



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Collagen-Inspired Self-Assembly of Twisted Filaments

Martin J. Falk, Amy Duwel, Lucy J. Colwell, and Michael P. Brenner

Phys. Rev. Lett. **123**, 238102 — Published 4 December 2019

DOI: [10.1103/PhysRevLett.123.238102](https://doi.org/10.1103/PhysRevLett.123.238102)

# Collagen-inspired self-assembly of twisted filaments

Martin J. Falk,<sup>1</sup> Amy Duwel,<sup>2</sup> Lucy J. Colwell,<sup>3</sup> and Michael P. Brenner<sup>4</sup>

<sup>1</sup>*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02138*

<sup>2</sup>*Charles Stark Draper Laboratory, Cambridge, MA 02138*

<sup>3</sup>*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge UK CB2 1EW*

<sup>4</sup>*School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138*

Collagen consists of three peptides twisted together through a periodic array of hydrogen bonds. Here we use this as inspiration to find design rules for programmed specific interactions for self assembling *synthetic* collagen-like triple helices, starting from disordered configurations. The assembly generically nucleates defects in the triple helix, the characteristics of which can be manipulated by spatially varying the enthalpy of helix formation. Defect formation slows assembly, evoking kinetic pathologies that have been observed to mutations in the primary collagen amino acid sequence. The controlled formation and interaction between defects gives a route for hierarchical self-assembly of bundles of twisted filaments.

There have been dramatic developments in functionalizing submicron scale objects with molecules enabling specific interactions, creating the opportunity to program the assembly of diverse structures. However, the structures that can be synthetically assembled still pale relative to those in biology. This gap between synthetic and natural self-assembly is surely in part due to the fact that nature has had more time to search through the parameter space of possible interactions to find clever design solutions. One approach towards learning how these solutions work is to try to mimic biological assembly strategies for specific structures.

Here we use this approach to study the self-assembly of nanoscale filamentous bundles. Such twisted bundles could find uses ranging from the fibers that comprise fabrics to Litz wires found in high-frequency electrical transmission cables [1, 2]. To robustly control the assembly of twisted filaments, we use a design scheme inspired by collagen, which is the most abundant protein on the planet, forming a hierarchical bundle of bundles. Collagen initially assembles into a triple helix consisting of three peptides, held together with hydrogen bonds. Although the amino acid sequence leading to this design varies both across its length and across species, x-ray crystallography suggests that the structure is quite regular[3–5]. With sequence and structure information, one might expect collagen formation to be a completely solved problem. However, while efforts from synthetic biology to experimentally produce collagen-like materials have had success[6, 7], we are still not able to fully control the formation of collagen [3].

The goal of this paper is explore a computational model of synthetic triple helix formation, to uncover the basic design rules of collagen assembly. To this end, we propose a scheme for the self-assembly of polymers into a triple helix, mediated by pair-wise specific interactions between the monomers. Each polymer consists of a string of monomers, whose surfaces are uniformly coated with a specific glue (e.g. DNA). This causes short-ranged, specific interactions between particles, with a programmable

binding strength between pairs of particles. Such systems have been previously shown to allow robust assembly of complex structures[8, 9], including systems which self-replicate[10, 11] and metabolize[12]. By choosing interactions between monomers to mimic collagen, our numerical simulations show robust formation of a long triple-helical ground state. However, if the interactions between monomers are isotropic and without internal chirality, the handedness of the helix can only be controlled only within a certain correlation length, arising from the switching of two filaments during the assembly process. Other forms of defects slow assembly, evoking kinetic pathologies previously ascribed to mutations in the primary collagen amino acid sequence. We show through explicit examples that it is possible to control the nucleation and location of defects.

