David DELLEMME[(1)]*, Maxime LECLERCQ[(1)], Mathieu FOSSEPRE[(1)], Jie LI[(2)], Qian QIN[(2)], Alain M. JONAS[(2)], Karine GLINEL[(2)], Mathieu SURIN[(1)]

[(1)] Laboratory for Chemistry of Novel Materials, Centre of Innovation and Research in Materials and Polymers (CIRMAP), University of Mons – UMONS, 20 Place du Parc, 7000 Mons, Belgium. E-mail: david.dellemme@umons.ac.be

[(2)] Institute of Condensed Matter and Nanosciences, Bio- and Soft Matter, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium.

Folding and self-assembly of sequence-defined oligomers containing nucleobases

Summary

The synthesis of oligomers and polymers is constantly improving since its inception. Nowadays, we are far from the uncontrolled, heterogeneous mixtures of polymer chains that used to be characteristic of the samples. Synthesis has now reached a stage where it is possible to precisely control the sequence of monomers and the length of each individual chain, allowing us to obtain homogeneous samples. This has long been achieved by Nature, exhibiting a myriad of tailor-made biopolymers, such as proteins or DNA, to perform various biological functions. Sequence-defined polymers (SDPs) and sequence-defined oligomers (SDOs) constitute thus a blooming field, finding applications in various domains.

However, even if it is now clear that the sequence of monomers is a powerful tool to tune the properties of polymeric materials, our understanding of the links between the sequence, the 3D structure and the function of sequence-defined macromolecules remains elusive. The present work aimed at studying short synthetic SDOs functionalized with nucleobases. During the research works of this Master's thesis, their conformations and self-assemblies were studied by means of circular dichroism spectroscopy, molecular dynamics simulations and atomic force microscopy. This joint experimental and theoretical approach helped to understand supramolecular assembly and the dynamics of interactions.

Keywords

Supramolecular chemistry, Sequence-defined oligomers, Folding, Self-assembly

1. Introduction

1.1. General introduction to sequence-defined macromolecules

Sequence-defined macromolecules are ubiquitous in Nature, whose functional aspects are built through complex and highly organized networks of interactions and self-assembly of discrete molecules [1]. Proteins or nucleic acids illustrate well the prime importance of the order of monomers within a chain, with a perfectly defined organization of amino acids and nucleotides, respectively. This definition in the primary structure, or sequence, allows these molecules to exhibit remarkable properties like, for example, biocatalysis for enzymes or genetic information storage for DNA. It is also clear that modifying

the primary structure, even a single monomer or only one atom, can sometimes lead to major alterations in the functions, as exemplified by post-translational modifications of proteins [2].

It is thus tempting to reproduce this absolute control of the sequence within synthetic chains to fine-tune the properties of soft materials. For years, research focused on finding efficient ways to synthesize these sequence-defined macromolecules, which led to the emergence of various synthetic routes [3]. Thanks to these progresses, sequence-defined polymers (SDPs) and sequence-defined oligomers (SDOs) can now be envisioned for a broad range of applications inspired by their natural counterparts, e.g. catalysis, information storage, controlled folding, and self-assembly [4].

1.2. Objectives and studied compounds

By controlling the position of each monomer unit along the chain, SDOs can be designed to reach specific 3D structures for targeted properties and functions. Furthermore, the self-assembly of chains, such as the formation of duplexes, can be harnessed by carefully pre-organizing complementary units in the primary structure along the backbones [5]. Following this idea, several studies reported the functionalization of molecules with nucleobases in order to drive the self-assembly through complementary hydrogen-bonding [6, 7, 8]. In this work, complementary SDOs were designed with a control in the sequence of pending nucleobases, which are thought to promote self-assembly through molecular recognition. SDOs have been functionalized with three canonical bases (guanine, **G**; thymine, **T**; cytosine, **C**) and one non-canonical base (a 2,6-diaminopyridine derivative, **D**), see Figure 1. They are complementary by pairs: thymine with diaminopyridine, and guanine with cytosine. The objective of my research during this Master's thesis was to better understand the relevant conformations of these SDOs, their folding and the dynamics of interactions. These SDOs are being synthesized in the group of UCLouvain, controlling both sequence definition and chirality at the monomer level [9, 10]. In this report, we focus only on the TG dimer, which shows a unique behaviour among other oligomers studied during our Master's thesis. The TG dimer is isotactic with a chirality *R* at the monomer level.

