

# Reliable Computational Prediction of the Supramolecular Ordering of Complex Molecules under Electrochemical Conditions

Benedikt Hartl,\* Shubham Sharma, Oliver Brügger, Stijn F. L. Mertens, Michael Walter, and Gerhard Kahl

Cite This: *J. Chem. Theory Comput.* 2020, 16, 5227–5243

Read Online

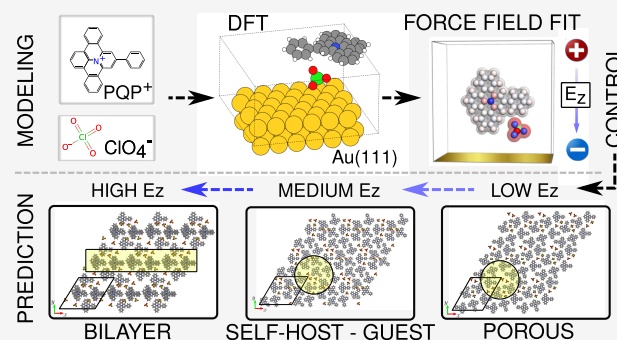
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** We propose a computationally lean, two-stage approach that reliably predicts self-assembly behavior of complex charged molecules on metallic surfaces under electrochemical conditions. Stage one uses ab initio simulations to provide reference data for the energies (evaluated for archetypical configurations) to fit the parameters of a conceptually much simpler and computationally less expensive force field of the molecules: classical, spherical particles, representing the respective atomic entities; a flat and perfectly conducting wall represents the metallic surface. Stage two feeds the energies that emerge from this force field into highly efficient and reliable optimization techniques to identify via energy minimization the ordered ground-state configurations of the molecules. We demonstrate the power of our approach by successfully reproducing, on a semiquantitative level, the intricate supramolecular ordering observed experimentally for PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> molecules at an Au(111)–electrolyte interface, including the formation of open-porous, self-host–guest, and stratified bilayer phases as a function of the electric field at the solid–liquid interface. We also discuss the role of the perchlorate ions in the self-assembly process, whose positions could not be identified in the related experimental investigations.



## 1. INTRODUCTION

Supramolecular chemistry deals with intermolecular interactions and structure formation beyond individual molecules and as such lies at the base of many nano- and mesoscopic structures found in biology. In recent decades, impressive progress in the experimental branches of this field has resulted in at least two Nobel Prizes in chemistry. By contrast, the theoretical understanding and especially the in silico prediction of supramolecular ordering has lagged behind somewhat. This is easily understood if one considers the sheer size of the systems under study, requiring in many cases consideration of a solid substrate, a sufficiently large number of molecular building blocks or tectons, and a condensed matter medium (i.e., a solvent or electrolyte solution). The interaction of these three components, each with its intrinsic properties and with optional extrinsic steering (e.g., by light, heat, electric field), will determine the observed supramolecular structures and govern the transitions between them.<sup>1,2</sup>

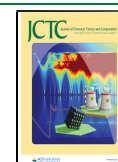
In this paper, we propose a new theoretical framework to predict the supramolecular ordering of complex molecules at an electrochemical solid–liquid interface. The calculations were inspired by recent experimental work<sup>3</sup> in which particularly clear-cut transitions between supramolecular structures were observed as a function of the applied electric field at a metal–electrolyte interface. The target molecules

whose supramolecular ordering is considered constitute an organic salt that consists of a large, disc-shaped polyaromatic cation (PQP<sup>+</sup>) and a much smaller, inorganic anion (perchlorate, ClO<sub>4</sub><sup>-</sup>).<sup>4–6</sup>

The concept of choice to investigate these scenarios would rely on (i) a faithful description of the properties of the system (notably a reliable evaluation of its energy) via ab initio simulations and (ii) in a subsequent step the identification of the optimized (ordered) arrangement of the molecules on the substrate by minimizing this energy via efficient and reliable numerical tools; this optimization has to be performed in a high-dimensional search space, spanning all possible geometries of the unit cell and all possible coordinates and orientations of the molecules within that cell. Both these approaches, considered separately from each other, are conceptually highly complex and from the numerical point of view very expensive, which precludes the application of this combined concept even for a single set of external parameters

Received: December 19, 2019

Published: June 15, 2020



(such as temperature, density, and external field); it is thus obvious that systematic investigations of the self-assembly scenarios of such systems are definitely out of reach.

In this contribution, we propose an approach to overcome these limitations via the following strategy: in the first step, we map the ab initio-based energies onto the energy of a related classical model (or a classical force field), where the atomistic units of the molecules are featured as spherical, charged units with Lennard-Jones-type interactions and where the electrolyte is treated as a homogeneous, dielectric medium; the interaction between the atomic entities and the metallic surface is modeled by a classical, perfectly conductive, Lennard-Jones-like wall potential. The as-yet open parameters of the resulting force field (energy- and length-scales, charges, etc.) are fixed by matching the ab initio energies of the system with the related energies of this force field: this is achieved by considering archetypical configurations of the system's building blocks (molecules and surface) and by systematically varying characteristic distances between these units over a representative range. These ab initio energies were then fitted along these "trajectories" by the parameters of the classical force field: the energy- and length-scales of the involved interatomic Lennard-Jones or Mie potentials as well as the atom-wall interaction parameters.

It turns out that this force field is indeed able to reproduce the ab initio-based energies along these "trajectories" faithfully and with high accuracy. Even though the emerging force field is still quite complex (as it features both short-range and long-range Coulomb interactions and involves mirror charges), it is now amenable to the aforementioned optimization techniques, which thus brings systematic investigations of the self-assembly scenarios of these molecules under the variations of external parameters within reach.

As a benchmark test for our approach, we have considered the above-mentioned system, studied in recent experimental investigations: the cation is  $\text{PQP}^+$  (9-phenylbenzo[1,2]quinolizino[3,4,5,6-*fed*]phenanthridinium, a disk-shaped polycyclic aromatic molecule), while the anion is perchlorate,  $\text{ClO}_4^-$ ; the self-assembly of these ions on a Au(111) surface under the influence of an external electric field was studied. The high accuracy with which the ensuing energies calculated from the force field reproduce the ab initio simulation data make us confident about the applicability of the force field for the subsequent optimization step: using an optimization technique that is based on ideas of evolutionary algorithms, we have then identified the self-assembly scenarios of the ions on the Au surface, for a given set of external parameters (temperature, density, and external field). These first results provide evidence that our approach is quite promising. This concept is furthermore completely flexible as it can easily be extended to other organic molecules of similar (or even higher) complexity. The computational cost of this optimization step is still substantial. Therefore, we postpone a detailed, quantitative, and, in particular, systematic investigation of the self-assembly scenarios of the  $\text{PQP}^+$  and the  $\text{ClO}_4^-$  ions on the Au surface for a broad range of external parameters to a later contribution. Instead, we demonstrate in this contribution for selected sets of parameters that our approach is indeed able to reproduce several of the experimentally identified self-assembly scenarios.

In this context, it has to be emphasized that such a type of optimization problem is highly nontrivial since the huge number of possible local minima in the potential energy

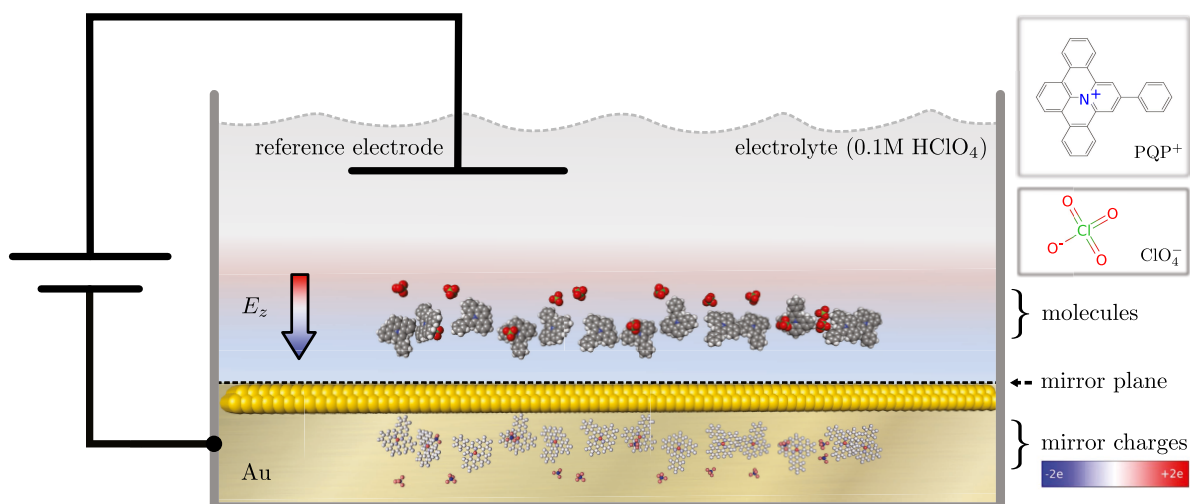
surface (embedded in a high-dimensional parameter space) increases exponentially with the number of particles (and their degrees of freedom) of the system;<sup>7</sup> thus, exhaustive search strategies hit the computational limits or even exceed the capacities of present-day supercomputers. Yet, another complication in structure prediction algorithms is caused by the fact that different polymorphs of a system can be kinetically trapped and a vast number of other minima, having values of the internal energy comparable to the global minimum, may also play an important role in structure formation processes.<sup>7,8</sup>

At this point, we owe an explanation to the reader why we have chosen the possibly unconventional approach. Of course, it is obvious that an optimization of the molecular configurations on the basis of full ab initio calculations is from the computational point of view by far out of reach. However, one can argue that suitable force fields (available in the literature) or machine-learning (ML) potentials such as high-dimensional neural network potentials,<sup>9–15</sup> kernel-based ML methods<sup>16</sup> (such as Gaussian approximation potentials<sup>17–21</sup>), or more specialized, effective potentials for selected molecular motives such as the SAMPLE approach<sup>8,22,23</sup> might represent a more conventional approach to tackle this problem (note that the field of ML potentials is rapidly growing and the above list is far from comprehensive). Such arguments represent fully legitimate objections against our approach.

The problem we are addressing in this contribution is, however, a nonstandard problem and thus requires to be treated with a custom force field: the justification for our strategy is that we wanted to endow the atomic units of the molecules with "real" physical properties (such as "size" or "charge"), which will help us to perform the second step in our structural search that we have envisaged (and that we are currently working on): as the computational costs of our approach are still considerable, large-scale investigations are still prohibitively expensive. In an effort to overcome these limitations, we plan to proceed to even more coarse-grained models, which grasp, nevertheless, the essential features of our complex molecules. On the basis of such models, we would then be able to identify with rather low computational costs first trends in structural identification processes. Investigations along this direction will be explored in forthcoming contributions.

Finally, we point out that we are well aware of the limitations and deficiencies of our present model. Features such as the response of the metallic electronic distribution of the gold surface due to the presence of an external bias, variable electrostatic properties of the molecular species (allowing thus for polarization effects), or a space-dependent permittivity cannot be included in our concept. However, at this point, it is fair to say that, to the best of our knowledge, none of the aforementioned alternative approaches (such as the use of conventional force fields or machine-learning frameworks) is able to take all of these effects faithfully into account, either.

The manuscript is organized as follows: In Section 2, we describe the essential features of the experimental setup, introduce an ab initio and a classical representation thereof, and discuss the mapping procedure between those different instances. In Section 3, we put forward the memetic optimization procedure based on ideas of evolutionary algorithms to identify ordered ground-state configurations of complex molecules under electrochemical conditions, and in Section 4, we present selected numerical results, which demonstrate a semiquantitative agreement with the exper-



**Figure 1.** Schematic visualization of the experimental setup to control the pattern formation of PQP<sup>+</sup> (and ClO<sub>4</sub><sup>-</sup>) molecules (structure formulas given in top-right insets) close to a Au(111) surface: two Au layers are explicitly shown; the golden, shiny area represents the conductive Au bulk, and the black dashed line marks the surface of the electronic density, which we interpret as a mirror plane. The ions are immersed into an electrolyte (gray, shaded region), which is considered as an effective, homogeneous medium. In the region close to the Au surface (red to blue shaded areas), a homogeneous, electrostatic field  $E_z$  (bold, colored arrow), oriented in the  $z$ -direction, features the electrostatic potential drop between the Au surface and the reference electrode inside the electrolyte. The colors of the atoms in the electrolyte correspond to their type, while the color of the mirror atoms (located in the Au bulk) specifies their partial charges, quantified by the color bar (see bottom right) in units of the electron charge,  $e$ .

imentally observed self-assembly scenarios of PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> ions on a Au(111)–electrolyte interface under the influence of an external electrostatic field. We conclude our findings in Section 5.

