Supporting information to:

How the choice of force-field affects the stability and self-assembly process of supramolecular CTA fibers

Tomasz K. Piskorz¹, Alex H. de Vries², and Jan H. van Esch¹,*

¹Department of Chemical Engineering, Delft University of Technology, van der Maasweg 9, Delft, 2629 HZ, The Netherlands

²Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

*E-mail: j.h.vanesch@tudelft.nl

Table of Contents

1. Methods 2
   1.1. Analysis 2
      1.1.1. Hydrogen bond analysis 2
   1.2. Parameterization 2
      1.2.1. Martini 2
      1.2.2. CGenFF Mod. 3
   1.3 Simulations 5
      1.3.1. Protocol 5
      1.3.2. Position restraints 8
      1.3.3. Performance 8
2. Additional results 9
   2.1. Low concentration 9
   2.2. Simulations in NPT 11
   2.3. Simulations in isotropic NPT ensemble 12
   2.4. Decomposition of enthalpy of creation 13
   2.5. Energy landscape of fiber 14
   2.6. Pseudo-phase approximation 15
3. Attachments 15
4. References 15
1. Methods

1.1. Analysis

1.1.1. Hydrogen bond analysis

Hydrogen bond analysis has been done using VMD and plugin HBonds¹. To set the criterion for hydrogen bonds, we have calculated the distribution of the distance between hydrogen donor and acceptor and the angle between donor-hydrogen-acceptor. We calculated the distributions for the first 1 ns of fiber simulation for those force-fields in which the fiber visually seemed stable, i.e., CHARMM Drude, GROMOS and MartiniP (see Figure S1). From the distribution, we established that for the atomistic force-fields, we use as H-bond criteria that the distance be below 0.33 nm and angle be below 40° and for MartiniP that the distance be below 0.4 nm and angle be below 40°.

![Figure S1. Distribution of distance between hydrogen acceptor and donor and angle of donor-hydrogen-acceptor for 1 ns of fiber simulation in CHARMM Drude, GROMOS and MartiniP force-fields.](image)

1.2. Parameterization

1.2.1. Martini

CTA molecule (Figure S2a) has been parameterized according to the official tutorial available on Martini force-field website. The bonded parameters have been parameterized on the basis of the GROMOS simulation of a single molecule. We converted the all-atomistic trajectory to a coarse-grained one by calculating the center of mass of group of atoms creating each bead. From such created coarse-grained trajectories, we have extracted bonded parameters, that is, bonds, angles, and torsion angles used for the coarse-grained force-field. The bead types were chosen on the basis of already parameterized molecules. We have tested several combinations of beads from which the best result yielded topologies presented in Figure S2b for Martini and Figure S2c for MartiniP. The topology file is attached at the end of this document.
1.2.2. CGenFF Mod.

Since CHARMM Drude showed that the structure of the pre-formed fibers was stable, we were interested to see if we can use the effective charges and find similar stability in simulations with the otherwise standard CGenFF force-field. We have constructed effective charges for the CGenFF model by redistributing the induced dipoles in the CHARMM Drude model over the atomic centers in a manner that preserves the local dipole moments optimally, akin to the Dipole Preserving Charge (DPC) introduced by Thole and van Duijnen\(^2\). The analysis, as implemented here, calculates the induced dipole for each Drude atom as the charge of the Drude particle times the displacement of the Drude charge from its parent atom. The analysis aims to recreate this induced dipole by placing partial charges on all parent atoms in the non-polarizable model. The partial charges obey the restraints (common in many charge-fitting schemes such as RESP\(^3\)) that the total charge is conserved and that the total dipole moment is conserved, but in addition, weight factors are introduced that favor placing the partial charges on the atoms nearest to the Drude charge that generates the local dipole. This idea is taken from the DPC analysis; the weighting scheme is exponential, i.e., the weight decreases exponentially with the distance of the atom to the Drude charge. The scheme is implemented in a Mathematica NOTEBOOK\(^4\) as a matrix inversion problem. The effective charges were obtained by averaging over three snapshots of a stack of eight CTA molecules taken from an MD simulation in the CHARMM Drude model and averaging over all equivalent atoms (three in each CTA) and snapshots (three). Table S1 reports the atoms whose charges were modified compared to the standard CGenFF. For these charges, we show the base charge (atom charge in Drude model + Drude (mobile) charge of that atom, i.e., sum of the charges when they are on top of each other), the polarized charge (the change in the charge of the atom: the modified charge used is base charge + polarized charge), and the standard deviation in the polarized charge, calculated from the pooled measurements over CTA molecules in the stack and multiple snapshots.
Table S1. Mapped charge distribution for CHARMM Drude force-field for selected atoms.

