

On the Physical Consistency of an Open Quantum Region with a Classical Reservoir in Molecular Simulation

Sara Panahian Jand,* Thomas D. Kühne, and Luigi Delle Site

The possibility of treating a molecular liquid in an open region at *ab initio* electronic resolution embedded in a classical reservoir of energy and particles, is investigated. Because of its challenging properties and its relevance in many field of current research, the system chosen as prototype of molecular liquid is water at room conditions. A numerical protocol based on the mathematical model of open particle system is applied and the results are compared with results of a full *ab initio* simulation of reference. The key conclusion is that one can claim the existence of a mandatory minimal size of the quantum region in which structural and electronic properties reproduce those of reference and, at the same time, the exchange of molecules with the environment takes place as expected. This work provides a proof of concept about the possibility to systematically define a physically well founded open quantum system embedded in a classical environment. In turn, the proof of concept is a key information for the design of numerically efficient algorithms for *ab initio* molecular dynamics simulations of open systems.

models have been implemented in the Adaptive Resolution numerical scheme (AdResS)^[7] and so far it describes classical systems, or quantum systems that can be mapped onto classical models, embedded in a simplified reservoir.^[8] The introduction of the electronic degrees of freedom adds a further constrain for assuring physical consistency, that is, the electronic chemical potential of the quantum open system embedded in the simplified molecular reservoir should be consistent with the electronic chemical potential of the equivalent full *ab initio* simulation.^[9,10] It must be underlined that current numerical approaches that treat open molecular systems with electronic degrees of freedom, known as A-QM/MM (Adaptive Quantum Mechanics/Classical Mechanics),^[11] often focus on the electronic structure aspects but do not additionally consider the aforementioned statistical mechanics principles. They are

constructed on frameworks where the open system is mimicked by the interpolation of atomic forces determined in separate, closed system, calculations each with a different number of quantum molecules. Although the numerical implementation may be straightforward and the corresponding electronic properties sufficiently accurate, the associated risk to sample artificial ensembles not compatible with first principles of statistical mechanics (see e.g. refs. [10, 12]) should be taken into account. The protocol to check the physical consistency of an open quantum system proposed in this work could be actually applied to strengthen the validity of such approaches as well. In this context the aim of this work is to apply the theoretical framework of refs. [3, 4, 9, 10] and its associated numerical protocol to an *ab initio* Molecular Dynamics (MD) simulation and show the possibility of identifying an open quantum molecular system embedded in a simplified reservoir; more specifically, one would like to define a physically consistent open quantum region in a large reservoir of classical molecules (see **Figure 1**). Due to the Born-Oppenheimer approximation the nuclei are classical particles and thus their exchange with the reservoir is regulated by the rules of a classical system as in refs.^[3,7]. Instead, the electrons must follow the rule of an open quantum systems thus one must require (necessary condition) that the electronic chemical potential is consistent with the electronic chemical potential that the full system would have if it was entirely at quantum resolution^[9,10]; of course, as a consequence, the molecular electronic properties of the quantum region should reproduce those of an *ab initio* simulation of reference.

1. Introduction

The theory and simulation of open molecular systems that exchange energy and particles with a simplified molecular reservoir has made a substantial progress in the last decade (see e.g., refs. [1, 2] and references therein). In particular the emergence of physico-mathematical models that provide clear rules for the numerical realization of the open system concept^[3,4] has significantly risen our capability of simulating a broad spectrum of physical situations in and out of equilibrium.^[5,6] One of the

S. Panahian Jand, L. Delle Site
Freie Universität Berlin
Institute of Mathematics
Arnimallee 6, 14195 Berlin, Germany
E-mail: sara.jand@fu-berlin.de

T. D. Kühne
CASUS - Center for Advanced Systems Understanding
Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR)
Untermarkt 20, D-02826 Görlitz, Germany

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adts.202400833>

© 2024 The Author(s). Advanced Theory and Simulations published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

