UC Davis
UC Davis Previously Published Works

Title

Foldamer Architectures of Triazine-Based Sequence-Defined Polymers Investigated with Molecular Dynamics Simulations and Enhanced Sampling Methods

Permalink

https://escholarship.org/uc/item/4ch039p3

Journal

The Journal of Physical Chemistry B, 123(44)

ISSN

1520-6106

Authors

Ahn, Surl-Hee

Grate, Jay W

Publication Date

2019-11-07

DOI

10.1021/acs.jpcb.9b06067

Copyright Information

Peer reviewed

Foldamer Architectures of Triazine-Based Sequence-Defined Polymers Investigated with Molecular Dynamics Simulations and Enhanced Sampling Methods

Surl-Hee Ahn[*,†] and Jay W. Grate[*,‡]

†Department of Chemistry and Biochemistry, University of California San Diego, La Jolla

‡Pacific Northwest National Laboratory, Richland

E-mail: s3ahn@ucsd.edu; jwgrate@pnnl.gov

Phone: +1 (650)521-7646; +1 (509)371-6500

1

Abstract

Triazine-based sequence-defined polymers have recently been developed that are biomimetic and robust. In molecular dynamics (MD) simulations, the triazine polymers were shown to form linear nanorod foldamers through hydrogen bonding and π-π interactions. The nanorod foldamers have motifs resembling those of DNA, α-helices, and β-sheets, and have potential to be useful building blocks for new macromolecules and materials. To understand the formation of nanorod foldamers, we investigate how linker structures in the middle of the triazine polymers lead to folding using MD simulations. We found that a variety of linkers can participate in folding, but that specific linker structures are more favorable than others, depending on the polymer length. Folding of hexamers into well-defined nanorod foldamers was most favorable with pentanediamine and *ortho*-xylenediamine linkers in the center of the polymers. Foldamers with *ortho*-xylenediamine linkers in the center were investigated for longer polymers, i.e., octamers and decamers, using two different enhanced sampling methods, since regular MD simulations had failed to show any folding for these longer polymers. In particular, the recently developed concurrent adaptive sampling (CAS) algorithm and replica exchange molecular dynamics (REMD) were used. We found that the two enhanced sampling methods did lead to the observation of foldamers, and that REMD revealed new foldamer architectures where *cis-trans* isomerizations had occurred. Foldamer formation, diversity, and the strengths and limitations of simulation techniques are discussed. These findings provide new insights into the diversity of foldamer architectures for a new type of biomimetic synthetic polymer.

Introduction

Sequence-defined polymers, epitomized in nature by peptides and poly(nucleic acids), are macromolecules composed of a multiplicity of monomers that are sequenced into the structure in a pre-determined order. The resulting polymer also has a defined length. The monomers in a sequence-defined polymer are distinguished from one another by having a different side

chain; in peptides, the amino acids each have a particular side chain, while in poly(nucleic acids), the side chains are the nucleic acid bases. In nature, sequence-defined polymers create biomaterials, encode information, perform biocatalysis, participate in molecular recognition, and shuttle species across membranes. There has been a recent resurgence in interest in synthetic sequence-defined polymers, in the expectation that features such as functionality, sequence, monodispersity, folding, and assembly will lead to useful molecules and materials; this is evident in an edited volume,[1] a number of reviews and topical articles,[2–10] as well as numerous papers on new syntheses and structures.[11–30]

Among these, the triazine-based sequence-defined polymers reported in 2016 claim a number of biomimetic features, while posing challenges for molecular dynamics (MD) simulations.[11–13] These polymers consist of aromatic triazine rings connected together by linker sections that are typically derived from diamines, such as ethylenediamine, as shown in Figure 1. Each aromatic triazine ring in the backbone has a pendant side chain. The iterative submonomer synthesis approach enables sequencing monomers with different side chains, like the amino acids in peptides. With the triazine polymers, it is also possible to sequence linker sections with different structures into the polymer backbone. The lone pairs of the NH groups from the linker sections, adjacent to the aromatic rings, are delocalized with the aromatic π electrons. This produces a barrier to rotation in regular spots along the backbone. This can also be regarded as biomimetic to peptides, which have amide groups in the backbone presenting rotational barriers. These energetic barriers limit the ability of simple MD simulations from fully exploring conformational structures within feasible computational time frames.

To see the possible conformational behavior and assembly of triazine polymers, molecular dynamics (MD) simulations were used in the initial report of triazine polymers.[11] Regular MD simulations showed that dimers of trimers form nanorod foldamers that are held together by motifs of paired hydrogen bonds and π-π interactions, and replica exchange molecular dynamics (REMD)[31] simulations also revealed similar nanorod foldamers from hexamers,[11]
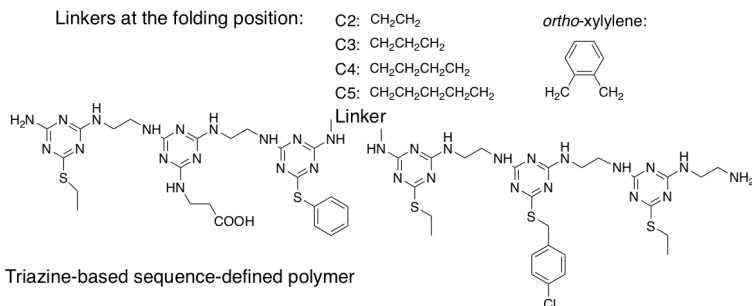
3

Figure 1: Triazine-based sequence-defined polymer example highlighting a linker in the central position. Linkers examined in simulations in this study, from ethylene(C2) to *ortho*-xylene, are shown. Side chains are varied to illustrate the sequence-defined nature of these polymers. In simulations in this study, however, homopolymers with ethyl side chains are used in order to focus on the role of backbone structure in folding.

which are shown in Figure 2a and Figure 2b. Hydrogen bonds drive conformation and assembly due to their to their complementarity, directionality, and strength,[14,32] and π-π interactions, which are weaker than hydrogen bonds, help stabilize self-assembled molecules and foldamers.[32–41] These structures seen in simulations have overall features and non-covalent interactions resembling those of DNA, peptide α-helices, and peptide β-sheets.[11]
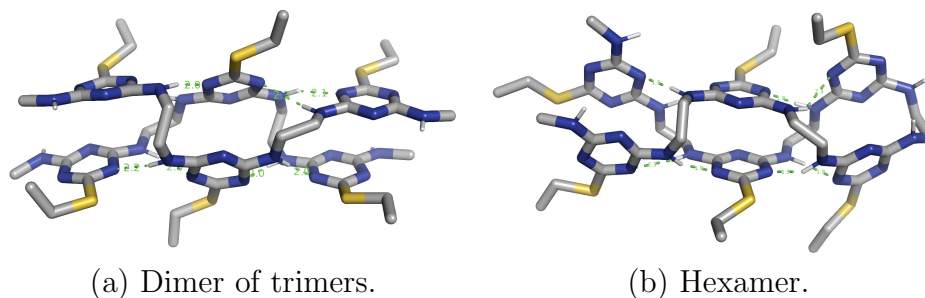

(a) Dimer of trimers. (b) Hexamer.

Figure 2: Nanorod structures seen from Ref. 11, which are held together by 8 hydrogen bonds and 3 π-π interactions. Both cases are shown with the same ethyl side chains here, but the hexamer foldamer had pentyl side chains in Ref. 11. The green dashed lines indicate hydrogen bonds and the numbers indicate their length. The depth cueing feature was used to add depth to the images, so some parts of the molecule are lighter than others.

In the specific structures seen, which we call nanorod foldamers, the side chains for each pair of π-bonded aromatic triazine rings project out opposite sides of the rod. The configuration of the bonds connecting the aromatic rings to the alpha amines were all *cis* along the length of the rod. The corresponding bonds at the fold site of folded hexamers were

trans. In this paper, foldamers with this specific structure, i.e., with regard to side chains and *cis* bond orientations and linear aspect, are called nanorods. Although these simulations used homohexamers (i.e., all monomers with the same side chain) and homotrimers, the observation of such foldamers suggests that in the future, sequence plus folding or assembly could create specific arrangements of side chain functional groups, just as sequence plus folding lead to functional group arrangements in peptide and protein conformations.

After the initial report on synthesis and simulation of triazine polymers, we focused on how the backbone structure and different side chains influenced the formation of nanorod foldamers in MD simulations.[13] Since MD simulations by themselves are limited in predictive power by routinely getting "stuck" in metastable states, enhanced sampling methods were used, namely the recently developed concurrent adaptive sampling (CAS) algorithm[12] and REMD. The CAS algorithm is a state-based enhanced sampling method that divides up the conformational space into small volume elements called "macrostates" and runs short simulations between the macrostates to efficiently sample both thermodynamic and kinetic properties. REMD is a thermodynamic enhanced sampling method that biases the system with different temperatures to overcome high energy barriers and efficiently samples thermodynamic properties. From this study, we found that the backbone structure needs to have hydrogen bonding ability to form nanorod foldamers; π-π interactions alone were not sufficient to stabilize the nanorod foldamers. In addition, the side chains' hydrogen bonding ability or lack thereof did not matter. This main result, the importance of backbone-backbone hydrogen bonding interactions, can provide guidance for materials scientists designing molecules intended to fold or assemble in biomimetic fashion.

Having explored backbone structures and different side chains, we now explore structural aspects that may affect folding. Specifically, we focus on how the structure of the central linker of the molecule allows or promotes folding. To consider possible linker structures for the folding location, physical molecular models of a dimer of trimers were first constructed. The bonding locations coming off the end of a pair of interacting trimers,

which we might bridge with a linker to create a hexameric foldamer, have a directionality. Potential central linkers must have bond numbers and angles that can connect these locations. While ethylenediamine linkers at the fold position were seen to be possible from past simulations, these physical models suggested that pentanediamine linkers might also be favorable for the folding of hexamers. In an effort to promote such folding, we also considered the *ortho*-xylylenediamine linker, in which the *ortho* arrangement might bring the chains into proximity, while the spacing and tetrahedral angles of the benzylic amines would fit to the triazine rings on the nanorod end. We also modeled the structures using the Avogadro software,[42] starting with a nanorod foldamer formed from a dimer of trimers and connecting the ends with the proposed linkers to make a folded hexamer. These models confirmed that the pentanediamine and *ortho*-xylylenediamine linkers could make a fold and keep the nanorod in order. For this paper, as shown in Figure 1, we considered ethane(C2)-pentane(C5)diamine linkers as well as the *ortho*-xylylenediamine linker. In addition, the triazine polymers (technically just oligomers) with different lengths, ranging from tetramers to decamers, were tested with these linkers.

