## **ARTICLE** OPEN Dimerization energetics of curli fiber subunits CsgA and CsgB

Martha Dunbar<sup>1</sup>, Elizabeth DeBenedictis D<sup>1</sup> and Sinan Keten D<sup>1,2</sup>

Curli fibers are functional amyloids that exhibit strong adhesion and robust self-assembly as biofilm structural components; however, the binding energetics and mechanical properties of wild-type curli are not well understood. To address this, we present dimer structures made up of the major and minor curli subunits (CsgA and CsgB), perform free energy calculations to obtain absolute binding energies, and estimate the Young's modulus and persistence length of curli fibers. Equilibrium molecular dynamics simulations are used to evaluate nonbonded interactions. Binding energies are most favorable for CsgB–CsgA, while CsgA–CsgA dimers have a higher binding energy than CsgB–CsgB despite possessing less favorable nonbonded interaction energies. Decomposing each potential of mean force of separation indicated that solvent effects positively impact CsgA–CsgA binding but not CsgB–CsgB and CsgB–CsgA. Charged residues and conserved polar residues were also notable contributors to attractive nonbonded interactions, underlining their importance in dimer assembly. Our findings elucidate sequence effects on binding energy contributions and establish theoretical limits for the elasticity, persistence length, and strength of curli fibers.

npj Computational Materials (2019)5:27; https://doi.org/10.1038/s41524-019-0164-5

## INTRODUCTION

Curli fibers are a type of functional amyloid found in the biofilm extracellular matrix for many bacteria phyla.<sup>1</sup> These fibers perform several roles within the biofilm, including attachment,<sup>2,3</sup> cell invasion,<sup>4</sup> protection against phage attacks,<sup>5,6</sup> and more.<sup>7</sup> Studies with biofilms of varying composition have found curli to be necessary for adhesion.<sup>8</sup> Curli fibers are composed of mostly CsqA, the major subunit, with CsgB, the minor subunit, incorporated in smaller ratios.<sup>9</sup> Mature CsgA and CsgB proteins share similar dimensions and structure: a beta-helix with five strand-loop-strand motifs (see Fig. 1). These five repeats make up the amyloid core, and an unstructured 22-residue domain is located at the N-terminus. During self-assembly, CsgB nucleates CsgA on the surface of the cell, inducing a conformational change from soluble to insoluble. Both proteins pass through the pore protein, CsqG, to be secreted from the surface. CsqG has been shown to be necessary for stabilizing CsgA and CsgB during curli growth in vivo,<sup>10</sup> and chaperone-like proteins CsgE and CsgF are needed for transporting CsgA to the CsgG pore and for linking the growing fibril to the cells, respectively.<sup>11,12</sup> The dispersion of CsgG on the cell surface dictates curli growth locations and can promote CsgA and CsgB interactions.<sup>1</sup>

CsgB not only aggregates faster than CsgA, but also speeds up CsgA aggregation through nucleation, in a process called "seeding." Seeding can also be achieved with preformed CsgA or CsgB fibers. In fact, even at substoichiometric levels, the presence of CsgB can speed up CsgA aggregation.<sup>14</sup> Although CsgB is required for nucleation in vivo, CsgA and CsgB alone can eventually self-polymerize in vitro,<sup>9,15–18</sup> although H/D exchange suggests CsgA fibrils may be more stable than CsgB fibrils.<sup>19,20</sup> Additionally, cells expressing only CsgB can grow curli by recruiting CsgA secreted from neighboring cells. This process of cross-seeding is called "interbacterial complementation" and highlights the unique strategy of isolating the nucleator function (CsgB) from fiber elongation (CsgA) organelles.<sup>21</sup> This has been proposed by Hammer et al. to be another strategy for reducing exposure to toxic folding intermediates.<sup>22</sup>

To further investigate the role of sequence in self-assembly, studies involving curli mutants and synthetic peptides have already begun to discern the roles of specific amino acids and repeat strands. For example, within the CsgA monomer, four residues have been found to be critical for self-assembly: Q29, N34, Q119, and N124. These are all inward-facing residues, within the repetitive polar zippers. Notably, although Asn at 34 and 124 could be replaced by Gln and still result in fiber formation, Gln at 29 and 119 could not be replaced by Asn, demonstrating the exclusivity of this position in self-assembly.<sup>23</sup> Alternatively, when studying sequence effects in repeat strands multiple aspartic acid and glycine residues in R2, R3, and R4 were categorized as "gatekeepers" for their role in modulating polymerization, effectively slowing curli growth.<sup>24</sup> Deamidation can also slow curli growth, and multiple areas were found in the amyloid core that are susceptible, particularly at Asn in locations 57, 87, and 102.<sup>25</sup>

Multiple studies have investigated the aggregation propensity of entire repeat units to determine their importance for curli growth. Within CsgA, R1, R3, and R5 have been found to be particularly amyloidogenic, while R1, R2, and R4 are the most amyloidogenic within CsgB.<sup>16</sup> Indeed, R5 from CsgB may not take part in the amyloid core but may mainly interact with the membrane surface. Deletion studies have found that CsgB can form fibers without R1, R2, or R3 repeats, but mutants missing R4 or R5 did not localize to the membrane and form fibrils.<sup>18</sup> However, CsgB without R5 can nucleate the major subunit although less effectively as it is no longer tethered to the cell. R5 from CsgB is less conserved than all other repeat units and has the most positive charge. This charged region plays an important role by neutralizing oppositely charged membrane lipid

<sup>1</sup>Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA and <sup>2</sup>Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL 60208, USA

Correspondence: Sinan Keten (s-keten@northwestern.edu)

Received: 24 August 2018 Accepted: 25 January 2019 Published online: 26 February 2019



**Fig. 1** a CsgB–CsgB dimer structure with highlighted beads representing the center of mass position for each frame-of-reference point. **b** The orientational ( $\Theta, \Phi, \Psi$ ) and positional restraints are listed ( $\theta, \phi, r$ ). The separation distance, r, is the distance between P1a and P2a, the center of mass of the beta-sheet carbonyl carbons of the two monomers. **c** Visualizations of dimer set-ups of each type

molecules, allowing a higher local concentration of proteins to spur aggregation.  $^{\rm 26}$ 