DOI: 10.1002/adts.202400833

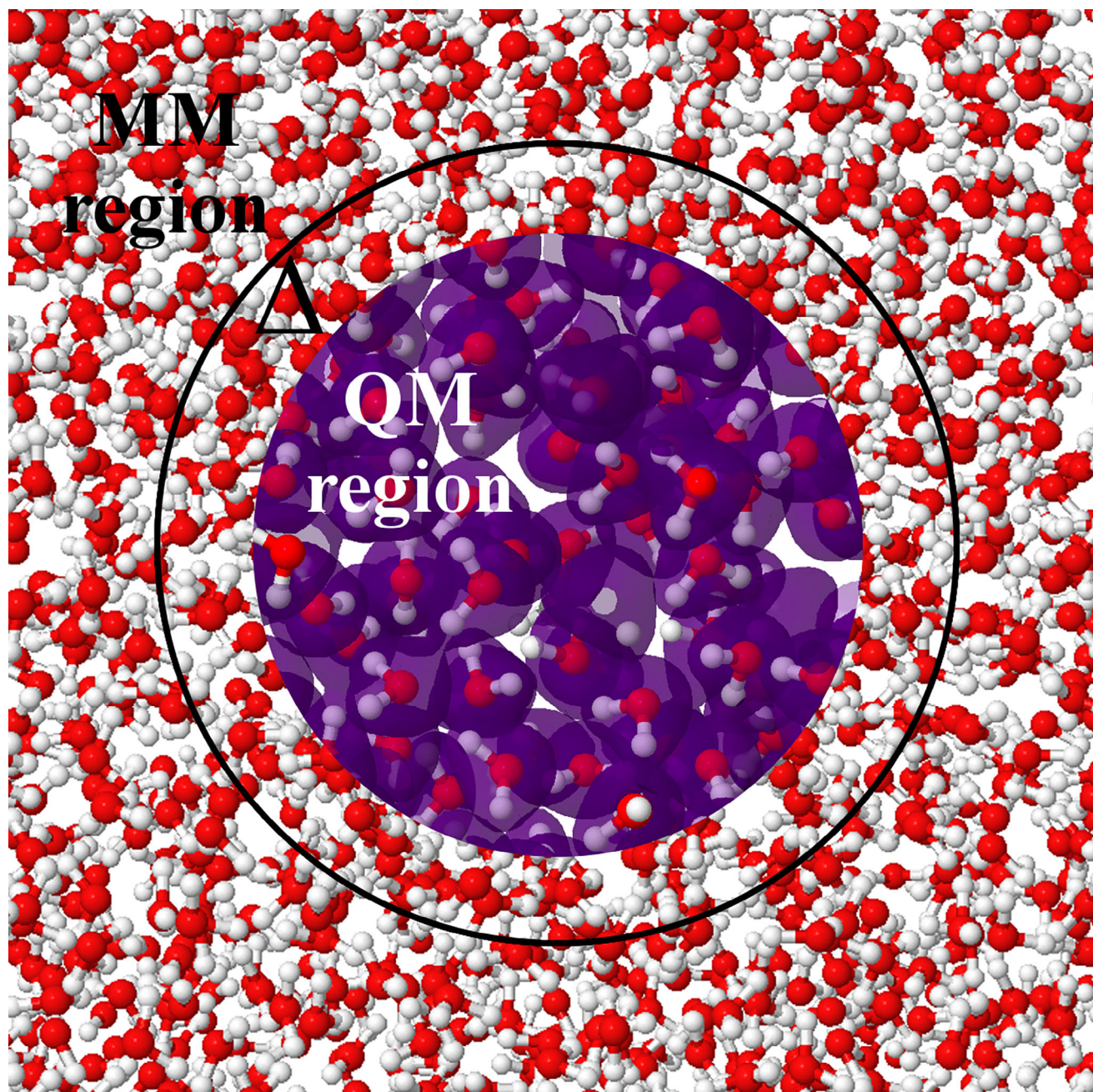


Figure 1. Pictorial representation of the open QM/MM system in the AdResS fashion. At the center is the QM region with radius of 14.0\AA . Around such region, in the MM domain, there is the Δ region of 9.0\AA , which usually in the AdResS set up is the region where the thermodynamic force is applied so that the corresponding average particle density is kept within a certain threshold w.r.t. the target density. Such force assures also the consistency of the chemical potential of the QM region with the reference full ab initio case. Outside there is the large MM environment.

For the specific system studied here (liquid water) we actually conclude, from a numerical estimate, that the constraint on the electronic chemical potential is not numerically relevant because its associated energy contribution is negligible compared to the total energy, independently of the size of the quantum system.

The application of the protocol defines the size of the open quantum region, which is large enough to sustain the key features of the liquid at electronic level and assures a macroscopic thermodynamic state consistent with the reference full ab initio

case; the thermodynamic consistency is provided by a compatible classical environment. The size of the quantum region found in this work contains about 440 molecules (in average) and such a size implies a relative inexpensive simulation. It must be underlined that one expects that the larger the QM region the better agreement will be achieved between the QM/MM and full ab initio calculations; one also expects that there is a minimum collection of the QM atoms that is required to capture the essential electronic-structure features of the region of interest. Thus,

it will not be surprising that a larger QM region reproduces the structural properties of the full ab initio system than a smaller QM region does. The key question is about the criteria to use to identify the minimal region required to reproduce not only structural and electronic properties of the ab initio system but also its fluctuations so that one achieves the physical consistency with respect to the statistical mechanics ensemble of an open system that exchanges particles and energy with the exterior.

As anticipated above, in this work we simulate liquid water at room conditions; this choice was made because water is the universal solvent and thus water-related systems are ubiquitous across the whole domain of chemical physics, thus liquid water is a relevant prototype system for future applications. For example, in studies of solvation of molecules with an open quantum region determined as in this work, one would have a sizable computational gain compared to the large systems usually required in full ab initio simulations. In simple words, one can define an optimal quantum region of solvation, embedded in a classical reservoir, and be assured to capture all the particle fluctuations of a large system as in a full ab initio calculation. In a broader perspective, this study can be considered a pilot work for the implementation of the open system schemes to generic systems that may be more complex than the example reported here. It must be added that within the same scheme proposed here one can, in perspective, treat both electrons and nuclei quantum-mechanically; these latter can be treated through the path integral representation. In fact, as mentioned above, the scheme has already successfully treated atoms represented in a path integral formalism^[8]; the extension to nuclei in presence of electrons is conceptually equivalent to the situation treated in this paper. The only difference is that the nuclei instead of classical localized spheres would be represented by flexible (classical) polymer rings that obey the classical rules of refs. [3, 4] regarding their coupling to the external reservoir.

2. System Studied and Application of the Protocol of Physical Consistency for the Quantum Region as Open System

We consider liquid water at room temperature. The system is defined by a cubic box with a central spherical region (QM region) where molecules are treated at quantum ab initio resolution and the rest (environment of the quantum region, named here MM region) is treated at classical level with a flexible water model (see Figure 1 for an illustrative explanation). Molecules can freely move from one region to the other and at each time step they are accordingly reclassified as classical or quantum according to the corresponding updated position in space. Following the mathematical model of open system implemented in the Adaptive Resolution approach (AdResS),^[3,4] the difference of resolution, due to the statistical and thermodynamic incompatibility of the models at two resolutions should produce a large difference of the average particle number density at the interface between the QM and the MM region, compared to the reference case. Here for difference is meant that the average particle number density, within the typical fluctuations found in the calculations of reference, systematically deviates from the target showing an overall different density distribution, e.g., non uniform as the reference of this case. If such a deviation takes place, one can then

define a transition region Δ in the MM region adjacent to the QM region and impose in such region the so-called thermodynamic force $F_{th}(x)$.^[13] The thermodynamic force is calculated self-consistently in the equilibration run: A first guess is taken, e.g., $F_{th}^{(0)}(x) = 0$, then at step k one has $F_{th}^{(k+1)}(x) = F_{th}^{(k)}(x) - c \nabla \rho_k(x)$, where $\rho_k(x)$ is the average molecular number density calculated in the simulation using $F_{th}^{(k)}(x)$; $c > 0$ is a coefficient that calibrates the convergence. The iteration is terminated when the deviation of $\rho_k(x)$ from a reference density profile is within a wished tolerance. Once $F_{th}(x)$ has been determined, it remains unchanged without any need for recalibration. It has been shown that this force balances the macroscopic chemical potential between the two regions and assures equilibrium (see e.g., ref. [14]). If the two different molecular representations are thermodynamically compatible, that is, if they are interfaced the thermodynamic state point remains unaltered, there may not be relevant differences for the average particle density along the radial direction between the QM and the MM region and thus the thermodynamic force is not needed. According to the general mathematical model of open system, the constraint on the particle density is a strong and yet only necessary condition for defining the QM region as a physically meaningful open system. From the statistical mechanics point of view, once the criterion on the density is satisfied, one should verify that the QM region delivers the same results obtained for the equivalent subregion in a large full (ab initio) simulation. This check is done following a protocol based on a posteriori checks on quantities of structural and statistical relevance; such quantities are the radial distribution functions and particle number distribution.

We do not treat dynamic properties here; they require long and expensive simulations that can be planned, at this stage, for future projects. However, it must be underlined that the current study about statistical and thermodynamic consistency represents a mandatory verification of the definition of a physically meaningful QM open region. While structural and thermodynamic properties are independent of the specific dynamics, consistency of the dynamics would require, as mandatory condition, structural consistency.

For this work we report that, for a QM region with 440 molecules and for a QM region with 192 molecules embedded in the classical environment of flexible TIP3P water model, the average particle number density fluctuates around the target value similarly to the result of the reference full ab initio simulations and thus there is no need of calculating and applying the thermodynamic force. However, while for the 440 molecules system we are able to reproduce the results of the reference full ab initio simulation with satisfactory accuracy, for the 192 molecules system an evident disagreement can be found for several properties; the disagreement is such that does not allow to qualify the QM region corresponding to 192 molecules as a physically consistent open system.