A series of regular MD simulations were run first, pre-arranging the *trans* bonds around the linker expected to serve as the folding point, while the other bonds were *cis*. Following these simulations, specific cases were further investigated using two enhanced sampling methods the CAS algorithm and REMD following Ref. 13. The CAS method with noncovalent interactions as the collective variables more widely explores conformational space for the starting arrangement of cis and trans bonding. REMD, on the other hand, does allow cis and trans isomerizations to occur within the simulation, and a greater diversity of foldamer architectures arise. In the course of these studies, we found new ways that foldamers can form, with different *cis* and *trans* arrangements, or folding sections that include a triazine monomer within the fold.

Methods

Concurrent adaptive sampling algorithm

The CAS algorithm is a recently developed enhanced sampling method[12] that is based on the weighted ensemble (WE) method,[43] which has been widely used to obtain thermodynamic and kinetic properties of various systems.[44–47] Details of the CAS algorithm can be found in Ref. 12 and details of the WE method can be found in Ref. 43. Essentially, the CAS algorithm is an adaptive version of the WE method that has additional features, which makes the WE method more powerful. In the CAS algorithm/WE method, many simulations or "walkers" are simultaneously run for a short period of time τ and they carry probabilities or "weights" w. After each run, the walkers are binned to "macrostates," or small volume elements of the conformational space, that is defined by the reaction coordinates, which can be non-differentiable, chosen for the system, e.g., dihedral angles and number of non-covalent interactions. Then the walkers are "resampled" or go through an unbiased process that maintains a certain target number number of walkers n_w for each macrostate in a statistically correct way. This way, the number of walkers decreases in low energy regions where sampling is easy and increases in high energy regions where sampling is difficult. How resampling is carried out in the CAS algorithm is slightly different from the original WE method's implementation: the CAS algorithm produces walkers with equal mean weights within each macrostate, which reduces statistical errors, compared to having walkers with varying weights. As a result, sampling becomes more uniform irrespective of the energy barrier heights. Since the process does not add any statistical bias to the system, we can obtain thermodynamic and kinetic properties directly from the simulation.

The CAS algorithm also has unique features that can speed up the sampling process. For instance, it can have adaptive macrostates that are newly created at every simulation step, which can be useful when exploring conformational spaces that would have an intractable number of macrostates if pre-defined. Additionally, the method can have more than one or

two reaction coordinates, which is the case for most enhanced sampling methods, without incurring exponentially increasing computational cost. This is because the macrostates are essentially n-dimensional spheres, where n denotes the number of reaction coordinates, and be fairly large in a high dimensional conformational space. Finally, the CAS algorithm can focus its computational effort in sampling the slowest non-stationary process in a system that has two major metastable states by using a technique denoted as "spectral clustering." It uses the committor function that is approximated by the normalized second dominant eigenvector from the transition matrix between macrostates and importance sampling. By doing so, the CAS algorithm keeps macrostates along the slowest pathways while depleting macrostates orthogonal to the slowest pathways.

To use the CAS algorithm in practice, a few parameters need to be set beforehand. First, the reaction coordinates need to be chosen, which should be able to describe the various sampled conformations and progress from reactant to product if those states are defined. Note that the selection of these reaction coordinates may influence which pathways are sampled and hence which structures are seen or not seen in the final results. Second, the simulation length τ needs be set to an appropriate length so that transitions are not inadvertently missed (if too long) and the computational cost is not too high from frequent resampling (if too short). Fortunately, we do not need to worry about setting τ so that transitions between macrostates are Markovian. Each walker keeps track of its past history so statistically unbiased estimates of kinetics can be directly obtained, regardless of the length of τ. Third, the target number of walkers per macrostate n_w needs to be high enough so that visited macrostates are constantly sampled throughout the simulation. Finally, macrostates need to be either pre-defined before the simulation or set to be adaptive so that new macrostates are created at every simulation step. The radius or size of the macrostates also needs to be chosen appropriately so that regions with different energy barrier heights can be effectively sampled.

Replica exchange molecular dynamics

REMD is a well-established enhanced sampling method[31] that has been widely used to obtain thermodynamic properties of various systems.[48–52] Details of REMD can be found in Ref. 31. Essentially, REMD simultaneously runs many simulations or "replicas" at different temperatures and exchanges the replicas periodically with a Metropolis criterion. The periodic exchanges occur so that the low temperature replicas can access high temperature replicas and vice versa, overcome energy barriers, and efficiently sample the conformational space. After the exchanges, momenta are rescaled to maintain the temperature condition for each replica and equilibrium canonical (NVT) ensemble for each temperature. Combined with the Metropolis criterion for the exchange probability, REMD ends up maintaining detailed balance for an extended ensemble of canonical states.

However, since bias, or temperature in this case, is added to the system, the weighted histogram analysis method (WHAM)[53] or multistate Bennett acceptance ratio (MBAR)[54] needs to be used during post-processing to obtain the actual free energy landscape of the system. In addition, since the system's kinetics are altered from the process, kinetics can only be obtained approximately after post-processing and constructing a master equation.[55,56] Nonetheless, REMD is effective at overcoming enthalpic barriers and sampling thermodynamic properties of the system. Additionally, there is no need to choose reaction coordinates beforehand, which allows the method to be applied to new and unfamiliar systems.

To use REMD in practice, a few parameters need to be set beforehand. First, the exchange probability needs to be high enough to mix low and high temperature replicas well. The exchange probability p between replica i and replica j is given by

$$p = \min\left(1, e^{(E_i - E_j)\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)}\right) \tag{1}$$

where k_B represents the Boltzmann constant, E represents energy, and T represents temperature. As seen from Eq. 1, the difference between the energies and/or the temperatures

9

need to be small in order for p to be high. Hence, the number of replicas and temperature range need to be set well so that we have the right spacing between replicas. Having uniform exchange probabilities for all replicas is a good indicator that we have an optimal temperature distribution. Finally, the number of simulation steps between exchanges needs to be long enough so that each replica has time to equilibrate.

Simulation protocol

The triazine polymers were all simulated with GROMACS 4.6.4[57] at temperature $T = 300$ K with time step $\Delta t = 2$ fs. Most simulation parameters were identical to the ones in Ref. 11 and Ref. 13. The simulations were in the non-periodic canonical (NVT) ensemble with velocity-rescale temperature coupling (0.2 ps coupling time).[58] The generalized Amber force field (GAFF)[59] and the programs AnteChamber PYthon Parser interfacE (ACPYPE)[60] and Antechamber[61] were used to generate parameters and topologies for the triazine polymers. The triazine polymers' initial conformations were constructed to be all *cis* and extended, except for the linkers at the fold position that were *trans* on both sides, using an in-house Python script used in Ref. 11. The Generalized Born/Surface Area implicit solvent with the Onufriev/Bashford/Case algorithm[62] for calculating Born radii, solvent dielectric constant of 78.3, and infinite van der Waals and Coulomb cutoffs were used. The hydrophobic solvent accessible surface area was calculated using an analytical continuum electrostatics (ACE)-type approximation[28] and the internal dielectric constant was set to 1,[63] which were the default settings on GROMACS.

The particular force field GAFF and implicit solvent were used since Ref. 64 reported that they were accurate at reproducing the torsional landscape and major conformers of peptoids, which are similar to triazine polymers, seen from quantum mechanical calculations and experiments. Additionally, GAFF was shown to have π-π interaction energy and equilibrium distance comparable to those from quantum mechanical calculations and reliable in simulating long-range structures for liquid benzene, which is the prototypical system for π-π

interactions.[65] GAFF was also shown to be reliable at reproducing parallel π-π interactions, which were the only type of π-π interactions that appeared in all of our simulations, at small molecular separation less than 5 $\mathring{A}$ for liquid benzene, which is similar to our protocol of counting ones less than 4.2 $\mathring{A}$, as noted later in the Data Analysis section. Other works that involve systems with π-π interactions have used GAFF.[66,67] In this paper, implicit solvent was used consistently across all of the simulation methods used and to stay consistent with simulations of single folding macromolecules in our previous work.[11–13] REMD is known to scale poorly for explicit solvent systems, so consistent use of implicit solvent allowed direct comparison among regular MD, the CAS algorithm, and REMD simulations.

The CAS algorithm simulations were run with conformations sampled every 100 ps from a 100 ns long regular MD simulation that started with the all cis and extended conformation. To be consistent with the 100 ns simulation trajectory, the CAS algorithm simulation had each conformation transition into its next conformation that occurred after 100 ps and resampled for the very first step and then regular CAS algorithm simulation steps were taken. The total simulation time was calculated by the cumulative number of macrostates $\times$ target number of walkers per macrostate n_w $\times$ simulation time τ. The target number of walkers per macrostate n_w was set to 10, and the simulation time τ was set to 100 ps so that the triazine polymer had sufficient time to transition into another conformation. The reaction coordinates were the number of hydrogen bonds and the number of π-π interactions, which could easily indicate whether the triazine polymer was a nanorod foldamer or not. The macrostates were pre-defined and fixed throughout the simulation and had centers with integer values with a radius of 0.5.

Following Ref. 64, which used REMD from 300 to 800 K to obtain free energy landscapes of peptoids, as a reference, the REMD simulations were run with 16 replicas that uniformly spanned from 300 to 800 K for each system. The highest temperature of 800 K is used to enable and observe cis to $trans$ and vice versa isomerizations (a single isomerization is experimentally measured to have an energy barrier of $\Delta G^{\ddagger} = 15$ kcal/mol),[14,68,69] as done

in Ref. 64. The system was stable at the highest temperature since numerical integration of the MD simulation was stable throughout the entire simulation. The number of simulation steps between exchanges was set to 1000 (2 ps), which resulted in an exchange probability around 50% for each replica. According to Ref. 70, exchange probabilities that are greater than 25% do not affect sampling greatly. Before running the REMD simulations, the replicas were equilibrated for 200 ps at each temperature. Conformations and potential energies were saved every 2 ps.

