Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Carbohydrate Self-Assembly at Surfaces: STM Imaging of Sucrose Conformation and Ordering on Cu(100)

Sabine Abb, Nathalie Tarrat, Juan Cortés, Bohdan Andriyevsky, Ludger Harnau, J. Christian Schön, Stephan Rauschenbach, and Klaus Kern*

anie_201901340_sm_miscellaneous_information.pdf

Table of Contents

SI-A: Experimental methods p. 2

SI-B: Theoretical methods p. 3

SI-C: Evaluation of assembly patterns p. 7

SI-A: Experimental methods

1. Sample Preparation

For the electrospray ion beam deposition (ES-IBD) experiment, sucrose (acquired from Sigma-Aldrich (6104) is dissolved in a water-ethanol 1:1 mixture, resulting in a concentration of 10^{-3} M. In order to enhance the ionization, 0.1 % ammonia hydroxide was added to promote deprotonation. Subsequently, these solutions were used with our home-built ES-IBD apparatus[1] to bring the molecule into the gas phase (2-3kV on the emitter, 20 µl/h flow rate, current up to -400 pA, about 1 h of deposition time). Before deposition in UHV, the composition of the ion beam is monitored by a time-of-flight (TOF) mass spectrometer (mass spectrum see Fig. S1), and the singly deprotonated sucrose molecules were mass selected by RF quadrupoles to ensure high-quality samples (reducing the current to 30%-50% of the initial value). During the deposition (at a pressure of $6 \cdot 10^{-10}$ mbar), the molecular ion is decelerated to 5 eV to avoid fragmentation upon deposition. This high quality and purity of the samples is essential for reliable STM measurements.

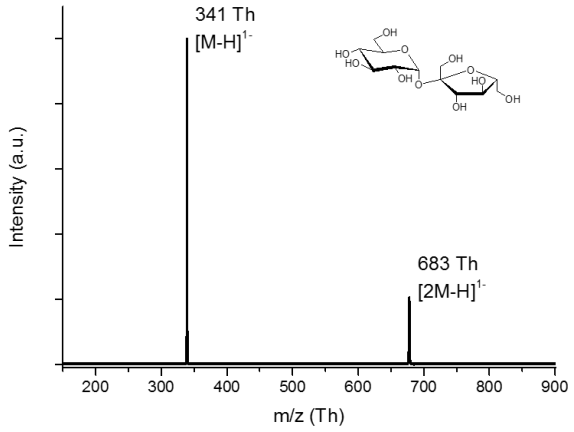


Figure S1. Mass spectrum of sucrose in the ion beam prior to mass selection and deposition.

Prior to deposition, the Cu(100) single crystal (Mateck GmbH)) is cleaned via several cycles of sputtering (1 kV, 30 min) and annealing (800 K, 10 min), and then transferred into the deposition chamber without breaking the vacuum. After deposition, the sample is transferred to the STM chamber into a variable temperature STM (Omicron VT-STM) where the measurements were conducted at 40 K to immobilize the adsorbents on the surface.

2. STM of sucrose

Due to the mobility of the molecule on the surface, only streaks are observed for sucrose on Cu(100) at room temperature. Upon cool-down to 40K, we observe agglomerations at the step edges as well as 2D islands, as shown in Figure S2. These islands consist of the periodic structure discussed in the paper. A few linear assemblies (marked with white rings) are also observed, exhibiting a second periodic pattern (not investigated in this study). STM parameters were -1.8 V and 50 pA current setpoint.

