ACS OMEGA

Open Access

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

Review

Tips and Tricks in the Modeling of Supramolecular Peptide Assemblies

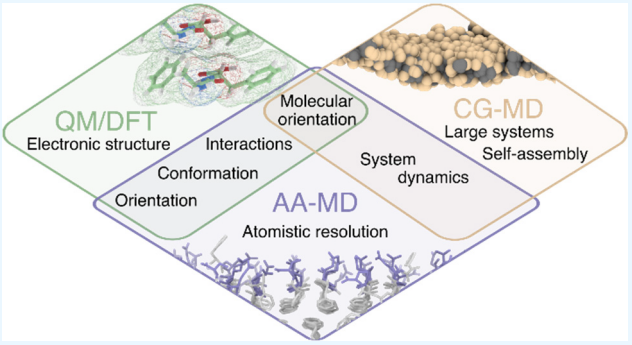
Tomasz K. Piskorz,[#] Laura Perez-Chirinos,[#] Baofu Qiao, and Ivan R. Sasselli*

Cite This: *ACS Omega* 2024, 9, 31254−31273

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Supramolecular peptide assemblies (SPAs) hold promise as materials for nanotechnology and biomedicine. Although their investigation often entails adapting experimental techniques from their protein counterparts, SPAs are fundamentally distinct from proteins, posing unique challenges for their study. Computational methods have emerged as indispensable tools for gaining deeper insights into SPA structures at the molecular level, surpassing the limitations of experimental techniques, and as screening tools to reduce the experimental search space. However, computational studies have grappled with issues stemming from the absence of standardized procedures and relevant crystal structures. Fundamental disparities between SPAs and protein simulations, such as the absence of experimentally validated initial structures and the importance of the simulation size, number of molecules, and concentration, have compounded these challenges. Understanding the roles of various parameters and the capabilities of different models and simulation setups remains an ongoing endeavor. In this review, we aim to provide readers with guidance on the parameters to consider when conducting SPA simulations, elucidating their potential impact on outcomes and validity.



I. INTRODUCTION

Supramolecular peptide assemblies (SPAs) are garnering significant attention as potential novel materials for applications in biomedicine and nanotechnology.[1−3] They emulate the capacity of proteins to self-assemble into well-ordered structures. This simplifies material synthesis, enabling the creation of structures using peptides containing fewer than 10 amino acids. Similar to the proteins they mimic, the assembled morphology and properties of the resulting material are encoded by their sequence.[4−6] Minor alterations in the amino acid sequence can induce significant structural changes and affect crystallinity, molecular mobility, and, consequently, bioactivity.[7−9] Sequence modifications also impact the material's surface, influencing interactions with cells, altering viscosity, and potentially triggering the formation of superstructures.[10−12] Therefore, rationalizing the connection between material properties and amino acid sequence has become a priority in developing customizable biomaterials.

Experimental efforts have clarified the impact of small variations in SPAs sequence, encompassing differences in amino acid nature (charged, polar or apolar, and aromatic or aliphatic), size, and position in the sequence.[12−16] While the sequence dependence of SPA properties is beyond doubt, the underlying rationale often remains elusive. Unlike the abundant repositories of proteins, crystal structures of short peptides are frequently unrepresentative of their assembled state.[17,18] Additionally, polydispersity in samples for cryogenic transmission electron microscopy (cryo-TEM) hampers the resolution of molecular details. Although techniques like Fourier-transformed infrared (FTIR), circular dichroism (CD) spectroscopies, and wide-angle X-ray scattering (WAXS) can be employed to construct models,[19−21] they are mostly unrealistic owing to the multitude of possible arrangements of peptides, and often overestimate intermolecular order. This challenge intensifies in coassemblies, where diverse peptides assemble into complex materials,[22] being often unattainable to discern the degree of mixture and the arrangement of each component in the coassembly.[23−25] Thus, despite numerous advances and the development of a rather extensive library of peptide-based supramolecular materials, the discovery of new materials often depends on serendipity. Currently, only a few design rules can be considered general, and the function-focused development of materials is not feasible, except through simple modifications in existing SPAs.

Received: March 18, 2024
Revised: June 17, 2024
Accepted: June 19, 2024
Published: July 8, 2024

Computational methods have gained popularity to bridge the experimental gap between the building block sequence and the supramolecular structures, as well as the material properties.[26,27] They facilitate a more profound understanding of the self-assembly process and the properties of the supramolecular structures, offering detailed insights into the underlying interactions, the impact of amino acid sequences, and specific environmental changes.[28−36] Their integration with experimental studies enhances the interpretation of experimental results and/or optimizes the development of materials with specific properties.[7,35−38]

The computational methods are often borrowed and adapted from protein-based methods. While this offers a range of existing tools and protocols, their applicability to SPAs is not always straightforward owing to issues such as structure uncertainty, system size, and concentration variations. Additional challenges arise in SPAs that include nonpeptide sections such as aliphatic or aromatic moieties in peptide amphiphiles (PAs) and aromatic peptide amphiphiles (APAs), respectively.[36,39] These challenges extend to validation, as the analysis protocols for computational studies in SPAs may not always align directly with experimental quantities. Consequently, a focused reassessment of these methods in their application to SPAs is imperative.

In this review, we aim to assist beginners by providing guidance on computational approaches, with specific attention to methods, system setup, and the information they yield. The focus is not to provide an overview of computational achievements or breakthroughs in the field of SPAs but to address the technical considerations when employing these methods. For a broader understanding of computational advancements, we recommend referring to reviews by Tuttle et al. and Marrink et al. in peptide self-assembly,[26,27] or by Pavan and Bochicchio, more general in the field of supramolecular polymers.[40]

II. SPAs AT DIFFERENT RESOLUTIONS: FROM ELECTRONS TO MOLECULAR ASSEMBLY

In supramolecular chemistry, computational methods are used to (a) give insights into the structural properties and self-assembly process, and (b) guide the design of supramolecular structures. In particular, computational methods allow access to spatial-temporal details on the nanoscale, inaccessible by experimental methods.

The choice of a computational method depends on factors such as the problem being addressed, the system size, and available resources (Figure 1). They can be categorized into quantum mechanical (QM) and classical (molecular mechanical, MM) approaches. Both provide energy values for a given set of atomic positions, which can be optimized to obtain a structure at a (local) minimum and can access dynamics of the system by molecular dynamics (MD), i.e., propagating the atom positions using Newtonian equations. The QM methods, provide an accurate and robust description of a system while being easy to setup. However, their high computational cost usually prevents MD simulations and limits their application to small, static systems, often unrepresentative for SPAs. On the other hand, classical approaches are efficient and suitable for studying systems on a scale of SPAs systems and MD on a scale of milliseconds. Therefore, in practice classical approaches are almost always associated with MD and its nondynamic version, MM, is rarely used. However, they require a strong setup and parametrization effort. In what follows, we briefly introduce the QM and classical MD methods, illustrating their strengths and weaknesses in modeling SPAs.

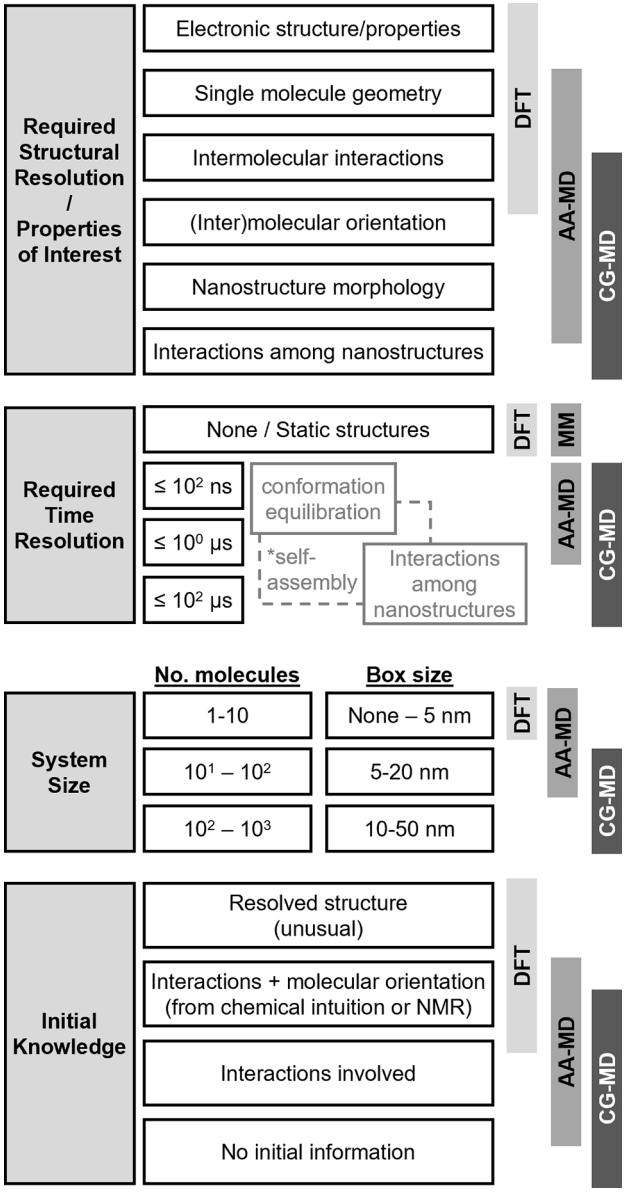


Figure 1. Scheme of the scope of different computational methods (DFT, MM, AA-MD, and CG-MD) as a function of required structural resolution or properties of interest, required time resolution, system size, and initial knowledge.

II.a. QM/DFT. At the heart of quantum chemistry lies the Schrödinger equation, typically in its time-independent form with the Born−Oppenheimer approximation, which, when solved, provides all possible observable information about the system. Unfortunately, exact solutions of the Schrödinger equation are only known for a few simple cases, such as the particle in a box, the harmonic oscillator, or the hydrogen atom. Consequently, QM methods employ approximation and numerical techniques to offer insights into physically interesting systems. These methods are broadly categorized into wave function-theory (WFT) based, density functional theory (DFT) approximations, and semiempirical methods (SQM).

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

WFT-based methods are based on solving Schrödinger equation in its original form and are renowned for their accuracy and robustness. However, they come with a high computational cost, often rendering them impractical for SPAs applications. An exemplary method within this category is the coupled-cluster approach, often regarded as the 'gold standard' and frequently used as a reference in numerous benchmark studies. On the other side, SQM comprise of QM methods with drastic approximations and parameters fitted to empirical data.[41] As a result, they can model medium systems ($\sim 10^3$ atoms), but with trade-offs in accuracy and robustness. Notable examples of SQM methods include PM6, PM7, DFTB3, and GFN2-xTB.[42−46]

DFT is probably the most popular QM method in the SPA field. It is based on solving the Schrödinger equation using electron density rather than wave function. They are more efficient than WFT and can be applied to study small-sized, such as small supramolecular assemblies ($\sim 10^2$ atoms). When used appropriately, they offer reliable and robust results, providing accurate energies and structural properties. While the theory of DFT is exact, the precise form of the key component of this theory, the exchange-correlation functional, is unknown, which leads to the formulation of at least 400 different approximations.[47] They can be categorized by Perdew's Jaccobi Ladder, in which at the bottom of accuracy lies the Hartree world/hell and each subsequent rung introduces an additional ingredient to the functional, finishing in "heaven" of chemical accuracy (Figure 2).[48] The functionals on the first three rungs focus on electron density and can be categorized as local density approximation (LDA), based only on electron density, generalized gradient approximations (GGA), which also includes electron density gradient, and meta-GGA, which includes higher-order derivatives of electron density. The fourth and fifth rungs, named hybrids and double-hybrid functionals, include nonlocal effects by combining DFT with WFT methods. In principle, the higher the rung on the ladder, the greater its accuracy and computational cost. Therefore, knowing where the chosen approximation is in the ladder is helpful as it gives rough guidance about the accuracy and computational efficiency of the method.[49] For a detailed introduction to DFT, we recommend reviews by Grimme et al. and Goerik et al.[50,51]

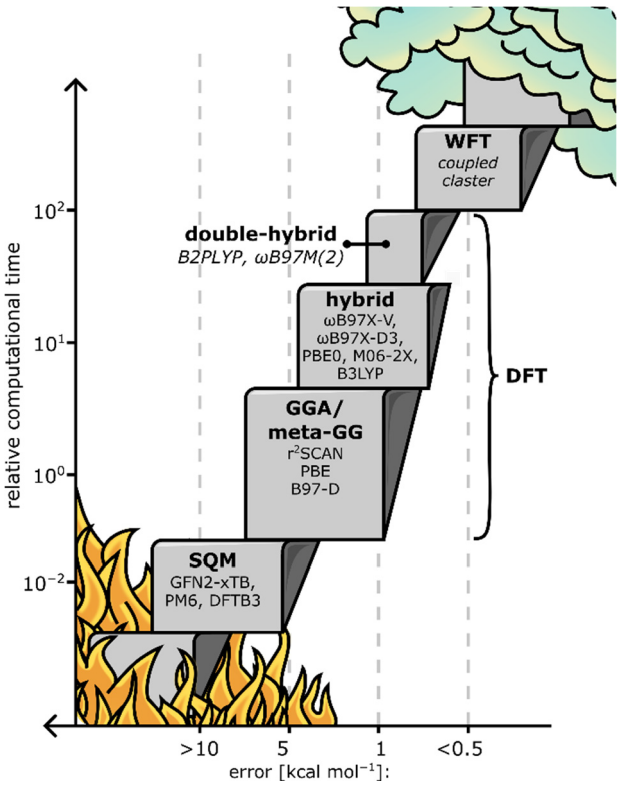


Figure 2. Typical accuracy and relative computational time for QM methods. Often, DFT methods are categorized (with increasing accuracy) as generalized gradient approximation (GGA), meta-GGA, hybrid, and double-hybrid. In general, methods with more components yield more accurate results but higher computational costs. Figure based on ref 50.

The two most common strategies for selecting the functional are based on (a) previous literature in the field/popularity and (b) benchmarks. While the former approach is commonly employed, caution is advised as the field continuously evolves, with new methods being developed and older ones potentially becoming outdated.[49] Alternatively, benchmarks offer an objective evaluation of the functional performance. The two most popular benchmarks are GMTKN55 and MGCDB84,[52,53] which contain 40% of data to test the noncovalent interaction of small organic molecules as well as the relative energies of amino acids and small peptide conformers. Both show that the family of ωB97 performed particularly well, with the best performing functionals ωB97X-V and ωB97M-V, followed by ωB97X-D3 or ωB97X-D. They emphasize the need of dispersion-corrected functions to describe weak nonbonded interactions accurately. It is worth noting that both benchmarks were published in 2017 and, since that time, new promising methods have been developed.[54]

In order to solve the Schrödinger equation, an accurate representation of electron density is essential. DFT calculations commonly employ basis sets, offering a practical and computationally efficient approach to approximate electron density. The basis sets are available in a variety of sizes, starting with the minimal basis sets (e.g., STO-3G), followed by split-valence (e.g., def2-SVP, cc-pVDZ, 6-31G*), triple-ζ (e.g., def2-TZVP,cc-pVTZ, 6-311G*), quadruple-ζ (e.g., def2-QZVP, cc-pVQZ), etc. The accuracy and computational cost of the calculations increase with the basis set size. Therefore, it is essential to balance the trade-off between precision and efficiency. A smaller basis set increases the probability of encountering issues associated with an incomplete basis set, known as basis set superposition error (BSSE), which leads to an artificial overestimation of noncovalent interactions. As a result, it is recommended to use at least a triple-ζ sized basis set or employ BSSE corrections.