To proceed, we carry out numerical simulations using a patterning that is inspired by collagen (Fig 1A). In collagen, each peptide is a repeating chain of three amino acids: G X Y, where G is always glycine, but X and Y vary across the length of a single collagen molecule, and across species[3]. Crystal structures show that G and Y form hydrogen bonds with their counterparts on the other peptides involved in the triple helix[3–5], whereas glycine provides structural flexibility and satisfies packing constraints, to enforce the stagger of the three collagen strands relative to each other[7]. The major difference between the collagen patterning and our colloidal system is that hydrogen bonds require a specific alignment between the atoms involved[13–15], which is not realizable with uniformly coated particles. Instead, we consider a scheme where the hydrogen-bonded amino acids are replaced with three flavors of monomers; A, B, and C, which are sticky; and one additional type of monomer, an inert monomer. Rather than using the non-uniform bending energy implied by the glycines in collagen, we simply add more inert monomers to each motif (Fig 1A), thereby increasing chain flexibility in a simple fashion. It is also known that collagen formation requires a nucleation site that holds three peptides together[16].

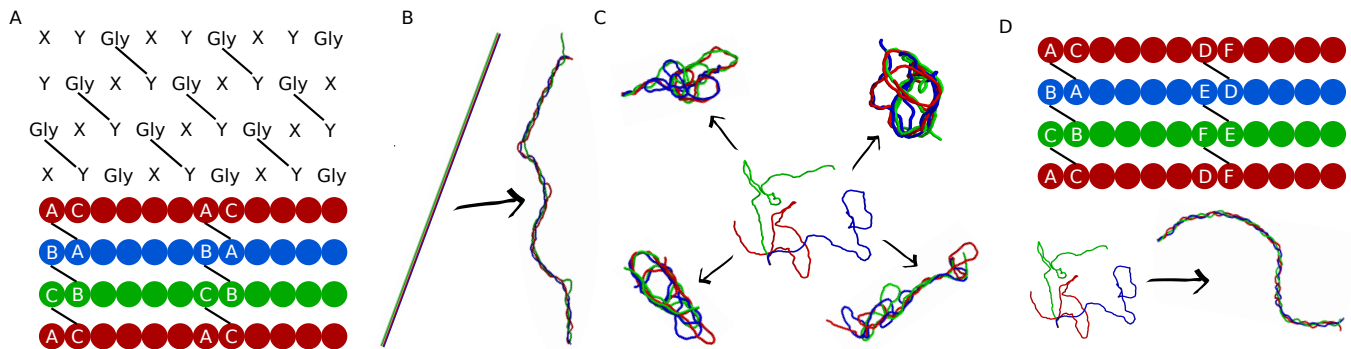


Figure 1. **Robust twisted filaments from a collagen-inspired scheme** A. Schematic showing three collagen peptides, and the network of hydrogen bonds that hold them together. We create an analogous network of bonds in our self-assembly schematic below, where each polymer is colored based on the different sequences of three stickers (A, B, and C) on the polymer. B. Initial condition and final configuration for a simulation showing successful self-assembly of a collagen-like triple helix, based on the schematic in A. C. Disordered configurations almost universally lead to a variety of loops and rackets lacking a triple-helical geometry. D. Schematic showing the maximal alphabet version of our self-assembly scheme, where each folding unit is comprised of a unique set of stickers. Crosstalk is prevented between different folding units, and helix formation robustly occurs.

This is incorporated into our simulations by binding the monomers at the end of each polymer chain to each other.

We simulate the assembly with Langevin dynamics using a publicly-available wrapper of OpenMM [17, 18], OpenMM-polymer (<http://bitbucket.org/mirnylab/openmm-polymer>).

Simulations were run at room temperature ( $T = 300K$ ). To generate initial configurations for simulations, we take three polymers aligned parallel to each other, held together on one end in a triangle via harmonic bonds. The polymers have 180 monomers, with 30 folding units. Each folding unit consists of 6 monomers, with two sticky monomers followed by four inert monomers. Simulations were run for  $3.5 \times 10^6$  time steps without the influence of any attractive potentials to generate disordered starting configurations for simulations with attractive or chiral potentials. At this point, any chiral and attractive energies were added to the simulation. The functional forms of the energies used are detailed in the Methods section.

