of the protractor with GLP-1 but not with the GLP-1R extracellular domain (ECD). However, the ECD is distal from both the GLP-1 binding pocket and the transmembrane regions of GLP-1R, producing extended protractor conformations with very low R_g^{-1} and R_{ee}^{-1} that lead to a lower overall shaded

https://doi.org/10.1021/acsomega.4c05857
ACS Omega 2024, 9, 36787−36794

surface area in Figure 3B. On the other hand, protraction at Y19 has a nontrivial effect on the GLP-1 RMSF between residues 12 and 20 (~2.5 Å, Figure 3D); i.e., it gets sterically locked in the N-terminal end of the GLP-1R binding interface (Figure 3A) and fails to extend into the solution as evidenced by higher values of normalized R_g^{-1} and R_{ee}^{-1} (Figure 3B).

N-terminal and C-terminal residues, exemplified by V33, often appear as promising protraction candidates upon the initial visual inspection of protein structures, a common practice in the pharmaceutical industry. However, the true suitability of these sites can be rigorously assessed only by considering the full structural model of the complex and the dynamic flexibility of the protractor chains. Molecular dynamics (MD) simulations complement RosettaMatch by revealing potentially unfavorable interactions that may not be apparent from static structural analysis alone, thereby refining the selection of protraction sites for subsequent experimental validation.

CONCLUSIONS

In this work, we introduced an in silico pipeline for predicting sites on drug-like peptide molecules where long-chain biopolymers may be grafted to provide half-life protraction by binding to serum albumin as an example. The first stage of the pipeline used the RosettaMatch algorithm to sample sterically compatible protractor conformations on different residues along a biologic molecule, whereas the second stage employed MD simulations to further understand the implication of the protractor within a system that closely mimics physiological conditions. We benchmarked our method on a candidate system of the GLP-1(7−36) peptide fragment in complex with full-length GLP-1R and a γGlu-2xOEG-C18-diacid protractor. The RosettaMatch stage filtered out N-terminal GLP-1 residues buried deep within the GLP-1R binding pocket and generated conformational ensembles for protractors attached to each of the remaining residues. In the second stage, structural parameters and interaction energies were calculated from MD simulations done in the presence of a lipid bilayer separating the extracellular and transmembrane domains of GLP-1R and simulating the effect of the cell membrane. The simulations further revealed the inadequacy of C-terminal GLP-1 residues for protraction due to deleterious interactions with the peptide and the binding interface. Eventually, residue K26 demonstrated low peptide, receptor, and membrane interaction and extended protractor conformations that may attach more easily with serum-abundant albumin while preserving the native GLP-1/GLP-1R interface, which is corroborated by previous protraction studies of GLP-1 derivatives.[15]

The success of our workflow is dependent on the quality of the input structure, and any inaccuracies in protein−protein interfaces of in silico predicted structures such as from AlphaFold-multimer[S4] will be inevitably reflected in protraction strategies predicted by our method.

Nevertheless, we provide a simple method with the potential to be fully automated for placement of protractors based on virtual scanning. Although this study focuses on a single peptide-receptor system, our computational pipeline is readily adaptable to other biologic formats of interest, including mini-proteins, antibodies (for antibody−drug conjugates), and various protractor and linkage chemistries. For molecules larger than peptides, the number of potential protraction sites to test experimentally is also much greater, and even running only the RosettaMatch stage can provide significant gains in experimental efficiency by screening the number of testable sites.

ASSOCIATED CONTENT

Data Availability Statement

All relevant code is provided in the repository: https://github.com/novonordisk-research/nnprotscan. The repository includes bash scripts to run the Matcher application; Rosetta XML scripts for relaxing the peptide−receptor interface; a Python script then wraps the above two scripts allowing to run the workflow on any protein complex of choice; GROMACS input files, force field files, and associated scripts for MD simulations of protracted GLP-1/GLP-1R system; and reference GLP-1/GLP-1R structures with and without the bilayer.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c05857.

