Article

Interaction Patterns for Staggered Assembly of Fibrils from Semiflexible Chains

Arnoud Jongeling [1], Carsten Svaneborg [2] and Renko de Vries [1],*

[1] Physical Chemistry and Soft Matter, Wageningen University & Research, Stippeneng 4,
6708 WE Wageningen, The Netherlands; arnoud.jongeling@uhasselt.be
[2] Department of Physics, Chemistry and Pharmacy, SDU, University of Southern Denmark, Campusvej 55,
DK-5230 Odense M, Denmark; science@zqex.dk
* Correspondence: renko.devries@wur.nl; Tel.: +31-317-484561

check for updates

Abstract: The design of colloidal interactions to achieve target self-assembled structures has especially been done for compact objects such as spheres with isotropic interaction potentials, patchy spheres and other compact objects with patchy interactions. Inspired by the self-assembly of collagen-I fibrils and intermediate filaments, we here consider the design of interaction patterns on semiflexible chains that could drive their staggered assembly into regular (para)crystalline fibrils. We consider semiflexible chains composed of a finite number of types of interaction beads (uncharged hydrophilic, hydrophobic, positively charged and negatively charged) and optimize the sequence of these interaction beads with respect to the interaction energy of the semiflexible chains in a number of target-staggered crystalline packings. We find that structures with the lowest interaction energies, that form simple lattices, also have low values of L/D (where L is chain length and D is stagger). In the low interaction energy sequences, similar types of interaction beads cluster together to form stretches. Langevin Dynamics simulations confirm that semiflexible chains with optimal sequences self-assemble into the designed staggered (para)crystalline fibrils. We conclude that very simple interaction patterns should suffice to drive the assembly of long semiflexible chains into staggered (para)crystalline fibrils.

Keywords: protein assembly; collagen; staggered assembly; protein design; self-assembly design; coarse-graining; computer simulations

1. Introduction

The design of soft materials that self-assemble into programmable structures with physical properties that can be precisely controlled, is now within reach thanks to technologies such as DNA nanotechnology [1], and the computational design of sequences for small globular proteins that self-assemble into precisely designed structures [2]. At larger length scales, and more coarse-grained levels of description, interaction potentials for soft matter components can be designed that assemble to stable target structures [3]. For example, isotropic interaction potentials have been designed that lead to the assembly of colloidal particles into quasi-crystals [4]. A large body of work, both experimental and theoretical, also deals with patchy colloidal particles, with anisotropic interaction potentials, which can be designed for assembly into, e.g., into monodisperse icosahedral clusters [5] or specific crystal structures [6].

Virtually no work has been carried out yet on patchy rods, or patchy semiflexible chains. No doubt this is related to the fact that as of yet there are not many experimental examples of such systems for which patterns on rods or semiflexible chains can be precisely controlled. However, in nature, patchy semiflexible chains are abundant, a particularly important example being those that exhibit staggered assembly into fibrils such as intermediate filaments and collagen-I. Both collagen fibrils

and intermediate filaments form by the staggered assembly of long, rod-shaped, semiflexible protein subunits. Presumably, some form of "interaction code" along the rod-shaped protein subunits directs the staggered assembly into fibrils. Indeed a regular pattern of electrostatic and other interactions for parallel triple helices has been implicated in the assembly of collagen-I triple helices [7,8], but only at a qualitative level, leaving the possibility that very local and specific interactions may in fact dominate the total driving force for the staggered assembly. The same holds for the staggered assembly of intermediate filaments: patterns of electrostatic interactions have been implicated, but only at a qualitative level [9]. An additional complication is that these fibrils assemble in the context of a living organism, such that it is not clear whether indeed the assembly is fully encoded in the primary sequence of these proteins, or that it is assembled under the influence of external forces. For example, the staggered assembly of collagen-I triple helices into fibrils and fibers is a highly regulated process that is influenced by a multitude of cellular processes. Indeed, while isolated collagen-I triple helices have been shown to be able to assemble into staggered fibrils in-vitro, showing the familiar "banding" pattern, structurally these fibrils are different from those formed in-vivo [10].

Nevertheless, the fact that the staggered assembly is also found for collagen-I in-vitro (albeit not the same as in-vivo) does indicate that it is indeed encoded to some extent in the primary sequence. Further evidence in this direction also comes from studies on the assembly of engineered concatemers of (non-hydroxylated) fragments of collagen-I sequences, that form staggered fibrils, but with different gap sizes than natural collagen-I fibrils [11–13].

Designed fibrillar systems have been explored widely as mimics of collagen-I fibrils of the extracellular matrix. A number of systems have been developed that consist of sticky semiflexible chains that bundle into fibrils and form dilute hydrogels [14–16]. These have shown great promise as scaffolds for cell growth. However, semiflexible chains that exhibit precise staggered assembly into (quasi) crystalline arrays have not been developed yet. For short, helical peptides, it has been discovered that simple patterns of rather long-range interactions, such as electrostatic interactions, suffice to direct staggered assembly into fibrils, showing the characteristic "banding" when imaged using electron microscopy. This was shown both for collagen-like triple-helical peptides [17] and for alpha-helical peptides [18].

The longer-ranged interactions that drive staggered assembly into fibrils for these short peptide chains can be fairly well represented by coarse-grained force-fields. Therefore, we reason that when aiming at the computational design of interaction patterns that lead to staggered self-assembly of long semiflexible chains, a fair starting point would be the use of a coarse-grained representation of the interaction patterns on the semiflexible chains. Needless to say, this leads to efficient computations for which one can screen a very large number of interaction patterns and target crystalline arrangements, even for long semiflexible chains. The resulting coarse-grained interaction patterns may be the starting point for more refined sequence optimization at less coarse-grained levels of descriptions, or they may be directly implemented using different kinds of chemistries such as collagen-like or alpha-helical amino-acid sequences, or, in the near future, possibly using other sequence-defined synthetic polymer chemistries [19]. With this in mind, we here consider the computational design of interaction patterns on semiflexible chains for their spontaneous staggered assembly into (para)crystalline fibrils.

2. Materials and Methods

The coarse-grained model for the patterned semiflexible chains is formulated in terms of 4 types of interaction beads, hydrophobic, uncharged hydrophilic, positively charged, and negatively charged. Note that the use of interaction beads is an approximation of at least two levels. First, the molecular structures that we have in mind such as peptide alpha-helices and triple helices are only approximately semiflexible rods, and secondly, additional errors are introduced by approximating semiflexible rods in terms of bead–spring models. Type numbers s associated with the bead types are: $s = 1$ (uncharged hydrophilic), $s = 2$ (hydrophobic), $s = 3$ (positively charged) and $s = 4$ (negatively charged). We consider M patterned semiflexible chains, each chain consists of $N = 40$ interaction beads, such that the total

number of interaction beads is $N_{tot} = M.N$. All semiflexible chains have the same sequence of bead types, $S = \{s_1, s_2, \ldots, s_N\}$, and all beads have the same diameter σ. Electrostatic interactions are represented by a screened Coulomb potential, hydrophobic attraction and excluded volume interactions by a Lennard–Jones potential. Hence, the non-bonded interaction between two beads i and j with bead types s_i and s_j is:

$$V_{s_i s_j}(r_{ij}) = 4\epsilon_{s_i s_j}\left((\sigma/r_{ij})^{12} - (\sigma/r_{ij})^6\right) + q_{s_i}q_{s_j}\frac{l_B}{r_{ij}}\exp(-\kappa r_{ij})k_B T \tag{1}$$

where $l_B = e^2/4\pi\epsilon_0\epsilon_r k_B T$ is the Bjerrum length, in terms of the elementary charge e, the absolute and relative dielectric permittivities ϵ_0 and ϵ_r, and the thermal energy $k_B T$. The Debye screening length is $\kappa^{-1} = (8\pi l_B n_s)^{-1/2}$, where n_s is the number density of (monovalent) ions. The effective bead charges q_{s_i} (number of elementary charges e) are:

$$q_{s_i} = \begin{cases} 0 & s_i = 1, 2 \\ +q & s_i = 3 \\ -q & s_i = 4 \end{cases} \tag{2}$$

and the Lennard–Jones interaction parameters are:

$$\epsilon_{s_i s_j} = \begin{cases} \epsilon_h & s_i = s_j = 2 \\ \epsilon_0 & \text{otherwise} \end{cases} \tag{3}$$

Bonded interactions used to describe the semiflexible chains are harmonic stretching- and bending potentials, such that the total potential energy function for the M chains is:

$$U(r_1, \ldots, r_{N_{tot}}) = \sum_{pairs\ i,j} V_{s_i s_j}(r_{ij}) + \frac{1}{2}k_s \sum_{bonds\ i,j}(r_{ij} - l_b)^2 + \frac{1}{2}k_b \sum_{bends\ i,j,k}\theta_{ijk}^2 \tag{4}$$

The bending angle θ_{ijk} is the angle between the $(i)\ldots(j)$ and $(j)\ldots(k)$ virtual bond vectors. The linear chains are assumed to be straight at equilibrium hence the equilibrium bending angle is zero, and $j = i + 1$, $k = i + 2$.