## 2. SYSTEM AND ITS REPRESENTATIONS

**2.1. System.** Both the density functional theory (DFT) calculations and the related force field are based on a framework that mimics the essential features of the experimental setup, put forward (and discussed) in ref 3; this framework is schematically depicted in Figure 1: PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> ions are immersed in an electrolyte (0.1 M aqueous perchloric acid). From below, the system is confined by a Au(111) surface, which in the experiment serves as the solid substrate for adsorption. An electric field,  $E_z$ , can be applied between a reference electrode located within the electrolyte and the Au surface. The PQP<sup>+</sup> and the ClO<sub>4</sub><sup>-</sup> ions are first treated via DFT-based ab initio calculations (see Section 2.2). The calculated energies are then used to fix the force fields of classical particles (notably their sizes, energy parameters, and charges), which represent the atomic entities of the respective ions; the interaction between the atomic entities and the Au(111) substrate is described by means of a classical wall-particle force field (see Section 2.3). Throughout, the electrolyte molecules are not considered explicitly. The electrolyte is rather assumed to be an effective homogeneous medium with a permittivity of water, i.e.,  $\epsilon_r = 78.36$ , at  $T = 25$  °C,<sup>24–26</sup> corresponding to the temperature at which the experiments by Cui et al.<sup>3</sup> were carried out and assuming that the low concentration of perchloric acid does not change the value of  $\epsilon_r$  substantially.<sup>27–30</sup> Hence, in this contribution, we use “electrolyte” as a synonym for a “solvent” unless explicit use is required.

We emphasize at this point that in the experiment, an exact specification of the electric field strength is not possible: as detailed in the supporting information of ref 3, the authors of

the related experimental investigations have estimated rather the degree of charge compensation on the Au surface by the adsorbed PQP<sup>+</sup> ions as a function of their changing coverage, which does not allow the estimation of the electric field directly. This fact limits the degree of quantitative comparison between experiment and theory.

**2.2. Ab Initio Simulations.** The density functional theory calculations were performed with the software package GPAW,<sup>31,32</sup> and the structures were handled by the atomic simulation environment.<sup>33</sup> The electronic density and the Kohn–Sham orbitals were represented within the projector-augmented wave method,<sup>34</sup> where the smooth parts were represented on real space grids with a grid spacing of 0.2 Å for the orbitals and 0.1 Å for the electron density. The exchange–correlation energy is approximated as proposed by Perdew, Burke, and Ernzerhof (PBE),<sup>35</sup> and weak interactions missing in the PBE functional are described as proposed by Tkatchenko and Scheffler (TS09).<sup>36</sup> The TS09 approximation assumes that long-range dispersive contributions are absent in the PBE functional, such that these can be applied as a correction. The total energy is written as

$$E = E_{\text{PBE}} + w_s E_{\text{vdW}} \quad (1)$$

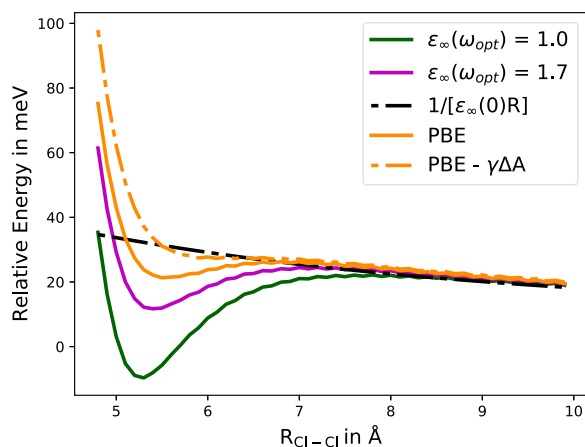
where  $E_{\text{PBE}}$  is the PBE energy and  $E_{\text{vdW}}$  is the TS09 correction. We have introduced a weight factor  $w_s$  that will allow the incorporation of electrolyte effects into the dispersive contributions as discussed below. For interactions in vacuum,  $w_s = 1$ . The presence of the aqueous environment on the electronic and nuclear degrees of freedom included in  $E_{\text{PBE}}$  is modeled by a continuum solvent model.<sup>37</sup>

Molecular interactions were studied on simulation grids with Dirichlet (zero) boundary conditions. Neumann (periodic) boundary conditions were applied in  $x$ - and  $y$ -directions in the surface plane for simulations involving the gold surface, while zero boundary conditions were applied in the perpendicular  $z$ -direction. The simulation grid was chosen such that at least 4 Å of space around the position of each atom in the nonperiodic

directions was ensured. The Au(111) gold substrate was modeled by two layers of 54 atoms, each using the experimental lattice constant of face-centered cubic (fcc) gold of  $a = 4.08 \text{ \AA}$ . These settings result in a rectangular unit cell of  $26.0 \times 15.0 \text{ \AA}^2$ . The Brillouin zone was sampled by  $3 \times 3$  Monkhorst–Pack<sup>38</sup> distributed  $k$ -points in the periodic directions.

Potentials were scanned by fixing all gold atoms and a central atom of  $\text{PQP}^+$  (the nitrogen atom) and/or of  $\text{ClO}_4^-$  (the chlorine atom) to given positions, while all other atoms were allowed to relax without any symmetry constraints until all forces were below  $0.05 \text{ eV/\AA}$ .

The interaction of two perchlorate anions in dependence of their distance is shown in Figure 2 for different approximations



**Figure 2.** Relative energy of two  $\text{ClO}_4^-$  anions as a function of the distance between their chlorine atoms,  $R_{\text{Cl-Cl}}$ , where the separated anions define the energy reference;  $\epsilon(\omega_{\text{opt}}) = 1$  with full van der Waals (vdW) corrections and  $\epsilon(\omega_{\text{opt}}) = 1.7$  with scaled vdW corrections. The dash-dotted line shows the PBE energy, where the energy contribution of effective surface tension  $\gamma\Delta A$  is subtracted (see the text).

for the total energy in eq 1. As expected, the potentials follow the screened electrostatic repulsion  $[\epsilon_{\infty}(0)R_{\text{Cl-Cl}}]^{-1}$  for large distances  $R_{\text{Cl-Cl}}$ , where  $\epsilon_{\infty}(0) = \epsilon_r$  is the static relative permittivity of water. There is a slight attractive part in the potential around  $R_{\text{Cl-Cl}} \simeq 5.2 \text{ \AA}$  already in the PBE potential, which leads to a very shallow local minimum. The main reason for this minimum is the decrease in the effective surface  $\Delta A$  when the solute cavities (the solvent-excluded regions) begin to overlap. This decreases the energetic cost to form the surface due to the effective surface tension  $\gamma = 18.4 \text{ dyn cm}^{-1}$  ( $\gamma$  also contains attractive contributions and is therefore much lower than the experimental surface tension of water<sup>37</sup>). Subtracting  $\gamma\Delta A$  nearly removes all of the minima as demonstrated in Figure 2. By including the full dispersion contribution [ $\epsilon_{\infty}(\omega_{\text{opt}}) = 1$ ], this local minimum substantially deepens and becomes the total minimum of the potential. An attractive contribution to the potential is not to be expected for the interaction of two anions and needs further discussion. We suspect an overestimation of dispersion interactions if these are treated as in vacuum and no screening through the electrolyte is considered.

The aqueous environment influences the van der Waals (vdW) interactions as these are of Coulombic origin.<sup>39</sup> To derive an approximate expression for the screened vdW interaction of two molecules A and B at distance  $R$  inside

the electrolyte, we express the  $C_6$  coefficient defining the vdW energy  $C_6/R^6$  by the Casimir–Polder integral<sup>36,40</sup>

$$C_6 = \frac{3}{8\pi^2\epsilon_0} \int_0^{\infty} \alpha_A^*(i\xi)\alpha_B^*(i\xi)\phi(i\xi) d\xi \quad (2)$$

where  $\alpha_{A,B}^*$  is the polarizability of the interacting molecules and  $\phi$  is determined by the propagation of the electric field through the embedding medium<sup>40</sup> with  $\phi = 1$  in vacuum. Both  $\alpha_{A,B}^*$  and  $\phi$  are modified relative to vacuum in solution. In the simplest model,<sup>40</sup> we may write  $\phi(i\xi) = \epsilon_{\infty}^{-2}(i\xi)$  with the frequency-dependent relative permittivity of the electrolyte  $\epsilon_{\infty}$ . The effects of the electrolyte on the polarizabilities  $\alpha_{A,B}^*$  should, at least partly, already be included in the TS09 description through the effective atomic polarizabilities derived from the self-consistent electron density calculated within the electrolyte. What is left is the effect of the permittivity entering through the function  $\phi(i\xi)$  in eq 2. We assume that the main contribution of  $\phi(i\xi)$  is at the resonance frequencies of  $\alpha_{A,B}^*$ , which are in the optical region for usual molecules. We further assume that  $\epsilon_{\infty}(\omega_{\text{opt}})$  is approximately constant in this frequency region such that we may pull  $\phi = [\epsilon_{\infty}(\omega_{\text{opt}})]^{-2}$  out of the integral. This factor scales the  $C_6$  coefficient and therefore the vdW contribution. In other words, we apply the weight  $w_S = [\epsilon_{\infty}(\omega_{\text{opt}})]^{-2}$  in eq 2 with the experimental permittivity of water in the optical region of  $\epsilon_{\infty}(\omega_{\text{opt}}) = 1.7$ ; see ref 41. This approach reduces considerably the depth of the suspiciously deep minimum as seen in Figure 2 such that only a shallow local minimum remains similar to the PBE potential. The reduction obtained is quite strong in respect of the small contributions of Axilrod–Teller–Muto interactions commonly assumed.<sup>42,43</sup> The quantitative connection between the screening of dispersive interactions in polarizable media and the many-body effects neglected in TS09<sup>44</sup> is not immediately clear and is certainly worth further investigation. In what follows, we use the same scaling for all of the vdW contributions of the DFT potentials in this work.

**2.3. Force Field Model.** In this section, we describe how we cast our setup into force fields where the atoms in the molecular constituents are described as spherical particles, each of them carrying a charge. The mapping is guided by the energies obtained via the ab initio simulations detailed above. The Au(111) surface is modeled as a flat and perfectly conducting surface involving mirror charges as detailed below. However, we note that the position of the corresponding surface in the DFT calculations does not coincide with the position of the atoms. Before proceeding, the following comment is in order: in this mapping procedure, the distance of a point charge to a metallic surface is unambiguously defined through the electrons leaking out of the potential defined by the nuclei.<sup>45,46</sup> This feature can explicitly be seen in jellium models<sup>47</sup> but also emerges in implicit calculations<sup>48</sup> where electrons spill out of the surface of metal clusters.<sup>49</sup> From the latter study, we estimate an effective spill out of the surface of  $0.5 \text{ \AA}$ , a value that agrees qualitatively with estimates from the jellium models, extrapolated to large structures.<sup>46</sup> This value will be used below for our problem.

**2.3.1. Atomistic Model.** In our atomistic model, the molecules are represented as rigid entities composed of atomistic constituents. The molecules are immersed in a microscopic electrolyte, which is treated as a continuous medium of given permittivity. From below, the system is confined by a conducting Au(111) surface (which is assumed to extend in the  $x$ - and  $y$ -directions), and an external field

(with respect to the electrolyte) can be applied in the  $z$ -direction, i.e., perpendicular to the surface (or wall). Figure 1 schematically depicts all details of this atomistic model for the  $\text{PQP}^+\text{ClO}_4^-$  system, confined by the Au surface.

To specify the different entities of the system and their force fields, we use the following notation:

- (i) Each of a total number of  $N$  molecules is uniquely labeled by capital Latin indices  $I$ : for each of these units, this index is assigned to its center-of-mass (COM) position vector,  $\mathbf{R}_I$ ; to a vector  $\mathbf{P}_I$ , specifying its orientation within the lab frame in terms of the angle-axis framework<sup>50,51</sup> (see Supporting Information (S.I.), Section S2.2 for more details), and to the set of coordinates,  $\mathbf{r}^{N_I}$ , of the respective  $N_I$  atomistic constituents of the molecule in its COM frame (to which we also refer as its blueprint). The set of COM positions and orientation vectors of all  $N$  molecules are denoted  $\mathbf{R}^N$  and  $\mathbf{P}^N$ , respectively. The set of all  $n = \sum_{I=1}^N N_I$  atom positions in the lab frame is given by  $\mathbf{r}^n$ , and the position of each atom in the lab frame is uniquely defined by a vector  $\mathbf{r}_i$ , labeled with Latin indices ( $i = 1, \dots, n$ ).
- (ii) Between all atoms, we consider long-range Coulombic interactions (index “C”)

$$U^{(C)}(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_i q_j}{r_{ij}}, \quad i \neq j \quad (3)$$

with the interatomic distance  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  and charges  $q_i$  and  $q_j$  of the units  $i$  and  $j$ , respectively; the dielectric constant  $\epsilon_0$  and the relative permittivity  $\epsilon_r$  specify the implicit electrolyte. Further, we introduce short-range force fields (index “sr”) for which we have considered two options: first, a Lennard-Jones potential (index “LJ”), i.e.

$$U^{(LJ)}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (4)$$

for the energy and length parameters,  $\epsilon_{ij}$  and  $\sigma_{ij}$ , we have opted for the standard Lorentz–Berthelot mixing rules,<sup>52</sup> i.e.,  $\sigma_{ij} = (1/2)(\sigma_i + \sigma_j)$  and  $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ , respectively. Alternatively, we have also considered for the short-range interactions the Mie potential<sup>53</sup> (index “Mie”), which can be considered as a generalization of the LJ interaction; its functional form is given by

$$U^{(Mie)}(r_{ij}) = C_{ij} \epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{\gamma_{ij}^{(R)}} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{\gamma_{ij}^{(A)}} \right] \quad (5)$$

and allows for a variation of the exponents of the repulsive and attractive contributions to the potential,  $\gamma_{ij}^{(R)}$  and  $\gamma_{ij}^{(A)}$ , respectively.  $\epsilon_{ij}$  and  $\sigma_{ij}$  are again parameters for the energy- and the length-scales. The  $C_{ij}$  is defined as a function of the exponents<sup>53</sup>

$$C_{ij} = \left( \frac{\gamma_{ij}^{(R)}}{\gamma_{ij}^{(R)} - \gamma_{ij}^{(A)}} \right) \left( \frac{\gamma_{ij}^{(R)}}{\gamma_{ij}^{(A)}} \right)^{\frac{\gamma_{ij}^{(A)}}{\gamma_{ij}^{(R)} - \gamma_{ij}^{(A)}}} \quad (6)$$

For the exponents, we apply arithmetic mixing laws, i.e.,  $\gamma_{ij}^{(R)} = (1/2)(\gamma_i^{(R)} + \gamma_j^{(R)})$  and  $\gamma_{ij}^{(A)} = (1/2)(\gamma_i^{(A)} + \gamma_j^{(A)})$ .