1.3. Methods

To study these SDOs, a joint experimental and theoretical approach was developed, as briefly described below.

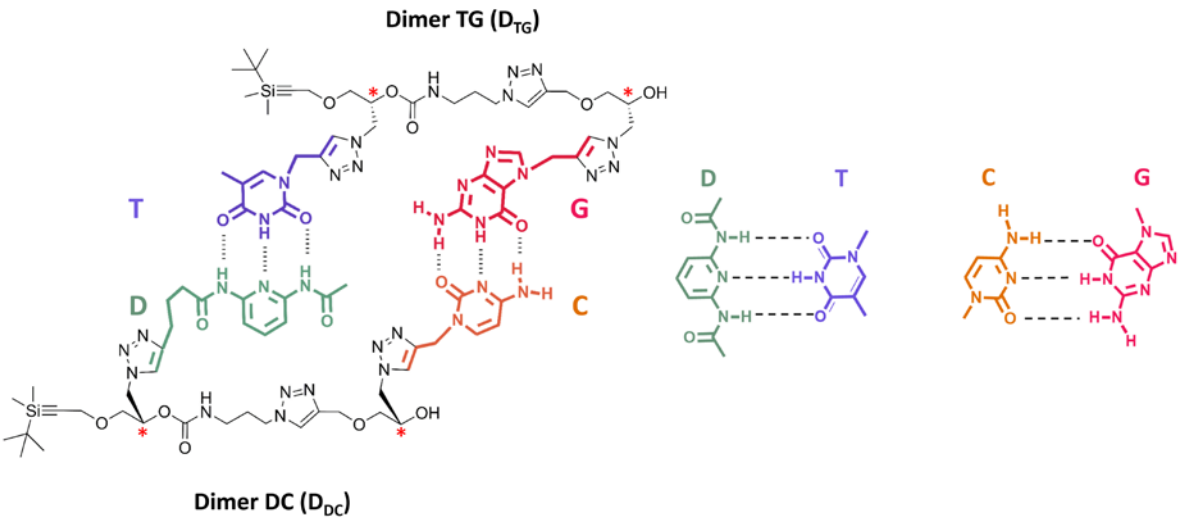

Figure 1. Chemical structures of some of the molecules studied in this work. The nucleobases are guanine, **G**, in pink; thymine, **T**, in purple; cytosine, **C**, in orange; and 2,6-diaminopyridine derivative, **D**, in cyan. The complementary hydrogen-bonding patterns are shown on the right as black dotted lines. Stereocenters are shown with a red asterisk.

As an experimental technique, circular dichroism (CD) spectroscopy is a powerful tool to study the conformations of chiral molecules in solution. It has been broadly used to study biomolecules, like proteins or DNA, for which the CD signals can be linked to specific secondary structures. However, it is much more difficult to clearly interpret the CD signal of a molecule of unknown structure [11]. Still, the CD signal is very sensitive to conformational changes and it is thus possible to detect the formation or deformation of a chiral self-assembled supramolecular structure while varying several parameters such as solvent, temperature, concentration. Circular dichroism is the difference in absorbance, ΔA, between left- and right-circularly polarized light and can be developed following the Beer-Lambert law:

$$CD = \Delta A = A_{left} - A_{right} = (\varepsilon_{left} - \varepsilon_{right})\, Cl \qquad \text{(Eq.1)}$$

Where ε is the extinction coefficient, C is the concentration and l is the optical path length. Chiral molecules have the particularity to absorb differently left- and right-circularly polarized light and thus to exhibit a non-zero CD signal, which can be positive or negative.

All spectra were recorded with a Chirascan Plus CD Spectrophotometer from *Applied Photophysics*, with a bandwidth of 1 nm and a time per point of 1 s.

In addition to CD spectroscopy, which was used to characterize the samples in solution, atomic force microscopy (AFM) allowed us to study the surface topography of thin deposits of SDOs. In brief, this technique involves a nanoscale tip carried by a cantilever. The tip scans the surface and measures its height through its interactions with the sample, thus giving a 3D view of the surface topography. A droplet of 10 µL of the same solutions as used in CD was deposited on freshly-cleaved mica surfaces and the solvent was evaporated for at least one night. We used AFM in Tapping-Mode, which reduces the contact between the tip and the sample and therefore reduces its deformation [12].