Regarding the specific quantities considered we start, as in classical AdResS, from the Oxygen–Oxygen, and Oxygen–Hydrogen radial distribution functions; these are the keys to check whether the QM region reproduces the liquid structure and hydrogen-bonding features of the reference full ab initio simulation. If the agreement is satisfactory one can proceed with checking that the exchange of molecules between the QM and the MM region occurs as in the reference case. To this aim

one calculates the probability distribution of the number of particles in the QM region, $P(N)$; the relevant feature is that the $P(N)$ of the open system and the $P(N)$ of reference must have a compatible overall shape. Instead from the quantitative point of view, they may slightly disagree along the axis of N (we have chosen 5% as acceptable threshold like in classical AdResS). If the $P(N)$ gives satisfactory results, then, for the quantum case with electrons, one needs to go a step beyond the classical rules of open system as used in the standard AdResS scheme. In fact in the QM region we deal with molecules that explicitly carry a distribution of electronic charge. The electronic structure properties are thus very relevant, and in a physically meaningful QM open system one should be able to reproduce also electronic structure features of the full ab initio simulation of reference.

In principle, as concluded from a statistical mechanics analysis,^[9,10] the strict way to proceed is to perform electronic structure calculations at constant electronic chemical potential, rather than the standard energy minimization, for the QM region.^[9,10] The electronic chemical potential should be fixed at the value found for the reference full ab initio calculations. In practice this approach implies non trivial technical problems of implementation in current (standard) QM/MM codes because the electronic charge may not be conserved.^[1] Future numerical developments should definitely go along such direction, with possible solution for the excess or lack of electronic charge.^[9] However, for molecular liquids like water, where there is no extreme case of electronic polarization, molecules display quite localized electronic orbitals around the molecular skeleton of the nuclei, thus one would expect that the statistical equilibrium of the (classical) nuclei with the reservoir drives automatically the statistical equilibrium of the electronic degrees of freedom. In order to justify the claim above it would be sufficient to prove that the constraint of the electronic chemical potential does not play a relevant role in this case.

The concept above can be formalized in a rigorous manner on the basis of recent results obtained by one of us together with A. Djurdjevac.^[15] In ref. [15] it has been rigorously derived an effective Hamiltonian of a quantum open system that exchanges particles and energy with a reservoir: $\hat{H}_{QM}^{eff} = \hat{H}_N - \mu \hat{N}$; where \hat{H}_N is the Hamiltonian of the open quantum system for an N -particle realization, μ the chemical potential of the quantum particles and \hat{N} the number operator that counts the number of particles in a given state. If one has the condition: $||\hat{H}_N|| \gg \mu ||\hat{N}||; \forall N$, then it automatically implies that \hat{H}_N alone, without the implication of μ , is sufficient to determine the statistical mechanics properties of the open particle system. In this work we have verified that the energy associated to the electronic chemical potential, $\mu ||\hat{N}||$, is negligible compared to the total energy of the system independently from the size of the system (for a statistically relevant number of molecules, see Table 1). As a consequence one would expect that the constraint on the electronic chemical potential does not play a key role in assuring consistency for the electronic structure properties of the open QM region. In any case, a posteriori checks of consistency for molecular electronic properties would definitively prove whether or not the estimate above holds.

In order to numerically verify that the electronic structure properties are preserved, we calculate the Wannier centers^[16,17] in each molecule and calculate the Wannier–Wannier and the Wannier–Oxygen radial distribution functions. A Wannier cen-

Table 1. The relevance of the energy associated to the electronic chemical potential, $\mu_e N_e$, with N_e the total number of electrons, with respect to the total energy of the system E_{tot} , as a function of the total size of the (full ab initio) system, expressed in number of molecules, N_m .

N_m	$\mu_e N_e / E_{tot}$
208	0.049
366	0.045
608	0.043
843	0.041
1112	0.039
1874	0.037
2454	0.035

ter corresponds to a maximally localized electron density of a molecule in a point of space, thus it carries the very essence of the electron distribution in liquids like water; for water each Wannier center localizes an electron pair, thus given the eight electrons of valence one would have four Wannier centers per water molecule.^[18]

The agreement of the Wannier–Wannier and the Wannier–Oxygen radial distribution functions of the QM region with those of reference implies that the QM region is indeed a physically meaningful open system embedded in a reservoir of classical particles. In fact the consistency of the liquid structure is expressed by the agreement of the Oxygen–Oxygen and Oxygen–Hydrogen radial distribution function of the QM region with the corresponding reference function, the consistency of the electronic charge distribution is expressed by the agreement of the Wannier–Wannier and the Wannier–Oxygen radial distribution function of the QM region with the corresponding reference function and the proper exchange of molecules between the QM and the MM region is assured by agreement of the $P(N)$ of the QM region with the $P(N)$ of reference.

3. Technical Details

The full ab initio molecular dynamics (AIMD) simulation of reference has been performed with a time step of 1fs for 10ps using DFT in CP2K code.^[19] The Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional^[20] has been employed, as well as Goedecker–Teter–Hutter pseudopotentials^[21] and a DZVP basis set^[22] with a wave function cutoff of 280 Ry. In addition, we have included D3 van der Waals correction.^[23] The Nose–Hoover thermostat^[24] was employed to assure thermal equilibrium at the target temperature of 300 K for a mass density of $1020 \frac{\text{kg}}{\text{m}^3}$. The size of the simulations box is $41.5156 \times 41.5156 \times 41.5156 \text{ \AA}$ containing 2454 water molecules. To calculate the electronic chemical potential, first the geometry optimization has been performed for the systems with 208, 366, 608, 843, 1112, and 1874 water molecules. The details of these computations are as above, i.e., using PBE exchange–correlation functional with D3 van der Waals correction and DZVP basis set with a wave function cutoff of 280 Ry. Then, the linear scaling method with curvy steps algorithm^[25] at 10^{-15} filtering threshold for the matrix multiplication has been applied for the energy minimization. The purification method is trace resetting algorithm (TRS4),^[26] which

calculates the electronic chemical potential without need for an initial estimation of μ_e . For the open QM/MM, we have used the AdConv-QM/MM^[27] implemented in CP2K with the same computational setting as the full ab initio simulations for the QM region and the flexible TIP3P (fTIP3P) potential^[28] for modeling water molecules in the MM region. The cutoff radius for Lennard-Jones potential is 9.0Å. The Molecular dynamics simulations of the open QM/MM system have been carried with a time step of 1fs for 10 ps at 300 K. The adaptive Langevin thermostat^[29] has been used to support adaptive QM/MM. The size of the whole simulation box is $50.3040 \times 50.3040 \times 50.3040\text{\AA}$ with 4320 water molecules, the larger QM region has the radius of 14.5Å, containing in average around 440 water molecules while the smaller QM region has a radius of 11.0Å with about 192 molecules. The localization methods of Crazy rotations and Jacobi rotations have been used to calculate the Wannier function centers. The convergence criterion of the localization method has been set to 10^{-5} . Given the cost of the simulation, the size of the full ab initio simulation is smaller than the size of the QM/MM simulation, however it is sufficient to reproduce the key features of the liquid. Instead the QM/MM simulation needs a bigger size to allow the classical region to behave as a proper thermodynamic reservoir.

