

USER MANUAL

Chitin Builder: a VMD tool for the generation of structures of chitin molecular crystals for atomistic simulations

David C. Malaspina* and Jordi Faraudo†
*Institut de Ciència de Materials de Barcelona (ICMAB-CSIC) ,
Campus UAB, E-08193 Bellaterra, Barcelona, Spain.*
(Dated: January 9, 2024)

Chitin is the second most abundant organic material in nature after cellulose and its study is now of great interest in the field of biocompatible and eco-friendly materials. Computational studies of natural polymers such as Cellulose and Chitin are often hindered by the practical difficulties in generating structures suitable for the simulations. This has motivated the recent introduction of a cellulose builder program [T. C. F. Gomes et al. J. Comput. Chem. 2012, 33, 1338–1346.] that generates coordinate, structure and topology files for atomistic simulations of cellulose crystal polymorphs. Here we present the Chitin Builder tool, a program that enhances the Visual Molecular Dynamics (VMD) environment with the ability to generate coordinate and structure files of chitin organic crystals. The program generates cartesian coordinates and atomic connectivity and structure files for crystalline structures of chitin polymorphs α and β . Crystal structures of any size with or without bonds in the periodic directions can be easily build. The resulting structure is automatically saved in PDB and PSF format (used by well known simulation packages such as NAMD) and it can be easily converted to many other file formats using VMD build-in features.

Keywords: VMD, Modelling of carbohydrates, Chitin, Atomistic simulation, Molecular Dynamics, CHARMM

I. INTRODUCTION

Chitin is a polysaccharide present in the exoskeleton and internal structure of many invertebrates like molluscs, crustaceans, insects, fungus, algae, and other related organisms [1, 2]. It is so prevalent in nature that it constitutes the second most abundant polymerized form of carbon in Earth. From the point of view of material sciences, chitin is biorenewable, environmentally friendly, biocompatible and biodegradable material. It has applications as a chelating agent, water treatment additive, drug carrier, biodegradable pressure-sensitive adhesive tape, wound-healing agents and many others [2–4].

Chemically, chitin is a polysaccharide composed of repeated units of (1-4) N-acetyl-D-glucosamine (2-acetamido-2-deoxy- β -D-glucopyranose). As other polysaccharides such as cellulose, chitin is usually found in crystal structures with several inter-chain hydrogen bonds that produce high order structures. There are three known crystal allomorphs [5]: α , β and γ chitin. The crystal structures of α and β chitin (the most abundant allomorphs) are known [6, 7] but the structure of the rare γ chitin allomorph is still under debate [5, 8, 9].

The knowledge of the crystal structures of chitin with atomistic resolution is important from both a fundamental and practical point of view, since it opens the

possibility to predict the properties of chitin based materials and derivatives (mechanical, thermal, interaction with solvents,...) from the knowledge of its molecular structure. Starting from the atomic coordinates provided by crystal structures, it is possible to perform Molecular Dynamics (MD) simulations of chitin and study its properties and its interactions with other materials. However, up to date, there are only a few works that deal with all-atomic MD simulations of chitin [10–13]. These studies explore important practical questions such as the interaction of chitin with proteins or the mechanical properties of chitin.

The lack of atomistic simulations of chitin, is even more surprising when we compare this situation with the case of cellulose, which is the other most abundant polysaccharide. In the case of cellulose, there are many atomistic simulation works, deriving the most diverse features of cellulose from the known crystal structure [14]. We think that one possible reason for this difference is the availability of a cellulose builder tool [15] that allows an easy build up of atomistic configurations and structure/topology files that can be used for MD simulations. Since these materials are complex materials, the build up of the files required for the simulations is not a trivial task. It is clear that the existence of tools that facilitate the build up of appropriate files for atomistic simulation of polymeric organic crystals will fuel the use of simulation techniques for the understanding of these important materials.