Molecular dynamics (MD) simulations were used to study the structure and dynamics of sequence-defined oligomers computationally. In MD simulations, the molecules are represented with a set of particles (the atoms) connected together by springs to model the covalent bonds. Molecular conformations can thus be defined by the coordinates of each atom. The dynamics of such conformations is calculated at a certain temperature in an iterative way by solving Newton's equations of motion, following the theories of classical mechanics. Each step is separated from the next one by a discrete timestep, Δt, usually fixed in the range of 0.5-2 fs (10^{-15} s) to follow the period of the highest frequency related to dynamics of molecules, i.e., the molecular vibration. The calculation of the forces that apply to each atom at a given step allow to predict the evolution of the atomic coordinates according to the Velocity Verlet algorithm [13]. The potential energy of the system, from which are derived the forces acting on the atoms, is calculated by means of a force field which takes into account the chemical nature and environment of the atoms as well as their coordinates. The total potential energy is calculated as a sum of energy terms describing bond stretching, angle bending, van der Waals interactions, etc. The result of a MD simulation is a trajectory, a set of conformations and their relative energy.

In this work, we used the General Amber Force Field (GAFF 2.1) for MD simulations of oligomers [14]. The GAFF force field has proven to be reliable to reproduce several experimental data of biomolecular and organic compounds, [15, 16] and was adapted in our group for the study of sequence-defined oligomers [10, 17].

2. The case of the dimer TG: detection of a self-assembled structure

The chemical structure of the TG molecule is shown in Figure 1. This dimer was studied by CD spectroscopy in acetonitrile at 20 °C at increasing concentrations (Fig. 2 a). The spectra recorded at low concentrations (between 10 and 25 µM) are dominated by a positive peak around 210 nm. The three other dimers studied, of sequence DC, CT and GD have similar spectra (results not shown).

However, at higher concentrations (between 50 and 200 µM, Figure 2 a right), the shape of the spectra completely changes with the appearance of other peaks, both positive and negative. While it is very difficult to relate this spectrum to a specific conformation, this result clearly indicates a conformational change of the dimer TG at high concentrations. This is not observed for the other dimers (not shown). A possible explanation would be the formation of a self-assembled supramolecular structure, where several dimers interact with each other. To verify this hypothesis, a temperature ramp was carried out: a sample of the dimer at 250 µM was first heated from 20 to 70 °C and then cooled from 70 to 20 °C (Fig. 2 b).

When the sample is heated, the signal changes gradually until 50 °C, a temperature at which the signals are stabilized. The two consecutive negative peaks centred around 290 nm and 265 nm progressively rise to zero. The two positive peaks following them continuously decrease, with a very slight, broad positive peak between 260 nm and 220 nm remaining at 50°C. Finally, the strong negative peak at 205 nm and the positive one at 195 nm reverse. Again, these modifications suggest a conformational change due to the increase in temperature. The resulting signal at 70 °C is in fact very similar to the one observed at low concentrations (Figure 2 a left). This could mean that the energy brought by temperature breaks low-energy interactions that stabilize the assemblies of dimers TG. Then, when the sample is cooled, the initial signal reappears when the temperature drops below 50 °C, which indicates the formation of the supramolecular structure in solution. This suggests that the interactions that stabilize the dimers are reversible. The solvent is also an important parameter. Indeed, in a mixture of acetonitrile/dimethyl sulfoxide 9/1 (v/v), the CD signal of the self-assembled structure disappears. Dimethyl sulfoxide is known to be a strong H-bonds competitor, which suggests the presence of this interaction to stabilize the assembly.