Although mutagenesis studies have revealed key information about the role of specific amino acids and repeats in curli subunits, details pertaining to structure and mechanical properties have not been well characterized. In recent years, atomic force microscopy (AFM) has been increasingly applied to amyloid systems to obtain high-resolution topological, mechanical, and even kinetic information. For example, high-resolution AFM data revealed that curli nucleation occurs in one step, with folding and oligomerization occurring together. Growth was found to be polar (faster in one direction) and displayed stop-and-go kinetics. Growing fibrils could also develop "scars" where structural perturbations appeared at the tip that could remain in the fiber or be resolved.<sup>27</sup> AFM has also been applied to study binding of CsgA monomers and curliated bacteria with fibronectin-functionalized cantilevers. This revealed the formation of multiple quantized bonds, requiring about ~50 pN to unbind.<sup>28</sup> On the network scale, indentation experiments using AFM have calculated a transverse Young's modulus of ~10 MPa for CsgA and CsgB networks, as well as modified networks.<sup>29</sup> Although these mechanical experiments provide insight into binding and elastic properties of curli, variability in experiments (such as inclusion of bacteria and other components, variable network thickness or architecture, and lack of monomer-scale knowledge) create a challenge when extracting general mechanical properties or characterizing single fibers. Conducting experiments such as tensile extension of curli subunits or fibers, using thermal fluctuations to get persistence length, or obtaining structural information for curli bundles and networks would be instrumental in better understanding curli's contribution to biofilm mechanics and its potential as an engineered biomaterial.

To progress beyond AFM-based mechanical testing and determine binding behavior at an atomistic resolution, in silico free energy calculations are necessary. Recently, we have presented atomistic models for individual CsgA and CsgB subunits,<sup>30</sup> which can be used as a foundation for investigations of fibrils using both all-atom<sup>31</sup> and coarse-grained molecular dynamics (MD) simulations.<sup>32</sup> From this work by DeBenedictis et al., the RobB-5 CsgB structure and the CsgA-map model based upon it will be utilized in this paper. However, determining the binding energies of protein–protein assemblies using computational methods can present a challenge. With these complex assemblies, the multitude of conformations to sample precludes

the use of equilibrium techniques, while methods that reduce degrees of freedom by coarse graining or implicit solvent assumptions are neither accurate nor precise enough to determine absolute binding energies. In order to study these complex interactions in finite computational time, two advanced sampling methods, extended adaptive biasing force<sup>33</sup> (eABF) and replica exchange umbrella sampling (REUS), were utilized per the method originally developed by Woo and Roux<sup>34</sup> and elaborated upon by Gumbart et al.<sup>35</sup> This method expands the mathematically simple but computationally indeterminable (for complex equilibrium binding systems) constant equation,  $K_{eq} = 4\pi \int_0^r dr r^2 e^{-\beta W(r)}$ , expressed in terms of the onedimensional radial potential of mean force (PMF) W(r), into multiple components that are evaluated either analytically or with a biased simulation. The equilibrium binding constant determined with this method can then be used to calculate the standard binding free energy,  $\Delta G_{\text{bind}} = k_{\text{B}}T \ln(K_{\text{eq}}C^{\circ})$ , where  $C^{\circ}$  is the standard concentration, 1 M. The expanded version of the equilibrium binding constant equation includes components that account for the effect of a series of conformational, orientational, and positional restraints applied to the bound protein system and removed in the unbound system. By including these restraints, all positions and orientations of the protein dimer at all separation distances do not need to be sampled, merely a subset, making this method feasible for determining binding energy. eABF is used to determine the energetic contribution of applying a restraint in the bound condition and removing it in the unbound condition, which is combined with the REUS-generated separation PMF using the expanded equilibrium binding constant equation to calculate the free energy. With this method, it is possible to compare the dimer systems and analyze the effect of interfacial residue pairs, solvation, and nonbonded energy on binding energy.

Here we present atomistic models of CsgA–CsgA, CsgB–CsgB, and CsgB–CsgA dimers and test them using equilibrium and nonequilibrium dynamics simulations. We are particularly interested in determining the binding energy and uncovering which amino acids contribute to dimerization. Dimer model structures are first constructed from subunits using all-atomistic implicit solvent calculations,<sup>36</sup> such that the internally conserved polar residues (Ser, Gln, and Asn) align across the interface, as shown in Fig. 2 and Figures S8, S9, and S10. For CsgB–CsgA dimers, the bottom monomer was CsgB, to reflect the typical nucleation behavior of curli. A strategically chosen frame of reference was applied to each dimer system for use in free energy calculations,

Polar

Hydrophobi



**Fig. 2** Residue alignment at dimer interface for CsgA–CsgA, CsgB–CsgB, and CsgB–CsgA dimers. The dotted line represents the interface between the upper and lower subunits, with R5 of the top monomer shown above and R1 of the bottom monomer below. Although CsgA (and B) have two turns, the R5 repeat is terminated at the end of the second strand (R5'), before the second turn. Thus only the first turn is included here for consistency. Polar amino acids are shown in light blue, hydrophobic residues in red, and charged amino acids are labeled by (+) or (-). Glycine and five- or six-member rings are not colored. Beta-strands and turns are marked for the CsgA–CsgA dimer; CsgB–CsgB and CsgB–CsgA have the same designations. For R1, residue numbering is from 23–41 for CsgA and 24–42 for CsgB. Residue numbering for R5 is from 113 to 131 for CsgA and from 112 to 130 for CsgB. Here we can see that CsgB–CsgB dimers contain no aromatic residues at the interface. CsgA has multiple aromatic residues in R1 and R5, although no pairs align. CsgA–CsgA and CsgB–CsgA dimers only contain one charged residue at their interface, while CsgB–CsgB has five charged residues at the interface. These differences in sequence at the interface may give rise to variations in oligomerization speed, binding energy, as well as relative contributions to the binding energy

**Table 1.** Free energy of binding,  $\Delta G^0_{\text{bind}'}$  calculated from the separation PMF, as well as the Young's modulus, *E*, and persistence length,  $L_{\text{pr}}$ , generated by fitting the curve with a harmonic function to find the spring constant, *k*, and the equilibrium distance,  $r_0$ 

| Property                              | CsgA–CsgA        | CsgB–CsgB        | CsgB–CsgA        |
|---------------------------------------|------------------|------------------|------------------|
| $\Delta G_{\text{bind}}^0$ (kcal/mol) | $-4.01 \pm 0.36$ | $-3.10 \pm 0.41$ | $-8.20 \pm 0.39$ |
| E (GPa)                               | 47.5             | 44.2             | 37.3             |
| $k (k_{\rm B}T/{\rm \AA}^2)$          | 156              | 132              | 107              |
| r <sub>o</sub> (Å)                    | 22.3             | 22.6             | 23.4             |
| L <sub>p</sub> (nm)                   | 2401             | 2064             | 1740             |
| PMF potential of mean force           |                  |                  |                  |

as shown in Fig. 1a. eABF, REUS, and equilibrium MD simulations were then performed in explicit solvent. We calculate binding energy for each set of subunits and compare these in the context of sequence differences. Additionally, using equilibrium energy calculations, we identify residues that are major contributors to nonbonded interaction energies between the two subunits.