It must be also clarified that of the code of ref. [27] we adopt only the scheme of reclassification of molecules in QM and MM at each step, but discharged the corrections at the interface of QM and MM done in that work. In fact our approach is based on a different procedure to assure the physical consistency of the results.

Finally, the particle density is calculated along the radial direction, $\rho(r)$. The binning consists of spherical shells whose thickness is such that the volume is the same for all of them so that for each point the statistics is done over the same number of molecules. The average distance of each shell from the center of the QM region defines r . Data are taken from independent configurations over a 10 ps trajectory. The length of the trajectory is long enough to assure that the relevant physics of the system is satisfactorily reproduced (see e.g., ref. [18]).

It must be taken into account that, due to the length of the simulation that combines with the low convergence rate in thinner shells, one cannot expect a perfectly converged flat density profile as, for example, in classical simulations. Even in these latter, with averages done over trajectories up to 1–2 ns and over an entire box with equally spaced bins, one still finds fluctuations in the density profile of the order of 5–7% (see e.g., ref. [30]). In our case we can see fluctuations of the density profile in the QM region of the order of 15% at worst (and about 18% at the very critical quantum-classical interface for the QM/MM system), that is not particularly large, which suggests that, despite the density has not converged to a perfect flat profile, there is not a physical anomaly in the simulation.

Moreover, the relevant message in this case is that one cannot see any systematic drift of the density, w.r.t. its average value and in addition the results satisfy a closure condition for the density, thus assuring physical consistency of the simulation. In fact we reproduce the Oxygen–Oxygen radial distribution expected; if there was any physical anomaly of the molecular density one would have seen anomalies in the Oxygen–Oxygen radial distribution function.

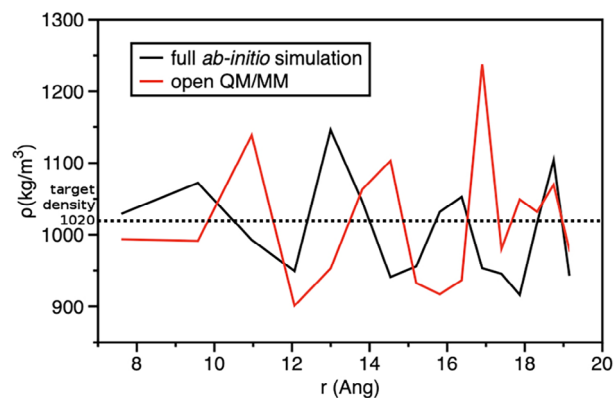


Figure 2. The red curve shows the molecular density along the radial direction from the center of the QM region up to the region defined as Δ in the MM region; the black curve shows the corresponding quantity in the full ab initio calculation of reference. As it can be seen the density fluctuates similarly around its expected value in both cases. There is not a systematic deviation from the average as it usually occurs in AdResS simulations when the molecular representation of the reservoir is not compatible with the molecular representation of the open system. This implies that there is no need of a thermodynamic force as in standard AdResS.

4. Determination of the QM Size: Results of the 440 Molecules System

In this section, we report the results of the numerical checks that have been made to test the physical consistency of the QM region as open system embedded in a reservoir of classical particles.

In particular, here we discuss the results for the successful case of the QM region with 440 molecules, the results for the non successful case of 192 molecules are reported in the Appendix.

The particle number density, with the corresponding fluctuations around the expected average, reproduces in a satisfactory manner the full ab initio target and retain the uniform character of the liquid (see **Figure 2**). This behavior is satisfactory enough to not even require the application of the thermodynamic force in the classical Δ region, thus we consider the behavior of the density close enough to the reference case. Once the mandatory condition on the density is satisfied, the next tests of consistency are the Oxygen–Oxygen and the Oxygen–Hydrogen radial distribution functions. Such quantities characterize uniquely the structure of the liquid, thus if the results for the QM open region are in agreement with the results of the reference calculation, the implication is that the key features of the liquid structure are reproduced, that is, w.r.t. such properties the QM region and the subregion of the full ab initio system of reference are physically equivalent. **Figures 3** and **4** show a satisfactory agreement.

Next test of consistency concerns the exchange of particles between the QM region and the classical environment, the statistical behavior of such an exchange should be close to that of a sub-region with its full quantum environment in the full ab initio simulation of reference. To such aim we calculate the $P(N)$, that is, the probability distribution of finding N molecules in the QM region. **Figure 5** shows a rather good agreement, the peak is shifted of only 5% w.r.t. the full ab initio calculation, this negligible shift may be due also to the smaller size of the full ab initio simulation, however this is not a relevant difference. More importantly, the

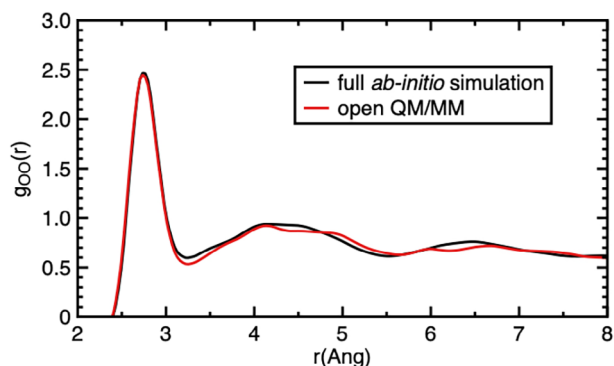


Figure 3. Oxygen–Oxygen local radial distribution function. The specification “local” is due to the fact that the calculation is performed only the QM region and in the equivalent subregion of the full *ab initio* simulation of reference. There is no need to normalize it because the calculation is performed in the same way in both cases and the relevant information regards the agreement of the results.

shape of the curve is close to a gaussian and very close to the one of reference. The similarity between the results of the QM region and the results of the reference calculation is very important because the shape of the curve expresses, in a statistical manner, the process of system-reservoir exchange of particles. So far, we have made use of the standard consistency tests employed in classical simulations of open systems, within the AdResS scheme, however we are now dealing with molecules with electrons and thus the corresponding electronic structure properties are of major relevance. As said before, to the aim of testing the validity of such properties in the QM region, we calculate the Wannier–Wannier and Wannier-Oxygen radial distribution functions; we underline once again that the Wannier centers express the localization of electrons (electron density) in a physical point, thus it carries a correspondent, average, classical-like point-particle structure of the quantum molecule. In this sense the Wannier–Wannier and Wannier-Oxygen uniquely express in essence the electron density correlation between molecules in a liquid like water.

Figures 6 and 9 show a satisfactory global agreement with the reference results, thus assuring that, on the large scale, in average, the electronic structure of the liquid in the QM region is physically consistent with an open system embedded in a full quantum environment, as required by the consistency tests.