Parameter values are chosen such that they represent one of the possible chemistries that can be used to implement patterned semiflexible chains, viz., the collagen triple helix. The bead diameter σ, therefore, corresponds to the diameter of a collagen triple helix, of around 1.5 nm. Experimental values reported for the persistence length l_p of collagen triple helices cover a wide range of values [20,21].

We assume thermal bending fluctuations have a destabilizing effect on fibril assembly, therefore we here use a value of $l_p \approx 35$ nm which is on the lower side of the reported values. We do so since we would rather like to overestimate the effect of bending fluctuations on the fibril stability than to underestimate it. Collagen amino acid sequences are repeats of G-Xaa-Yaa triplets, where Xaa is often proline and Yaa often hydroxyproline. If we let each bead represent the helical rise 0.85 nm of a single G-Xaa-Yaa triplet of amino acids, we arrive at a bond distance $l_b = 0.57\sigma$. A single interaction bead then represents 9 amino acids (3 strands, a G-Xaa-Yaa triplet on each strand). Given a persistence length $l_p \approx 35$ nm, we then arrive at a bending elastic constant $k_b = 10k_B T$. We assume a typical Debye length of $\kappa^{-1} = 1$ nm, or $\kappa\sigma = 1.5$, corresponding to an equivalent concentration of monovalent electrolyte of approximately 0.1 M. Values for the effective charges of q of the positively and negatively charged beads, and for the Lennard–Jones energy parameter ϵ_h for hydrophobic beads are chosen such that the attraction between the beads (at a distance corresponding to the minimum of the Lennard–Jones potential) is approximately equal to the thermal energy $k_B T$. The Lennard–Jones energy parameter ϵ_0 for the charged beads and for the uncharged hydrophilic beads is chosen to be small such that the

Lennard–Jones attraction between these beads is negligible as compared to the thermal energy $k_B T$. Numerical values of all parameters are given in Table 1.

Table 1. Parameter values for the coarse-grained patterned semiflexible chain model.

Parameter	Value	Unit
κ	1.5	σ^{-1}
l_B	0.47	σ
q	2.0	–
ϵ_h	2.0	$k_B T$
ϵ_0	0.1	$k_B T$
l_b	0.57	σ
k_s	100	$k_B T / \sigma^2$
k_b	10	$k_B T$

For energy minimalization of the total interaction energy $E_{rod}(S)$ of a central chain with the surrounding chains in a staggered crystalline arrangement, we use a Metropolis Monte Carlo search algorithm. Sequences S are assumed to be palindromic. The initial sequence is a random (but palindromic) sequence with 30% hydrophobic, 30% uncharged hydrophilic and 40% charged beads (half of which are positively and half of which are negatively charged). Two random beads are selected in the first half of the sequence. Random selection is repeated until two beads are found of a different type. A Monte Carlo trial move consists of exchanging these two beads in both halves of the palindromic (symmetric) sequence, yielding a new sequence S_{new}. If $\Delta E_{rod} = E_{rod}(S_{new}) - E_{rod}(S) < 0$, the trial move is accepted. If $\Delta E_{rod} > 0$, it is accepted with probability $\exp(-\Delta E_{rod}/\Delta E)$. We choose $\Delta E = 2k_B T$. All sequences S and corresponding interaction energies $E_{rod}(S)$ produced by the runs are stored for later analysis. To speed up the calculations, before the start of the minimalization, neighbor lists are prepared and used for each packing, for a cut-off interaction radius of the beads of $r_{cutoff} = 5\sigma$.

Langevin Dynamics simulations of assembly into fibrils of optimized and a reference random sequences S_0 and S_{rand}

$S_0 = \{4,4,4,4,2,2,2,2,1,1,1,1,1,3,3,3,3,2,2,2,2,3,3,3,3,1,1,1,1,1,2,2,2,2,4,4,4,4\}$

$S_{rand} = \{2,2,4,3,2,3,4,1,1,2,3,1,2,1,3,2,4,1,1,4,4,1,1,4,2,3,1,2,1,3,2,1,1,4,3,2,3,4,2,2\}$

(each having $N = 40$ beads) were performed using the LAMMPS software package [22,23] using the Grønbech–Jensen–Farago time-discretization of the Langevin equation. [24,25]. Lennard–Jones (LJ) units were used, with lengths in terms of the bead diameter σ, and energies in terms of a LJ interaction energy scale ϵ. To connect to the parameter values of Table 1 used for sequence optimization (which are not in LJ units), we choose $\epsilon = k_B T^*$, where T^* is a temperature on the order of the temperature below which the chains start self-assembling. For the simulations, we take parameter values from Table 1 with $T = T^*$, such that the numerical parameter values in Table 1 are also the parameter values in LJ units. With this convention, the dimensionless temperature $\widetilde{T}$ in LJ units is $\widetilde{T} = T/T^*$. The simulation timestep was set to $\Delta t = 0.05\tau$ (where τ is the LJ unit of time), the bead friction constant was set to $\zeta_0 = 100(m/\tau)$, where m and t are the LJ units of mass and time. Numbers of simulation steps are fairly arbitrary numbers, but the time in units of τ, or better yet, in units of the rotational diffusion time τ_r of the rods, has physical meaning. For a rod consisting of N beads of diameter σ, spaced at a distance l_b, in the absence of bead–bead hydrodynamic interactions, $\tau_r = \zeta_0 N^3 l_b^2 / 12 \, k_B T$, where ζ_0 is the hydrodynamic friction of a single bead. For our parameters this gives $\tau_r = 8.67 \times 10^4 \tau$ (at $\widetilde{T} = 1$). The cut-off for both the Lennard–Jones and Debye–Hückel interactions was set to 5σ. Non-bonded interactions were only taken into account between beads on different chains. For assembly simulations a rectangular simulation box was used with dimensions $(l_x, l_y, l_z) = (20\sigma, 20\sigma, 120\sigma)$. Reflective rather than periodic boundary conditions were used in order to enforce the formation of finite rather than infinite fibrils. Initially, $M = 40$ rods were placed regularly in the box, with their long axis pointing in the z-direction. Particles were given random initial velocities

corresponding to a LJ temperature $\widetilde{T} = 5$ and a simulation was run for $58\tau_r$ (or 10^8 simulation steps). Next, particles were given random initial velocities corresponding to a LJ temperature $\widetilde{T} = 2.5$ and a simulation was again run for $58\tau_r$ (or 10^8 simulation steps). Finally, particles were given random initial velocities corresponding to a LJ temperature $\widetilde{T} = 1.5$ and a simulation was run for $46\tau_r$ (or 8×10^7 simulation steps). During equilibration, temperatures for all Langevin dynamics runs, decreases somewhat. For the run with $\widetilde{T} = 1.5$, the final equilibrium temperature was $\widetilde{T} = 1.0$. By running the simulations for a large number of rotational diffusion times, we guarantee that thermal equilibrium can indeed be reached during the simulations.

For cross-sectional assembly simulations, only those parts of the sequence S_0 are used that together form the non-gapped region. These are:

$S_{0,A}$ = {4,4,4,4,2,2,2,2,1,1,1,1}

$S_{0,B}$ = {3,3,3,3,2,2,2,2,3,3,3,3},

$S_{0,C}$ = {1,1,1,1,2,2,2,2,4,4,4,4}.

A 10×10 hexagonal array was prepared of A, B, and C chains, corresponding to the P_1 packing of the S_0 sequence, with a lattice spacing $l_c = 1.3\sigma$. For these simulations, a rectangular simulation box was used with dimensions $(l_x,l_y,l_z) = (24\sigma,18\sigma,11.84\sigma)$. Again, reflective boundary conditions were used. Particles were given random initial velocities corresponding to an LJ temperature $\widetilde{T} = 1.4$. After equilibration, this leads to an average temperature during the simulations of $\widetilde{T} = 1$. The rotational diffusion time for these much shorter rods with $N = 12$ is $\tau_r = 2.34 \times 10^3 \tau$. Simulations were run for $21\tau_r$ (or 10^6 simulation steps).