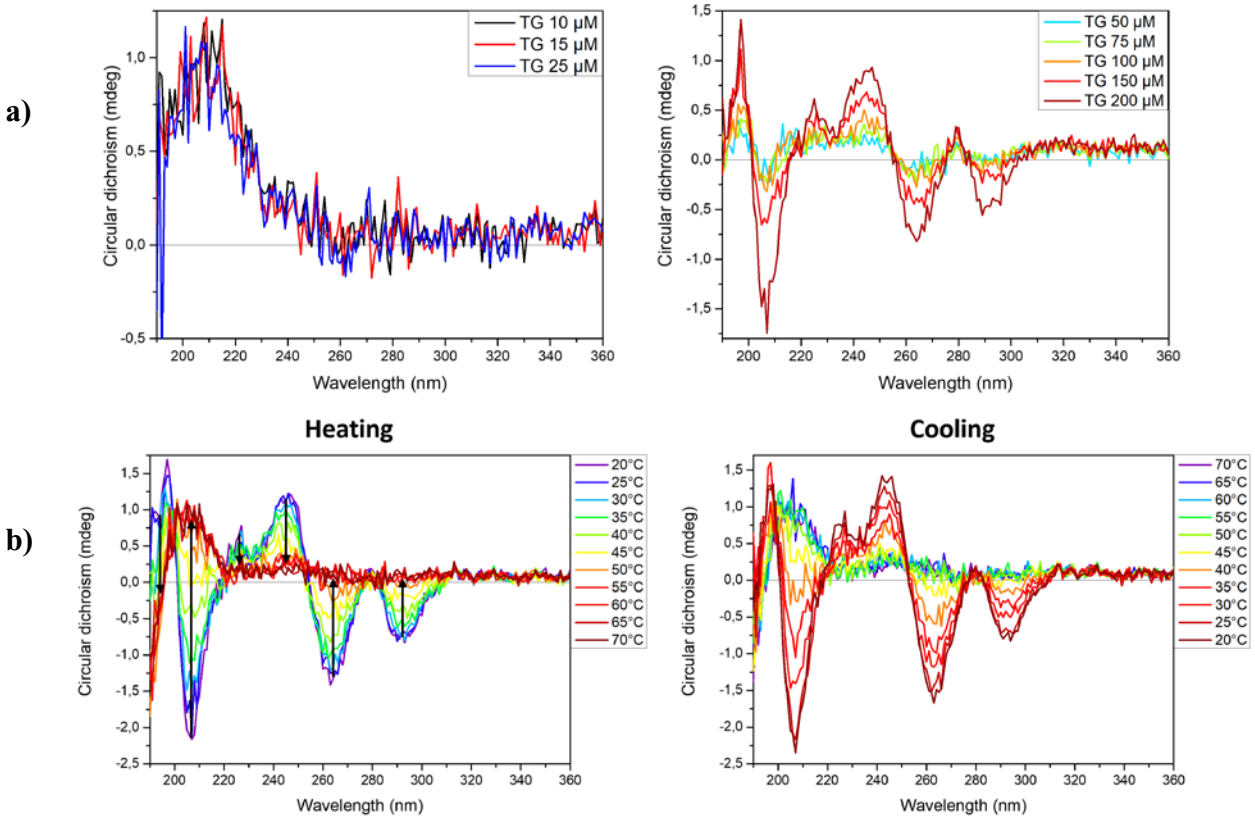

Figure 2. **a)** CD spectra of dimer TG at increasing concentrations. Spectra measured in acetonitrile, at 20 °C. **b)** CD spectra of dimer TG at various temperatures, in acetonitrile, at 250 µM. The heating/cooling rate is 1 °C/min, with a spectrum measured every 5 °C.