Data analysis

For all of the simulations, the nanorod foldamers were detected by examining and noting the number of hydrogen bonds between the triazine rings and the rest of the triazine polymer, which were defined to be 2.5 $\mathring{A}$ or shorter in order to be counted, and the number of π-π interactions, which was detected by measuring the center of mass distance between the triazine rings and had to be 4.2 $\mathring{A}$ or shorter in order to be counted. Our previous work[11,13] had used a distance metric to count hydrogen bonds and π-π interactions and this protocol was again followed here. The protocol in counting π-π interactions is similar to the work where π-π interactions were counted for MD simulations of single-stranded DNA binding to graphene oxide,[71] which in turn followed the work where π-π interactions were counted for the same system.[72] Other work in bioinfomatics had used the distance metric to count π-π interactions.[73] On the other hand, there are other works with MD simulations and molecular docking that had used an additional metric along with the distance metric, i.e., measuring the angle between the two aromatic ring planes.[74,75] Having counted π-π interactions without this additional criterion, our results might have overestimated the number of π-π interactions present in the MD simulations. That being said, in our prior paper,[13] we showed that the hydrogen bonding interactions played the key role in assembly, not π-π interactions. Without hydrogen bonding interactions, π-π interactions are not strong enough to keep foldamers together and are not the key driving force.

For the CAS algorithm and REMD simulations with even-numbered triazine polymers (octamers, decamers), only the distances between the triazine rings that should be paired up to form a full linear nanorod foldamer were measured, e.g., for the octamers, the distance between the first and the eighth triazine rings, the distance between the second and the seventh triazine rings, the distance between the third and the sixth triazine rings, and the distance between the fourth and the fifth triazine rings were measured. If the triazine rings pair up otherwise, then there would be a shift and the triazine polymer would not form a full linear nanorod foldamer. Additionally, a triazine polymer would have to have a certain number of hydrogen bonds and π-π interactions to be considered as a nanorod foldamer and the numbers would depend on the polymer length, e.g., a tetramer nanorod foldamer would have at least 4 hydrogen bonds and 2 π-π interactions. Similarly, a hexamer nanorod foldamer would have 8 hydrogen bonds and 3 π-π interactions like the dimer of trimers in Figure 2a. See Table 1 for the number of π-π interactions and hydrogen bonds needed to form a nanorod foldamer for each triazine polymer depending on the polymer length.

Table 1: Number of π-π interactions and hydrogen bonds needed for nanorod foldamer formation

	No. of π-π interactions	No. of hydrogen bonds[a]
Tetramer	2	4-5
Pentamer	2	4-6
Hexamer	3	8-9
Heptamer	3	8-11
Octamer	4	12
Nonamer	4	12-13
Decamer	5	16

[a] The lower bound indicates the minimum number of hydrogen bonds needed. This column ended up being a range for most triazine polymers after counting the number of hydrogen bonds for the observed nanorod foldamers in MD simulations.

The regular MD simulations were run for 1 μs, unless otherwise noted. Conformations were sampled every 100 ps and the first 100 ns were thrown out (equilibrium time). The CAS algorithm and REMD simulations were run for 2 μs and the first 100 ns were thrown out as well. The conformations and potential energies were sampled every 100 ps in the

REMD simulations to produce the free energy landscapes using the multistate Bennett acceptance ratio (MBAR), specifically the Python implementation by Shirts and Chodera,[54] to be consistent with the CAS algorithm simulations that had a simulation time of $\tau = 100$ ps. But for convergence analysis, the finest replica data that were produced from sampling every 2 ps were used to calculate the average number of round-trips in a given observation time per replica.

To identify major conformations for each triazine polymer from the regular MD simulations and REMD simulations, we used k-means clustering[76] using the R program[77] as done in Ref. 11. Specifically, with the conformations that were sampled every 100 ps from the last 900 ns, we first computed a matrix with entries that indicated the root mean square deviation (RMSD) of the triazine rings' position with respect to each other. Then the matrix was used as input for an in-house R script used in Ref. 11 that identifies k, or the number of clusters, for the particular triazine polymer. The optimal k was found by identifying the "knee," or the point where the slope greatly decreases, in the within-group sum of squares vs. k plot. As stated in Ref. 11, clusters with small internal RMSDs likely represent stable, folded conformations, whereas clusters with big internal RMSDs likely represent transient, unfolded conformations. Hence, the fraction of the total population that is in a "compact" cluster with a small internal RMSD gives an estimate of how much the triazine polymer is likely to form stable, folded conformations.

Results

Regular molecular dynamics simulation

As stated in the Introduction, several triazine polymers with different lengths, ranging from tetramers to decamers, with four different aliphatic linkers, namely ethylenediamine, propanediamine, butanediamine, and pentanediamine, and an *ortho*-xylylenediamine linker, as shown in Figure 1, were run for 1 μs in implicit solvent. For the present studies focus-

ing on backbone-backbone interactions and conformational structures, the triazine polymers had ethyl side chains on all monomers and were symmetrical end-to-end with methylamino terminal substituents (i.e., this means that the final monomer did not have a complete linker section dangling at the end). The percentage of nanorod foldamers in the total population from k-means clustering is listed in Table 2 for each triazine polymer. The same results are plotted in Figure 3 to visualize the results more easily. Surprisingly, the linkers leading to nanorod foldamers depended on the polymer length.

The nanorod foldamers, formed from the linkers leading to foldamers, are shown in Figure 4a, Figure 4b, Figure 4c, and Figure 4d for tetramers, pentamers, hexamers, and heptamers, respectively. First, note that all of the nanorod foldamers shown formed with the ends evenly aligned with paired triazine rings. Therefore, tetramers and hexamers have a single linker making the fold, whereas with pentamers and heptamers, the fold region consists of a triazine monomer plus the linker. The tetramer and pentamer each form a rod section with 2 pairs of π-π bonded triazine rings with side chains projecting toward different sides of the rod. Similarly, the hexamer and heptamer each form a rod section with 3 pairs of π-π bonded triazine rings with side chains projecting toward different sides of the rod. Thus, they conform to the nanorod structure as set out in the Introduction and are overall linear in shape.

For tetramers, the ethylenediamine, butanediamine, and pentanediamine linkers lead to high percentages of nanorod foldamers, i.e., ethylenediamine; 79.1 %, butanediamine; 69.0 %, and pentanediamine; 75.8 %. The *ortho*-xylylenediamine linker was not favorable for nanorod foldamers in this case. For the longer hexamers, the pentanediamine linker yielded the highest percentage of nanorod foldamers at 99 %, while the *ortho*-xylylenediamine linker resulted in 89 %. In these even-numbered cases, the fold occurs entirely with the central linker, and the triazine rings are paired in the nanorod architecture. The hexamer formed the expected well-defined nanorod foldamer. The prediction from physical models of hexamers, which was described in the Introduction, was that the pentanediamine and *ortho*-

xylylenediamine linkers would likely be compatible with folding, and this was confirmed for the hexamers. The design of the *ortho*-xylylenediamine to favor folding appears to be valid, but the pentanediamine linker is also quite favorable and perhaps advantageous.

For pentamers, the ethylenediamine linker plus a triazine monomer were most favorable (45.7 %), but the butanediamine plus a triazine monomer were also suitable for nanorod foldamers (36.8 %). The middle monomer in the fold of the pentamer with the enthylenediamine linker is able to participate in one additional hydrogen bond with an adjacent monomer ring along the chain, which may help stabilize the fold. For the longer heptamers, the *ortho*-xylylenediamine linker yielded the highest percentage of nanorod foldamers, albeit only 23.4 %, from regular MD simulations, and the folding region is also a triazine ring plus the *ortho*-xylylenediamine linker. In the heptamer, the main portion of the nanorod foldamer has 3 π-π interactions and 4 paired hydrogen bonds like the hexamer nanorod foldamer.

Beyond heptamers, nanorod foldamers according to the defined criteria and clustering methods were not found in the regular MD simulations. None of the linkers yielded folding for the octamers and decamers. For nonamers, the nanorod foldamer appeared at several points in the simulation with the *ortho*-xylylenediamine linker. However, it was not significant enough to be detected with k-means clustering, so the percentage of nanorod foldamers for the nonamer with *ortho*-xylylenediamine linker is shown as zero in Table 2. This nanorod foldamer conformation is shown in Figure 5. The fold region consists of the *ortho*-xylylenediamine linker as intended, such that the terminal monomers are not aligned as a π-π bonded pair. The odd monomer dangles beyond the organized nanorod, although it is able to form an additional hydrogen bond back to the prior monomer in the sequence. These nanorod foldamers were seen when the *ortho*-xylylenediamine linker was placed in the middle between the fourth and fifth triazine rings. The other fold tested was between the third and fourth triazine rings so that there would be three and six triazine rings on each side; nanorod foldamers were not observed at all in this case. The results suggest that folding is most likely when there are more or less equal number of triazine rings on both

16

sides of the folding region.

Since nanorod foldamers were not found for the octamers and decamers with any of the linkers, we extended the regular MD simulations of these cases up to 2 μs. However, nanorod foldamers were still not detected by our criteria. The regular MD simulations were biased to look for the nanorod architecture because the *cis* and *trans* bonds were pre-arranged to enable them. The nanorod architecture was indeed seen, but new diversity was also revealed. The fold location can involve a triazine ring in addition to the linker and the terminal ends may be aligned or not. For longer polymers, however, the regular MD simulations did not lead to nanorod foldamers as defined by our criteria within the extended times. This left a question unanswered, i.e., whether longer nanorods do not form, or alternatively, are simply not found with regular MD simulations but could form. Prior simulations had suggested that the dimers of pentamers, once formed, would be stable throughout the simulation.[11] Therefore, our efforts proceeded to use enhanced sampling methods for selected cases to better explore conformational space with regard to folding of longer structures.