The presence of a solvent in QM can be considered (a) implicitly by using one of the continuum solvation models (e.g., CPCM, SMD, or COSMO) or (b) explicitly by adding solvent molecules to the system.[50] The first approach is more common due to its simplicity and lower computational cost. Alternatively, the second might offer some advantages in cases of strong solvent−solute interactions. However, this approach often requires lengthy and laborious optimization due to a larger system and the need to explore the optimal position of solvent molecules. Therefore, the computational cost might outweigh the advantage of an explicit solvent and should be considered only if necessary.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

Regarding their application to SPAs, QM methods are highly suitable for single molecules and small assemblies as, for such a system, the calculations can be relatively fast and provide accurate energies and structural properties, such as intra- and intermolecular noncovalent interactions and molecular electrostatic potential.[55,56] For example, Xu and co-workers related the handedness of helical fiber in final assemblies from single peptide change by comparing conformers optimized by DFT initially obtained from MD simulations.[57] DFT has been utilized to investigate small assemblies of F-based peptides examining arrangement of the β-sheets,[58−62] and sequences protected with large aromatic moieties, capitalizing on the straightforward implementation of these methods, and obtaining critical intermolecular details of these systems.[63−65]

The limitation of the system size imposed by DFT can be addressed by multiscale modeling, particularly the QM/MM approach. In this approach, a small region of the system is treated using QM methodologies, while the remainder is handled with MM.[66,67] There are several protocols that vary in how the interface between QM and MM regions is implemented (using additive and subtractive schemes) and in the interaction mechanisms (mechanical, electrostatic, and polarizable embedding).[68] It is important to note that the bottleneck of these methods lies primarily in the QM regions; the efficiency of these methods largely depends on the size and level of theory applied to the QM region, as the computational cost associated with MM regions is relatively negligible. Despite their potential for SPA simulations, QM/MM approaches are infrequently used due to their complex setup (compared to standalone QM calculations), higher computational costs (compared to MM or MD), and the limited availability of compatible software. These and other multiscale examples are discussed in Section IV.b.

II.b. MD Approaches. SPAs have been extensively studied using MD approaches at all-atom (AA) and coarse-grained (CG) resolutions. MD simulations compute acceleration, velocity, and subsequent displacement of atoms by applying Newton's equation of motion, where forces arise from interactions with surrounding particles. At AA resolution, each particle describes individual atoms, while at the CG level a set of atoms are grouped into one particle.

The atomistic resolution of AA-MD allows the study of molecular and intermolecular features, such as molecular conformation and nonbonded interactions, which underlie the self-assembly process and the stability of different assemblies.[69−72] The lower resolution CG model greatly elevate computational performance by reducing the number of particles needed to model specific systems. Additionally, the use of longer time steps in CG, facilitated by the larger size of the particles employed, can speed up simulations by around 3 orders of magnitude.[73] Consequently, CG approaches enable the study of larger systems and longer time scales. In quantitative terms, AA-MD can typically simulate up to 10^1−10^2 short peptides and times of 10^1−10^2 ns, whereas CG-MD can easily handle 10^3 molecules for 10^0−10^1 μs in even shorter calculation times, often referred to as *wall-clock time* (Figure 1).

With the loss of resolution, CG models often fail to reproduce the enantiomeric nature of amino acids, a factor known to influence the self-assembling behavior and resulting morphology of the structures formed.[74−76] In contrast, AA resolution has proven its efficacy in modeling chirality, capturing the effects of substituting L-amino acids with D-amino acids in SPAs.[28,76,77]

AA-MD. With several decades of development, all-atom (AA) force fields, such as OPLS-AA,[78,79] CHARMM,[80] and AMBER,[81] support simulations of proteins/peptides/DNA/RNA and organic molecules, e.g., hydrocarbon tails and aromatic moieties. Specifically, OPLS-AA seamlessly supports peptides and hydrocarbons. CHARMM and AMBER released the CHARMM General Force Field (CGenFF)[82,83] and General AMBER force field (GAFF),[84] respectively, which were developed for organic molecules to be compatible with the original biomolecules parameters. In conclusion, all three force fields are suitable for SPA simulations. United atoms (UA) force-fields, which reduce the number of particles by omitting nonpolar hydrogens, have also been successful in modeling SPAs.[85,86]

Nevertheless, the accuracy of these force fields for SPAs needs to be quantitatively assessed. The parameters for (oligo)peptides, hydrocarbon tails, and aromatic groups, as well as their combinations, in addition to the water model, the force field of salt ions/counterions, and the simulation protocols must be quantitatively examined to ensure the validity of the method.

Peptides. The parameters for peptides were generally borrowed from those from folded proteins, with further improvements for unstructured peptides. The AA force fields of proteins have been extensively developed and validated with experimental approaches (X-ray structure, NMR, CD, etc.) and generally display minor differences in accuracy.[87,88] Unlike folded proteins, peptides are mostly unstructured and have much less experimental data to validate the force field parameters, raising questions about applying force fields traditionally designed for folded proteins directly to them. The progress of these atomistic force fields for peptide simulations up to 2019 was reviewed by Georgoulia and Gluykos.[89] In conclusion, all three force fields are qualified for SPA simulations, which generally provide comparable accuracy, with some observable differences that are barely experimentally validated.[90,91] Note that the CHARMM 36m force field has been better supported in recent studies for peptides, which play a crucial role in SPAs owing to their polarity nature.[80,92−94]

In most of the SPA simulations the charge of the amino-acids is chosen based on their pK_a. However, the highly condensed structures of SPAs involves additional challenges, such as the local pK_a, which might differ from that in the bulk solution, leading to variations in the ionization of the amino acids.[95] For instance, glutamic acids were found to be only partially charged at the surface of SPA nanofibers.[96] Addressing these ionization variations requires constant-pH simulations, implemented in GROMACS,[97,98] NAMD,[99] and AMBER.[100] However, given their high computational demand, large systems, such as SPA structures, are still challenging for constant-pH simulations.[97]

Hydrocarbon Tails. The hydrocarbon tails of PAs are also crucial in their assembly and the morphology formed. While their length can vary, longer chains of 12 to 16 carbons have demonstrated greater capacity for enhancing self-assembly.[85,86] As these tails are equal or shorter than those in phospholipids, their CHARMM and AMBER force field parameters can be utilized for SPAs. Note that for OPLS-AA, recent versions OPLS/L[101] or OPLS/2020[102] are required for the simulations of long-chain alkanes to reproduce the experimental data of gel-to-liquid phase transition temperature, diffusion coefficient, viscosity, gauche−trans ratio, and other features.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

The Peptide-Tail Connector. Although both peptides and hydrocarbon chains have been extensively developed and validated, the accuracy of their covalent link, specifically the bonded interactions between the peptide headgroup and the hydrocarbon tails, has not been appropriately addressed to the best of our knowledge and their parameters are barely described in the literature. Historically these parameters were generally obtained based on similar bonds, angles, or dihedrals in the existing force field database and were considered transferrable. Nowadays, automated tools have been developed to facilitate parametrization. For the CHARMM force field, the CGenFF web servers (https://charmm-gui.org) generate parameters for an input structure, including the joint section of SPAs with an accuracy score indicating the quality of each parameter. Similarly, AmberTools[103] and LigParGen[104] have been developed for AMBER and OPLS-AA, respectively.

Aromatic Moieties. APAs harness both the hydrophobic effect and π-stacking interactions to enhance self-assembly. However, the electronically complex nature of aromatic groups poses additional challenges for parametrization. While the servers introduced above can be applied, manual parametrization can be advantageous. For example, the parametrization of the Fmoc moiety with optimized partial charges, utilizing interactions with water and dimerization energies in different conformations to using DFT benchmarks, reproduced sequence-dependent self-assembly with the CHARMM force field.[105]

Water Models. Explicit solvent (water) models generally perform better than implicit solvents.[106,107] There exist around 100 water models, among which SPC, SPC/E, TIP3P are the most frequently employed 3-site models[108] as along with the 4-site models (TIP4P).[89] The 3-site models dominate in atomistic simulations owing to compromised accuracy and computational performance. Generally, force fields were originally developed using specific water models, and it is highly recommended to adhere to this pairing:[109] GROMOS/SPC,[110] OPLS-AA/TIP3P (OPLS-AA/TIP4P),[111] AMBER/TIP3P (AMBER/TIP4P/OPC),[89] and CHARMM/$TIP3P_{CHARMM}$.[94]

Salt Ions. In SPA simulations, most molecules are charges and monovalent ions (mainly Na^+ or Cl^-) are used to neutralize the system.[34,37,71,112] Additionally, 0.10−0.15 M NaCl is generally used to describe salt levels in biology or buffer solutions.[113] The effects of ions arrangement on the SPA surface due to the exposed charges in PA nanofibers,[112,114] or the stability changes induced by the addition of functional groups into the PAs[115] have been explored. Regarding differences among force fields, recent work by Qiao et al. reported that elevated consistency was indeed reached for the CHARMM force field.[116]

CG-MD. Coarse-grained (CG) models sacrifice some of the resolution of AA models to accelerate calculations by representing groups of atoms with a single particle or bead. Consequently, amino acids are represented by 1 to 6 beads (Figure 3). While the reduced resolution compromises the accuracy necessary to fully replicate the behavior and interactions of molecules, the overall structures formed are not significantly affected. In fact, reproducing certain features often necessitates large systems rather than focusing on fine-grained intricacies. Thus, CG models have demonstrated great performance in modeling the self-assembly of different types of SPAs, including short peptides, PAs, and APAs,[39,117,118] with great reproducibility in the formation of various sequence-dependent architectures, even between isomers.[118−120]

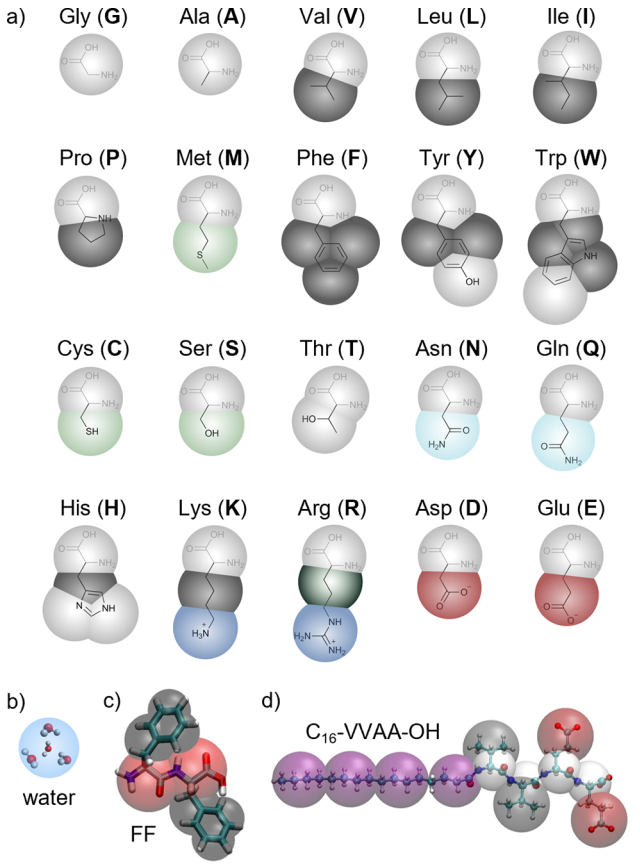


Figure 3. MARTINI mapping representation of (a) the 20 natural amino acids, (b) water, (c) the FF dipeptide, and (d) the C_{16}−VVAA−OH PA. Panel (c) reproduced from ref 121 and (d) from ref 120. Copyright 2024 and 2022, respectively, American Chemical Society.

Although different CG models exist, we focus on those that maintain enough resolution to model the amphiphilic nature of SPAs. The maximum approximation involves using beads to represent full amino acids. Olvera de la Cruz et al. employed such model to study the self-assembly of PAs as a balance between hydrophobic interactions and hydrogen bonding, predicting its effect on morphology (Figure 4).[122] A more recent study on FFF tripeptide supported that a higher resolution of the F side chains leads to better agreement with the experimental formation of solid-core nanospheres.[123]

The MARTINI force field stands out as the most widely adopted CG model, employing a 4−1 mapping approach where 4 non-hydrogen atoms are grouped per bead. This ratio may be adjusted to 3−1 or even 2−1 for aromatics, enhancing their shape replication (Figure 3a). Within MARTINI, water is explicitly represented, with each water bead corresponding to 4 water molecules (Figure 3b). Unlike atomistic models, beads in MARTINI feature only net charges and not partial charges, with the rest of the nonbonded interactions modeled by Lennard-Jones terms. Originally developed for lipid bilayers,[124,125] it was extended to proteins in 2008,[126] opening up opportunities for SPA studies. In 2011, the Tuttle group harnessed MARTINI's capacity to model short peptide self-assembly, screening 400 dipeptides for SPA sequences and design rules, later extending to 8000 tripeptides.[127,128] This

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

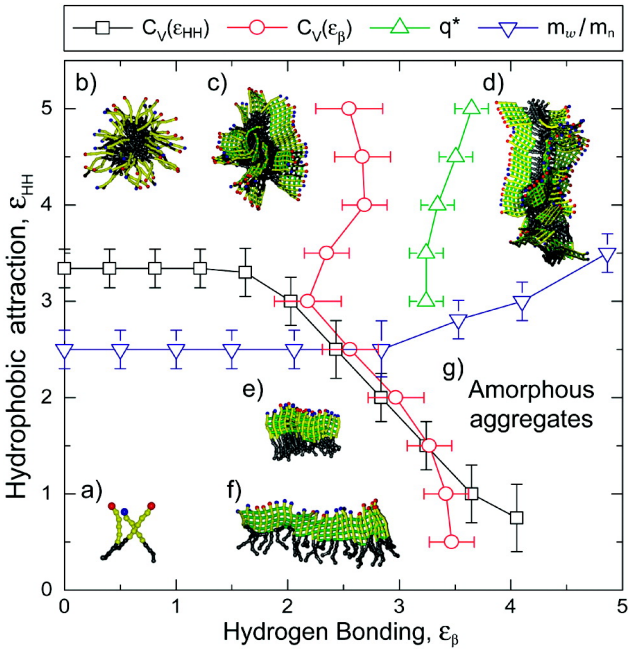


Figure 4. Schematic phase diagram presented by Olvera de la Cruz et al. as a function of the hydrogen bonding and hydrophobic terms in CG-MD simulations. The diagram includes regions of (a) free molecules, (b) spherical micelles, (c) micelles with β-sheets on the outside forming the corona, (d) long cylindrical fibers, (e) β-sheets stacks, (f) single β-sheets, and (g) amorphous aggregates. Figure reproduced from ref 122. Copyright 2008 American Chemical Society.

approach led to the discovery of new self-assembling sequences, the development of design rules, and, lately, the selection of candidates to act as emulsifiers.[129] The same procedure was successfully adapted to predict the self-assembly of dipeptide amphiphiles and liquid−liquid phase separation (LLPS) in dipeptides.[130,131] In 2012, Wei et al. studied the self-assembly mechanism of FF.[118] The MARTINI force field has demonstrated significant success in modeling self-assembly[119,120,127−132] and coassembly.[133−138] Specifically, this method has enabled the prediction of conformational changes and even full phase diagrams of coassemblies,[133,134] as well as the understanding of the inhibition of self-assembly in mixtures of components and the unraveling of the exact location of monomers depending on the sequence.[137,138]

The MARTINI team has been dedicated to refining the force field for improved representation of protein dynamics. However, these improvements have ironically impacted its efficacy in peptide self-assembly. In 2016, the Wei group recognized the need to adhere to version 2.1 for short peptides,[126] the initial version designed for proteins, as the subsequent version 2.2[139] could not replicate the self-assembly of FF or FFF.[133] Recent studies also identified certain limitations for the most recent MARTINI 3 akin to those observed in version 2.2, which exhibited heightened solubility of peptides.[140] Nevertheless, a recent study suggested that these limitations diminish with peptide length, making it potentially negligible above 4−5 amino acids. Moreover, we also proposed corrections that address these issues, enabling accurate modeling of the self-assembly of various dipeptides and tripeptides.[121] Despite the initial challenges, the improvements introduced in MARTINI 3 are expected to contribute positively to the field of SPAs.

CG models are versatile in replicating various environmental conditions beyond aqueous solutions. Nguyen and colleagues, for instance, utilized an implicit solvent model to investigate the impact of hydrophobicity in a PA's self-assembly.[141] The MARTINI force field offers the flexibility to incorporate other explicit solvents to explore the behavior of tripeptides in water/organic biphasic systems, providing insights into sequence-dependent surfactant behavior.[129] The influence of ions studied using the more detailed polarizable version of the solvent,[142] demonstrated the inhibition of FF self-assembly in the presence of NaCl.[143] Moreover, pH can be introduced combining MD with Monte Carlo calculations, as described by Tuttle's lab,[144] or by using the MARTINI's constant pH model, developed by Marrink's lab, which allows proton transfer between the beads.[145] Other studies have adopted diverse strategies, such as modeling pH-dependent self-assembly by running simulations across the full range of charge states for pH sensitive side chains.[20,120]

III. SYSTEM SIZE, CONCENTRATION, TIME SCALE, AND INITIAL STRUCTURES

As well as the level of accuracy and details that can be achieved, the method choice also influences the size of the system that can be simulated within the constraints of current computational power. DFT, which does not account for time, is usually limited to no more than 10 molecules. Generally, atomistic simulations are typically constrained to 10^2 ns in boxes below 20 nm. CG can extend up to 10^1 μs, and covers tens of nanometers. As such, CG-MD can model the self-assembly process from scratch, requiring minimal or no initial information from experiments,[127,128,130,131] while AA-MD often necessitates an initial input structure to provide atomistic detail on the structures.[36,146] Additionally, recent studies have demonstrated the influence of concentration in modeling the self-assembly behavior of SPAs.[121,147]

III.a. System Size. The time required for any computational calculation is proportional to the number of particles involved. DFT demands considerable resources, limiting its application to providing insights to the interactions and preferential molecular orientation of single peptides and small assemblies, usually with no more than 10 molecules, allowing for the comparison of different arrangements.[19,59,60,148] In contrast, AA-MD simulations can operate on the scale of ~10^2 SPA molecules and CG-MD can handle systems up to 10 times larger in one-tenth the time compared to AA.