- (iii) We assume the Au surface to be perfectly conductive. Consequently, we need to explicitly consider mirror charges in our model; when further assuming  $z = 0$  as the plane of reflection, the Coulombic interaction becomes

$$U^{(C)}(r_{ij}) = U^{(C)}(r_{ij}) + U^{(C)}(r_{i'}) + U^{(C)}(r_{i'j}) + U^{(C)}(r_{i'j'}) \quad (7)$$

with the mirror charges  $q_{i'} = -q_i$  and their positions  $\mathbf{r}_{i'} = (x_{i'}, y_{i'}, z_{i'}) = (x_i, y_i, -z_i)$ .

- (iv) We describe the solid–liquid interface in terms of a slab geometry with a lower confining wall, i.e., we assume periodicity in the  $x$ - and  $y$ -directions, but a finite extent,  $c_z$ , of the geometry in the  $z$ -direction, which is chosen such that no restriction in the orientation of any molecule occurs; thus,  $c_z \approx 1.2$ – $2$  nm, given their size and the slab width. We define the (orthorhombic) lattice vectors,  $\mathbf{a} = (a_x, 0, 0)$ ,  $\mathbf{b} = (b_x, b_y, 0)$ , and  $\mathbf{c} = (0, 0, c_z)$ , which, without the loss of generality, define the volume of the unit cell,  $V = a_x b_y c_z$ , and which we collect within the matrix  $\mathbb{V} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ . Together with the molecular basis, given by  $\mathbf{R}^N$ ,  $\mathbf{P}^N$ , and all  $N$  (rigid) molecular blueprints,  $\mathbf{r}^{N_I}$ , we now define the supramolecular lattice

$$\mathcal{G} = \mathcal{G}(\mathbf{R}^N, \mathbf{P}^N, \mathbb{V}) = \{\mathbf{R}^N, \mathbf{P}^N, \mathbb{V}\} \quad (8)$$

which gives rise to all atomic coordinates in the lab frame,  $\mathbf{r}^n$ , i.e., the molecular crystal structure of the system (see S.I. Section S2.2).

- (v) The force field between the atomic entities and the Au surface is described via an LJ-type wall potential<sup>54</sup>

$$U^{(\text{wall})}(z_i) = 2\pi\epsilon_{wi} \left[ \frac{2}{5} \left( \frac{\sigma_{wi}}{z_i} \right)^{10} - \left( \frac{\sigma_{wi}}{z_i} \right)^4 - \frac{\sqrt{2}\sigma_{wi}^3}{3(z_i + (0.61/\sqrt{2})\sigma_{wi})^3} \right] \quad (9)$$

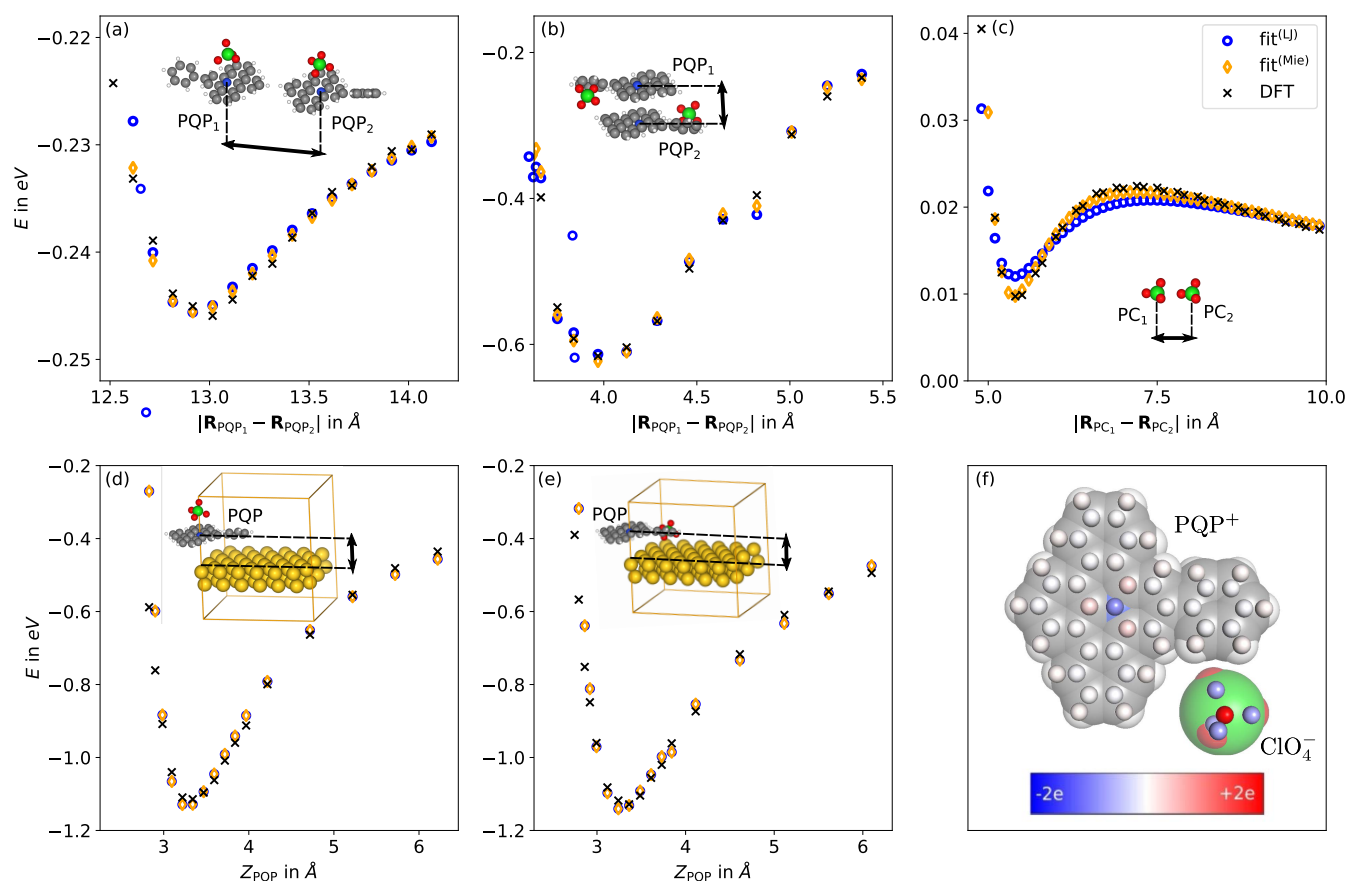
In the above relation,  $z_i$  is the height of atom  $i$  above the surface and  $\sigma_{wi}$  and  $\epsilon_{wi}$  are the length and energy parameters of the interactions of each atom  $i$  with the wall, respectively.

- (vi) Finally, we express the electrostatic interfacial potential between the electrode and the Au surface by an external, homogeneous electrostatic field,  $E_z$  (i.e., oriented perpendicular to the surface); we account for this potential via  $U^{(\text{field})}(z_i) = z_i q_i E_z$ .<sup>55</sup>

Thus, and eventually, the total potential energy of our model is given by the expression

$$U(\mathbf{r}^n, \mathbb{V}; E_z) = \sum_{i \neq j}^n * [U^{(C)}(r_{ij}) + U^{(\text{sr})}(r_{ij})] + \sum_{i=1}^n [U^{(\text{wall})}(z_i) + U^{(\text{field})}(z_i)] \quad (10)$$

with “sr” standing for “LJ” or “Mie”; we recall that  $\mathbf{r}^n$  is the set of all  $n$  atomic positions  $\mathbf{r}_i$  in a lattice with slab geometry (defined by the unit cell  $\mathbb{V}$ ). If not present (and not explicitly addressed), the electric field will be dropped in the argument



**Figure 3.** Energies as obtained in ab initio simulations (black crosses) and fitted data, using the force field (involving LJ interactions (open blue circles) or Mie interactions (open orange diamonds)); see Section 2.3.1. Also shown are, with labels (a)–(e), five schematic sketches of the five archetypical configurations of the molecules (along with their relative displacements, schematically indicated via the arrows as the distances vary along the abscissa); the related energy curves are used to fit the parameters of the force field, as outlined in the text; the labels correspond to the itemization (a)–(e) used in Section 2.3.2. (f)  $\text{PQP}^+$  and the  $\text{ClO}_4^-$  molecules, drawn to scale and using the Mie force field for the short-range interactions: atomic entities are shown as transparent spheres with their diameters fixed by their respective optimized  $\sigma_i$  values and their Bader charges (see the color code).

list of eq 10, that is,  $U(\mathbf{r}^n, \mathbb{V}; E_z = 0) \equiv U(\mathbf{r}^n, \mathbb{V})$ . The notation “ $\sum^*$ ” indicates that summation is only carried out over atoms, labeled with Latin indices  $i$  and  $j$ , which belong to different molecules  $I$  and  $J$  (with  $I \neq J$ ), i.e., molecules being labeled with capital indices. The energy given in eq 10 and the corresponding force fields are efficiently evaluated using the software package LAMMPS.<sup>56</sup>

To evaluate the long-range Coulomb term,  $\sum_{i \neq j}^* U^{(C)}(r_{ij})$ , in the given slab geometry, we use numerically reliable and efficient slab-corrected three-dimensional (3D) Ewald-summation techniques.<sup>57–59</sup> The other terms in eq 10 are evaluated via direct lattice summation techniques.

**2.3.2. Parametrizing the Classical Model via Ab Initio Calculations.** In this work, the blueprint of each molecule  $\mathbf{r}^{N_i}$  is obtained from electronic structure calculations based on density functional theory (DFT), using dispersion-corrected ab initio structure optimization,<sup>35,36</sup> as described in Section 2.2. The partial charges of the atoms,  $q_i$ , are parameterized via a Bader analysis<sup>60</sup> and are collected in Tables S2 and S3 in the S.I. Section S2.3. These charges are directly transferred to the atomic entities. We repeat that throughout the electrolyte molecules have not been considered explicitly: instead, we treat within the force field the electrolyte as an effective,

homogeneous medium, introducing the electric permittivity of water  $\epsilon_r$ .

To fix the remaining model parameters that specify the interactions in eq 10, we search for each atomistic entity (labeled  $i$ ) the set of atomistic model parameters (specified below), which reproduces via eq 10 the ab initio energies as good as possible. On the one side, we consider either the length and the energy parameters of the LJ potential (denoted  $\mathcal{L} = \{\sigma_i, \epsilon_i\}$ ) or the length and the energy parameters together with the exponents of the Mie potentials (denoted  $\mathcal{M} = \{\sigma_i, \epsilon_{ij}, \gamma_i^{(R)}, \gamma_i^{(A)}\}$ ), as well as the wall parameters,  $\mathcal{W} = \{\sigma_{wi}, \epsilon_{wi}\}$ . To fix these parameters, we proceed as follows:

- (i) We first perform ab initio structure optimization for different characteristic molecular configurations specified below. Here, molecules are either positioned next to each other (without considering the wall) or above the Au surface: in the former case, we fix the positions of two selected atoms belonging to different molecules, with the atoms being separated by  $r_{ij}$ ; in the latter case, we keep the height,  $z_k$ , of one selected atom above the surface constant. Relaxation of all other degrees of freedom leads in the ab initio simulations to spatially and orientationally optimized molecular structures; they are de-

**Table 1.** Numerical Results for the Optimized LJ and Mie Parameters,  $\mathcal{L} = \{\sigma_i, \epsilon_i\}$  and  $\mathcal{M} = \{\sigma_i, \epsilon_i, \gamma_i^{(R)}, \gamma_i^{(A)}\}$ , for Each Element  $i$ , for the Results Depicted in Figure 3 ( $\sigma_i$  in Å and  $\epsilon_i$  in meV)<sup>a</sup>

	$\sigma_{\text{H}}$	$\sigma_{\text{C}}$	$\sigma_{\text{N}}$	$\sigma_{\text{O}}$	$\sigma_{\text{Cl}}$	$\epsilon_{\text{H}}$	$\epsilon_{\text{C}}$	$\epsilon_{\text{N}}$	$\epsilon_{\text{O}}$	$\epsilon_{\text{Cl}}$
$\mathcal{L}$	2.243	3.658	3.743	2.865	5.953	3.052	1.204	3.311	7.396	0.172
$\mathcal{M}$	2.236	3.703	3.328	2.428	4.956	3.999	0.946	2.021	11.481	5.289
	$\gamma_{\text{H}}^{(R)}$	$\gamma_{\text{C}}^{(R)}$	$\gamma_{\text{N}}^{(R)}$	$\gamma_{\text{O}}^{(R)}$	$\gamma_{\text{Cl}}^{(R)}$	$\gamma_{\text{H}}^{(A)}$	$\gamma_{\text{C}}^{(A)}$	$\gamma_{\text{N}}^{(A)}$	$\gamma_{\text{O}}^{(A)}$	$\gamma_{\text{Cl}}^{(A)}$
$\mathcal{L}$	12	12	12	12	12	6	6	6	6	6
$\mathcal{M}$	6.263	7.136	8.659	8.743	15.455	7.500	12.299	13.854	17.193	4.684

<sup>a</sup>Reference values from the literature are listed in Table S1 in the Supporting Information.