| Element       | Base charge | Polarized charge | St. dev. of polarized charge |
|---------------|-------------|------------------|------------------------------|
| Hydroxyl group| O           | -0.46            | -0.16                        | 0.06        |
|               | H           | 0.36             | 0.07                         | 0.03        |
| Amide group   | C           | 0.50             | -0.04                        | 0.04        |
|               | N           | -0.38            | 0.03                         | 0.11        |
|               | H           | 0.27             | 0.06                         | 0.05        |
| Ester group   | C           | 0.70             | 0.04                         | 0.04        |
|               | =O          | -0.34            | 0.02                         | 0.03        |
|               | -O-         | -0.61            | -0.01                        | 0.03        |

Table S1 shows some insights into charge and polarization from CHARMM Drude for selected atoms. The most polarized atoms are the hydroxyl groups at the end of the side chains. Their polarization probably comes from interaction with water molecules. Amide groups, which we expected to have the strongest polarization due to the creation of an array of aligned hydrogen bonds, also are polarized, but the polarization level is not close to that of the hydroxyl groups.

Our motivation for the use of CHARMM Drude force-field in this work comes from the work of Albuquerque et al., which have shown that amide polarization could be a crucial factor in the stabilization of CTA type fibers. We have also calculated the interactions and dipole moments for the molecules from their works, and the results are presented in Figure S3. We have stacked molecules on top of each other (translated by 0.47 nm as in reference 5), minimized the energy, and calculated the average dipole moment of the structure. The results for side chain group R1 are presented in Figure S3c. The dipole moment of molecules in the stack, which were simulated using GROMOS force-field, is constant, which contrasts with results presented by Albuquerque, which showed that with an increasing number of monomers also the dipole moment increases. The increased dipole moment of molecules in the stack could result in increased electrostatic interactions between monomers. Therefore, we have tested a polarizable CHARMM Drude force-field, which can take into account the polarizability of atoms. It can be seen in Figure S3c that for this force-field that upon increasing the number of monomers, the dipole moment also increases, and it reproduces a similar trend as for the results reported in the literature.
Figure S3. Analysis of fiber formation in vacuum for a molecule from the work of Albuquerque et al.³, which showed that amide polarization could be crucial for the stability of CTA type molecules.

1.3 Simulations

1.3.1. Protocol

Simulations in this work have been performed using GROMACS. The simulation procedure for different force-fields followed the scheme: (a) energy minimization, (b) short equilibration in NVT ensemble, (c) short equilibration in NPT ensemble, and (d) produce run (depending on the process in NVT or NPT ensemble). CHARMM Drude production runs were performed only in NVT (simulations in NPT are not efficient). We used different barostats and thermostats, timestep, and number of steps for different force-fields. The parameters are presented in Table S2, Table S3, and Table S4.
Table S2. Simulations of spontaneous self-assembly.