In terms of self-assembly, AA-MD simulations typically focus on the initial stages of aggregation to gain an understanding of the forces driving the process.[69−71] Although there are instances where the entire formation of nanoscale assemblies is reproduced,[76,105,149] it is computationally expensive due to the need of a large number of SPAs molecules and accompanying environment (e.g., solvent and ions) and the long time-scale of the process.

Instead, using CG models, simulations ranging from 10^2 ns to μs can easily contain hundreds of solute molecules, enabling reliable reproduction of the entire self-assembly process.[38,39,119,133] Furthermore, given the system size that CG simulations can achieve, they allow for the reproduction of nanoscale morphologies at a level that can be correlated with experimental microscopy techniques, creating a synergy between both approaches that has enhanced the understanding of SPAs.[150]

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

Despite the preference for larger systems with more molecules in modeling self-assembly and SPAs, certain problems can prioritize resolution over size. For example, small AA systems can be employed to study intricate features at the atomistic level, even comprising single or a few-molecules systems within the scale affordable for DFT calculations. However, unlike the latter, AA-MD reproduces experimental dynamic conditions and entropic contributions. There are instances where even single-molecule AA-MD simulations are employed to elucidate preferred conformations influencing self-assembly.[151,152]

The system size can also influence the ability to reproduce the formation of certain assemblies in CG simulations. For instance, Tuttle et al. increased the box side from 12.5 to 24.5 nm and the number of dipeptides from 300 to 1600 to successfully replicate the formation of FF tubes.[127] A more recent study also supported the formation of tube structures using a larger system (1600 FF molecules with simulation box length of 24.5 nm) than a small system (600 molecules with a box length of 17 nm), at the same concentration.[121]

III.b. Time Scale. First, QM methods and MM are usually utilized to give insights into the static system and therefore do not account for time. MM is often used only to minimize assemblies obtained from idealized models, as it allows for larger systems than QM methods.[21,153,154]

In contrast, MD methods account for thermal motion and incorporate a time dependence. There exist notable distinctions in terms of time affordability between CG and AA approaches. Except for a few examples,[28,105,149] AA simulations may fall short in modeling the entire self-assembly process, which typically extends beyond reachable time scale of microsecond. As such, AA models are frequently employed to assess the stability of proposed models. These models can range from small stacks, necessitating only brief simulations to evaluate their stability, to larger structures that require extended simulations.[155−159] Small stacks may suffer from a lack of stability due to their limited size and low number of molecules, given that self-assembly is a collaborative process. Consequently, results must be interpreted with caution and always in a relative context.[30,157−159] Larger models necessitate more extended periods to reach equilibrium, which is contingent on the initial structure's proximity to equilibrium, ranging from less than 40 ns,[146,155] to 100 ns.[36,72,156,160] The implications of using different starting structures, along with some relevant examples, are discussed in Section III.d.

The acceleration in CG simulations is not solely attributed to the simplified topologies, making each step computationally cheaper, but also to the longer times steps allowed by the larger size of the particles. While AA-MD is constrained to time steps of 1−2 fs, a CG model can afford time steps of 20−40 fs. Moreover, in CG models, due to the smoother nature of the potential energy surface, molecules diffuse unrealistically faster.[125] For example, in the MARTINI model, an additional speed-up factor of 4 is often assumed, resulting in up to 160 times fewer steps to simulate the same time than AA-MD. With these steps being faster to calculate, CG simulations, on a ∼ 24 cores/node, can run 1 μs simulations with boxes of 10−20 nm per side within a day. On the other hand, simulating a 10 nm box for 100 ns in AA may take weeks. In general, the most significant advantage of the time scale in CG models is the capability to reproduce the entire self-assembly process, enabling the assessment of self-assembly tendencies in simulation ranging from 1 to 5 μs.[33,132,149,161] This also enables the study of different stages of the process along the simulation, identifying early aggregation stages and their coalescence to form nanostructures in pathways of up to 14 μs.[39,118,162] Finally, CG simulations have also even been employed to assess the dynamic properties of resulting structures across several microseconds and interactions between structures in simulations of 10 μs.[7,20,120]

III.c. Concentration. While the concentration dependence of self-assembly in solution is widely acknowledged, this parameter is often overlooked in computational works. As in the case of lipids, SPAs present critical concentrations at which structural transitions occur, namely critical aggregation concentration (CAC), critical micelle concentration (CMC), or critical fiber concentration (CFC).[163−165] However, due to the computational cost, researchers typically minimize the simulation box, resulting in concentrations 1 order of magnitude above what is commonly employed experimentally for SPAs, often surpassing any critical concentration. For instance, a concentration of 10 mM necessitates boxes with 25.5 nm side to accommodate at least 100 peptides, which only requires 12 nm box for a concentration of 100 mM. However, although this 10-fold increase in concentration is justified as a means of expediting the self-assembly process at an affordable computational cost,[39,119,128] its importance becomes more apparent when assessing the effect of concentration in self-assembling simulations. The concentration-dependent polymorphism of the tripeptide KFG using the MARTINI force field, exhibited a correlation at computational concentrations 15 to 50 times higher than the experimental ones (Figure 5).[147] This dependence was also observed by Guo et al. with

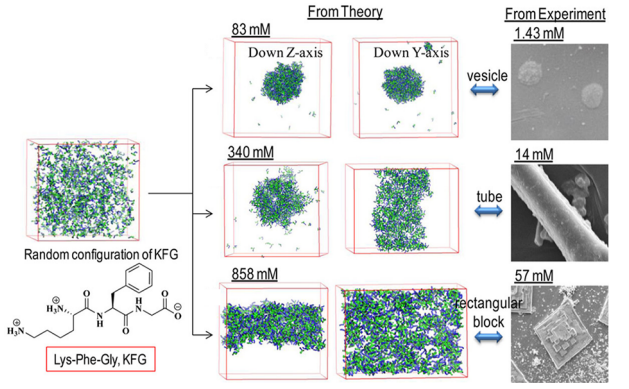


Figure 5. CG-MD self-assembly simulations of the tripeptide KFG at 83, 340, and 858 mM and the correlation of their formed shapes with TEM images at concentrations of 1.43, 14, and 57 mM, respectively. Figure reproduced from ref 147, Copyright 2017 American Chemical Society.

the FF dipeptide, which formed vesicles instead of tubes at low concentrations.[118] A more extended study demonstrated that simulations require concentrations above 100 mM to replicate the formation of tubes experimentally observed at 10 mM.[2,121] These studies suggest that the 10-fold increment in concentration does not only accelerate the self-assembly, but is also required to reproduce the thermodynamic equilibrium of self-assembly. Hence, while systematic investigations on CACs or other critical concentrations involving morphological transformations are needed to ascertain the exact concentration scaling factor, we note that the 10-fold has emerged as a practical approximation.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

The use of significantly higher concentrations is notably emphasized in AA simulations, driven by the substantial computational costs that imposes limits on box size. Concentrations exceeding 1660 mM (216 molecules in 6 nm side boxes) were employed by Garcia et al.[28,76] on the self-assembly of tripeptides, experimentally investigated within the range of 10 to 20 mM. Remarkably, despite the elevated concentrations, Garcia et al. successfully delve into this process at the atomistic level, extracting intricate details on the interactions and the impact of backbone conformation comparing the assemblies when introducing one D-amino in tripeptides with simulations lasting 300 ns. Tuttle et al. explored sequence-dependent morphological differences in two protected amino acids at a lower concentration, ∼350 mM (120 molecules in 8 nm side boxes), although still 35 times more concentrated than the corresponding experimental system.[105] While employing such heightened concentrations has yielded positive results in AA-MD simulations, there is no established correlation between experimental and computational values as in CG-MD. Moreover, in AA simulations, it is common practice to employ a specific initial arrangement that mitigates potential concentration effects, allowing the system to be treated as representative of the local high concentration where the assemblies are localized.

III.d. Initial Structures in AA. AA-MD simulations of the assembly process of SPAs are accessible only for selected systems. Therefore, an initial structure is often built using a combination of experimental knowledge and chemical intuition to provide a starting point. Subsequent MD simulations will evolve into the energetically most favorable structure within the time scale affordable.

Schatz and co-workers conducted numerous early studies on SPA nanofibers using both AA and CG simulations. They proposed the optimal initial structures for atomistic simulations on SPA nanofibers,[146] a framework widely employed in subsequent studies (Figure 6a).[155,166] In this initial structure, the nanofiber consisted of layers (cross-section) of PA chains, with each layer featuring 9 PA chains radially and evenly distributed (40° between neighboring PA chains). The neighboring layers were rotated by half of the PA−PA angle (20°), and the distance between adjacent layers was set at 5 Å, as the distance between β-sheet strands (Figure 6). Tekin et al. explored other initial structures for PA nanofibers, varying the numbers of PAs in each layer from 7 to 12 and the number of layers in the nanofiber from 9 to 21, in relatively short simulations, 30−50 ns.[167,168] They concluded that 12 PAs/layer with 19 layers displayed enhanced stability in the cylindrical nanofiber composed of C_{12}−VVAGERGD.[167] A subsequent study also assessed how the coassembly of functionalized PAs could affect the preferential composition of these layers.[169] These types of stacks provided information on atomistically resolved structures, interactions responsible for the stability of the assembly, and the distribution of peptide secondary structures.[146,155]

Further studies have employed AA to model the homochirality of amino acids to evaluate the twist in PA-built nanofilaments, gaining insights into curvature and dependence on the amino acid sequence.[156,170] The construction of these initial structures of PAs, however, is relatively straightforward due to the design of these SPAs. PAs are designed with an aliphatic tail, usually at their N-terminus, driving aggregation, and the peptide region that shows a gradient of hydrophobicity until the charged C-terminus (Figure 6b). This design restricts the arrangement of the molecules perpendicular to the fiber axis. This arrangement was further confirmed by MARTINI CG simulations that showed the formation of a cylindrical nanofiber after 16 μs.[39]

In addition to PAs, there exist other SPA designs that facilitate the task of building initial structures, though with relatively limited experimental information. This is exemplified by the self-replicating macrocycles developed by the Otto group.[171] These are APAs that form cycles through reversible disulfide bonds, creating an aromatic core with the peptides providing a gradient of hydrophobicity that stabilizes their interaction with water (Figure 7).[37] The task of building models is facilitated by the evident molecular orientation and

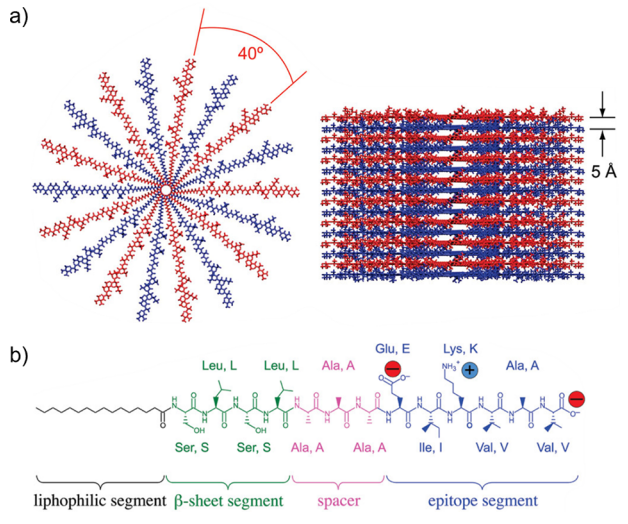


Figure 6. (a) Proposed initial structure for atomistic simulations by Lee et al.[146] showing top (left) and side (right) view of the nanofiber structure with two separate layers (interlayer distance of 5 Å), in red and blue, rotated by 20°. (b) PA chemical structure. Figure reproduced from ref 146. Copyright 2011 American Chemical Society.

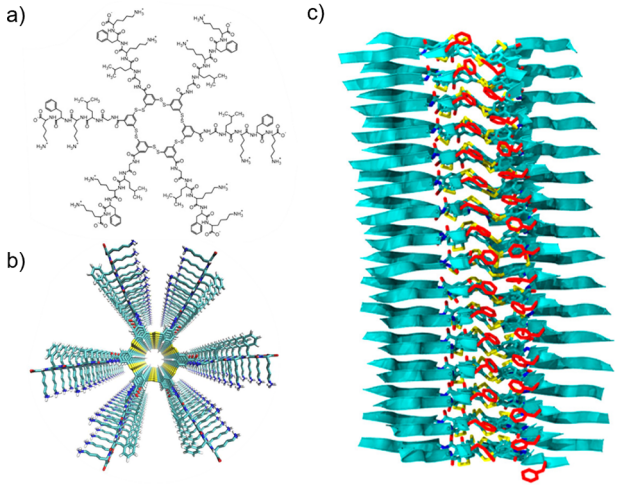


Figure 7. (a) Chemical structure, (b) top view of the initial configuration, and (c) side view of the AA-MD equilibrated structure of the self-replicating macrocycles developed by the Otto group. Figure adapted from ref 37. Copyright 2017 American Chemical Society.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

the limited number of possibilities to stack molecules, enabling the combination of peptide−peptide hydrogen bonding and π-stacking. The equilibrated structure was employed to gain a deep understanding of the intermolecular order. Other peptides with a surfactant structure (Figure 8), such as A_NK

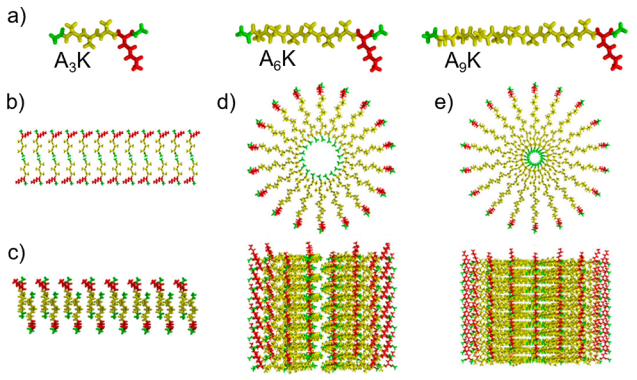


Figure 8. A_NK (N = 3, 6, or 9) (a) structure and some proposed initial configurations for the (b, c) A_3K bilayer, (d) A_6K tube, and (e) A_9K fiber, with the latter two showing top (top) and side (bottom) views. Figure reproduced from ref 34. Copyright 2014 American Chemical Society.