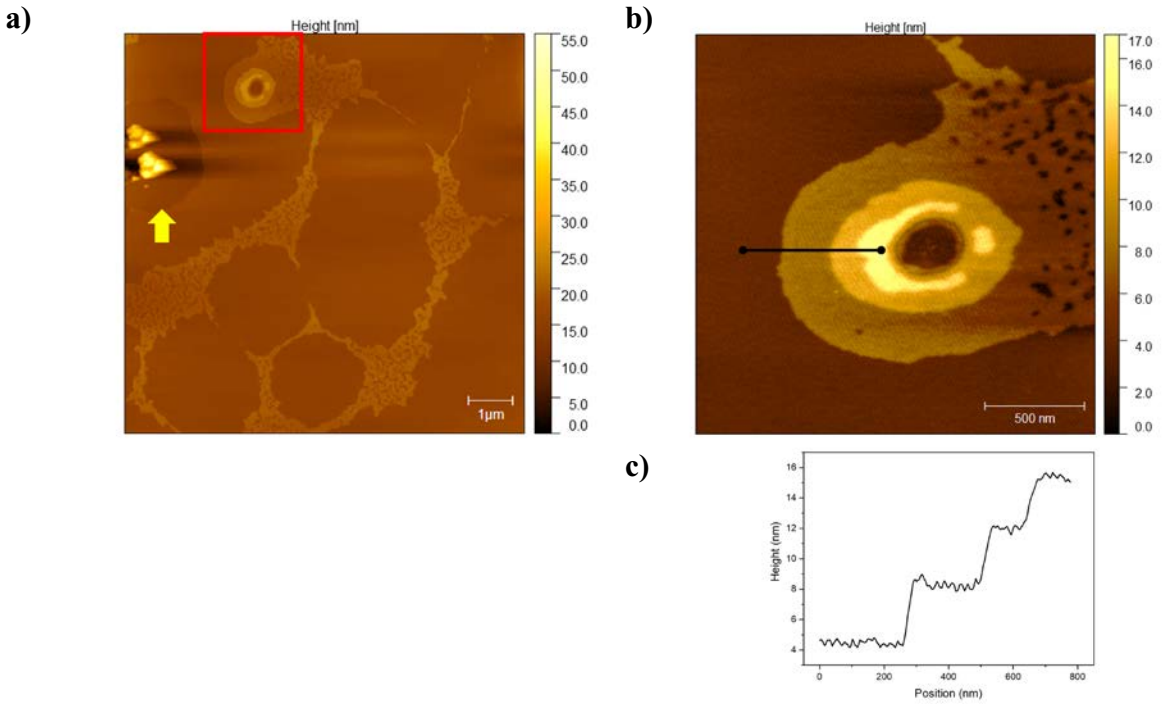

Figure 3. Tapping-Mode AFM height image of a TG dimer deposit obtained from a 500 µM solution. **a)** 9 x 9 µm² image. **b)** 2 x 2 µm² image of the area inside the red square in the image on the left. **c)** Height profile measured along the black line depicted in b).

Thin deposits of dimer TG shows an organized solid-state morphology, as observed by AFM (Fig. 3). The yellow arrow indicates the darkest area, i.e. the one with the lowest height, which is probably the substrate surface. A homogeneous layer lies on this surface. In some places, superimposed layers with a greater height are visible. For example, the area inside the red square (Fig. 3 b) shows three superimposed layers on top of the first homogeneous layer. The height profile (Fig. 3 c) measured along the black line on b) indicates a layer-by-layer organization with a regular height of the layers, on average of 3.5 ± 0.5 nm. Knowing that the end-to-end distance of the extended dimer is about 2 nm, one layer could consist of at least two molecules in its width. A hypothesis would be the formation of molecular bilayers, where the dimers face each other and bring in close proximity their nucleobases. When densely stacked, they could form both hydrogen-bonds between nucleobases and π-stacking of aromatic cycles to stabilize their assembly. Further work is needed to identify the precise layering scheme. The formation of 2D bilayers on mica has already been observed for sequence-defined peptoids bearing hydrophilic and hydrophobic substituents [18].

These results led us to study the dimer TG in solution by MD simulations. The results indicate that the molecule is very flexible and has the tendency to fold into compact structures, as shown by the mean radius of gyration of about 6 Å, whereas the end-to-end distance of the extended structure is about 22 Å. Flexibility arises from the presence of methylene groups in the backbone and allows the formation of intramolecular H-bonds and π-stacking of the aromatic cycles. The strongest H-bond occurs around 20% of the simulated time, between the guanine and the carbamate group of the backbone. This shows that not only the bases but also other polar and aromatic moieties participate in the interaction network.

An assembly of two TG molecules in interaction was simulated for a MD time of 1 ms to understand how the molecules could interact in stable self-assembled structures. Despite the high flexibility of the dimers, persistent interactions are observed. The guanine and thymine of a chain remain π-stacked once for more than 30 consecutive ns while the two thymine bases pile up for more than 20 ns (not consecutive) in the same time interval. These π-stacking interactions coincide

with a hydrogen-bonding interaction between the two guanine moieties. Moreover, the most stable conformation detected during the simulation, reported in Figure 4, is found during the same time interval, being 5.3 kcal/mol more stable than the second most stable conformation. The interactions mentioned above are shown on this snapshot. Three nucleobases are π-stacked and the two guanine moieties, facing each other, interact by H-bonds. These results are in good agreement with the experimental ones, showing the possible formation of a self-assembled structure of TG molecules stabilized by hydrogen-bonding interactions and π-stacking of aromatic cycles.