3. Results

### 3.1. Coarse-Grained Model for Patterned Semiflexible Chains

Driving forces for self-assembly that are taken into account in the coarse-grained model are electrostatic interactions, hydrophobic interactions, and excluded volume interactions. More specific, shorter-ranged and directional interactions such as hydrogen bonds, can possibly be implemented in later, more refined further sequence optimalizations, that start from patterns optimized at the most coarse-grained level, as obtained here. Semiflexible chains are represented using a simple bead–spring model (bead diameter σ, equilibrium bond length l_b) with harmonic bond stretching energy and harmonic bending energy. Electrostatic interactions are represented using a screened Coulomb interaction potential, steric and hydrophobic interactions using a Lennard–Jones potential. We focus on sequence variation and therefore choose a single set of interaction parameter values representative for semiflexible collagen-like triple helices (see Section 2). Specifically, we assume a persistence length of the semiflexible chains of $l_p = 20\sigma$.

A difference with the helical peptide systems for which staggered assembly has already been demonstrated [17,18] is that we focus on longer semiflexible chains that can have more complicated interaction patterns. Specifically, for the numerical examples, we use chains of $N = 40$ interaction beads. Taking collagen triple helices as an example, this would translate to a length of 40 G-Xaa-Yaa triplets or 120 amino acids per strand, much longer than typical collagen-like peptides, but still much shorter than the triple-helical part of human collagen I, which would correspond to $N \approx 350$. A shorter length was chosen to ensure that we can still adequately sample the interaction patterns and perform the Langevin Dynamics simulations of fibril assembly.

Interaction patterns are represented in terms of four characteristic bead types: hydrophobic (type 1, represented as green beads), hydrophilic uncharged (type 2, represented as white beads), positively charged (type 3, represented as red beads) and negatively charged beads (type 4, represented as blue beads). Interaction patterns therefore correspond to coarse-grained sequences $S = \{s_1,s_2,\ldots,s_N\}$, where the bead type $s_i = 1 \ldots 4$. For the values of the non-bonded interaction parameters that we use (Table 1), contact interactions between the different types of beads are on the order of the thermal

energy $k_B T$, such that multiple attractive contact interactions along the chains will be necessary to stabilize the fibrillar state.

Numerical values of parameters are chosen to roughly match triple-helical collagen, with three G-Xaa-Yaa triplets (one for each strand), making up a single bead in the coarse-grained model. For collagen, a single bead would therefore correspond to nine amino acids. We do not attempt to make an explicit mapping of the bead types to underlying microscopic chemistries here. The present very coarse-grained approach is mainly meant to guide future more detailed computational design approaches for specific microscopic chemistries, such as triple helical collagen. A qualitative mapping however is still possible. For example, staying with the case of triple-helical collagen, GPO triplets (where O is hydroxylated proline) that occur frequently in the sequence of triple-helical human collagen-I would probably be best represented by hydrophilic uncharged beads, whereas GPR triplets, known to occur at quite high frequencies in collagen sequences due to its stabilizing effect on collagen triple helices [26], would be best represented by a positively charged bead. An example comparison of a microscopic structure of a collagen-like triple-helical peptide and its coarse-grained representation is shown in Figure 1.
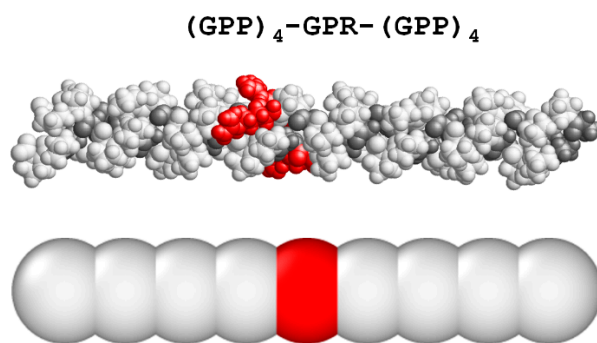
$$(GPP)_4-GPR-(GPP)_4$$

Figure 1. Illustration of coarse-graining of microscopic collagen triple-helical structure, to the coarse-grained representation in terms of interaction beads. Shown is the ideal triple helical structure for a $(GPP)_4$-GPR-$(GPP)_4$ homotrimer, with three GPP triplets (one in each strand) being represented by a single hydrophilic (white) bead, and the three central GPR triplets (one in each strand) being represented by a positively charged (red) bead.

### 3.2. Target-Staggered Crystalline Arrangements

While for some collagen-like peptides is was found that simple patterns of just electrostatic interactions can lead to cubic arrangements of the helices inside fibrils and two-dimensional membranes [17], we anticipate that for engineering stable fibrils using weak interactions, it is helpful to allow for as many contacts as possible. For these reasons, we here focus on the staggered assembly of semiflexible chains into hexagonal (para)crystalline arrays. For pattern design we focus on minimizing the interaction energy $E_{rod}(S)$ of patterned rods arranged in staggered hexagonal lattices, with respect to their sequence S. If the ground state that is found in this manner is sufficiently deep, we anticipate it will be stable against the thermally excited shape fluctuations of the semiflexible chains. This will be checked explicitly by verifying that semiflexible chains with the optimized sequences do indeed assemble into the designed crystal structure, using Langevin Dynamics simulations. Similar ground state design approaches are very successful for designing, e.g., de-novo designed self-assembling proteins [2].

Target structures that we consider are a family of staggered crystalline arrangements of straight chains with length $L = N \cdot l_b$, aligned with their long axis in the *z*-direction, with a hexagonal arrangement in the *x-y* plane. The hexagonal lattice spacing is l_c. In the *x-z* plane, neighbouring rods are shifted

in the *z*-direction by the stagger *D*. The geometry of the family of lattices is illustrated in Figure 2. The ratio *L/D* determines both the periodicity $N_\parallel$ of the staggered arrangement and the gap width *g*:

$$N_\parallel = \text{int}(L/D) + 1 \tag{5}$$

$$g = N_\parallel D - L \tag{6}$$
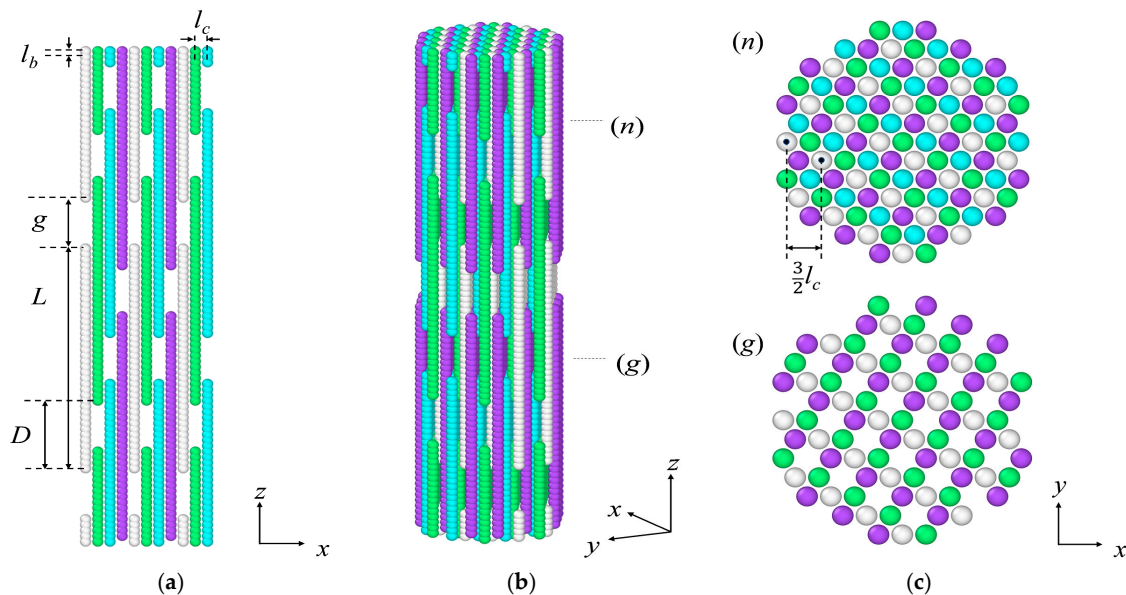
Figure 2. Representative packing chosen from the family of simple hexagonal staggered packings of (straight) chains considered. Shown is the packing P_5, for (straight) chains of $N = 40$ interaction beads and length $L = 40l_b$, arranged with a stagger $D = 12l_b$ such that the periodicity of the lattice is $N_\parallel = 4$ and the gap is $g = 8l_b$. (**a**) Projection onto *x-z* plane, showing a single staggered layer of rods. The rod length *L*, stagger *D* and gap width *g* are indicated, as well as the crystal spacing l_c and the bond length l_b (separation between beads in the chains). (**b**) Three-dimensional representation of cylindrical subsection of the crystalline lattice. Gapped and non-gapped regions of the fibril for which *y-x* cross-sections are shown in (**c**), are indicated by (*g*) and (*n*). (**c**) Projections on the *x-y* plane for non-gapped (*n*) and gapped region (*g*) indicated in (**b**). Additionally indicated is the shift of $3/2l_c$ in the x-direction, between successive layers of rods (in the *x-z* direction).