(N = 3, 6, or 9), also simplify the formation of initial configurations, with the most hydrophobic side always oriented toward the core of the structures. However, even in this case, different conformations were tested to reproduce the stability of membranes, nanotubes, and nanorods at N equals 3, 6, and 9, respectively.[34] Amyloid-like SPAs mimic the design of toxic folds formed by these proteins, alternating hydrophobic and hydrophilic residues, limiting the potential molecular conformations to replicate the experimentally observed fibers. This allows for the reliable construction of fiber models for AA simulations, enabling the study of preferential conformations, relative stabilities of the structures, and the arrangements of coassembled molecules.[72,172−175]

In some cases, there is only limited information on certain interactions that do not provide enough constraints to determine the specific arrangement of monomers in the nanostructures. This is often the case for short peptides or short aromatic peptide amphiphiles (APAs), where the amphiphilicity of the molecules is not as well-defined as for PAs to favor their core−shell structures (Figure 9a, b). Given the limited information, scientists have often analyzed relative stabilities of rather small stacks, composed of a limited number of molecules (2 to 21). The relatively high dynamics of such small structures facilitate their equilibration within 25 ns.[152,157,158] These systems can be simple enough to run DFT calculations to compare binding energies and calculate spectroscopic magnitudes.[19,57−60,148] Still, these stacks have proven to be useful even for assessing the effects of sequence mutations and environmental factors, such as pH and solvent variation.[30,32,159]

Caution must be exercised to minimize bias when building initial configurations. Insufficient conformational space sampling can introduce errors. Furthermore, the limited size of some of these stacks also introduces a bias toward parallel conformations, as antiparallel ones do not exhibit appropriate distances for π-stacking of moieties placed in one of the termini (Figure 9c).[19,157,158] However, simulating fiber models constructed by expanding similarly simple stacks concurred on

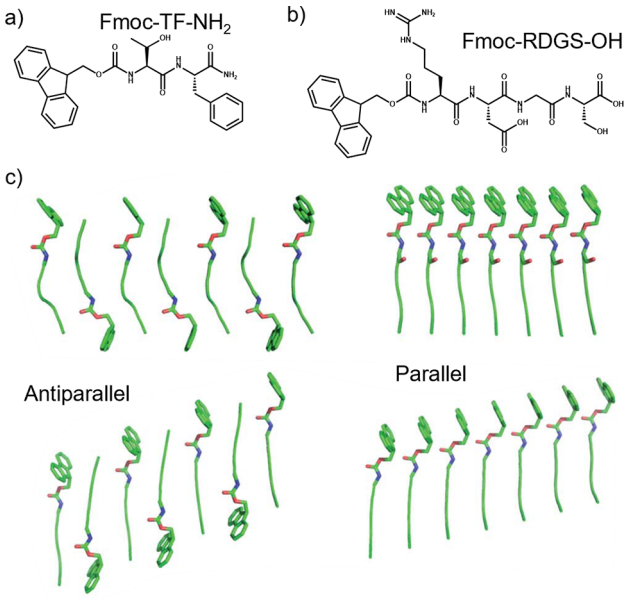


Figure 9. Chemical structures of two APAs (a) Fmoc-TF-NH_2[36] and (b) Fmoc-RDGS−OH, and (c) the four stacks arrangements studied by Aleman et al.[158] Panel (c) reproduced with permission from ref 158. Copyright 2013 The Royal Society of Chemistry.

the parallel arrangement of APAs, which was inconsistent with experimental findings.[29,160,176]

An alternative approach is to combine chemical intuition with an interpretation of experimental results. Such approach was used to build models of Fmoc-protected dipeptides nanostructures.[21,153] The correlation between experiments and simulations in an initially proposed model, allowed the refinement of the molecular arrangement to better fit the experimental observations.[36] However, this approach required long times of 150 ns to converge the simulations with the wrong initial model. Nevertheless, in certain cases, NMR, in combination with other experimental information, can provide constraints enough to build atomistic models.[14] Additionally, although single-crystal molecular arrangements are usually not representative of the supramolecular assembly in solution,[17,18] stabilities of stacks based on the crystal structure, with different stereoisomers, were compared with MM calculations by Kralj et al.[154] However, except in the latter case, where they extrapolated the structures directly from crystallographic data, these studies require a deep understanding of both experimental and computational techniques and are highly time-consuming.

Therefore, initial structures in AA give us the ability to study the stability, chirality, and supramolecular interactions, among other features, of self-assembled structures by constructing initial models based on observations from experimental data and human reasoning. However, due to the limited experimental information available, caution is advised, and analyzing various initial structures may provide insights into the proper orientation of the molecules within the assemblies. The validity and accuracy of the results depend on the number and variety of structures and arrangements included in each study. With only two different initial structures, we can assess the preferential molecular orientation—e.g., parallel versus antiparallel. However, adding various possible structures for each arrangement will further validate the conclusions and potentially unlock more details about the specific interactions

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

involved.[158,167] Even in cases where molecular arrangements are constrained by experimental observations and chemical intuition, proposing different possibilities and fine-tuning among them enhances understanding at the molecular level. These adjustments should focus on specific interactions, such as hydrogen bonding among specific residues, to fully exploit AA-MD to its highest potential.[34,36,37] In certain instances, when experiments provide sufficient constraints, when great detail is not required, or when the study is focused in evaluating the effect of sequence or environmental conditions, assessing the stability of just one structure may be sufficient.[30,32,156,159,170] Nevertheless, we must recognize that the results are only relevant within the context of the proposed structure and should not be considered as proof of preferential arrangement,[160] except in cases where the final structure significantly deviates from the initial one. In such cases, this change can serve as evidence of preferential intermolecular binding arrangements. However, it is still advisible to rerun the system with the newly suggested conformation as a final check of stability to validate the atomistic level of detail.[36]

III.e. Initial Structures in CG Simulations. Although CG-MD simulations can reproduce the entire self-assembly process of SPAs, there are cases where it is pertinent to employ CG-MD simulations with specific initial structures, such as coassemblies, cointeraction with biomolecular systems, and the study of certain mechanisms.

Coassembly is frequently modeled considering the different monomers in solution (Figure 10).[133,134,137] While it is often a

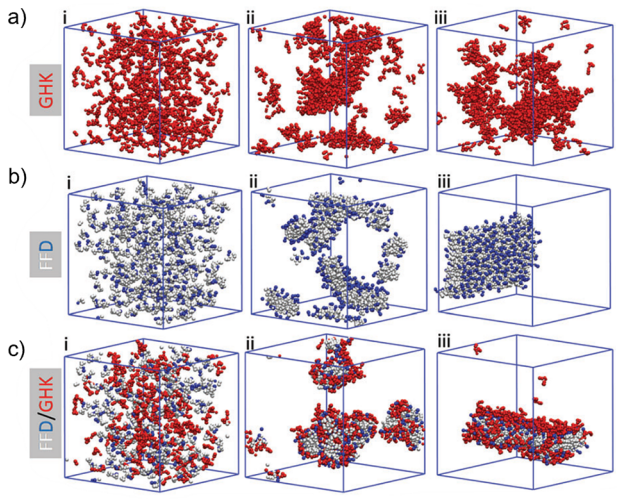


Figure 10. CG-MD simulation snapshots of the self-assembly of (a) GHK and (b) FFD, and (c) the coassembly between them at (i) 0 μs (initial simulation setup), (ii) ≈3 μs, and (iii) ≈9.6 μs. Figure reproduced with permission from ref 134. Copyright 2017 The Royal Society of Chemistry.

valid simplified approach, it is only an accurate representation of the experimental system if, for instance, monomers are mixed in a solvent where they are not preassembled/aggregated, such as hexafluoroisopropanol (HFIP).[138] Otherwise, simulating the coassembly of preassembled components provides a more accurate representation.[7,177] This has been employed to incorporate functionalized monomers into filaments, built or assembled in a previous step.[7,177,178] The results provided insights into the percentage of incorporation and the presentation of the bioactive sequences, demonstrating excellent correlation with biological *in vitro* and *in vivo* experiments.

Initial structures have been instrumental in directing the self-organization of PAs from diverse starting configurations.[120] This study compares the structures formed by molecules in solution with those derived from an initial bilayer configuration, inspired by a model proposed years earlier for the experimental system.[12] The simulations revealed the formation of filaments through both configurations, while the bilayer's initial structure retained a distinctly 2D nature due to robust interactions among filaments, resulting in the formation of superstructures (Figure 11a, b). Although the trend toward

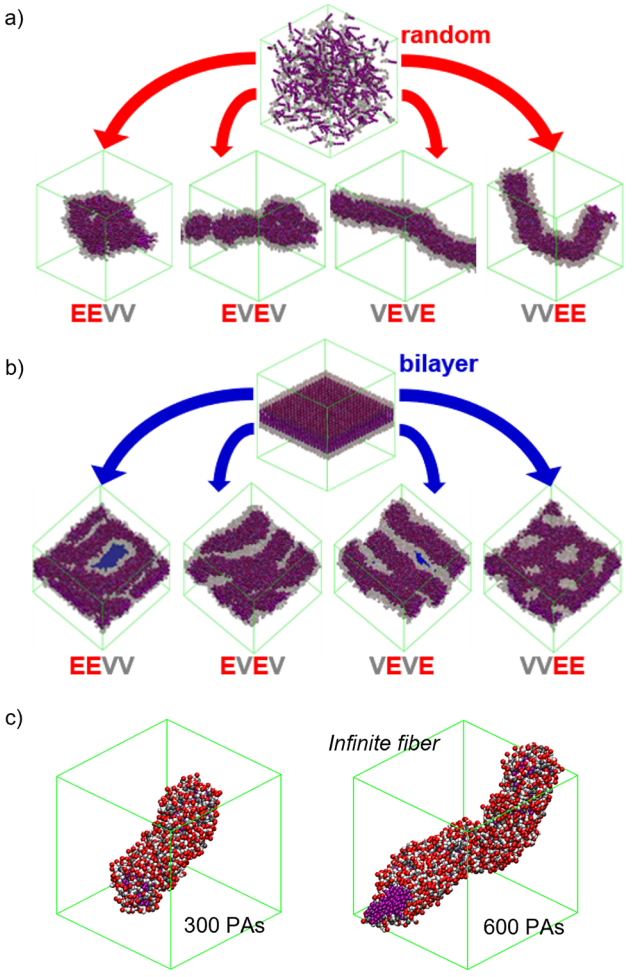


Figure 11. Structures obtained from CG-MD simulations with four PA isomers initially disposed (a) randomly and (b) in a bilayer configuration. (c) Comparison of the fibers obtained from a random initial configuration with 300 and 600 PA molecules, with the latter giving an *infinite* fiber across opposite edges of the box (in green). Figure reproduced from ref 120. Copyright 2022 American Chemical Society.

superstructures was also evident in simulations from scratch, it was less pronounced. A comparison with interfilamentous interactions among pre-equilibrated single filaments suggested that the hierarchical superstructure formation involved simultaneous, rather than sequential, processes. Applying the same protocol to study interactions between filaments in this and a previous study demonstrated a strong correlation between interfilamentous binding energies and the viscosity of the materials, as influenced by the sequence and charge,

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

experimentally tuned through pH variations.[20] While interactions of SPAs with other biomolecules can also be investigated with atomistic resolution,[28] the addition of more components intensifies the challenge of system size affordability. However, in specific cases, such studies can provide crucial mechanistic insights, such as in understanding the superstructure nature of twisted ribbons.[36]

The examples of interaction studies among nanostructures and with other biomolecular systems are poised to see a significant increase in the coming years. These examples underscore that studying the interaction between a structure A and a structure B, being A a SPA and B either SPA or another biomolecular system, e.g., lipid bilayer, protein, or protein assembly, necessitates the use of pre-equilibrated molecules. After independently equilibrating structures A and B, they are placed in the same simulation box, ideally positioned to facilitate interaction among them to optimize simulation time.[20] This aligns with existing protocols for various interactions, such as peptide and other molecules interacting with bilayers.[179−182] However, challenges may arise when the size of the SPA nanostructure needs to match that of the lipid bilayer, as demonstrated by Schroer et al. in the case of a protein filament.[180] Although each component of these multicomponent systems must be pre-equilibrated, SPAs can still be formed in a preceding step when utilizing CG resolution. As the number of examples increases, there is a growing opportunity to establish robust protocols and gain a deeper understanding of the potential challenges that may arise.

III.f. Practical Aspects of Building Initial Structures. The use of starting structures involves additional technical challenges. The approach to building these initial structures can be systematic, employing, for instance, Packmol,[112,183] to position molecules with specified orientations (angles) and at appropriate intermolecular distances.[146] For example, hydrogen bonding is favored at a 5 Å interpeptide distance (center to center). On the other hand, more complex architectures may require manual construction of the models,[36] utilizing software such as Avogadro software or Visual Molecular Dynamics (VMD).[184,185] In many cases, initial fiber structures need to be strategically placed to exploit periodic boundaries across the simulation box (Figure 6a) to model an *infinite* fiber, fixing the z-side to its length. This makes that additional box variations to adjust concentrations require modification of *x*- and *y*-axis only.[177] Alternatively, CG self-assembly simulations to result in *infinite* fibers, at least as long as the box side, require adjustments of the simulation box and number of molecules (Figure 11c). Usually, we place the fiber in the *z*-axis because certain simulation packages, e.g., GROMACS, facilitate decoupling this dimension from the *xy*-plane. This approach, known as semi-isotropic pressure coupling, allows the system to vary the *z*-axis independently of x and *y*-axes, facilitating the equilibration of a 1D fiber. However, exploiting periodic boundaries induce other constrains, specifically in the twisting of the fibers, making simulations seeking understanding on this matter to set up initial structures smaller than box dimensions, with enough separation to avoid unrealistic elongation.[156,170]

IV. FURTHER TIPS TO OBTAIN VALUABLE INFORMATION FROM SPA CALCULATIONS

IV.a. Sampling Methods. The constraints on AA-MD simulations in replicating the self-assembly process are primarily bound by the time scale, which is ascribed to the kinetic barriers of complex structures. In this regard, advanced sampling methods have been developed. Replica exchange molecular dynamics (REMD) enhances sampling by running sequences of simulations in various temperatures, periodically exchanging the explored configurations. The conformational search, ranging from monomers to hexamers, has enabled the formation of well-ordered stacks and the assessment of variations in molecular conformation with the degree of arrangement, yielding results more reliable than brute-force AA-MD self-assembly simulations.[152,186,187] The Xu group utilized the resulting preferred conformations to construct reliable starting structures for assessing fiber stability in different solvents.[35] Employing REMD to propose more accurate starting structures for nonbiased AA-MD simulations holds significant promise as an alternative for SPAs where intuitive construction is challenging, such as APAs.[160] Umbrella sampling is another method to overcome kinetic barriers and enhance the sampling toward a reaction coordinate. The Schatz group examined the disassembly process of a PA as a coordinate to calculate the free energy of the process.[188] Despite these advantages, the adoption of these enhanced sampling techniques is not widespread, partly due to their higher computational cost, difficulties of finding appropriate reaction coordinates to describe self-assembly, and the fact that their advantages can be offset by chemical intuition and experimental information.

IV.b. Multiresolution Methods. Given that different computational approaches present their own strengths and weaknesses, methods exist to combine them synergistically to leverage information provided by each. It is common to use DFT to enhance accuracy in studying interactions using structures formed, or at least equilibrated, by AA-MD.[176,189−191] For example, binding energies calculated with DFT are employed as benchmarks to optimize the AA force field.[105] In certain cases, this offers the advantage of spectroscopic validation using, for example, FT-IR experimental results as a reference for the calculated vibrational spectra with DFT from MD-formed structures.[191] Furthermore, DFT is sometimes employed to link the AA-MD structure to electronic structure-dependent features. For example, it can be used to identify effects on the reaction mechanism depending on MD conformation in catalytically active SPAs,[192] or to determine the charge distribution in donor−acceptor interactions of SPAs with large aromatics like porphyrins.[190]

Hybrid AA-CG MD simulations can provide detailed molecular information from the former while benefiting from the scalability inherent to the latter. This process can be executed sequentially, beginning with the assembled state formed through CG-MD simulations and backmapping it into AA resolution. Subsequently, the CG structure undergoes re-equilibration with AA resolution in a short simulation. Therefore, the output of the CG simulation serves as the starting structure, mitigating human bias, while the information acquired from the simulations entails the AA resolution.[193−195] Sankaranarayanan and co-workers employed a more intricate approach, commencing with AA to assess the initial stages of aggregation and its interactions, particularly with water within the SPA. The system was then converted to CG to facilitate scaling up and enable the aggregates to coalescence into fibers, before being backmapped to AA to elucidate the intermolecular interactions.[196]

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

The combination of methods can be integrated into the same calculation by delineating regions with varying resolutions, thereby enhancing detail in specific groups or molecules of interest. For instance, employing QM/MD calculations to treat water with AA force field and the solute molecules as QM allows for the computation of optical Raman activities with an MD trajectory.[197] Similarly, AA-CG approaches stand for a comparable strategy, representing solute molecules with AA resolution and CG for the surrounding water molecules, thereby significantly reducing the computation time.[198] Nonetheless, these methodologies often pose additional challenges. The energy scales are intrinsic to each force field, and their absolute values cannot be directly compared across them. Hence, when integrating competing interactions calculated with various force fields—like simultaneously using CG and AA for different interacting pairs—proper scaling factors need to be benchmarked to prevent bias toward specific interaction types. For instance, if the CG force field representing interactions with water shows higher binding energies compared to the AA force field representing solute interactions, this inconsistency might lead to artificially high solubilities.