- noted  $\mathbf{d}^n(r_{ij})$  and  $\mathbf{d}^n(z_k)$ , respectively, with corresponding energies  $U_{\text{DFT}}(\mathbf{d}^n(r_{ij}), \mathbb{V})$  and  $U_{\text{DFT}}^{(\text{wall})}(\mathbf{d}^n(z_k), \mathbb{V})$ ; they are, themselves, functions of the interatomic distance,  $r_{ij}$ , and the atom–wall separation,  $z_k$ , of the selected atoms.
- (ii) For every optimized ab initio structure,  $\mathbf{d}^n(r_{ij})$  and  $\mathbf{d}^n(z_k)$ , obtained in this manner, we define a corresponding molecular configuration  $\mathbf{r}^n(r_{ij})$  and  $\mathbf{r}^n(z_k)$ , which is based on the above-introduced rigid molecular blueprint model  $\mathbf{r}^n$  (with the index  $I$  running now over all  $N$  molecules present in the respective DFT structure). To this end, we synchronize the COM positions of each molecule  $I$  in the ab initio simulation with the corresponding COM positions  $\mathbf{R}_I$  of its classical counterparts and align their orientation  $\mathbf{P}_I$  accordingly.
- (iii) Finally, we evaluate the corresponding energies with the help of the force field via eq 10 at zero electric field, i.e.,  $U_{\mathcal{L}/\mathcal{M}}(\mathbf{r}^n(r_{ij}), \mathbb{V})$  and  $U_{\mathcal{L}/\mathcal{M},\mathcal{W}}^{(\text{wall})}(\mathbf{r}^n(z_k), \mathbb{V})$ . We search for the best set of parameters  $\mathcal{L}$  (or  $\mathcal{M}$ ) and  $\mathcal{W}$  via simultaneously minimizing

$$\mathcal{F}_{\mathcal{L}/\mathcal{M}} = \sum_{\{r_{ij}\}} |U_{\text{DFT}}(\mathbf{d}^n(r_{ij}), \mathbb{V}) - U_{\mathcal{L}/\mathcal{M}}(\mathbf{r}^n(r_{ij}), \mathbb{V})|^2 \quad (11a)$$

$$\mathcal{F}_{\mathcal{L}/\mathcal{M},\mathcal{W}}^{(\text{wall})} = \sum_{\{z_k\}} |U_{\text{DFT}}^{(\text{wall})}(\mathbf{d}^n(z_k), \mathbb{V}) - U_{\mathcal{L}/\mathcal{M},\mathcal{W}}^{(\text{wall})}(\mathbf{r}^n(z_k), \mathbb{V})|^2 \quad (11b)$$

Of course, in the model, the same unit cell,  $\mathbb{V}$ , and the same number of particles,  $n$ , as in the respective ab initio simulations have to be used. Note that in eq 11a the wall term included in eq 10 is obsolete since the surface atoms are not considered.

These fits are based on five particularly chosen, archetypical configurations, to be discussed below. In the panels of Figure 3, we display schematic sketches of these configurations of the PQQ<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> molecules; these panels show the corresponding energy curves obtained from the force field, with parameters based on a fitting procedure to the ab initio energy profiles.

- (a) Tail-to-tail configuration (see the inset in Figure 3a): We have considered a series of ab initio structure optimizations at constant but successively increasing nitrogen–nitrogen distances,  $r_{\text{NN}}$ , in the  $x$ -direction (while keeping  $y_{\text{NN}}$  and  $z_{\text{NN}}$  constant) of an antiparallel oriented pair of PQQ<sup>+</sup> molecules; both cations are vertically decorated with a ClO<sub>4</sub><sup>-</sup> molecule. The aromatic parts of the PQQ<sup>+</sup> molecules lie flat in the  $x$ - and  $y$ -directions such that their tails face each other.

- (b) Face-to-face configuration (see the inset in Figure 3b): In this case, we consider antiparallel oriented but vertically stacked PQQ<sup>+</sup> molecules (both being horizontally decorated by ClO<sub>4</sub><sup>-</sup> molecules) under the variation of the nitrogen–nitrogen distance,  $r_{\text{NN}}$ , in the  $z$ -direction (while now keeping  $x_{\text{NN}}$  and  $y_{\text{NN}}$  constant). Again, the aromatic parts of the PQQ<sup>+</sup> molecules lie flat in  $x$ - and  $y$ -directions; however, and in contrast to case (a), these units face each other.
- (c) ClO<sub>4</sub><sup>-</sup>–ClO<sub>4</sub><sup>-</sup> configuration (see the inset in Figure 3c): Here, two ClO<sub>4</sub><sup>-</sup> molecules are considered, varying the chlorine–chlorine  $x$ -distance,  $r_{\text{ClCl}}$ , while keeping  $y_{\text{ClCl}}$  and  $z_{\text{ClCl}}$  constant.
- (d) Face-to-wall topped configuration (see the inset in Figure 3d): In this case, a single PQQ<sup>+</sup> molecule, lying flat and parallel to the  $(x, y)$ -plane, is located above two layers of Au and is vertically decorated by a ClO<sub>4</sub><sup>-</sup> molecule. The cell geometry is assumed to be periodic in the  $x$ - and  $y$ -directions and finite along the  $z$ -axis; in an effort to scan along the  $z$ -direction, we have performed a series of ab initio-based structure optimizations for selected fixed values of  $z_{\text{N}}$ , i.e., the  $z$ -position of the nitrogen in PQQ<sup>+</sup> above the Au surface. The LJ 10-4-3 potential<sup>54</sup> was used between the Au(111) surface and the molecules.
- (e) Face-to-wall beside configuration (see the inset in Figure 3e): In contrast to case (d), the PQQ<sup>+</sup> cation is now horizontally decorated by the ClO<sub>4</sub><sup>-</sup> anion such that both molecules are adsorbed on the Au surface. Again, the LJ 10-4-3 potential<sup>54</sup> defined in eq 9 was used between the Au(111) surface and the molecules.

In practice, we first optimize  $\mathcal{F}_{\mathcal{L}/\mathcal{M}}$ , given in eq 11a, involving thereby all interatomic force field parameters; their values are listed in Table 1 for the LJ and the Mie models. These parameters are then kept fixed and are used in the subsequent calculations to optimize the wall force field parameters via optimizing  $\mathcal{F}_{\mathcal{L}/\mathcal{M},\mathcal{W}}^{(\text{wall})}$ , specified in eq 12; the emerging parameters are listed in Table 2. In Figure 3f, we present a visualization of the molecules PQQ<sup>+</sup> and ClO<sub>4</sub><sup>-</sup>, using these optimized parameters and providing information about the charge of the atomic entities via the color code.

### 3. IDENTIFICATIONS OF SELF-ASSEMBLY SCENARIOS

With the classical force field for the PQQ<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> molecules introduced in Section 2.3.1 at hand, we are now ready to identify the ordered ground-state configurations of these molecules as they self-assemble on the Au surface, immersed into an electrolyte and exposed to an electric field. While we leave a more comprehensive and systematic investigation of these self-assembly scenarios to a future publication,<sup>61</sup> we focus

**Table 2. Numerical Results for LJ Length and Well-Depth Parameters,  $\sigma_{wi}$  in Å and  $\epsilon_{wi}$  in meV, between the Wall and Each Element  $i = [\text{H}, \text{C}, \text{N}]$  and  $j = [\text{O}, \text{Cl}]$ , Grouped by the Molecules They Belong to (PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup>), for Intermolecular Short-Range LJ Parameters Listed in Table 1 and Corresponding  $\sigma_{wi}$  and  $\epsilon_{wi}$  Parameters for Intermolecular Short-Range Mie Parameters Also Listed in Table 1**

	$\sigma_{w[\text{H,C,N}]}$	$\sigma_{w[\text{O,Cl}]}$	$\epsilon_{w[\text{H,C,N}]}$	$\epsilon_{w[\text{O,Cl}]}$
$\mathcal{W}^{(\text{LJ})}$	3.197	3.625	3.741	15.781
$\mathcal{W}^{(\text{Mie})}$	3.208	3.630	3.698	20.167

in this contribution on the technical details of our approach and on a few selected sets of external parameters (i.e., the electric field strength and the particle density).

Our overall objective is to find for our system the global minimum of the total free energy,  $F$ , at  $T = 0$  K as a function of the positions and orientations of all molecules per unit cell for a given value of cell volume and  $E_z$ ; at  $T = 0$ , this task reduces to the minimization of the internal energy  $U$ . The minimum has to be found in a huge-dimensional parameter space, spanning the positions and orientations of the molecules, and by the parameters specifying the unit cell. To be more specific, the dimensionality is set by the number of parameters to be optimized, which read 64, 76, and 88 for five, six, and seven molecules per unit cell, respectively. It is a particular strength of our optimization algorithm (as detailed below) to identify in an efficient and reliable manner minima in such high-dimensional search spaces.

For this purpose, we use a memetic search algorithm that combines evolutionary search strategies (EA)<sup>62–68</sup> and local, steepest gradient descent procedures (LG):<sup>69,70</sup> first, a total number of  $N_{\text{EA}}$  different lattice configurations,  $\mathcal{G} = \mathcal{G}(\mathbf{R}^N, \mathbf{P}^N, \mathbb{V})$  as defined in eq 8, is generated. We note that among those configurations we have also intentionally included as “educated guesses” molecular configurations, inspired by the experimental self-assembly scenarios identified in ref 3; however, it should be emphasized that this information is only available for the PQP<sup>+</sup> ions, as the experiment does not provide any information about the positions of the perchlorate ions. This population,  $\mathcal{G}^{N_{\text{EA}}}$ , is exposed to concepts of natural (or, rather, artificial) selection. At every iteration step of the EA, a new configuration, i.e., an offspring, is created from existing configurations of the most recent population via crossover and mutation operation. This new configuration is then subjected to an LG optimization, an operation that represents by far the most time-consuming task in our algorithm and is performed in parallel using the “mpi4py” framework.<sup>71–73</sup> For an optimal load balance, we additionally spawn a master thread on one of the mpi processes to asynchronously distribute optimization tasks of offspring configurations among all idle mpi processes. The relaxed configurations are gathered by this master thread, which then decides, via a criterion primarily based on the respective internal energy of the configurations, whether the new relaxed particle arrangements are accepted or rejected.

Since the experimental observations<sup>3,5</sup> provide evidence of a structural organization of the molecules into supramolecular lattices, the center-of-mass coordinates of the molecules,  $\mathbf{R}^N$ , and their orientations,  $\mathbf{P}^N$ , as well as the parameters defining the unit cell,  $\mathbb{V}$ , (see Section 2.3.1 and Figure 1 for details), are

the variables that have to be optimized for the search of ground-state configurations: we minimize  $U(\mathbf{r}^N, \mathbb{V}; E_z)$ , defined in eq 10, with respect to  $\mathbf{R}^N$ ,  $\mathbf{P}^N$ , and  $\mathbb{V}$ , keeping the number of molecules  $N$ , the unit-cell volume  $V$  (with fixed slab width  $c_z$ ), and the electrostatic field strength  $E_z$  constant.

In more detail, we proceed as follows:

- (i) It is common in evolutionary algorithms to define a genome representation of the entity that is subject to optimization.<sup>74,75</sup> In our case, we represent a supra-molecular lattice configuration phenotypically (rather than genotypically<sup>75</sup>) by the set  $\mathcal{G} = \{\mathbf{R}^N, \mathbf{P}^N, \mathbb{V}\}$  as defined by eq 8, i.e., the set of all COM coordinates and orientations of all molecules as well as the lattice vectors.
- (ii) In the first step, two configurations (labeled henceforward by Latin indices),  $\mathcal{G}_i$  and  $\mathcal{G}_j$ , are chosen at random or via the “roulette wheel” method (see item (v) below) from the evolutionary population;<sup>62,63,76–78</sup> this strategy favors parents of high quality, hence making them more likely to be used for reproduction than “weaker” configurations, i.e., configurations with higher energy from the evolutionary population. Then, these two configurations are combined via a crossover operation (i.e., a cut-and-splice process; for more details, see below), creating thereby an offspring configuration,  $\mathcal{G}_{i\oplus j}$ , with the subscript “ $i \oplus j$ ” emphasizing the executed crossover operation between  $\mathcal{G}_i$  and  $\mathcal{G}_j$ . The purpose of this operation is to save high-quality blocks of the genetic material (e.g., the relative positions and orientations of molecules within the unit cell) to efficiently sample the parameter space.<sup>62,63,74–78</sup>
- (iii) In the second step, the newly generated offspring configuration,  $\mathcal{G}_{i\oplus j}$ , is then exposed to random mutation moves: these are either translations or rotations of single molecules, swaps of center-of-mass positions or orientations of pairs of molecules, or deformations of the unit cell, each of them with a certain probability and within preset numerical boundaries. This step of the algorithm has the purpose of exploring disconnected areas in the parameter space, a feature that is indispensable in global minimization techniques.
- (iv) After these two steps, and assuming that the offspring configuration,  $\mathcal{G}_{i\oplus j}$ , does not represent a local minimum with respect to the potential energy, a local energy minimization is performed. Here, we mainly rely on the “scipy” implementation of the SLSQP gradient descent algorithm<sup>69,70</sup> (allowing us to define numerical boundaries and constraints on the parameters during the optimization), which minimizes the forces and torques between the molecules as well as the stress of the unit cell. These tools are very helpful to keep the unit-cell volume fixed and to prevent re-orientations of the molecules where some of their atomic constituents would be transferred into positions outside the slab geometry, ensuring thus that  $z_i > 0$  for all atoms. Subsequently, we perform several “basin-dropping” (BD) steps, where we further try to lower the energy of the configuration by applying several small random “moves” in the parameter space of the LG-optimized offspring; from the emerging configurations, only the ones with low energies are accepted. This specific operation turned out to considerably improve the



convergence rate of the local optimization, in particular if multiple and alternating sequences of LG and BD runs are applied.