With this model, we find that the formation of a triple helix is possible (Fig 1B). However, robust assembly only occurs for special initial conditions; random initial conditions rarely folds into a triple helix. The failure modes include polymers forming loops and rackets (Fig 1C), with monomers that simultaneously interact with multiple other monomers. In contrast, collagen robustly avoids these misfolded configurations, in large part because of unidirectional, monovalent, hydrogen bonds, a feature which has also been noted in the context of designing synthetic heteropolymers with specific folds[19–21].

There are several potential strategies to avoid yield-destroying configurations. Yield can be improved in limited circumstances without introducing valency and directionality. Monomers within a persistence length only

infrequently come into contact, so one strategy is to completely change the set of chemical glues every persistence length to prohibit these multiple contacts. This makes connections needed for forming loops more costly due to the increased elastic energy of a loop, and so yield improves slightly (Sup Fig 1). This strategy ultimately provided limited benefit in terms of increasing yield while keeping the number of sticker types small (Sup Fig 1).

Alternatively, we can achieve a robust self-assembly scheme with a number of glues that does not grow with the length of the helix if we introduce directionality to our scheme. By functionalizing only a sector of the polymer, we introduce valency via packing constraints, and the fraction of correctly assembled structures can approach unity (Sup Fig 2). However, the individual filaments in the triple helix can stick together out of alignment. We suggest that misregistration can be prevented by breaking translational symmetry along the filament and introducing distinct linker points where the filaments must align, still with a sub-extensive number of glues. Such a scheme would require at least four distinct glues: three for the binding, and at least one for breaking translational symmetry along the fiber. Four glues is comparable to numbers seen in recent experimental work[22].

We turned to a simple scheme in which every set of sticky monomers was identified with a distinct set of glues. Individual glues serve as a proxy for the more complicated methods by which non-helical configurations might be removed. This scheme is equivalent to the Go model in protein folding [23] or the maximal alphabet in colloidal self-assembly simulations [8]. As anticipated, simulations of our scheme with a maximal alphabet case robustly assemble a triple-helical ground state even if assembly begins from disordered initial conditions (Fig 1D).

Within this model, the pitch, handedness, and radius of the helix are encoded by the interaction parameters. We find that the magnitude of the pitch can be controlled via the spacing of stickers along the contour length of the bundle (Fig 2A). The longer the gap between stickers, the more the filaments are able to relax from the twist created at the stickers, and hence the pitch decreases.

The radius of the helix measures whether or not the helix has successfully formed; for successfully assembled helices, the radius of the helix is small and nearly constant across the entire length of the filament. However, at fixed bending stiffness, if the sticker strength is too low, the helix does not form (Fig 2B).

The sign of the pitch has a more complicated dependence on material parameters. Recall that in mapping our scheme from collagen to a system of synthetic fibers functionalized with isotropic specific glues, it was necessary to break symmetry between the filaments in the pattern of interactions. As such, we might expect our bundles to form helices of a single hand along their entire lengths. However, this is not the case, as can be seen by computing the decay in correlation of chirality between folding units as a function of distance (Fig 2E).

Indeed, when two polymers are spatially permuted at a particular location along the bundle, the handedness of the energetically favored twist switches (Fig 2C). This results in the formation of regions where the helix is left- or right-handed, with domain walls of undefined chirality (perversions) necessarily appearing between the two. We can control the characteristic length of this decay by increasing the bending stiffness of our three polymers at fixed sticker strength, with decreasing bending stiffness corresponding to a decreasing chirality correlation length (Fig 2E). This is likely due to the fact that chiral defects require polymers to cross, and the statistics of those crossings are set by the bending stiffness of the polymers.

In our simulations, we have so far neglected any energetic penalty related to chirality; these must be present in collagen since twists form of only one handedness[3]. We implement a chiral energy penalty with an interaction between monomers depending on the dihedral angle between adjacent polymers (Methods). In collagen, the chirality of amino acids conceivably provides a source for a chiral energy penalty. Penalizing configurations where this angle is negative causes left-handed twists to have a higher energy than right handed twists. Increasing the energy difference at fixed bending stiffness and sticker strength preferentially drives the system to right-handed twists, even at modest energetic differences between the two chiralities (Fig 2D).