Table 2: Percentage of nanorod foldamers in regular MD simulations

	Tetramer	Pentamer	Hexamer	Heptamer	Octamer	Nonamer	Decamer
Ethylenediamine	79.1	45.7	0.0	0.0	0.0	0.0	0.0
Propanediamine	0.0	0.0	46.7	15.4	0.0	0.0	0.0
Butanediamine	69.0	36.8	0.0	0.0	0.0	0.0	0.0
Pentanediamine	75.8	0.0	99.3	0.0	0.0	0.0	0.0
ortho-Xylylenediamine	0.0	0.0	88.9	23.4	0.0	0.0	0.0

Concurrent adaptive sampling algorithm

The octamer and decamer structures were further simulated with the *ortho*-xylylenediamine linker in the middle using the CAS algorithm. This linker had produced high percentages of nanorod foldamers for the hexamers, a significant percentage for the heptamers, and observable transient nanorod foldamers for the nonamers. In the latter case, it was the only linker that yielded a nanorod foldamer. Octamers and decamers were not observed to yield
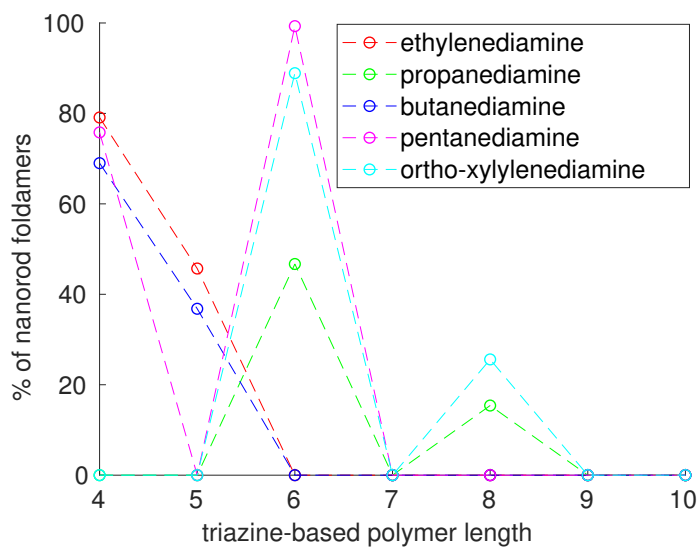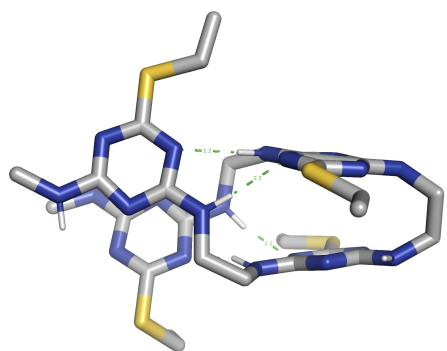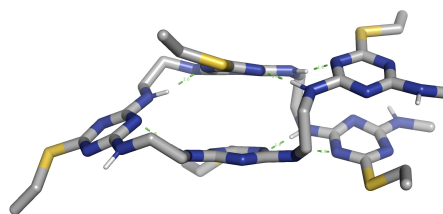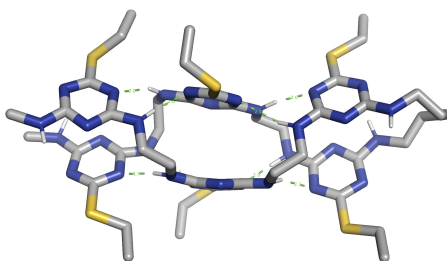
Figure 3: Percentage of nanorod foldamers in regular MD simulations
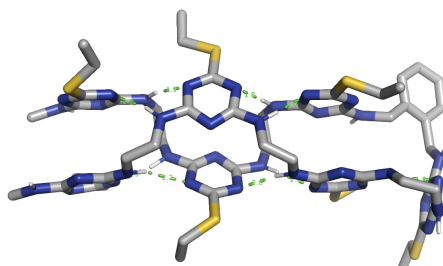

(a) Tetramer with ethylenediamine linker.

(b) Pentamer with ethylenediamine linker.

(c) Hexamer with pentanediamine linker.

(d) Heptamer with *ortho*-xylylenediamine linker.

Figure 4: Same as Figure 2 but from each polymer with its most effective linker in regular MD simulations.
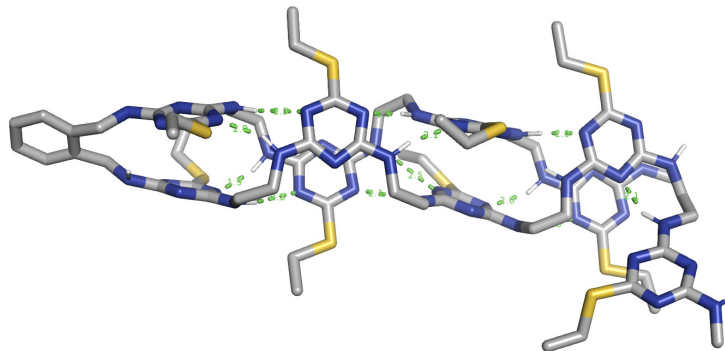
Figure 5: Same as Figure 2 but from nonamer with *ortho*-xylylenediamine linker in the middle in regular MD simulation.

nanorod foldamers in regular MD simulations.

As stated in the Methods section, the reaction coordinates were the number of hydrogen bonds and the number of π-π interactions. The starting structures were extended conformations with all *cis* bonds except around the linker in the middle folding position, just as in regular MD simulations. Dihedral angles were not followed as reaction coordinates in these simulations; therefore, pathways involving energetically unfavorable *cis-trans* isomerizations were unlikely to be sampled. The CAS algorithm thus serves as a method to further explore nanorod foldamers and other structures with these bond configurations but does not explore all of the possible conformational space. Since the free energy landscape, defined by the selected non-covalent interactions as reaction coordinates, was two-dimensional with a small range for both reaction coordinates, the macrostates were pre-defined and fixed throughout the simulation. Hence, the unique features of the CAS algorithm, such as having adaptive macrostates and spectral clustering, were not used, and the CAS algorithm simulations essentially became regular WE method simulations.

After running the octamer with the *ortho*-xylylenediamine linker for 825 ns, the nanorod foldamer with 4 π-π interactions and 12 hydrogen bonds first appeared, which is shown in Figure 6. This is the expected foldamer, i.e., the nanorod, with the side chains of π-π bonded aromatic rings projecting to opposite sides of the rod. The nanorod foldamer was constantly sampled throughout the 2 μs simulation. In addition, even though the other conformations

did not strictly have 4 π-π interactions and 12 hydrogen bonds with our criteria, some of the conformations clearly had the nanorod structure, i.e., conformations with 0 π-π interactions and 11 or more hydrogen bonds, 1 π-π interaction and 10 or more hydrogen bonds, 2-3 π-π interactions and 7 or more hydrogen bonds, and 4 π-π interactions and 10 or more hydrogen bonds. The rest of the conformations were globular. The representative conformations with various numbers of π-π interactions and hydrogen bonds are provided in the SI as PSE files that can be viewed with PyMOL.[78] Combining the probabilities of all of the conformations that corresponded to the nanorod structure, the probability of observing the nanorod foldamer is 7.96×10^{-3} % (the probability of observing the conformation with 4 π-π interactions and 12 hydrogen bonds is 5.68×10^{-8} %). Hence, observing the nanorod foldamer is not very likely, but the CAS algorithm was able to successfully sample this rare conformation.

The free energy landscape of the octamer, as defined by the reaction coordinates used, from the CAS algorithm simulation is shown in Figure 7a. The conformers in the "blue" region have higher probabilities of being observed and thus have lower relative free energies, whereas the conformers in the "red" region have lower probabilities of being observed and thus have higher relative free energies. To check for convergence, the relative standard error for each macrostate was calculated, which ranged from 0.88 to 19.71 %. In addition, since the transition matrices between macrostates were calculated at every simulation step, the equilibrium weights were also obtained, which is shown in Figure 7b. Comparing the two free energy landscapes, the two were equivalent. Having low relative standard errors and matching weights to equilibrium weights indicated that the CAS algorithm simulation had well-converged. In addition, the CAS algorithm simulation sampled much more widely compared to regular MD simulation, which is shown in Figure 8.

Although the CAS algorithm simulation sampled the nanorod foldamer for the octamer, it did not do so with the decamer. As shown in Figure 9a and Figure 9b, both the CAS algorithm simulation and regular MD simulation failed to explore much farther than con-

formations that had 1 π-π interaction during 2 μs of simulation time. The CAS algorithm simulation was slightly more successful at exploring the free energy landscape compared to the regular MD simulation, however. The observed conformations were all globular, however, and did not have obvious signs that they were on their way to form nanorod foldamers.

The results show that the CAS algorithm *can* sample the well-defined nanorod foldamer for the octamer with the *ortho*-xylylenediamine linker. Thus, a nanorod foldamer can form from the extended conformation containing pre-defined *cis* and *trans* bond arrangements from a longer macromolecule. The CAS algorithm is more effective than regular MD simulations at sampling this free energy landscape.
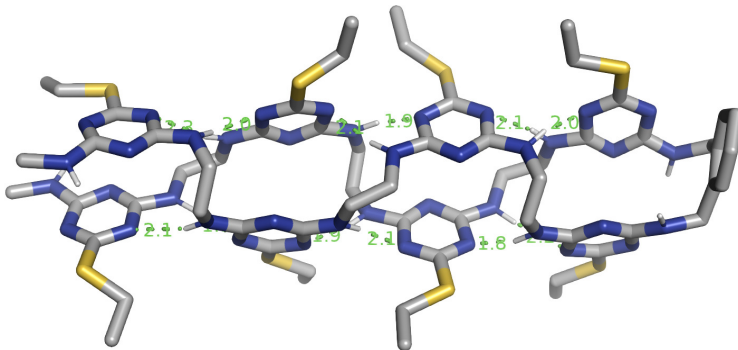
Figure 6: Same as Figure 2 but from the octamer with *ortho*-xylylenediamine linker in the CAS algorithm simulation.

Replica exchange molecular dynamics

Simulations using REMD were carried out on the octamer and decamer with the *ortho*-xylylenediamine linker, as we did for the CAS algorithm simulations. When using REMD, *cis-trans* isomerizations can occur, so a larger conformational space can be potentially sampled. After running the REMD simulation for 1 μs, a new octamer foldamer was found that had 4 π-π interactions and 7-12 hydrogen bonds according to our original criteria. We call this the "zigzag" foldamer, which is shown in Figure 10. The "zigzag" foldamer has an extensive rearrangement of *cis* and *trans* configurations, and the creation of an orderly *cis-trans* between every pair of aromatic rings along the foldamer. After extending the REMD simu-
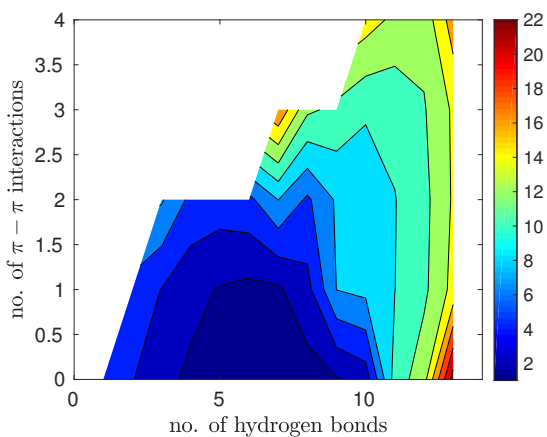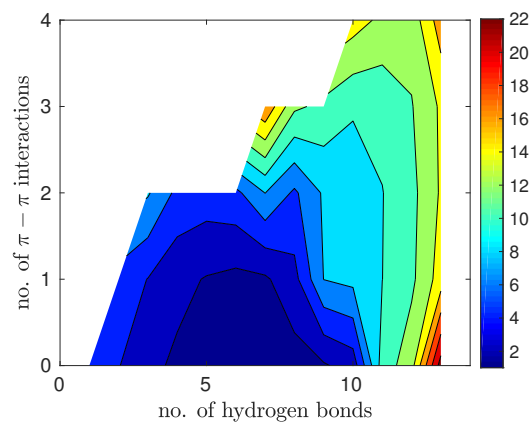
(a) From the CAS algorithm simulation. (b) Equilibrium weights from transition matrix.