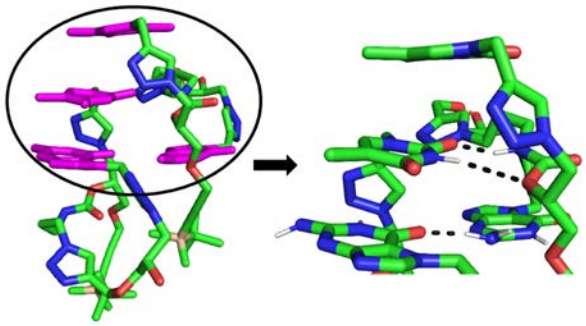

Figure 4. 3D-views of the most stable conformation detected during the MD simulation of the assembly of two dimers TG. Left: Only heavy atoms are shown. The nucleobases are coloured in pink. Right: Only heavy atoms and polar hydrogen atoms are shown. Hydrogen-bonds are shown in black dotted lines.

3. Conclusion and perspectives

Our joint experimental and theoretical study of sequence-defined oligomers brought insights on their conformations and assembly in solution. In the case of the TG dimer, a reversible supramolecular structure was observed by CD spectroscopy, with stabilization by hydrogen-bonding and stacking of aromatic cycles, as suggested by MD simulations. To determine more precisely the 3D structure of this self-assembly, it would be interesting to calculate CD spectra by quantum-chemical calculations on relevant conformations from MD. On an experimental point of view, X-ray reflectometry could be useful to get insights on the layer thickness of TG dimer, to complement the AFM measurements. The fact that TG dimers seem to be able to self-assemble shows that the strong flexibility of these oligomers allows them to interact even without complementary base pairs. The assembly of longer SDOs, complementary or not by their sequence of nucleobases, is currently under investigation.

4. Acknowledgements

The collaboration between University of Mons and Université Catholique de Louvain is supported by the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek under EOS project no. 30650939. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Wallonia Region. D.D. thanks the French Community of Belgium for the FRIA grant supporting a doctoral scholarship. K.G. and M.S. are Senior Research Associates of the F.R.S.-FNRS.

References

[1] J.-F. Lutz, J.-M. Lehn, E. W. Meijer, K. Matyjaszewski, *Nat. Rev. Mater.* **2016**, *1*, 16024.

[2] K. Zhao, Y. Ju, S. Li, Z. Altaany, R. Wang, G. Yang, *EMBO Rep.* **2014**, *15*, 792.

[3] S. C. Solleder, R. V. Schneider, K. S. Wetzel, A. C. Boukis, M. A. R. Meier, *Macromol. Rapid Commun.* **2017**, *38*, 1600711.

[4] R. Aksakal, C. Mertens, M. Soete, N. Badi, F. Du Prez, *Adv. Sci.* **2021**, *8*, 2004038.

[5] S. C. Leguizamon, T. F. Scott, *Nat. Commun.* **2020**, *11*, 784.

[6] R. McHale, R. K. OReilly, *Macromolecules* **2012**, *45*, 7665.

[7] A. del Prado, D. González-Rodríguez, Y. L. Wu, *ChemistryOpen* **2020**, *9*, 409.

[8] M. Surin, *Polym. Chem.* **2016**, *7*, 4137.

[9] J. C. Barnes, D. J. C. Ehrlich, A. X. Gao, F. A. Leibfarth, Y. Jiang, E. Zhou, T. F. Jamison, J. A. Johnson, *Nat. Chem.* **2015**, *7*, 810.

[10] J. Li, M. Leclercq, M. Fossepré, M. Surin, K. Glinel, A. M. Jonas, A. E. Fernandes, *Polym. Chem.* **2020**, *11*, 4040.

[11] A. Rodger, B. Nordén, in "Circular Dichroism and Linear Dichroism", *Oxford University Press* **1997**.

[12] P. Eaton, P. West, in "Atomic Force Microscopy", *Oxford University Press*, **2010**.

[13] L. Verlet, *Phys. Rev.* **1967**, *159*, 98.

[14] J. Wang, R. M. Wolf, J. W. Caldwell, P.A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.

[15] S. Zhu, *J. Chem. Inf. Model.* **2019**, *59*, 4239.

[16] S. Kashefolgheta, M. P. Oliveira, S. R. Rieder, B. A. C. Horta, W. E. Acree Jr., P. H. Hünenberger, *J. Chem. Theory Comput.* **2020**, *16*, 7556.

[17] M. A. Reith, S. Kardas, C. Mertens, M. Fossépré, M. Surin, J. Steinkoenig, F .E. Du Prez, *Polym. Chem.* **2021**, *12*, 4924.

[18] E. J. Robertson, A. Battigelli, C. Proulx, R. V. Mannige, T. K. Haxton, L. Yun, S. Whitelam, R. N. Zuckermann, *Acc. Chem. Res.* **2016**, *49*, 379.