Positions $\mathbf{r}_{ijk}$ of the origins of the rods in the lattices are:

$$\mathbf{r}_{ijk} = \text{i} \cdot \boldsymbol{e}_i + \text{j} \cdot \boldsymbol{e}_j + \text{k} \cdot \boldsymbol{e}_k \tag{7}$$

where *i*, *j* and *k* are the lattice indices. The lattice vectors are:

$$\boldsymbol{e}_\text{i} = l_c\,\hat{x} \tag{8}$$

$$\boldsymbol{e}_\text{j} = -\frac{3}{2}l_c\hat{x} + \frac{1}{2}\sqrt{3}l_c\hat{y} \tag{9}$$

$$\boldsymbol{e}_\text{k} = D\hat{z} \tag{10}$$

For a given value of *k*, only certain values of *i* and *j* correspond to the origin of a rod:

$$(i+j)\bmod N_\parallel = k \bmod N_\parallel \tag{11}$$

Different lattices in the family have different values for the stagger *D*, the periodicity $N_\parallel$ and the gap width *g*. For the numerical calculations, we choose the stagger *D* to be an integer number of times

the equilibrium bond length l_b. Furthermore, we restrict ourselves to values of the gap g, $2 \leq g/l_b \leq 8$, and lattice periodicities up to $N_{\parallel} = 6$. This allows for eight distinct staggered hexagonal crystalline packings of rods to be used in the numerical calculations P_1–P_8, and these are listed in Table 2. As an example, Figure 2 shows the three-dimensional structure of the P_5 packing.

Table 2. Parameters for the target-staggered hexagonal packings P_1–P_8 of (straight) chains considered in the numerical calculations.

Packing	L/l_b	D/l_b	$N_{\parallel}$	g/l_b
P_1	40	14	3	2
P_2	40	15	3	5
P_3	40	16	3	8
P_4	40	11	4	4
P_5	40	12	4	8
P_6	40	9	5	5
P_7	40	8	6	8
P_8	40	7	6	2

3.3. Sequence Optimization

For the staggered hexagonal packings P_1–P_8 that we consider here, each chain has the same environment. Hence, as the objective function for the sequence design we can use the interaction energy $E_{rod}(S)$ of the central rod with lattice indices $(i,j,k) = (0,0,0)$, with all the other chains in the crystalline arrangement:

$$E_{rod}(S) = \sum_{(i,j,k) \neq (0,0,0)} V^{(ijk)}_{rod-rod}(S) \tag{12}$$

The potential of the interaction of the central chain at $(0,0,0)$ with another chain at lattice position (i,j,k) is a sum of bead–bead interactions:

$$V^{(ijk)}_{rod-rod}(S) = \sum_{n,m=1,N} V_{s_n s_m}\left(r^{ijk}_{nm}\right) \tag{13}$$

$$\mathbf{r}^{ijk}_{nm} = \mathbf{r}_{ijk} + (m-n)l_b \hat{z} \tag{14}$$

where $V_{s_n s_m}\left(r^{ijk}_{nm}\right)$ is the interaction potential for two beads: bead n on the central rod and bead m on the rod at lattice position (i,j,k), with bead types s_n and s_m, and for a bead–bead distance r^{ijk}_{nm} (see Section 2). For sequence optimization, we minimize the rod interaction energy $E_{rod}(S)$ with respect to the sequence S, for each of the packings P_1–P_8. We do so for a fixed composition in terms of bead types. Bead compositions we choose do not reflect natural collagen-like sequences, but rather reflect our expectations as to a typical bead composition that would readily form fibrils. We expect fibril assembly will be opposed by a net charge, while both charged and hydrophobic beads are necessary to drive assembly into staggered hexagonal packings. Too many hydrophobic beads could lead to kinetically trapped assembly into off-target packings. These considerations lead us to choose, for the study here, a composition of 30% of hydrophobic beads, 30% uncharged hydrophilic beads, and 40% charged beads (half of which are positively and half of which are negatively charged).

We search for sequences S with a minimal rod interaction energy $E_{rod}(S)$ using a Metropolis Monte Carlo search algorithm that generates random sequences S with probability:

$$p(S) \sim \exp(-E_{rod}(S)/\Delta E) \tag{15}$$

For a value of ΔE that we set to $\Delta E = 2$. We run the Metropolis Monte Carlo search for a large number of steps, and save the list of generated sequences and associated rod interaction energies, for later analysis. The Monte Carlo move that is used to generate a new sequence from a previous

sequence is a simple swap of the two beads in the sequence, provided the two bead types are different. For simplicity, we restrict ourselves to palindromic sequences, such that the orientation of the rods in the lattice does not matter.