Additional details on the RosettaMatch protocol and MD simulations, including a summary of MD simulations (PDF)

AUTHOR INFORMATION

Corresponding Author

Kristine Deibler − *Digital Science and Innovation, Novo Nordisk Research Center Seattle Inc., Seattle, Washington 98101, United States;* orcid.org/0000-0002-2376-8456; Email: ktdb@novonordisk.com

Authors

Nidhin Thomas − *Digital Science and Innovation, Novo Nordisk Research Center Seattle Inc., Seattle, Washington 98101, United States*

Tanmoy Sanyal − *Digital Science and Innovation, Novo Nordisk Research Center Seattle Inc., Seattle, Washington 98101, United States;* orcid.org/0000-0002-6009-9431

Per Greisen, Jr. − *Digital Science and Innovation, Novo Nordisk Research Center Seattle Inc., Seattle, Washington 98101, United States;* Present Address: Per Greisen Jr. is currently at BioMap, Palo Alto, California 94303, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c05857

Author Contributions

P.G. and K.D. conceived the original approach and tested the preliminary code of the Matcher. T.S. executed and streamlined Matcher code and tested the preliminary MD work. N.T. developed and executed the MD workflow and analysis. The manuscript was written through contributions of all authors. The manuscript was edited by K.D. and P.G. All authors have given approval to the final version of the manuscript.

Author Contributions

‡N.T. and T.S. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We'd like to acknowledge Ta-Yi Yu for testing the code. We'd also like to thank François-Yves Dupradeau for providing

guidance in the derivation of RESP charges from https://upjv.q4md-forcefieldtools.org.

REFERENCES

(1) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic peptides: current applications and future directions. *Signal Transduction Targeted Ther.* **2022**, *7* (1), 48.

(2) Kurtzhals, P.; Ostergaard, S.; Nishimura, E.; Kjeldsen, T. Derivatization with fatty acids in peptide and protein drug discovery. *Nat. Rev. Drug Discovery* **2023**, *22* (1), 59−80.

(3) Strohl, W. R. Fusion proteins for half-life extension of biologics as a strategy to make biobetters. *BioDrugs.* **2015**, *29* (4), 215−39.

(4) Cao, S-j; Lv, Z-q; Guo, S.; Jiang, G-p; Liu, H-l An Update-Prolonging the action of protein and peptide drugs. *Journal of Drug Delivery Science and Technology.* **2021**, *61*, No. 102124.

(5) Myers, S. R.; Yakubu-Madus, F. E.; Johnson, W. T.; Baker, J. E.; Cusick, T. S.; Williams, V. K.; Tinsley, F. C.; Kriauciunas, A.; Manetta, J.; Chen, V. J.; et al. Acylation of human insulin with palmitic acid extends the time action of human insulin in diabetic dogs. *Diabetes* **1997**, *46* (4), 637−642.

(6) Markussen, J.; Havelund, S.; Kurtzhals, P.; Andersen, A. S.; Halstrøm, J.; Hasselager, E.; Larsen, U. D.; Ribel, U.; Schäffer, L.; Vad, K.; Jonassen, I.; et al. Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs. *Diabetologia* **1996**, *39*, 281−288.

(7) Kurtzhals, P.; Havelund, S.; Jonassen, I.; Kiehr, B.; Ribel, U.; Markussen, J. Albumin binding and time action of acylated insulins in various species. *Journal of pharmaceutical sciences.* **1996**, *85* (3), 304−8.

(8) Kurtzhals, P.; Havelund, S.; Jonassen, I.; Kiehr, B.; Larsen, U.; Ribel, U.; et al. Albumin binding of insulins acylated with fatty acids: characterization of the ligand-protein interaction and correlation between binding affinity and timing of the insulin effect in vivo. *Biochem. J.* **1995**, *312* (3), 725−31.

(9) Knudsen, L. B.; Nielsen, P. F.; Huusfeldt, P. O.; Johansen, N. L.; Madsen, K.; Pedersen, F. Z.; et al. Potent derivatives of glucagon-like peptide-1 with pharmacokinetic properties suitable for once daily administration. *Journal of medicinal chemistry.* **2000**, *43* (9), 1664−9.