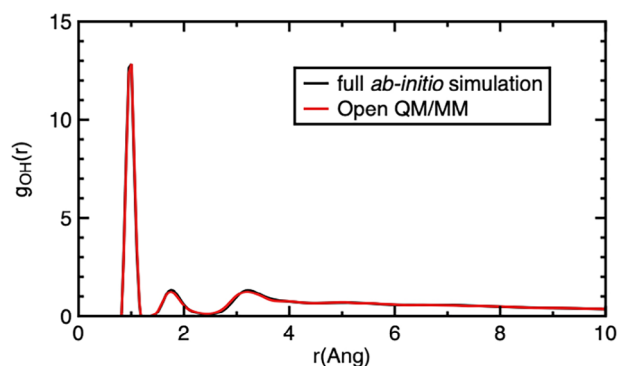


Figure 4. As the previous figure but for the Oxygen–Hydrogen radial distribution function.

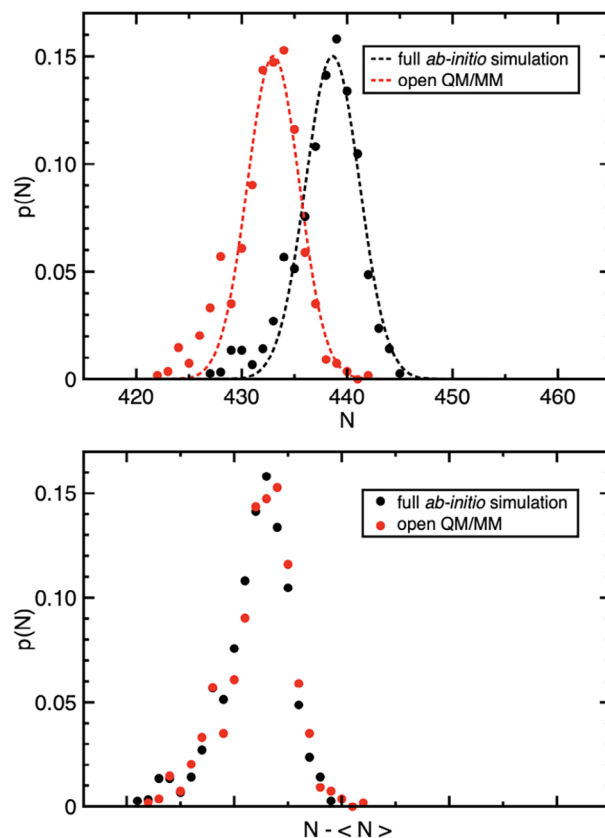


Figure 5. Probability of finding N particles in the QM region and in the equivalent subregion of the full *ab initio* simulation of reference, $P(N)$, (top). (Bottom) the shifted $N - \langle N \rangle$ curves so that one can appreciate the similarity of the shape.

Figures 7 and 10, show with a very fine binning the key feature of charge–charge localization and Oxygen-charge localization inside the single molecule, while **Figures 8 and 11** show the zooming of the charge–charge and Oxygen-charge correlations between different molecules. The results are indeed satisfactory.

In order to show that for the QM region the consistency of the liquid structure and of the electronic structure leads to the consistency of a relevant physical quantity, we calculated the probability distribution of the molecular dipole. For a dipolar liquid like

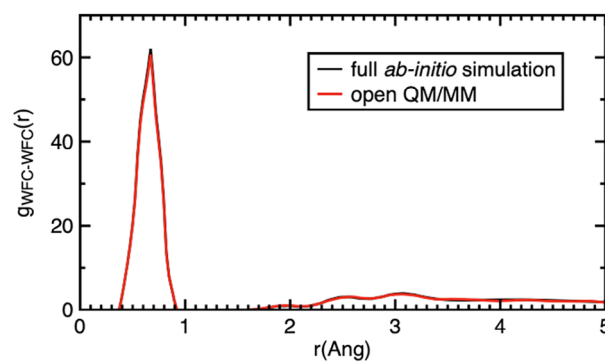


Figure 6. Global radial distribution function for the Wannier–Wannier centers.

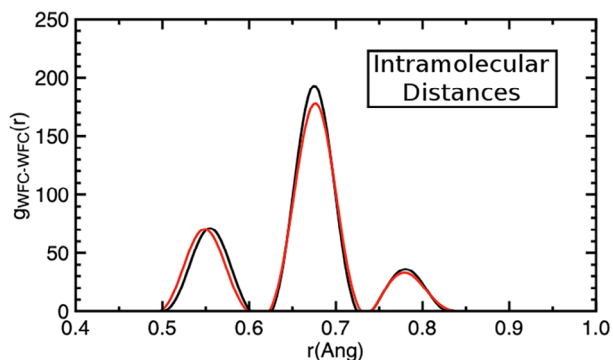


Figure 7. As the previous figures but for the Wannier–Wannier centers for the intramolecular distance.

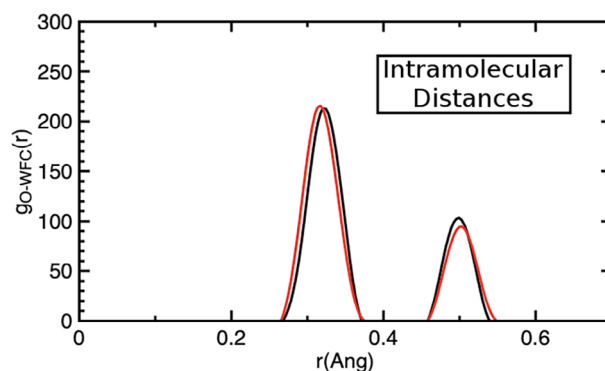


Figure 10. As the previous figures but for the Wannier–Oxygen points for the intramolecular distance.

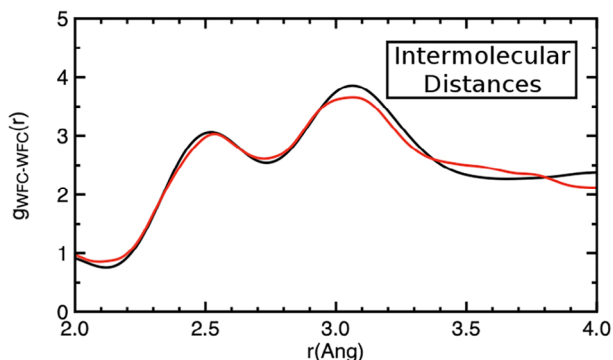


Figure 8. As the previous figures but for the Wannier–Wannier centers for the inter-molecular distance.

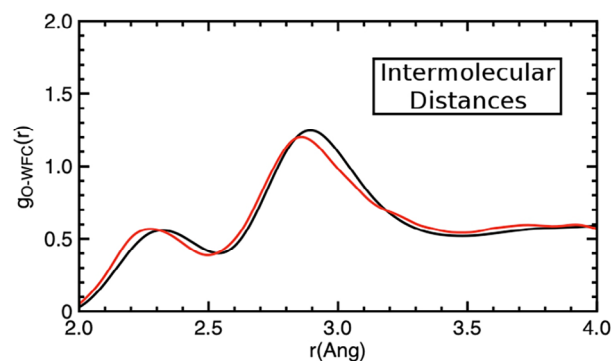


Figure 11. As the previous figures but for the Wannier–Oxygen points for the inter-molecular distance.

water such a property is of central relevance.^[18] **Figure 12** shows the agreement of the curve calculated in the QM region with the curve of reference; such an agreement is satisfactory.