## RESULTS

## Free energy calculations

The free energy of binding for all dimer systems are listed in Table 1, while the contributions of the conformational, orientational, and positional restraints on each system can be found in Table S1. As expected, the separation contribution to the free energy is by far the largest, accounting for -18 kcal/mol in CsgA-CsgA, -18 kcal/mol in CsgB-CsgB, and -21 kcal/mol in CsgB-CsgA system. The plateau regions of the separation PMFs in Fig. 3 correspond closely to the free energy contribution of separation but will be higher than experimentally generated PMFs because of the restraints applied in the simulation. When comparing dimer species, the



**Fig. 3** Separation potential of mean force (PMF) calculated with replica exchange umbrella sampling (REUS) for CsgA–CsgA, CsgB–CsgB, and CsgB–CsgA dimers after 3 ns. All REUS simulations were run longer than 3.0 ns to ensure convergence. Figures S1 and S2 show the convergence and variance of each PMF. Inset: A representative harmonic fit of the separation PMF well, used to determine the Young's modulus with a bead-spring approximation. The harmonic fit or each dimer can be found in Figure S3

largest difference in binding energy is due to the magnitude of the separation contribution and not the orientational and positional restraints. Additionally, both CsgA–CsgA and CsgB–CsgB have energetic transition barriers to assembly in the separation PMF. This is due to the balance between the longrange electrostatic repulsion acting between the equivalent total charges on each monomer and solvation effects. Notice that the oppositely charged CsgB–CsgA does not have an energetic barrier,



**Fig. 4** Decomposition of the **a** CsgA–CsgA, **b** CsgB–CsgB, and **c** CsgB–CsgA separation potential of mean force (PMF) curves into electrostatic (blue), Van der Waals (red), and solvation (yellow) contributions with the molecular mechanics Poisson–Boltzmann surface area (MMPBSA) method. Error bars indicate standard deviation and values are shifted so that the last point along the reaction coordinate is zeroed. The unshifted bound state electrostatic and Van der Waals values match those found in Fig. 5a. When comparing the sum of the contributions calculated with MMPBSA (purple) to the separation PMF generated with replica exchange umbrella sampling (green), the values are usually within 1 standard deviation of the MMPBSA total energy, which can be seen in detail in Figure S4

and CsgB–CsgB, which possesses the largest number of charged interface residues, has the largest transition barrier. When observing MD trajectories of separation, the peak of the transition barrier corresponds approximately to the point at which the first water molecule enters the interface. This can also be seen in Fig. 4, where the separation PMFs are decomposed into electrostatic, Van der Waals (VDW), and solvation contributions. In the case of CsgA–CsgA that has a total charge of -6 q per monomer, the electrostatic contribution reaches a peak value ~7 Å from the bound state, has a local minimum at the bound state, and increases again as the dimer is forced together.

The contribution to the free energy from conformational restraints applied to the interface is the next largest, with a total of +4.6 kcal/mol for CsgA-CsgA, +5.1 kcal/mol for CsgB-CsgB, and +9.2 kcal/mol for CsgB-CsgA. This indicates that interfacial residues are important for binding, with the interface pairs identified in Fig. 2 playing a significant role. The interface conformational PMF minimum along the reaction coordinate for the bound state is always smaller than the minimum for the unbound state, indicating that the pair interactions across the interface restrain one another into energetically stable configurations despite being entropically costly. The same is true for the backbone conformational PMF, where the bound state is more tightly held than the unbound state. Knowing this, it is important to note that this computational method assumes the conformation of the bound state of the protein does not significantly differ from the unbound state, which justifies the use of root mean square deviation (RMSD) as an appropriate measure of conformational change. However, this approach seeks to primarily determine the binding energy and does not attempt to resolve the folding process, an approximation making it a more suitable measurement of separating dimers, rather than a growing fibril. Experimental evidence hints at concurrent folding and oligomerization,<sup>27</sup> but this has not been confirmed. Additionally, no method is currently known to the authors that could accurately evaluate the energetic cost of folding during binding, making this the most suitable approach available.

The energetic contribution of the positional and  $\Theta$ ,  $\Phi$  orientational restraints in the bound state for all the dimer types are small as compared to the conformation and separation components and of similar magnitude regardless of species. The exception to this is the  $\Psi$  orientational restraint, which describes the helical "twist" of the dimer and is equal to -18.1 kcal/mol for CsgB–CsgA, -1.0 kcal/mol for CsgA–CsgA, and -7.3 kcal/mol for CsgB–CsgB. This restraint generally has a larger effect than others as sampling the "twist" of the dimer disrupts nonbonded

interactions throughout the interface and pushes sterically repulsed residues together. In contrast, the other positional and orientational restraints primarily measure the tensile breakage of interface interactions. In the case of CsgB–CsgA, the opposite overall charge and subsequent strong electrostatics result in a very high restraint contribution. The contributions of all orientational and positional restraints in the unbound state are determined analytically according to the equations in Methods and Supplementary Information sections and given as  $G_0^{\text{bulk}}$  and  $S^*$ , respectively.  $G_0^{\text{bulk}}$  is equal for all dimer species, as the spring constants are consistent across all simulations.