IV.c. Analysis Tools. The most captivating result of MD is to examine the morphology of supramolecular assemblies, which can be visualized with VMD,[185] Chimera,[199] Pymol,[200] or Morphoscanner,[201] providing qualitative information. In self-assembling simulations, they serve as proof of self-assembly and their morphology can be compared with experimental techniques such as TEM, SEM, AFM, and SAXS.[119,120,177,201] Their morphological structures can be quantitatively studied through geometry maps (Figure 12), obtaining parameters that

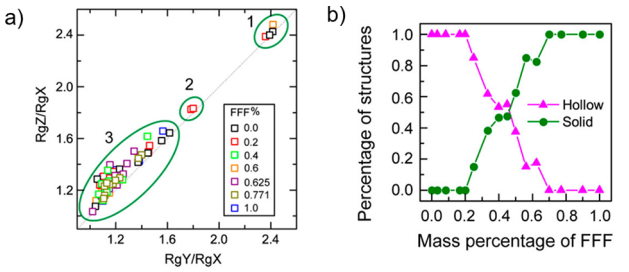


Figure 12. (a) Geometry map of the nanostructures formed at different FFF mass percentages. The results are grouped in 3 clusters according to the morphology: (1) nanotubes or nanorods (RgZ ≈ RgY ≫ RgX); (2) ellipsoids (RgZ ≈ RgY > RgX); and (3) sphere-like nanovesicles or nanospheres on the diagonal line, and toroids or ellipsoids in upper left of the diagonal. (b) Percentage of hollow or solid structure as a function of the FFF mass percentage. Figure reproduced from ref 133. Copyright 2016 American Chemical Society.

can be directly validated with experiments.[133] The tools are derived from protein data, making the obtention of specific parameters for multimolecular systems highly challenging. For example, using moment of inertia for quantitative shape determination requires pretreatment of self-assembled structures.[128] Therefore, using these methods for SPAs poses some limitations that must be considered, including their lower order and subsequent heterogeneity.

Numerous features could be analyzed from MD simulations. First, the evolution of structural features such as the root-mean-square deviation (RMSD) or solvent accessible surface are (SASA) are advisible to assess equilibration in SPA simulations. The quantification for the self-assembly tendency was first described through the Aggregation Propensity (AP) parameter employing CG-MD simulations,[127] and was also employed to study the distribution of the different molecules within coassemblies.[135] A recent study has demonstrated an inverse correlation between simulated AP values and experimental CAC.[202] However, AP quantified aggregation rather than self-assembly, with higher values for insoluble sequences, not directly implying good self-assembling properties. Therefore, the measure was improved by introducing a bias toward the hydrophilic residues to compensate for the aggregation excess on hydrophobic peptides, called hydrophilicity-corrected score (AP_H).[128] However, these parameters are not suitable for differentiating between PAs with a low aggregation capability or with no aggregation propensity. In a subsequent study, Tang et al. determined the clustering degree parameter, so-called exchange rate of interaction parameter (ER), that predicts the low-aggregating PAs liquid—liquid phase transition propensity, offering insight into their ability to form supramolecular self-assembled structures.[131] In addition to the self-assembling tendency of PAs, it is also possible to study the mobility of the peptides through the root-mean-square fluctuation (RMSF),[7] the molecular orientation of the different sections of the PAs within assemblies by calculating the exposure of the molecule to the solvent,[203] the distribution of the water within the assemblies,[39] or even to analyze the bundling between fibers by calculating the free energy between two fibers.[20]

Furthermore, atomistic simulations provide us with information on the peptide secondary structure, conformational changes, and thermodynamics. The Morphoscanner software was developed for the visualization and identification of β-structuring and β-sheet formation in CG simulations.[201] However, despite the hint provided by this software on β-sheet structure in CG models, detailed secondary structure is more adequately obtained from AA simulations by using the DSSP or STRIDE algorithms.[70,146,155,160,163] Conformational changes can be easily inferred using the molecular orientation of SPAs or the position of individual amino acids within the assemblies by calculating the radial density profile, density profile, Ramachandran maps, probability maps, or radius of gyration, among others.[71,95,115,151,152] Additionally, specific interactions can be calculated. Hydrogen bonding can be calculated straightforwardly through tools in different software packages, but π-stacking can be derived from proximity, radial distribution function, or contacts analyses. These calculations can be qualitatively correlated with spectroscopic experiments such as CD, UV—vis, NMR, or FT-IR.[72,204] Spectra can also be derived from MD structures using empirical-based algorithms.[37,76] Although they often require specific simulation set ups, the calculation of thermodynamics sheds light on the mechanism of self-assembly and structure stability. For instance, the entropy or free energy of SPAs, between monomers, or with the solvent can also be calculated.[166,188]

Alternatively, the combination with ab initio methods such as DFT permits the calculation of features to compare with experimental spectroscopic techniques such as CD, UV—vis, Overhauser dynamic nuclear polarization relaxometry (ODNP), NMR, or FT-IR.[19,37,76,197,205]

IV.d. Artificial Intelligence and Machine Learning. In recent years, significant advancements in artificial intelligence (AI) and machine learning (ML) applied to proteins, exemplified by tools like RoseTTAFold and AlphaFold, have

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

garnered considerable attention in the scientific community.[206−208] These consist of the prediction of protein structures from their amino acid sequence based on their homology with already-known architectures. They capitalize on the vast repository of crystallographic data available for protein structures, enabling algorithms to achieve remarkable accuracy, particularly in protein folding prediction tasks. However, the application of AI and ML for SPAs presents distinct challenges. Historically, the development of quantitative structure−activity relationship and quantitative structure−property relationship (QSAR/QSPR) models was widely used to screen chemical compounds and evaluate their properties or activity. These models were trained using mathematical models that established correlations between the structures of the compounds and properties or activity.[209−212] With the advent of AI and ML, these methodologies have been integrated into QSAR/QSPR models, significantly enhancing their predictive performance.[213] However, given the ongoing scarcity of resolved structures and the lack of standardized procedures, comparing results across different research groups remains challenging in the SPA field. Consequently, AI and ML approaches have primarily been utilized in SPAs to expedite the screening process in conjunction with other computational methods. These approaches are continuously being improved. For instance, Wang et al. undertook the synthesis and evaluation of 160 peptides to assess their self-assembling capabilities, providing essential data for their ML approach.[214] Another example is the work performed by Perez and Jones, who observed improved predictive performance when incorporating experimental NMR information into their ML-based predictions for self-assembling and coassembling peptides.[215]

The Tuttle group utilized an active ML method to extend their material discovery approach from dipeptides and tripeptides to peptides up to hexapeptides.[216] In their study, they highlighted the impracticality of simulating all possible sequences due to the exponential growth in number with increasing peptide length. Consequently, while many aggregation propensities (APs) were predicted rather than obtained from simulations, the vast search space necessitated a simpler model compared to the MARTINI model initially employed. Subsequently, in the second screening step, CG-MD techniques were employed to enhance the accuracy of AP calculations for promising candidates. This information was then utilized iteratively to refine the sequence selection process in the initial step. By applying a convergence criterion, the Tuttle group successfully proposed different sequences, validated using existing experimental data.

This approach also enabled Ferguson et al. to screen the tripeptide search space, predicting self-assembling APAs with a π-conjugated core by simulating only 2.3% of the overall sequences.[217] Another ML method was employed by Xiao et al. to identify amyloid-like structures, exhibiting excellent agreement with experimental validation, confirming the presence of β-sheets typical of amyloid order in all selected candidates.[218] They utilized an assisted method selection based on sequence comparison with known amyloids, leveraging existing experimental information for these protein-inspired materials. Additionally, Batra et al. introduced Monte Carlo steps to randomize sequence selection, aiming to reduce human bias favoring the search for β-sheet forming peptides.[219] These approaches share a common goal of expediting material discovery, employing diverse acceleration techniques and search criteria.

V. CONCLUSIONS AND REMARKS

Computational methods play a crucial role in filling the gaps left by experiments in SPAs. They pose unique challenges arising from their supramolecular nature underscored by the scarcity of experimentally resolved structures. To overcome these limitations, it is crucial to highlight that correlation with experiments serves not only as a means to validate models but also as an integral approach. Many studies in the field present both experimental and computational works in parallel, often in an iterative manner with each method enhancing the implementation and interpretation of the other, providing details not accessible by any single approach alone.

We have examined examples of applications of DFT, AA-MD, and CG-MD. To simplify, DFT provides insights into structure stability, interactions, and spectroscopic features. However, it lacks dynamism and is highly resource-demanding, limiting its application to a small number of molecules, much below the experimental structures composed of 10^2−10^3 molecules.

MD methods capture system dynamics over time, allowing stability calculations at various temperatures and conditions and can afford a larger number of molecules depending on the resolution (AA vs CG). AA models offer full resolution for representing molecular conformation and interactions but face challenges in modeling the self-assembly process. CG sacrifices some of these details to allow for longer and larger simulations that can simulate the entire self-assembly from scratch. Thus, if sufficient knowledge is available to construct a starting structure−making it unnecessary to model the entire self-assembly process − AA will provide finer details of SPAs. Even if the information is not enough to propose a single conformation, this approach can be used to compare among them, but care must be taken to avoid biases introduced by the limited size of the system. Enhanced sampling has addressed some problems in surpassing kinetic barriers to access the more stable molecular conformation in AA, leading to more reliable conformations, ensuring efficient sampling of the conformational space.

The ability of CG-MD to model the entire self-assembly process, allows it to assess structure formation or study self-assembly mechanisms. The MARTINI force field is the most extensively used, but the validity of the MARTINI 3 for SPAs is still under discussion. Proper validation should be carried out, and extreme care must be taken when modeling peptides below 5 amino acids long.

In addition to the model, factors such as box size, the number of molecules, and concentration significantly affect the output. Small boxes can be employed for an initial assessment of self-assembly, but large boxes are necessary to replicate the self-assembly process reliably. Additionally, the actual correlation between computational and experimental concentrations is currently unknown, but a scaling factor of 10 is commonly applied in CG simulations. Under adequate conditions, CG-MD holds the potential to screen self-assembling molecules, recently boosted by its combinations with machine learning (ML) methods. However, these methods are limited in the SPA field to enhance the efficiency of MD approaches, due to the lack of an experimental database of SPAs for training these methods. However, despite this

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

challenge, computational methods have achieved success in aiding the discovery of novel self-assembling sequences.

The contribution of computational methods to SPAs is indisputable. It has enabled a comprehensive understanding of these structures, their properties, and the determining factors. These results complement experimental findings, significantly advancing our ability to design and optimize materials for specific applications. The ongoing enhancement of models and computational resources holds the potential to further amplify their application in the coming years. Beyond the scope of self-assembly and structural features, these methods can be applied to comprehend complex systems. This encompasses the study of interactions among SPAs and with other systems related to their function, such as bilayers and proteins. This broader application allows for insights into bioactivity mechanisms and the parameters influencing them. The immense potential of computational methods in understanding supramolecular peptide systems underscores the importance of meticulous attention to calculation parameters and the necessity for adequate correlation with experimental data to ensure their validity and reliability. As these methods continue to evolve, they will persistently contribute to unraveling the intricacies of supramolecular peptide systems, providing unprecedented insights into their structures, properties, and functions.

AUTHOR INFORMATION

Corresponding Author

Ivan R. Sasselli − *Centro de Física de Materiales (CFM), CSIC-UPV/EHU, 20018 San Sebastián, Spain;* orcid.org/0000-0001-6062-2440; Email: i.sasselli@csic.es

Authors

Tomasz K. Piskorz − *Department of Chemistry, University of Oxford, Oxford OX1 3QZ, U.K.;* orcid.org/0000-0003-0716-6874

Laura Perez-Chirinos − *Center for Cooperative Research in Biomaterials (CIC biomaGUNE), Basque Research and Technology Alliance (BRTA), 20014 Donostia-San Sebastián, Spain*

Baofu Qiao − *Department of Natural Sciences, Baruch College, City University of New York, New York, New York 10010, United States;* orcid.org/0000-0001-8870-5985

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c02628

Author Contributions

#T.K.P. and L.P.-C. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

I.R.S. acknowledges financial support from Ramon y Cajal Program (RYC2021-033294-I), and Spanish State Research Agency (PID2022-136392NA-I00). B.Q. is thankful for the support from the National Science Foundation (2328095). L.P.-C. acknowledges financial support from the Spanish State Training Subprogram (PRE2019-090076) and the Maria de Maeztu Units of Excellence Program from the Spanish State Research Agency (MDM-2017-0720). We dedicate this work to all those advocates of science for whom the system failed, compelling them to diverge. May your excellence be more duly recognized now.

REFERENCES

(1) Lampel, A.; McPhee, S. A.; Park, H.-A.; Scott, G. G.; Humagain, S.; Hekstra, D. R.; Yoo, B.; Frederix, P. W.; Abzalimov, R. R.; Greenbaum, S. G.; Tuttle, T.; Hu, C.; Bettinger, C. J.; Ulijn, R. V. Polymeric peptide pigments with sequence-encoded properties. *Science* **2017**, *356*, 1064−1068.

(2) Reches, M.; Gazit, E. Casting Metal Nanowires within Discrete Self-assembled Peptide Nanotubes. *Science* **2003**, *300*, 625−627.

(3) Freeman, R.; Han, M.; Álvarez, Z.; Lewis, J. A.; Wester, J. R.; Stephanopoulos, N.; McClendon, M. T.; Lynsky, C.; Godbe, J. M.; Sangji, H.; Luijten, E.; Stupp, S. I. Reversible self-assembly of superstructured networks. *Science* **2018**, *362*, 808−813.

(4) Pappas, C. G.; Shafi, R.; Sasselli, I. R.; Siccardi, H.; Wang, T.; Narang, V.; Abzalimov, R.; Wijerathne, N.; Ulijn, R. V. Dynamic Peptide Libraries for Discovery of Supramolecular Nanomaterials. *Nat. Nanotechnol.* **2016**, *11*, 960−967.

(5) Alvarez, Z.; Ortega, J. A.; Sato, K.; Sasselli, I. R.; Kolberg-Edelbrock, A. N.; Qiu, R.; Marshall, K. A.; Nguyen, T. P.; Smith, C. S.; Quinlan, K. A.; Papakis, V.; Syrgiannis, Z.; Sather, N. A.; Mussumeci, C.; Engel, E.; Stupp, S. I.; Kiskinis, E. Artificial Extracellular Matrix Scaffolds of Mobile Molecules Enhance Maturation of Human Stem Cell-Derived Neurons. *Cell Stem Cell* **2023**, *30*, 219.

(6) Carlini, A. S.; Gaetani, R.; Braden, R. L.; Luo, C.; Christman, K. L.; Gianneschi, N. C. Enzyme-responsive progelator cyclic peptides for minimally invasive delivery to the heart post-myocardial infarction. *Nat. Commun.* **2019**, *10*, 1735.

(7) Álvarez, Z.; Kolberg-Edelbrock, A.; Sasselli, I.; Ortega, J.; Qiu, R.; Syrgiannis, Z.; Mirau, P.; Chen, F.; Chin, S.; Weigand, S.; Kiskinis, E.; Stupp, S. Bioactive scaffolds with enhanced supramolecular motion promote recovery from spinal cord injury. *Science* **2021**, *374*, 848−856.

(8) Hughes, M.; Frederix, P. W. J. M.; Raeburn, J.; Birchall, L. S.; Sadownik, J.; Coomer, F. C.; Lin, I. H.; Cussen, E. J.; Hunt, N. T.; Tuttle, T.; Webb, S. J.; Adams, D. J.; Ulijn, R. V. Sequence/Structure Relationships in Aromatic Dipeptide Hydrogels Formed Under Thermodynamic Control by Enzyme-assisted Self-assembly. *Soft Matter* **2012**, *8*, 5595−5602.

(9) Newcomb, C. J.; Sur, S.; Ortony, J. H.; Lee, O.-S.; Matson, J. B.; Boekhoven, J.; Yu, J. M.; Schatz, G. C.; Stupp, S. I. Cell death versus cell survival instructed by supramolecular cohesion of nanostructures. *Nat. Commun.* **2014**, *5*, 3321.

(10) Alakpa, E. V.; Jayawarna, V.; Lampel, A.; Burgess, K. V.; West, C. C.; Bakker, S. C.; Roy, S.; Javid, N.; Fleming, S.; Lamprou, D. A.; Yang, J.; Miller, A.; Urquhart, A. J.; Frederix, P. W. J. M.; Hunt, N. T.; Peault, B.; Ulijn, R. V.; Dalby, M. J. Tunable supramolecular hydrogels for selection of lineage-guiding metabolites in stem cell cultures. *Chem.* **2016**, *1*, 298−319.