Figure 7: Free energy landscapes of the octamer with *ortho*-xylylenediamine linker. The conformations are colored according to their relative free energy values or $-k_B T \ln(P)$ (kcal/mol), where k_B denotes the Boltzmann constant, T denotes the temperature ($T = 300\ K$), and P denotes the probability. Hence, the "blue" regions represent more probable conformations, whereas the "red" regions represent less probable conformations.
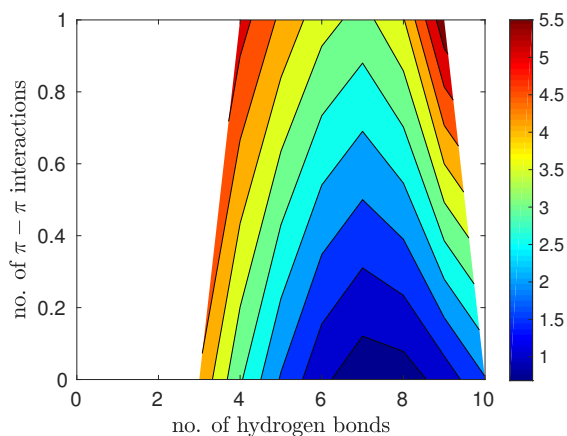
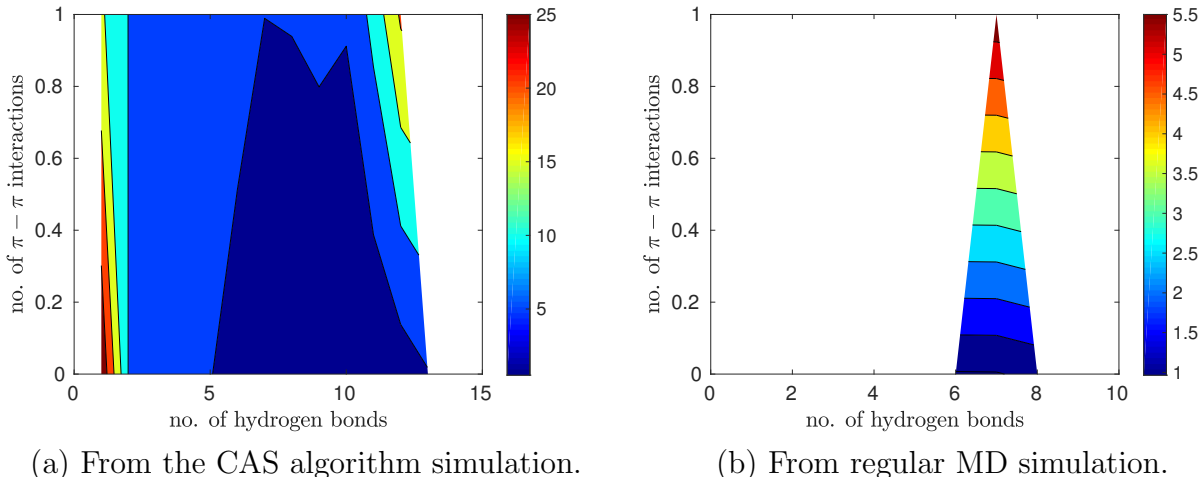Figure 8: Same as Figure 7 but for the octamer with *ortho*-xylylenediamine linker from regular MD simulation.

22

(a) From the CAS algorithm simulation.

(b) From regular MD simulation.

Figure 9: Same as Figure 7 but for the decamer with *ortho*-xylylenediamine linker.

lation up to 2 μs, another new octamer foldamer was found that had 4 π-π interactions and 6-11 hydrogen bonds according to our original criteria. Due to a shift from the linear axis, we call this the "staircase" foldamer, which is shown in Figure 11. However, the well-defined nanorod foldamer did not appear in the REMD simulation; the conformation that had 4 π-π interactions and 12 hydrogen bonds, which appeared only once after extending the REMD simulation up to 2 μs, corresponded to the "zigzag" foldamer. The closest octamer foldamer to the nanorod foldamer had 4 π-π interactions and 10-11 hydrogen bonds after extending the REMD simulation up to 2 μs, which is shown in Figure 12.

As for the decamer, REMD was able to sample conformations that had up to 4 π-π interactions but not 5. Most of the conformations that had 3-4 π-π interactions were foldamers that were partly "zigzag" as shown in Figure 13. The rest of the conformations were globular. As done for the CAS algorithm simulation of the octamer, the representative conformations from the REMD simulations with various numbers of π-π interactions and hydrogen bonds are provided in the SI as PSE files that can be viewed with PyMOL.

From k-means clustering, the "zigzag" foldamer was the only foldamer mentioned previously that was detected for the octamer and comprised 20.08 % of the total population. Only globular structures were detected for the decamer from k-means clustering. Hence, the

rest of the foldamers were not major conformations in the REMD simulations. In addition, a foldamer that had a nanorod foldamer as part of its structure was detected for the octamer through k-means clustering and comprised 53.87 % of the total population, which is shown in Figure 14. Since, as previously mentioned in Methods/Data Analysis, only the distances between the triazine rings that should be paired up to form a full linear nanorod foldamer were measured for the octamers and decamers, so these foldamers were indicated to have 0 π-π interactions, even though they clearly have 3 π-π interactions. Hence, the free energy landscapes under-reported π-π interactions since the "expected ones" were only counted.

One common and noticeable feature of these foldamers was that they were all non-linear, unlike the previously observed nanorod foldamers that were linear. This is due to the triazine polymers having *cis* to *trans* and vice versa isomerizations occurring to them with the temperature change in the REMD simulations. After temperature is added to the system, the octamer is able to overcome energy barriers going from *cis* to *trans* and vice versa. In the CAS algorithm simulations, we did not keep track of dihedral angles or any other reaction coordinates that indicated *cis* to *trans* and vice versa isomerizations. Hence, we did not observe any conformations with *cis* to *trans* and vice versa isomerizations and only observed linear nanorod foldamers. Additionally, the new non-linear foldamers had side chains on the same side for any given pair of π-π bonded aromatic rings, whereas the linear nanorod foldamers had side chains on opposite sides.

To compare the REMD simulation findings with those of the CAS algorithm and regular MD simulation, we plotted the free energy landscapes along the two reaction coordinates (number of hydrogen bonds and number of π-π interactions), which are shown in Figure 15a and Figure 15b. Compared to Figure 7a and Figure 9a, the free energy landscapes from the REMD simulations are vastly different since the CAS algorithm simulations explored the free energy landscape by probing the two reaction coordinates, whereas the REMD simulations explored the free energy landscapes more uniformly with different temperatures. To check for convergence, we calculated the average number of round-trips in a given observation time

per replica, following Ref. 79 and Ref. 80. A round-trip indicates a replica visiting both the lowest and the highest temperature states. The given observation time was set to be 10 ns and the average number of round-trips was measured as simulation time increased. For error bars, the standard deviation of the number of round-trips was multiplied by 2, which approximately represents a 95% confidence interval. As seen in Figure 16a and Figure 16b, the average numbers of round-trips converge to stable values with small error bars as the simulation time increases, indicating that the REMD simulations had converged.
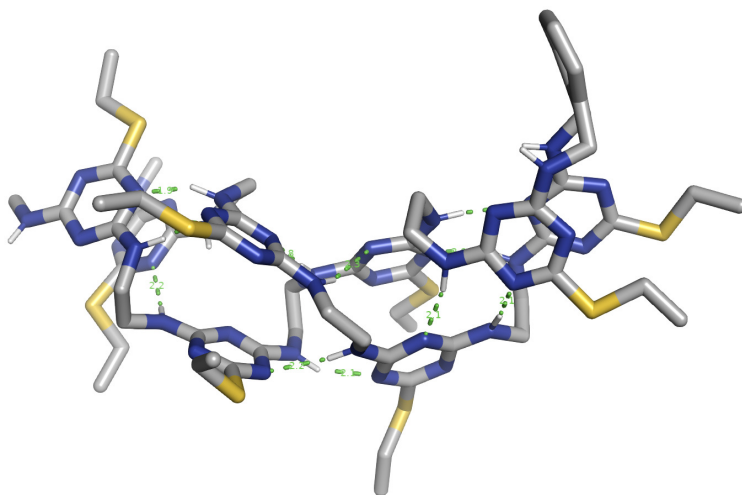
Figure 10: Same as Figure 2 but from the octamer with *ortho*-xylylenediamine linker in the REMD simulation. Denoted as "zigzag" foldamer.
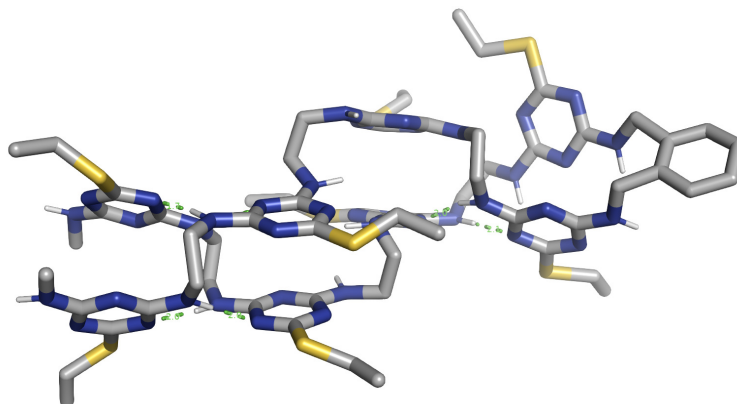

