Biomacromolecules

Article

Cite This: Biomacromolecules 2020, 21, 435−443

pubs.acs.org/Biomac

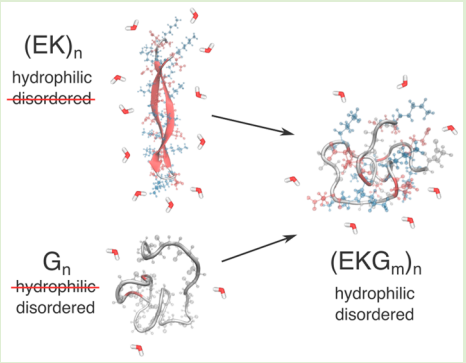
Elucidating Molecular Design Principles for Charge-Alternating Peptides

Josh Smith, Patrick McMullen, Zhefan Yuan, Jim Pfaendtner,* and Shaoyi Jiang*

Department of Chemical Engineering, University of Washington, Seattle, Washington 98195-1750, United States

Supporting Information

ABSTRACT: The therapeutic potential of protein drugs has been hindered by difficulties with long-term stability and rapid clearance from the body. Recombinant fusion proteins provide a scalable platform for engineered biologics, whereby a polypeptide domain is appended to alter the physical characteristics of a therapeutic protein and enhance its pharmaceutical viability. Two simple design principles for recombinant fusion proteins, based on the physical properties of the polypeptide domain, have been separately applied to address issues with the stability and delivery of biologics. "Conformationally disordered" peptides, exemplified by the homo amino acid peptide polyG, have been shown to increase the circulation half-life and bioactivity of protein therapeutics *in vivo*. Superhydrophilic peptides, exemplified by the alternating-charge peptide poly(EK), have been shown to increase the thermostability of proteins *in vitro*. The combination of superhydrophilicity and conformational disorder in a single fusion peptide could simultaneously address concerns regarding the stability and therapeutic lifetime of biologics. In the current work, we use enhanced sampling molecular dynamics (MD) simulations to investigate the conformational ensemble of poly(EK) and glycine-substituted poly(EK) variants and validate our structural predictions with circular dichroism (CD). We find the $(EK)_{15}$ peptide exhibits a high propensity for forming antiparallel β-strand secondary structures, which are stabilized by extensive salt bridging of the positive and negative side chains. MD simulations predict that limited glycine substitutions effectively disrupt the secondary structure and promote disordered conformations at physiologically relevant temperatures. We conclude that the conformational disorder of alternating-charge peptides should be taken into account to improve their suitability for drug delivery applications. We also contribute a computational approach to quantify conformational disorder in polypeptides, which should facilitate the *de novo* design of effective fusion proteins.



1. INTRODUCTION

Protein drugs (or biologics) have great potential for precision medicine due to their high affinity and specificity for biological targets.[1] The therapeutic potential of biologics has been hindered by their poor long-term stability and relatively short therapeutic half-life.[2] The chemical conjugation of biocompatible polymers to proteins of interest has been among the most successful and widely used approaches to overcoming these issues.[1,3−6] Bioinert, synthetic polymers, such as poly(ethylene glycol) (PEG) and polycarboxybetaine (PCB), have been shown to improve the pharmacokinetics and shelf life of various biologics.[7−9] However, polymer conjugation requires complicated synthesis and purification.

Recombinant fusion proteins provide a scalable synthesis platform for engineered biologics, whereby a polypeptide domain is appended to alter the physical characteristics of a therapeutic protein and enhance its pharmaceutical viability. Genetic fusion of an engineered polypeptide domain to the biologic eliminates the additional conjugation and purification steps required with a synthetic polymer. Using a polypeptide chain, composed of natural amino acids, is also inherently biodegradable. However, in dealing with biological building blocks, extra care must be taken to avoid unintended interactions with the biological machinery *in vivo*. The vast combinatorial space of protein sequences poses an additional challenge, as it provides a virtually endless design landscape for protective polypeptide domains.

Preliminary attempts to design protective polypeptides have been directly inspired by the most successful polymers in the chemical conjugation literature. Several conformationally disordered peptides have been designed to mimic the random structure and large hydrodynamic radius of PEG.[10−12] This class of peptide aims to reduce renal clearance, via increased size, and reduce immunogenicity, via entropic repulsion of antibodies. Conformationally disordered peptides include HAP (homo amino acid polymer, polyG), PAS, and XTEN.[10−13] These peptides have been shown to modestly improve the circulation half-life of hGH in a mouse model.[11] Much like PEG, these peptides are amphiphilic. Alternating-charge peptides have been designed to mimic the superhydrophilicity

Received: August 28, 2019
Revised: November 9, 2019
Published: November 18, 2019

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

of the zwitterionic polymers PCB.[8,9] Superhydrophilicity, or extremely strong affinity for water, can also provide a large hydrodynamic radius for reduced renal clearance and provides an enthalpic repulsion of antibodies.[14,15] The alternating-charge peptide, polyEK, has been shown to increase the thermal stability of β-lactamase enzyme without decreasing its catalytic rate.[8] More recently, polyEK was shown to increase the efficiency of an organophosphate hydrolase.[9] The conformation of polyEK has not yet been characterized in detail. A polypeptide domain combining disorder and superhydrophilicity would have the desirable feature of entropic and enthalpic resistances to unintended biological interactions.

While polyEK achieves very high water affinity, there is reason to believe it will not achieve the same degree of conformational disorder as other protective peptides. Conformationally disordered peptides such as HAP were designed with amphiphilic amino acid residues with low potential for side chain−side chain interactions.[10] PCB, the polymeric inspiration for polyEK, assumes disordered conformations due to repulsive interactions between its zwitterionic headgroups. However, salt bridging between the positive and negative side chains of polyEK could potentially stabilize regular secondary structures. Indeed, related peptide sequences have been used to create ionic self-complementary peptides (EAK and RAD), which form β-sheet structures in water and macroscopic aggregates at physiological salt concentrations.[16,17] The combination of electrostatic interactions between charged side chains and hydrophobic interactions between aliphatic alanine residues stabilizes the β-sheet structures formed by ionic-self-complementary peptides. It is unclear whether similar structures will be formed by polyEK peptides without the hydrophobic contributions of alanine side chains. The structure of "neat" alternating charge peptides (composed of only charged residues) is determined by a delicate balance between entropy and desolvation enthalpy on one side and the enthalpy of favorable side-chain interactions on the other. The balance between order and disorder for alternating charge peptides has not been extensively studied.

