Accurate SCC-DFTB Parametrization of Liquid Water with Improved Atomic Charges and Iterative Boltzmann Inversion

Nicolas Cinq, Aude Simon, Fernand Louisnard, Jérôme Cuny

An Accurate SCC-DFTB Parametrization of Liquid Water with Improved Atomic Charges and Iterative Boltzmann Inversion

Nicolas Cinq, Aude Simon,* Fernand Louisnard, and Jérôme Cuny*

Laboratoire de Chimie et Physique Quantiques (LCPQ), FeRMI Institute, Université de Toulouse [UT3] and CNRS, F-31062 Toulouse, France

E-mail: aude.simon@irsamc.ups-tlse.fr; jerome.cuny@irsamc.ups-tlse.fr

Abstract

This work presents improvements of the description of liquid water within the self-consistent-charge density-functional based tight-binding scheme combining the use of Weighted Mulliken (WMull) charges and optimized O-H repulsive potential through the Iterative Boltzmann Inversion (IBI) process. The quality of the newly developed models is validated considering pair radial distribution functions (RDFs), as well as other structural, energetic, thermodynamic, and dynamic properties. The use of WMull charges certainly improves the agreement with experimental data, however leading to over-structured RDFs at short distance, that can be further improved by considering an optimized O-H repulsive potential obtained by the IBI process. Three different schemes were used to optimize this potential: (i) optimization including short O-H distances. This led to accurate RDFs as well as improved self-diffusion coefficient and heat of vaporization while proton transfer energy barrier is severely deteriorated ; (ii) optimization starting at long distance. The proton transfer energy barrier is recovered while the heat of vaporization is deteriorated and the O-H RDF is less accurate at short distance ; (iii) optimization within the path-integral molecular dynamics scheme which allows to exclude nuclear quantum effects from the repulsive potential. The latter potential, in conjunction with the WMull improved atomic charges, provides similar results as (i) for structural, dynamic and thermodynamic properties while recovering a large part of the proton transfer energy barrier. It therefore offers a good compromise to study both dynamic properties and chemistry within liquid water at a quantum chemical level.

1 INTRODUCTION

Due to its tremendous importance on Earth, liquid water has raised a lot of interest all the more as it displays a variety of anomalous behaviors that makes it a fascinating substance. It has motivated a considerable amount of studies under various conditions,[1–7] among which theoretical investigations focusing on its structural and dynamic properties.[2–16] Despite con-

tinuous theoretical efforts and methodological improvements, the simulation of liquid water remains challenging. The variety of theoretical issues that arises always necessitates to find an appropriate compromise between computational cost and accuracy. Force fields (FFs) have been developed to model liquid water under various intensive parameters (T,P) with a low computational cost. Both non-polarizable FFs, for instance TIPnP[17–20] and SPC/E[21] models, and polarizable FFs, for instance Thole-type models (TTM),[22–25] AMOEBA models,[26–28] and others,[29–31] lead to accurate results for macroscopic properties of liquid water considering systems containing up to thousands of atoms for up to microseconds simulations. Among them, the MB-pol water model[32–34] has been shown to quantitatively reproduce a large variety of properties of liquid water and can be considered to date as the most accurate FF potential for modelling liquid water.[30,35] However, due to the lack of an explicit description of electronic structure, a proper description of solvation properties and chemical reactivity involving the solvent water molecules are out of reach, or becomes computationally very demanding, and FF potentials also suffer from transferability issues. A way to introduce the description of electronic structure at a low computational cost is to use density functional theory (DFT). DFT methods have been widely used to model liquid water at the atomic level and it has been shown that an accurate description at this level requires dispersion corrections,[36–38] an hybrid functional[38–40] as well as the inclusion of nuclear quantum effects (NQEs)[7,16,41–44] using, for instance, the path-integral molecular dynamics (PIMD) scheme.[45–48] The use of an hybrid functional and dispersion corrections already leads to very accurate pair radial distribution functions (RDFs), as demonstrated in the work of DiStasio *et al.*,[38] but at a rather high computational cost although the NQEs were not taken into account. Molecular dynamics (MD) simulations using DFT would thus be hardly feasible to obtain converged dynamic properties due to a tremendous computational cost.

A way for improvement has recently been explored by benefiting from machine learning (ML) algorithms. For instance, Dasgupta *et al.* proposed to increase the accuracy of the strongly constrained and appropriate normed (SCAN) functional,[49] a meta-GGA functional

known to provide good prediction of liquid water properties, through a density-corrected many-body formalism based on a data-driven approach, namely MB-SCAN(DC).[15] This method allows to reproduce coupled cluster accuracy for structures and binding energies of water clusters. The interest of this method is that MD simulations carried out with MB-SCAN(DC) also reproduce the properties of liquid water. A deep neural-network (DNN) potential for liquid water was also trained with data obtained from the SCAN0 functional, that includes 10% of exact exchange, improving the description of liquid water with respect to the SCAN functional.[50] It was used to conduct 2 ns MD simulations and 500 ps PIMD simulations, which is much longer than common DFT MD and PIMD simulations, while maintaining DFT accuracy. This leads to a very good extrapolated value of the self-diffusion coefficient of liquid water, 0.29 $Å^2.ps^{-1}$ at 300 K, as compared to the experimental value of 0.24 $Å^2.ps^{-1}$. However, the predicted H-bond network remains overstructured and the density remains slightly overestimated (1.041 $g.cm^{-3}$ with PIMD and 1.030 $g.cm^{-3}$ with MD) with respect to experimental values.[50] Other studies have recently proposed DNN potentials for liquid water trained on high level DFT or wavefunction data,[51–53] but as for FFs, their range of applicability and transferability is yet to be tested.

Apart from the combination with ML algorithms, in between FF and DFT, the density-functional based tight-binding (DFTB) formalism[54–57] appears as an intermediate method of choice in terms of computational cost and accuracy. It provides an explicit electronic structure description while reducing the computational cost with respect to standard DFT methods, allowing to model large molecular systems. Furthermore, the transferability of its parameters is a considerable advantage over FF approaches, which makes it possible to study a variety of systems and chemical reactions without need of a specific parametrisation procedure. Obtaining a good description of liquid water using a DFTB hamiltonian has thus been the subject of several studies. Unfortunatly, the original formulation of DFTB in its self-consistent-charge formulation (SCC-DFTB),[54] was shown to provide a poor description of this system.[58–61] The comparison between the computed $g_{OO}(r)$, $g_{OH}(r)$ and