First, for each of the crystal packings P_1–P_8 we search for sequences S with minimal interaction energy $E_{rod}(S)$ in runs of 10^5 Monte Carlo steps (per packing). Results are shown in Figure 3. The lowest energy sequences are found for packings P_1–P_3 with the lowest periodicity, $N_{\parallel}$ = 3. Differences in results between these lattices are very small. We choose the P_1 lattice for a more detailed investigation. Next, to get a complete histogram of the distribution of $E_{rod}(S)$ for all possible sequences S arranged in a P_1 lattice, we performed a simple random sampling search of 10^6 steps. The resulting energy distribution is shown in Figure 4a. For the bead compositions that we chose, we find that the most probable value for E_{rod} is negative. Even random sequences with this bead type composition might therefore have a tendency to stick to each other and bundle up, albeit probably not in an ordered array. This we will check later using Langevin Dynamics simulations for a sequence S_{rand} was randomly chosen from the set of sequences with interaction energy close to the peak of the energy distribution (given in Figure 5a).
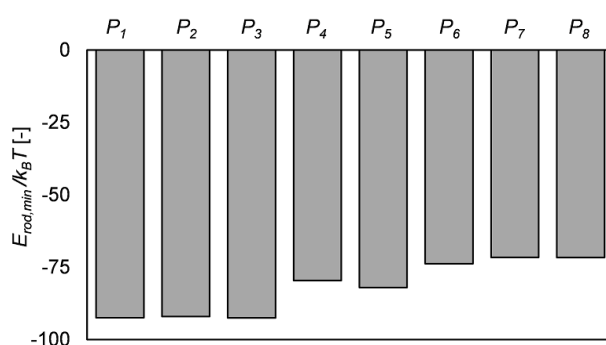
Figure 3. Lowest interaction energies $E_{rod,min}$ (units of $k_B T$) found in Monte Carlo sequence optimization for chains of $L = 40$ beads (30% hydrophobic, 30% hydrophilic, 40% charged) arranged in staggered hexagonal packings P_1–P_8.
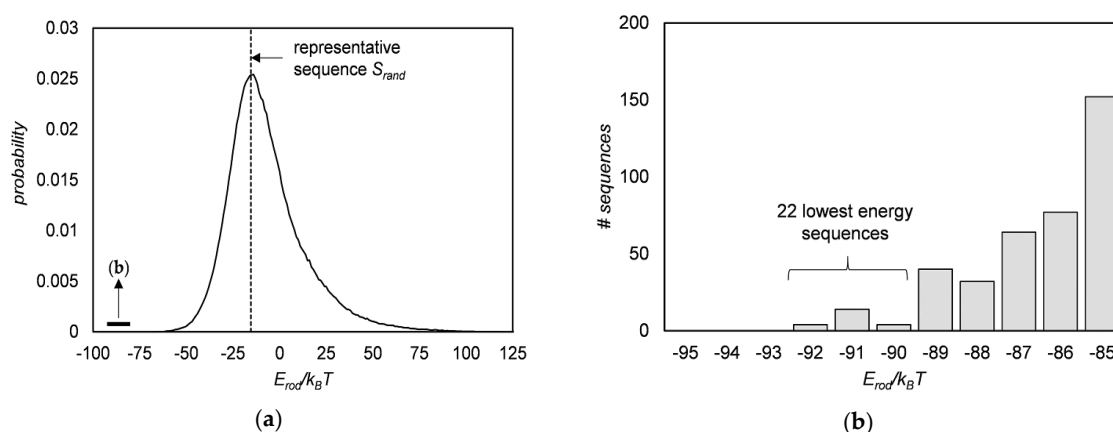
Figure 4. Sequence optimization for staggered hexagonal packing in P_1 lattice. (**a**) Distribution of rod interaction energies E_{rod} (units of $k_B T$) from simple random sequence search (10^6 steps). A representative random sequence S_{rand} is selected randomly from the sequences with energy for which the distribution peaks (vertical dotted line). (**b**) Low energy tail of the distribution of rod interaction energies, obtained with Metropolis Monte Carlo sequence search (10^6 steps).
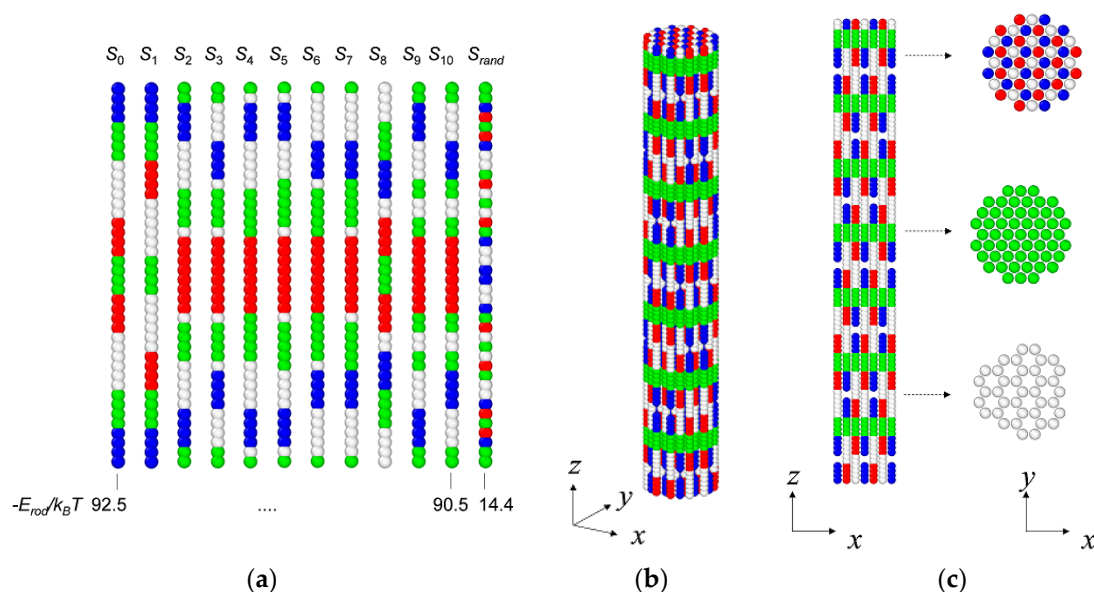
Figure 5. (**a**) Lowest rod interaction energy sequences S_0–S_{10} from Metropolis Monte Carlo sequence search. Additionally shown is the representative random sequence S_{rand} drawn from sequences with the rod interaction energy for which the probability distribution peaks, see Figure 4a. (**b,c**) Packing of optimized sequence S_0 into the staggered hexagonal P_1 target lattice. (**b**) Cylindrical subsection of crystal lattice. (**c**) slice in *x-z* plane: single layer of the lattice of chains, with staggered arrangement and perpendicular slices, in the *x-y* plane, demonstrating the three different perpendicular packings, one gapped and two non-gapped. Color code for interaction beads: green = hydrophobic (type 1), white = hydrophilic uncharged (type 2), red = positively charged (type 3), blue = negatively charged (type 4).

For finding the tail of the distribution at the side of low interaction energies, the simple random search is not efficient enough, and we need the Monte Carlo sampling, which zooms in on all sequences up to a certain energy above the minimum, as controlled by the parameter ΔE. The tail of the distribution found from a Monte Carlo search of 10^6 steps is shown in Figure 4b, the 22 lowest energy sequences are illustrated in Figure 5a. The interaction energy is invariant with respect to exchange of the positive and negative beads, so in fact there are only 11 distinct minimum sequences for packing the chains into the P_1 packing, that we label in order of ascending rod interaction energy, S_0–S_{10} The striking feature of these minimum sequences, compared to the random sequence, is that all show clustering of similar bead types. The two lowest energy sequences, S_0 and S_1, exhibit the most regular and simple patterns of stretches of bead types. The packing of chains with the sequence S_0 in the P_1 lattice is illustrated in Figure 5b,c.

Many of the low energy sequences S_0–S_{10} might correctly assemble into the designed P_1 structure also in the presence of thermal motion. A possible reason for exploring multiple sequences might be that the physical properties of P_1 fibrils, such as the mechanical properties, need not be the same. This, however, we leave as a topic for further research. Here, we restrict ourselves to establishing that designed sequences indeed assemble into the designed structures even if thermal motion is switched on. This we do for the lowest energy sequence S_0.
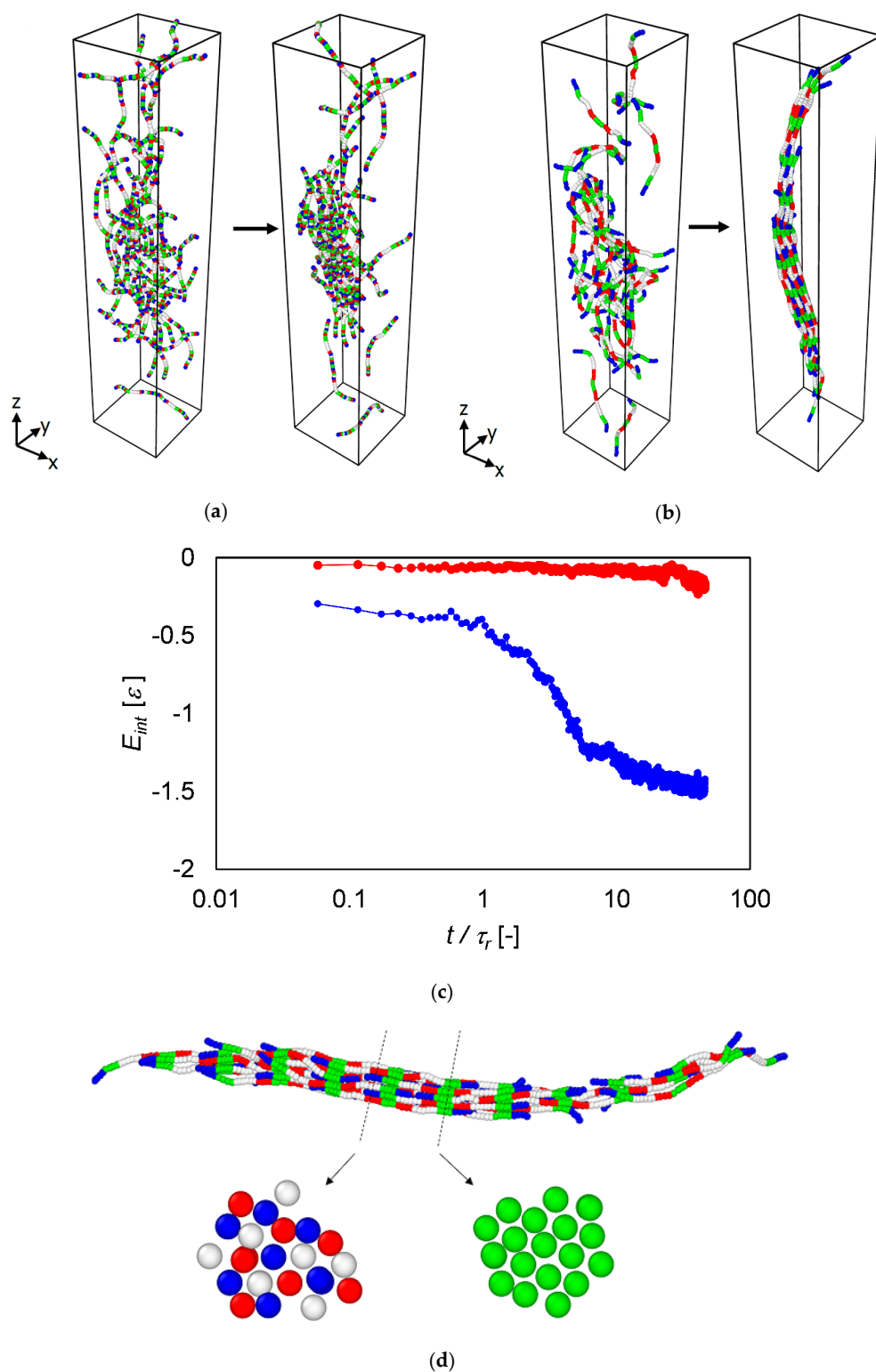
3.4. Langevin Dynamics Simulations