# Young's modulus and persistence length with a bead-spring approximation

Experimental dimer-scale mechanical tests have not been performed and are currently not accessible using techniques such as AFM. Larger-scale tensile tests have been done on amyloid fibers and networks, obtaining Young's moduli from 0.2 to 20 GPa.<sup>37-39</sup> However, network results are influenced by the behavior of fibrils in shear with neighbors, viscoelasticity due to fluid moving through the network, network architecture, and other factors that are not easily eliminated experimentally. Even if all these factors could be accounted for, the mechanics of networks are fundamentally different than the mechanics of isolated fibers. As a result, there is a gap in knowledge surrounding the mechanical behavior of isolated single fibrils, which hinders coarse grain modeling efforts of curli fibers and networks. Other types of amyloid fibers have been characterized at the single fiber level and can be used as a qualitative comparison to the mechanical properties we have calculated computationally. For example, insulin fibrils studied by Smith et al. with force spectroscopy were found to have a Young's modulus of  $\sim$ 3.3 GPa and a persistence length of  $\sim$ 22 µm. When Adamcik et al. studied a variety of amyloid fibers with the peak force quantitative nanomechanical technique, they found all fibers ranged from 2 to 4 GPa and a persistence length of 3.5–18.5 µm depending on fiber packing. It is also important to note that computational studies of amyloid nanomechanics commonly calculate higher Young's moduli than those found experimentally.<sup>40</sup> We calculate the Young's modulus, E, by fitting a harmonic spring constant to the energy well of the separation PMF (example found in the inset of Fig. 3), using the approximation  $k = \frac{AE}{r_0}$ , where A is the cross-sectional area of a single fibril (~2.76 nm<sup>2</sup>) (ref. <sup>28</sup>) and  $r_0$  is the bond length (center of mass distance) listed in Table 1. It is important to note that this calculation represents an upper limit of E, as this method does not include the effect of monomer



**Fig. 5 a** The nonbonded interaction energy of CsgA–CsgA, CsgB–CsgA, and CsgB–CsgB calculated from explicit solvent simulations for the entire protein. The magnitude of interaction energy is CsgB–CsgA > CsgB–CsgB > CsgA–CsgA. In each case, Van der Waals (VDW) contributions are similar due to similar dimer geometry, but electrostatic contributions can vary greatly. CsgA and CsgB have opposite overall charges, resulting in the lowest (most attractive) interaction energy. For CsgB–CsgB dimers, charged and polar residues are positioned favorable to get lower interaction energies than CsgA–CsgA dimers. Error bars indicate standard deviation. **b** Residue content in CsgA and CsgB by repeat strand. The amino acid content of each repeat strand from R1 (N-terminal) to R5 (C-terminal). Although both CsgA and CsgB have oplar and hydrophobic residues distributed throughout all strands, their aromatic and charge content differs by interior repeats (R2–R4) and interface repeats (R1, R5)

unfolding and is sensitive to the chosen A. However, it would be possible to integrate the effect of unfolding into the calculation in future work. It was found that all dimers have a modulus around 43 GPa (see Table 1), which is significantly higher than experimental results for reasons discussed above. Using the calculated Young's modulus, we find the persistence length using  $L_p = \frac{E}{k_B T}$ assuming each fiber behaves as a flat tape with an area moment of inertia equal to  $I = \frac{bh^3}{12}$ . Each fiber has a persistence length of ~ 2.0 µm, with the additional assumption that the long-range electrostatics of a homomeric fiber are not significantly different from that of a single dimer. Despite this assumption, approximations of both the modulus and persistence length of single fibers can help further refine future coarse-grained models for curli nanofibers. These findings are in very good agreement with recent computational estimates on the persistence lengths of various beta-solenoids.<sup>41</sup>

#### MD—energetic analysis

The stability, dynamics, and nonbonded energetic interactions of dimer structures in equilibrium explicit solvent simulations were also investigated. All dimer structures remained stable throughout the MD simulations and retained their secondary structure (see Dimer Stability Analysis section of Supplementary Information). CsgA and CsgB models have similar beta-sheet content, with 45–55% beta-sheet structure on average of the entire monomer to begin the simulation. Throughout equilibrium MD simulations, there was some fluctuation in beta-sheet content at either terminus of the top (unrestrained) monomer, resulting in 41–53% beta-sheet structure on average at the end of the simulation.

To complement binding energy calculations, we also calculated the total nonbonded interaction energy between each monomer using the NAMDEnergy tool, as seen in Fig. 5a. When considering the entire protein, CsgA–CsgA dimers have the largest deviation in measurement of all set-ups. Owing to similar geometry, all dimer types have similar levels of contribution arising from VDW interactions. For electrostatic contributions, the CsgB–CsgA dimer has the strongest interaction, which is expected because the CsgA and CsgB monomers have overall opposite charges. CsgB–CsgB dimers have a stronger electrostatic contribution than CsgA–CsgA dimers. CsgB–CsgB dimers also have more favorable placement of charged residues leading to interactions on average lower than within CsgA–CsgA dimers.

### Structure production and analysis—interface analysis

To further delineate sequence-based differences in CsgA and CsgB dimers, we focus on the aligned residue pairs at the dimer interface. A schematic showing how residues at the interface align can be seen in Fig. 2, and the number of each type of pair interactions at the interface can be found in Fig. 6a. We note that, for aromatic interactions, only pairs located within the beta-sheet face are included, as the flexible turn regions allow for greater mobility and reduced alignment and contain smaller residues such glycine. Although the number of polar-polar, as hydrophobic-hydrophobic, and glycine-X pairs are similar, CsgB-CsgB dimers have fewer interface pairs containing aromatic residues and more pairs containing charged residues compared to dimers including CsgA. Here glycine-X indicates a pair that has at least one glycine with another residue (glycine or not). Although CsgA-CsgA dimers have four pairs containing an aromatic residue, none of these residues align for possible stacking of side-chain rings. CsgB-CsgB contains multiple interface pairs containing charged residues, all of which are positively charged. CsgA-CsgA and CsgB-CsgA dimers only have one interface pair with a charged residue each. Part of this discrepancy between CsgB-CsgA and CsgB-CsgB is because the C-terminal R5 repeat of CsgB contains the most (+) charged residues; CsgB's R1 and CsgA's R5 only have one charged residue each, see Fig. 5b. While charged residues in R5 of CsgB has been shown to be important for membrane association, aromatic residues in CsgA have not proven critical in assembly based on mutagenesis experiments.<sup>23</sup> Similarities between all three dimers indicate the importance of polar-polar and hydrophobic-hydrophobic interface pairs in stacking.