In the present work, we provide evidence from both simulation and experiment that $(EK)_{15}$, a computationally tractable polyEK surrogate, forms stable secondary structure in aqueous solution. We identify a high propensity for β-strand structures in $(EK)_{15}$ using enhanced sampling molecular dynamics (MD) simulations. We also find that the dilution of the superhydrophilic $(EK)_{15}$ construct with glycine (G) residues results in conformationally disordered, alternating-charge peptides. We confirm with circular dichroism spectroscopy the presence and absence of regular secondary structure in $(EK)_{15}$ and $(EKGG)_7$, respectively. We conclude that polyEK, while near optimal in terms of hydrophilicity, explores a restricted conformational space due to strong side-chain interactions. Adding glycine or small polar amino acids, such as serine or asparagine, can be used as a strategy to add flexibility and prevent the formation of stable or metastable secondary structures. Further research should be performed to explore the seeming trade-off between superhydrophilicity and structural diversity.

2. METHODS

2.1. Computational Methods. Classical molecular dynamics (MD) simulations cannot effectively capture the long charcteristic time and length scales of protein conformational changes. Enhanced sampling techniques for the characterization of protein and peptide structural ensembles, however, are well-established.[18−20] In the present work, we use parallel tempering metadynamics in the well-tempered ensemble (PTMetaD-WTE), which has previously been used to characterize the kinetics of folding pathways for small proteins and the ensemble of structures adopted for inherently unstructured peptides.[20−24] Our simulation protocol included several equilibration steps to improve sampling efficiency, including selecting starting structures in different areas of conformational phase space, unbiased equilibration MD, parallel tempering in the well-tempered ensemble (PT-WTE) to improve exchange, and production parallel tempering metadynamics (PTMetaD). We separately describe each step of the enhanced sampling simulation pipeline below and summarize these simulations in Table S1.

Selection and Equilibration of Diverse Starting Structures. Starting structures were selected from 30 residue segments of real protein structures deposited in the PDB.[25] Twelve structures were selected for diversity in shape, size, and secondary structure. Backbone coordinates were saved from VMD, and residue names were changed to match the appropriate EK-based peptide sequence. The backbone conformations of the selected structures are depicted in Figure 1. Side chains were added for E and K residues based on the Amber ff14SB force field[26] with the tleap application.

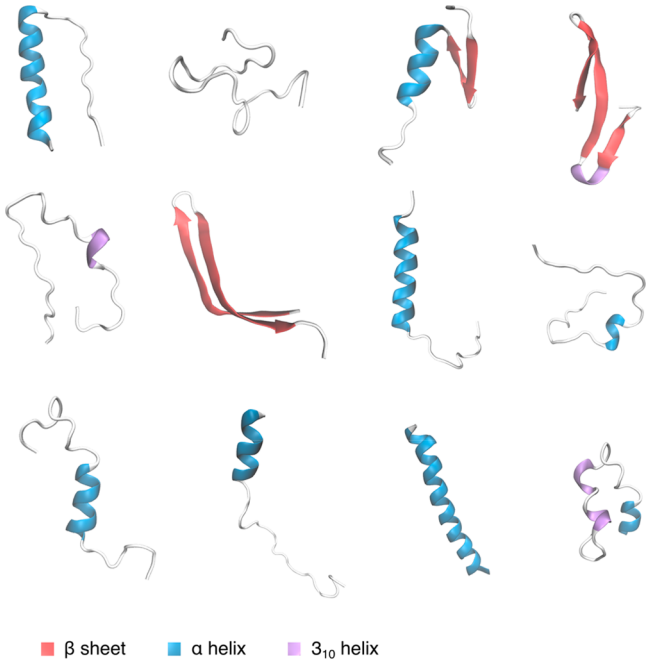


Figure 1. Starting structures for PTMetaD-WTE. $(EK)_{15}$ and G-substituted variants started with the same 12 backbone structures. These backbone segments were taken from arbitrarily selected protein crystal structure in the protein data bank, after being identified by visual inspection as diverse in secondary structure and radius of gyration. Secondary structures not explicitly labeled at the bottom of the figure (i.e., coil, turn, and bridge) are colored white.

Each subsystem was solvated with TIP3P water.[27] All simulations were performed with the GROMACS 5.1.2 simulation engine,[28,29] with PT-WTE and PTMetaD implemented through the Plumed 2.3 plugin.[30,31] Energy minimization was performed with 10000 steps of steepest descent with 1.0 nm cutoff for VDW and coulomb interactions. Solvent equilibration was performed for 250 ps in the NPT ensemble with the Bussi−Donadio−Parrinello (v-rescale) thermostat[32] and Berendsen barostat.[33] The largest box volume of any subsystem at the end of NPT equilibration was assigned to all NPT outputs to ensure identical subsystems for replica exchange in PT-WTE. From NPT output, a 250 ps annealing simulation was performed using the v-rescale thermostat which smoothly increased

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

the coupling temperature for each replica from 5 K to its PT-WTE starting temperature (between 300 and 450 K). This preliminary annealing step ensures configurations and velocities fed to PT-WTE are appropriate for initial temperatures and prevents the deposition of large potential energy bias for unequilibrated structures. All the data and PLUMED input files required to reproduce the results reported in this paper are available on PLUMED-NEST (www.plumed-nest.org), the public repository of the PLUMED consortium,[34] as plumID:19.062.

Parallel-Tempering in the Well-Tempered Ensemble (PT-WTE). PT-WTE was performed as described by Bonomi and Parrinello.[20] The 12 replicas were initialized at 12 temperatures geometrically distributed (T_{n+1}/T_n = constant) between 300 and 450 K. Every 250 MD steps (0.5 ps), the MD engine attempts to exchange the atomic coordinates of replicas at neighboring temperatures based on a Boltzmann exchange probability. Gaussian bias hills with height 4.18 kJ/mol and bias factor 30 are added to the potential energy of the system every 250 steps (0.5 ps). PT-WTE runs for EK and EKGGG peptides were performed for initial exploration of accessible conformational space and extended 300 ns. Because the exchange probability for these simulations leveled off before 100 ns, PT-WTE was limited to 100 ns to build bias potentials for EKG, EGKG, EKGG, and GG. Once this step is completed, the final energy biases are used to improve the efficiency of the subsequent parallel tempering simulations, all of which begin from the end point(s) of this initial 100 PT simulation. Figure S1 shows that all replicas in the PTMetaD simulation regularly visited each of the 12 temperature levels.