- (v) After the local search procedure, the optimized offspring configuration,  $\mathcal{G}_{i\oplus}$ , becomes a new candidate to enter the evolutionary population,  $\mathcal{G}^{N_{EA}}$ . The objective of the EA is to retain the best configurations (i.e., the energetically most favorable ones) within the population and to include only the candidates with energy values better or comparable to those of the current population. In an effort to quantify the quality of the candidates, their so-called fitness is evaluated,<sup>62,63,75–78</sup> for which we have used in this contribution the function

$$F(U) = \exp\left(-s \frac{U - U_{\min}}{U_{\max} - U_{\min}}\right) \quad (12)$$

$F(U)$  is a monotonic function of the energy  $U$  of the candidates, whose value ranges within the interval  $0 \leq F(U) \leq F(U_{\min}) = 1$ ;  $U_{\min}$  and  $U_{\max}$  are the minimal and maximal energies appearing in the population, respectively. The selection parameter  $s$  quantifies the reproduction rate for configurations within the population in the sense that large values of  $s$  tend to exclude configurations with low fitness from reproduction; following ref 78, we commonly use  $s = 3$ . The aforementioned “roulette wheel” method for choosing suitable parent configurations also relies on the fitness function (and hence the selection parameter): assuming that the configurations within the population  $\mathcal{G}^{N_{EA}}$  are sorted by their respective fitness values in descending order,  $F(U_i) > F(U_{i+1})$ , the probability,  $f(U_i)$ , of a configuration,  $\mathcal{G}_i$ , to be selected for reproduction is given in terms of the relative fitness<sup>62,76–78</sup>

$$f(U_i) = \frac{F(U_i)}{\sum_{j=1}^{N_{EA}} F(U_j)} \cdot \left[ \sum_{k=1}^{N_{EA}} \sum_{j=k}^{N_{EA}} F(U_j) \right]^{-1} \quad (13)$$

$N_{EA}$  is the total number of configurations within the population. With a certain probability (commonly in 20% of all crossover moves), we allow reproduction between randomly chosen configurations.

- (vi) Once a new configuration is accepted to enter the population, another configuration has to be eliminated. The probability  $p(U_i)$  for a configuration,  $\mathcal{G}_i$ , to be replaced is given by

$$p(U_i) = \exp[-sF(U_i)] \left[ \sum_{j=1}^{N_{EA}} \exp[-sF(U_j)] \right]^{-1} \quad (14)$$

a value that is again related to the fitness of the configuration,  $F(U_i)$ , and the selection parameter  $s$ . Thus, configurations with low fitness are more likely to be eliminated. In any case, a few of the best configurations within the population are retained in an effort to keep the so far best solutions as appropriate parent candidates for the above-mentioned crossover procedures (a strategy referred to in the literature as elitism<sup>76</sup>). It should be emphasized that this strategy does not follow biological selection mechanisms,<sup>79</sup> where populations are replaced entirely once that new

generations have been formed; however, our strategy ensures to protect the best genetic material from extinction during the entire search procedure.<sup>76,78</sup>

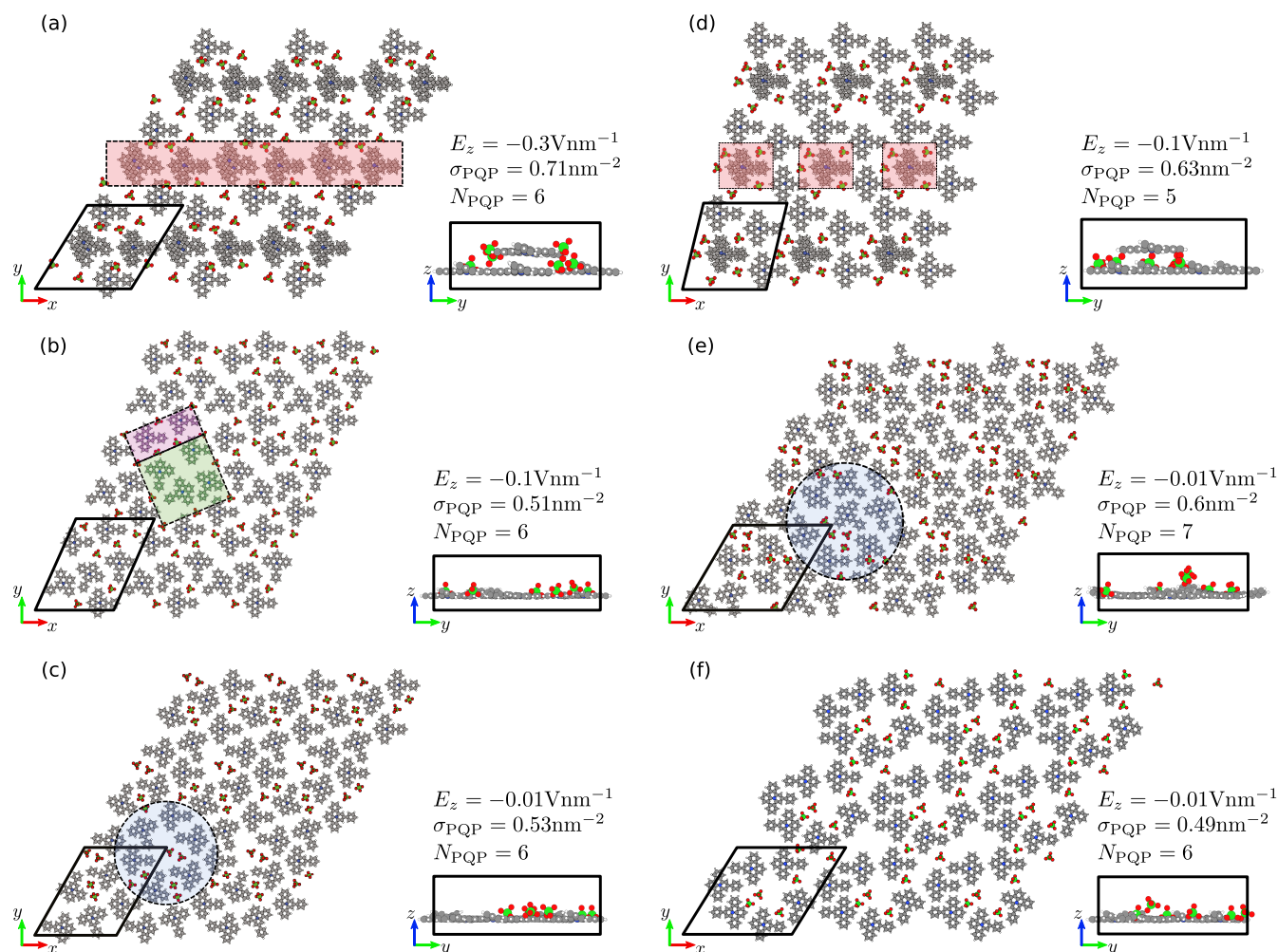
- (vii) In an effort to maintain diversity within the population, an additional operation (in the literature referred to as nichening<sup>80,81</sup>), is applied: locally optimized offspring configurations will be discarded if the values of their energy are too close to the energy of any configuration currently in the population, avoiding thereby that the population is overrun by structurally identical configurations. At the same time, the maintenance of genetic diversity is guaranteed.

However, this procedure alone cannot cope with “degenerate” configurations, i.e., if structurally distinct configurations have essentially the same energy values (within the specified nichening tolerance). In our approach, we allow configurations to enter the population only if their structures differ significantly from those of the competing, degenerate configurations. To quantify the structural difference between configurations, we associate a feature vector,  $\mathbf{f}_i$ , which collects a set of order parameters pertaining to configuration  $\mathcal{G}_i$  (see S.I. Section S3.2 for details). The degree of similarity between two configurations,  $\mathcal{G}_i$  and  $\mathcal{G}_j$ , is then evaluated by taking the Euclidean distance between the corresponding feature vectors, i.e.,  $\Delta_{ij} = |\mathbf{f}_i - \mathbf{f}_j|$ ; similar configurations will have a small distance, while unlike configurations will have a large distance. If  $\Delta_{ij}$  is above a certain threshold value, the offspring configuration,  $\mathcal{G}_{i\oplus}$ , will not be discarded by the energy-nichening operation.

To offer the reader an insight into the computational complexity of our project, we outline via a few characteristic numbers the computational limitations: the bottlenecks of the identifications of self-assembly scenarios are (i) the huge number of calls of energy evaluations in the optimization step, as underlined via some example: per generation, we have at least  $10^4$  calls of the energy kernel; for each state point, we need at least  $10^4$  generations, which leads to an absolute minimum of  $10^8$  calls of the energy kernel for one (!) set of system parameters and (ii) the optimization of the energy in a high-dimensional search space (as specified above), ranging from  $\sim 60$  to  $90$ , depending on the number of molecules.

Summarizing, the complexity of the problem at hand forces us to use all of the above-listed advanced optimization tools, including a basin-hopping memetic approach combining the heuristic nature of evolutionary strategies with deterministic local gradient descent algorithms.<sup>62</sup> The gradient descent method deterministically evaluates every local minimum of the basin with high accuracy (which is additionally sped up by the “basin-dropping” procedure), while the evolutionary search gradually adapts its population to the energetically most favorable solution, exploring the search space for the global optimum.

To round up this section, it should be noted that a variety of techniques have been used in the literature for related optimization problems; among those are Monte Carlo or molecular dynamics-based techniques such as simulated annealing,<sup>82,83</sup> basin-hopping,<sup>84–86</sup> minima-hopping,<sup>87,88</sup> and eventually evolutionary approaches such as genetic algorithms.<sup>62–68,80,81,89–92</sup> The decision on the method of choice relies on the specific problem: for instance, as Hofmann et al.



**Figure 4.** Results for the ground-state configurations of PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> molecules, adsorbed on a Au(111) surface under the influence of an external electrostatic field  $E_z$ , as they are obtained via the numerical procedure, as specified in Section 3; calculations are based on the classical model for the molecules, involving the Mie potential (for details, see Section 2.3.1). In the main panels, configurations are shown in a periodically extended view as projections onto the  $(x, y)$ -plane and in the respective insets as projections onto the  $(y, z)$ -plane; in the main panels, the respective unit cells are highlighted by thick black lines. Results are shown for different values of the number of PQP<sup>+</sup> molecules, the surface density  $\sigma_{\text{PQP}}$ , and the electrostatic field  $E_z$ ; see labels in different panels and Table 3 for details. The red (gray) shaded areas, framed with dashed lines, in (a) and (d) emphasize PQP<sup>+</sup> molecules that sit on top of other cations, forming a bilayer structure. The dashed, shaded, magenta (gray) rectangular, and green (gray) square areas in (b) represent tilings formed by perchlorate molecules within the dense PQP<sup>+</sup> monolayer configuration. The dashed, shaded, and blue (gray) circles in (c) and (e) emphasize, quantitatively, the porous and auto-host-guest tiles identified in the experiment (see Figure 4A,C in ref 3, respectively).

used the SAMPLE technique (see refs 8, 22 and 23), relying on a discretization of the search space into limited, archetypical, intermolecular motives and elaborate data fitting of emerging force fields to describe intermolecular interactions. To the best of our knowledge, this approach has neither been applied to molecular motives beyond monolayer configurations or to charged molecules, so far, nor has it been used in combination with an external control parameter, such as an electric field or systems composed of multiple components. In general, the fact that the number of archetypical intermolecular motives grows rapidly with the increasing size of the molecules bears the risk of hitting very soon the limits of computational feasibility. However, suitable adaptations of this strategy and/or a combination with evolutionary search strategies or with reinforcement learning, which has, for instance, very successfully been applied to protein folding problems<sup>93</sup> in a similar way as AlphaGo<sup>94</sup> was able to master the infamous board game, might represent a viable route to circumvent the

aforementioned limitations; thus, future investigations of such intricate problems as the complex monolayer-to-bilayer transition, addressed in this contribution, might come within reach.