|                        | GROMOS | CGenFF | CHARMM Drude | CGenFF Mod. | GAFF | Martini | MartiniP |
|------------------------|--------|--------|--------------|-------------|------|---------|----------|
| Number of CTA molecules| 8      |        |              |             |      |         |          |
| Initial simulation box size | 4.54 x 4.00 x 4.08 nm³ (dodecahedron) |        |              |             |      |         |          |
| Number of water molecules | 2092 | 2058  | 1480         | 2058        | 2051 | 370     | 500      |
| Water model            | SPC    | TIP3P  | SWM4-NDP     | TIP3P       | TIP3P| Martini | Polar Martini |
| Energy minimization    | Algorithm | Steps | Steps | Steps | Steps | Steps | Steps | Steps |
| NVT ensemble           | Berendsen | 5000  | 1000  | 50000 | 5000 | 5000   | 10000    |
| Thermostat             | Berendsen | Berendsen | Berendsen | V-rescale | V-rescale | V-rescale |          |
| Temperature coupling constant | 1.0 | 1.0 | 1.0 | 0.1 | 1.0 | 1.0 |
| Steps                  | 5000  | 5000  | 1000  | 10000 | 5000 | 5000 | 10000 |
| dt [ps]                | 0.0005 | 0.0005 | 0.0005 | 0.002 | 0.01 | 0.03 ps |
| NPT ensemble           | Berendsen | Berendsen | V-rescale | V-rescale | V-rescale | V-rescale |          |
| Thermostat             | Berendsen | V-rescale | V-rescale | V-rescale | V-rescale | V-rescale |          |
| Temperature coupling constant | 1.0 | 0.1 | 0.1 | 0.1 | 1.0 | 1.0 |
| Steps                  | 5000  | 5000  | 10000 | 10000 | 10000 | 5000 | 5000 |
| dt [ps]                | 0.0005 | 0.002 | 0.001 | 0.002 | 0.002 | 0.03 ps |
| Production run         | Berendsen | Berendsen | V-rescale | V-rescale | V-rescale | V-rescale |          |
| Thermostat             | Berendsen | V-rescale | V-rescale | V-rescale | V-rescale | V-rescale |          |
| Temperature coupling constant | 1.0 | 0.1 | 0.1 | 0.1 | 1.0 | 1.0 |
| barostat               | Berendsen | Parrinello-Rahman | None (NVT ensemble) | Parrinello-Rahman | Parrinello-Rahman | Berendsen | Berendsen |
| Pressure coupling constant | 0.1 | 2.0 | 1.0 | 2.0 | 3.0 | 3.0 |
| Steps                  | 50000000 | 50000000 | 100000000 | 50000000 | 50000000 | 3333334 | 3333334 |
| dt [ps]                | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.03 | 0.03 |
|                           | GROMOS | CGenFF | CHARMM Drude | CGenFF Mod. | GAFF | Martini | MartiniP |
|---------------------------|--------|--------|--------------|-------------|------|---------|----------|
| **Number of CTA molecules** | 24     |        |              |             |      |         |          |
| **Initial simulation box size** | 4.00 x 4.00 x 11.82 nm³ | 4.15 x 10.97 x 3.73 nm³ (dodecahedron) | 4.00 x 4.00 x 11.82 nm³ | 5.00 x 5.00 x 11.52 nm³ | 11.82 x 4.99 x 4.40 nm³ (dodecahedron) |
| **Number of water molecules** | 5373   | 5270   | 4563         | 5272        | 5270 | 2464    | 1831     |
| **Energy minimization**     | Algorithm | Steepest | Steepest | Steepest | Steepest | Steepest | Steepest |
|                            | Steps   | 5000   | 50000       | 1000        | 50000| 50000   | 10000    |
|                           | Thermostat | Berendse n | Berendsen | Berendse n | Berendsen | V-rescale | V-rescale |
|                           | Temperature coupling constant | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
|                           | Steps   | 10000  | 5000       | 5000        | 5000 | 5000   | 5000     |
|                           | dt [ps] | 0.0005 | 0.0005     | 0.0005      | 0.0005 | 0.01   | 0.03 ps  |
| **NVT ensemble**           | Thermostat | Berendse n | Berendsen | Berendse n | Berendsen | V-rescale | V-rescale |
|                           | Temperature coupling constant | 1.0 | 0.1 | 0.1 | 0.1 | 0.1 | 1.0 |
|                           | Barostat | Berendse n | Parrinello-Rahman | Parrinello-Rahman | Parrinello-Rahman | Berendsen | Berendsen |
|                           | Pressure coupling constant | 0.1 | 2.0 | 1.0 | 2.0 | 2.0 | 3.0 |
|                           | Steps   | 10000  | 5000       | 10000       | 50000 | 100000 | v-rescale |
|                           | dt [ps] | 0.0005 | 0.002      | 0.001       | 0.002 | 0.01  | 0.03 ps  |
| **NPT ensemble**           | Thermostat | Berendse n | Berendsen | Berendse n | Berendsen | V-rescale | V-rescale |
|                           | Temperature coupling constant | 1.0 | 0.1 | 0.1 | 0.1 | 1.0 | 1.0 |
|                           | Barostat | none | none | none | none | None | none |
|                           | Steps   | 100000000 | 50000000 | 100000000 | 50000000 | 50000000 | 3333334 |
|                           | dt [ps] | 0.001  | 0.001      | 0.001       | 0.002 | 0.002 | 0.03    |
Table S4. Simulations of dimers and pentamer