$g_{HH}(r)$ and experimental RDFs demonstrated an overcoordinated first solvation shell and an overall lack of structuration in the second and third solvation shells. A few proposals were made to improve this description. Maupin *et al.* first proposed a modified scheme which includes a hydrogen bonding damping function (HBD).[57,58] However, the main defaults of the original SCC-DFTB scheme remain. In 2013, Doemer *et al.* first proposed to optimize, following the Iterative Boltzmann Inversion (IBI) method, the O-O or O-H repulsive potentials of the SCC-DFTB hamiltonian to fit RDFs obtained at the DFT level and obtained considerable improvements with respect to original SCC-DFTB.[62] More recently, Lourenço *et al.* improved significantly the description of liquid water tuning both the O-O and O-H repulsive potentials by means of the same IBI process.[60] Based on detailed comparisons with experiments and other simulations, the authors discussed the achievements and failures of the proposed SCC-DFTB potentials. Meanwhile, Cuny *et al.* also proposed to use improved atomic charges that go beyond the Mulliken charges used in the original formulation of the SCC-DFTB hamiltonian.[61] The proposed scheme, referred to as weighted Mulliken (WMull) charges, introduces a nonsymmetric distribution of the overlapping densities between two centers. It was originally proposed to improve the SCC-DFTB description of water ice.[63] This description of atomic charges was shown to significantly improve the liquid water structuring with respect to the original formulation of SCC-DFTB. Furthermore, the authors also considered for the first time the influence of NQEs using the PIMD scheme in conjunction with SCC-DFTB. As in the DFT framework, this leads to a slight decrease of the structuring of $g_{OH}(r)$ and $g_{HH}(r)$ with respect to classical SCC-DFTB simulations, but without significantly improving the overall agreement with experimental RDFs. Finally, Goyal and co-workers proposed to go beyond second-order DFTB by considering various formulations of third-order DFTB (DFTB3) to model liquid water.[64,65] None of them succeeds in a real improvement of the description of liquid water except when the O-H repulsive potential is optimised following the aforementioned procedure.

Prompted by the hereabove mentioned studies that allowed to significantly improve the

DFTB description of liquid water modifying (i) the O-H repulsive potential using the IBI process and (ii) the definition of atomic charges using the WMull scheme, we demonstrate in the present article that combining both approaches enables us to go a step further in the improvement of the description of liquid water at the SCC-DFTB level. In particular, we discuss in details technical issues related to the IBI process and its range of applications. We also demonstrate for the first time how a O-H repulsive potential can be obtained from PIMD simulations without double-counting NQEs. The paper is organized as follows: the computational methods and simulations details employed in the article are described in section 2. The different potentials we obtained and their performances on various properties are presented and discussed in section 3. We finally conclude on the actual status of the SCC-DFTB description of liquid water and ways of potential improvements.

2 METHODS

SCC-DFTB Framework and Potential. The SCC-DFTB method is an approximate DFT scheme in the Kohn–Sham orbital based formalism of which a detailed description can be found in the original founding papers.[54,55] The SCC-DFTB hamiltonian can be derived from DFT through a second-order Taylor expansion of the total DFT energy around a reference density. By doing so, one obtains the following equation for the SCC-DFTB total energy:

$$E^{SCC-DFTB} = \sum_{i,\mu,\nu} n_i c_{i\mu} c_{i\nu} H^0_{\mu\nu} + \frac{1}{2} \sum_{a,b} \gamma_{ab} q_a q_b + \sum_{ab} E^{rep}(R_{ab}) \tag{1}$$

where the first term is usually referred to as the band energy term, with n_i the occupation of orbital i and $H^0_{\mu\nu}$ the DFT Kohn–Sham operator matrix element computed at the reference density. These matrix elements are pre-calculated parametrized integrals which provides to SCC-DFTB, in addition to the use of a minimal atomic valence basis set $\{\phi_\mu\}$, its computational efficiency. In this study, we used the mio-set for the Slater–Koster tables

of integrals.[54]

We detail in the following the last two terms of eq. 1, relevant for the present work. The second term of eq. 1 is the second-order term in the Taylor expansion, which is expressed as a function of the atomic charges q_a and q_b. γ_{ab} is the γ matrix element that depends on the atomic Hubbard parameters U_a and U_b. In the original formulation of SCC-DFTB, the atomic charges q_a and q_b are computed within the Mulliken scheme. This can be refined using an improved definition of atomic charges such as the WMull scheme described below. The last term of eq. 1 is the repulsive energy expressed as a sum over interatomic repulsive terms $E^{rep}(R_{ab})$ that, in the mio-set of the Slater–Koster tables, take the form of spline functions. They can be obtained by fitting the difference between an energy reference calculation and the band and second-order energy terms. Finally, to improve description of intermolecular interactions, empirical terms has been added to the SCC-DFTB potential to account for dispersion interactions,[66,67] with atomic C_6 parameters derived from the study by Wu and Yang[68].

As discussed in the introduction, SCC-DFTB in its original formulation provides a poor description of liquid water. In the following, we describe in details the two parameters that are optimized in the present work in order to improve this description (i) the atomic charges and (ii) the repulsive O-H potential.

-(i)- Weighted Mulliken Charges. The use of the default Mulliken charges in the SCC-DFTB scheme does not allow for a satisfactory description of liquid water. As mentioned in the introduction, in order to improve its description, Cuny *et al.* proposed to replace them by the Weighted Mulliken charges (quoted hereater WMull and initially proposed by Michoulier *et al.*[63]) which introduce a non symmetric bias directly into the Mulliken repartition of the overlapping density:[61]

$$\phi_\mu(r)\phi_\nu(r) \approx \frac{1}{2}S_{\mu\nu}((1+t_{\mu\nu})|\phi_\mu(r)|^2 + (1-t_{\mu\nu})|\phi_\nu(r)|^2) \tag{2}$$

where the parameter $t_{\mu\nu}$ is introduced to modify the distribution of charges between the

two centers μ and ν. $t_{\mu\nu}$ is a free empirical parameter that ranges from -1 to 1 and describes the polarity of a given chemical bond. For liquid water, the impact of the t_{OH} value has been studied in details by Cuny et al.[61] and has been shown to be optimal for t_{OH}=0.28. Even though the use of t_{OH} improves the structuring of liquid water at medium and long range distances, it leads to a too high first peak of $g_{OO}(r)$. This results in an overstructuration at short distance that needs to be compensated for.