Parallel Tempering Metadynamics (PTMetaD). PTMetaD was used with the same temperature distribution and with output conformations and bias potentials from PT-WTE. An identical exchange protocol was employed for parallel tempering, and the radius of gyration was biased to force peptides to explore extended and collapsed configurations. This approach encourages each system to explore a range of compact and extended states, without biasing toward a specific type of secondary structure. Bias was deposited with a stride of 250 steps (0.5 ps), an initial hill height of 2.0 kJ/mol, and a bias factor of 15 for the radius of gyration; the width of the Gaussians was 0.005 nm. Sprenger et al. showed infrequently updating the potential energy bias can improve sampling efficiency,[35] so we additionally deposited potential energy bias every 250000 steps (500 ps), with an initial hill height of 4.18 kJ/mol and a bias factor of 30. The necessary input files to reproduce these simulations are available at https://github.com/UWPRG/ek-conformation-project.

Unbiased Production MD Simulations. Long unbiased production MD simulations were performed to assess the stability of likely EK structures emerging from the PT and MetaD simulations. Microsecond long NPT simulations using the Parrinello−Rahman barostat[36] and v-rescale thermostat were performed for EK at 300 and 450 K in a cubic box with a side length of 8 nm and packed with 16766 water molecules. Figure S2 shows that the β-strand starting structure for the unbiased simulations fits well within the cubic box and does not interact with its nearest periodic image. Shorter simulations (500 ns) were also conducted with identical MD inputs excepting mutations in the peptide sequence to DK and EKG. All peptides studied were net neutral and simulated in pure water. The equilibration and production simulations are summarized in Table S2. We also checked the stability effects of physiologically relevant salt concentrations by simulating EK in 0.154 and 0.308 M solutions of NaCl (corresponding to approximately once and twice the concentration of normal saline, respectively) and found no destabilization in salt solutions on the microsecond time scale (results not shown).

Analysis of Biased Trajectories. The PTMetaD method produces biased statistical ensembles of the conformations sampled at each temperature. To obtain equilibrim ensemble averages of unbiased quantities in the metadynamics simulations, we used the time-independent free energy estimator and reweighting scheme proposed by Tiwary and Parrinello to recover the unbiased statistics for each conformational ensemble.[37] We applied the resultant frame weights to calculate ensemble average observables for each peptide at each temperature. All trajectory analysis (including reweighting) was performed using in-house python scripts and analysis functions in the MDTraj python library.[38] We used the DSSP algorithm for residue-wise structure assignment, as implemented in "mdtraj.compute_dssp".[39] We used the Shrake and Rupley algorithm to calculate solvent accessible surface area (SASA), as implemented in "mdtraj.shrake_rupley". We used the "mdtraj.rmsd" function to calculate the Cα root-mean-squared deviation (RMSD) for unbiased and biased trajectories and an in-house implementation of the clustering method introduced by Daura et al. (sometimes called the gromos method) to identify the lowest energy structures.[40] Conformational entropy was calculated as the mean of the residue-wise conformational entropy described by Baruah et al.[41] This formulation is based on the Shannon entropy of dihedral angles, giving higher entropy scores to residues that sample a broader distribution of angles. These analysis scripts are publicly available at https://github.com/UWPRG/ek-conformation-project.

2.2. Experimental Materials and Methods. *Peptide Materials and Synthesis.* Fmoc-Glu(tBu)-OH, H-Lys(Boc)-OH, Fmoc-Gly-OH, ethyl cyanohydroxyiminoacetate (Oxyma), and Fmoc-Rink Amide resin (0.54 mmol/g) were purchased from AAPPTec. *N*,*N*′-Diisopropylcarbodiimide (DIC) and piperidine were obtained from Chemimpex. Trifluoroacetic acid (TFA) and 3,6-dioxa-1,8-octanedithiol (DODT) were purchased from TCI America. Diisopropylethylamine (DIEA), *N*,*N*-dimethylformamide (DMF), and triisopropylsilane (TIS) were obtained from Sigma-Aldrich.

An amino acid dimer (Fmoc-Glu(tBu)-Lys(Boc)-OH) was first obtained by reacting Fmoc-Glu(tBu)-OH and H-Lys(Boc)-OH and used as EK monomer during the synthesis of (EK)10 and (EKGG)7. Microwave-assisted peptide syntheses were performed using Fmoc solid-phase peptide synthesis (SPPS) strategy on a microwave peptide synthesizer (CEM Corporation, Matthews, NC). Default standard 90 °C deprotection and coupling methods were used for synthesis. Peptide preparation was set at scale of 0.1 mmol and starting on Fmoc-Rink Amide resin. Fmoc deprotection was achieved in 20% piperidine−DMF solution, and the final peptides cleavage were performed in a TFA/scavenger cocktail (TFA/TIS/H_2O/DODT = 92.5/2.5/2.5/2.5) for 3 h at room temperature.

Circular Dichroism. Peptides were dissolved in neat water to a concentration of 150 μM and pH adjusted to 7.4. The resulting solutions were transferred to a 1.0 mm thickness quartz cuvette for analysis on a Jasco J-720 circular dichroism machine. For each condition, eight replicate spectra were obtained and averaged with wavelengths ranging from 215 to 250 nm. Urea introduces strong background absorbance and prevents the reliable determination of CD spectra below this range, but the spectra for peptide solutions in neat water are provided in Figure S3. The resulting spectra were smoothed by using the Jasco circular dichroism software. For conditions with urea, urea was added in 2 M increments to the peptide-containing solutions.

3. RESULTS AND DISCUSSION

We used PTMetaD simulations to assess the conformational freedom of polyEK. We performed atomistic enhanced sampling simulations of an EK 30-mer, $(EK)_{15}$, as a surrogate for longer polyEK peptides because it would be computationally infeasible to simulate the 30 kDa peptide used in experiments. The conformational ensemble of this relatively short peptide provides insight into the structural preferences of polyEK. We performed identical PTMetaD simulations for a 30-residue glycine peptide (G_{30}) for comparison. We chose to use this polyG surrogate as a conformationally disordered control because glycine-rich peptides are known to adopt predominantly random coil structures in aqueous solution. Also, the glycine-based homo amino acid peptide (HAP), the first reported conformationally disordered protective polypeptide domain, was designed based on this observation.[10] We routinely refer to the set of structures visited during the