|                        | GROMOS | CGenFF | CHARM M Drude | CGenFF Mod. | GAFF   | Martini | MartiniP |
|------------------------|--------|--------|---------------|-------------|--------|---------|---------|
| **Dimer**              |        |        |               |             |        |         |         |
| Number of CTA molecules| 2      |        |               |             |        |         |         |
| Initial simulation box size | 4.00 x 4.00 x 4.00 nm³ | 4.00 x 4.00 x 2.83 nm³ (dodecahedron) | 4.00 x 4.00 x 2.83 nm³ (dodecahedron) | 4.54 x 4.00 x 4.08 nm³ (dodecahedron) | 4.00 x 4.00 x 4.00 | 4.00 x 4.00 x 2.83 nm³ (dodecahedron) |
| Number of water molecules | 2000  | 2000   | –1400 (1399-1412) | –1400 (1396-1412) | 2200  | 400     | 240     |
| **Pentamer**            |        |        |               |             |        |         |         |
| Number of CTA molecules| 5      |        |               |             |        |         |         |
| Initial simulation box size | 5.00 x 5.00 x 5.00 nm³ | 5.00 x 5.00 x 5.00 nm³ (dodecahedron) | 5.00 x 5.00 x 5.00 nm³ (dodecahedron) | 5.00 x 5.00 x 5.00 nm³ (dodecahedron) | 5.00 x 5.00 x 5.00 nm³ | 5.00 x 5.00 x 5.00 nm³ |
| Number of water molecules | 2400  | 2400   | 2400          | 2400        | 2400  | 1100    | 900     |
| **Energy minimalization** | Algorithm | Steepest | Steepest | Steepest | Steepest | Steepest | Steepest |
| Steps                  | 5000   | 5000   | 1000         | 50000       | 5000   | 10000   | 20000   |
| **NVT ensemble**       | Thermmostat | Berendsen | Berendsen | Berendsen | V-rescale | V-rescale | V-rescale |
| Temperature coupling constant | 1.0      | 1.0    | 1.0          | 1.0         | 1.0    | 1.0     | 1.0     |
| Steps                  | 5000   | 5000   | 5000         | 10000       | 5000   | 5000    | 5000    |
| dt [ps]                | 0.0005 | 0.0005 | 0.0005       | 0.0005      | 0.002  | 0.01    | 0.03 ps |
| **NPT ensemble**       | Thermmostat | V-rescale | V-rescale | V-rescale | V-rescale | V-rescale | V-rescale |
| Temperature coupling constant | 1.0      | 0.1    | 0.1          | 0.1         | 0.1    | 1.0     | 1.0     |
| Barostat               | Berendsen | Parrinello-Rahman | Parrinello-Rahman | Parrinello-Rahman | Parrinello-Rahman | Berendsen | Berendsen |
| Pressure coupling constant | 0.1      | 2.0    | 1.0          | 2.0         | 2.0    | 3.0     | 3.0     |
| Steps                  | 5000   | 5000   | 5000         | 5000        | 10000  | 1000    | 5000    |
| dt [ps]                | 0.002  | 0.002  | 0.001        | 0.002       | 0.002  | 0.01    | 0.03 ps |
| **Product run**        | Thermmostat | Berendsen | V-rescale | Nose-Hoover | V-rescale | V-rescale | V-rescale |
| Temperature coupling constant | 1.0      | 0.1    | 0.1          | 0.1         | 0.1    | 1.0     | 1.0     |
| Barostat               | Berendsen | Parrinello-Rahman | none | Parrinello-Rahman | Parrinello-Rahman | Berendsen | Berendsen |
| Pressure coupling constant | 0.1      | 2.0    | 2.0          | 2.0         | 3.0    | 3.0     | 3.0     |
| Steps                  | 5000000 | 5000000 | 5000000      | 5000000     | 5000000 | 333334  | 1000000 |
| dt [ps]                | 0.002  | 0.002  | 0.001        | 0.002       | 0.002  | 0.03    | 0.01    |
1.3.2. Position restraints
The simulations of the formation of pentamers have been done by adding one molecule to the tetramer system. The tetramer has been created from the crystal structure. The tetramer has been kept together by putting position restraints on three atoms of cyclohexane rings (the strength of the restraints is 100 kJ/mol/nm$^2$).