In this work we present a chitin builder tool implemented as a VMD plugin, that produces files containing atomic coordinates and topology information of pure α

* davidcmalaspina@gmail.com

† jfaraudo@icmab.es

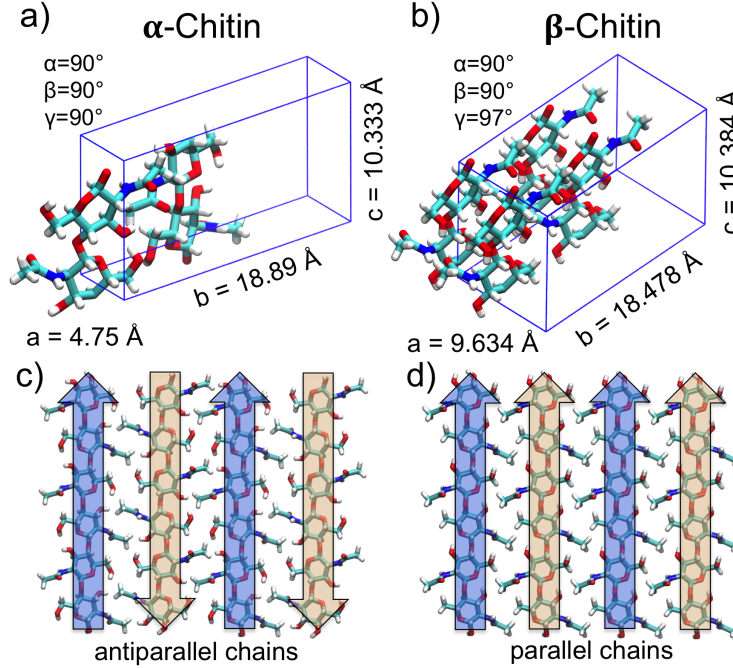


Figure 1. Structure of chitin crystal allomorphs α and β . a) α chitin primitive unit cell. b) β chitin expanded unit cell (corresponding to 2 primitive unit cells). c) scheme of the antiparallel organization of the chains in α chitin, indicated by arrows over a crystal generated with the input for the number of replicas $a=1$ $b=2$ $c=3$. d) scheme of the parallel chain organization in β chitin, indicated by arrows over a crystal generated with the input for the number of replicas $a=1$ $b=2$ $c=3$ (note that an extra plane of molecules is below the image in this case). Chitin molecules are shown in bond representation in all cases (O in red, H in white, N in blue and H in white). Figure made with VMD. [16].

and β chitin crystals of arbitrary size from experimental chitin crystal unit cells determined from X-ray crystallography [6, 7].

This plugin will greatly facilitate the process of generation of input files (coordinates, structures, topology) for atomistic simulations and we expect that it will fuel the use of these techniques in the study of these materials. Future developments of these plugin will incorporate the generation of crystal structures of other polymeric crystals.

II. METHODS

A. Crystal structures of Chitin

As we mentioned in the Introduction, there are three known chitin crystal allomorphs, α , β and γ chitin. However, we will not consider γ chitin since its structure is not well known and in fact it is sometimes considered as a mixture of α and β chitin [5, 8, 9]. The crystal structures of α and β chitin allomorphs are shown in Figure 1. In the case of α chitin, we show in Figure 1a the primitive cell as determined by from X-ray diffraction in Ref. [6]. This is the unit cell employed in our

program for α chitin. Note that α chitin crystals have a characteristic antiparallel organization of chitin chains, as shown in Figure 1c.

In the case of β chitin, it is important to note that β chitin is partially soluble in water and in fact interaction with water seems to modify the crystal structure [5]. For this reason, we consider its anhydrous form, as determined by X-ray diffraction in Ref. [7]. The structure of the anhydrous β chitin allomorph is shown in Figure 1. The unit cell of β chitin employed in our program is the cell shown in Figure 1b. This is an expanded cell, corresponding to double of the primitive cell in both the a and b directions, that facilitates the understanding of the structure and also simplifies the replication algorithm employed in our program. This is due to the characteristic parallel organization of β chitin chains in the crystal (see Figure 1d). This organization of the extended crystal is more easily generated by simple replications of the unit cell shown in Figure 1b, without introducing any practical limitation in simulation studies.

An important feature determined in experiments is the presence of deacetylated groups (chitosan groups) in the fibril structure. This degree of deacetylation in chitin is of approximately between 5 to 15% [2] although

there are species that produce chitin of high purity. In our program, we will consider pure chitin so we do not consider the presence of deacetylated groups in the structures. This possibility is too complex to be incorporated in our program at the present time until further modelling studies clarify a realistic way to incorporate this chemical feature.