Figure 11: Same as Figure 2 but from the octamer with *ortho*-xylylenediamine linker in the REMD simulation. Denoted as "staircase" foldamer.
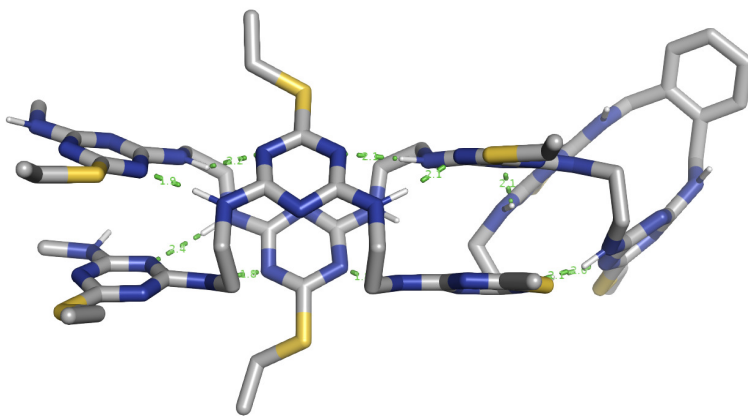
Figure 12: Same as Figure 2 but from the octamer with *ortho*-xylylenediamine linker in the REMD simulation. Closest foldamer to the nanorod foldamer.
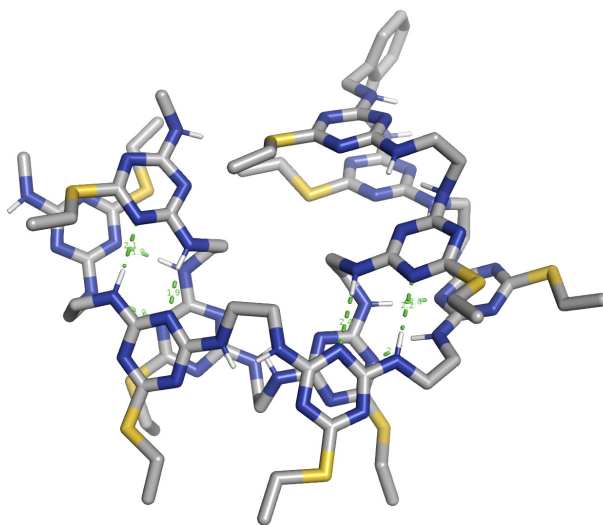

Figure 13: Same as Figure 2 but from the decamer with *ortho*-xylylenediamine linker in the REMD simulation. Foldamer that is partly "zigzag."
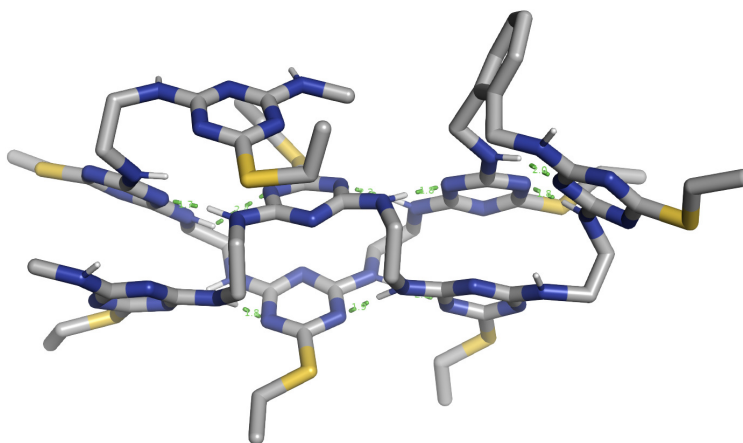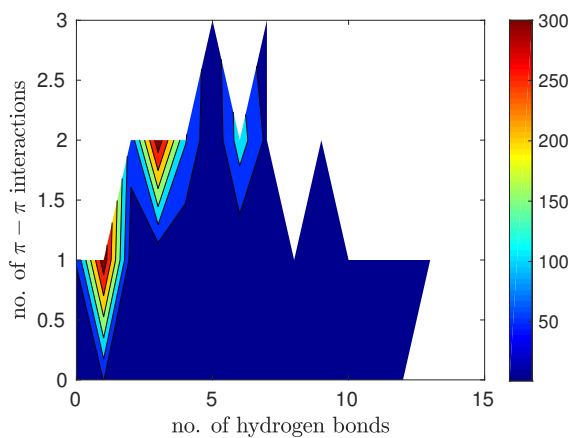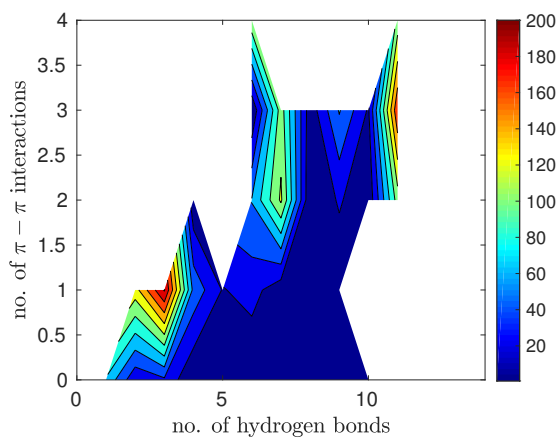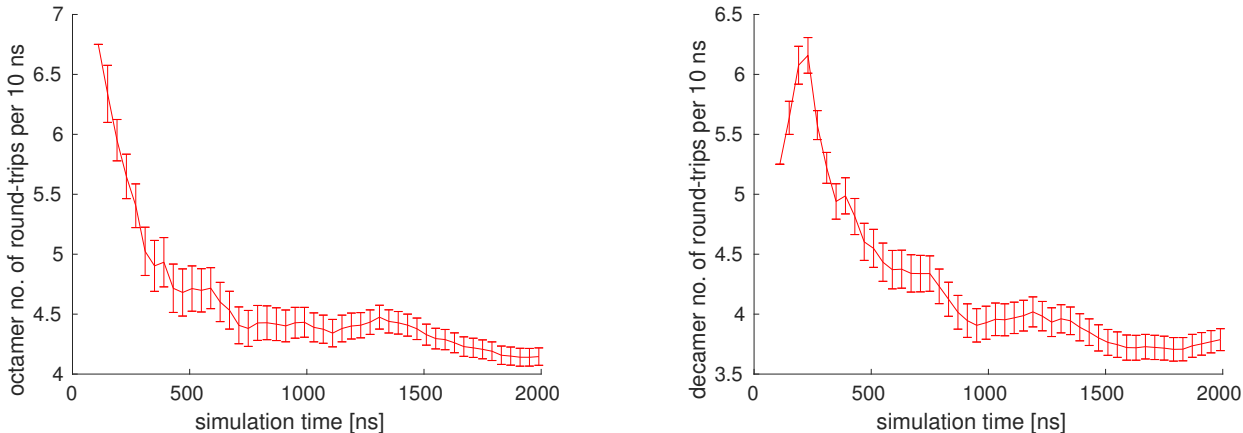
Figure 14: Same as Figure 2 but from the octamer with *ortho*-xylylenediamine linker in the REMD simulation. Foldamer that has a nanorod foldamer as part of its structure.

(a) Octamer with *ortho*-xylylenediamine linker.

(b) Decamer with *ortho*-xylylenediamine linker.

Figure 15: Same as Figure 7 but from REMD simulations.

(a) Octamer with *ortho*-xylylenediamine linker. (b) Decamer with *ortho*-xylylenediamine linker.

Figure 16: Plots of average number of round-trips in a given observation time (10 ns) per replica vs. simulation time for the REMD simulations.

Discussion

Apart from investigating and observing new and diverse foldamers with various linkers, our study revealed two main observations. One is that the triazine polymers do not appear to prefer to form nanorod foldamers longer than folded hexamers or more than 3 π-π interactions. To see if this was the really the case, we ran a 2 μs regular MD simulation of a dimer of tetramers. The dimer of tetramers formed a foldamer similar to the one in Figure 14 that had a nanorod foldamer as part of its structure with 3 π-π interactions with one triazine ring dangling on each side, stabilized by "along the chain" or "turn" hydrogen bonds, as shown in Figure 17. This result supports our observation but the question remains to be answered: whether this is a pathway problem to get more than 3 triazine rings lined up or a geometry/stability issue that the triazine rings cannot all line up. Since the CAS algorithm was able to sample the linear nanorod foldamer with 4 π-π interactions and Ref. 11 showed that an artificially set-up nanorod foldamer from a dimer of pentamers remained stable throughout a 500 ns MD simulation, this might be a pathway problem. Nonetheless, it is also possible that as the triazine polymer gets longer, there are so many possible conformations, many of which still provide many stabilizing hydrogen bonds and abundant dispersion interactions,

28

that on balance, the linear nanorod foldamer is not the most favorable. In addition, as the triazine polymer gets longer, the accessible area for unfavorable solvent interactions gets larger in a linear nanorod structure. Globular structures may be more favorable for longer polymers, which can reduce the accessible area for unfavorable solvent interactions and can still have abundant non-covalent interactions.

In the course of these studies, each additional simulation method expanded the conformational space explored within the given simulation time frame. With the given rotational barriers in going from *cis* to *trans* and vice versa, regular MD explores foldamers with the given preset *cis* and *trans* configurations. The CAS algorithm, on the other hand, probes and samples extensively along the chosen reaction coordinates, as it is a reaction coordinate-based method, with its many walkers. As a result, the CAS algorithm can sample many small energy minima in a relatively flat free energy landscape or a "golf-course" potential[81] and overcome entropic barriers. The linear nanorod foldamer from the octamer with *ortho*-xylylenediamine linker, which the CAS algorithm successfully sampled, was probably was one of the many small energy minima in a relatively flat free energy landscape. However, since the CAS algorithm is heavily dependent on the choice of reaction coordinates, it did not sample other foldamers with *cis* to *trans* and vice versa isomerizations, like the "zigzag" and "staircase" foldamers found with REMD. In other words, if the chosen reaction coordinates do not include dihedral angles for the *cis* and *trans* bonds with rotational barriers, then the CAS algorithm, like regular MD, does not find other foldamers requiring *cis* to *trans* and vice versa isomerizations. Hence, the CAS algorithm is effective in exploring a particular class of foldamers for these macromolecules, but the obtained free energy landscape does not represent the total free energy landscape of the possible conformations. In other words, the CAS algorithm has limited exploratory ability depending on the reaction coordinates chosen.