-(ii)- Iterative Boltzmann Inversion and Repulsive Potential. Another way to improve the description of liquid water is to tune repulsive potentials by means of the iterative Boltzmann inversion process (IBI).[59,60,62] Applied in the present context, the aim of the IBI process is to self-consistently fit a RDF computed at the SCC-DFTB level to a reference RDF by adjusting one or several repulsive potentials used in the SCC-DFTB simulation. If successful, this process allows to determine the repulsive potentials that lead to computed RDFs equal to the reference RDFs. Although several reference RDFs can be considered, as done for instance by Lourenço et al.[60], we have considered only the experimental $g_{OH}(r)$ of Soper[69] as the reference RDF and thus only optimized the repulsive O-H potential.The initial repulsive O-H potential is from the *mio* set of parameters.[54]

In the IBI process, the central quantity to evaluate is g_{OH}(r), which is calculated by:

$$g_{OH}(r) = \frac{\Omega}{4\pi r^2} p_{OH}(r) \tag{3}$$

where $p_{OH}(r)$ is the probability density of finding a hydrogen atom at a certain distance r from an oxygen atom and Ω is the volume of the simulated unit cell. The probability density $p_{OH}(r)$ is extracted from MD trajectories and the Boltzmann inversion then relates $g_{OH}(r)$ to the Helmholtz free energy:[70]

$$F_{OH}(r) = -k_B T ln(g_{OH}(r)) + C \tag{4}$$

Therefore, E_{rep} is modified using the IBI scheme following:

$$E^{i+1}_{rep}(r) = E^{i}_{rep}(r) + k_B T ln(\frac{g^{i}_{OH}(r)}{g^{exp}_{OH}(r)}) \quad (5)$$

where $g^{i}_{OH}(r)$ and $g^{exp}_{OH}(r)$ are the computed and reference RDFs, respectively, at distance r. $E^{i}_{rep}(r)$ is the O-H repulsive potential at distance r at the ith iteration, k_B the Boltzmann constant and T the temperature.

We proceeded to the iterative update of $E^{i}_{rep}(r)$ using the following procedure: using $E^{i}_{rep}(r)$, an MD simulation is run (see below for computational details). From this simulation, $g^{i}_{OH}(r)$ is evaluated as well as the $\frac{g^{i}_{OH}(r)}{g^{exp}_{OH}(r)}$ ratio. According to eq. 5, a new $E^{i+1}_{rep}(r)$ function is calculated and used to perform a new MD simulation at iteration $i+1$. The repulsive potential is defined as consecutive splines of 3^{rd} degree polynomial functions, each defined by four parameters. The constant value of each polynomial function is first updated, then the derivative constant is updated to follow the previous modifications and the last two parameters of the polynomial functions are modified to insure continuity and derivative continuity of the repulsive potential. The iterative procedure is stopped after few modifications of the potential (~ 12 iterations) and the potential that leads to the closest $g^{i}_{OH}(r)$ to $g^{exp}_{OH}(r)$ is kept to go further.

Achieving convergence is not straightforward due to stability issues, resulting from the presence of the $\frac{g^{i}_{OH}(r)}{g^{exp}_{OH}(r)}$ term, that have to be considered. Two main problems may indeed occur. First, $g^{i}_{OH}(r)$ is or becomes equal to 0 at a distance r that is inside the fitting range, *i.e.* the range of distances where $E^{i}_{rep}(r)$ is adjusted. $E^{i}_{rep}(r)$ can not be modified any more at this distance r and the IBI procedure fails. Second, $g^{i}_{OH}(r)$ is too small as compared to $g^{exp}_{OH}(r)$. This results in a large negative modification of the potential for the considered distance, and the IBI procedure fails too. Considering the overall shape of $g^{i}_{OH}(r)$ and $g^{exp}_{OH}(r)$, both issues occur mainly at short distances which necessitate to define with care the fitting range of the repulsive potential. For completeness, we thus considered two distinct fitting ranges over which $E^{i}_{rep}(r)$ could be adjusted. In both cases, an automatic procedure does not necessarily insure continuity of the repulsive potential at the junction between the

fitting range and shorter distances. In order to correct this tedious point, continuity was maintained manually after the aforementioned ∼12 modifications of the potential and a new round of iterative procedure is performed. Convergence is achieved when both the shape of the repulsive potential and the behaviour at the junction of the fitting range do not evolve any more.

The IBI approach was first employed for r values between 1.59 and 4.00 Å. Considering only distances larger than 1.59 Å ensures the stability of the algorithm by excluding the beginning of the first intermolecular peak of $g_{OH}(r)$. Continuity was maintained between 1.05 and 1.59 Å considering a third degree polynomial function and using a variable change to fit the shape of the initial potential. In the following, the repulsive potential obtained using this approach will be noted E_{rep}^{opt}. Second, the IBI process was employed in the range 1.93 to 5.54 Å. Continuity was maintained between 1.73 and 1.97 Å in the same way. This second approach was considered in order to counteract issues appearing in the evaluation of proton transfer energy barrier using E_{rep}^{opt} (see Section 3.2). The repulsive potential obtained using this approach is noted $\mathrm{E'}_{rep}^{opt}$ in the following.

Several studies have demonstrated the significant impact of NQEs on $g_{OH}(r)$ of liquid water.[7] Consequently, when considering as reference RDF the experimental $g_{OH}(r)$, NQEs are intrinsically accounted for in this reference. Using a classical MD trajectory in the IBI process thus biases the shape of the repulsive potential by artificially including the contribution of NQEs in this potential. If used to perform a PIMD simulation, NQEs will then be double counted. This can be avoided using PIMD trajectories in the optimization procedure of the repulsive potential. This was done following the first recipe described above and the resulting potential is noted $\mathrm{E}_{rep}^{opt-PIMD}$.

Computational Details. All simulations were performed with the deMonNano code,[71] and consisted in 64 water molecules in a 12.42 Å cubic cell. The optimization of the repulsive potential follows the methodology described above and was performed in the canonical ensemble at 300 K.

In the case of classical simulations, *i.e.* to optimize E_{rep}^{opt} and E'^{opt}_{rep}, a Nose-Hoover chain of five thermostats with frequencies of 800 cm^{-1} and a 0.5 fs timestep were used.[72–74] A preliminary rough optimization of the repulsive O-H potential was performed considering 10 ps between the first 5 iterations. Then a finer optimization was made considering 40 ps between each optimization of the repulsive potential. This led to an evaluation of $g^i_{OH}(r)$ considering 20000 and 80000 frames, respectively. The converged $g_{OH}(r)$, $g_{OO}(r)$ and $g_{HH}(r)$ presented and discussed in section 3 were calculated in the NVT ensemble using the same thermostat, a 0.2 fs timestep, 10 ps of equilibration, 50 ps of production, and the final repulsive O-H potentials.

The $E_{rep}^{opt-PIMD}$ repulsive potential was optimized using PIMD simulations.[7] They were performed using the interface between the deMonNano code and the i-PI code developed by Ceriotti *et al.*[61,75] For the optimization, simulations were performed in the canonical ensemble using the PIGLET thermostat,[76,77] 8 replicas and a 0.5 fs timestep. This formalism allows to use a lower number of replicas as compared to conventional PIMD simulations. The friction and diffusion matrix defining the PIGLET thermostat were automatically generated using the GLE4MD website.[78] The optimization process was performed with 10 ps between each iteration and discontinuity issues were handled as detailed above. 20000 frames were considered for the evaluation of each RDF. The optimized $E_{rep}^{opt-PIMD}$ potential was further used to performed classical simulations following the computational details as for classical MD and PIMD simulations as described above.