1.3.3. Performance
Table S5 shows the performance of the simulation for different force-fields for simulation of spontaneous self-assembly.

Table S5. Performance of MD for simulation of self-assembly of eight molecules for different force-fields measured for 10 ps simulations of fiber on single core of Intel® Core™ i7-5600U CPU @ 2.60 GHz.

| Force-field    | Performance [hours/ns] |
|----------------|------------------------|
| CHARMM Drude   | 28.088                 |
| CGenFF         | 8.274                  |
| GROMOS         | 7.660                  |
| MartiniP       | 0.053                  |
| Martini        | 0.026                  |

2. Additional results
2.1. Low concentration
To test the influence of concentration on self-assembly, we simulated 8 CTA molecules using GROMOS, GAFF, and MartiniP force-fields in different volumes of the simulation box: 74 nm$^3$ (shown in the main text), 150 nm$^3$, and 300 nm$^3$. The results for 100 ns simulation in GROMOS and GAFF are presented in Figure S4 and 1 µs simulations for MartiniP in Figure S5. All except one system behave similarly: they form one unordered cluster (as indicated by fast decrease of SASA) and slowly rearrange into an ordered structure (as indicated by the slow increase of H-bonds). There seems little dependence on ordering on the size of the simulation box. The only exception is self-assembly using GAFF in the 300 nm$^3$ simulation box. Increased size of the simulation box results in lowering the probability of immediate formation of the large unordered cluster. As a result, small clusters are formed in the first 20 ns (see SASA), which can quickly rearrange into an ordered structure. We anticipate that further increase of the systems would lead to speed up of self-assembly on simulation timescale. Unfortunately, it will also significantly increase computational costs.
Figure S4. Simulations of 100 ns self-assembly of 8 CTA molecules in different sizes of simulation box:
(a) 74 nm$^3$, GROMOS (shown in the main text), (b) 150 nm$^3$, GROMOS, (c) 300 nm$^3$, GROMOS, (d) 74 nm$^3$, GAFF (shown in the main text), (e) 150 nm$^3$, GAFF, (f) 300 nm$^3$, GAFF.
(g) Progression of the number of hydrogen bonds between amide groups. (h) Progression of solvent accessible surface area.
2.2. Simulations in NPT