(11) Martin, A. D.; Wojciechowski, J. P.; Du, E. Y.; Rawal, A.; Stefen, H.; Au, C. G.; Hou, L.; Cranfield, C. G.; Fath, T.; Ittner, L. M.; et al. Decoupling the effects of hydrophilic and hydrophobic moieties at the neuron-nanofibre interface. *Chem. Sci.* **2020**, *11*, 1375−1382.

(12) Cui, H.; Cheetham, A. G.; Pashuck, E. T.; Stupp, S. I. Amino acid sequence in constitutionally isomeric tetrapeptide amphiphiles dictates architecture of one-dimensional nanostructures. *J. Am. Chem. Soc.* **2014**, *136*, 12461−12468.

(13) Hu, Y.; Lin, R.; Zhang, P.; Fern, J.; Cheetham, A. G.; Patel, K.; Schulman, R.; Kan, C.; Cui, H. Electrostatic-driven lamination and untwisting of β-sheet assemblies. *ACS Nano* **2016**, *10*, 880−888.

(14) Lau, C. Y. J.; Fontana, F.; Mandemaker, L. D.; Wezendonk, D.; Vermeer, B.; Bonvin, A. M.; De Vries, R.; Zhang, H.; Remaut, K.; Van Den Dikkenberg, J.; et al. Control over the fibrillization yield by varying the oligomeric nucleation propensities of self-assembling peptides. *Commun. Chem.* **2020**, *3*, 164.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

(15) Bakota, E. L.; Sensoy, O.; Ozgur, B.; Sayar, M.; Hartgerink, J. D. Self-assembling multidomain peptide fibers with aromatic cores. *Biomacromolecules* **2013**, *14*, 1370−1378.
(16) Zhao, Y.; Wang, J.; Deng, L.; Zhou, P.; Wang, S.; Wang, Y.; Xu, H.; Lu, J. R. Tuning the self-assembly of short peptides via sequence variations. *Langmuir* **2013**, *29*, 13457−13464.
(17) Houton, K. A.; Morris, K. L.; Chen, L.; Schmidtmann, M.; Jones, J. T. A.; Serpell, L. C.; Lloyd, G. O.; Adams, D. J. On Crystal *versus* Fiber Formation in Dipeptide Hydrogelator Systems. *Langmuir* **2012**, *28*, 9797−9806.
(18) Adams, D. J.; Morris, K.; Chen, L.; Serpell, L. C.; Bacsa, J.; Day, G. M. The Delicate Balance Between Gelation and Crystallisation: Structural and Computational Investigations. *Soft Matter* **2010**, *6*, 4144−4156.
(19) Fleming, S.; Frederix, P. W. J. M.; Sasselli, I. R.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Assessing the Utility of Infrared Spectroscopy as a Structural Diagnostic Tool for beta-Sheets in Self-Assembling Aromatic Peptide Amphiphiles. *Langmuir* **2013**, *29*, 9510−9515.
(20) Sather, N. A.; Sai, H.; Sasselli, I. R.; Sato, K.; Ji, W.; Synatschke, C. V.; Zambrotta, R. T.; Edelbrock, J. F.; Kohlmeyer, R. R.; Hardin, J. O.; Berrigan, J. D.; Durstock, M. F.; Mirau, P.; Stupp, S. I. 3D Printing of Supramolecular Polymer Hydrogels with Hierarchical Structure. *Small* **2021**, *17*, 2005743.
(21) Smith, A. M.; Williams, R. J.; Tang, C.; Coppo, P.; Collins, R. F.; Turner, M. L.; Saiani, A.; Ulijn, R. V. Fmoc-Diphenylalanine Self Assembles to a Hydrogel via a Novel Architecture Based on π-π Interlocked β-Sheets. *Adv. Mater.* **2008**, *20*, 37−41.
(22) Sasselli, I. R.; Syrgiannis, Z. Small Molecules Organic Co-Assemblies as Functional Nanomaterials. *Eur. J. Org. Chem.* **2020**, *2020*, 5305−5318.
(23) Fleming, S.; Debnath, S.; Frederix, P. W.; Hunt, N. T.; Ulijn, R. V. Insights into the coassembly of hydrogelators and surfactants based on aromatic peptide amphiphiles. *Biomacromolecules* **2014**, *15*, 1171−1184.
(24) Edelbrock, A. N.; Àlvarez, Z.; Simkin, D.; Fyrner, T.; Chin, S. M.; Sato, K.; Kiskinis, E.; Stupp, S. I. Supramolecular Nanostructure Activates TrkB Receptor Signaling of Neuronal Cells by Mimicking Brain-Derived Neurotrophic Factor. *Nano Lett.* **2018**, *18*, 6237−6247.
(25) Sur, S.; Tantakitti, F.; Matson, J. B.; Stupp, S. I. Epitope topography controls bioactivity in supramolecular nanofibers. *Biomater. Sci.* **2015**, *3*, 520−532.
(26) Frederix, P. W.; Patmanidis, I.; Marrink, S. J. Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments. *Chem. Soc. Rev.* **2018**, *47*, 3470−3489.
(27) Ramakrishnan, M.; van Teijlingen, A.; Tuttle, T.; Ulijn, R. V. Integrating Computation, Experiment, and Machine Learning in the Design of Peptide-Based Supramolecular Materials and Systems. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202218067.
(28) Garcia, A. M.; Melchionna, M.; Bellotto, O.; Kralj, S.; Semeraro, S.; Parisi, E.; Iglesias, D.; D'Andrea, P.; De Zorzi, R.; Vargiu, A. V.; et al. Nanoscale assembly of functional peptides with divergent programming elements. *ACS Nano* **2021**, *15*, 3015−3025.
(29) Eckes, K. M.; Mu, X.; Ruehle, M. A.; Ren, P.; Suggs, L. J. β Sheets Not Required: Combined Experimental and Computational Studies of Self-Assembly and Gelation of the Ester-Containing Analogue of an Fmoc-Dipeptide Hydrogelator. *Langmuir* **2014**, *30*, 5287−5296.
(30) Caruso, M.; Gatto, E.; Placidi, E.; Ballano, G.; Formaggio, F.; Toniolo, C.; Zanuy, D.; Alemán, C.; Venanzi, M. A single-residue substitution inhibits fibrillization of Ala-based pentapeptides. A spectroscopic and molecular dynamics investigation. *Soft Matter* **2014**, *10*, 2508−2519.
(31) Ozkan, A. D.; Tekinay, A. B.; Guler, M. O.; Tekin, E. D. Effects of temperature, pH and counterions on the stability of peptide amphiphile nanofiber structures. *RSC Adv.* **2016**, *6*, 104201−104214.
(32) Nir, S.; Zanuy, D.; Zada, T.; Agazani, O.; Aleman, C.; Shalev, D. E.; Reches, M. Tailoring the self-assembly of a tripeptide for the formation of antimicrobial surfaces. *Nanoscale* **2019**, *11*, 8752−8759.
(33) Thota, N.; Ma, Y.; Jiang, J. Molecular insights into the self-assembly of short amphiphilic peptides F m D n and F m K n. *RSC Adv.* **2014**, *4*, 60741−60748.
(34) Colherinhas, G.; Fileti, E. Molecular dynamics study of surfactant-like peptide based nanostructures. *J. Phys. Chem. B* **2014**, *118*, 12215−12222.
(35) Zhao, Y.; Deng, L.; Wang, J.; Xu, H.; Lu, J. R. Solvent controlled structural transition of KI4K self-assemblies: from nanotubes to nanofibrils. *Langmuir* **2015**, *31*, 12975−12983.
(36) Sasselli, I. R.; Pappas, C. G.; Matthews, E.; Wang, T.; Hunt, N.; Ulijn, R.; Tuttle, T. Using experimental and computational energy equilibration to understand hierarchical self-assembly of Fmoc-dipeptide amphiphiles. *Soft Matter* **2016**, *12*, 8307−8315.
(37) Frederix, P. W.; Idé, J.; Altay, Y.; Schaeffer, G. l.; Surin, M.; Beljonne, D.; Bondarenko, A. S.; Jansen, T. L.; Otto, S.; Marrink, S. J. Structural and spectroscopic properties of assemblies of self-replicating peptide macrocycles. *ACS Nano* **2017**, *11*, 7858−7868.
(38) Moreira, I. P.; Piskorz, T. K.; van Esch, J. H.; Tuttle, T.; Ulijn, R. V. Biocatalytic self-assembly of tripeptide gels and emulsions. *Langmuir* **2017**, *33*, 4986−4995.
(39) Lee, O.-S.; Cho, V.; Schatz, G. C. Modeling the Self-Assembly of Peptide Amphiphiles into Fibers Using Coarse-Grained Molecular Dynamics. *Nano Lett.* **2012**, *12*, 4907−4913.
(40) Bochicchio, D.; Pavan, G. M. Molecular modelling of supramolecular polymers. *Adv. Phys. X* **2018**, *3*, 1436408.
(41) Dral, P. O.; Hourahine, B.; Grimme, S. Modern semiempirical electronic structure methods. *J. Chem. Phys.* **2024**, *160*, 040401.
(42) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, No. e1493.
(43) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(44) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931−948.
(45) Stewart, J. J. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(46) Stewart, J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1−32.
(47) Lehtola, S.; Steigemann, C.; Oliveira, M. J.; Marques, M. A. Recent developments in libxc—A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1−5.
(48) Perdew, J. P.; Schmidt, K. In *Jacob's ladder of density functional approximations for the exchange-correlation energy*; AIP Conf. Proc., American Institute of Physics: 2001; pp 1−20.
(49) Kruse, H.; Goerigk, L.; Grimme, S. Why the standard B3LYP/6-31G* model chemistry should not be used in DFT calculations of molecular thermochemistry: understanding and correcting the problem. *J. Org. Chem.* **2012**, *77*, 10824−10834.
(50) Bursch, M.; Mewes, J. M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202205735.
(51) Goerigk, L.; Mehta, N. A trip to the density functional theory zoo: warnings and recommendations for the user. *Aust. J. Chem.* **2019**, *72*, 563−573.
(52) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.
(53) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob's ladder: Defining and testing the

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

ωB97M (2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.
(54) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.
(55) Piana, F.; Case, D. H.; Ramalhete, S. M.; Pileio, G.; Facciotti, M.; Day, G. M.; Khimyak, Y. Z.; Angulo, J.; Brown, R. C.; Gale, P. A. Substituent interference on supramolecular assembly in urea gelators: synthesis, structure prediction and NMR. *Soft Matter* **2016**, *12*, 4034−4043.
(56) Liyanage, W.; Nilsson, B. L. Substituent Effects on the Self-Assembly/Coassembly and Hydrogelation of Phenylalanine Derivatives. *Langmuir* **2016**, *32*, 787−799.
(57) Wang, M.; Zhou, P.; Wang, J.; Zhao, Y.; Ma, H.; Lu, J. R.; Xu, H. Left or right: how does amino acid chirality affect the handedness of nanostructures self-assembled from short amphiphilic peptides? *J. Am. Chem. Soc.* **2017**, *139*, 4185−4194.
(58) Mayans, E.; Ballano, G.; Casanovas, J.; Díaz, A.; Pérez-Madrigal, M. M.; Estrany, F.; Puiggalí, J.; Cativiela, C.; Alemán, C. Self-Assembly of Tetraphenylalanine Peptides. *Chem.—Eur. J.* **2015**, *21*, 16895−16905.
(59) Pérez-Madrigal, M. M.; Gil, A. M.; Casanovas, J.; Jiménez, A. I.; Macor, L. P.; Alemán, C. Self-assembly pathways in a triphenylalanine peptide capped with aromatic groups. *Colloids Surf., B* **2022**, *216*, 112522.
(60) Gil, A. M.; Casanovas, J.; Mayans, E.; Jiménez, A. I.; Puiggalí, J.; Alemán, C. Heterochirality Restricts the Self-Assembly of Phenylalanine Dipeptides Capped with Highly Aromatic Groups. *J. Phys. Chem. B* **2020**, *124*, 5913−5918.
(61) Casanovas, J.; Mayans, E.; Díaz, A.; Gil, A. M.; Jiménez, A. I.; Cativiela, C.; Puiggalí, J.; Alemán, C. Amyloid fibrils from organic solutions of an amphiphilic dipeptide. *Chem. Commun.* **2019**, *55*, 8556−8559.
(62) Mayans, E.; Casanovas, J.; Gil, A. M.; Jiménez, A. I.; Cativiela, C.; Puiggalí, J.; Alemán, C. Diversity and hierarchy in supramolecular assemblies of triphenylalanine: from laminated helical ribbons to toroids. *Langmuir* **2017**, *33*, 4036−4048.
(63) Mayans, E.; Ballano, G.; Casanovas, J.; Del Valle, L. J.; Pérez-Madrigal, M. M.; Estrany, F.; Jiménez, A. I.; Puiggalí, J.; Cativiela, C.; Alemán, C. Hierarchical self-assembly of di-, tri-and tetraphenylalanine peptides capped with two fluorenyl functionalities: from polymorphs to dendrites. *Soft Matter* **2016**, *12*, 5475−5488.
(64) Kuila, S.; Singh, A. K.; Shrivastava, A.; Dey, S.; Singha, T.; Roy, L.; Satpati, B.; Nanda, J. Probing Molecular Chirality on the Self-Assembly and Gelation of Naphthalimide-Conjugated Dipeptides. *J. Phys. Chem. B* **2023**, *127*, 4808.
(65) Misra, S.; Singh, P.; Singh, A. K.; Roy, L.; Kuila, S.; Dey, S.; Mahapatra, A. K.; Nanda, J. Tuning of the Supramolecular Helicity of Peptide-Based Gel Nanofibers. *J. Phys. Chem. B* **2022**, *126*, 10882−10892.
(66) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.
(67) Groenhof, G. Introduction to QM/MM simulations. *Biomolecular simulations: methods and protocols* **2013**, *924*, 43−66.
(68) Cao, L.; Ryde, U. On the difference between additive and subtractive QM/MM calculations. *Front. Chem.* **2018**, *6*, 89.
(69) Jeon, J.; Mills, C. E.; Shell, M. S. Molecular insights into diphenylalanine nanotube assembly: all-atom simulations of oligomerization. *J. Phys. Chem. B* **2013**, *117*, 3935−3943.
(70) Tang, J. D.; Mura, C.; Lampe, K. J. Stimuli-responsive, pentapeptide, nanofiber hydrogel for tissue engineering. *J. Am. Chem. Soc.* **2019**, *141*, 4886−4899.
(71) Zanuy, D.; Puiggalí-Jou, A.; Conflitti, P.; Bocchinfuso, G.; Palleschi, A.; Alemán, C. Aggregation propensity of therapeutic fibrin-homing pentapeptides: insights from experiments and molecular dynamics simulations. *Soft Matter* **2020**, *16*, 10169−10179.
(72) Noble Jesus, C.; Evans, R.; Forth, J.; Estarellas, C.; Gervasio, F. L.; Battaglia, G. Amphiphilic histidine-based oligopeptides exhibit pH-reversible fibril formation. *ACS Macro Lett.* **2021**, *10*, 984−989.
(73) Marrink, S. J.; Periole, X.; Tieleman, D. P.; de Vries, A. H. Comment on "On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models" by M. Winger, D. Trzesniak, R. Baron and WF van Gunsteren. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2254.
(74) Sato, K.; Ji, W.; Álvarez, Z.; Palmer, L. C.; Stupp, S. I. Chiral recognition of lipid bilayer membranes by supramolecular assemblies of peptide amphiphiles. *ACS Biomater. Sci. Eng.* **2019**, *5*, 2786−2792.
(75) Clover, T. M.; O'Neill, C. L.; Appavu, R.; Lokhande, G.; Gaharwar, A. K.; Posey, A. E.; White, M. A.; Rudra, J. S. Self-Assembly of Block Heterochiral Peptides into Helical Tapes. *J. Am. Chem. Soc.* **2020**, *142*, 19809−19813.
(76) Garcia, A. M.; Iglesias, D.; Parisi, E.; Styan, K. E.; Waddington, L. J.; Deganutti, C.; De Zorzi, R.; Grassi, M.; Melchionna, M.; Vargiu, A. V.; et al. Chirality effects on peptide self-assembly unraveled from molecules to materials. *Chem.* **2018**, *4*, 1862−1876.
(77) Vargiu, A. V.; Iglesias, D.; Styan, K.; Waddington, L.; Easton, C.; Marchesan, S. Design of a hydrophobic tripeptide that self-assembles into amphiphilic superstructures forming a hydrogel biomaterial. *Chem. Commun.* **2016**, *52*, 5912−5915.
(78) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.
(79) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.
(80) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Meth.* **2017**, *14*, 71−73.
(81) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.
(82) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.
(83) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.
(84) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(85) Leung, C.-Y.; Palmer, L. C.; Kewalramani, S.; Qiao, B.; Stupp, S. I.; Olvera de la Cruz, M.; Bedzyk, M. J. Crystalline polymorphism induced by charge regulation in ionic membranes. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16309−16314.
(86) Leung, C.-Y.; Palmer, L. C.; Qiao, B.; Kewalramani, S.; Sknepnek, R.; Newcomb, C. J.; Greenfield, M. A.; Vernizzi, G.; Stupp, S. I.; Bedzyk, M. J.; Olvera de la Cruz, M. Molecular Crystallization Controlled by pH Regulates Mesoscopic Membrane Morphology. *ACS Nano* **2012**, *6*, 10901−10909.
(87) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PLoS One* **2012**, *7*, No. e32131.
(88) Petrov, D.; Zagrovic, B. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *PLoS Comput. Biol.* **2014**, *10*, No. e1003638.
(89) Georgoulia, P. S.; Glykos, N. M. Molecular simulation of peptides coming of age: Accurate prediction of folding, dynamics and structures. *Arch. Biochem. Biophys.* **2019**, *664*, 76−88.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