The energetic contribution of each pair interaction during explicit solvent MD simulations was calculated by measuring the pairwise interaction energy for each set of aligned residues using the NAMDEnergy tool and can be found in Fig. 6b. By sum, polar–polar interactions followed by aromatic–X interactions contribute mostly to CsgA–CsgA interface interactions. Hydrophobic and charged interactions both contributed <20 kcal/mol total, and glycine interactions contributed <10 kcal/mol. For



**Fig. 6** The number of interface pairs of each type is represented in **a**, with the predominant pair types being polar-polar and hydrophobic-hydrophobic. Note that the aromatic-X pairs within the turn are not included in this sum; the turn regions are more flexible than the beta-sheet face, resulting in less consistent alignment. The energetic contribution of each type of pair interaction is shown in **b**. Note that some pair interactions belong to more than one group. Charged interactions contributed more than any other pair type for all three dimers, followed by polar-polar residue pairs. Error bars indicate standard deviation

CsgB-CsgB interactions, polar-polar interactions had the greatest total contribution (although still less in sum than CsgA-CsgA), followed by hydrophobic interactions, then charged interactions. On average, each pair type had similar interaction strengths, except for charged interactions: the one charged pair at the CsgA-CsgA interface had a stronger interaction energy than the CsgB-CsgB charged pairs. Additionally, pairwise interactions were calculated for all possible residue pairs to explain energetic interactions in greater detail. We look specifically at pair interactions with the most negative interaction energy, as seen in Fig. 7. For all dimers, pairwise interactions were dominated by electrostatic interactions, and charged and polar residue pairs were often the strongest. In all cases, the charged C-terminal residues incurred both strong repulsive and attractive interactions. CsgA-CsgA and CsgB-CsgA dimers had strong interactions between both the N- and C-terminus and charged residues, while CsgB-CsgB only had one pair with a strong interaction energy containing a terminal residue.

## DISCUSSION

np

6

These results shed light on how differences in sequence impact dimer interactions despite very similar geometries. Although residue pairs at the interface vary between different dimer set-ups, the interaction energies for each case are, on average, similar ( $-5.3 \pm 4.3$  kcal/mol for CsgA–CsgA vs  $-4.6 \pm 4.3$  kcal/mol for CsgB–CsgB and  $-5.1 \pm 4.3$  kcal/mol for CsgB–CsgA). While all dimers had similar numbers of polar–polar, hydrophobic and



**Fig. 7** Pairwise interactions under -10 kcal/mol for CsgA-CsgA dimers **a**, CsgB-CsgB dimers **b**, and CsgB-CsgA dimers (**c**). Note that both residues in each pair are not necessarily found at the dimer interface. Although CsgA-CsgA and CsgB-CsgA dimers had strong residue pairs containing N- and C-terminal residues, oppositely charged residues, and polar residues, the CsgB-CsgB dimer contained several more oppositely charged pairs. This accentuates the dominance of electrostatic interactions arising from charged residues in CsgB, and their favorable distribution throughout the core to form highly attractive residue pairs

glycine pairs, CsgA dimers had multiple aromatic pairs, which contributed second most by total to interaction energy at the interface, and CsgB had multiple charged pairs due to the positively charged R5 at the C-terminus. However, these pairs contributed less to the interface interaction energy than hydrophobic pairs, making it the third largest contributor. In future work, it would be interesting to look at mutations in charged residues to see the impact on aggregation speed.

In the CsgA–CsgA dimer, charged, polar, and the C-terminal Tyr all contribute strongly through pairwise interactions. Gln–Gln and Asn–Asn interactions between internally conserved polar residues underline their importance in fiber formation and stability. Three of the strongest pairwise interactions for CsgA–CsgA dimers are between charged residues and the positively charged N-terminus, and one is between a polar residue and the negatively charged Cterminus. Three other residue pairs have opposite charges and are not both located at the interface. The remaining four pairs are between polar and one charged residue at the interface. In the CsgB–CsgB dimers, of the strongest pairwise interaction energies, nine pairs are made up of oppositely charged residues, four containing one or both within the unstructured N-terminal 22 residues, and the remaining five containing a residue in an internal repeat (R2 or R3). Additionally, an Arg–Arg (C) pair also contributes significantly to the interaction energy. Only two of these strong pairs are directly aligned at the interface, one containing a conserved Gln–Gln stack. This underlines the role of charged residues in CsgB–CsgB interactions and their favorable distribution throughout the amyloid core (not necessarily at the interface). In CsgB–CsgA dimers, three of the strongest interactions are between negatively charged residues and the positively charged N-terminus. Two pairs contain oppositely charged residues, and two pairs are made up of Gln–Gln and Asn–Asn. The negatively charged C-terminus formed strong pairs with two positively charged residues and one nearby glycine.

Repulsive pairwise interactions occurred for all dimer types and were often among charged residues. However, CsgA–CsgA dimers had much higher repulsive electrostatic interactions than CsgB–CsgB dimers, underlining the differences in charge distribution. The CsgB–CsgA dimer had the fewest repulsive pairs, due to opposite overall charges of CsgB and CsgA. While CsgA–CsgA and CsgB–CsgA had similar representations of residue pairs contributing most strongly, CsgB–CsgB contained markedly more pairs containing oppositely charged residues. These differences highlight the varied pairwise interactions between CsgA and CsgB subunits arising from differences in sequence chemistry, despite similar geometry.

When decomposing the separation PMF for each dimer, similar interaction trends can be seen acting along the reaction coordinate. CsqB-CsqA was the only dimer with strong electrostatic contributions to dimerization, as was predicted initially due to its overall opposite charges. CsgB-CsgB had a much smaller electrostatic contribution, which was guickly screened as the interface separated. This small attractive electrostatic contribution near the bound state is most likely due to the large number of charged residues at the interface. CsgA-CsgA had the weakest nonbonded interaction along the reaction coordinate and in the bound state, which was balanced by strong solvation effects contributing to dimerization. Energy decomposition was also performed on mutated versions of CsgA-CsgA dimers to investigate the role of gatekeeper residues and can be found in Figure S11. Removing all seven gatekeeper residues decreased the electrostatic repulsion between separated monomers and increased the solvation energy.

Using computational modeling and simulation, we have presented structural models for CsgA and CsgB dimers and performed equilibrium and nonequilibrium MD to shed light on the nuanced differences between CsgA and CsgB. We find that, although protein geometry is similar, the difference in amino acid distribution gives rise to differences in binding energy. As CsgA and CsgB have overall opposite charges, CsgB-CsgA dimers predictably have the strongest binding energy. We were able to obtain quantitative values for the binding energy and estimations for the Young's modulus and persistence length of each dimer. We find that the absolute binding energies for CsgB-CsgA = 8.2 kcal/ mol, CsqA-CsqA = 4.0 kcal/mol, and CsqB-CsqB = 3.1 kcal/mol. All three dimers had a Young's modulus of ~43 GPa and a persistence length of ~2.0 µm. Intriguingly, although CsgB–CsgB had stronger nonbonded interactions between monomers than CsgA-CsgA dimers, CsgA-CsgA had a stronger binding energy. This hints at the importance of not only electrostatic and VDW contributions but also solvent effects. By decomposing the separation PMF into protein-protein and protein-solvent interactions, the long-range effects of solvation were elucidated. The protein-solvent contribution to binding free energy reaches the maximum at the bound state when the last of the residues have been dehydrated. We find charged residues contribute strongly to attractive interactions within CsgB-CsgB, while charged interactions in CsgA-CsgA contribute less and can also be repulsive. Particularly, charged, terminal, and polar residues could elicit strong interaction energies between dimers. Overall, these findings provide additional detail about the dimers involved in curli biogenesis, the driving forces behind their assembly, and the subtle differences between CsgA and CsgB. For future mechanical studies, these dimer structures provide a starting point for atomistic studies at the oligomer/fibril scale. To develop models on larger length scales, the values derived from this analysis can readily be used for coarse-grained modeling.