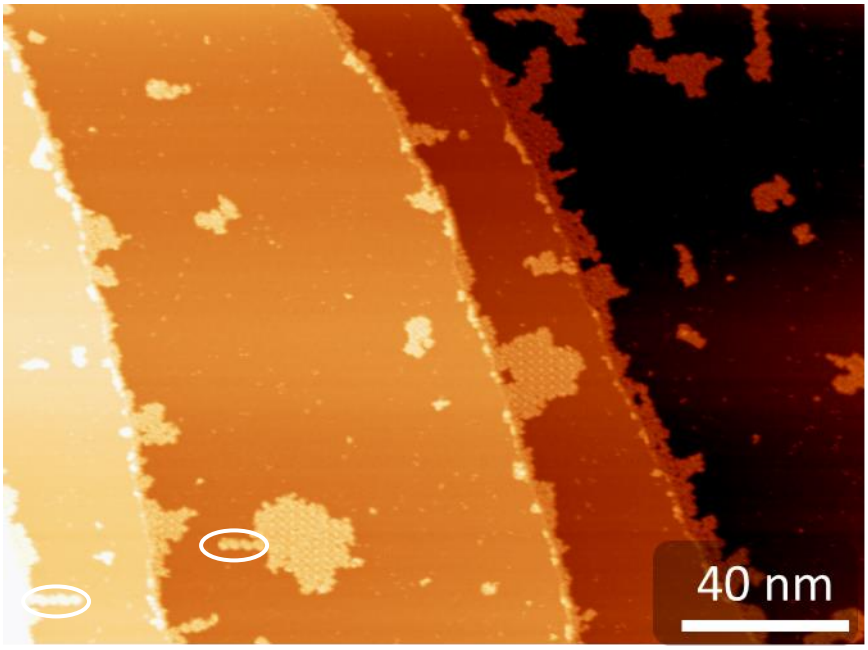


Figure S2: Overview image of sucrose on Cu(100) at 40 K.

SI-B: Theoretical methods

The prediction of probable, thermodynamically stable conformations of molecular systems can be formulated as a global optimization problem where the objective function corresponds to a model energy. Depending on the size of the system and the accuracy/complexity of the energy function, solving this optimization problem can be extremely challenging or even impossible using existing methods and available computational power.

To treat the problems addressed in this work, we applied a two-stage approach during the identification of optimal configurations, using different energy functions and different algorithms at each stage. In a first stage, a suitable global search algorithm was employed with the goal to find the most representative minima on the conformational energy landscape for a single molecule, or a whole monolayer, on the surface, using an iterative global exploration and local optimization (IGLOO) scheme (see below) or the basin hopping simulated annealing module in the G42+ code[2] respectively. For the sake of computational efficiency, a simple empirical energy function and a reduced moveclass were used at this stage (note: a moveclass defines the set of allowed variations of the system configuration during the landscape exploration). The set of minima identified by the global search method were then subject to an unconstrained local energy re-minimization using a more accurate energy function, based on Density Functional Theory (DFT). The computational complexity of DFT prevents its application at the global search level for systems of the size considered here.

Nevertheless, its use to locally refine representative conformations in different energy basins is essential to provide a more accurate picture at the atomic scale - both regarding the atomic configuration of the molecules and the charge density distribution -, and for a better understanding of the interactions between the molecule and the surface.

As a further approximation we treat the unchanged molecule ($C_{12}H_{21}O_{12}$) in the calculations instead of the deprotonated anion. To justify this assumption, we consider the properties of the anion. Firstly, in gas phase studies, it was shown that the deprotonation site migrates rapidly and therefore delocalizes the excess charge over the molecule.[10] Therefore we have no information regarding the position of the deprotonation site within the molecule on the surface. Furthermore, the presence of the metal surface will promote an effective screening or partial charge transfer and hence we are essentially dealing with a dehydrogenated neutral sugar molecule with the excess charge spread over the underlying metal slab. The surface mobility of the adsorbate, apparent in the propensity to form ordered structures, would not be in line with strong chemisorption. Evaluation by DFT of individual deprotonated sucrose molecules on a Cu-surface in vacuum confirmed the relatively easy movement of the proton-vacancy inside the molecule, but also showed that only minor changes in the overall conformations of the molecule on the surface are induced by a single deprotonation. Finally, we do not observe puckered non-standard ring-deformations in the minimum conformations of the dehydrogenated molecule on the surface. Thus, we conclude that the neutral molecule approximation is a reasonable choice for the purpose of modelling the conformations that help us in interpreting the observed network formation.