B. Generation of coordinates and topology of crystals

The program generates atomic coordinates of α and β chitin crystals with dimensions entered by the user. The dimensions of the crystal must be a multiple of the unit cell shown in Figure 1a for α chitin and in Figure 1b for β chitin crystals. The atomic coordinates and cell vectors of the cells shown in Figures 1a and 1b were converted by us from the cif files containing X-ray crystallography data supplied in Refs. [6, 7] and are distributed with our program in PDB format in order to be understood by the VMD program. These files are given in the folder */structures* (alpha-4residues.pdb and beta-4residues.pdb). The user specifies how many replicas wants to obtain in each of the three main axis (a , b and c) and the program (which is a TCL script running inside VMD) loads the required PDB file with the atomic coordinates of the unit cell and replicates the coordinates. Once the coordinates are calculated, an intermediate file is generated containing the coordinates of the atoms in the new crystal cell in PDB format (crystal-alpha-temp.pdb or crystal-beta-temp.pdb). These coordinate files can be used inside input files of certain calculations (for example, geometry optimizations or structural relaxations with DFT packages) but they do not contain enough information for force-field based molecular dynamics (MD) simulation packages. For these MD simulations, topology and structural information (such as atom type, atom connectivity, partial charges,...) is required. In order to generate this additional information, we employ the topology from the carbohydrates section of CHARMM36 force field [17]. Chitin monomers correspond to the standard CHARMM36 residue BGLCNA (beta-N-acetyl-glucosamine). But here, we modify the name of this residue to BGNA in order to have a 4 characters residue name which is the preferred residue name length for several simulation and analysis packages and it is also the residue length specified in the standard PDB format. The corresponding topology file from CHARMM36 "top_all36_carb.rtf", modified with the new residue name, is included in the */structures* folder. Our program automatically calls the psfgen package included in VMD to generate a new PDB coordinate file and a PSF structure file combo containing the same crystal coordinates generated before but now includ-

ing the BGNA residue name, CHARMM36 atom types, charges, bonds, angles and dihedrals. In the generation of the structure PSF file, each chitin chain in the new crystal cell is treated as an individual segment. Each chitin residue in each chain is linked by the "14bb" patch in the topology file, that correspond to the (1-4)beta glycosidic bond. Finally, an additional bond between the first and last residue of the chain can be optionally included, by selection of the option "periodic bonds" by the user (see details of the user interface below). This possibility allows the simulation of infinitely long chains with the use of periodic boundary conditions. The generated pdb and psf files are finally saved into the working folder and are ready to be employed in force field MD simulations. For completeness, we also provide a parameters file ("par_all36_carb.prm" in the ForceField folder) with CHARMM36 carbohydrate standard Lennard-Jones, bonds, angles and dihedrals sections of the CHARMM36 force field needed for the simulation.

III. PROGRAM DESCRIPTION, USE AND FEATURES

A. Brief description of the program

The code is a plugin for VMD written in tcl/tk v8.4 programming language. The source code contains two main parts: the code that obtains the atomic coordinates of the atoms of the crystal and the structure and topology of the crystal (see Methods section) and a code responsible for the graphical user interface (gui). The program generates two main outputs: a coordinate file (PDB) containing the position of all the atoms in the generated structure and a topology file (PSF) containing all the bonds, angles and dihedrals according to CHARMM36 carbohydrate section [17]. These output files are named crystal-alpha-psf.pdb/psf or crystal-beta-psf.pdb/psf depending on the allomorph, and are stored in the working folder chosen by the user. This two files, plus the included CHARMM36 parameters file (located in the */ForceField/* folder) allows the user to easily start a molecular dynamics simulation using the NAMD simulation program [18] that accompanies VMD. Also, using VMD, the users can easily export the data from these two files (PDB and PSF) to other coordinates formats or convert the topology to the formats required by other programs such as GROMACS [19, 20] using the topotools plugin included in VMD.

B. Installation

As we mentioned before the code is a plugin for VMD. In the Readme.md file we detailed the installation process. Basically, the plugin can be easily installed by simply moving the downloaded plugin folder inside the standard plugin folder of VMD (the folder /plugins/noarch/tcl/ inside the folder where VMD is installed). Usually administrator privileges are needed to copy into the VMD installation folder. Also note that the specific location at which VMD is installed vary depending on the operating system and the user VMD installation preferences, as explained in VMD documentation. We also included a bash script that performs an automatic installation for Ubuntu Linux system.

C. Graphical user interface

All the options of our plugin can be controlled from a graphical user interface (GUI), shown in figure 2. This GUI allows the user to switch between the different options for the chitin crystal generation (Figure 2). It also provides a quick view to the unit cell data as well as the cell vectors describing the generated crystal, to be employed in a MD simulation.