(90) Smith, M. D.; Rao, J. S.; Segelken, E.; Cruz, L. Force-field induced bias in the structure of Aβ21−30: A comparison of OPLS; AMBER, CHARMM, and GROMOS force fields. *J. Chem. Inf. Model.* **2015**, *55*, 2587−2595.
(91) Kashefolgheta, S.; Wang, S.; Acree, W. E.; Hünenberger, P. H. Evaluation of nine condensed-phase force fields of the GROMOS; CHARMM, OPLS, AMBER, and OpenFF families against experimental cross-solvation free energies. *Phys. Chem. Chem. Phys.* **2021**, *23*, 13055−13074.
(92) Man, V. H.; He, X.; Gao, J.; Wang, J. Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of PHF6 Peptide of Tau Protein. *J. Chem. Theory Comput.* **2021**, *17*, 6458−6471.
(93) Man, V. H.; He, X.; Derreumaux, P.; Ji, B.; Xie, X. Q.; Nguyen, P. H.; Wang, J. Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of Aβ(16−22) Dimer. *J. Chem. Theory Comput.* **2019**, *15*, 1440−1452.
(94) Samantray, S.; Yin, F.; Kav, B.; Strodel, B. Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 6462−6475.
(95) Nap, R. J.; Qiao, B.; Palmer, L. C.; Stupp, S. I.; Olvera de la Cruz, M.; Szleifer, I. Acid-Base Equilibrium and Dielectric Environment Regulate Charge in Supramolecular Nanofibers. *Front. Chem.* **2022**, *10*, 852164.
(96) Li, Y.; Kim, M.; Pial, T. H.; Lin, Y.; Cui, H.; Olvera de la Cruz, M. Aggregation-Induced Asymmetric Charge States of Amino Acids in Supramolecular Nanofibers. *J. Phys. Chem. B* **2023**, *127*, 8176−8184.
(97) Buslaev, P.; Aho, N.; Jansen, A.; Bauer, P.; Hess, B.; Groenhof, G. Best Practices in Constant pH MD Simulations: Accuracy and Sampling. *J. Chem. Theory Comput.* **2022**, *18*, 6134−6147.
(98) Aho, N.; Buslaev, P.; Jansen, A.; Bauer, P.; Groenhof, G.; Hess, B. Scalable Constant pH Molecular Dynamics in GROMACS. *J. Chem. Theory Comput.* **2022**, *18*, 6148−6160.
(99) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems. *J. Chem. Theory Comput.* **2017**, *13*, 5933−5944.
(100) Harris, R. C.; Shen, J. GPU-Accelerated Implementation of Continuous Constant pH Molecular Dynamics in Amber: pKa Predictions with Single-pH Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4821−4832.
(101) Siu, S. W.; Pluhackova, K.; Böckmann, R. A. Optimization of the OPLS-AA force field for long hydrocarbons. *J. Chem. Theory Comput.* **2012**, *8*, 1459−1470.
(102) Ghahremanpour, M. M.; Tirado-Rives, J.; Jorgensen, W. L. Refinement of the Optimized Potentials for Liquid Simulations Force Field for Thermodynamics and Dynamics of Liquid Alkanes. *J. Phys. Chem. B* **2022**, *126*, 5896−5907.
(103) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(104) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331−w336.
(105) Sasselli, I. R.; Ulijn, R. V.; Tuttle, T. CHARMM force field parameterization protocol for self-assembling peptide amphiphiles: the Fmoc moiety. *Phys. Chem. Chem. Phys.* **2016**, *18*, 4659−4667.
(106) Lang, E. J.; Baker, E. G.; Woolfson, D. N.; Mulholland, A. J. Generalized Born implicit solvent models do not reproduce secondary structures of de novo designed Glu/Lys peptides. *J. Chem. Theory Comput.* **2022**, *18*, 4070−4076.
(107) Marcisz, M.; Samsonov, S. A. Solvent model benchmark for molecular dynamics of glycosaminoglycans. *J. Chem. Inf. Model.* **2023**, *63*, 2147−2157.
(108) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.
(109) Boonstra, S.; Onck, P. R.; van der Giessen, E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J. Phys. Chem. B* **2016**, *120*, 3692−3698.
(110) Reif, M. M.; Winger, M.; Oostenbrink, C. Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 1247−1264.
(111) Robertson, M. J.; Qian, Y.; Robinson, M. C.; Tirado-Rives, J.; Jorgensen, W. L. Development and Testing of the OPLS-AA/M Force Field for RNA. *J. Chem. Theory Comput.* **2019**, *15*, 2734−2742.
(112) Deiss-Yehiely, E.; Ortony, J. H.; Qiao, B.; Stupp, S. I.; Olvera de La Cruz, M. Ion condensation onto self-assembled nanofibers. *J. Polym. Sci., Part B: Polym. Phys.* **2017**, *55*, 901−906.
(113) Sun, H.; Qiao, B.; Choi, W.; Hampu, N.; McCallum, N. C.; Thompson, M. P.; Oktawiec, J.; Weigand, S.; Ebrahim, O. M.; de la Cruz, M. O.; Gianneschi, N. C. Origin of Proteolytic Stability of Peptide-Brush Polymers as Globular Proteomimetics. *ACS Cent. Sci.* **2021**, *7*, 2063−2072.
(114) Iscen, A.; Schatz, G. C. Hofmeister Effects on Peptide Amphiphile Nanofiber Self-Assembly. *J. Phys. Chem. B* **2019**, *123*, 7006−7013.
(115) Lin, Y.-A.; Kang, M.; Chen, W.-C.; Ou, Y.-C.; Cheetham, A. G.; Wu, P.-H.; Wirtz, D.; Loverde, S. M.; Cui, H. Isomeric control of the mechanical properties of supramolecular filament hydrogels. *Biomater. Sci.* **2018**, *6*, 216−224.
(116) Qiao, B.; Skanthakumar, S.; Soderholm, L. Comparative CHARMM and AMOEBA Simulations of Lanthanide Hydration Energetics and Experimental Aqueous-Solution Structures. *J. Chem. Theory Comput.* **2018**, *14*, 1781−1790.
(117) Zhang, H.; Wang, H.; Xu, G.; Yuan, S. A molecular dynamics simulation of N-(fluorenyl-9-methoxycarbonyl)-dipeptides supramolecular hydrogel. *Colloids Surf., A* **2013**, *417*, 217−223.
(118) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Probing the Self-Assembly Mechanism of Diphenylalanine-Based Peptide Nanovesicles and Nanotubes. *ACS Nano* **2012**, *6*, 3907−3918.
(119) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Triphenylalanine peptides self-assemble into nanospheres and nanorods that are different from the nanovesicles and nanotubes formed by diphenylalanine peptides. *Nanoscale* **2014**, *6*, 2800−2811.
(120) Sasselli, I. R.; Syrgiannis, Z.; Sather, N. A.; Palmer, L. C.; Stupp, S. I. Modeling Interactions within and between Peptide Amphiphile Supramolecular Filaments. *J. Phys. Chem. B* **2022**, *126*, 650−659.
(121) Sasselli, I. R.; Coluzza, I. Assessment of the MARTINI 3 Performance for Short Peptide Self-Assembly. *J. Chem. Theory Comput.* **2024**, *20*, 224−238.
(122) Velichko, Y. S.; Stupp, S. I.; de la Cruz, M. O. Molecular simulation study of peptide amphiphile self-assembly. *J. Phys. Chem. B* **2008**, *112*, 2326−2334.
(123) Hooten, M.; Banerjee, A.; Dutt, M. Multiscale, Multiresolution Coarse-Grained Model via a Hybrid Approach: Solvation, Structure, and Self-Assembly of Aromatic Tripeptides. *J. Chem. Theory Comput.* **2024**, *20*, 1689−1703.
(124) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(125) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(126) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(127) Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. Virtual Screening for Dipeptide Aggregation: Toward Predictive

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

Tools for Peptide Self-Assembly. *J. Phys. Chem. Lett.* **2011**, *2*, 2380−2384.
(128) Frederix, P. W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Exploring the Sequence Space for (Tri-)peptide Self-assembly to Design and Discover New Hydrogels. *Nat. Chem.* **2015**, *7*, 30−37.
(129) Scott, G. G.; McKnight, P. J.; Tuttle, T.; Ulijn, R. V. Tripeptide Emulsifiers. *Adv. Mater.* **2016**, *28*, 1381−1386.
(130) Hu, T.; Zhang, Z.; Hu, H.; Euston, S. R.; Pan, S. A comprehensive study on self-assembly and gelation of C13-dipeptides—from design strategies to functionalities. *Biomacromolecules* **2020**, *21*, 670−679.
(131) Tang, Y.; Bera, S.; Yao, Y.; Zeng, J.; Lao, Z.; Dong, X.; Gazit, E.; Wei, G. Prediction and characterization of liquid-liquid phase separation of minimalistic peptides. *Cell Rep. Phys. Sci.* **2021**, *2*, 100579.
(132) Manandhar, A.; Chakraborty, K.; Tang, P. K.; Kang, M.; Zhang, P.; Cui, H.; Loverde, S. M. Rational Coarse-Grained Molecular Dynamics Simulations of Supramolecular Anticancer Nanotubes. *J. Phys. Chem. B* **2019**, *123*, 10582−10593.
(133) Guo, C.; Arnon, Z. A.; Qi, R.; Zhang, Q.; Adler-Abramovich, L.; Gazit, E.; Wei, G. Expanding the Nanoarchitectural Diversity Through Aromatic Di-and Tri-Peptide Coassembly: Nanostructures and Molecular Mechanisms. *ACS Nano* **2016**, *10*, 8316−8324.
(134) Abul-Haija, Y. M.; Scott, G. G.; Sahoo, J. K.; Tuttle, T.; Ulijn, R. Cooperative, ion-sensitive co-assembly of tripeptide hydrogels. *Chem. Commun.* **2017**, *53*, 9562−9565.
(135) Moreira, I. P.; Scott, G. G.; Ulijn, R. V.; Tuttle, T. Computational prediction of tripeptide-dipeptide co-assembly. *Mol. Phys.* **2019**, *117*, 1151−1163.
(136) Bera, S.; Mondal, S.; Tang, Y.; Jacoby, G.; Arad, E.; Guterman, T.; Jelinek, R.; Beck, R.; Wei, G.; Gazit, E. Deciphering the Rules for Amino Acid Co-Assembly Based on Interlayer Distances. *ACS Nano* **2019**, *13*, 1703−1712.
(137) Sasselli, I. R.; Moreira, I. P.; Ulijn, R. V.; Tuttle, T. Molecular dynamics simulations reveal disruptive self-assembly in dynamic peptide libraries. *Org. Biomol. Chem.* **2017**, *15*, 6541−6547.
(138) Qiu, R.; Sasselli, I. R.; Álvarez, Z.; Sai, H.; Ji, W.; Palmer, L. C.; Stupp, S. I. Supramolecular Copolymers of Peptides and Lipidated Peptides and Their Therapeutic Potential. *J. Am. Chem. Soc.* **2022**, *144*, 5562−5574.
(139) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.
(140) van Teijlingen, A.; Smith, M. C.; Tuttle, T. Short Peptide Self-Assembly in the Martini Coarse-Grain Force Field Family. *Acc. Chem. Res.* **2023**, *56*, 644−654.
(141) Fu, I. W.; Markegard, C. B.; Nguyen, H. D. Solvent effects on kinetic mechanisms of self-assembly by peptide amphiphiles via molecular dynamics simulations. *Langmuir* **2015**, *31*, 315−324.
(142) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.
(143) Kwon, J.; Lee, M.; Na, S. Sodium chloride's effect on self-assembly of diphenylalanine bilayer. *J. Comput. Chem.* **2016**, *37*, 1839−1846.
(144) van Teijlingen, A.; Swanson, H. W.; Lau, K. H. A.; Tuttle, T. Constant pH coarse-grained molecular dynamics with stochastic charge neutralization. *J. Phys. Chem. Lett.* **2022**, *13*, 4046−4051.
(145) Grünewald, F.; Souza, P. C.; Abdizadeh, H.; Barnoud, J.; de Vries, A. H.; Marrink, S. J. Titratable Martini model for constant pH simulations. *J. Chem. Phys.* **2020**, *153*, 024118.
(146) Lee, O.-S.; Stupp, S. I.; Schatz, G. C. Atomistic Molecular Dynamics Simulations of Peptide Amphiphile Self-Assembly into Cylindrical Nanofibers. *J. Am. Chem. Soc.* **2011**, *133*, 3677−3683.
(147) Moitra, P.; Subramanian, Y.; Bhattacharya, S. Concentration dependent self-assembly of TrK-NGF receptor derived tripeptide: new insights from experiment and computer simulations. *J. Phys. Chem. B* **2017**, *121*, 815−824.
(148) Erdogan, H.; Babur, E.; Yilmaz, M.; Candas, E.; Gordesel, M.; Dede, Y.; Oren, E. E.; Demirel, G. B.; Ozturk, M. K.; Yavuz, M. S.; et al. Morphological Versatility in the Self-Assembly of Val-Ala and Ala-Val Dipeptides. *Langmuir* **2015**, *31*, 7337−7345.
(149) Moreira, I. P.; Sasselli, I. R.; Cannon, D.; Hughes, M.; Lamprou, D. A.; Tuttle, T.; Ulijn, R. Enzymatically activated emulsions stabilised by interfacial nanofibre networks. *Soft Matter* **2016**, *12*, 2623−2631.
(150) Maity, S.; Ottelé, J.; Santiago, G. M.; Frederix, P. W.; Kroon, P.; Markovitch, O.; Stuart, M. C.; Marrink, S. J.; Otto, S.; Roos, W. H. Caught in the Act: Mechanistic Insight into Supramolecular Polymerization-Driven Self-Replication from Real-Time Visualization. *J. Am. Chem. Soc.* **2020**, *142*, 13709−13717.
(151) Levine, Z. A.; Larini, L.; LaPointe, N. E.; Feinstein, S. C.; Shea, J.-E. Regulation and aggregation of intrinsically disordered peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2758−2763.
(152) Zhou, P.; Deng, L.; Wang, Y.; Lu, J. R.; Xu, H. Interplay between intrinsic conformational propensities and intermolecular interactions in the self-assembly of short surfactant-like peptides composed of leucine/isoleucine. *Langmuir* **2016**, *32*, 4662−4672.
(153) Hughes, M.; Xu, H. X.; Frederix, P. W. J. M.; Smith, A. M.; Hunt, N. T.; Tuttle, T.; Kinloch, I. A.; Ulijn, R. V. Biocatalytic self-assembly of 2D peptide-based nanostructures. *Soft Matter* **2011**, *7*, 10032−10038.
(154) Kralj, S.; Bellotto, O.; Parisi, E.; Garcia, A. M.; Iglesias, D.; Semeraro, S.; Deganutti, C.; D'Andrea, P.; Vargiu, A. V.; Geremia, S.; De Zorzi, R.; Marchesan, S. Heterochirality and Halogenation Control Phe-Phe Hierarchical Assembly. *ACS Nano* **2020**, *14*, 16951−16961.
(155) Lee, O.-S.; Liu, Y.; Schatz, G. C. Molecular dynamics simulation of β-sheet formation in self-assembled peptide amphiphile fibers. *J. Nanopart. Res.* **2012**, *14*, 1−7.
(156) Xiong, Q.; Stupp, S. I.; Schatz, G. C. Molecular Insight into the β-Sheet Twist and Related Morphology of Self-Assembled Peptide Amphiphile Ribbons. *J. Phys. Chem. Lett.* **2021**, *12*, 11238−11244.
(157) Castelletto, V.; Moulton, C.; Cheng, G.; Hamley, I.; Hicks, M. R.; Rodger, A.; López-Pérez, D. E.; Revilla-López, G.; Alemán, C. Self-assembly of Fmoc-tetrapeptides based on the RGDS cell adhesion motif. *Soft Matter* **2011**, *7*, 11405−11415.
(158) Lopez-Perez, D. E.; Revilla-Lopez, G.; Hamley, I. W.; Aleman, C. Molecular insights into aggregates made of amphiphilic Fmoc-tetrapeptides. *Soft Matter* **2013**, *9*, 11021−11032.
(159) Wang, J.; Tao, K.; Zhou, P.; Pambou, E.; Li, Z.; Xu, H.; Rogers, S.; King, S.; Lu, J. R. Tuning self-assembled morphology of the Aβ (16−22) peptide by substitution of phenylalanine residues. *Colloids Surf., B* **2016**, *147*, 116−123.
(160) Mu, X. J.; Eckes, K. M.; Nguyen, M. M.; Suggs, L. J.; Ren, P. Y. Experimental and Computational Studies Reveal an Alternative Supramolecular Structure for Fmoc-Dipeptide Self-Assembly. *Biomacromolecules* **2012**, *13*, 3562−3571.
(161) Mazza, M.; Notman, R.; Anwar, J.; Rodger, A.; Hicks, M.; Parkinson, G.; McCarthy, D.; Daviter, T.; Moger, J.; Garrett, N.; et al. Nanofiber-based delivery of therapeutic peptides to the brain. *ACS Nano* **2013**, *7*, 1016−1026.
(162) Sun, F.; Chen, L.; Ding, X.; Xu, L.; Zhou, X.; Wei, P.; Liang, J. F.; Luo, S.-Z. High-resolution insights into the stepwise self-assembly of nanofiber from bioactive peptides. *J. Phys. Chem. B* **2017**, *121*, 7421−7430.
(163) Tantakitti, F.; Boekhoven, J.; Wang, X.; Kazantsev, R. V.; Yu, T.; Li, J.; Zhuang, E.; Zandi, R.; Ortony, J. H.; Newcomb, C. J.; Palmer, L. C.; Shekhawat, G. S.; Olvera de la Cruz, M.; Schatz, G. C.; Stupp, S. I. Energy landscapes and functions of supramolecular systems. *Nat. Mater.* **2016**, *15*, 469−476.
(164) Hamley, I. W.; Castelletto, V.; Dehsorkhi, A.; Torras, J.; Aleman, C.; Portnaya, I.; Danino, D. The conformation and aggregation of proline-rich surfactant-like peptides. *J. Phys. Chem. B* **2018**, *122*, 1826−1835.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