(10) Chen, B.; Sun, Y.; Niu, J.; Jarugumilli, G. K.; Wu, X. Protein lipidation in cell signaling and diseases: function, regulation, and therapeutic opportunities. *Cell chemical biology.* **2018**, *25* (7), 817−31.

(11) van Witteloostuijn, S. B.; Pedersen, S. L.; Jensen, K. J. Half-Life Extension of Biopharmaceuticals using Chemical Methods: Alternatives to PEGylation. *ChemMedChem.* **2016**, *11* (22), 2474−95.

(12) Pilati, D.; Howard, K. A. Albumin-based drug designs for pharmacokinetic modulation. *Expert Opin Drug Metab Toxicol.* **2020**, *16* (9), 783−95.

(13) Zaman, R.; Islam, R. A.; Ibnat, N.; Othman, I.; Zaini, A.; Lee, C. Y.; et al. Current strategies in extending half-lives of therapeutic proteins. *J. Controlled Release* **2019**, *301*, 176−89.

(14) Menacho-Melgar, R.; Decker, J. S.; Hennigan, J. N.; Lynch, M. D. A review of lipidation in the development of advanced protein and peptide therapeutics. *J. Controlled Release* **2019**, *295*, 1−12.

(15) Lau, J.; Bloch, P.; Schäffer, L.; Pettersson, I.; Spetzler, J.; Kofoed, J.; et al. Discovery of the once-weekly glucagon-like peptide-1 (GLP-1) analogue semaglutide. *Journal of medicinal chemistry.* **2015**, *58* (18), 7370−80.

(16) Jiang, H.; Zhang, X.; Chen, X.; Aramsangtienchai, P.; Tong, Z.; Lin, H. Protein lipidation: occurrence, mechanisms, biological functions, and enabling technologies. *Chemical reviews.* **2018**, *118* (3), 919−88.

(17) Ryberg, L. A.; Sønderby, P.; Bukrinski, J. T.; Harris, P.; Peters, G. H. J. Investigations of albumin−insulin Detemir complexes using molecular dynamics simulations and free energy calculations. *Mol. Pharmaceutics* **2019**, *17* (1), 132−144.

(18) Frederiksen, T. M.; Sønderby, P.; Ryberg, L. A.; Harris, P.; Bukrinski, J. T.; Scharff-Poulsen, A. M.; Elf-Lind, M. N.; Peters, G. H. Oligomerization of a glucagon-like peptide 1 analog: bridging experiment and simulations. *Biophys. J.* **2015**, *109* (6), 1202−1213.

(19) Østergaard, S.; Paulsson, J. F.; Kofoed, J.; Zosel, F.; Olsen, J.; Jeppesen, C. B.; Spetzler, J.; Ynddal, L.; Schleiss, L. G.; Christoffersen, B. Ø.; Raun, K.; Sensfuss, U.; Nielsen, F. S.; Jørgensen, R.; Wulff, B. S. The effect of fatty diacid acylation of human PYY3−36 on Y2 receptor potency and half-life in minipigs. *Sci. Rep.* **2021**, *11* (1), 21179.

(20) Landrum, G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* **2013**, *8*, 5281.

(21) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf Model.* **2015**, *55* (12), 2562−74.

(22) Zhang, X.; Belousoff, M. J.; Zhao, P.; Kooistra, A. J.; Truong, T. T.; Ang, S. Y.; et al. Differential GLP-1R binding and activation by peptide and non-peptide agonists. *Mol. Cell* **2020**, *80* (3), 485−500.

(23) Huang, P.-S.; Ban, Y-EA; Richter, F.; Andre, I.; Vernon, R.; Schief, W. R.; et al. RosettaRemodel: a generalized framework for flexible backbone protein design. *PloS one.* **2011**, *6* (8), No. e24109.

(24) Mandell, D. J.; Coutsias, E. A.; Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods.* **2009**, *6* (8), 551−2.

(25) Stein, A.; Kortemme, T. Improved sampling of near-native protein conformations. *PLoS One* **2013**, *8* (5), No. e63090.

(26) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **2011**, *487*, 545−74.

(27) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031−48.