## 4. RESULTS

**4.1. General Remarks and System Parameters.** Below, we present selected results for self-assembly scenarios of PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> molecules on an Au(111)–electrolyte interface under the influence of an external electrostatic field, as obtained via the algorithm presented in the preceding sections. Our choice of parameters is guided by the experimentally observed molecular configurations.<sup>3</sup> We demonstrate that our proposed strategy is indeed able to reproduce on a semi-quantitative level the experimentally observed self-assembly scenarios.<sup>3</sup> As a consequence of the still sizable costs of the numerical calculations, we leave more detailed investigations (where we systematically vary the system parameters) and a

quantitative comparison of our results with the related experimental findings<sup>3</sup> to a future contribution.<sup>61</sup>

To be more specific, we have used the following values for the (external) system parameters:

- An indication for the number of molecules per unit cell is provided by the experiment:<sup>3</sup> we have considered unit cells containing 10, 12, and 14 pairs of PQP<sup>+</sup> and ClO<sub>4</sub><sup>-</sup> molecules. These numbers of molecules include, of course, also the related mirror molecules and correspond to 630, 636, and 742 atomic entities per unit cell, respectively (which interact via short-range and long-range potentials, which are subject to particle wall interaction and are sensitive to an external electrostatic field).
- Also, the actual values of the surface area *A* are motivated by estimates taken from the experiment:<sup>3</sup> we have varied *A* within the range of 6.5–12.25 nm<sup>2</sup>, assuming a step size of typically 0.5 nm<sup>2</sup>; systems will be characterized by the surface density of the PQP<sup>+</sup> molecules, defined as  $\sigma_{\text{PQP}} = N_{\text{PQP}}/A$ , with *N*<sub>PQP</sub> being the number of PQP<sup>+</sup> molecules per unit cell.
- The range of the experimentally realized values for the electrostatic field strength *E<sub>z</sub>* is, however, difficult to be estimated since the major drop in voltage occurs near the negatively charged Au surface and the nearby layers of cations,<sup>55</sup> which is not directly accessible in the experiment. Therefore, we have covered, at least in this first contribution, several orders of magnitude in the value for *E<sub>z</sub>* within a range that extends (on a logarithmic grid) from *E<sub>z</sub>* = −10 to −10<sup>−3</sup> V/nm; in addition, we have also performed calculations at zero electrostatic field.

It should be mentioned that we have used in all of these calculations the Mie potential within the classical model, since the related LJ model is not able to fit the ab initio data with a comparable and sufficient accuracy (see also the discussion in Section 2.3.1).

We have covered in total approximately 176 combinations of these parameters (that is the unit-cell volume *V* with a constant slab width *c<sub>z</sub>*, the number of molecules *N*, and the electrostatic field strength *E<sub>z</sub>*); for each of these, we performed independent evolutionary searches with a population size of typically  $\mathcal{G}^{\text{NEA}} = 40$  configurations. Some details about the numerical costs of our calculations can be found in S.I. Section S4.1.

**4.2. Discussion of the Results.** **4.2.1. Lateral Particle Arrangements.** In this section, we discuss the lateral self-assembly scenarios of the PQP<sup>+</sup> and of the ClO<sub>4</sub><sup>-</sup> molecules. Selected results for our numerical investigations are presented in Figure 4 and in, on a more quantitative level, in Table 3. The actual values have been chosen in an effort to reproduce, at least on a qualitative level, the results obtained in the experimental investigations. Indeed, the sequence of the obtained ordered ground-state configurations (shown in panels a–c) clearly indicates the transition from a stratified bilayer configuration (identified at a rather strong electrostatic field strength of *E<sub>z</sub>* = −0.3 V/nm) to a self-host–guest monolayer structure (obtained by reducing the field down to *E<sub>z</sub>* = −0.1 V/nm) and eventually to an open-porous configuration (identified at *E<sub>z</sub>* = −0.01); similar observations have been reported in the related experimental study.<sup>3</sup>

From the results of our investigations (which are shown only selectively), we learn that an electrostatic field strength of *E<sub>z</sub>* =

**Table 3. Results of Evolutionary Ground-State Search for Different Electric Field Strengths, *E<sub>z</sub>*, for Different Unit Cell Areas, *A*, and Number of PQP<sup>+</sup> Molecules, *N*<sub>PQP</sub><sup>a</sup>**

	<i>E<sub>z</sub></i> (V/nm)	<i>U</i> / <i>N</i> <sub>PQP</sub> (eV)	<i>N</i> <sub>PQP</sub>	<i>A</i> (nm <sup>2</sup> )	<i>N</i> <sub>PQP</sub> / <i>A</i> (nm <sup>−2</sup> )
(a)	−0.30	−1.5804	6	8.5	0.705882
(b)	−0.10	−1.7276	6	11.75	0.510638
(d)	−0.10	−1.7274	5	8.0	0.625000
(c)	−0.01	−1.6445	6	11.25	0.533333
(f)	−0.01	−1.6364	6	12.25	0.489796
(e)	−0.01	−1.6332	7	11.75	0.595745

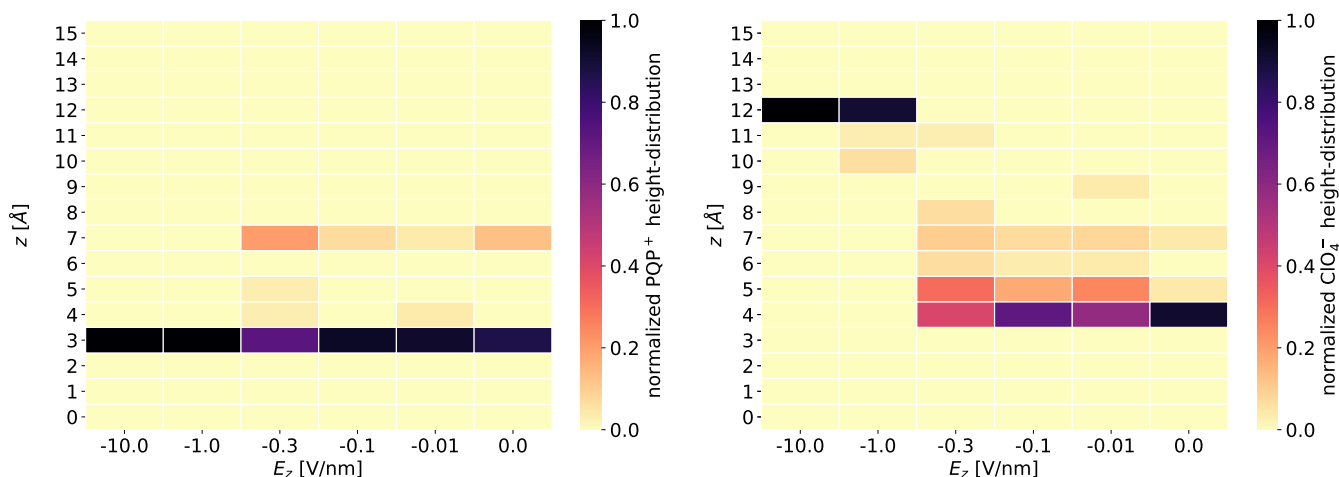
<sup>a</sup>Each line represents an evolutionary search. The respective structures are presented in Figure 4.

−0.3 V/nm always leads to bilayer configurations, similar to the one shown in Figure 4a. This stratified bilayer configuration represents the energetically most favorable one as we vary at fixed *E<sub>z</sub>* the volume of the unit cell and the number of molecules within the respective ranges, specified in the preceding section; the numerical data of the related internal energy are compiled in Table 3.

As we proceed to *E<sub>z</sub>* = −0.1 V/nm, we observe self-assembly scenarios as the ones depicted in Figure 4b,d, which correspond to self-hosts–guest configurations observed in the experiment;<sup>3</sup> for the data presented in these panels, two different values for *N*<sub>PQP</sub> (and hence for  $\sigma_{\text{PQP}}$ ) have been considered: the monolayer configuration shown in panel b has a slightly better value for the internal energy (per molecule) than the rhombohedral bilayer configuration shown in panel d; however, as can be seen in Table 3, the energy differences are very tiny: differences of the order of 10<sup>−4</sup> eV correspond to values where we hit the numerical accuracy of the ab initio-based energy values.

Eventually, we arrive at the so-called open-porous structures, observed in the experiment:<sup>3</sup> the ground-state configurations depicted in Figure 4c,e,f are evaluated at the same electrostatic field strength of *E<sub>z</sub>* = −0.01 V/nm, assuming different values for *N*<sub>PQP</sub> and  $\sigma_{\text{PQP}}$ ; the open-porous pattern emerging in panel c is the most favorable one in terms of energy per molecule (see Table 3 for the numerical details). There are, however, several serious competing structures with minute energy differences at this value of the electric field strength: another open-porous structure, depicted in panel f, with an energy penalty of less than 8.1 meV per PQP<sup>+</sup> molecule compared to case (c) and also a considerably denser configuration, depicted in panel e, with an internal energy value worse by only 11.3 meV compared to case (c) and by 3.2 meV compared to case (f).

From the numerical point of view, the following comments are in order: for a fixed state point, the energy differences of competing structures attain values that hit the limits of the accuracy of the ab initio-based simulations, which can be estimated to be of the order of 0.1–0.01 eV per molecule for dispersive interactions.<sup>36,95–97</sup> These values set the limits of our numerical accuracy. For completeness, we note that for the results for the energies obtained via the classical force field (which are based on LAMMPS calculations), we estimate that our results are numerically reliable down to ~10<sup>−6</sup> eV per atom; within the range of such minute energy differences, no competing structures have been found in our investigations. In general, we observe that the energy differences for the energetically optimal ground-state configurations become smaller as the electrostatic field tends toward zero. Even



**Figure 5.** Height distribution of  $\text{PQP}^+$  (left) and  $\text{ClO}_4^-$  (right) molecules as functions of the considered values of the electrostatic field,  $E_z$ , normalized to the number of respective molecules (see the color code on the right-hand side of the panels; the value of one means that all respective molecules in all considered configurations are counted in one specific bin; see the text for the energy considered in this analysis). Along the vertical axes, the binning is performed in steps of 1 Å:  $z = 0$  Å marks the position of the gold surface, and the slab width amounts to 12 Å.

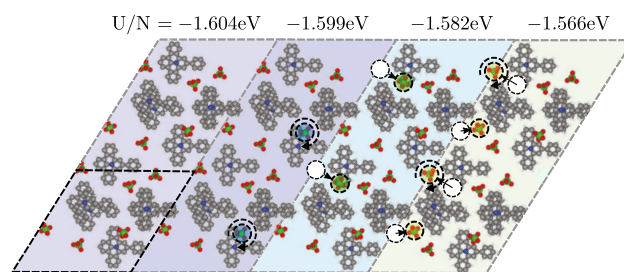
though the optimization algorithm (as outlined in Section 3) has turned out to be very efficient and reliable, we observe (in particular for smaller values of the external field) that new configurations are included in the population of the best individuals even after a large number of optimization steps.

**4.2.2. Vertical Particle Arrangements.** In Figure 5, we present in separate panels the height distributions of  $\text{PQP}^+$  and  $\text{ClO}_4^-$  as functions of the electrostatic field,  $E_z$  (which is binned for the six different values of  $E_z$  that were investigated); along the vertical axis, we count (for a given value of  $E_z$ ) the occurrence of the respective molecules in bins of 1 Å, and normalize by the total number of all considered configurations identified by the evolutionary algorithm, which are located within an interval of at most  $1 k_B T$  (or 43 meV) above the configuration with the best energy, which is of the same order of magnitude as the values presented in Table 3 for different electric field strengths,  $E_z$ . Note in this context that the distance of the first layer of  $\text{PQP}^+$  molecules can be directly estimated by the vertical equilibrium position of carbon atoms,  $z_{\text{C}}^{(\text{eq})} = 3.166$  Å, obtained by minimizing eq 9 for a single carbon atom with  $\sigma_{\text{wC}}^{(\text{Mie})} = 3.208$ , taken from Table 2.

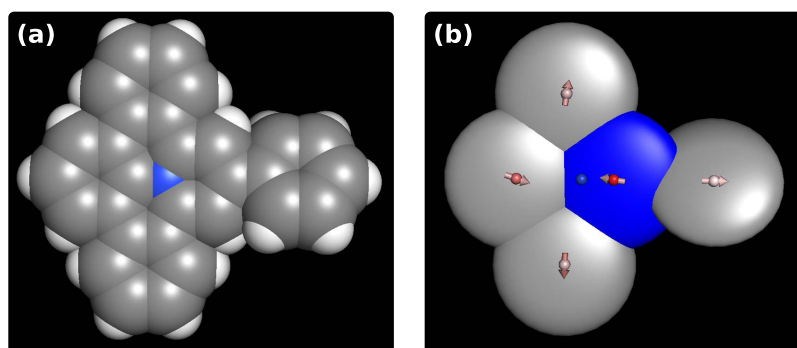
For the high values of the field strength (i.e., for  $E_z = -10$  and  $-1$  V/nm), the  $\text{PQP}^+$  ions are preferentially adsorbed onto the gold surface as a closely packed monolayer (see the left panel of Figure 5), while the perchlorate anions are strongly dissociated and assemble as far from the gold surface as possible (corresponding in our investigation to the numerical value of the slab height, which we fixed to 12 Å); see the right panel in Figure 5. This situation represents an extreme case in the sense that neither the Coulomb nor the short-range Mie interactions between ions and anions can compensate for the strong negative surface potential.