5. Discussion and Conclusion

We have applied a statistical mechanics protocol for defining a physically consistent open region at electronic resolution embedded in a reservoir of classical molecules. We have shown that for a QM size containing about 440 molecules the results of the calculations in the QM region agree in a satisfactory manner

with those obtained in the equivalent subregion of a full *ab initio* simulation of reference. As a consequence, in essence, this study shows two different aspects: i) the *FTIP3P* model of water that forms the reservoir is physically compatible with the *ab initio* description of the molecules at electronic level and thus it can assure thermodynamic equilibrium to the QM region; ii), with the thermodynamic equilibrium assured by the reservoir, for liquid water a spherical subregion of about 440 molecules represents a size sufficient to support the key features of the intra- and intermolecular electronic structure (see also refs.^[31–34]

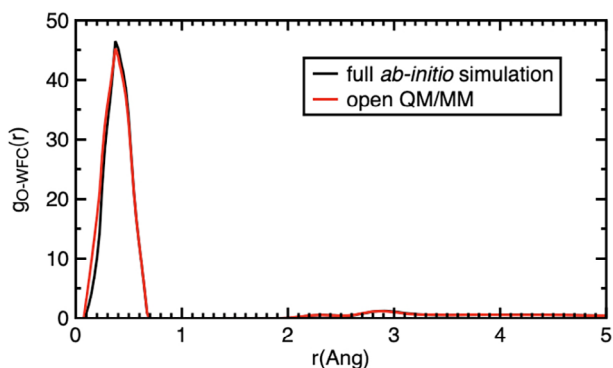


Figure 9. Global radial distribution function for the Wannier–Oxygen points.

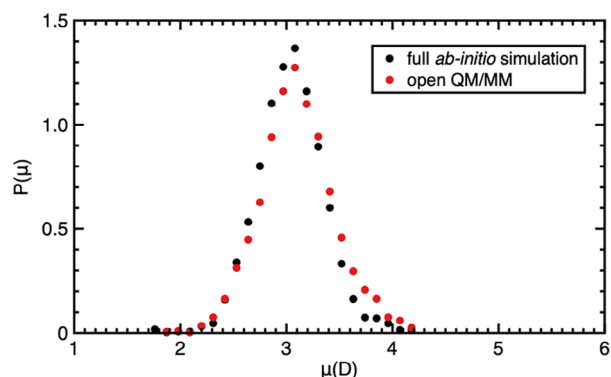


Figure 12. $P(\mu)$ in the QM region and in the equivalent subregion of the reference full *ab initio* calculation. The dipole μ , is measured in Debye. The agreement is highly satisfactory.

for a general discussion about the minimal size of a system). For smaller QM regions the reproduction of several properties become questionable. The possibility of systematically checking the physical consistency of an open quantum region with a classical reservoir is a relevant information for the developing of a well founded open QM/MM approach. In fact the study shows that the protocol of AdResS for classical systems can be extended to the quantum ab initio electronic calculations and it is possible to define a meaningful open region.

At this point, two main concerns about the conclusions above may rise: i) does the definition of an optimal QM region require always a case-by-case expensive simulation of reference?

If it is so, then one would not have any computational or conceptual convenience; ii) is the existence of a classical model of molecule, compatible with the quantum model, available for all systems of interest?

Point (i) can be addressed with the very same arguments of the equivalent classical AdResS, that is, one reference system can be used to tune the QM region for an entire class of applications, under the same QM level of calculations. For example, in the case studied here one could safely conclude that for the solvation of small molecules a QM size of the order of 500 molecules is sufficient to assure (a) structural and electronic properties and, at the same time, (b) any possible thermodynamic effect due to the fluctuation of the number of water molecules. This means that one can be assured about the physical consistency of both local microscopic molecular properties and global thermodynamic properties. For the solvation of larger molecules one may consider a system with a prototype molecule as solute that can be used as reference for molecules of similar or smaller size. What one can certainly conclude by applying the protocol is whether the size of the system used for an ab initio or the size of the QM region of a standard QM/MM simulation is physically meaningful.

The same view applies to ab initio approaches with continuum embedding^[35] or more sophisticated Drude oscillators for the environment.^[36] Such approaches, certainly elegant and efficient to detect local structural and spectroscopic properties, cannot, by statistical mechanics principles, take into account thermodynamic effects of particle fluctuations, unless specifically designed with some Monte Carlo approach, which is usually computationally expensive.^[37,38] However, even if insertion and deletion of molecules is present, the question about the minimal size of the QM region with statistical mechanics consistency still holds. In any case, for highly fluctuating systems like water the correct dynamic exchange of molecules with the environment may be a key feature for many properties of interest.

For example in chemical reactions in water of small molecules like NaCl^[39] or Formamide,^[40] the systems used for the calculations have a smaller size than the one determined here and there is not the possibility of exchanging particles with a reservoir. The current standard computational resources allow nowadays to routinely treat QM/MM systems with the QM size found here, thus the advantage of using our method for such systems is that if there is any relevant thermodynamic effect due to the particle density fluctuations then one would certainly capture it. Specific chemical reactions are outside the scope of the current paper since the proof of concept regards the solvent, however a systematic revision of the subject for small molecules is certainly the first step in future application of our approach.

Regarding point (ii), the development of force-fields from ab initio data with modern techniques of machine learning can essentially provide highly compatible force-fields for any molecular system treated at ab initio level.^[41,42]

In conclusion, the application of the protocol for a physically consistent definition of an open QM region is certainly not the final solution to build a systematic tool for the simulation of open quantum molecular systems. It certainly assures that properties calculated in the QM region defined by the method are physically meaningful, thus it can be certainly used as a basis for the design of a simulation tool for open quantum system. It must be said that already at the current stage it allows for meaningful applications to a large class of systems, as discussed above.

Appendix: Results for the System with 192 Molecules in the QM Region

In this section, we report the results of the simulation for the QM region containing in average 192 molecules. As it can be seen the particle density in **Figure A1** fluctuates around the target value in a sufficiently similar manner to the reference case, thus also in this case the application of the thermodynamic force is negligible. Also the Oxygen–Oxygen and

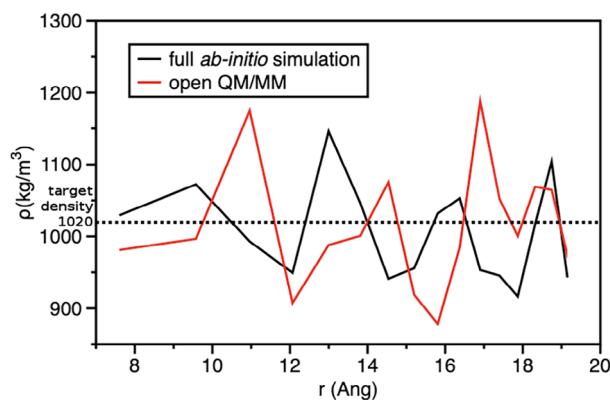


Figure A1. As for the case with the larger QM region discussed before, here the red curve corresponds to the molecular density along the radial direction from the center of the QM region up to the region defined as Δ in the MM region; the black curve shows the corresponding quantity in the full ab initio calculation of reference. As for the case with the larger QM region, the density fluctuates around its expected value, thus, in this case there is no need of a thermodynamic force as in standard AdResS.