1. Global search for low energy configurations of a single molecule on a metal surface

1.1 The IGLOO algorithm

We have applied a new algorithm called IGLOO (Iterative Global exploration and LOcal Optimization) to globally explore the energy landscape of the molecule (+surface) system aiming to find a representative set of low-energy conformations. IGLOO belongs to a family of stochastic methods that perform global optimization by iterating local searches.[4] Well-known representative methods of this family are e.g. the Monte-Carlo minimization approach known as Basin-Hopping[5], or the Thermal Cycling method.[6] Based on our numerical experiments, IGLOO provides a more complete characterization of the energy landscape in comparison to other methods we have tested.

IGLOO iterates three main steps until a convergence criterion is satisfied or a maximum number of iterations is reached: 1) global sampling/exploration, 2) local minimization, 3) clustering. These three steps are sketched below:

1. Global sampling/exploration: At each iteration, IGLOO tries to cover the regions of the conformational space below a given energy threshold using multiple exploration trees. For the first iteration, the trees are rooted at n uniformly sampled random points being far from each other (the number of points, the minimum distance in configuration space between them, and the maximum energy are defined by the user). For subsequent iterations, the roots and the energy threshold are determined from the result of the previous iteration. The method used to construct the exploration trees is a variant of the Rapidly-exploring Random Trees (RRT) algorithm, originating from robotics,[7] and which has been extended to the exploration of energy landscapes of molecules [TRRT].[8]

2. Local minimization: Once the exploration phase is finished, the configurations corresponding to the nodes of the search trees are subject to local energy minimization. In order to reduce computing time, a representative subset of configurations is selected based on proximity filtering: in densely explored regions, only one configuration is selected within a small ball whose radius is determined from the density of points. In this work, we have used a randomized gradient descent approach for the local minimization.
3. Clustering: The minimized configurations are clustered aiming to identify the main basins on the energy landscape. For this, we apply a density-based approach.[9] The lowest-energy configuration in each cluster is a representative local minimum, and becomes one root for the next iteration of IGLOO. Convergence is assumed when the energy and the distance variations with respect to the previous iteration are below given threshold values for all the cluster representatives.

Before the "production" runs, IGLOO was run several times in order to identify good parameter settings for the algorithm, which ensured an optimal balance between computational cost and diversity in the generation of low-energy minima. The main parameters involved are related to the step size for the exploration (using the RRT algorithm), and the cut-off distances in configuration space that are used to select a subset of configurations to be minimized and are also employed during the clustering stage at the end of each iteration. These preliminary tests already showed the reliability of the algorithm, since similar results were obtained from all the test runs.

Once good parameter settings had been established, six production runs of IGLOO were performed. The first iteration of each run started from 6 randomly sampled distant configurations serving as roots for the search trees (the optimal number of these roots was established during the preliminary test runs). Each iteration of the algorithm (exploration + local minimization + clustering) produced on average around 3000 configurations, and about 10% of these were minimized. Note that as the algorithm proceeds, these configurations incrementally concentrate on the main energy basins of the energy landscape. Convergence was reached after 3 - 4 iterations. Thus, the total number of configurations generated from each run of IGLOO was around 10,000, of which ca. 1,000 were locally optimized.

All the runs of IGLOO identified the same three clustered sets of configurations that included the minima with the lowest-energies, indicating that the algorithm globally explores the low-energy regions of the landscape in a satisfactory fashion. The lowest energy minima of each of these three basins were selected for further refinement at the DFT level (shown in Fig. 2). The corresponding cartesian coordinates can be found in the SI. Note that in addition to these three minima basins, IGLOO produced several other minima with higher energy during each run, but these configurations were not considered for the subsequent analysis on ab initio level.

1.2 Force fields used during the global search

The models of the sucrose molecules in vacuum and on the surface were generated using the GLYCAM Molecular Modelling Library (GMML)[10] and AMBER[11] for the energy evaluation. The accuracy of the parameters and coefficients of this simple molecular-mechanics-based force-field were evaluated and tuned using DFT calculations. The interaction between the molecule and the metal surface consisted of Lennard-Jones type interactions (tuned via comparison with DFT calculations) plus electrostatic terms modelled via mirror charges.