When the material parameters spatially vary along the filaments, we open up the possibility of controlling the position of defects. One option is to spatially vary sticker strengths, allowing the creation of localized bubbles where the three filaments unbind; this introduces a

defect to the triple-helical ground state at a well-defined position (Fig 3A). This type of defect is structurally distinct from the chiral defects discussed above, and furthermore slows down the kinetics of self-assembly (Fig 3A, inset), similarly to defects implicated in osteogenesis imperfecta[24–26]. We can also localize chiral defects by spatially varying the chiral bias. This robustly creates sites along the bundle where chiral defects are likely to occur (Fig 3B). Similarly, the presence of a ‘weak’ spot in the bending stiffness of the bundle increases the probability of observing a chiral defect at that position (Fig 3C).

*Implications for Collagen Assembly.*—Our results, though based on a very different set of interactions than those found in collagen, nevertheless imply the presence of several constraints for collagen self-assembly. Firstly, the implicit valency of hydrogen bonding is crucial for robustly avoiding non-helical configurations in an assembly process driven solely by thermal fluctuations. Secondly, though the nature of our interaction pattern explicitly breaks chiral symmetry once a particular permutation of our three filaments has been fixed, no such chiral symmetry breaking occurs in collagen. Hence, if collagen is to form helices of only one handedness, it is crucial that collagen implements a bias between left-handed and right-handed configurations. It is possible to realize a Hamiltonian with this behavior if the monomeric constituents of natural collagen themselves break chiral symmetry; this occurs with amino acids, which generally tend to be right-handed, potentially leading to an energy gap between left-handed and right-handed supra-molecular structures. Alternatively, chiral symmetry breaking can occur via an elastic mechanism in coiled-coil proteins[27], which may also be relevant to the case of the three left-handed peptides which form collagen.

Our results also have parallels to the phenomenology of collagen folding in osteogenesis imperfecta (OI), commonly known as brittle bone disease. In this family of disorders, glycine normally becomes substituted for a larger amino acid, making the packing of the three strands locally more frustrated and less energetically favorable. While the distance of the substitution from the nucleation site and the identity of the amino acids around the substitution affect the lethality of the substitution, the overall effect is to produce structures similar to the localized bubble defects we observe in our artificial collagen assembly simulation when we create regions of lower binding energy[3, 4, 24, 26, 28]. Furthermore, the presence of these defects in our simulations delays the formation of the triple helix as it twists from the nucleation site, recapitulating what is thought to be an important causal feature of OI[24–26].

OI is a particularly dramatic manifestation of sequence variation in collagen. However, even in non-disease states, collagen’s variable amino acid sequence is known to translate to a variable stability and flexibility of the