## METHODS

#### Dimer structures

The CsgA and CsgB subunit structures used were predicted in our previous work.<sup>30</sup> Subunits were initially placed in proximity to each other such that their beta-strands are perpendicular to the fiber axis and subjected to implicit solvent simulations to allow docking to occur. Briefly, two subunit structures were positioned such that strands R1 and R5 were aligned, strands R1' and R5' were aligned, and internally conserved polar residues (Ser, Gln, and Asn) maintained alignment across the interface. The initial space between adjacent subunits was about 5–7 Å wide. For CsgB–CsgA dimers, the CsgB was the bottom monomer because it is necessary for membrane association<sup>9</sup> and CsgA nucleates atop CsgB.

While building the dimer structures, all possible alignments were attempted (i.e., N–N-terminus stacking, C–C-terminus stacking, N–C-terminus stacking, for all combinations of CsgA or CsgB). The set-ups that were not included in this paper were found to have an inadequate number of hydrogen bonds across the interface, with some alignments forming hydrogen bonds across only one strand.

In the implicit solvent simulations, Generalized Born Implicit Solvent was used as implemented in NAMD.<sup>42</sup> The alpha cutoff used was 12.0 Å and the ion concentration 0.2 molar, chosen to ensure protein stability,42 with solvent accessible surface area (SASA) calculations on. The SASA calculation is used to calculate nonpolar/hydrophobic energy from the implicit solvent. In these simulations, the alpha carbons of the bottom monomer amyloid core (residues >22) were constrained by soft springs ( $k = 0.05 \text{ kcal/mol/Å}^2$ ). Structures were first minimized for 12,000 steps and run for at least 35 ns, at a 1 fs/step timestep. All simulation energy minimizations were performed using the conjugate gradient method. Docked structures were extracted from these simulations when at least 12 backbone hydrogen bonds existed at the R1/R5 interface between the two monomers. The implicit solvent simulations were used to obtain dimer structures, and trajectory information was not involved in any analysis presented here.

#### Equilibrium simulations

Dimer structures obtained through implicit simulations were then solvated with TIP3P<sup>44</sup> water molecules and ionized such that the net charge is zero.<sup>45</sup> For CsgA–CsgA dimers, 12 sodium ions are needed. For CsgB–CsgB dimers, six chlorine ions neutralize the system, and for CsgB-CsgA dimers, three sodium ions are required. Explicit solvent simulations were also run in NAMD, with periodic boundary conditions under the NPT ensemble at a constant pressure of 1 atm and temperature of 300 K. The latest CHARMM 36 parameter set is used, with the particle mesh Ewald technique for electrostatics calculations and the standard Lennard-Jones potential for nonbonded interactions.<sup>46</sup> Prior to production simulations, models underwent energy minimization and equilibration. For each model, first a simulation with all alpha carbons fixed was conducted to allow side-chain relaxation. An energy minimization of 10,000 steps was conducted, followed by 1 ns equilibration. Next, each system underwent 1 ns equilibration with only the alpha carbons of the amyloid core (residues >22) of the bottom monomer lightly restrained (spring constant of 0.05 kcal/mol/Å<sup>2</sup>). In lieu of using multiple short simulations, one simulation of each set-up was run for 50 ns, with simulation data recorded every 10 ps.

#### Free energy calculations

Each dimer system was solvated identically to the equilibrium simulations, with a minimum of 12 Å to the edge of the periodic box. Initial structures were chosen that had no interacting N-termini. The entire system was minimized for 50,000 steps using the conjugate gradient method and equilibrated for a minimum of 1 ns while the beta-sheet carbonyl carbons (all carbonyl carbons excluding turns or unstructured terminus regions) were harmonically restrained with a 2 kcal/mol/Å<sup>2</sup> spring. Then all

restraints were removed for an additional 0.5 ns of equilibration. This system was used as the starting configuration for all orientational and positional eABF simulations. This equilibrated protein structure was then copied and solvated in three additional boxes, one containing the dimer with 30 Å of length added in the direction of the separation reaction coordinate, one containing only the bottom monomer and one containing only the top monomer. For each set-up, the protein structure was fixed in place as the entire system was minimized for 50,000 steps and equilibrated for 1 ns. These additional systems were constructed to perform the REUS simulation, quantify the effect of removing conformational restraints on the top monomer in bulk with eABF, and quantify the effect of removing conformational restraints on the top monomer in bulk with eABF.