Decreasing now the magnitude of the electrostatic field to the more moderate value of  $E_z \sim -0.3$  V/nm reveals the emergence of a bilayer structure, formed by the  $\text{PQP}^+$  molecules, with pronounced peaks located at  $z_{\text{PQP}^+}^{(1)} \sim 3$  Å and  $z_{\text{PQP}^+}^{(2)} \sim 7$  Å, with relative weights of 72 and 21%, respectively. Such stratified bilayer configurations (as depicted in Figure 4a) are in competition with structures similar to ones shown in Figure 4d. Note that in parallel a far more complex height distribution of the perchlorate molecules sets in as soon

as the now moderate electrostatic field allows them to proceed toward the interior of the slab: now, more than half of the  $\text{ClO}_4^-$  anions are located “in between” the  $\text{PQP}^+$  “layers”, trying on the one side to compensate the charges of one or several  $\text{PQP}^+$  “partners” in the slab region and “filling spatial holes” wherever they can, on the other side. A large portion of perchlorate ions is even allowed to adsorb on the surface at a distance of  $z_{\text{ClO}_4^-}^{(1)} \sim 4.074$  Å; note that these COM positions above the interface are larger than the minimal height of the  $\text{PQP}^+$  cations due to two reasons: if one face of the oxygen tetrahedron is oriented toward the interface (i.e., parallel to the gold surface), the COM of the  $\text{ClO}_4^-$  ion is increased by a value of  $z_{\text{Cl}} - z_{\text{O}} \approx 0.492$  Å, with respect to the oxygen atoms. These atoms themselves have an equilibrium distance to the surface of  $z_{\text{O}}^{(\text{eq})} \approx 3.582$  (evaluated by minimizing eq 9 for a single oxygen atom with  $\sigma_{\text{wO}}^{(\text{Mie})} = 3.630$ , cf. Table 2), summing up to the presented minimal COM distance of the adsorbed  $\text{ClO}_4^-$  molecules from the interface. Note that the height distribution of the  $\text{ClO}_4^-$  ions is now rather broad (see Figure 6), which is



**Figure 6.** Four structurally different configurations of perchlorate ions (framed by gray, dashed lines and differently shaded areas) at optimized, fixed cell geometry (indicated in the bottom-left corner by the black, dashed line) and optimized, fixed positions and orientations of the  $\text{PQP}^+$  ions; the energies of these four configurations range within an interval of 38 meV (per  $\text{PQP}^+-\text{ClO}_4^-$  pair), as obtained after an evolutionary energy minimization of solely the degrees of freedom of the  $\text{ClO}_4^-$  ions, starting from the configuration depicted in Figure 4a. Changes in the structure as one proceeds from the left to right are highlighted by respective circles (specifying the position of the “moving” perchlorate ion) and arrows.



**Figure 7.** (a) Atomistic model of the  $\text{PQP}^+$  molecule as used in this contribution (white: hydrogen atoms, gray: carbon atoms, and blue: nitrogen atom); see also Figure S3 in the S.I. (b) Related coarse-grained model in a hierarchy of ever simpler models, using, e.g., Gay–Berne potentials to account for the van der Waals interaction of all atoms in the specific rings and a multipole expansion to second order (monopole as colored points, dipole moments as small arrows) for the electrostatic interaction.

definitely attributed to their relatively smaller size and their considerably higher mobility, as compared to their cationic counterparts (see also the discussion below); these features make a conclusive interpretation of the roles of the  $\text{ClO}_4^-$  ions in the structure formation of the entire system rather difficult. Our interpretation is that the perchlorate ions are, due to their small spatial extent and their high mobility, able to compensate for local charge mismatches and to act as spatial spacers between the cations.

Decreasing further the magnitude of the electrostatic field down to  $E_z \sim -0.1$  and  $-0.01$  V/nm provides unambiguous evidence that the formation of the  $\text{PQP}^+$  ions into bilayer structures become energetically more and more unfavorable, as the upper peak in the height distribution of the cations vanishes gradually. Concomitantly, an increasing number of perchlorate molecules approach the gold surface and are predominantly located there; possibly, they act as a space filler on the surface itself, while at the same time the small values of the electrostatic field keep the  $\text{PQP}^+$  molecules near to the surface. In this context, it should be noted that decreasing the electrostatic field is equivalent to decreasing the surface potential; thus, and in combination with the adsorbed perchlorate molecules, a transition from an auto-host–guest to a porous structure is plausible.

Eventually, at zero electric field, the system exclusively gains energy from intramolecular interactions and adsorption on the gold surface. Since the perchlorate molecules are rather spherical in their shape, they can efficiently adsorb onto the gold surface (in an orientation explained above), while the  $\text{PQP}^+$  molecules are able to efficiently stack, especially without a guiding electrostatic field.

**4.2.3. Role of the Perchlorate Anions.** We come back to the above-mentioned volatility of the perchlorate ions: in Figure 6, we present results from yet another evolutionary analysis, i.e., we now fix the positions and orientations of  $\text{PQP}^+$  molecules as well as the extent and the shape of the unit cell of some optimized configuration (as, for instance, depicted in Figure 4a) and vary only the degrees of freedom of the perchlorate anions. Figure 6 shows, for fixed cell geometry and fixed  $\text{PQP}^+$  positions and orientations, four structurally different perchlorate arrangements whose energy ranges within an interval of 38 meV (per  $\text{PQP}^+ - \text{ClO}_4^-$  pair): the fact that we obtain completely different configurations of the perchlorates (with essentially comparable energies) undoubtedly indicates the high mobility of the  $\text{ClO}_4^-$  ions. Changes in the structure, as one proceeds from the left to right, are highlighted by

respective circles (specifying the position of the “moving” perchlorate ion) and related arrows. The  $\text{ClO}_4^-$  molecules exhibit a remarkable freedom in their rotation without (or only marginally) changing the energy of a configuration; this fact has rendered the minimization of the energy very difficult. However, it should also be noted that even translations can be performed without a substantial change in energy.

We note that the analysis of these different structures was achieved using a so-called t-SNE<sup>98</sup> analysis on the leading five principal components of a principal component analysis (PCA)<sup>99</sup> of order parameters of all configurations identified by the evolutionary algorithm; for more detailed information on this rather technical issue, we refer to Figure S5 in Section S3.2 of the Supporting Information.

## 5. CONCLUSIONS AND OUTLOOK

The prediction of the supramolecular ordering of complex molecules at a metal–electrolyte interface using DFT-based ab initio calculations is in view of the expected gigantic computational costs, and despite the availability of petascale computers, still an elusive enterprise. In this contribution, we have proposed a two-stage alternative approach: (i) DFT-based ab initio simulations provide reference data for the energies introduced in a classical model for the molecules involved, where each of their atomic entities are represented by a classical, spherical particle (with respective size, energy parameters, and charges). We modeled the interaction between the atomic entities and the metallic surface by a classical, perfectly conductive, Lennard-Jones-like wall potential; the electrolyte is treated as a homogeneous, dielectric medium.

The interparticle and particle–wall parameters were obtained via the following procedure: considering archetypical configurations (involving pairs of ions and/or ions located close to the surface), DFT energies were fitted by the related energy values of the classical model. (ii) The second step identifies the ordered ground-state configurations of the molecules by minimizing the total energy of the now classical system. This optimization is based on evolutionary algorithms, which are known to operate efficiently and reliably even in high-dimensional search spaces and for rugged energy surfaces.

Our new two-stage strategy overcomes the hitherto prohibitive computational cost of modeling the full system while reproducing the key observations of a well-documented experimental system consisting of disc-shaped  $\text{PQP}^+$  cations and  $\text{ClO}_4^-$  anions: as a function of increasing electric field at

the metal–electrolyte interface, the molecular building blocks are seen to self-organize into an open-porous structure, a self-host–guest pattern, and a stratified bilayer. Future work will focus on verifying the extent of the predictive power of our method toward molecular self-assembly under electrochemical conditions and on strategies to further streamline and reduce the computational cost of our approach, without sacrificing the reliability of the predicted results.

In view of the high computational costs and the conceptual challenges encountered in our investigations, we have pondered the question whether the complexity of the current model (which, as a classical model, is comprehensive in the sense that it contains all atomistic features) could possibly be further reduced, avoiding thereby conceptual and computational bottlenecks. The idea behind this strategy is to develop, starting from the present model, a hierarchy of ever simpler models where, for instance, larger subunits of the molecule (such as aromatic rings) are replaced by disk-shaped units carrying higher electrostatic moments, as schematically visualized in Figure 7. Such a model might provide a first, semiquantitative prediction of the self-assembly of the PQP<sup>+</sup> and of the ClO<sub>4</sub><sup>-</sup> ions at considerably reduced costs and might help in prescreening possibly promising portions of the huge parameter space for subsequent investigations of the full model. Efforts in this direction are currently pursued.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.9b01251>.

Introduction, system and its representations, identifications of self-assembly scenarios, and results (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Benedikt Hartl** – Institute for Theoretical Physics and Center for Computational Materials Science (CMS), TU Wien, 1040 Wien, Austria; [orcid.org/0000-0001-7787-4839](https://orcid.org/0000-0001-7787-4839); Email: [benedikt.hartl@tuwien.ac.at](mailto:benedikt.hartl@tuwien.ac.at)

### Authors

**Shubham Sharma** – FIT Freiburg Centre for Interactive Materials and Bioinspired Technologies, 79110 Freiburg, Germany

**Oliver Brügger** – FIT Freiburg Centre for Interactive Materials and Bioinspired Technologies, 79110 Freiburg, Germany

**Stijn F. L. Mertens** – Department of Chemistry, Lancaster University, Lancaster LA1 4YB, United Kingdom; Institute of Applied Physics, TU Wien, 1040 Wien, Austria; [orcid.org/0000-0002-5715-0486](https://orcid.org/0000-0002-5715-0486)

**Michael Walter** – FIT Freiburg Centre for Interactive Materials and Bioinspired Technologies, 79110 Freiburg, Germany; Cluster of Excellence livMatS@FIT—Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, D-79110 Freiburg, Germany; MikroTribologie Centrum  $\mu$ TC, Fraunhofer IWM, 79108 Freiburg, Germany; [orcid.org/0000-0001-6679-2491](https://orcid.org/0000-0001-6679-2491)

**Gerhard Kahl** – Institute for Theoretical Physics and Center for Computational Materials Science (CMS), TU Wien, 1040 Wien, Austria; [orcid.org/0000-0002-4375-4684](https://orcid.org/0000-0002-4375-4684)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.9b01251>

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

B.H. acknowledges a DOC Fellowship of the Austrian Academy of Sciences; further, he gratefully acknowledges helpful discussions with Moritz Antlanger (Wien), Benjamin Rotenberg (Paris), and Martial Mazars (Paris-Sud) on several issues of this project and thanks the Freiburg Centre for Interactive Materials and Bioinspired Technologies (Universität Freiburg) for the kind hospitality, where part of the work has been carried out. B.H. also acknowledges the Christiana Hörbiger prize for covering travel expenses to Paris. B.H. and G.K. acknowledge financial support by E-CAM, an e-infrastructure center of excellence for software, training, and consultancy in simulation and modeling funded by the EU (Project no. 676531). The computational results presented have been achieved (in part) using the Vienna Scientific Cluster (VSC). S.S. and M.W. acknowledge funding from Deutsche Forschungsgemeinschaft (WA 1687/10-1) and the computing time granted by the John von Neumann Institute for Computing (NIC) within project HFR08 as well as the computational resource bwUni-Cluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC. M.W. is thankful for an enlightening discussion with Johannes Fiedler and Stefan Buhmann. S.F.L.M. and G.K. gratefully acknowledge financial support within the TU Wien funded Doctoral College BIOINTERFACE. S.F.L.M. further acknowledges support from the Austrian Science Fund (FWF, “Boron Nitride Nanomesh for Actuated Self-Assembly” (I3256-N36)) and Energy Lancaster.

## ■ REFERENCES

- (1) Cui, K.; Dorner, I.; Mertens, S. F. L. Interfacial supramolecular electrochemistry. *Curr. Opin. Electrochem.* **2018**, *8*, 156–163.
- (2) Mertens, S. F. L. *Encyclopedia of Interfacial Chemistry: Surface Science and Electrochemistry*, 1st ed.; Wandelt, K., Ed.; Elsevier, 2018; Vol. 4, pp 13–23.
- (3) Cui, K.; Mali, K. S.; Ivasenko, O.; Wu, D.; Feng, X.; Walter, M.; Müllen, K.; De Feyter, S.; Mertens, S. F. L. Squeezing, then stacking: From breathing pores to three-dimensional ionic self-assembly under electrochemical control. *Angew. Chem., Int. Ed.* **2014**, *53*, 12951–12954.
- (4) Cui, K.; Ivasenko, O.; Mali, K. S.; Wu, D.; Feng, X.; Müllen, K.; De Feyter, S.; Mertens, S. F. L. Potential-driven molecular tiling of a charged polycyclic aromatic compound. *Chem. Commun.* **2014**, *50*, 10376–10378.
- (5) Cui, K.; Mali, K. S.; Wu, D.; Feng, X.; Müllen, K.; Walter, M.; De Feyter, S.; Mertens, S. F. L. Reversible Anion-Driven Switching of an Organic 2D Crystal at a Solid-Liquid Interface. *Small* **2017**, *13*, No. 1702379.
- (6) Cui, K.; Mali, K. S.; Wu, D.; Feng, X.; Müllen, K.; Walter, M.; De Feyter, S.; Mertens, S. F. L. Ambient bistable single dipole switching in a molecular monolayer. *Angew. Chem., Int. Ed.* **2020**, DOI: [10.1002/anie.202004016](https://doi.org/10.1002/anie.202004016).
- (7) Stillinger, F. H. Exponential multiplicity of inherent structures. *Phys. Rev. E* **1999**, *59*, 48–51.
- (8) Scherbela, M.; Hörmann, L.; Jeindl, A.; Obersteiner, V.; Hofmann, O. T. Charting the energy landscape of metal/organic interfaces via machine learning. *Phys. Rev. Mater.* **2018**, *2*, No. 043803.
- (9) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129–4137.