REMD, on the other hand, does not need pre-defined reaction coordinates and is effective in overcoming enthalpic barriers uniformly, as it is a reaction coordinate-free method, with

its different temperatures. As a result, *cis* to *trans* and vice versa isomerizations occurred in REMD simulations and REMD was able to explore more of the total free energy landscape and sample more structures.[82] Unfortunately, REMD does not work well for entropic barriers, including "golf-course" potentials, and can miss sampling small energy minima. Hence, REMD was able to sample new non-linear foldamers but unable to sample linear nanorod foldamers. Taken together, the CAS algorithm only thoroughly explored a part of the total free energy landscape with the chosen reaction coordinates. REMD explored the total free energy landscape more widely by overcoming enthalpic barriers but missed sampling small energy minima since it is ineffective in overcoming entropic barriers.
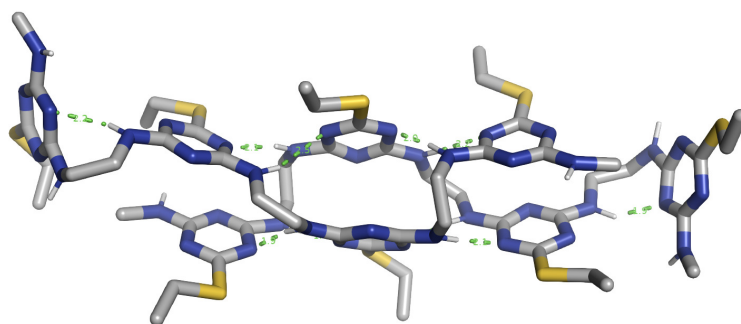

Figure 17: Same as Figure 2 but from the dimer of tetramers in regular MD simulation. Foldamer that has a nanorod foldamer as part of its structure.

Conclusions

Our original motivation of this study was to understand foldamer architectures of length-defined triazine-based polymers; and to investigate specific linker structures in the middle of triazine polymers that might influence folding into a conformation discovered in prior work, i.e., the nanorod foldamer. While nanorod foldamer conformations do arise for shorter polymers up to hexamers, they appear to be less likely to persist for longer polymers. Using the CAS algorithm, it was possible to sample the expected nanorod foldamer for an octamer with the *ortho*-xylylenediamine linker. However, the expected nanorod foldamer did not dominate the conformer population over other possible conformations of octamers.

In addition, the simulations in this paper revealed a diversity of foldamer structures that are possible. For example, foldamers containing π bonded pairs of triazine rings in a nanorod architecture may form with folding sections that include a full triazine monomer, i.e., with the aromatic triazine ring as part of the folding section. The SI also showed a transient foldamer seen from an undecamer with two *ortho*-xylylenediamine linkers, which are placed 3 triazine rings away from the end. The folding section is composed of a central section of triazine monomers and has the aromatic groups of the two *ortho*-xylylenediamine linker aligned. From a molecular design point of view, this suggests that the linkers that are sequenced into a triazine polymer might influence structure along the foldamer arms, as opposed to primarily working at the fold position.

The simulations also showed new foldamer architectures, especially when *cis* and *trans* bond configurations were interchangeable in the REMD simulations. The "zigzag" foldamer of the octamer in Figure 10, comprising 20.08 % of the total population, represents a new well-organized foldamer architecture. The SI also showed a "zigzag" foldamer found for a hexamer with an ethylenediamine linker at the fold.

The nanorod and "zigzag" motifs can be seen as portions of larger structures. The "zigzag" motif can be found, for example, as a portion of the decamer conformation seen in Figure 13. Linear or near linear arrangements of hydrogen and π bonded triazine rings up to 6, similar to the hexamer nanorod, can be seen as portions of octamer structures in Figure 12 and Figure 14. Indeed, the octamer foldamer with 6-ring nanorod motif comprised 53.87 % of the total population. In addition, the triazine monomer in the fold section, which is not in a π-π interaction, does have a set of paired hydrogen bonds to an adjacent triazing ring that is in a π-π interaction as part of a nanorod motif.

Overall, these simulations, using three different simulation techniques, showed that the foldamer behavior is potentially quite rich. It may be challenging to control or predict by synthetic design. On the other hand, the specific linker structures investigated for hexamers, i.e., those based on pentanediamine and *ortho*-xylylenediamine, were favorable for nanorod

31

structures of hexamers as expected. In addition, we found that the motif of the nanorod structure can arise within a larger structures, and thus may be a useful motif to explore as a potential building block.

Supporting Information Available

Additional foldamers from additional simulations are shown and described. The PSE files for the representative conformations from the CAS algorithm and REMD simulations are available and can be viewed with PyMOL. The CAS algorithm code is available at `http://github.com/shirleyahn/CAS_Code`. Simulation files are available upon request. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Acknowledgement

References

(1) Lutz, J.-F.; Meyer, T. Y.; Ouchi, M.; Sawamoto, M. *Sequence-Controlled Polymers: Synthesis, Self-Assembly, and Properties*; ACS Publications, 2014.

(2) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341*, 1238149.

(3) Szymaski, J. K.; Abul-Haija, Y. M.; Cronin, L. Exploring Strategies to Bias Sequence in Natural and Synthetic Oligomers and Polymers. *Accounts of Chemical Research* **2018**, *51*, 649–658.

(4) Trinh, T. T.; Laure, C.; Lutz, J.-F. Synthesis of Monodisperse Sequence-Defined Polymers Using Protecting-Group-Free Iterative Strategies. *Macromolecular Chemistry and Physics* **2015**, *216*, 1498–1506.

(5) Zhang, Z.; You, Y.; Hong, C. Multicomponent Reactions and Multicomponent Cascade Reactions for the Synthesis of Sequence-Controlled Polymers. *Macromolecular Rapid Communications* **2018**, *39*, 1800362.

(6) Hill, S. A.; Gerke, C.; Hartmann, L. Recent Developments in Solid-Phase Strategies Towards Synthetic, Sequence-Defined Macromolecules. *Chemistry–An Asian Journal* **2018**, *13*, 3611–3622.

(7) Cole, J. P.; Hanlon, A. M.; Rodriguez, K. J.; Berda, E. B. Protein-Like Structure and Activity in Synthetic Polymers. *Journal of Polymer Science Part A: Polymer Chemistry* **2017**, *55*, 191–206.

(8) Lyon, C. K.; Prasher, A.; Hanlon, A. M.; Tuten, B. T.; Tooley, C. A.; Frank, P. G.; Berda, E. B. A Brief User's Guide to Single-Chain Nanoparticles. *Polymer Chemistry* **2015**, *6*, 181–197.

(9) Aida, T.; Meijer, E.; Stupp, S. I. Functional Supramolecular Polymers. *Science* **2012**, *335*, 813–817.

(10) Solleder, S. C.; Schneider, R. V.; Wetzel, K. S.; Boukis, A. C.; Meier, M. A. Recent Progress in the Design of Monodisperse, Sequence-Defined Macromolecules. *Macromolecular Rapid Communications* **2017**, *38*, 1600711.

(11) Grate, J. W.; Mo, K.-F.; Daily, M. D. Triazine-Based Sequence-Defined Polymers with Side-Chain Diversity and Backbone-Backbone Interaction Motifs. *Angewandte Chemie International Edition* **2016**, *55*, 3925–3930.

(12) Ahn, S.-H.; Grate, J. W.; Darve, E. F. Efficiently Sampling Conformations and Pathways Using the Concurrent Adaptive Sampling (CAS) Algorithm. *The Journal of Chemical Physics* **2017**, *147*, 074115.

(13) Ahn, S.-H.; Grate, J. W.; Darve, E. F. Investigating the Role of Non-Covalent Interactions in Conformation and Assembly of Triazine-Based Sequence-Defined Polymers. *The Journal of Chemical Physics* **2018**, *149*, 072330.

(14) Archer, E. A.; Krische, M. J. Duplex Oligomers Defined via Covalent Casting of a One-Dimensional Hydrogen-Bonding Motif. *Journal of the American Chemical Society* **2002**, *124*, 5074–5083.

(15) Cavallo, G.; Poyer, S.; Amalian, J.-A.; Dufour, F.; Burel, A.; Carapito, C.; Charles, L.; Lutz, J.-F. Cleavable Binary Dyads: Simplifying Data Extraction and Increasing Storage Density in Digital Polymers. *Angewandte Chemie International Edition* **2018**, *57*, 6266–6269.

(16) Cole, J.; Lessard, J.; Rodriguez, K.; Hanlon, A.; Reville, E.; Mancinelli, J.; Berda, E. Single-Chain Nanoparticles Containing Sequence-Defined Segments: Using Primary Structure Control to Promote Secondary and Tertiary Structures in Synthetic Protein Mimics. *Polymer Chemistry* **2017**, *8*, 5829–5835.

(17) de Rochambeau, D.; Sun, Y.; Barlog, M.; Bazzi, H. S.; Sleiman, H. F. Modular Strategy to Expand the Chemical Diversity of DNA and Sequence-Controlled Polymers. *The Journal of Organic Chemistry* **2018**, *83*, 9774–9786.

(18) Espeel, P.; Carrette, L. L.; Bury, K.; Capenberghs, S.; Martins, J. C.; Du Prez, F. E.;

Madder, A. Multifunctionalized Sequence-Defined Oligomers from a Single Building Block. *Angewandte Chemie International Edition* **2013**, *52*, 13261–13264.

(19) Hartmann, L.; Krause, E.; Antonietti, M.; Börner, H. G. Solid-Phase Supported Polymer Synthesis of Sequence-Defined, Multifunctional Poly(amidoamines). *Biomacromolecules* **2006**, *7*, 1239–1244.

(20) Kanasty, R. L.; Vegas, A. J.; Ceo, L. M.; Maier, M.; Charisse, K.; Nair, J. K.; Langer, R.; Anderson, D. G. Sequence-Defined Oligomers from Hydroxyproline Building Blocks for Parallel Synthesis Applications. *Angewandte Chemie International Edition* **2016**, *55*, 9529–9533.

(21) Karamessini, D.; Simon-Yarza, T.; Poyer, S.; Konishcheva, E.; Charles, L.; Letourneur, D.; Lutz, J.-F. Abiotic Sequence-Coded Oligomers as Efficient In Vivo Taggants for the Identification of Implanted Materials. *Angewandte Chemie International Edition* **2018**, *57*, 10574–10578.

(22) König, N. F.; Telitel, S.; Poyer, S.; Charles, L.; Lutz, J.-F. Photocontrolled Synthesis of Abiotic Sequence-Defined Oligo(Phosphodiester)s. *Macromolecular Rapid Communications* **2017**, *38*, 1700651.

(23) Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable Synthesis of Sequence-Defined, Unimolecular Macromolecules by Flow-IEG. *Proceedings of the National Academy of Sciences* **2015**, *112*, 10617–10622.