Determination of Structural, Dynamics and Thermodynamics Properties. We first calculated oxygen-oxygen-oxygen triplet angular distribution functions, $P_{OOO}(\theta)$. To do so, for each oxygen atom O_i, we considered the angles involving atom O_i and the pair of oxygen atoms displaying distances lower than 3.25 Å from atom O_i. The distribution is then normalised such as $\int_0^\pi P_{OOO}(\theta) sin(\theta) d\theta = 1$.

Two other properties were also considered, the self-diffusion coefficient and heat of vaporization, that were obtained from simulations performed in the micro-canonical ensemble. To

do so, random $(\mathbf{r}, \mathbf{p})$ initial conditions were selected from 100 ps NVT simulations and those with an instantaneous temperature close to 300 K were kept as starting configurations to further conduct 100 ps NVE trajectories. Between 6 and 8 NVE simulations were conducted for each set of parameters considered below. The self-diffusion coefficient and the heat of vaporization were then calculated for each NVE simulation and the final values reported in this work result from the average of all values obtained from the individual NVE simulations.

The heat of vaporization ΔH_{vap} was computed according to the following equation:

$$\Delta H_{vap} = -E_{int} + RT \tag{6}$$

where E_{int} is the mean intermolecular interaction energy computed as:

$$E_{int} = \frac{< E_n > -(nE_{gas})}{n} \tag{7}$$

where $< E_n >$ is the total mean SCC-DFTB potential energy extracted from the NVE simulations, E_{gas} is the energy of an isolated optimized water molecule computed at the same level of theory, n is the number of water molecules in the simulation. The self-diffusion coefficient was determined as the slope of the mean-square displacement as a function of time:

$$D = \lim_{t \to \infty} \frac{\langle |r(t) - r(0)|^2 \rangle}{6t} \tag{8}$$

3 Results and Discussion

3.1 Radial Distribution Functions

In this section, we present and discuss the $g_{OO}(r)$, $g_{OH}(r)$ and $g_{HH}(r)$ RDFs obtained from the different sets of repulsive potentials E_{rep}^{opt}, $\mathrm{E'}_{rep}^{opt}$ and $\mathrm{E}_{rep}^{opt-PIMD}$ optimized as described in section 2.

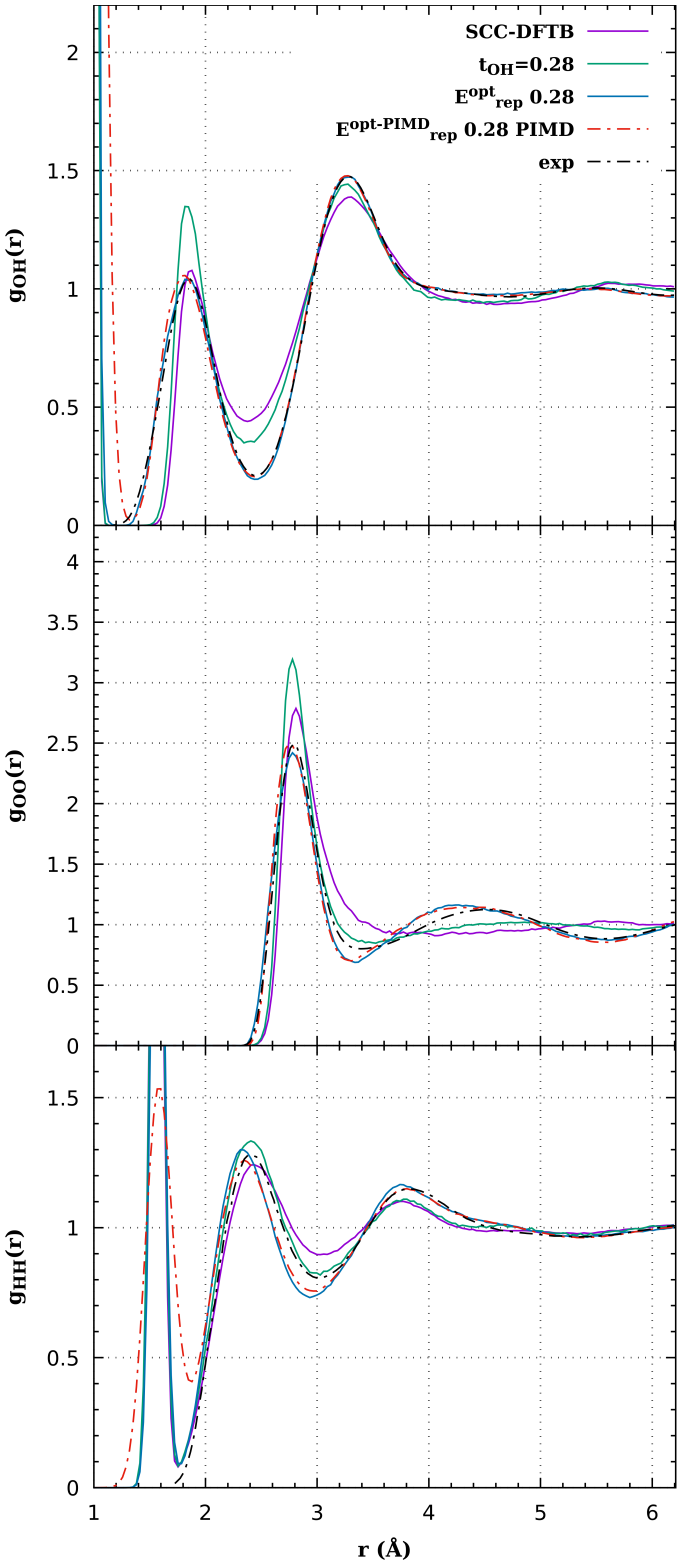

Figure 1: $g_{OH}(r)$ (top), $g_{OO}(r)$ (middle), $g_{HH}(r)$ (bottom) obtained at 300 K for liquid water using the original SCC-DFTB scheme (purple), the WMull scheme with t_{OH}=0.28 (green), the WMull scheme with t_{OH}=0.28 and E_{rep}^{opt} (blue) compared to experimental results from Soper (dashed black lines).[69] The RDFs computed with the same WMull charges (t_{OH}=0.28) and $E_{rep}^{opt-PIMD}$ are also displayed (dashed red lines).