In Figure 2, we also indicate the different sections of the GUI which are the following:

Section 1) of the gui correspond to the "help" menu. It contains two items: "About", that contains information of the developers and "Documentation", that redirects the user to the repository code webpage with the latest version of the user manual. Also, the last version of the code is available at this repository.

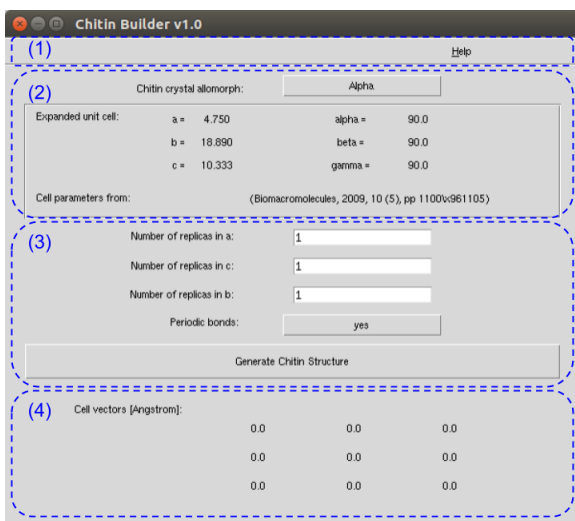


Figure 2. Chitin builder graphical user interface sections

Section 2) of the gui corresponds to the selection of the crystal allomorph. It contains a button that allows the user to choose between α and β chitin allomorphs as explained in the Method section. Under this button a description of the unit cell data including a , b and c cell vectors and α , β and γ angles is shown (corresponding to the unit cells in Figures 1 a and c) as well as the reference to the primary publication where the experimental data (from which this unit cell was build) can be found.

Section 3) of the gui allows the user to input the number of replicas in a , b and c crystallographic directions of the new crystal that is going to be generated. The inputs need to be an integer and the maximum number that the calculation allow is such that the number of replicas do not exceed 9.999×10^3 residues (the maximum allowed by VMD). Under the input for the number of replicas, a button allows the user to select between a structure with ("yes") or without ("no") periodic bonds in the longitudinal direction of the polysaccharide chain. This type of bonds allow the simulation of infinitely long molecular crystals by using periodic boundary conditions. At the bottom of this section the button "Generate Chitin Structure" will prompt dialog box where the user need to choose the destination folder and after this the code will initialize for the replication and topology generation. Once the calculation is done the generated structure will appear in the VMD display and the output files crystal-alpha-psf.psf/pdb or crystal-beta-psf.psf/pdb will be generated on the destination folder selected by the user. Data of the crystal cell will be printed on the gui, on the main console and in the log file "crystal.log".

Finally, section 4) of the gui displays the cell data of the generated chitin structure. The cell data is displayed as a matrix of the cell vector data (the employed units are Å). The first row correspond to the 3 cartesian components (i, j, k) of the a vector of the cell, second row correspond to the 3 components of the b vector of the cell and third row to the 3 components of the c vector of the cell. This data is provided in this format so the user can use it directly in a configuration file for a MD simulation (as in a simulation with NAMD for example). This useful data is also printed on the main console and on the log file.

D. Use from the command line (API mode)

Actually, it is possible to invoke the Chitin Builder functionality without using the Graphical User Interface described in the previous subsection. Once the software is installed, from the VMD menu select the Tk Console instead of selecting the Graphical user interface. Now, at the Tk Console the Chitin Builder can be called by typing:

```
package require chitin
```

We also need to set the working directory for chitin builder:

```
set chitin::fname myfolder
```

where "myfolder" is the desired working directory (both for temporary usage and final data storage). This adds a new "replicate" functionality that allows the creation of chitin crystals. The command has the following usage:

```
chitin::replicate n1 n2 n3 crys1 per1
```

The arguments $n1$, $n2$ and $n3$ should be integers representing the number of replicas in x , y and z respectively. The argument $cris1$ should be the string "Alpha" or "Beta" depending on the chitin crystal structure to replicate. The argument $per1$ should be the string "yes" or "no" depending if you want periodic bonds between the first and last residue.

After using the Chitin Builder through the VMD command line, an error appears if afterwards you try to generate a crystal using the GUI. If you wish to use the GUI, restart VMD.

IV. EXAMPLES

Using the chitin builder plugin we have generated several structures, shown in Figure 3. All the generated structures (PDB + PSF files) are included in the /examples/ folder as well as a screen capture of the gui that generated the respective structure.