- (10) Lorenz, S.; Groß, A.; Scheffler, M. Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. *Chem. Phys. Lett.* **2004**, *395*, 210–215.
- (11) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, No. 146401.
- (12) Behler, J.; Lorenz, S.; Reuter, K. Representing molecule-surface interactions with symmetry-adapted neural networks. *J. Chem. Phys.* **2007**, *127*, No. 014705.
- (13) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, No. 170901.
- (14) Schütt, K. T.; Kindermans, P. J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K. R. In *SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions*, Advances in Neural Information Processing Systems, 31st Annual Conference on Neural Information Processing Systems, NIPS 2017, Long Beach, 2017; pp 992–1002.
- (15) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
- (16) Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press, 2000.
- (17) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, No. 136403.
- (18) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, No. 058301.
- (19) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, No. 184115.
- (20) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.
- (21) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.
- (22) Obersteiner, V.; Scherbela, M.; Hörmann, L.; Wegner, D.; Hofmann, O. T. Structure Prediction for Surface-Induced Phases of Organic Monolayers: Overcoming the Combinatorial Bottleneck. *Nano Lett.* **2017**, *17*, 4453–4460.
- (23) Hörmann, L.; Jeindl, A.; Egger, A. T.; Scherbela, M.; Hofmann, O. T. SAMPLE: Surface structure search enabled by coarse graining and statistical learning. *Comput. Phys. Commun.* **2019**, *244*, 143–155.
- (24) Kaatze, U. Complex permittivity of water as a function of frequency and temperature. *J. Chem. Eng. Data* **1989**, *34*, 371–374.
- (25) Hamelin, J.; Mehl, J. B.; Moldover, M. R. The Static Dielectric Constant of Liquid Water Between 274 and 418 K Near the Saturated Vapor Pressure. *Int. J. Thermophys.* **1998**, *19*, 1359–1380.
- (26) Vargaftik, N. B.; Volkov, B. N.; Voljak, L. D. International Tables of the Surface Tension of Water. *J. Phys. Chem. Ref. Data* **1983**, *12*, 817–820.
- (27) Hasted, J. B.; Ritson, D. M.; Collie, C. H. Dielectric Properties of Aqueous Ionic Solutions. Parts I and II. *J. Chem. Phys.* **1948**, *16*, 1–21.
- (28) Gavish, N.; Promislow, K. Dependence of the dielectric constant of electrolyte solutions on ionic concentration: A microfield approach. *Phys. Rev. E* **2016**, *94*, No. 012611.
- (29) Mollerup, J. M.; Breil, M. P. Modeling the permittivity of electrolyte solutions. *AIChE J.* **2015**, *61*, 2854–2860.
- (30) Fumagalli, L.; Esfandiari, A.; Fabregas, R.; Hu, S.; Ares, P.; Janardanan, A.; Yang, Q.; Radha, B.; Taniguchi, T.; Watanabe, K.; Gomila, G.; Novoselov, K. S.; Geim, A. K. Anomalously low dielectric constant of confined water. *Science* **2018**, *360*, 1339–1342.
- (31) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **2005**, *71*, No. 035109.
- (32) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dulak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsarolis, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter* **2010**, *22*, No. 253202.
- (33) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dulak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, No. 273002.
- (34) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979.
- (35) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (36) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, No. 073005.
- (37) Held, A.; Walter, M. Simplified continuum solvent model with a smooth cavity based on volumetric data. *J. Chem. Phys.* **2014**, *141*, No. 174108.
- (38) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **1976**, *13*, 5188–5192.
- (39) Eischenschitz, R.; London, F. Über das Verhältnis der van der Waalschen Kräfte zu den homöopolaren Bindungskräften. *Z. Phys.* **1930**, *60*, 491–527.
- (40) Fiedler, J.; Thiyam, P.; Kurumbail, A.; Burger, F. A.; Walter, M.; Persson, C.; Brevik, I.; Parsons, D. F.; Boström, M.; Buhmann, S. Y. Effective Polarizability Models. *J. Phys. Chem. A* **2017**, *121*, 9742–9751.
- (41) Beneduci, A. Which is the effective time scale of the fast Debye relaxation process in water. *J. Mol. Liq.* **2008**, *138*, 55–60.
- (42) Bukowski, R.; Szalewicz, K. Complete ab initio three-body nonadditive potential in Monte Carlo simulations of vapor-liquid equilibria and pure phases of argon. *J. Chem. Phys.* **2001**, *114*, 9518–9531.
- (43) McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A Universal Approach. *J. Phys. Chem. B* **2014**, *118*, 8042–8053.
- (44) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, No. 236402.
- (45) Perdew, J. P. Energetics of charged metallic particles: From atom to bulk solid. *Phys. Rev. B* **1988**, *37*, 6175–6180.
- (46) Serena, P. A.; Soler, J. M.; García, N. Self-consistent image potential in a metal surface. *Phys. Rev. B* **1986**, *34*, 6767–6769.
- (47) Lang, N. D.; Kohn, W. Theory of Metal Surfaces: Induced Surface Charge and Image Potential. *Phys. Rev. B* **1973**, *7*, 3541–3550.
- (48) Finnis, M. W.; Kaschner, R.; Kruse, C.; Furthmüller, J.; Scheffler, M. The interaction of a point charge with a metal surface: theory and calculations for (111), (100) and (110) aluminium surfaces. *J. Phys.: Condens. Matter* **1995**, *7*, 2001.
- (49) Held, A.; Moseler, M.; Walter, M. Charging properties of gold clusters in different environments. *Phys. Rev. B* **2013**, *87*, No. 045411.
- (50) Chakrabarti, D.; Kusumaatmaja, H.; Rühle, V.; Wales, D. J. Exploring energy landscapes: From molecular to mesoscopic systems. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5014–5025.
- (51) Chakrabarti, D.; Wales, D. J. Simulations of rigid bodies in an angle-axis framework. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1970–1976.
- (52) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids: With Applications to Soft Matter*; Academic Press, 2013.

- (53) Mie, G. Zur kinetischen Theorie der einatomigen Körper. *Ann. Phys.* **1903**, *316*, 657–697.
- (54) Magda, J. J.; Tirrell, M.; Davis, H. T. Molecular dynamics of narrow, liquid-filled pores. *J. Chem. Phys.* **1985**, *83*, 1888–1901.
- (55) Jacob, T.; Sabo, L. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, 2015.
- (56) Plimpton, S. Fast Parallel Algorithms for Short – Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (57) Mazars, M. Long ranged interactions in computer simulations and for quasi-2D systems. *Phys. Rep.* **2011**, *500*, 43–116.
- (58) Smith, E. R. Electrostatic Energy in Ionic Crystals. *Proc. R. Soc. London, Ser. A* **1981**, *375*, 475–505.
- (59) Yeh, I. C.; Berkowitz, M. L. Ewald summation for systems with slab geometry. *J. Chem. Phys.* **1999**, *111*, 3155.
- (60) Bader, R. F. W. Atoms in Molecules. *Acc. Chem. Res.* **1985**, *18*, 9–15.
- (61) Hartl, B.; Sharma, S.; Walter, M.; Mertens, S. F. L.; Kahl, G. Phase Behaviour of Supramolecular Ordering of Complex Molecules under Electrochemical Conditions, to be submitted for publication, 2020.
- (62) Gottwald, D.; Kahl, G.; Likos, C. N. Predicting equilibrium structures in freezing processes Predicting equilibrium structures in freezing processes. *J. Chem. Phys.* **2005**, *122*, No. 204503.
- (63) Fornleitner, J.; Lo Verso, F.; Kahl, G.; Likos, C. N. Genetic algorithms predict formation of exotic ordered configurations for two-component dipolar monolayers. *Soft Matter* **2008**, *4*, 480–484.
- (64) Fornleitner, J.; Lo Verso, F.; Kahl, G.; Likos, C. N. Ordering in Two-Dimensional Dipolar Mixtures. *Langmuir* **2009**, *25*, 7836–7846.
- (65) Doppelbauer, G.; Bianchi, E.; Kahl, G. Self-assembly scenarios of patchy colloidal particles in two dimensions. *J. Phys.: Condens. Matter* **2010**, *22*, No. 104105.
- (66) Fornleitner, J.; Kahl, G. Pattern formation in two-dimensional square-shoulder systems. *J. Phys.: Condens. Matter* **2010**, *22*, No. 104118.
- (67) Doppelbauer, G.; Noya, E. G.; Bianchi, E.; Kahl, G. Self-assembly scenarios of patchy colloidal particles. *Soft Matter* **2012**, *8*, 7768–7772.
- (68) Antlanger, M.; Kahl, G.; Mazars, M.; Šamaj, L.; Trizac, E. Rich Polymorphic Behavior of Wigner Bilayers. *Phys. Rev. Lett.* **2016**, *117*, No. 118002.
- (69) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors. SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, 261.
- (70) Kraft, D. *A Software Package for Sequential Quadratic Programming*, DFVLR-FB-88-28; DFVLR, 1988.
- (71) Dalcín, L. D.; Paz, R. R.; Storti, M. MPI for Python. *J. Parallel Distrib. Comput.* **2005**, *65*, 1108–1115.
- (72) Dalcín, L. D.; Paz, R. R.; Storti, M.; Elía, J. D. MPI for Python: Performance improvements and MPI-2 extensions. *J. Parallel Distrib. Comput.* **2008**, *68*, 655–662.
- (73) Dalcín, L. D.; Paz, R. R.; Kler, P. A.; Cosimo, A. Parallel distributed computing using Python. *Adv. Water Resour.* **2011**, *34*, 1124–1139.
- (74) Mascato, P. *On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms*; California Institute of Technology, 1989; p 67.
- (75) Gottwald, D. Genetic Algorithms in Condensed Matter Theory. Ph.D. Thesis, TU Wien, 2005.
- (76) Doppelbauer, G. Ordered Equilibrium Structures of Patchy Particle Systems. Ph.D. Thesis, TU Wien, 2012.
- (77) Fornleitner, J.; Kahl, G. Lane formation vs. cluster formation in two-dimensional square-shoulder systems A genetic algorithm approach Lane formation vs. cluster formation in two-dimensional square-shoulder systems A genetic algorithm approach. *EPL* **2008**, *82*, No. 18001.
- (78) Antlanger, M. Ordered Equilibrium Structures in Systems with Long-Range Interactions. Ph.D. Thesis, TU Wien, 2015.
- (79) Darwin, C. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*; John Murray: London, 1859.
- (80) Hartke, B. Global geometry optimization of clusters using genetic algorithms. *J. Phys. Chem. A* **1993**, *97*, 9973–9976.
- (81) Hartke, B. Global cluster geometry optimization by a phenotype algorithm with Niches: Location of elusive minima, and low-order scaling with cluster size. *J. Comput. Chem.* **1999**, *16*, 1752–1759.
- (82) Pannetier, J.; Bassas-Alsina, J.; Rodriguez-Carvajal, J.; Caignaert, V. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* **1990**, *346*, 343–345.
- (83) Schön, J. C.; Jansen, M. In *Modern Methods of Crystal Structure Prediction*; Oganov, A. R., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA, 2015; pp 67–105.
- (84) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- (85) Panosetti, C.; Krautgasser, K.; Palagin, D.; Reuter, K.; Maurer, R. J. Global Materials Structure Search with Chemically Motivated Coordinates. *Nano Lett.* **2015**, *15*, 8044–8048.
- (86) Krautgasser, K.; Panosetti, C.; Palagin, D.; Reuter, K.; Maurer, R. J. Global structure search for molecules on surfaces: Efficient sampling with curvilinear coordinates. *J. Chem. Phys.* **2016**, *145*, No. 084117.
- (87) Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.
- (88) Goedecker, S. In *Modern Methods of Crystal Structure Prediction*; Oganov, A. R., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA, 2015; pp 131–145.
- (89) Abraham, N. L.; Probert, M. I. J. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Phys. Rev. B* **2006**, *73*, No. 224104.
- (90) Deaven, D. M.; Ho, K. M. Molecular Geometry Optimization with a Genetic Algorithm. *Phys. Rev. Lett.* **1995**, *75*, 288–291.
- (91) Lyakhov, A. O.; Oganov, A. R.; Valle, M. In *Modern Methods of Crystal Structure Prediction*; Oganov, A. R., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA, 2015; pp 147–180.
- (92) Supady, A.; Blum, V.; Baldauf, C. First-Principles Molecular Structure Search with a Genetic Algorithm. *J. Chem. Inf. Model.* **2015**, *55*, 2338–2348.
- (93) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Zidek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710.
- (94) Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; Chen, Y.; Lillicrap, T.; Hui, F.; Sifre, L.; van den Driessche, G.; Graepel, T.; Hassabis, D. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359.
- (95) Hanke, F. Sensitivity analysis and uncertainty calculation for dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1424–1430.
- (96) Berland, K.; Cooper, V. R.; Lee, K.; Schröder, E.; Thonhauser, T.; Hyldgaard, P.; Lundqvist, B. I. van der Waals forces in density functional theory: a review of the vdW-DF method. *Rep. Prog. Phys.* **2015**, *78*, No. 066501.
- (97) Stöhr, M.; Voorhis, T. V.; Tkatchenko, A. Theory and practice of modeling van der Waals interactions in electronic-structure calculations. *Chem. Soc. Rev.* **2019**, *48*, 4118–4154.
- (98) Maaten, L. V. D.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.



(99) Jolliffe, I. *Principal Component Analysis*; Springer Verlag: New York, 2002.