E_{rep}^{opt} Potential and WMull Scheme. Figure 1 provides a comparison between $g_{OO}(r)$, $g_{OH}(r)$ and $g_{HH}(r)$ functions obtained using the original SCC-DFTB hamiltonian, the WMull scheme with t_{OH}=0.28, the WMull scheme with t_{OH}=0.28 in combination E_{rep}^{opt} or $E_{rep}^{opt-PIMD}$. For comparison, experimental $g_{OO}(r)$, $g_{OH}(r)$ and $g_{HH}(r)$ RDFs from Soper [69] are also drawn. Only the RDFs for t_{OH}=0.28 are reported in Figure 1 but other t_{OH} values were tested (0.0, 0.1, 0.2, 0.28, 0.3, 0.4 and 0.5, see Figures S1, S2 and S3 in the SI).

As previously stated, original SCC-DFTB (using Mulliken charges) exhibits issues in the description of liquid water that we remind here. First, considering the $g_{OO}(r)$ curve, it leads to a lack of structuring at medium and long range distance resulting in the absence of minimum around 3.3 Å, an ill-defined second solvation shell and a too wide first peak. Overall, $g_{OH}(r)$ displays the correct shape but the maximum of the first peak is slightly too low, the first minimum is too high, there is a small shift in the position of the second maximum, which also has an overestimated value. As for $g_{HH}(r)$ there are some differences with a first peak slightly shifted and an underestimated value, a first minimum too high and a slight shift for the second maximum which is also too low. These discrepancies where previously discussed,[58–61] and lead to poor dynamical properties such as the self-diffusion coefficient of water. It was computed to be 1.11±0.04 $Å^2.ps^{-1}$ although it was measured at 0.23 $Å^2.ps^{-1}$ experimentally,[79] which can be related to the lack of structuration of $g_{OO}(r)$ at medium and long distances. As explained in section 2, the replacement of the Mulliken charges by the WMull scheme improved the description of liquid water,[61] the t_{OH} parameter being empirical and describing the polarity of the O-H covalent bond. t_{OH}=0.28 was found to be the best value, leading to a better structuring at medium and long range distances at the cost of an over-structuring of the first solvation shell, which corresponds to the first peak at around 2.75 Å on $g_{OO}(r)$.

Further optimizing the O-H repulsive potential finally leads to a very accurate $g_{OH}(r)$ RDF even at short distance, *i.e.* in between 1.4 and 1.8 Å, which is a very tedious region regarding the IBI optimization procedure (see section 2). Indeed, even in the $g_{OH}(r)$ curves

provided in the studies by Goyal et al.[65] and Lourenço et al.[60], this region of the $g_{OH}(r)$ RDF is not properly reproduced, which certainly results from the convergence issues of the IBI process discussed in section 2. Overall, we reach a better agreement between the theoretical and reference $g_{OH}(r)$ than in those two previous studies. As expected, the IBI process applied on the O-H repulsive potential combined with the WMull scheme also greatly improves the $g_{OO}(r)$ function, without explicitly modifying the O-O repulsive potential as performed for instance by Lourenço et al.[60]. Combination of E_{rep}^{opt} and WMull scheme corrects the over- structuring at short distance while increasing the structuration at medium and long distances. This results in a $g_{OO}(r)$ curve that fits well the experimental one. However, the agreement is not perfect, as the second maximum is shifted and slightly too high while the first minimum is too low as compared to the reference RDF. This could be corrected by also fitting the O-O repulsive potential as done by Lourenço et al.[60] , who obtained a good agreement between theoretical and reference $g_{OO}(r)$. The curves of Goyal et al.[65] display similar discrepancies as ours but in a seemingly less pronounced way. $g_{HH}(r)$ presents the most differences with the experimental curve. Indeed, the first maximum is slightly too high and is shifted towards smaller distances. More importantly, the calculated $g_{HH}(r)$ curve is significantly lower than the experimental one in between 1.5 and 3.2 Å. It is worth mentioning that over this range, the WMull potential with t_{OH}=0.28 and no modification of the repulsive potential performs much better. For larger distances, the two sets of functions become similar with E_{rep}^{opt} while a clear difference exists when considering WMull only. Previous studies did not discuss the shape of $g_{HH}(r)$, it is thus not possible to conclude how the present potential performs as compared to others. However one can observe an overall good agreement despite an obvious higher sensitivity of $g_{HH}(r)$.

Similar RDFs were obtained using other t_{OH} values in combination with E_{rep}^{opt}. In that case, a E_{rep}^{opt} repulsive potential is optimized specifically for each t_{OH} value. The results are presented in Figures S1, S2 and S3 of the SI. As can be seen, despite the significant impact of the t_{OH} value on the three sets of RDFs (see reference[61]), providing an optimized

E_{rep}^{opt} potential specifically optimized for each value leads to very similar $g_{OO}(r)$, $g_{OH}(r)$ and $g_{HH}(r)$ functions although the specific shape of each E_{rep}^{opt} potential varies. E_{rep}^{opt} potentials obtained for different t_{OH} values are reported in Figure 2. As can be seen, the IBI process results in a non-physical attractive part in the O-H repulsive potential at short distances. This also appears in the potentials generated in the work of Doemer et al.[62]. The higher the t_{OH} parameter, the deeper is the hollow at short distance of the potential centred at ∼1.35 Å. This distance corresponds exactly to the beginning of the experimental $g_{OH}(r)$ curve, it thus results from the fitting of the O-H intermolecular region at short distance that is not described by the original SCC-DFTB potential. This non-physical behaviour highlights a failure of the SCC-DFTB model to accurately describe hydrogen bonds in liquid water, probably due to the lack of atomic polarisability in this model, that is compensated for in the fitting procedure. As expected, the higher the t_{OH} parameter, the higher is the maximum of the repulsive potential, in between 1.75 and 1.8 Å, to compensate for the increasing structuring of the first solvation shell resulting from the WMull scheme. Finally, although repulsive potentials are expected to be short-range in the original SCC-DFTB formulation, a repulsive contribution up to 3.5 Å is observed after optimization, again, compensating for short-comings of the SCC-DFTB formalism.