Sequence optimized sequences S are those which have the lowest rod interaction energy $E_{rod}(S)$ for a given target packing P. This "ground-state" approximation for sequence design is shown to be very powerful in other cases, such as the de-novo computational design of protein sequences [2], but it cannot guarantee that the target structure will be adopted by the optimized sequence. First of all, entropy competes with ordering and may preclude the ordered packing from being actually realized at

finite temperatures. Furthermore, the optimized sequence may have yet lower rod interaction energies in arrangements other than the target packing. Finally, kinetic barriers may prevent the target structure from being adopted. Therefore, we check using Langevin Dynamics simulations, whether semiflexible chains with the optimized sequences indeed spontaneously assemble in the target packing or not. We do so for the S_0 sequence, which was computationally optimized to assemble into the P_1 structure, and compare with a typical random sequence, S_{rand}. Random non-assembled initial states for $M = 40$ semiflexible chains were prepared at a temperature $\widetilde{T} = 1.4$ by stepwise cooling from still higher temperatures. A rectangular, closed simulation box was employed. To start assembly, the system was cooled to $\widetilde{T} = 1$, and simulated for $t = 46\tau_r$ (or 8×10^7 simulation steps).

　　Results are shown in Figure 5. For the random sequence S_{rand} we observe clustering of the chains into a very loosely associated anisotropic aggregate, but no clear fibril formation (Figure 6a). This is associated with just a small decrease in the total interaction energy E_{int} per interaction bead (Figure 6c). On the other hand, for the S_0 sequence, we observe rapid staggered assembly into fibrils (Figure 6b), associated with a large decrease in the interaction energy E_{int} per interaction bead (Figure 6c). A zoom of the final fibril structure for the optimized S_0 sequence is shown in Figure 6d. How close is this to the target P_1 packing it was designed to assemble into, and which is shown in Figure 5b,c?

　　The staggered assembly in the parallel direction clearly matches that of the P_1 target structure. For the cross-sectional order, in the plane perpendicular to the long axis of the fibril, it is more difficult to establish whether this matches the P_1 packing or not, because of the relatively low number of chains in the cross-sections. Representative cross-sections are shown in Figure 6d. The left cross-section, through the charged part of the chains, is expected to show the structure indicated in Figure 5c (top): positively charged beads should have a surrounding consisting of alternating negatively charged and neutral beads, etc. This can indeed be qualitatively recognized, but the numbers of chains in the cross-section are too low to be sure about the regularity of the pattern. Both this cross-section and the other cross-section, through the hydrophobic part of the chain, should show hexagonal order, but again, the numbers of chains in the cross-sections are too low to be absolutely sure.

　　As an additional test to establish whether the preferred cross-sectional order for the S_0 sequence is indeed the hexagonal P_1 packing, we performed Langevin-Dynamics simulations for the subsequences of the S_0 sequence, which together form the non-gapped part of the fibril. These were arranged in a 10×10 hexagonal lattice, and Langevin-Dynamics at $\widetilde{T} = 1$ was used to establish the stability of this lattice against the thermal motion. Results are shown in Figure 7. The 10×10 hexagonal starting lattice is shown in Figure 6a. The final configuration, at $t = 23\tau_r$ (or 10^6 simulation steps) is shown in Figure 7b. The top view shown in Figure 6c more clearly shows the ordering that remains in the presence of thermal fluctuations. The interaction energy per bead during the simulation run is shown in Figure 7d. During the simulation, the initial hexagonal lattice contracts somewhat, lowering the interaction energy per bead, and then remains constant. Especially from Figure 7d, it is evident that at $\widetilde{T} = 1$, thermal motion does not destroy the initial hexagonal ordering for these larger 10×10 cross-sectional lattices. Some disorder is observed at the edges, but not in the central part of the lattice. We conclude therefore that at to $\widetilde{T} = 1$, the S_0 sequence indeed spontaneously assembles into the target-staggered hexagonal P_1 structure, as designed.

Figure 6. Langevin Dynamics simulation of assembly of chains with sequences S_0 and S_{rand} after quench from $\widetilde{T} = 1.4$ to $\widetilde{T} = 1$. (**a**) Initial ($t = 0$) and final ($t = 75\tau_r$) configurations for sequence S_{rand}, where τ_r is the rotational diffusion time of the corresponding rigid rods. (**b**) Initial ($t = 0$) and final ($t = 75\tau_r$) configurations for sequence S_0, where τ_r is the rotational diffusion time of the corresponding rigid rods. (**c**) Chain–chain interaction energy E_{int} (per bead, in Lennard–Jones units, ϵ) during the assembly, for chains with sequences S_{rand} (red symbols) and S_0 (blue symbols), as a function of the simulation time in Lennard Jones units, t, scaled by the diffusional rotation time τ_r. (**d**) Zoom-in of the final structure for the chains with sequence S_0, with cross sections at the two different non-gapped regions.
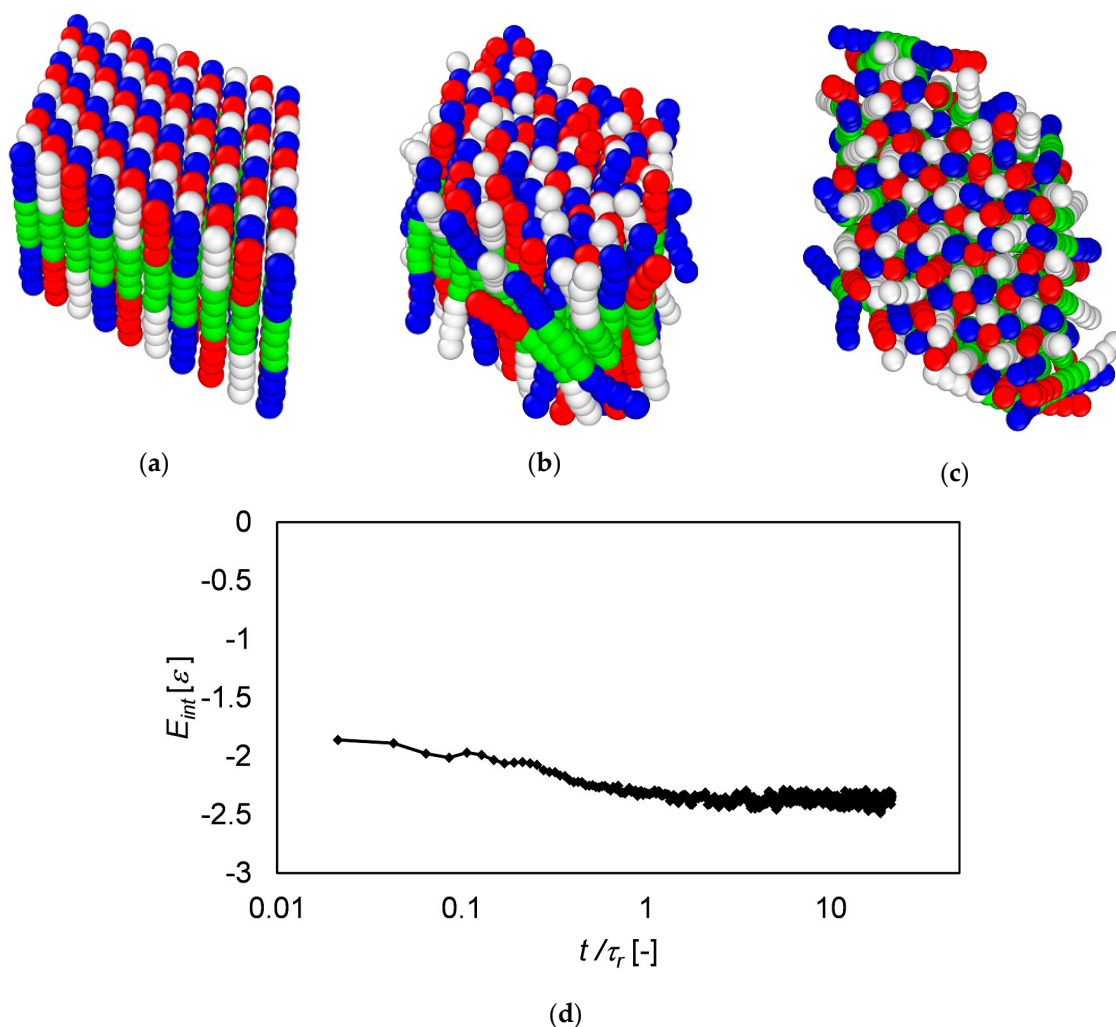
Figure 7. Langevin Dynamics simulation of hexagonally packed assembly of chains with sequences $S_{0,A}$, $S_{0,B}$, $S_{0,C}$, which are the subsequences of S_0 that together make up the non-gapped regions of fibrils assembled from chains with the sequence S_0. The simulation demonstrates that the hexagonal cross-sectional structure expected for the P_1 packing, is stable against thermal fluctuations for the S_0 sequence. (**a**) 10×10 hexagonally packed initial configuration. (**b**) relaxed final configuration at $t = 190\tau_r$. (**c**) relaxed final configuration, top view, which more clearly shows that the hexagonal ordering is maintained in the presence of thermal fluctuations. (**d**) Chain–chain interaction energy E_{int} (Lennard Jones units, ϵ) versus simulation time t scaled by rotational diffusion time τ_r of the rods.