We note that in the literature, Marianski et al.[12] investigate the accuracy of various methods (force fields, DFT, etc.) with respect to coupled cluster benchmarks at the example of a single isolated maltose molecule. They suggest that the GLYCAM force field employed in the current study has an expected mean absolute error of about 3 kcal/mol (ca. 13 kJ/mol) with a maximal error for computing conformational energy differences of about 15 kcal/mol (ca. 63 kJ/mol). Considering the fact that each molecule contains 45 atoms, this would correspond to a maximal error of about 0.014 ev/atom (or a temperature of about 150 K). Nevertheless, this study points towards the obvious limitations of the empirical potential employed during the global optimizations of the molecules in vacuum and on the surface. However, we have found that the local minimization on DFT level for individual molecules, both in vacuum and on the surface, change the minimum configurations we have identified only by very small amounts by slightly shifting the hydrogen atom positions. Of course, the details of the molecule-surface interaction change, too, but this does not lead to large deformations of the molecules. Thus, the potential catches quite well the general features of the conformations, and we expect that we are not going to miss major minimum basins of the energy landscape of the sucrose molecules.

Nevertheless, the precise ranking of the minima by energy may change when switching from empirical to DFT energies, just as the ranking might depend on whether we consider the same overall molecular conformation for a neutral or deprotonated molecule on the surface. Thus, we consider several of these single molecule conformation minima when generating the multi-molecular arrangements.

1.3 Moveclass employed during the global search

Only the most important degrees of freedom were considered as variables for the exploration of the global energy landscape of the single molecule on the substrate: the six variables defining the pose (position and orientation) of the molecule with respect to the surface, and the 13 bond torsions of the sucrose molecule. Since the conformation of the glucose and fructose rings is known to be relatively stable at room temperature and below,[13] the rings were considered as rigid bodies at this stage. For the generic starting choice regarding the conformation of the sucrose molecule, results from a global optimization study of the shape of the molecule in vacuum were employed; [14] we chose the chair conformation for the six-ring moiety and a planar configuration for the fructose moiety, because it is the energetically favorable form. We used a perfect model of the Cu(100) surface involving three layers of Cu atoms for which no mobility was allowed.

2. Local refinement: DFT minimization

Our post-IGLOO DFT minimizations were performed using the Vienna ab initio simulation package (VASP) with PAW pseudopotentials, under periodic boundary conditions.[15] The simulation box size was optimized in order to avoid any spurious interaction between the sucrose molecule and its periodic images (x and y directions: 5 × 5 unit cells) and the subsurface (at least 18 Å of vacuum). Four copper layers were taken into account (i.e. 200 Cu atoms). The lowest layer was kept fixed and all other atoms (three Cu layers + molecule) were allowed to relax without constraints. A conjugate-gradient algorithm (CGA) was used to relax the ions. Since the size of the supercell was large enough, the Brillouin zone sampling in reciprocal space restricted to the Γ point was sufficient to ensure good convergence of the total energy. A plane-wave kinetic energy cut-off of 500 eV was employed and the

convergence criterion was set up so that the maximum atomic force was less than 0.01 eV Å−1. For dealing with the partial occupancies around the Fermi level, a Methfessel–Paxton smearing was used with σ = 0.01 eV.[16]

Although GGA functionals provide acceptable performance for describing both structure and energies of solids,[17] their local character implies the neglect of London dispersion interactions which should be properly accounted for when studying adsorption of molecules on metallic systems.[18] In the framework of DFT, many schemes are available to treat dispersion. Among them, the opt-B86b functional, developed by Dion et al,[19] was chosen because of its ability to describe structural properties of adsorbates with high accuracy.[20] Although this functional slightly overestimates the adsorption energies, the effect of this overestimation is cancelled when comparing different adsorption conformations of a given sucrose molecule. In this functional, the van der Waals part is directly obtained from the electron density by adding a non-local term to the local correlation functional.