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

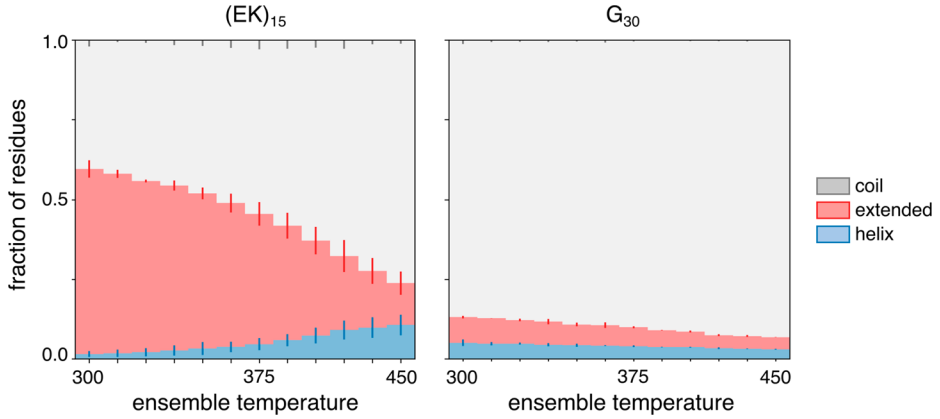


Figure 2. Average fraction of $(EK)_{15}$ and G_{30} residues assigned to each of the simplified structure classes by using the DSSP algorithm. Stacked bars are included for the structural ensemble sampled at each of the 12 temperatures during the PTMetaD simulation. The gray, red, and blue fractions of the bar represent the fraction of residues assigned to coil, extended, and helix structure classes, respectively. For example, the average fraction of residues assigned to coil, extended, and helix for $(EK)_{15}$ at 300 K were 0.41, 0.58, and 0.01, respectively. Error bars represent ±1 standard deviation of the mean structure fraction calculated for three separate 100 ns blocks.

PTMetaD simulation for a given sequence, at a given temperature, as a "conformational ensemble". For example, we constructed 12 conformational ensembles for $(EK)_{15}$ by sampling in parallel at 12 temperatures. While the low-temperature ensembles are of the most practical interest, corresponding to ambient experimental and physiological conditions, high-temperature ensembles can provide additional informational about the thermal stability of secondary structure motifs.

$(EK)_{15}$ Samples a Restricted Number of Structures Relative to Disordered G_{30}. Our primary objective was to characterize the relative stability of collapsed versus extended and structured versus unstructured conformations of alternating-charge EK-based peptides. We performed PTMetaD simulations for accelerated sampling of the vast conformational space, biasing the radius of gyration of the peptide backbone and exchanging replicas between 12 temperatures geometrically distributed from 300 to 450 K. This approach provided a biased conformational ensemble at each temperature for $(EK)_{15}$ and the G_{30} control. We recovered unbiased statistics from each biased conformational ensemble by reweighting with the time-independent estimator proposed by Tiwary and Parrinello.[37] We calculated the ensemble-averaged secondary structure content, the backbone conformational entropy, the number of salt bridges, and the hydrophilic solvent accessible surface area (SASA) for each ensemble. We have tabulated these ensemble-averaged physical characteristics across the range of temperatures for $(EK)_{15}$ and for G_{30} in Table S3. Figure 2 shows the secondary structure content of the conformational ensemble generated at all 12 temperatures for $(EK)_{15}$ and G_{30}. In this case, we used the simplified DSSP structure classification scheme, so that residues with loop, turn, and irregular structures are assigned to "coil", residues participating in isolated β-bridges or extended β-strands (ladders) are assigned to "extended", and residues in α helix, π helix, or 3_{10} helix are assigned to "helix". We provide the detailed breakdown of these structural classes to their component parts in Figure S4.

In line with theoretical[42] and experimental[10,43] characterizations of the structure of glycine-rich peptides in aqueous solution, an average of 87.5% of G_{30} residues in the conformations sampled at 300 K were assigned to the coil structure class. Practically, this means that an average of only 3.75 out of 30 residues were assigned to the extended or helix class for each conformation in the 300 K conformational ensemble. The small fraction of residues assigned to extended and helix structures for G_{30} are likely due to chance alignments of a few residues in the primarily disordered peptide chain.[44] Indeed, performing DSSP with the expanded list of secondary structure types shows the "extended" fraction of G_{30} is primarily composed of β-bridges rather than β-strands (Figure S4).

In contrast, only 41% of $(EK)_{15}$ residues were assigned to the coil structure class at 300 K. For $(EK)_{15}$, about 17.4 of 30 residues were assigned to the extended class, suggesting the presence of persistent, extended β-strands. The high fraction of $(EK)_{15}$ residues participating in β-strands is seemingly at odds with empirical evidence that suggests E and K are underrepresented in β-sheets compared to most amino acids.[45−47] However, it has been observed experimentally and explained theoretically that β-strand propensities are largely context dependent.[48−50] Because long, uninterrupted runs of charged amino acids are uncommon in natural proteins, empirical estimates of the β-sheet propensities for E and K underestimate the likelihood of β-sheet formation in the context of charge-alternating peptides. It has also been suggested that pairwise side-chain interactions are required to stabilize β-strands.[49] Because salt bridges represent the strongest noncovalent side chain interactions, it is reasonable to suspect that E−K side-chain interactions provide stabilizing energy for the observed β-strand frequency. The potential for $(EK)_{15}$ to form β-strand configurations would also be consistent with the observation that the ionic self-complementary peptide EAK forms β-sheets in aqueous solution, despite the high propensity of alanine for α-helix (and slight preference of E and K for α-helix) observed in globular proteins and polyalanine peptides.[17,46,51]

Hyperstable β-Strand Moiety Dominates $(EK)_{15}$ Conformation in Aqueous Solution. We performed a clustering analysis to identify the molecular features of stable extended structures in the conformational ensemble of $(EK)_{15}$ at 300 K. We calculated the pairwise C_α root-mean-squared distance (RMSD) for every pair of structures in the ensemble, resulting in an $N \times N$ distance matrix, where N is the number