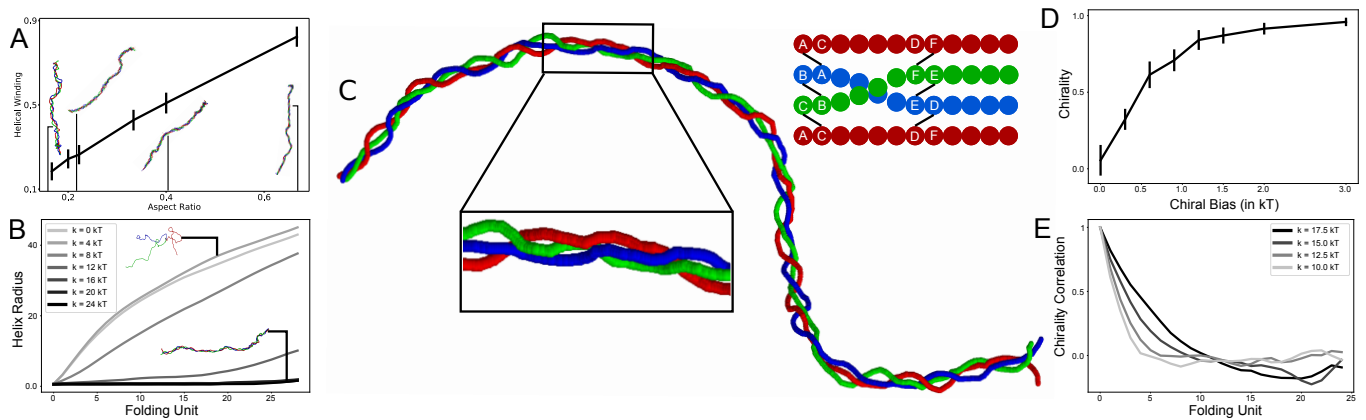


Figure 2. **Dependencies of helix geometry on material parameters** A. Pitch of the helix, as a function of the aspect ratio of the folding unit, defined as the ratio of the length of the sticky section of each folding unit and the length of each folding unit. The pitch generally increases as a function of aspect ratio, as can be seen from the snapshots taken from simulations at specific values of the aspect ratio. B. The correlation of the helix handedness as a function of distance along the filament. Length is measured in units of the repeats of the pattern. Decay is mediated by the presence of chiral defects where two strands switch their positions relative to the third strand, as shown in the inset of the simulation snapshot. The reason strand-switching corresponds to changes in chirality is illustrated in the schematic. C. Average radius of the helix, as a function of distance from the nucleation site. A successful assembly of a helical ground state implies a constant, small radius over the entire length of the bundle. D. Chirality of the filament as a function of chiral bias. Increasing the bias for right-handed configurations shifts the average chirality towards 1. E. Decay in the correlation of chirality along the filament.

triple helix[29–32]. These sequence variations are also thought to contribute to collagen’s ability to assemble many different hierarchical superstructures[33], potentially mediated by the same shifts in helix geometry that we have identified in our simulations[28].

*Implications for Nanoscale Assembly.*—Thinking beyond collagen, there are extensions of our work to bundles of more than three filaments. For instance, having  $n$  filaments changes the nature of the chiral defects, as permuting two filaments no longer simply reverses the handedness of the preferred twist direction.

There is additional work to be done in terms of devising strategies to avoid yield-destroying configurations without moving to the maximal alphabet regime. For example, the strategy we tested to reduce the number of stickers by repeating sets of stickers could be improved by designing optimal sequences in which to repeat them. We also tested schemes with directional interactions, and these could be improved by the random repetition of stickers to break translational symmetry. Approaches we did not consider in this paper would be to arrange the initial conditions of the self-assembly scheme to be more favorable to helix formation, to introduce an external forcing or energy in order to avoid loops, or to implement more geometrically complex pair-wise interactions.

In general, the technical challenges posed by our scheme are high. However, continued progress in engineering colloids with directional interactions and designed chiralities[22, 34], and varying degrees of rotational friction between bonds and colloidal substrate[35,

36], provides an ideal framework for implementation. For a scheme with directional interactions, the twist degree of freedom along the polymer backbone also becomes an important axis within the design space. We remark that the friction between bond and substrate is particularly important for exploration of the twist in a colloidal setting[37]. Given these advances, it is important that we theoretically explore how the interactions can be used to create technologically useful materials. This is the primary aim of the present study.

Returning to our collagen inspiration, the ability to localize defects may open opportunities for synthetic hierarchical self-assembly. As previously mentioned, collagen is thought to assemble a large number of different superstructures as a consequence of sequence variability. Correspondingly, the functionalization of unbound regions of the bundle may offer a pathway to a synthetic version of the hierarchical self-assembly biology has perfected.

*Conclusion*— In summary, we have articulated a scheme for self-assembling twisted helices. We find that successful self-assembly can be achieved with a limited number of stickers, but this requires either directionality or a high bending stiffness. A modest energetic bias between left- and right-handed twisting is sufficient to form helices of only one chirality. Spatial variation of chiral bias, stiffness, and sticker energy allows for the control of various defects. We point towards how these defects might play a role in collagen self-assembly, and suggest that strategies for the hierarchical self-assembly of synthetic filaments should follow similar lines.