4. Discussion

We found that in optimal sequences for assembly into staggered hexagonal packings, interaction beads of the same type tend to occur in stretches. Furthermore, the lowest interaction energy sequences were found for the simplest staggered hexagonal packings, with the lowest periodicity and L/D values.

Bead–bead interactions were chosen to be of the order of the thermal energy such that multiple favorable bead–bead contacts would be necessary to stabilize the chains in the target staggered hexagonal arrangement. Most likely the range of the interaction of the beads plays an important role in this observation: each bead not only interacts with its nearest neighbor (contact interactions), but also with next-nearest neighbors and beads that are still further away (non-contact interactions). Non-contact interactions will be enhanced for stretches of similar bead types, for example hydrophobic, or parallel stretches of charged beads with opposite charge sign, thus explaining our result.

For real systems such as helical peptides or proteins, the range of the hydrophobic interactions may not be as long as in our coarse-grained model, but the electrostatic interactions will have the same

long-range as in our model. Hence, we expect that the patterns of electrostatic interactions that we found will certainly be helpful in driving staggered fibrillar assembly also in experimental systems such as helical peptides or proteins.

It should be noted again that the translation of the coarse-grained model to real systems such as collagen-like helices is non-trivial. For example, the contact energy of pairs of oppositely charged ions is a very sensitive function of the distance of the charged groups and the number of intervening water molecules. Here we simply took a single average value for the contact energy of ion-pairs whereas in reality there will most likely be quite a broad spread of the contact energies of different ion-pairs.

The pattern of the S_0 sequence can also be considered as a more general pattern for staggered assembly into hexagonal (quasi) crystalline fibrils, independent of the specific type of bead interactions. This is illustrated in Figure 8. The S_0 sequence is palindromic (non-polar) and consists of two elements: a polar α domain, and a non-polar β domain, as shown in Figure 8a,b. Favorable interactions for the domains are antiparallel interactions of α-domains, and interactions of β-domains with α domains in any orientation. Real systems such as helical peptides and proteins are polar. A possible domain structure for a generic polar interaction pattern for driving staggered assembly into a hexagonal lattice immediately follows from the non-polar case of Figure 8b, and is shown in Figure 8c. It is a sequence of three polar domains, α, β and γ. Staggered assembly into a hexagonal lattice will ensue if interactions of the domains are such that the individual parallel α, β and γ domains form a hexagonal sheet with a pattern of alternating α, β and γ domains. This will happen, for example, if the only attractive interactions are α–β, β–γ and γ–α.
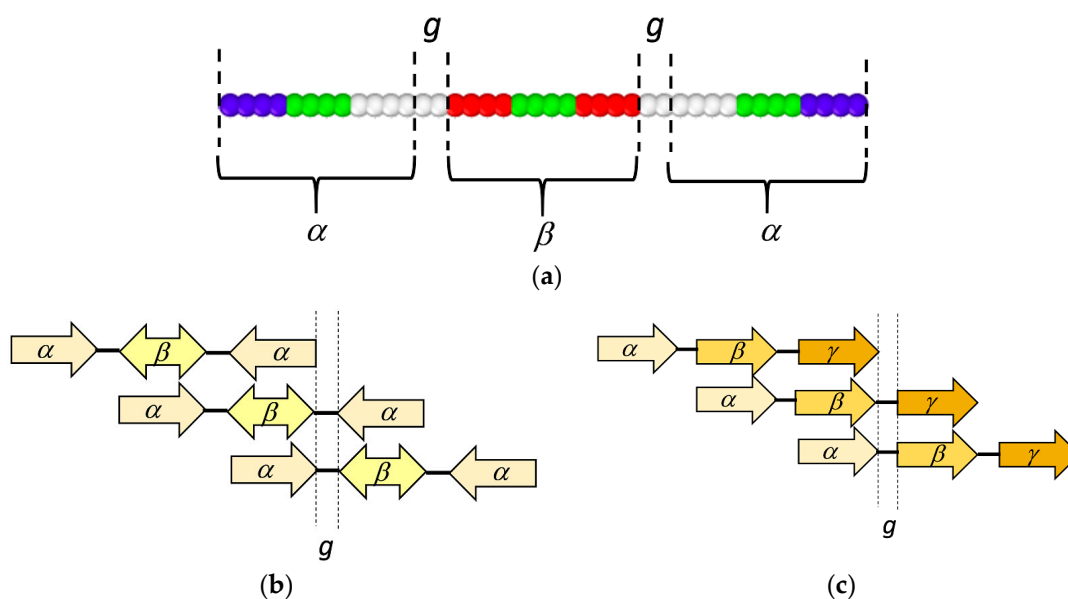
Figure 8. Generalization of the patterned chain. (**a**) For chains with the S_0 sequence, which have the simple staggered P_1 lattice as their lowest interaction energy state, the sequence is composed of two polar α-domains and a non-polar β-domain, separated by inert gap sequences (*g*). (**b**) The non-polar β-domain interacts with two of the polar α-domains, in the opposite orientation, forming a two-dimensional hexagonal lattice (**c**) generalization to polar chains with α, β, γ domains putatively assembling into hexagonal staggered P_1 lattices if there is a preference of the α, β, and γ domains for assembly in a two-dimensional hexagonal lattice in which the α, β and γ domains alternate.

The most likely experimental system for which interaction patterns on semiflexible chains such as those we discussed here can be realized are collagen-like proteins. Indeed, short collagen-like peptide sequences have already been designed, that make use of the kind of charge complementarity that we also discuss here, for driving staggered fibrillar assembly [17,18]. Our results suggest that a fruitful starting point for designing novel self-assembling collagen-like proteins would be the design

of sets of short collagen-like helices with controlled mutual interactions. These can then be used as elements in longer collagen-like helices to drive self-assembly. This is very similar to the notion of sets of orthogonal coiled coil-peptides and other peptide building blocks [27,28], that are being used to construct larger assemblies [29,30].

Successful de-novo design of self-assembling collagens is currently also hampered by our limited understanding of the thermal stability of collagen triple helices, such that a priori we do not know which G-Xaa-Yaa sequences will give stable triple helices and which will not. Only limited data and models are available, and are based on the substitution of single Xaa or Yaa residues in sequences that are otherwise composed of only G-P-P or G-O-P triplets [31–33]. These models fail to predict the stability of collagen-like triple helices for which the majority of triplets are not G-P-P or G-O-P, such as for example, collagen-like triple helices found in bacteria [34,35].

Given this limitation, a reasonable starting point to start exploring the atomistic-level computational design of self-assembling collagen-like proteins with current protein design software such as Rosetta [36], would therefore be to fix a triple-helical main-chain conformation, and to focus on Xaa and Yaa side-chain choice and conformation in designing sets of short collagen-like helices, with well-defined interaction interfaces. G-Xaa-Yaa triplets could, for example, be chosen to be those that are most frequent in natural collagen-like helices. These shorter sequences can then be used to drive self-assembly in longer designed collagen-like helices.