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

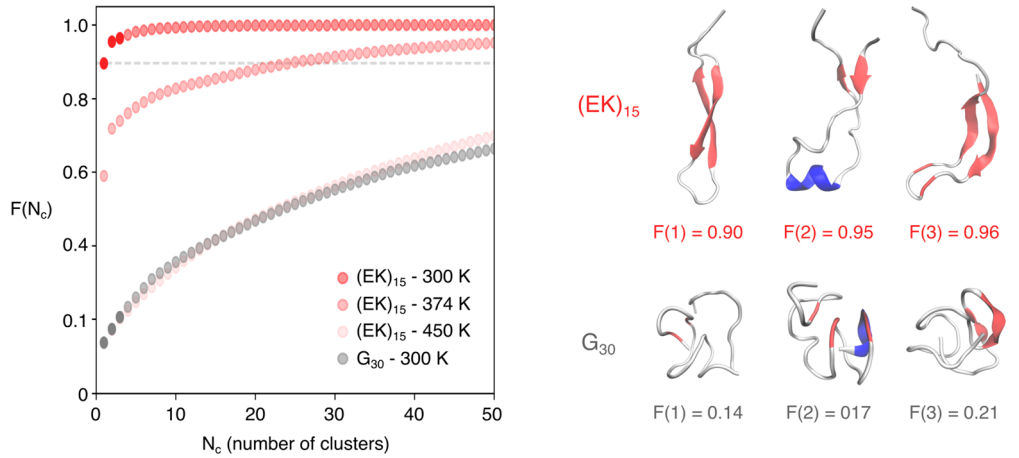


Figure 3. Cumulative distribution function (CDF) for structure clusters at each temperature from PT-MetaD simulation. Each cluster contains structures with a C_α RMSD less than 0.5 nm from the central structure of the cluster, identified by using the gromos clustering method.[40] The horizontal dashed line corresponds to the probability that structures in the $(EK)_{15}$ 300 K ensemble fall into the most likely cluster (~0.90). For the conformational ensembles of $(EK)_{15}$ and G_{30} at 300 K, the central structures of the top three clusters are included on the right. These structures correspond to the first three points in the respective CDFs on the left, which are highlighted with increased opacity.

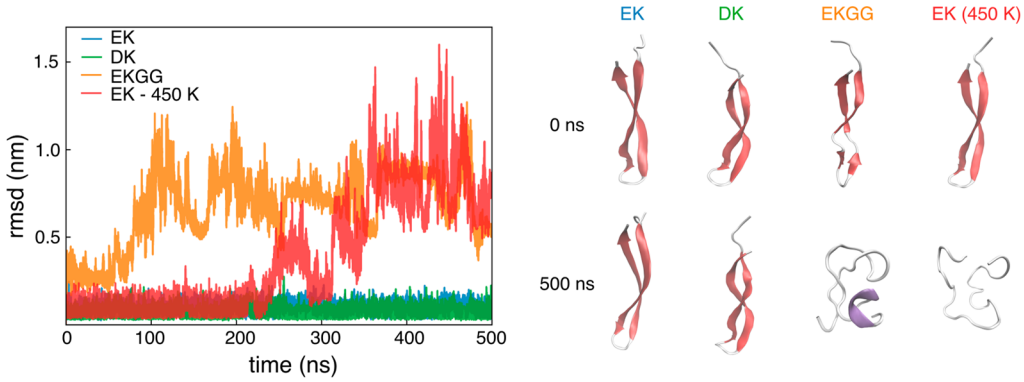


Figure 4. C_α RMSD for $(EK)_{15}$ and DK/EKGG variants from 500 ns of unbiased simulations. Starting from the stable antiparallel β-strand conformation identified as stable in the PTMetaD simulations. The unchanged $(EK)_{15}$ peptide remained stable for the whole simulation at 300 K (green) and began to unfold after 200 ns at 450 K (red). The $(DK)_{15}$ peptide also remained folded for the entire simulation at 300 K (blue). The EKGG mutant, with half as many E−K salt bridges to start, unfolded in less than 100 ns at 300 K (orange). Conformations for each sequence at 0 and 500 ns are included on the right.

of structures in the ensemble. We followed the peptide structure clustering algorithm of Daura et al.[40] (sometimes termed the "gromos" method) with a RMSD cutoff of 0.5 nm. This algorithm identifies the structure with the greatest number of neighbors (other structures with C_α RMSD < 0.5 nm) and designates that structure as the "central structure" of a cluster composed of all its neighbors. These structures are removed from consideration, and the algorithm iteratively identifies clusters and their central structures until no neighboring structures remain. After clustering, each cluster was assigned a probability based on the number of constituent structures and their associated unbiased weights. Figure 3 shows the cumulative distribution function (CDF) of structures for $(EK)_{15}$ (at 300, 374, and 450 K) and for G_{30} (at 300 K), in terms of the number of clusters required to account for each structure given the 0.5 nm RMSD cutoff. Figure 3 also shows representative structures for the three most populated clusters for $(EK)_{15}$ and G_{30} at 300 K.

The clustering results show that $(EK)_{15}$ explored a very restricted set of conformations at 300 K due to the predominance of a twisted, antiparallel β-strand moiety. The likelihood that a structure selected at random from the 300 K ensemble would contain this moiety was ~0.9. As expected, persistent secondary structures appear to be stabilized by extensive side-chain salt bridging. By inspecting the central structure of the most populous cluster for $(EK)_{15}$ at 300 K, we found that the β-strand moiety was stabilized by alternating salt bridges forming a zipper-like pattern on either side of the strand. The presence of antiparallel β-strand moieties in the second and third most populous clusters further supports the propensity of $(EK)_{15}$ for antiparallel β-strand structure. While transient β-bridges and short helices will unavoidably occur in polypeptide domains, the persistent secondary structure, like the β-strand moiety identified here, could be deleterious for fusion protein applications in terms of immune recognition and quaternary aggregation. Aggregation has not been reported for *in vitro* studies of polyEK fusion proteins, but the concentrations used in these studies were orders of magnitude lower than the protein concentrations currently targeted in therapeutic protein formulations (2−20 nM vs 1−3 mM).[8,9] Further engineering of the polyEK sequence may be required to disrupt stable secondary structures and encourage a more entropically favorable disordered conformations in challenging real-world conditions.