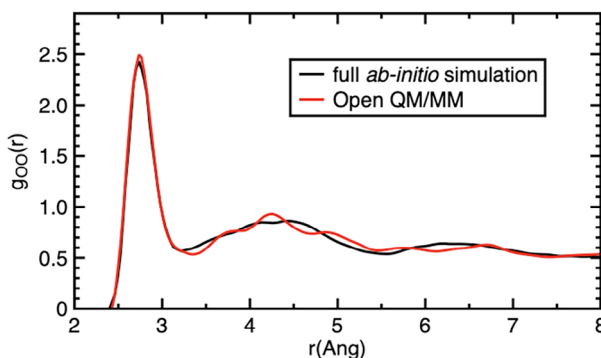


Figure A2. Oxygen–Oxygen local radial distribution function for the system with the smaller QM region. The agreement with reference data is similar to that obtained for the system with a larger QM region.

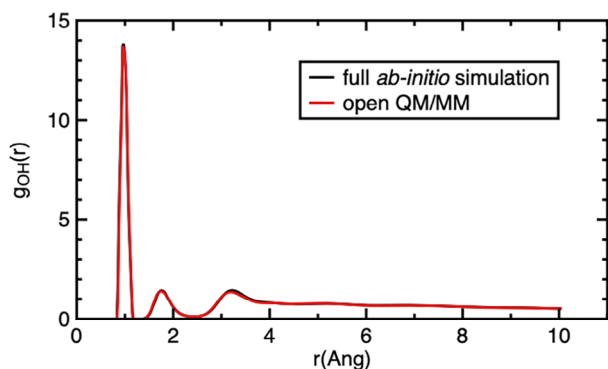


Figure A3. As the previous figure but for the Oxygen–Hydrogen radial distribution function. Also in this case the agreement with reference data is similar to that obtained for the system with a larger QM region.

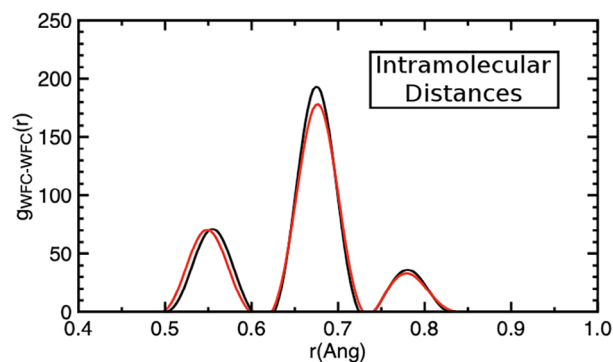


Figure A6. Wannier–Wannier centers radial distribution function for the intramolecular distance. There is a good agreement in such case.

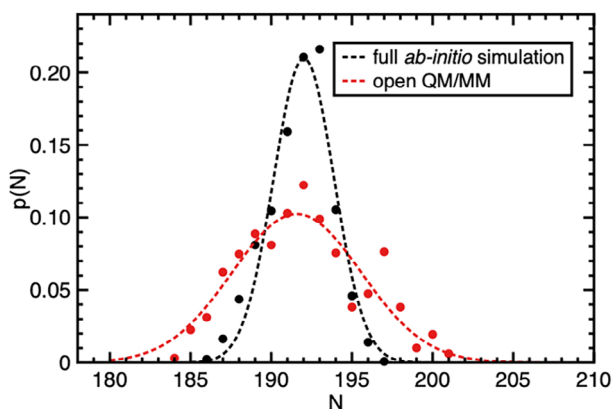


Figure A4. Probability of finding N particles in the QM region and in the equivalent subregion of the full *ab initio* simulation of reference, $P(N)$, (top) for the system with smaller QM region. Differently from the case with a larger QM region, although one gets somehow a slightly better agreement regarding the location of the peak, the shape is rather different from that of reference, implying a different process of particle exchange w.r.t. the reference case. The qualification of the QM region as physically consistent open system is not justified as much as the case of the system with a larger QM region.

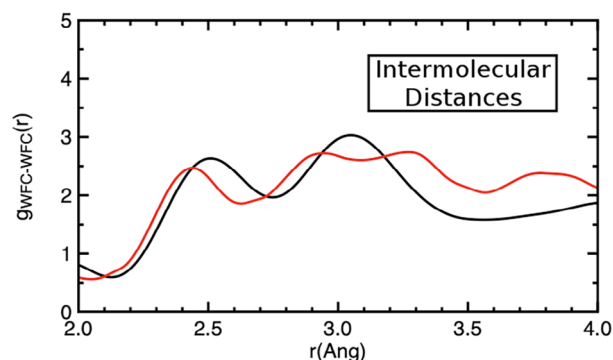


Figure A7. Wannier–Wannier centers radial distribution function for the inter-molecular distance. The disagreement with the reference full *ab initio* case is more pronounced than in the case of the system with larger QM region.

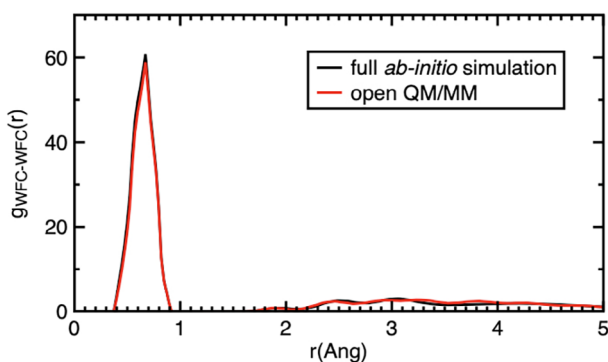


Figure A5. Global radial distribution function for the Wannier–Wannier centers. The figures below show the details of the comparison with the reference full *ab initio* calculations.

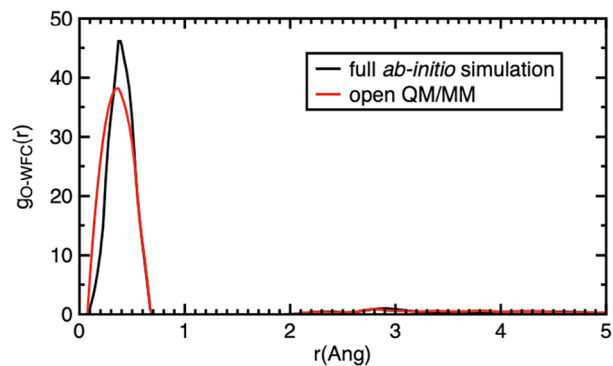


Figure A8. Global radial distribution function for the Wannier–Oxygen points. The disagreement with the reference full *ab initio* case is also in this case more pronounced than in the case with a larger QM region.