(24) Martens, S.; Landuyt, A.; Espeel, P.; Devreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional Sequence-Defined Macromolecules for Chemical Data Storage. *Nature Communications* **2018**, *9*, 4451.

(25) Nowalk, J. A.; Fang, C.; Short, A. L.; Weiss, R. M.; Swisher, J. H.; Liu, P.; Meyer, T. Y. Sequence-Controlled Polymers Through Entropy-Driven Ring-Opening

Metathesis Polymerization: Theory, Molecular Weight Control, and Monomer Design. *Journal of the American Chemical Society* **2019**, *141*, 5741–5752.

(26) Porel, M.; Alabi, C. A. Sequence-Defined Polymers via Orthogonal Allyl Acrylamide Building Blocks. *Journal of the American Chemical Society* **2014**, *136*, 13162–13165.

(27) Roy, R. K.; Meszynska, A.; Laure, C.; Charles, L.; Verchin, C.; Lutz, J.-F. Design and Synthesis of Digitally Encoded Polymers that Can Be Decoded and Erased. *Nature Communications* **2015**, *6*, 7237.

(28) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *The Journal of Physical Chemistry* **1996**, *100*, 1578–1599.

(29) Schneider, R. V.; Waibel, K. A.; Arndt, A. P.; Lang, M.; Seim, R.; Busko, D.; Bräse, S.; Lemmer, U.; Meier, M. A. Sequence-Definition in Stiff Conjugated Oligomers. *Scientific Reports* **2018**, *8*, 17483.

(30) Zydziak, N.; Feist, F.; Huber, B.; Mueller, J. O.; Barner-Kowollik, C. Photo-Induced Sequence Defined Macromolecules via Hetero Bifunctional Synthons. *Chemical Communications* **2015**, *51*, 1799–1802.

(31) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314*, 141–151.

(32) Rehm, T.; Schmuck, C. How to Achieve Self-Assembly in Polar Solvents Based on Specific Interactions? Some General Guidelines. *Chemical Communications* **2008**, 801–813.

(33) Gabriel, G. J.; Iverson, B. L. Aromatic Oligomers that Form Hetero Duplexes in Aqueous Solution. *Journal of the American Chemical Society* **2002**, *124*, 15174–15175.

(34) Sebaoun, L.; Kauffmann, B.; Delclos, T.; Maurizot, V.; Huc, I. Assessing Stabilization

Through π-π Interactions in Aromatic Oligoamide β-Sheet Foldamers. *Organic Letters* **2014**, *16*, 2326–2329.

(35) Sebaoun, L.; Maurizot, V.; Granier, T.; Kauffmann, B.; Huc, I. Aromatic Oligoamide β-Sheet Foldamers. *Journal of the American Chemical Society* **2014**, *136*, 2168–2174.

(36) Ikkanda, B.; Iverson, B. Exploiting the Interactions of Aromatic Units for Folding and Assembly in Aqueous Environments. *Chemical Communications* **2016**, *52*, 7752–7759.

(37) Goodman, C. M.; Choi, S.; Shandler, S.; DeGrado, W. F. Foldamers as Versatile Frameworks for the Design and Evolution of Function. *Nature Chemical Biology* **2007**, *3*, 252.

(38) Gellman, S. H. Foldamers: A Manifesto. *Accounts of Chemical Research* **1998**, *31*, 173–180.

(39) Guichard, G.; Huc, I. Synthetic Foldamers. *Chemical Communications* **2011**, *47*, 5933–5941.

(40) Hecht, S.; Huc, I. *Foldamers: Structure, Properties and Applications*; John Wiley & Sons, 2007.

(41) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. A Field Guide to Foldamers. *Chemical Reviews* **2001**, *101*, 3893–4012.

(42) 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. *Journal of Cheminformatics* **2012**, *4*, 17.

(43) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophysical Journal* **1996**, *70*, 97–110.

(44) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The Weighted Ensemble Path Sampling Method is Statistically Exact for a Broad Class of Stochastic Processes and Binning Procedures. *The Journal of Chemical Physics* **2010**, *132*, 054107.

(45) Adelman, J. L.; Grabe, M. Simulating Rare Events Using a Weighted Ensemble-Based String Method. *The Journal of Chemical Physics* **2013**, *138*, 01B616.

(46) Abdul-Wahid, B.; Feng, H.; Rajan, D.; Costaouec, R.; Darve, E.; Thain, D.; Izaguirre, J. A. AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. *Journal of Chemical Information and Modeling* **2014**, *54*, 3033–3043.

(47) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M. et al. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *Journal of Chemical Theory and Computation* **2015**, *11*, 800–809.

(48) Rhee, Y. M.; Pande, V. S. Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophysical Journal* **2003**, *84*, 775–786.

(49) Sanbonmatsu, K.; Garcia, A. Structure of Met-Enkephalin in Explicit Aqueous Solution Using Replica Exchange Molecular Dynamics. *Proteins: Structure, Function, and Bioinformatics* **2002**, *46*, 225–234.

(50) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica Exchange Molecular Dynamics Simulations of Amyloid Peptide Aggregation. *The Journal of Chemical Physics* **2004**, *121*, 10748–10756.

(51) Johnson, R. R.; Kohlmeyer, A.; Johnson, A. C.; Klein, M. L. Free Energy Landscape of a DNA- Carbon Nanotube Hybrid Using Replica Exchange Molecular Dynamics. *Nano Letters* **2009**, *9*, 537–541.

(52) Mori, T.; Miyashita, N.; Im, W.; Feig, M.; Sugita, Y. Molecular Dynamics Simulations of Biological Membranes and Membrane Proteins Using Enhanced Conformational Sampling Algorithms. *Biochimica et Biophysica Acta-Biomembranes* **2016**, *1858*, 1635–1651.

(53) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

(54) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *The Journal of Chemical Physics* **2008**, *129*, 124105.

(55) Buchete, N.-V.; Hummer, G. Peptide Folding Kinetics from Replica Exchange Molecular Dynamics. *Physical Review E* **2008**, *77*, 030902.

(56) Stelzl, L. S.; Hummer, G. Kinetics from Replica Exchange Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2017**, *13*, 3927–3935.

(57) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. et al. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(58) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101.

(59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(60) da Silva, A. W. S.; Vranken, W. F. ACPYPE-Antechamber Python Parser Interface. *BMC Research Notes* **2012**, *5*, 367.

(61) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

(62) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383–394.

(63) Calimet, N.; Schaefer, M.; Simonson, T. Protein Molecular Dynamics with the Generalized Born/ACE Solvent Model. *Proteins: Structure, Function, and Bioinformatics* **2001**, *45*, 144–158.

(64) Voelz, V. A.; Dill, K. A.; Chorny, I. Peptoid Conformational Free Energy Landscapes from Implicit-Solvent Molecular Simulations in AMBER. *Peptide Science* **2011**, *96*, 639–650.

(65) Fu, C.-F.; Tian, S. X. A Comparative Study for Molecular Dynamics Simulations of Liquid Benzene. *Journal of Chemical Theory and Computation* **2011**, *7*, 2240–2252.

(66) Kang, M.; Zhang, P.; Cui, H.; Loverde, S. M. π–π Stacking Mediated Chirality in Functional Supramolecular Filaments. *Macromolecules* **2016**, *49*, 994–1001.

(67) Han, G.; Guo, Y.; Song, X.; Wang, Y.; Yi, Y. Terminal π–π Stacking Determines Three-Dimensional Molecular Packing and Isotropic Charge Transport in an A–π–A Electron Acceptor for Non-Fullerene Organic Solar Cells. *Journal of Materials Chemistry C* **2017**, *5*, 4852–4857.

(68) Amm, M.; Platzer, N.; Guilhem, J.; Bouchet, J. P.; Volland, J. P. Structural and Conformational Study of Substituted Triazines by 15N NMR and X-Ray Analysis. *Magnetic Resonance in Chemistry* **1998**, *36*, 587–596.

(69) Birkett, H. E.; Harris, R. K.; Hodgkinson, P.; Carr, K.; Charlton, M. H.; Cherryman, J. C.; Chippendale, A. M.; Glover, R. P. NMR Studies of Exchange Between Triazine Rotamers. *Magnetic Resonance in Chemistry* **2000**, *38*, 504–511.

(70) Earl, D. J.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Physical Chemistry Chemical Physics* **2005**, *7*, 3910–3916.

(71) Xu, Z.; Lei, X.; Tu, Y.; Tan, Z.-J.; Song, B.; Fang, H. Dynamic Cooperation of Hydrogen Binding and π Stacking in ssDNA Adsorption on Graphene Oxide. *Chemistry–A European Journal* **2017**, *23*, 13100–13104.

(72) Kabeláč, M.; Kroutil, O.; Předota, M.; Lankaš, F.; Šíp, M. Influence of a Charged Graphene Surface on the Orientation and Conformation of Covalently Attached Oligonucleotides: a Molecular Dynamics Study. *Physical Chemistry Chemical Physics* **2012**, *14*, 4217–4229.

(73) Sivasakthi, V.; Anitha, P.; Kumar, K. M.; Bag, S.; Senthilvel, P.; Lavanya, P.; Swetha, R.; Anbarasu, A.; Ramaiah, S. Aromatic-Aromatic Interactions: Analysis of π-π Interactions in Interleukins and TNF Proteins. *Bioinformation* **2013**, *9*, 432.

(74) Wang, Y.-L.; Laaksonen, A.; Fayer, M. D. Hydrogen Bonding Versus π–π Stacking Interactions in Imidazolium–Oxalatoborate Ionic Liquid. *The Journal of Physical Chemistry B* **2017**, *121*, 7173–7179.

(75) Brylinski, M. Aromatic Interactions at the Ligand–Protein Interface: Implications for the Development of Docking Scoring Functions. *Chemical Biology & Drug Design* **2018**, *91*, 380–390.

(76) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **1979**, *28*, 100–108.

(77) R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2018.

(78) DeLano, W. L. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter On Protein Crystallography* **2002**, *40*, 82–92.

(79) Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *Journal of Chemical Theory and Computation* **2010**, *6*, 2804–2808.

(80) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. Simulating Replica Exchange Simulations of Protein Folding with a Kinetic Network Model. *Proceedings of the National Academy of Sciences* **2007**, *104*, 15340–15345.

(81) Loper, J.; Zhou, G.; Geman, S. Capacities and the Free Passage of Entropic Barriers. 2018.

(82) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced Sampling in Molecular Dynamics. *The Journal of Chemical Physics* **2019**, *151*, 070902.

Graphical TOC Entry