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

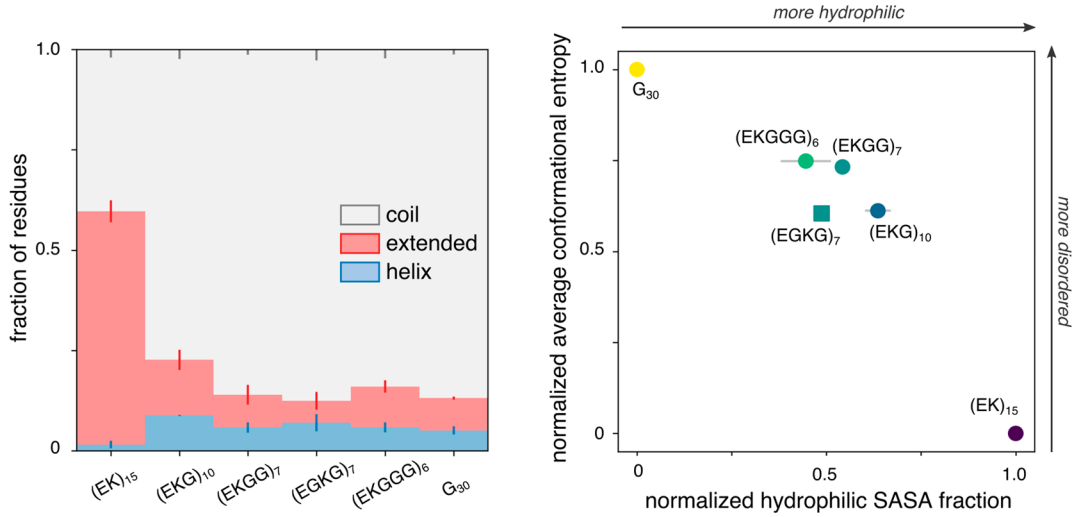


Figure 5. Ensemble average secondary structure fractions and physical properties of $(EK)_{15}$, $(EKG)_{10}$, $(EKGG)_7$, $(EGKG)_7$, $(EKGGG)_6$, and $(G)_{30}$ at 300 K. (left) The gray, red, and blue fractions of the bar represent the fraction of residues assigned to coil, extended, and helix structure classes, respectively. (right) Average conformational entropy vs fraction hydrophilic SASA. Conformational entropy is normalized so that the maximum and minimum values correspond to G_{30} and $(EK)_{15}$, respectively. Conversely, the fraction hydrophilic SASA is normalized so that the maximum and minimum values correspond to $(EK)_{15}$ and G_{30}, respectively. High values of disorder and hydrophilicity are desirable. Error bars represent ±1 standard deviation of the mean calculated for three separate 100 ns blocks.

Glycine Mutations Disrupt Stable β-Strand Structure in Unbiased Simulations. To further investigate the stability of the β-strand structure identified in PTMetaD simulations, we performed long classical MD simulations of $(EK)_{15}$ at 300 and 450 K. We also simulated two peptides derived from $(EK)_{15}$ to assess the potential for amino acid substitutions to disrupt the stable conformation. For each of these peptides, we started with a β-sheet backbone conformation taken from the dominant cluster of the 300 K conformational ensemble for $(EK)_{15}$ and then mutated some of the side chains. For the first mutant, we replaced all the negatively charged E side chains to aspartate (D). We refer to this mutant as DK. We chose not to mutate positively charged K side chains to arginine (R) side chains considering the potential role of R residues in protein−protein binding, evidenced by their prevalence in protein−protein interfaces and their role as hot spot residues in such interfaces.[52,53] For the second peptide, we removed approximately half of the EK side chains, resulting in the sequence $(EKGG)_4G_2(EKGG)_3$. We refer to this peptide as EKGG, although the EKGG repeat sequence is disrupted at the turn of the β-strand to preserve the maximum number of initial E−K salt bridges in the β-strand portion. Figure 4 shows the C_α RMSD for the β-sheet residues over the first 500 ns of each unbiased simulation. The EK and DK simulations at 300 K were continued for a full microsecond to verify they remained folded in their initial configurations.

The β-strand conformation for $(EK)_{15}$ remained stable for over 1 μs at 300 K. Figure 4 shows it took 200 ns to disrupt this conformation of $(EK)_{15}$ at 450 K, explaining the persistence of this structure in the 300 ns PTMetaD simulation (Figure 2). The DK peptide also remained folded for over 500 ns, while the EKGG peptide unfolded completely in less than 100 ns. With slightly greater than 50% of the EK pairs mutated to GG, the likelihood of cooperative salt bridging and the formation of extended β-strands is much lower. These results suggest that dilution with G is an effective design strategy for protective polypeptide domains that are both disordered and superhydrophilic.

To test how the dilution factor affects the conformational disorder and hydrophilicity of EK-based peptides, we performed PTMetaD simulations for several G-substituted variants of $(EK)_{15}$. These peptides were $(EKG)_{10}$, $(EKGG)_7$, $(EGKG)_7$, and $(EKGGG)_6$. For each sequence, we used the same PTMetaD protocol (and the same 12 starting structures) as described for $(EK)_{15}$ and G_{30}. We also performed identical analyses to calculate ensemble averages for secondary structure content, backbone entropy, and hydrophilic SASA. Figure 5 shows a comparison of these physical features, at 300 K, for each sequence. The backbone entropy was scaled so that the values for G_{30} and $(EK)_{15}$ were set to 1 and 0, respectively. The hydrophilic SASA fraction was scaled so that the values for G_{30} and $(EK)_{15}$ were set to 0 and 1, respectively.

The 300 K conformational ensemble for each G-diluted sequence contained significantly less ordered secondary structure than for $(EK)_{15}$. $(EKGG)_7$ and $(EGKG)_7$ were especially close to G_{30} (0.87 ± 0.10) in terms of coil structure fraction with 0.86 ± 0.11 and 0.88 ± 0.09, respectively. The drastic decrease in ordered secondary structure, even for $(EKG)_{10}$, suggests that a small fraction of G substitutions effectively destabilizes β-strands in EK-based peptides (Figure S5). As expected, the increased conformational freedom came at the cost of hydrophilicity. While progressive substitution with glycine increased the ensemble average backbone entropy relative to $(EK)_{15}$, the average hydrophilic SASA fraction was decreased. However, we also found that for $(EK)_{15}$ salt bridges generally decreased hydrophilic SASA (Figure S6). Therefore, we expect that substituting small polar amino acids, like serine and glutamine, or proline in place of glycine would similarly promote disorder with less detriment to hydrophilicity.

The slight increase in the mean backbone entropy for $(EKGGG)_6$ relative to $(EKGG)_7$ suggests there are diminishing entropic returns for more than 50% G substitution of the EK-based peptide sequence. Also noteworthy is the difference between $(EKGG)_7$ and $(EGKG)_7$ in terms of backbone entropy, which suggests that sequence and composition must be considered separately even within the simplest contexts of

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

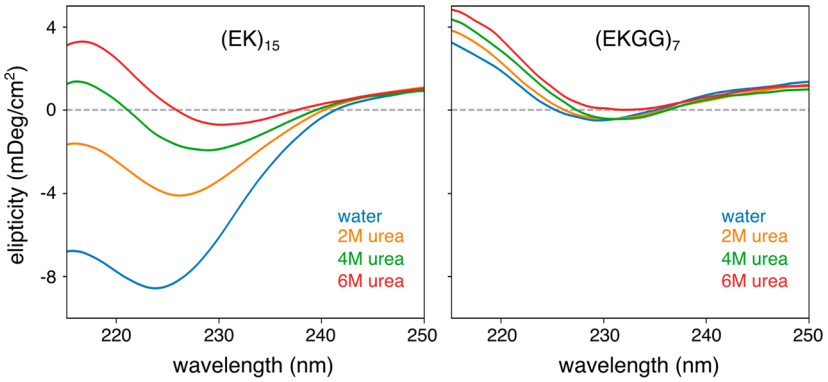


Figure 6. CD spectra for $(EK)_{15}$ and $(EKGG)_7$ in neat water and solutions with various urea concentrations. (left) CD spectra for $(EK)_{15}$ shows the peptide is assumes some stable secondary structure in neat water, which is progressively disrupted as urea concentration is increased. (right) Spectra for $(EKGG)_7$ are relatively insensitive to urea, suggesting the peptide is completely disordered across the concentration range tested here.