(28) Park, H.; Bradley, P.; Greisen, P., Jr.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; et al. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016**, *12* (12), 6201−12.

(29) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E. M.; Khare, S. D.; Koga, N.; et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One.* **2011**, *6* (6), No. e20161.

(30) Schrödinger L. L. C. *The PyMOL Molecular Graphics System*; Version 1.8. (No Title). 2015.

(31) Newport, T. D.; Sansom, M. S. P.; Stansfeld, P. J. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic acids research.* **2019**, *47* (D1), D390−D7.

(32) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; et al. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of chemical theory and computation.* **2020**, *16* (1), 528−52.

(33) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153* (11), 114502.

(34) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **2010**, *114* (23), 7830−43.

(35) Barca, G. M. J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J. E.; Fedorov, D. G.; Gour, J. R.; Gunina, A. O.; Guidez, E.; Harville, T.; Irle, S.; Ivanic, J.; Kowalski, K.; Leang, S. S.; Li, H.; Li, W.; Lutz, J. J.; Magoulas, I.; Mato, J.; Mironov, V.; Nakata, H.; Pham, B. Q.; Piecuch, P.; Poole, D.; Pruitt, S. R.; Rendell, A. P.; Roskop, L. B.; Ruedenberg, K.; Sattasathuchana, T.; Schmidt, M. W.; Shen, J.; Slipchenko, L.; Sosonkina, M.; Sundriyal, V.; Tiwari, A.; Galvez Vallejo, J. L.; Westheimer, B.; Włoch, M.; Xu, P.; Zahariev, F.; Gordon, M. S. Recent developments in the general atomic and

https://doi.org/10.1021/acsomega.4c05857
ACS Omega 2024, 9, 36787−36794

molecular electronic structure system. *J. Chem. Phys.* **2020**, *152* (15), 154102.
(36) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; et al. The REd. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821−39.
(37) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y. RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39* (suppl_2), W511−W517.
(38) Wang, F, Becker, J-P, Cieplak, P, Dupradeau, F-Y, editors. *RED Python: Object oriented programming for Amber force fields*; Abstracts of Papers of the American Chemical Society: American Chemical Society: 1155 16TH ST, NW, WASHINGTON, DC 20036 USA, 2014.
(39) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; et al. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 147−61.
(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of chemical physics.* **1983**, *79* (2), 926−35.
(41) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry.* **2002**, *23* (16), 1623−41.
(42) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation.* **2008**, *4* (3), 435−47.
(43) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* **2015**, *1*, 19−25.
(44) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **2020**, *153* (13), 134110.
(45) Thomas, N.; Agrawal, A. A lateral electric field inhibits gel-to-fluid transition in lipid bilayers. *Soft Matter.* **2022**, *18* (34), 6437−42.
(46) Koynova, R.; Caffrey, M. Phases and phase transitions of the phosphatidylcholines. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1998**, *1376* (1), 91−145.
(47) 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* (8), 3684−3690.
(48) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics.* **1981**, *52* (12), 7182−90.
(49) Salerno, K. M.; Bernstein, N. Persistence Length, End-to-End Distance, and Structure of Coarse-Grained Polymers. *J. Chem. Theory Comput.* **2018**, *14* (4), 2219−29.
(50) Tow, G. M.; Maginn, E. J. Fully Atomistic Molecular Dynamics Simulations of Hydroxyl-Terminated Polybutadiene with Insights into Hydroxyl Aggregation. *Macromolecules.* **2020**, *53* (7), 2594−605.
(51) Curry, S.; Brick, P.; Franks, N. P. Fatty acid binding to human serum albumin: new insights from crystallographic studies. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids.* **1999**, *1441* (2−3), 131−40.
(52) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *Journal of molecular biology.* **2005**, *353* (1), 38−52.
(53) Petitpas, I.; Grüne, T.; Bhattacharya, A. A.; Curry, S. Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. *Journal of molecular biology.* **2001**, *314* (5), 955−60.
(54) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* **2021**, *10* (04), 463034.

https://doi.org/10.1021/acsomega.4c05857
ACS Omega 2024, 9, 36787−36794