$E_{rep}^{opt-PIMD}$ Potential and WMull Scheme. As expected, the $g_{OH}(r)$ function obtained from a PIMD simulation using the $E_{rep}^{opt-PIMD}$ potential, obtained through fitting considering PIMD simulations, in combination with WMull, is very close to the $g_{OH}(r)$ curve discussed above, *i.e.* obtained with a classical MD simulation using the E_{rep}^{opt} potential (see Figure 1). The only difference is a slight shift towards smaller distances of the first maximum. The $g_{OO}(r)$ functions are also very close with a slight increase of the first maximum. Finally, the differences are more pronounced for $g_{HH}(r)$. As expected, the peak below 2.0 Å, which is intermolecular, is broader with PIMD simulations thanks to the inclusion of NQEs. At larger distances, differences appear mainly in terms of amplitudes. The two maxima are slightly lower with PIMD simulations while the minimum at ∼3 Å is slightly higher which

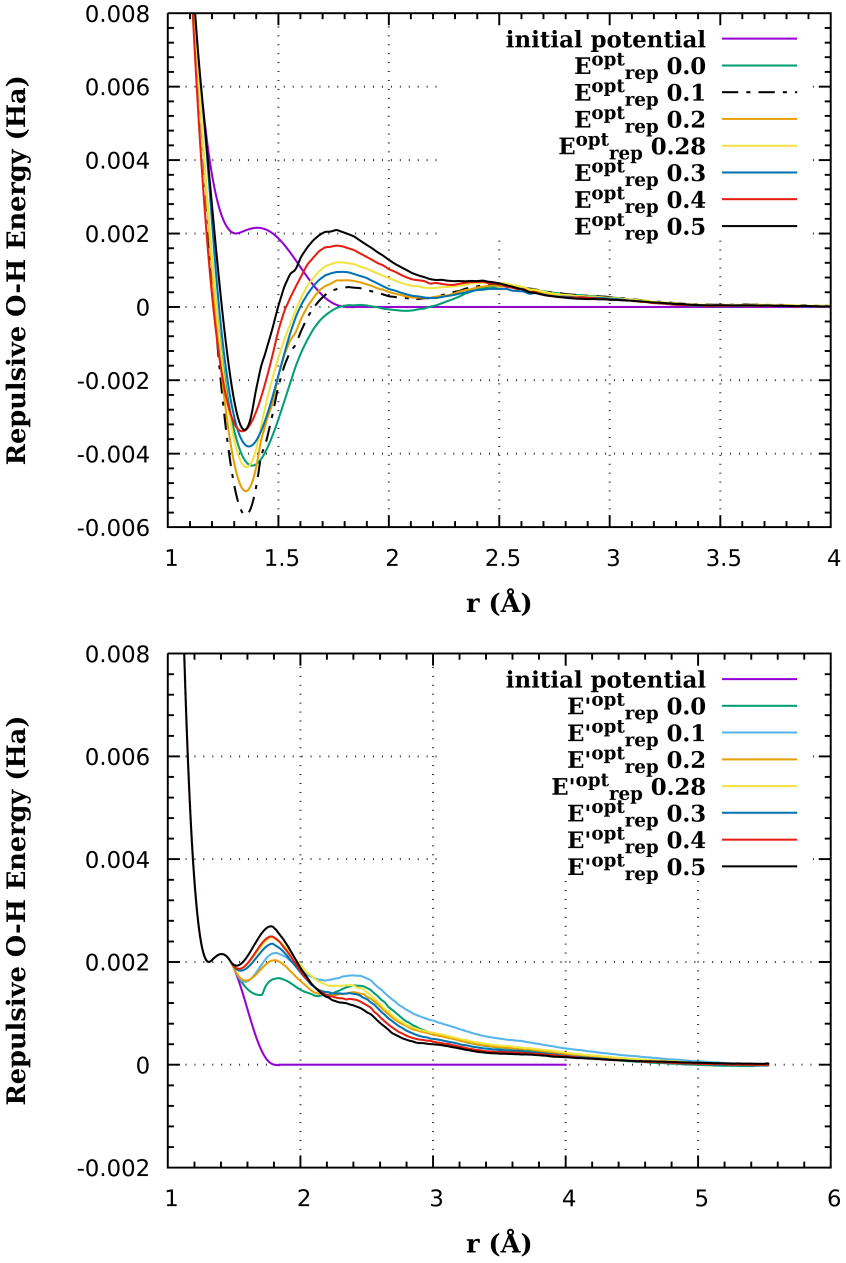

Figure 2: (Top) O-H repulsive potential E_{rep}^{opt} optimized using the IBI procedure within the WMull scheme with t_{OH}=0.0, 0.1, 0.2, 0.28, 0.3, 0.4 and 0.5 compared to the initial O-H repulsive potential E_{rep}. (Bottom) O-H repulsive potential E'^{opt}_{rep} optimized using the IBI procedure within the WMull scheme with t_{OH}=0.0, 0.1, 0.2, 0.28, 0.3, 0.4 and 0.5 compared to the initial O-H repulsive potential E_{rep}.

make the $g_{HH}(r)$ curve obtained with the $E_{rep}^{opt-PIMD}$ potential closer to the experimental one. However, the main discrepancies remain.

Although the RDFs are overall similar, the shape of $E_{rep}^{opt-PIMD}$ significantly differs from E_{rep}^{opt} at the end of the IBI procedure. Indeed, as seen on Figure 3, below ∼1.8 Å, the non-physical attractive part of $E_{rep}^{opt-PIMD}$ is significantly reduced as compared to E_{rep}^{opt}. The enhanced structural fluctuations resulting from the NQEs introduced by the PIMD formalism leads to a broader first intermolecular peak in $g_{OH}(r)$. This partly corrects the difference between theory and experimental reference in between 1.4 and 1.8 Å without modifying the repulsive potential. This leads to a significantly weaker spurious attractive contribution in $E_{rep}^{opt-PIMD}$ than in E_{rep}^{opt}. Beyond 1.8 Å, the two repulsive potentials are rather similar.

For completeness, we have performed a classical MD simulation using the $E_{rep}^{opt-PIMD}$ potential and a PIMD simulation using E_{rep}^{opt}. The results are reported in Figure S4 in the SI. As can be seen, the potentials are hardly transferable from one kind of simulation to another. Indeed, using E_{rep}^{opt} in a PIMD simulation leads to an enormous shift of the first peak of $g_{OH}(r)$ and $g_{OO}(r)$ resulting from an overpopulation of the non-physical attractive part of E_{rep}^{opt}. In contrast, the use of $E_{rep}^{opt-PIMD}$ in a classical MD simulation is less of an issue except for the first peak of $g_{OH}(r)$ that displays a bimodal shape due to an underpopulation of the attractive part of $E_{rep}^{opt-PIMD}$. This further demonstrates the sensitivity of the results to the shape of the repulsive potential in the 1.4-1.8 Å range.