peptide design. It is expected that the conformational ensemble of $(EGKG)_7$ is more restricted than that for $(EKGG)_7$ because the isolated E and K residues are more amenable to tightly packed random coils than the EK pair. This hypothesis is supported by the lower average radius of gyration and greater number of G−G contacts for $(EGKG)_7$ relative to $(EKGG)_7$ (Table S4). The mean scaled backbone entropy and scaled hydrophilic SASA fraction were both greater than 0.5 for $(EKG)_{10}$ and for $(EKGG)_7$. Given that $(EKG)_{10}$ contains the repetitive X−G−Y moiety characteristic of the collagen triple helix, its potential for intermolecular interactions may be greater than predicted by single-molecule simulations.

Circular Dichroism Confirms the Presence of β Sheet in $(EK)_{15}$ and Structure Breaking in $(EKGG)_7$. We performed circular dichroism spectroscopy (CD) to verify the structural predictions from our simulations. We synthesized $(EK)_{15}$ and $(EKGG)_7$ and measured their CD spectra in neat water, 2 M urea, 4 M urea, and 6 M urea solutions. Changes in the CD spectra with additional urea should reflect a shift toward a disordered, denatured state. We selected $(EKGG)_7$ as the G-substituted variant of $(EK)_{15}$ based on our observation that 50% glycine substitutions disrupted the $(EK)_{15}$ β-strand in the classical MD simulations (Figure 5). The CD spectra for $(EK)_{15}$ and $(EKGG)_7$ are provided in Figure 6.

Consistent with previous studies of self-complementary, EK-based peptides, the β-strand character of $(EK)_{15}$ is apparently based on the negative band near 225 nm. Typically, β-sheets exhibit a negative band near 218 nm.[54] However, $(EK)_{15}$ exhibits a red-shifted negative band that is consistent with the CD spectra for β-turns.[55] As the denaturant urea is added to solution at increasing concentrations, the β-strand signature of the CD spectrum disappears. The spectrum for $(EKGG)_7$ in water, on the other hand, shows no sign of secondary structure other than random coil. The relatively small effect of urea on the CD spectrum supports this conclusion.

4. CONCLUSIONS

In the present work, we find the alternating-charge peptide $(EK)_{15}$ has a high propensity for self-interaction which manifests in the formation of persistent secondary structure motifs. We expect this tendency to be present for other purely alternating-charge sequences (such as DK investigated here), but the addition of glycine reliably disrupts secondary structure. We provide molecular level insights about the mechanism of disruption and suggest lead sequence candidates for future protective peptides.

While the glycine-substituted EK sequences studied here were intermediate to $(EK)_{15}$ and $(G)_{30}$ in terms of hydrophilicity and conformational disorder, it is possible to engineer mutated sequences with greater hydrophilic SASA by using small polar amino acids, like serine and glutamine, in place of glycine. The hydrodynamic radius may be tuned by introducing proline, as demonstrated for PAS-based peptides. The current work shows how molecular modeling can provide useful molecular insights to support the *de novo* design of more complicated and performant protective polypeptides.

An important consideration is that these simulations are of peptides with chain lengths 2−6 times shorter than typically used in experiment. For example, protective polyEK has been reported for peptides from 10 to 30 kDa in weight, whereas the $(EK)_{15}$ peptide is slightly less than 5 kDa. Atomistic simulations of peptides on the experimental length scale are still computationally intractable, even with enhanced sampling methods. Advances in reliable coarse-grained force fields should bring this length scale into the realm of possibility in the coming years. However, given the consistency of our G_{30} control with experimental observations for longer glycine-rich polypeptides, we are confident that our simulations reflect true structural tendencies observable in the macroscopic world.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biomac.9b01191.

Additional figures and tables that provide further simulation details and analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail jpfaendt@uw.edu.
*E-mail sjiang@uw.edu.