In the first example (Figure 3a), we have created an α chitin crystal with the input parameters for the replicas $a=4$, $b=2$ and $c=4$ and the option of periodic bonds was turn to "no". In Figure 3a we show the resulting crystal produced by the program. Using this input, the program produces a crystal cell of 19.0 Å in the a direction, 37.78 Å in the b direction and 41.33 Å in the c direction. This structure is provided to the user in the folder /examples/example-a-alpha-4-2-4/.

In the second example (Figure 3b) we created a β chitin crystal with the input parameters for the replicas $a=2$, $b=2$ and $c=4$. The option periodic bonds was also turn to "no". The crystal generated with these options is shown in Figure 3b. It has a length of 19.276 Å, a height of 36.956 Å, and depth of 41.536 Å. Note that this structure is equivalent to the α in Figure 3a, but the number of replicas needed to generate it is different due to the differences in the unit cells, see Figure 1. This structure is provided as an example to the user in the folder /examples/example-b-beta-2-2-4/.

In the third example (Figure 3c) we repeated the creation of a β chitin crystal with the input parameters

for the replicas $a=2$, $b=2$ and $c=4$, but this time the periodic bond option was set to "yes". In Figure 3c) we can observe the effect of setting the periodic bond option to "yes" bonds as a direct link between the first residue of chitin chain with the last residue. In a periodic boundary condition representation in VMD this would produce an infinitely long chitin chain, as seen in this figure, since it will link the chain with its periodic representation. This structure is located in the folder /examples/example-c-beta-2-2-4-periodic/.

In the fourth example (Figure 3d) we created a β chitin crystal with the input parameters for the replicas $a=2$, $b=2$ and $c=2$, resulting in a crystal with the following dimensions: 19.276 Å length, 36.956 Å height and a depth of 20.768 Å. The option periodic bonds was turn to "no". Note that this structure is half the size of the second example (Fig. 3b) in the c direction. The files for this structure are located in the folder /examples/example-d-beta-2-2-2/.

In the fifth example (Figure 3e) we created a β chitin crystal with the input parameters for the replicas $a=4$, $b=2$ and $c=4$, resulting in a crystal with dimensions 38.552 Å, 36.956 Å and 41.536 Å. The option periodic bonds was turn to "no". Note that this structure is twice the size of the second example (Fig. 3b) in the a direction. The files for this structure are located in the

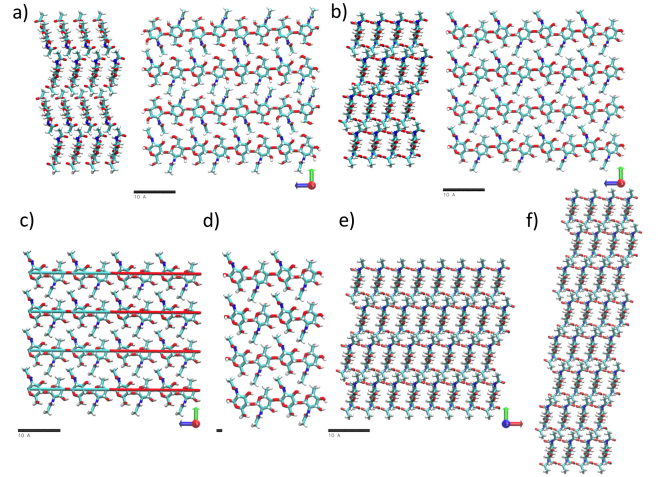


Figure 3. Examples of crystal structures generated by chitin builder. a) α chitin crystal (input parameters $a=4$, $b=2$, $c=4$), x - y plane on the left and y - z plane on the right. b) β chitin crystal (input parameters $a=2$, $b=2$, $c=4$), x - y plane on the left and y - z plane on the right. c) β chitin crystal (input parameters $a=2$, $b=2$, $c=4$) with periodic bonds, y - z plane. d) β chitin crystal (input parameters $a=2$, $b=2$, $c=2$) y - z plane. e) β chitin crystal (input parameters $a=4$, $b=2$, $c=4$) x - y plane. f) β chitin crystal (input parameters $a=2$, $b=4$, $c=4$) x - y plane. Chitin molecules are shown in bond representation in all cases (O in red, H in white, N in blue and C in grey). Figure made with VMD. [16].

folder `/examples/example-e-beta-4-2-4/`.