This research was funded by the National Science

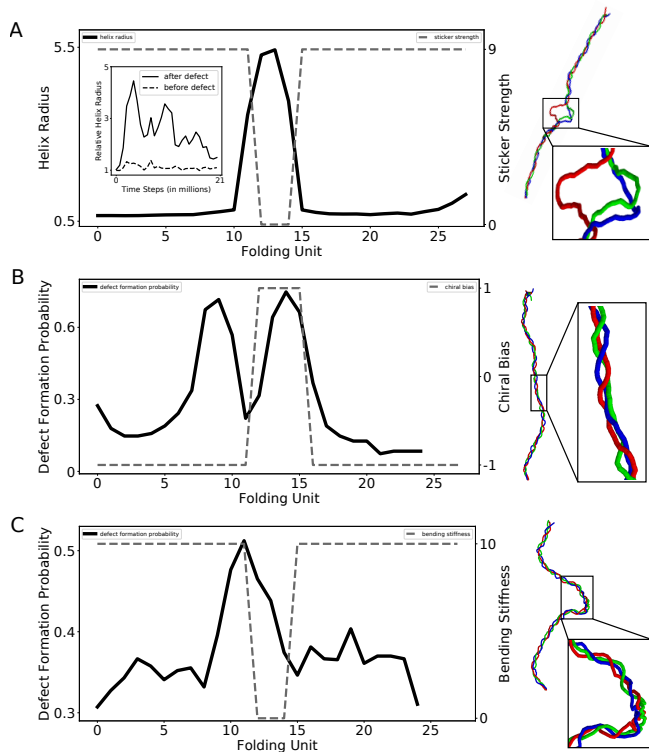


Figure 3. **Localization and kinetics of defects** A. Radius of the helix along the filament (solid line). Sticker strength along the filament (dotted line). The peak in average radius indicates local unbinding in the region of lower sticker strength, as shown in the simulation snapshot. (inset) We compare the helix radius of regions in simulations with bubble defects to similar regions in simulations without bubble defects. For regions before the defect (dotted line), there is no change in the progression of folding. For regions after the defect (solid line) there is an initial delay in folding. B. Defect formation probability along the filament, as measured by the fraction of simulations in which the folding unit three units ahead of the indicated position is of the opposite sign of the folding unit at the indicated position (solid line). Chiral bias as a function of distance along the filament (dotted line). The two peaks (solid line) correspond to the boundaries of the right-handed segment of the helix (dotted line). C. Defect formation probability (solid line) and bending stiffness (dotted line) along the filament. The peak (solid line) indicates that defects form across mechanically weaker regions with higher probability.

Foundation through DRAPER GRANT, the Harvard Materials Research Science and Engineering Center Grant DMR-1420570, DMREF Grant DMR-123869 and ONR Grant N00014-17-1-3029. MF acknowledges funding from the DoD through the NDSEG and A2P programs. MPB is an investigator of the Simons Foundation. We would like to thank Jorge Fallas for helpful discussions on the self-assembly of collagen-like peptides

and OI.