To determine the protein-protein binding energy of our system, a methodology very similar to Gumbart et al. was used.<sup>35</sup> A series of conformational, positional and orientational restraints were applied to improve convergence of the PMF between the two proteins as they are separated. The orientation and position of the two proteins are fully defined with three points per protein as shown in Fig. 1, where P1a and P2a are the location of the center of mass of the beta-sheet region carbonyl carbons in each respective protein. P1b, P2b, P1c, and P2c were chosen to minimize alignment to neighboring beads and are measured from the center of mass of four carbonyl carbons in the amyloid core (see Table S2 for details on each dimer). Four conformational restraints total, each with a force constant of 100 kcal/mol/Å<sup>2</sup> are applied to the interface and backbone of each protein to limit the RMSD of each. Restraints applied to the orientational and positional components had a force constant of 1 kcal/mol/deg<sup>2</sup>. Each orientational and positional restraint equilibrium angle/dihedral is listed in Table S3. In total, two backbone conformational restraints ( $\Delta G_{P1,b} \Delta G_{P2,b}$ ), two interface conformational restraints ( $\Delta G_{P1,i}$  $\Delta G_{P2,i}$ ), three orientational restraints ( $\Delta G_{\Theta}$ ,  $\Delta G_{\Psi}$ ,  $\Delta G_{\Psi}$ ), and two positional restraints  $(\Delta G_{\theta}, \Delta G_{\phi})$  were applied sequentially in the bound state and removed sequentially in the unbound state to account for the contribution of each to the absolute binding energy. The PMF for each of these restraints were generated using  $eABF^{47}$  and the CZAR free energy estimator<sup>33</sup> and can be seen in Figures S5, S6, and S7. The binding constant can be calculated with the equation,  $\mathcal{K}_{eq} = \mathsf{S}^* \mathsf{I}^* e^{-\beta \left[ \left( \mathsf{G}_{\mathsf{P2},b}^{\mathsf{bulk}} - \mathsf{G}_{\mathsf{P2},b}^{\mathsf{site}} \right) + \left( \mathsf{G}_{\mathsf{P1},b}^{\mathsf{bulk}} - \mathsf{G}_{\mathsf{P1},i}^{\mathsf{site}} \right) + \left( \mathsf{G}_{\mathsf{P2},i}^{\mathsf{bulk}} - \mathsf{G}_{\mathsf{P2},i}^{\mathsf{site}} \right) + \left( \mathsf{G}_{\mathsf{P1},i}^{\mathsf{bulk}} - \mathsf{G}_{\mathsf{P1},i}^{\mathsf{site}} \right) + \left( \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{site}} \right) + \left( \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} \right) + \left( \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bulk}} - \mathsf{G}_{o}^{\mathsf{bu$ where o and a denote the orientational and positional restraints, respectively.  $S^*$  is an integral over the positional restraints  $\boldsymbol{\theta}$  and  $\boldsymbol{\varphi}$  to account for the one-dimensional pathway taken from the binding site to bulk. I\* is an integral over the separation reaction coordinate r.

The separation PMF was generated with REUS. The coordinates of 48 replicas were generated from an explicit solvent-steered MD simulation along the reaction pathway moving at 1 Å/ns. The replicas were evenly spaced 0.5 Å apart with a 54 kcal/mol/Å<sup>2</sup> harmonic spring constant applied to the first 10 replicas, which is sufficient to span the transition region. The remaining replicas had a spring constant of 4 kcal/mol/Å<sup>2</sup>. The REUS simulations were run for a minimum of 3 ns/window to ensure convergence.

#### Free energy decomposition

The separation PMF was decomposed into electrostatic, Van der Waals, and protein–solvent contributions using the molecular mechanics Poisson–Boltzmann surface area technique. This was done using the Calculation of Free Energy tool,<sup>48</sup> which efficiently implements NAMD,<sup>42</sup> visual molecular dynamics (VMD),<sup>49</sup> and APBS<sup>50</sup> to calculate the molecular mechanics, non-polar solvation component by surface area and polar solvation component by the Poisson–Boltzmann equation, respectively. A trajectory (minimum of 0.2 ns long) from each of the 48 windows was used to calculate the contributions along the reaction coordinate.

#### Analysis

Simulations were visualized using VMD and analyzed using tcl scripts in VMD.<sup>49</sup> Hydrogen bond calculations were calculated for backbone atoms using a 4 Å distance cutoff and 30 degree angle cutoff. All secondary structure assessment was calculated using the STRIDE<sup>51</sup> algorithm. iRMSD was computed by calculating the RMSD of each frame from the first simulation frame. For this calculation, only backbone atoms were included, and the N-terminal 22 residues were excluded as these are unstructured.

#### Energetic analysis

Interaction energies were computed using the NAMDEnergy plugin. This calculates the VDW and electrostatic energies between groups of atoms. Results were averaged over the entire course of the trajectory. In these calculations, a dielectric constant is assumed in the space between the two groups of atoms, which was uniformly applied as 1.

## DATA AVAILABILITY

The datasets generated during the current study are available from the corresponding author on reasonable request.

### ACKNOWLEDGEMENTS

This research was sponsored by an award from the Office of Naval Research Young Investigator Program (grant #N00014-15-1-2701). E.D.B. was additionally sponsored by the DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. The authors acknowledge a supercomputing grant from the Northwestern University High Performance Computing Center and the Department of Defense Supercomputing Resource Center. E.D.B. gratefully acknowledges support from the Ryan Fellowship and the Northwestern University International Institute for Nanotechnology.

#### AUTHOR CONTRIBUTIONS

All authors designed and conceptualized the research, analyzed and interpreted the results, and wrote the paper. M.D. and E.D.B. performed simulations and data processing.

#### **ADDITIONAL INFORMATION**

**Supplementary information** accompanies the paper on the *npj Computational Materials* website (https://doi.org/10.1038/s41524-019-0164-5).

Competing interests: The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### REFERENCES

- Dueholm, M. S., Albertsen, M., Otzen, D. & Nielsen, P. H. Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. *PLoS ONE* 7, e51274 (2012).
- Cegelski, L. et al. Small-molecule inhibitors target *Escherichia coli* amyloid biogenesis and biofilm formation. *Nat. Chem. Biol.* 5, 913 (2009).
- Prigent-Combaret, C. et al. Developmental pathway for biofilm formation in curliproducing Escherichia coli strains: role of flagella, curli and colanic acid. *Environ. Microbiol.* 2, 450–464 (2001).
- Gophna, U. et al. Curli fibers mediate internalization of *Escherichia coli* by eukaryotic cells. *Infect. Immun.* 69, 2659 (2001).
- Vidakovic, L., Singh, P. K., Hartmann, R., Nadell, C. D. & Drescher, K. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nat. Microbiol.* 3, 26 (2018).
- Kikuchi, T., Mizunoe, Y., Takade, A., Naito, S. & Yoshida, S. Curli fibers are required for development of biofilm architecture in *Escherichia coli* K-12 and enhance bacterial adherence to human uroepithelial cells. *Microbiol. Immunol.* 49, 875–884 (2013).
- Gallo, PaulM. et al. Amyloid-DNA composites of bacterial biofilms stimulate autoimmunity. *Immunity* 42, 1171–1184 (2015).
- Prigent-Combaret, C. et al. Developmental pathway for biofilm formation in curliproducing Escherichia coli strains: role of flagella, curli and colanic acid. *Environ. Microbiol.* 2, 450–464 (2000).
- Bian, Z. & Normark, S. Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli. EMBO J. 16, 5827–5836 (1997).
- Loferer, H., Hammar, M. & Normark, S. Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG. *Mol. Microbiol.* 26, 11–23 (1997).
- Schubeis, T. et al. Structural and functional characterization of the Curli adaptor protein CsgF. FEBS Lett. 592, 1020–1029 (2018).