ORCID

Jim Pfaendtner: 0000-0001-6727-2957

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

REFERENCES

(1) Pisal, D. S.; Kosloski, M. P.; Balu-Iyer, S. V. Delivery of Protein Therapeutics. *J. Pharm. Sci.* **2010**, *99* (6), 2557−2575.
(2) Vaishya, R.; Khurana, V.; Patel, S.; Mitra, A. Long-Term Delivery of Protein Therapeutics. *Expert Opin. Drug Delivery* **2015**, *12* (3), 415−440.
(3) Turecek, P. L.; Bossard, M. J.; Schoetens, F.; Ivens, I. A. PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *J. Pharm. Sci.* **2016**, *105* (2), 460−475.
(4) Banerjee, S. S.; Aher, N.; Patil, R.; Khandare, J. Poly(Ethylene Glycol)-Prodrug Conjugates: Concept, Design, and Applications. *J. Drug Delivery* **2012**, *2012*, 1−17.
(5) Veronese, F. M.; Mero, A. The Impact of PEGylation on Biological Therapies. *BioDrugs* **2008**, *22* (5), 315−329.
(6) Swierczewska, M.; Lee, K. C.; Lee, S. What Is the Future of PEGylated Therapies? *Expert Opin. Emerging Drugs* **2015**, *20* (4), 531−536.
(7) Kozlowski, A.; Milton Harris, J. Improvements in Protein PEGylation: Pegylated Interferons for Treatment of Hepatitis C. *J. Controlled Release* **2001**, *72* (1−3), 217−224.
(8) Liu, E. J.; Sinclair, A.; Keefe, A. J.; Nannenga, B. L.; Coyle, B. L.; Baneyx, F.; Jiang, S. EKylation: Addition of an Alternating-Charge Peptide Stabilizes Proteins. *Biomacromolecules* **2015**, *16* (10), 3357−3361.
(9) Liu, E. J.; Jiang, S. Expressing a Monomeric Organophosphate Hydrolase as an EK Fusion Protein. *Bioconjugate Chem.* **2018**, *29* (c), 3686−3690.
(10) Schlapschy, M.; Theobald, I.; Mack, H.; Schottelius, M.; Wester, H. J.; Skerra, A. Fusion of a Recombinant Antibody Fragment with a Homo-Amino-Acid Polymer: Effects on Biophysical Properties and Prolonged Plasma Half-Life. *Protein Eng., Des. Sel.* **2007**, *20* (6), 273−284.
(11) Schlapschy, M.; Binder, U.; Börger, C.; Theobald, I.; Wachinger, K.; Kisling, S.; Haller, D.; Skerra, A. PASylation: A Biological Alternative to PEGylation for Extending the Plasma Half-Life of Pharmaceutically Active Proteins. *Protein Eng., Des. Sel.* **2013**, *26* (8), 489−501.
(12) Breibeck, J.; Skerra, A. The Polypeptide Biophysics of Proline/Alanine-Rich Sequences (PAS): Recombinant Biopolymers with PEG-like Properties. *Biopolymers* **2018**, *109* (1), e23069.
(13) Schellenberger, V.; Wang, C. W.; Geething, N. C.; Spink, B. J.; Campbell, A.; To, W.; Scholle, M. D.; Yin, Y.; Yao, Y.; Bogin, O.; Cleland, J. L.; Silverman, J.; Stemmer, W. P. A Recombinant Polypeptide Extends the in Vivo Half-Life of Peptides and Proteins in a Tunable Manner. *Nat. Biotechnol.* **2009**, *27* (12), 1186−1190.
(14) Cao, Z.; Jiang, S. Super-Hydrophilic Zwitterionic Poly-(Carboxybetaine) and Amphiphilic Non-Ionic Poly(Ethylene Glycol) for Stealth Nanoparticles. *Nano Today* **2012**, *7* (5), 404−413.
(15) Liu, S.; Jiang, S. Zwitterionic Polymer-Protein Conjugates Reduce Polymer-Specific Antibody Response. *Nano Today* **2016**, *11* (3), 285−291.
(16) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. Spontaneous Assembly of a Self-Complementary Oligopeptide to Form a Stable Macroscopic Membrane. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (8), 3334−3338.
(17) Zhang, S.; Holmes, T. C.; DiPersio, C. M.; Hynes, R. O.; Su, X.; Rich, A. Self-Complementary Oligopeptide Matrices Support Mammalian Cell Attachment. *Biomaterials* **1995**, *16* (18), 1385−1393.
(18) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Chem. Phys.* **2000**, *113* (15), 6042−6051.
(19) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128* (41), 13435−13441.
(20) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104* (19), 1−4.
(21) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *J. Chem. Theory Comput.* **2012**, *8* (7), 2189−2192.
(22) Bernetti, M.; Masetti, M.; Pietrucci, F.; Blackledge, M.; Jensen, M. R.; Recanatini, M.; Mollica, L.; Cavalli, A. Structural and Kinetic Characterization of the Intrinsically Disordered Protein SeV NTAIL through Enhanced Sampling Simulations. *J. Phys. Chem. B* **2017**, *121* (41), 9572−9582.
(23) Burney, P. R.; White, N.; Pfaendtner, J. Structural Effects of Methionine Oxidation on Isolated Subdomains of Human Fibrin D and AC Regions. *PLoS One* **2014**, *9* (1), e86981.
(24) Barducci, A.; Bonomi, M.; Prakash, M. K.; Parrinello, M. Free-Energy Landscape of Protein Oligomerization from Atomistic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (49), E4708−E4713.
(25) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, J.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58* (6), 899−907.
(26) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.
(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(28) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(29) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(30) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961−1972.
(31) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604−613.
(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
(33) 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.
(34) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16* (8), 670−673.
(35) Sprenger, K. G.; Pfaendtner, J. Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces. *Langmuir* **2016**, *32* (22), 5690−5701.

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443

(36) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182−7190.
(37) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736−742.
(38) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109* (8), 1528−1532.
(39) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577−2637.
(40) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38* (1−2), 236−240.
(41) Baruah, A.; Rani, P.; Biswas, P. Conformational Entropy of Intrinsically Disordered Proteins from Amino Acid Triads. *Sci. Rep.* **2015**, *5*, 1−12.
(42) Gō, M.; Hesselink, F. T.; Gō, N.; Scheraga, H. A. Molecular Theory of the Helix-Coil Transition in Polyamino Acids. III. Evaluation and Analysis of Sand u for Polyglycine and PolY-L-Alanine in Water. *J. Chem. Phys.* **1971**, *54* (10), 4489−4503.
(43) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. Helix-Coil Stability Constants for the Naturally Occurring Amino Acids in Water. III. Glycine Parameters from Random Poly (Hydroxybutylglutamine-Co-Glycine). *Macromolecules* **1971**, *4* (4), 417−424.
(44) Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. The Concept of a Random Coil: Residual Structure in Peptides and Denatured Proteins. *Folding Des.* **1996**, *1* (5), R95−106.
(45) Chou, P. Y.; Fasman, G. D. Conformational Parameters for Amino Acids in Helical, β-Sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* **1974**, *13* (2), 211−222.
(46) Costantini, S.; Colonna, G.; Facchiano, A. M. Amino Acid Propensities for Secondary Structures Are Influenced by the Protein Structural Class. *Biochem. Biophys. Res. Commun.* **2006**, *342* (2), 441−451.
(47) Bhattacharjee, N.; Biswas, P. Position-Specific Propensities of Amino Acids in the β-Strand. *BMC Struct. Biol.* **2010**, *10* (1), 29.
(48) Minor, D. L., Jr.; Kim, P. S. Context Is a Major Determinant of Beta-Sheet Propensity. *Nature* **1994**, *371*, 264−267.
(49) Yang, A. S.; Honig, B. Free Energy Determinants of Secondary Structure Formation: II. Antiparallel B-Sheets. *J. Mol. Biol.* **1995**, *252* (3), 366−376.
(50) Fujiwara, K.; Toda, H.; Ikeguchi, M. Dependence of Alpha-Helical and Beta-Sheet Amino Acid Propensities on the Overall Protein Fold Type. *BMC Struct. Biol.* **2012**, *12* (1), 18.
(51) Yang, A. S.; Honig, B. Free Energy Determinants of Secondary Structure Formation: I. A-Helices. *J. Mol. Biol.* **1995**, *252* (3), 351−365.
(52) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280* (1), 1−9.
(53) Jones, S.; Marin, A.; Thornton, J. M. Protein Domain Interfaces: Characterization and Comparison with Oligomeric Protein Interfaces. *Protein Eng., Des. Sel.* **2000**, *13* (2), 77−82.
(54) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure Norma. *Nat. Protoc.* **2006**, *1* (6), 2876−2890.
(55) Bush, C. A.; Sarkar, S. K.; Kopple, K. D. Circular Dichroism of (β Turns in Peptides and Proteins. *Biochemistry* **1978**, *17* (23), 4951−4954.

DOI: 10.1021/acs.biomac.9b01191
Biomacromolecules 2020, 21, 435−443