- [1] F. Tourkhani and P. Viarouge, *IEEE Trans. Magn.* **37**, 538 (2001).
- [2] C. R. Sullivan and L. W. Losses, in *2014 IEEE 15th Workshop on Control and Modeling for Power Electronics (COMPEL)* (2014) pp. 1–10.
- [3] M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.* **78**, 929 (2009).
- [4] J. Bella, M. Eaton, B. Brodsky, and H. M. Berman, *Science* **266**, 75 (1994).
- [5] A. Rich and F. H. Crick, *J. Mol. Biol.* **3**, 483 (1961).
- [6] J. A. Fallas, L. E. R. O’Leary, and J. D. Hartgerink, *Chem. Soc. Rev.* **39**, 3510 (2010).
- [7] J. A. Fallas and J. D. Hartgerink, *Nat. Commun.* **3**, 1087 (2012).
- [8] Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, *Proceedings of the National Academy of Sciences* **111**, 15918 (2014).
- [9] A. V. Tkachenko, *Phys. Rev. Lett.* **106**, 255501 (2011).
- [10] Z. Zeravcic and M. P. Brenner, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1748 (2014).
- [11] H. Tanaka, Z. Zeravcic, and M. P. Brenner, *Phys. Rev. Lett.* **117**, 238004 (2016).
- [12] Z. Zeravcic and M. P. Brenner, *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4342 (2017).
- [13] P. A. Kollman, *J. Am. Chem. Soc.* **94**, 1837 (1972).
- [14] J. Ireta, J. Neugebauer, and M. Scheffler, *J. Phys. Chem. A* **108**, 5692 (2004).
- [15] G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond: In Structural Chemistry and Biology* (Oxford University Press, 2001).
- [16] D. J. Prockop and K. I. Kivirikko, *Annu. Rev. Biochem.* **64**, 403 (1995).
- [17] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, B. R. Brooks, and V. S. Pande, “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics,” (2016).
- [18] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, and V. S. Pande, *J. Chem. Theory Comput.* **9**, 461 (2013).
- [19] C. Cardelli, V. Bianco, L. Rovigatti, F. Nerattini, L. Tubiana, C. Dellago, and I. Coluzza, *Sci. Rep.* **7**, 4986 (2017).
- [20] L. Feng, L.-L. Pontani, R. Dreyfus, P. Chaikin, and J. Brujic, *Soft Matter* **9**, 9816 (2013).
- [21] E. Bianchi, B. Capone, I. Coluzza, L. Rovigatti, and P. D. J. van Oostrum, *Phys. Chem. Chem. Phys.* **19**, 19847 (2017).
- [22] M. Y. Ben Zion, X. He, C. C. Maass, R. Sha, N. C. Seeman, and P. M. Chaikin, *Science* **358**, 633 (2017).
- [23] R. D. Hills, Jr and C. L. Brooks, 3rd, *Int. J. Mol. Sci.* **10**, 889 (2009).
- [24] T. J. Hyde, M. A. Bryan, B. Brodsky, and J. Baum, *J. Biol. Chem.* **281**, 36937 (2006).
- [25] M. Raghunath, P. Bruckner, and B. Steinmann, *J. Mol. Biol.* **236**, 940 (1994).

- [26] M. Bhate, X. Wang, J. Baum, and B. Brodsky, *Biochemistry* **41**, 6539 (2002).
- [27] S. Neukirch, A. Goriely, and A. C. Hausrath, *Phys. Rev. Lett.* **100**, 038105 (2008).
- [28] J. C. Marini, A. Forlino, W. A. Cabral, A. M. Barnes, J. D. San Antonio, S. Milgrom, J. C. Hyland, J. Kórkó, D. J. Prockop, A. De Paepe, P. Coucke, S. Symoens, F. H. Glorieux, P. J. Roughley, A. M. Lund, K. Kuurila-Svahn, H. Hartikka, D. H. Cohn, D. Krakow, M. Mottes, U. Schwarze, D. Chen, K. Yang, C. Kuslich, J. Troendle, R. Dalglish, and P. H. Byers, *Hum. Mutat.* **28**, 209 (2007).
- [29] A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick, and B. Brodsky, *Biochemistry* **44**, 1414 (2005).
- [30] A. V. Persikov, J. A. M. Ramshaw, and B. Brodsky, *J. Biol. Chem.* **280**, 19343 (2005).
- [31] W. Y. Chow, D. Bihan, C. J. Forman, D. A. Slatter, D. G. Reid, D. J. Wales, R. W. Farndale, and M. J. Duer, *Sci. Rep.* **5**, 12556 (2015).
- [32] W. Yang, M. L. Battineni, and B. Brodsky, *Biochemistry* **36**, 6930 (1997).
- [33] J. Khoshnoodi, J.-P. Cartailier, K. Alvares, A. Veis, and B. G. Hudson, *J. Biol. Chem.* **281**, 38117 (2006).
- [34] Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, and D. J. Pine, *Nature* **491**, 51 (2012).
- [35] L. Feng, L.-L. Pontani, R. Dreyfus, P. Chaikin, and J. Brujic, *Soft Matter* **9**, 9816 (2013).
- [36] A. McMullen, M. Holmes-Cerfon, F. Sciortino, A. Y. Grosberg, and J. Brujic, *Phys. Rev. Lett.* **121**, 138002 (2018).
- [37] J. P. Lee-Thorp and M. Holmes-Cerfon, *Soft Matter* **14**, 8147 (2018).