- Klein, R. D. et al. Structure-function analysis of the curli accessory protein CsgE defines surfaces essential for coordinating amyloid fiber formation. *mBio* 9, e01349–18 (2018).
- Epstein, E. A., Reizian, M. A. & Chapman, M. R. Spatial clustering of the curlin secretion lipoprotein requires curli fiber assembly. *J. Bacteriol.* **191**, 608–615 (2009).
- Shu, Q. et al. The *E. coli* CsgB nucleator of curli assembles to β-sheet oligomers that alter the CsgA fibrillization mechanism. *Proc. Natl Acad. Sci.* **109**, 6502–6507 (2012).
- 15. Chapman, M. R. et al. Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295**, 851–855 (2002).
- Wang, X., Smith, D. R., Jones, J. W. & Chapman, M. R. In vitro polymerization of a functional Escherichia coli amyloid protein. *J. Biol. Chem.* 282, 3713–3719 (2007).
- Dueholm, M. S. et al. Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation. *Biochemistry* 50, 8281–8290 (2011).
- Hammer, N. D. et al. The C-terminal repeating units of CsgB direct bacterial functional amyloid nucleation. J. Mol. Biol. 422, 376–389 (2012).
- 19. Yuan, P. Characteristics and Chemical Properties of Escherichia coli Curli. Doctoral Thesis, Technische Universität Braunschweig (2015).
- Nagaraj, M. Structural Analysis of Nucleation Mechanism in Curli Biogenesis using Nuclear Magnetic Resonance Spectroscopy. Doctoral Thesis, Technische Universität Braunschweig (2013).
- Hammar, M., Bian, Z. & Normark, S. Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli. Proc. Natl Acad. Sci.* **93**, 6562–6566 (1996).
- Hammer, N. D., Schmidt, J. C. & Chapman, M. R. The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proc. Natl Acad. Sci.* **104**, 12494–12499 (2007).
- Wang, X. & Chapman, M. R. Sequence determinants of bacterial amyloid formation. J. Mol. Biol. 380, 570–580 (2008).
- Wang, X., Zhou, Y., Ren, J.-J., Hammer, N. D. & Chapman, M. R. Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis. *Proc. Natl Acad. Sci.* **107**, 163 (2010).
- Wang, H., Shu, Q., Frieden, C. & Gross, M. L. Deamidation slows curli amyloidprotein aggregation. *Biochemistry* 56, 2865–2872 (2017).
- Swasthi, H. M. & Mukhopadhyay, S. Electrostatic lipid-protein interactions sequester the curli amyloid fold on the lipopolysaccharide membrane surface. J. Biol. Chem. 292, 19861–19872 (2017).
- 27. Sleutel, M. et al. Nucleation and growth of a bacterial functional amyloid at single-fiber resolution. *Nat. Chem. Biol.* **13**, 902 (2017).
- Oh, Y. J. et al. Curli mediate bacterial adhesion to fibronectin via tensile multiple bonds. *Sci. Rep.* 6, 33909 (2016).
- Abdelwahab, M. T., Kalyoncu, E., Onur, T., Baykara, M. Z. & Seker, U. O. S. Genetically-tunable mechanical properties of bacterial functional amyloid nanofibers. *Langmuir* 33, 4337–4345 (2017).
- DeBenedictis, E., Ma, D. & Keten, S. Structural predictions for curli amyloid fibril subunits CsgA and CsgB. RSC Adv. 7, 48102–48112 (2017).
- DeBenedictis, E. P., Liu, J. & Keten, S. Adhesion mechanisms of curli subunit CsgA to abiotic surfaces. Sci. Adv. 2, e1600998 (2016).
- Zhang, Y., Wang, A., DeBenedictis, E. P. & Keten, S. Bending energy penalty enhances the adhesive strength of functional amyloid curli to surfaces. *Nanotechnology* 28, 464002 (2017).
- Lesage, A., Lelièvre, T., Stoltz, G. & Hénin, J. Smoothed biasing forces yield unbiased free energies with the extended-system adaptive biasing force method. J. Phys. Chem. B 121, 3676 (2017).
- Woo, H.-J. & Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl Acad. Sci.* **102**, 6825–6830 (2005).

- Gumbart, J. C., Roux, B. & Chipot, C. Efficient determination of protein-protein standard binding free energies from first principles. J. Chem. Theory Comput. 9, 3789–3798 (2013).
- Tanner, D. E., Chan, K.-Y., Phillips, J. C. & Schulten, K. Parallel generalized born implicit solvent calculations with NAMD. J. Chem. Theory Comput. 7, 3635–3642 (2011).
- Fitzpatrick, A. W., Park, S. T. & Zewail, A. H. Exceptional rigidity and biomechanics of amyloid revealed by 4D electron microscopy. *Proc. Natl Acad. Sci.* 110, 10976–10981 (2013).
- Knowles, T. P. et al. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318**, 1900 (2007).
- Smith, J. F., Knowles, T. P. J., Dobson, C. M., MacPhee, C. E. & Welland, M. E. Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Natl Acad. Sci.* 103, 15806 (2006).
- 40. Adamcik, J. et al. Measurement of intrinsic properties of amyloid fibrils by the peak force QNM method. *Nanoscale* **4**, 4426–4429 (2012).
- Heinz, L. P., Ravikumar, K. M. & Cox, D. L. In silico measurements of twist and bend moduli for β-solenoid protein self-assembly units. *Nano Lett.* **15**, 3035–3040 (2015).
- Phillips, J. C. et al. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26, 1781–1802 (2005).
- Ibragimova, G. T. & Wade, R. C. Importance of explicit salt ions for protein stability in molecular dynamics simulation. *Biophys. J.* 74, 2906–2911 (1998).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- MacKerell, A. D. Jr et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616 (1998).
- 46. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi$ 1 and  $\chi$ 2 dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
- Fu, H., Shao, X., Chipot, C. & Cai, W. Extended adaptive biasing force algorithm. an on-the-fly implementation for accurate free-energy calculations. *J. Chem. Theory Comput.* 12, 3506–3513 (2016).
- Liu, H. & Hou, T. CaFE: a tool for binding affinity prediction using end-point free energy methods. *Bioinformatics* 32, 2216–2218 (2016).
- Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38 (1996).
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci.* 98, 10037–10041 (2001).
- Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Protein* 23, 566–579 (1995).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons. org/licenses/by/4.0/.

© The Author(s) 2019