E'^{opt}_{rep} Potential and WMull Scheme. As mentioned in section 2, we also optimized the O-H repulsive potential restricting the modifications to distances larger than 1.73 Å in order to avoid the difficulties discussed above. This potential is referred to E'^{opt}_{rep}. E'^{opt}_{rep} avoids the non-physical attractive part of E_{rep}^{opt} and, as will be explained in section 3.2, restores the proton energy transfer barrier. Figure 4 reports only results for t_{OH}=0.28 but the values 0.0, 0.1, 0.2, 0.28, 0.3, 0.4 and 0.5 were also tested. The corresponding curves are provided in the SI (Figure S5, S6 and S7). It is worth mentioning that no optimization of E'^{opt}_{rep} was performed using PIMD simulations. Indeed, as the main effect of including NQEs in

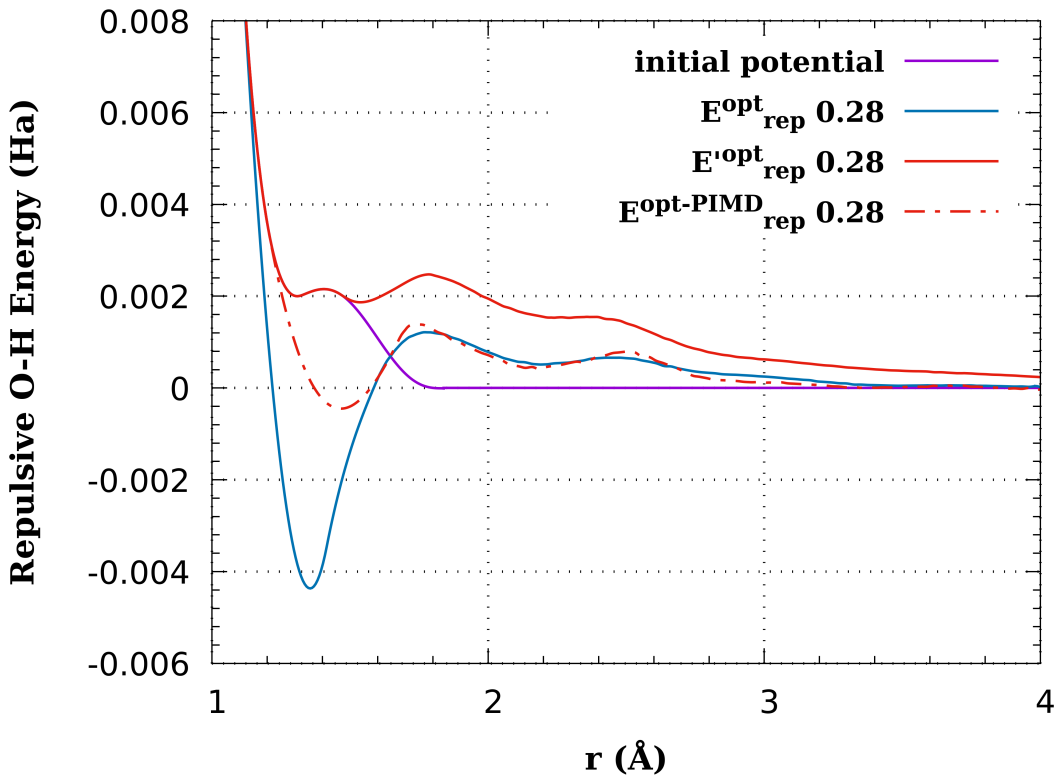

Figure 3: Optimized repulsive O-H potentials E_{rep}^{opt} (blue curve), $E_{rep}^{opt-PIMD}$ (dashed red curve) and E'^{opt}_{rep} (red curve) obtained within the WMull scheme with t_{OH}=0.28. The initial repulsive potential E_{rep} is drawn in purple.

the optimisation process is to correct spurious behaviours in the 1.4-1.8 Å range, it is not relevant to consider it for E'^{opt}_{rep}.

Overall, considering this second range of fitting of the repulsive potential, the IBI process leads to similar effects on the $g(r)$ curves as seen on Figure 4. First, as compared to the WMull scheme alone, the first peak of $g_{OH}(r)$ is less structured and the first minimum is lower. The resulting $g_{OH}(r)$ curve agrees very well with the experimental one except for distances below 1.6 Å. For $g_{OO}(r)$, the structuration at medium and long distance is increased compared to the WMull scheme and the overstructuration at short distance is damped. The $g_{OO}(r)$ curve displays similar defects as the one obtained with E^{opt}_{rep} in the 3.2-5.0 Å range. The agreement between the theoretical and experimental $g_{HH}(r)$ curves is much better as compared to E^{opt}_{rep} as E'^{opt}_{rep} even improves the results obtained with the WMull scheme alone. This demonstrates that the modification of the repulsive O-H potential at short distances deteriorates in turn $g_{HH}(r)$. As seen in Figures S5, S6 and S7 in the SI, the same results are obtained whatever the t_{OH} value.

As seen on the bottom of Figure 2, E'^{opt}_{rep} is more repulsive than the original potential over all the fitting range. In particular, it displays a stronger repulsive contribution in between 1.6-2.0 Å, which becomes stronger when t_{OH} increases. This potential does not differ from the original SCC-DFTB potential in the short distance range involved in the proton transfer energy barrier. Consequently, it does not contain any attractive part and is overall more repulsive than E^{opt}_{rep}.

3.2 Other Properties

In order to assess the quality of the newly proposed O-H repulsive potentials (E^{opt}_{rep}, $E^{opt-PIMD}_{rep}$ and E'^{opt}_{rep}) in combination with WMull charges, they were applied to the calculation of additional properties. We considered the oxygen-oxygen-oxygen triplet angular distribution function $P_{OOO}(\theta)$ (see Figure 5), the self-diffusion coefficient (see Table 1), the heat of vaporization (see Table 2) and the proton transfer energy barrier (see Figure 6).