References

1. Seeman, N.C.; Sleiman, H.F. DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3*, 1–23.
2. Huang, P.-S.; Boyken, S.E.; Baker, D. The coming of age of de novo protein design. *Nat. Cell Biol.* **2016**, *537*, 320–327. [CrossRef]
3. Torquato, S. Inverse optimization techniques for targeted self-assembly. *Soft Matter* **2009**, *5*, 1157–1173. [CrossRef]
4. Engel, M.; Damasceno, P.F.; Phillips, C.L.; Glotzer, S.C. Computational self-assembly of a one-component icosahedral quasicrystal. *Nat. Mater.* **2015**, *14*, 109–116. [CrossRef]
5. Wilber, A.W.; Doye, J.P.K.; Louis, A.A.; Noya, E.G.; Miller, M.A.; Wong, P. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J. Chem. Phys.* **2007**, *127*, 085106. [CrossRef] [PubMed]
6. Romano, F.; Sciortino, F. Patterning symmetry in the rational design of colloidal crystals. *Nat. Commun.* **2012**, *3*, 975–976. [CrossRef] [PubMed]
7. Hulmes, D.J.S.; Miller, A.; Parry, D.A.D.; Piez, K.A.; Woodhead-Galloway, J. Analysis of the primary structure of collagen for the origins of molecular packing. *J. Mol. Biol.* **1973**, *79*, 137–148. [CrossRef]
8. Doyle, B.B.; Hukins, D.W.L.; Hulmes, D.J.S.; Miller, A.; Rattew, C.J.; Woodhead-Galloway, J. Origins and implications of the D stagger in collagen. *Biochem. Biophys. Res. Commun.* **1974**, *60*, 858–864. [CrossRef]
9. Meng, J.J.; Khan, S.; Ip, W. Charge interactions in the rod domain drive formation of tetramers during intermediate filament assembly. *J. Biol. Chem.* **1994**, *269*, 18679–18685.
10. Eikenberry, E.F.; Brodsky, B. X-ray diffraction of reconstituted collagen fibers. *J. Mol. Biol.* **1980**, *144*, 397–404. [CrossRef]
11. Kaur, P.J.; Strawn, R.; Bai, H.; Xu, K.; Ordas, G.; Matsui, H.; Xu, Y. The Self-assembly of a Mini-fibril with Axial Periodicity from a Designed Collagen-mimetic Triple Helix. *J. Biol. Chem.* **2015**, *290*, 9251–9261. [CrossRef] [PubMed]
12. Strawn, R.; Chen, F.; Haven, P.J.; Wong, S.; Park-Arias, A.; De Leeuw, M.; Xu, Y. To achieve self-assembled collagen mimetic fibrils using designed peptides. *Biopolymers* **2018**, *109*, e23226. [CrossRef] [PubMed]

13. Chen, F.; Strawn, R.; Xu, Y. The predominant roles of the sequence periodicity in the self-assembly of collagen-mimetic mini-fibrils. *Protein Sci.* **2019**, *28*, 1640–1651. [CrossRef] [PubMed]

14. Kouwer, P.H.J.; Koepf, M.; Le Sage, V.A.A.; Jaspers, M.; Van Buul, A.M.; Eksteen-Akeroyd, Z.H.; Woltinge, T.; Schwartz, E.; Kitto, H.J.; Hoogenboom, R.; et al. Responsive biomimetic networks from polyisocyanopeptide hydrogels. *Nat. Cell Biol.* **2013**, *493*, 651–655. [CrossRef]

15. Włodarczyk-Biegun, M.K.; Slingerland, C.J.; Werten, M.W.T.; Van Hees, I.A.; De Wolf, F.A.; De Vries, R.; Stuart, M.A.C.; Kamperman, M. Heparin as a Bundler in a Self-Assembled Fibrous Network of Functionalized Protein-Based Polymers. *Biomacromolecules* **2016**, *17*, 2063–2072. [CrossRef] [PubMed]

16. Freeman, R.; Han, M.; Álvarez, Z.; Lewis, J.A.; Wester, J.R.; Stephanopoulos, N.; McClendon, M.T.; Lynsky, C.; Godbe, J.M.; Sangji, H.; et al. Reversible self-assembly of superstructured networks. *Science* **2018**, *362*, 808–813. [CrossRef]

17. Rele, S.; Song, Y.; Apkarian, R.P.; Qu, Z.; Conticello, V.P.; Chaikof, E.L. D-Periodic Collagen-Mimetic Microfibers. *J. Am. Chem. Soc.* **2007**, *129*, 14780–14787. [CrossRef]

18. Sharp, T.H.; Bruning, M.; Mantell, J.; Sessions, R.B.; Thomson, A.R.; Zaccai, N.R.; Brady, R.L.; Verkade, P.; Woolfson, D.N. Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of de novo design. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13266–13271. [CrossRef]

19. Rosenbaum, D.M.; Liu, D.R. Efficient and Sequence-Specific DNA-Templated Polymerization of Peptide Nucleic Acid Aldehydes. *J. Am. Chem. Soc.* **2003**, *125*, 13924–13925. [CrossRef]

20. Vesentini, S.; Redaelli, A.; Gautieri, A. Nanomechanics of collagen microfibrils. *Muscles Ligaments Tendons J.* **2013**, *3*, 23–34. [CrossRef]

21. Kirkness, M.W.; Lehmann, K.; Forde, N.R. Mechanics and structural stability of the collagen triple helix. *Curr. Opin. Chem. Biol.* **2019**, *53*, 98–105. [CrossRef] [PubMed]

22. LAMMPS Molecular Dynamics Simulator. Available online: https://lammps.sandia.gov/ (accessed on 1 September 2020).

23. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19. [CrossRef]

24. Grønbech-Jensen, N.; Farago, O. A simple and effective Verlet-type algorithm for simulating Langevin dynamics. *Mol. Phys.* **2013**, *111*, 983–991. [CrossRef]

25. Grønbech-Jensen, N.; Hayre, N.R.; Farago, O. Application of the G-JF discrete-time thermostat for fast and accurate molecular simulations. *Comput. Phys. Commun.* **2014**, *185*, 524–527. [CrossRef]

26. Ramshaw, J.A.M.; Shah, N.K.; Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host–Guest Triple-Helical Peptides. *J. Struct. Biol.* **1998**, *122*, 86–91. [CrossRef]

27. Bromley, E.H.C.; Channon, K.; Moutevelis, E.; Woolfson, D.N. Peptide and Protein Building Blocks for Synthetic Biology: From Programming Biomolecules to Self-Organized Biomolecular Systems. *ACS Chem. Biol.* **2008**, *3*, 38–50. [CrossRef]

28. Reinke, A.W.; Grant, R.A.; Keating, A.E. A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering. *J. Am. Chem. Soc.* **2010**, *132*, 6025–6031. [CrossRef]

29. Fletcher, J.M.; Harniman, R.L.; Barnes, F.R.H.; Boyle, A.L.; Collins, A.; Mantell, J.; Sharp, T.H.; Antognozzi, M.; Booth, P.J.; Linden, N.; et al. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013**, *340*, 595–599. [CrossRef]

30. Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.S.; Šali, A.; Klavžar, S.; Jerala, R. Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **2013**, *9*, 362–366. [CrossRef]

31. Locardi, E.; Kwak, J.; Scheraga, H.A.; Goodman, M. Thermodynamics of Formation of the Triple Helix from Free Chains and from Template-Constrained Chains of Collagen-like Monodisperse Poly(Gly-Pro-Hyp) Structures. *J. Phys. Chem. A* **1999**, *103*, 10561–10566. [CrossRef]

32. Persikov, A.V.; Ramshaw, J.A.M.; Brodsky, B. Prediction of Collagen Stability from Amino Acid Sequence. *J. Biol. Chem.* **2005**, *280*, 19343–19349. [CrossRef] [PubMed]

33. Doig, A.J. Statistical Thermodynamics of the Collagen Triple-Helix/Coil Transition. Free Energies for Amino Acid Substitutions within the Triple-Helix. *J. Phys. Chem. B* **2008**, *112*, 15029–15033. [CrossRef] [PubMed]

34. Yu, Z.; Brodsky, B.; Inouye, M. Dissecting a Bacterial Collagen Domain from Streptococcus pyogenes: Sequence and length-dependent variations in triple helix stability and folding. *J. Biol. Chem.* **2011**, *286*, 18960–18968. [CrossRef] [PubMed]

35. Yu, Z.; An, B.; Ramshaw, J.A.M.; Brodsky, B. Bacterial collagen-like proteins that form triple-helical structures. *J. Struct. Biol.* **2014**, *186*, 451–461. [CrossRef]

36. Alford, R.F.; Leaver-Fay, A.; Jeliazkov, J.R.; O'Meara, M.J.; DiMaio, F.P.; Park, H.; Shapovalov, M.V.; Renfrew, P.D.; Mulligan, V.K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048. [CrossRef] [PubMed]