3. Modeling of sucrose assemblies on Cu(100).

Due to the complexity of the systems, the extended system size, and the difficulty to choose a definitive assembly from among the multitude of structurally and energetically very similar local minima found during preliminary empirical potential global optimizations of multi-molecule systems (consisting of four sucrose molecules in a periodic simulation cell of variable size, each exhibiting the optimal conformations found for the single molecule on a copper surface) performed with the G42+ code,[21] we decided to construct models of the sucrose assembly by hand (c.f. SI-C). To this end, we employed the lowest-energy conformations of the single sucrose molecule on Cu(100) that resulted from IGLOO exploration and DFT relaxation, and imposed constraints from the STM data to the model, such as orientation, dimensions and symmetry. Specifically, copies of the single molecule were placed on the surface in arrangements that optimally agree with the STM image, while still keeping the same orientation with respect to the surface. Thus, the position of each molecule on the surface is the one found by the DFT minimization. This procedure was performed for the three lowest-energy sucrose conformations. As the energy differences between the different conformations are very small, an additional estimate of the stability and energy of the assembly was obtained by the number of possible hydrogen bonds. The simulated STM images were generated using the HIVE code.[22]

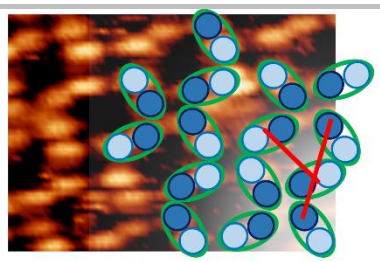
SI-C: Evaluation of the assembly patterns obtained using three different minimum-energy conformations of the sucrose molecule

As described above (section 3 in SI-B), three DFT optimized conformations of sucrose are employed to manually form models that resemble the pattern in the STM image. It is important to note that the orientation of the molecule on the surface is kept in its optimized state, so that only constrained translations by the Cu-unit cell parameters as well as rotations of 90° and 180 ° are possible. In the STM image, we observe two different nodes, one with bright features and one containing only the lower intensity feature. Therefore, the model is built containing nodes of either fructose or glucose units. These nodes exhibit a rotational direction (chirality) due to the offset of the molecules, which is important to replicate in the model. Naturally, in addition to these constraints, the distances between the molecules also have to conform to the experimental data.

SUPPORTING INFORMATION

The models were built in such a way that one node was formed by translation and rotation of the molecule. Subsequently this node was copied and translated to form the other node. By this procedure, the position of the molecules always stays the same. Fig. S3 shows a multiple possibilities for the assembly pattern. For s-min_1^{DFT}, the two patterns that are close to the reference distances of the measured assemblies are rejected due to the strong asymmetry of the pores. This is also the reason why one of the patterns of s-min_2^{DFT} is rejected (last row, panel two in Fig. S3). The other two assembly patterns for s-min_2^{DFT} do not posses chiral nodes, which are clearly visible in the STM image, and both are therefore discarded, too. For s-min_3^{DFT}, we found 3 possible assembly patterns. However, the first one also does not exhibit chiral nodes, while the second one does not show any node for fructose and therefore does not resemble the structure measured in STM. In contrast, the third structure of s-min_3^{DFT} fits very well to the experimentally observed structure. The distances are very similar to the distances in the STM image, while the pores are nearly quadratic and the nodes are chiral. As this is the only assembly pattern that is in agreement with the constraints, we discuss this pattern in the main article and assume that s-min_3^{DFT} is the conformation of sucrose on Cu(100).

We note that despite the important role of the H-bonds in the intermolecular interactions, they do not significantly alter the conformation of the individual saccharide molecules. This is supported by DFT simulations we performed for a rather close-packed periodic arrangement of four sucrose molecules per cell. One of these molecules was rotated by 360° about the z-axis (orthogonal to the surface) in a step-wise fashion. After each step (30°), the whole structure was locally optimized, but no large deformations of the sucrose molecules were observed (although the energy varied noticeably as function of rotation angle, indicating the presence of energy barriers that inhibit the rotation of the sucrose molecules when packed closely).