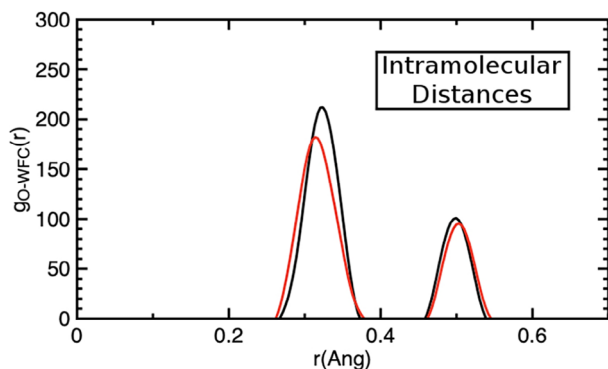


Figure A9. As the previous figures but for the Wannier-Oxygen points for the intramolecular distance.

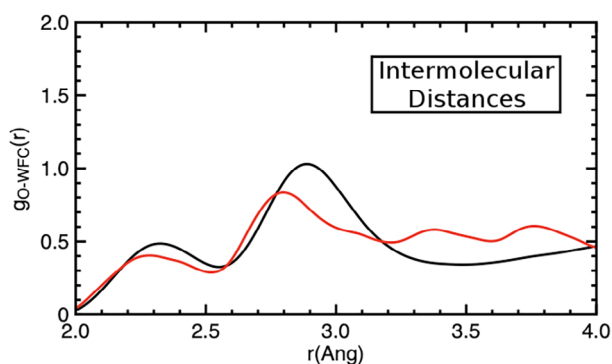


Figure A10. As the previous figures but for the Wannier-Oxygen points for the inter-molecular distance.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), project No.DE 1140/11-1, "Grand Canonical Adaptive Resolution Simulation for Molecules with Electrons". The simulations presented here were performed using HPC resources provided by Nationales Hochleistungsrechnen alliance (NHR), Project No. beb00028.

[Correction added on 24 September 2024, after first online publication: the Acknowledgment section was revised in this version.]

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Received: August 12, 2024
Published online: August 27, 2024

- [1] L. Delle Site, M. Praprotnik, *Phys. Rep.* **2017**, 693, 1.
[2] K. R. Cortes-Huerta, M. Praprotnik, L. Delle Site, *Eur.Phys.Jour.B* **2021**, 94, 189.

- [3] L. Delle Site, R. Klein, *J. Math. Phys.* **2020**, 61, 083102.
[4] R. Klein, L. Delle Site, *J.Phys. A:Math.Th.* **2022**, 55, 155002.
[5] A. Gholami, R. Klein, L. Delle Site, *Phys. Rev. Lett.* **2022**, 129, 230603.
[6] S. Panahian Jand, Z. Nourbakhsh, L. Delle Site, *Front. Chem.* **2022**, 10, 1072665.
[7] L. Delle Site, C. Krekeler, J. Whittaker, A. Agarwal, R. Klein, F. Höfling, *Adv.Th.Sim.* **2019**, 2, 1900014.
[8] A. Evangelakis, S. Panahian Jand, L. Delle Site, *ChemistryOpen* **2022**, 11, e202100286.
[9] L. Delle Site, *Comp.Phys.Comm.* **2018**, 222, 94.
[10] L. Delle Site, *Adv.Th.Sim.* **2018**, 1, 1800056.
[11] A. Duster, C.-H. Wang, C. Garza, D. Miller, H. Lin, *WIREs Comp. Mol. Sci.* **2017**, 7, e1310.
[12] R. Miranda-Quintana, P. Ayers, *J.Chem.Phys.* **2016**, 144, 244112.
[13] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, K. Kremer, *Phys. Rev. Lett.* **2012**, 108, 170602.
[14] A. Gholami, F. Höfling, R. Klein, L. Delle Site, *Adv.Th.Sim.* **2021**, 4, 2000303.
[15] L. Delle Site, A. Djurdjevac, *J. Phys. A: Math. Th.* **2024**, 57, 255002.
[16] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, D. Vanderbilt, *Rev. Mod. Phys.* **2012**, 84, 1419.
[17] P. Silvestrelli, M. Parrinello, *J. Chem. Phys.* **1999**, 111, 3572.
[18] P. L. Silvestrelli, M. Parrinello, *Phys. Rev. Lett.* **1999**, 82, 3308.
[19] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Chütt, F. Schiffrmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Wever, U. Borstnik, M. tallefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, *J. Chem. Phys.* **2020**, 152, 194103.
[20] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, 77, 3865.
[21] S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **1996**, 54, 1703.
[22] J. VandeVondele, J. Hutter, *J. Chem. Phys.* **2007**, 127, 114105.
[23] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, 132, 154104.
[24] D. J. Evans, B. L. Holian, *J. Chem. Phys.* **1985**, 83, 4069.
[25] Y. Shao, C. Saravanan, M. Head-Gordon, C. A. White, *J. Chem. Phys.* **2003**, 118, 6144.
[26] A. M. Niklasson, C. J. Tymczak, M. Challacombe, *J. Chem. Phys.* **2003**, 118, 8611.
[27] L. Mones, A. Jones, A. W. Götz, T. Laino, R. C. Walker, B. Leimkuhler, G. Csany, N. Bernstein, *J. Comp. Chem.* **2015**, 36, 633.
[28] D. J. Price, C. L. Brooks, *J. Chem. Phys.* **2004**, 121, 10096.
[29] A. Jones, B. Leimkuhler, *J. Chem. Phys.* **2011**, 135, 084125.
[30] A. Agarwal, L. Delle Site, *J. Chem. Phys.* **2015**, 143, 094102.
[31] L. Delle Site, G. Ciccotti, C. Hartmann, *J. Stat. Mech.: Theory Exp.* **2017**, 2017, 8.
[32] B. Reible, C. Hartmann, L. Delle Site, *Lett. Math. Phys.* **2022**, 112, 97.
[33] B. M. Reible, J. F. Hille, C. Hartmann, L. Delle Site, *Phys. Rev. Res.* **2023**, 5, 023156.
[34] C. Hartmann, L. Delle Site, *Phys. Rev. A* **2024**, 109, 022209.
[35] O. Andreussi, G. Fiscaro, *Int. J. Quant. Chem.* **2019**, 119, e25725.
[36] M. Ditte, M. Barborini, L. Medrano Sandonas, A. Tkatchenko, *Phys. Rev. Lett.* **2023**, 131, 228001.
[37] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, Academic Press, Cambridge **1996**.
[38] H. Wang, C. Hartmann, C. Schütte, L. Delle Site, *Phys. Rev. X* **2013**, 3, 011018.
[39] J. Timko, D. Bucher, S. Kuyucak, *J. Chem. Phys.* **2010**, 132, 114510.
[40] F. Pietrucci, M. Saitta, *Proc. Natl. Am. Soc.* **2015**, 112, 15030.
[41] F. Noe, A. Tkatchenko, K.-R. Müller, C. Clementi, *Ann. Rev. Phys. Chem.* **2020**, 71, 361.
[42] L. Zhang, H. Wang, R. Car, Weinan E, *Phys. Rev. Lett.* **2021**, 126, 236001.