(165) Hesser, M.; Thursch, L.; Lewis, T.; DiGuiseppi, D.; Alvarez, N. J.; Schweitzer-Stenner, R. The tripeptide GHG as an unexpected hydrogelator triggered by imidazole deprotonation. *Soft Matter* **2020**, *16*, 4110−4114.
(166) Yu, T.; Lee, O.-S.; Schatz, G. C. Steered Molecular Dynamics Studies of the Potential of Mean Force for Peptide Amphiphile Self-Assembly into Cylindrical Nanofibers. *J. Phys. Chem. A* **2013**, *117*, 7453−7460.
(167) Tekin, E. D. Molecular dynamics simulations of self-assembled peptide amphiphile based cylindrical nanofibers. *RSC Adv.* **2015**, *5*, 66582−66590.
(168) Dana, A.; Tekinay, A. B.; Tekin, E. D. A comparison of peptide amphiphile nanofiber macromolecular assembly strategies. *Eur. Phys. J. E* **2019**, *42*, 63.
(169) Mehralitabar, H.; Taghdir, M.; Naderi-Manesh, H. A combination of bioactive and nonbioactive alkyl-peptides form a more stable nanofiber structure for differentiating neural stem cells: a molecular dynamics simulation survey. *J. Biomol. Struct. Dyn.* **2019**, *37*, 3434−3444.
(170) Lai, C.-T.; Rosi, N. L.; Schatz, G. C. All-Atom Molecular Dynamics Simulations of Peptide Amphiphile Assemblies That Spontaneously Form Twisted and Helical Ribbon Structures. *J. Phys. Chem. Lett.* **2017**, *8*, 2170−2174.
(171) Carnall, J. M. A.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C. A.; Peyralans, J. J.-P.; Otto, S. Mechanosensitive Self-Replication Driven by Self-Organization. *Science* **2010**, *327*, 1502−1506.
(172) Zhou, P.; Deng, L.; Wang, Y.; Lu, J. R.; Xu, H. Different nanostructures caused by competition of intra-and inter-β-sheet interactions in hierarchical self-assembly of short peptides. *J. Colloid Interface Sci.* **2016**, *464*, 219−228.
(173) Raz, Y.; Rubinov, B.; Matmor, M.; Rapaport, H.; Ashkenasy, G.; Miller, Y. Effects of mutations in de novo designed synthetic amphiphilic β-sheet peptides on self-assembly of fibrils. *Chem. Commun.* **2013**, *49*, 6561−6563.
(174) Ivnitski, D.; Amit, M.; Silberbush, O.; Atsmon-Raz, Y.; Nanda, J.; Cohen-Luria, R.; Miller, Y.; Ashkenasy, G.; Ashkenasy, N. The strong influence of structure polymorphism on the conductivity of peptide fibrils. *Angew. Chem. Int. Ed.* **2016**, *55*, 9988−9992.
(175) Paul, A.; Li, W.-H.; Viswanathan, G. K.; Arad, E.; Mohapatra, S.; Li, G.; Jelinek, R.; Gazit, E.; Li, Y.-M.; Segal, D. Tryptophan-glucosamine conjugates modulate tau-derived PHF6 aggregation at low concentrations. *Chem. Commun.* **2019**, *55*, 14621−14624.
(176) Zanuy, D.; Poater, J.; Sola, M.; Hamley, I. W.; Aleman, C. Fmoc-RGDS based fibrils: atomistic details of their hierarchical assembly. *Phys. Chem. Chem. Phys.* **2016**, *18*, 1265−1278.
(177) Smith, C. S.; Álvarez, Z.; Qiu, R.; Sasselli, I. R.; Clemons, T.; Ortega, J. A.; Vilela-Picos, M.; Wellman, H.; Kiskinis, E.; Stupp, S. I. Enhanced Neuron Growth and Electrical Activity by a Supramolecular Netrin-1 Mimetic Nanofiber. *ACS Nano* **2023**, *17*, 19887−19902.
(178) Li, C.; Iscen, A.; Sai, H.; Sato, K.; Sather, N. A.; Chin, S. M.; Álvarez, Z.; Palmer, L. C.; Schatz, G. C.; Stupp, S. I. Supramolecular-covalent hybrid polymers for light-activated mechanical actuation. *Nat. Mater.* **2020**, *19*, 900−909.
(179) López, C. A.; De Vries, A. H.; Marrink, S. J. Computational microscopy of cyclodextrin mediated cholesterol extraction from lipid model membranes. *Sci. Rep.* **2013**, *3*, 2071.
(180) Schroer, C. F.; Baldauf, L.; van Buren, L.; Wassenaar, T. A.; Melo, M. N.; Koenderink, G. H.; Marrink, S. J. Charge-dependent interactions of monomeric and filamentous actin with lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 5861−5872.
(181) Arnarez, C.; Marrink, S.; Periole, X. Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. *Chem. Sci.* **2016**, *7*, 4435−4443.
(182) Su, J.; Marrink, S. J.; Melo, M. N. Localization Preference of Antimicrobial Peptides on Liquid-Disordered Membrane Domains. *Front. Cell Dev. Biol.* **2020**, *8*, 350.
(183) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(184) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17.
(185) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(186) Deng, L.; Zhou, P.; Zhao, Y.; Wang, Y.; Xu, H. Molecular origin of the self-assembled morphological difference caused by varying the order of charged residues in short peptides. *J. Phys. Chem. B* **2014**, *118*, 12501−12510.
(187) Jonnalagadda, S. V. R.; Ornithopoulou, E.; Orr, A. A.; Mossou, E.; Forsyth, V. T.; Mitchell, E. P.; Bowler, M. W.; Mitraki, A.; Tamamis, P. Computational design of amyloid self-assembling peptides bearing aromatic residues and the cell adhesive motif Arg-Gly-Asp. *Mol. Syst. Des. Eng.* **2017**, *2*, 321−335.
(188) Yu, T.; Schatz, G. C. Free Energy Profile and Mechanism of Self-Assembly of Peptide Amphiphiles Based on a Collective Assembly Coordinate. *J. Phys. Chem. B* **2013**, *117*, 9004−9013.
(189) Vijayaraj, R.; Sundar Raman, S.; Mahesh Kumar, R.; Subramanian, V. Studies on the structure and stability of cyclic peptide based nanotubes using oligomeric approach: a computational chemistry investigation. *J. Phys. Chem. B* **2010**, *114*, 16574−16583.
(190) Liu, K.; Kang, Y.; Ma, G.; Möhwald, H.; Yan, X. Molecular and mesoscale mechanism for hierarchical self-assembly of dipeptide and porphyrin light-harvesting system. *Phys. Chem. Chem. Phys.* **2016**, *18*, 16738−16747.
(191) Ji, W.; Yuan, C.; Chakraborty, P.; Makam, P.; Bera, S.; Rencus-Lazar, S.; Li, J.; Yan, X.; Gazit, E. Coassembly-induced transformation of dipeptide amyloid-like structures into stimuli-responsive supramolecular materials. *ACS Nano* **2020**, *14*, 7181−7190.
(192) Song, R.; Wu, X.; Xue, B.; Yang, Y.; Huang, W.; Zeng, G.; Wang, J.; Li, W.; Cao, Y.; Wang, W.; et al. Principles governing catalytic activity of self-assembled short peptides. *J. Am. Chem. Soc.* **2019**, *141*, 223−231.
(193) Ye, X. W.; Tian, W.; Han, L.; Li, Y. J.; Liu, S.; Lai, W. J.; Liu, Y. X.; Wang, L.; Yang, P. P.; Wang, H. High-Throughput Screening of pH-Dependent β-sheet Self-Assembling Peptide. *Small* **2024**, *20*, 2307963.
(194) Slyngborg, M.; Fojan, P. A computational study of the self-assembly of the RFFFR peptide. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30023−30036.
(195) Rad-Malekshahi, M.; Visscher, K. M.; Rodrigues, J. P.; De Vries, R.; Hennink, W. E.; Baldus, M.; Bonvin, A. M.; Mastrobattista, E.; Weingarth, M. The supramolecular organization of a peptide-based nanocarrier at high molecular detail. *J. Am. Chem. Soc.* **2015**, *137*, 7775−7784.
(196) Deshmukh, S. A.; Solomon, L. A.; Kamath, G.; Fry, H. C.; Sankaranarayanan, S. K. Water ordering controls the dynamic equilibrium of micelle-fibre formation in self-assembly of peptide amphiphiles. *Nat. Commun.* **2016**, *7*, 12367.
(197) Urago, H.; Suga, T.; Hirata, T.; Kodama, H.; Unno, M. Raman optical activity of a cyclic dipeptide analyzed by quantum chemical calculations combined with molecular dynamics simulations. *J. Phys. Chem. B* **2014**, *118*, 6767−6774.
(198) Cai, X.; Han, W. Development of a Hybrid-Resolution Force Field for Peptide Self-Assembly Simulations: Optimizing Peptide-Peptide and Peptide-Solvent Interactions. *J. Chem. Inf. Model.* **2022**, *62*, 2744−2760.
(199) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(200) DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82−92.
(201) Saracino, G. A.; Fontana, F.; Jekhmane, S.; Silva, J. M.; Weingarth, M.; Gelain, F. Elucidating self-assembling peptide

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273

aggregation via Morphoscanner: a new tool for protein-peptide structural characterization. *Adv. Sci.* **2018**, *5*, 1800471.
(202) Marciano, Y.; Nayeem, N.; Dave, D.; Ulijn, R. V.; Contel, M. N-Acetylation of Biodegradable Supramolecular Peptide Nanofilaments Selectively Enhances Their Proteolytic Stability for Targeted Delivery of Gold-Based Anticancer Agents. *ACS Biomater. Sci. Eng.* **2023**, *9*, 3379−3389.
(203) Chakraborty, P.; Oved, H.; Bychenko, D.; Yao, Y.; Tang, Y.; Zilberzwige-Tal, S.; Wei, G.; Dvir, T.; Gazit, E. Nanoengineered Peptide-Based Antimicrobial Conductive Supramolecular Biomaterial for Cardiac Tissue Engineering. *Adv. Mater.* **2021**, *33*, 2008715.
(204) Hiew, S. H.; Guerette, P. A.; Zvarec, O. J.; Phillips, M.; Zhou, F.; Su, H.; Pervushin, K.; Orner, B. P.; Miserez, A. Modular peptides from the thermoplastic squid sucker ring teeth form amyloid-like cross-β supramolecular networks. *Acta Biomater.* **2016**, *46*, 41−54.
(205) Ortony, J. H.; Qiao, B.; Newcomb, C. J.; Keller, T. J.; Palmer, L. C.; Deiss-Yehiely, E.; Olvera de la Cruz, M.; Han, S.; Stupp, S. I. Water Dynamics from the Surface to the Interior of a Supramolecular Nanostructure. *J. Am. Chem. Soc.* **2017**, *139*, 8915−8921.
(206) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(207) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millan, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; J, R. R.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871−876.
(208) Magi Meconi, G.; Sasselli, I. R.; Bianco, V.; Onuchic, J.; Coluzza, I. Key aspects of the past 30 Years of protein design. *Rep. Prog. Phys.* **2022**, *85*, 086601.
(209) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.* **1995**, *24*, 279−287.
(210) Hansch, C.; Leo, A.; Hoekman, D. H. *Exploring QSAR*. American Chemical Society: 1995.
(211) Yousefinejad, S.; Hemmateenejad, B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 177−204.
(212) Thurston, B. A.; Ferguson, A. L. Machine learning and molecular design of self-assembling-conjugated oligopeptides. *Mol. Simul.* **2018**, *44*, 930−945.
(213) Zhao, Y.; Mulder, R. J.; Houshyar, S.; Le, T. C. A review on the application of molecular descriptors and machine learning in polymer design. *Polym. Chem.* **2023**, *14*, 3325−3346.
(214) Xu, T.; Wang, J.; Zhao, S.; Chen, D.; Zhang, H.; Fang, Y.; Kong, N.; Zhou, Z.; Li, W.; Wang, H. Accelerating the prediction and discovery of peptide hydrogels with human-in-the-loop. *Nat. Commun.* **2023**, *14*, 3880.
(215) Jones, S. J.; Perez, A. Molecular Modeling of Self-Assembling Peptides. *ACS Appl. Bio Mater.* **2024**, *7*, 543.
(216) van Teijlingen, A.; Tuttle, T. Beyond Tripeptides Two-Step Active Machine Learning for Very Large Data sets. *J. Chem. Theory Comput.* **2021**, *17*, 3221−3232.
(217) Shmilovich, K.; Mansbach, R. A.; Sidky, H.; Dunne, O. E.; Panda, S. S.; Tovar, J. D.; Ferguson, A. L. Discovery of self-assembling π-conjugated peptides by active learning-directed coarse-grained molecular simulation. *J. Phys. Chem. B* **2020**, *124*, 3873−3891.
(218) Xiao, X.; Robang, A. S.; Sarma, S.; Le, J. V.; Helmicki, M. E.; Lambert, M. J.; Guerrero-Ferreira, R.; Arboleda-Echavarria, J.; Paravastu, A. K.; Hall, C. K. Sequence patterns and signatures: Computational and experimental discovery of amyloid-forming peptides. *PNAS nexus* **2022**, *1*, pgac263.
(219) Batra, R.; Loeffler, T. D.; Chan, H.; Srinivasan, S.; Cui, H.; Korendovych, I. V.; Nanda, V.; Palmer, L. C.; Solomon, L. A.; Fry, H. C. Machine learning overcomes human bias in the discovery of self-assembling peptides. *Nat. Chem.* **2022**, *14*, 1427−1435.

https://doi.org/10.1021/acsomega.4c02628
ACS Omega 2024, 9, 31254−31273