Glu-Glu: 1.0 nm
Fruct-Fruct: 1.1 nm
Between junctions: 1.7 nm
Antiparallel distance: 1.2 nm

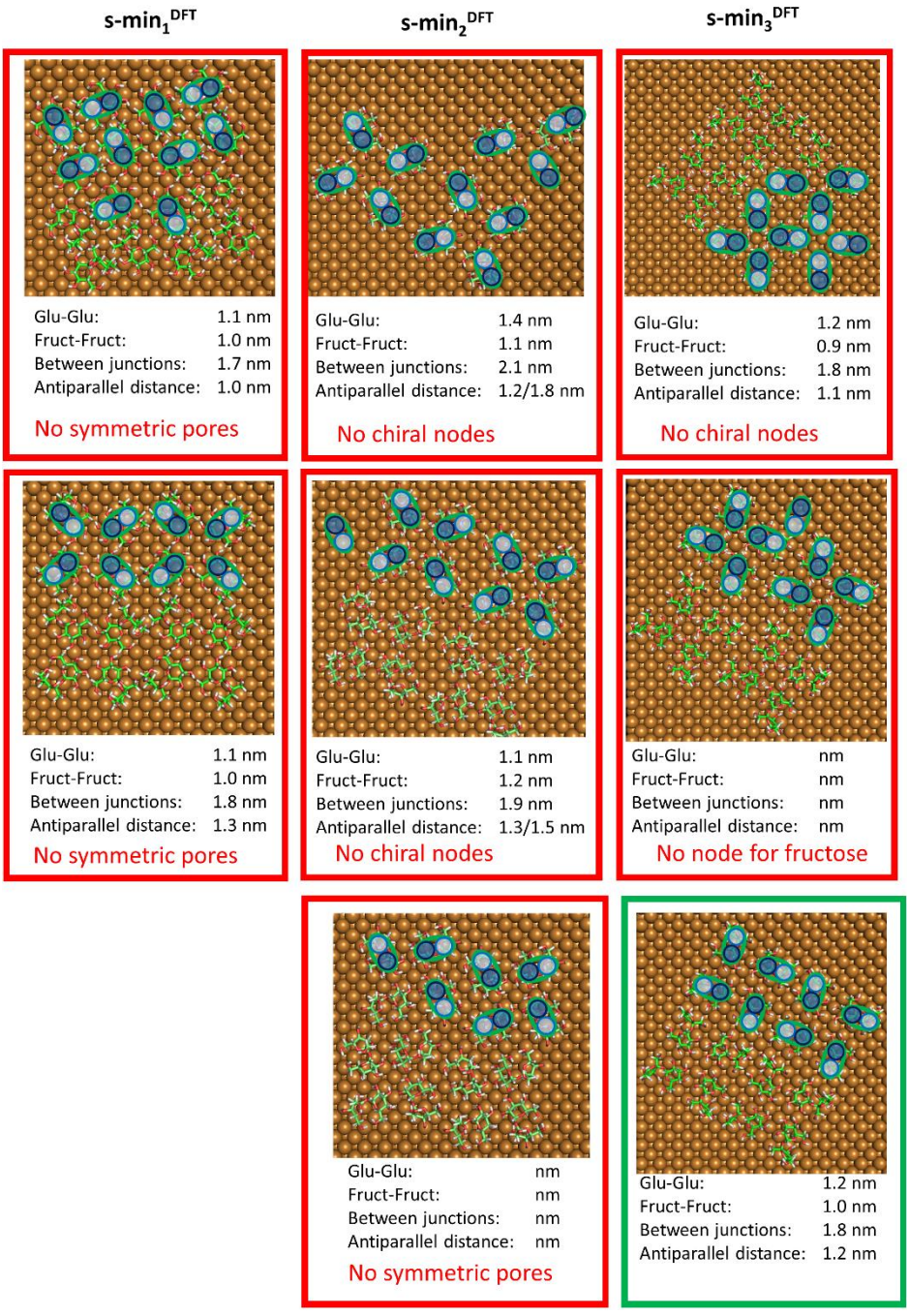


Figure S3: Models of the assembly of sucrose on Cu(100) for the three different lowest energy structures. Only the assembly with the green frame (last row, s-min$_3$) fits to the constraints of the experiment.