We have performed simulations of fiber stability also in the NPT ensemble (with semi-isotropic pressure coupling). The results are presented in Figure S6 and quantitatively are similar to results for the NVT ensemble. There are no results for CHARMM Drude due to the lack of an efficient way of solving the equations of motion in the NPT ensemble. Similarly, as in the NVT ensemble, the fiber is stable in GROMOS, GAFF, and MartiniP, and it disassembles in CGenFF, CgenFF Mod. and Martini as it can be seen from the progression of the number of hydrogen bonds (see Figure S6h). The solvent-accessible surface area is less informative for trajectories in NPT (see Figure S6i).
2.3. Simulations in isotropic NPT ensemble

We simulated fiber in an isotropic NPT ensemble. Except for MartiniP, in all other cases, the fiber breaks. However, in force-fields that show more stable fiber in NVT/semiisotropic NPT ensemble (GROMOS and GAFF), they maintain internal order to some extent (see Figure S7). We speculate that the decrease of stability in an isotropic ensemble results from a mismatch of pressures in xy-plane (dominated by diffusion of water) and z-directions (dominated by the motion of, relatively static, fiber).
2.4. Decomposition of enthalpy of creation

In the main text, energy decompositions for bonded, electrostatic, and L-J interaction is shown only for GROMOS and CHARMM Drude. Figure S8 presents energy decomposition for CGenFF, CGenFF Mod., GAFF, Martini, MartiniP, and GROMOS (as a reference).
Figure S8. (a) Enthalpy of creation of dimer, trimer etc. per molecule. (b-c) Decomposition of enthalpy of creation dimers, trimers etc. for (b) GROMOS, CGenFF, CGenFF Mod. and GAFF, and for (c) GROMOS, Martini, and MartiniP.

2.5. Energy landscape of fiber
Figure S9 shows 2D histogram of the position distribution of a molecule with respect to the other molecule or the tetramer for CGenFF, CGenFF Mod, and Martini. Relative positions were measured over final 1ns from 10 ns multiple simulations.
2.6. Pseudo-phase approximation

According to the pseudo-phase approximation, the difference in standard chemical potential between monomer and aggregate of size N, depends on the critical aggregation concentration (CAC) and can be expressed as:

\[ \Delta G = (\mu_1^0 - \mu_N^0) = RT\ln(X_{CAC}) \]

We approximate CAC by the critical gelation concentration (CGC), which for CTA molecules is 0.36 mM. Therefore, the difference of the potentials \( \Delta G \approx 30 \cdot \frac{k}{mol} \). This potential shows the difference of chemical potential between monomer and molecule in the aggregate, therefore it can be interpreted as the free energy of elongation studied in the main text.

3. Attachments

Compressed folder topologies.zip contains parameterization of the CTA molecule in all used force-fields.

4. References

(1) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 1996, 14 (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(2) Thole, B. T.; van Duijnen, P. T. A General Population Analysis Preserving the Dipole Moment. *Theor. Chim. Acta* 1983, 63 (3), 209–221. https://doi.org/10.1007/BF00569246.

(3) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* 1993, 97 (40), 10269–10280. https://doi.org/10.1021/j100142a004.
(4) Wolfram Research, I. Mathematica. Wolfram Research, Inc.: Champaign, Illinois.

(5) Albuquerque, R. Q.; Timme, A.; Kress, R.; Senker, J.; Schmidt, H. W. Theoretical Investigation of Macrodipoles in Supramolecular Columnar Stackings. Chem. - A Eur. J. 2013, 19 (5), 1647–1657. https://doi.org/10.1002/chem.201202507.

(6) Israelachvili, J. N. Thermodynamic Principles of Self-Assembly. In Intermolecular and Surface Forces; 2011; pp 503–534. https://doi.org/https://doi.org/10.1016/C2009-0-21560-1.

(7) van Bommel, K. J. C.; van der Pol, C.; Muizebelt, I.; Friggeri, A.; Heeres, A.; Meetsma, A.; Feringa, B. L.; Esch, J. van. Responsive Cyclohexane-Based Low-Molecular Weight Hydrogelators with Modular Architecture. Angew. Chem. Int. Ed. 2004, 43 (13), 1663 –1667. https://doi.org/10.1002/anie.200352396.