In the sixth example (Figure 3f) we created a β chitin crystal with the input parameters for the replicas $a=2$, $b=4$ and $c=4$ resulting in a crystal with dimensions 19.276 Å, 36.956 Å and 41.536 Å. The option periodic bonds was turned to "no". Note that this structure is twice the size of 3b) in the b direction. This structure is located in the folder `/examples/example-f-beta-2-4-4/`.

These structures can also be generated from the command line (VMD Tk console), as described in the previous subsection. As an example, the commands needed to generate the second example (Figure 3b) are given in Figure 4.

Figure 4. Example of generation of the crystal structure of Figure 3b from the VMD Tk console. Top: input screen, bottom: output screen.

V. COMPATIBILITY AND DISTRIBUTION

The chitin builder tool is written in tcl/tk as VMD plugin; it will therefore work on any of the platforms supported by the VMD program (Linux, OSX, Windows). The latest version of the package can be obtained from the repository URL https://github.com/soft-matter-theory-at-icmab-csic/chitin_builder/. The repository includes the code, installation instructions for different platforms and examples.

ACKNOWLEDGMENTS

We acknowledge financial support from the Spanish Government through the RTI2018-096273-B-I00 grant and the “Severo Ochoa” Grant SEV-2015-0496 for Research Centres of Excellence awarded to ICMAB. D.C. Malaspina is supported by the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 6655919. We acknowledge discussions with Prof. Lars A. Berghlund and Dr Yamila Garcia about chitin materials.

- [1] P. K. Dutta, M. N. V. Ravikumar, and J. Dutta, *Journal of Macromolecular Science, Part C: Polymer Reviews* **42**, 307 (2002).
- [2] V. Zargar, M. Asghari, and A. Dashti, *ChemBioEng Reviews* **2**, 204 (2015).
- [3] J. L. Shamshina, P. Berton, and R. D. Rogers, *ACS Sustainable Chemistry & Engineering* **7**, 6444 (2019).
- [4] M. N. Ravi Kumar, *Reactive and Functional Polymers* **46**, 1 (2000).
- [5] M.-K. Jang, B.-G. Kong, Y.-I. Jeong, C. H. Lee, and J.-W. Nah, *Journal of Polymer Science Part A: Polymer Chemistry* **42**, 3423 (2004).
- [6] P. Sikorski, R. Hori, and M. Wada, *Biomacromolecules* **10**, 1100 (2009).
- [7] Y. Nishiyama, Y. Noishiki, and M. Wada, *Macromolecules* **44**, 950 (2011).
- [8] M. Kaya, M. Muftaba, H. Ehrlich, A. M. Salaberria, T. Baran, C. T. Amemiya, R. Galli, L. Akyuz, I. Sargin, and J. Labidi, *Carbohydrate Polymers* **176**, 177 (2017).
- [9] D. G. Ramírez-Wong, M. Ramírez-Cardona, R. J. Sánchez-Leija, A. Rugerio, R. A. Mauricio-Sánchez, M. A. Hernández-Landaverde, A. Carranza, J. A. Pojman, A. M. Garay-Tapia, E. Prokhorov, J. D. Mota-Morales, and G. Luna-Bárcenas, *Green Chemistry* **18**, 4303 (2016).
- [10] Z. Yu and D. Lau, *Carbohydrate Polymers* **174**, 941 (2017).
- [11] M. T. McDonnell, D. A. Greeley, K. M. Kit, and D. J. Keffer, *The Journal of Physical Chemistry B* **120**, 8997 (2016).
- [12] Z. Štřelcová, P. Kulháněk, M. Friák, H.-O. Fabritius, M. Petrov, J. Neugebauer, and J. Koča, *RSC Advances* **6**, 30710 (2016).

- [13] K. Jin, X. Feng, and Z. Xu, *BioNanoScience* **3**, 312 (2013).
- [14] D. C. Malaspina and J. Faraudo, *Advances in Colloid and Interface Science* **267**, 15 (2019).
- [15] T. C. F. Gomes and M. S. Skaf, *Journal of Computational Chemistry* **33**, 1338 (2012).
- [16] W. Humphrey, A. Dalke, and K. Schulten, *Journal of Molecular Graphics* **14**, 33 (1996).
- [17] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, and A. D. MacKerell, *Journal of Chemical Theory and Computation* **7**, 3162 (2011).
- [18] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, *Journal of Computational Chemistry* **26**, 1781 (2005).
- [19] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *Journal of Computational Chemistry* **26**, 1701 (2005).
- [20] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *Journal of Chemical Theory and Computation* **4**, 435 (2008).