SUPPORTING INFORMATION

References

[1] a) S. Rauschenbach, F. L. Stadler, E. Lunedei, N. Malinowski, S. Koltsov, G. Costantini, K. Kern, *Small* **2006**, *2*, 540-547; b) S. Rauschenbach, M. Ternes, L. Harnau, K. Kern, *Annu. Rev. Anal. Chem.* **2016**, *9*, 473-498.

[2] J. C. Schön, *Processing and Application of Ceramics* **2015**, *9*, 157-168.

[3] W. B. Struwe, C. Baldauf, J. Hofmann, P. M. Rudd, K. Pagel, *Chem. Commun.* **2016**, *52*, 12353-12356.

[4] M. Locatelli, F. Schoen, *4OR* **2013**, *11*, 301-321.

[5] a) Z. Li, H. A. Scheraga, *Proc. Natl. Ac. Sci.* **1987**, *84*, 6611-6615; b) M. C. Prentiss, D. J. Wales, P. G. Wolynes, *J. Chem. Phys.* **2008**, *128*, 225106.

[6] A. Möbius, A. Neklioudov, A. Díaz-Sánchez, K. H. Hoffmann, A. Fachat, M. Schreiber, *Phys. Rev. Lett.* **1997**, *79*, 4297-4301.

[7] S. M. LaValle, *Planning Algorithms*, Cambridge University Press, **2006**.

[8] a) L. Jaillet, F. J. Corcho, J.-J. Pérez, J. Cortés, *J. Comput. Chem.* **2011**, *32*, 3464-3474; b) D. Devaurs, K. Molloy, M. Vaisset, A. Shehu, T. Siméon, J. Cortés*, *IEEE Trans. NanoBiosci.* **2015**, *14*, 545-552.

[9] a) F. Chazal, L. J. Guibas, S. Y. Oudot, P. Skraba, *J. ACM* **2013**, *60*, 1-38; b) A. Rodriguez, A. Laio, *Science* **2014**, *344*, 1492-1496.

[10] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622-655.

[11] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *WIREs Comput. Mol. Sci.* **2013**, *3*, 198-210.

[12] M. Marianski, A. Supady, T. Ingram, M. Schneider, C. Baldauf, *J. Chem. Theory Comput.* **2016**, *12*, 6157-6168.

[13] a) Z. Kan, X. Yan, J. Ma, *J. Phys. Chem. A* **2015**, *119*, 1573-1589; b) J. Xia, D. A. Case, *Biopolymers* **2012**, *97*, 276-288.

[14] S. Neelamraju, R. L. Johnston, J. C. Schön, *J. Chem. Theory Comput.* **2016**, *12*, 2471-2479.

[15] a) G. Kresse, J. Hafner, *Physical Review B* **1993**, *47*, 558-561; b) G. Kresse, J. Hafner, *Physical Review B* **1994**, *49*, 14251-14269; c) G. Kresse, J. Furthmüller, *Computational Materials Science* **1996**, *6*, 15-50; d) G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169-11186; e) P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953-17979; f) G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758-1775.

[16] M. Methfessel, A. T. Paxton, *Phys. Rev. B* **1989**, *40*, 3616-3621.

[17] a) J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865-3868; b) J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1997**, *78*, 1396-1396; c) A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098-3100.

[18] A. Tkatchenko, *Adv. Funct. Mater.* **2015**, *25*, 2054-2061.

[19] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **2004**, *92*, 246401.

[20] J. Björk, S. Stafström, *ChemPhysChem* **2014**, *15*, 2851-2858.

[21] J. C. Schön, C. Oligschleger, J. Cortes, in *Z. Naturforsch. B, Vol. 71*, **2016**, p. 351.

[22] D. E. P. Vanpoucke, G. Brocks, *Phys. Rev. B* **2008**, *77*, 241308.