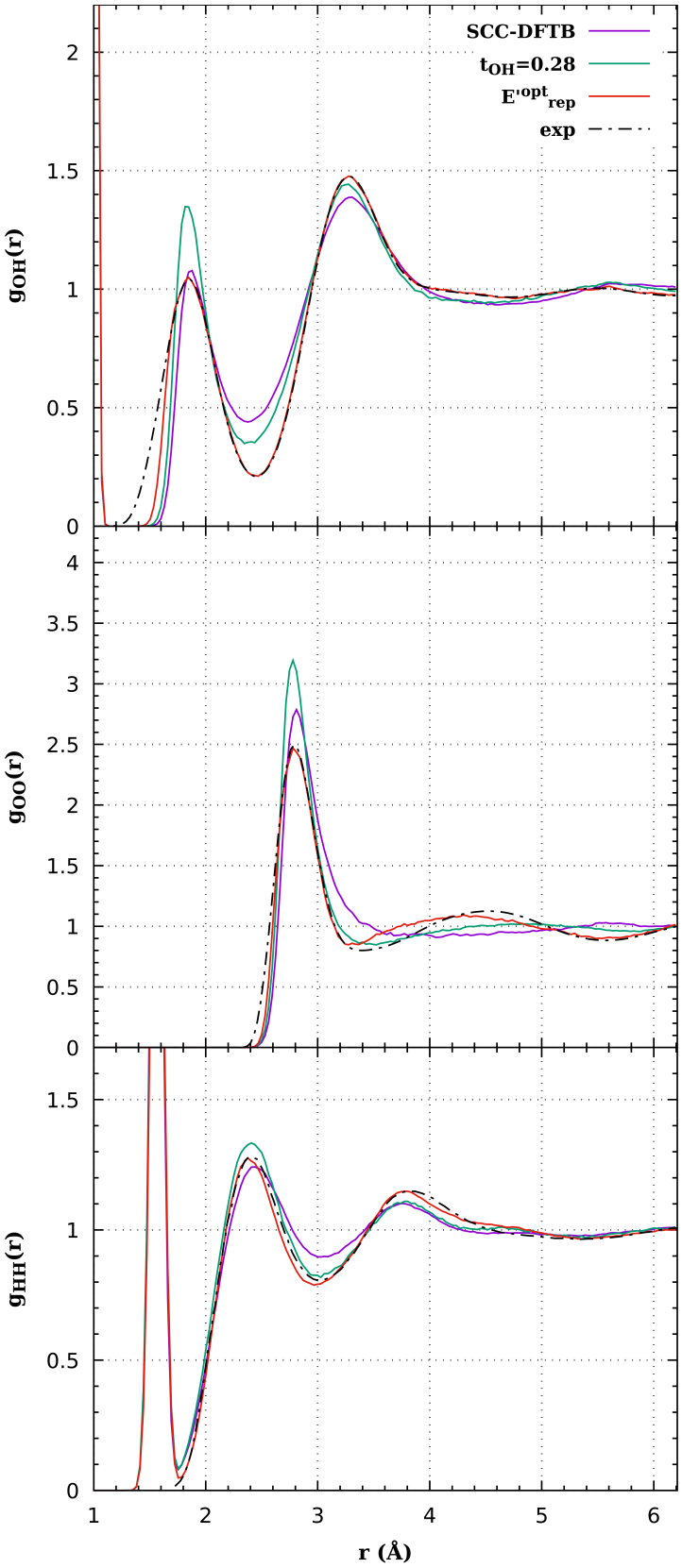

Figure 4: $g_{OH}(r)$ (top), $g_{OO}(r)$ (middle), $g_{HH}(r)$ (bottom) obtained at 300 K for classical liquid water with SCC-DFTB, WMull scheme with t_{OH}=0.28, WMull scheme with t_{OH}=0.28 and E'$^{opt}_{rep}$ compared to experimental results from Soper.[69]

Oxygen-Oxygen-Oxygen Triplet Angular Distribution Function. As can be seen in Figure 5, the experimental oxygen-oxygen-oxygen triplet angular distribution function, $P_{OOO}(\theta)$,[80] shows a broad peak centred at ∼101°. This evidences that water molecules undergo an overall tetrahedral environment in their first solvation shell although this local tetrahedral network is more distorted in liquid water than in crystalline ice. The distribution also displays a characteristic shoulder at ∼60° which originates from highly distorted hydrogen bond configurations in the first coordination shell of the water molecules. For the sake of completeness, $P_{OOO}(\theta)$ obtained at the PBE level and reported by DiStasio *et al.* is also reported in Figure 5 in order to compare the present results to the signature of an overstructured liquid water.[38] So, at the PBE level, the main peak is too close to the perfect tetrahedral angle of 109.5°, the distribution is too narrow, and the shoulder at ∼60° is significantly reduced, which is reminiscent of the overstructuration of liquid water at this level of theory.[81] In contrast, at the original SCC-DFTB level, the distribution is too wide and the local tetrahedral network is too disordered. By considering the WMull scheme only with t_{OH}=0.28, the network is more structured but $P_{OOO}(\theta)$ is still too wide as compared to the experimental one. However, the combination of E_{rep}^{opt} and WMull charges leads to a $P_{OOO}(\theta)$ in very good agreement with the experiment. Indeed, the maximum of the distribution is at the correct position, the shoulder at ∼60° is well reproduced and the two curves, although not perfectly equal, are very close to each other over all the angular range. The same behaviour is obtained considering the E'^{opt}_{rep} potential that leads to a slightly wider distribution with a reduced main peak. The $E_{rep}^{opt-PIMD}$ potential combined with the WMull scheme and a PIMD simulation leads to a $P_{OOO}(\theta)$ curve that is very close to the one obtained with E_{rep}^{opt} and a classical MD simulation. Overall, the three potentials lead to a significant improvement with respect to original SCC-DFTB with distributions that are quantitatively close to the experimental one. This statement is also true for other t_{OH} values that all lead to similar $P_{OOO}(\theta)$ curves as can be seen in Figure S8 in the SI. This demonstrates that optimizing the O-H repulsive potential based on RDFs reference also improves three-body correlations and

the tetrahedrality around water molecules. This can be further demonstrated by looking at the tetrahedral order parameter q calculated as:

$$q = 1 - \frac{3}{8}\sum_{i=1}^{3}\sum_{j=i+1}^{4}\left(\cos\theta_{ij} + \frac{1}{3}\right)^2 \tag{9}$$

The experimental q value is 0.576,[80] and the PBE value reported by DiStasio *et al.* is 0.78.[38] As expected, the original SCC-DFTB q value is 0.27, reminiscent of the lake of structuration at this level of theory, and 0.31 when considering WMull only (t_{OH}=0.28). When E_{rep}^{opt}, E'^{opt}_{rep} and $E_{rep}^{opt-PIMD}$ are considered in combination with WMull, the q value becomes equal to 0.55, 0.54 and 0.57, respectively, so quite close to the experimental value.

Self-Diffusion Coefficient. The self-diffusion coefficient values (D) drop when using WMull charges as compared to the original SCC-DFTB potential (1.1 $Å^2.ps^{-1}$, see Table 1) and this is true with or without considering an optimized O-H repulsive potential. This is in line with the increase in the structuring of the $g_{OO}(r)$ curve at short and medium distances (see Figure 1), which leads to an enhanced difficulty for the molecules to move. The effect of the optimized O-H repulsive potential is more difficult to rationalize. Indeed, as can be noticed in Table 1, variation of D values does not follow any law as a function of t_{OH}. This shows that the specific optimization of the O-H repulsive potential for each t_{OH} removes any direct correlation between t_{OH} and D. In spite of this, when considering SCC-DFTB with WMull only (t_{OH}=0.28), D falls to 0.50 $Å^2.ps^{-1}$, *i.e.* slightly more than twice the experimental value (0.23 $Å^2.ps^{-1}$).[79] When considering E_{rep}^{opt}, this value decreases in all cases but one (t_{OH}=0.1, D=0.58 $Å^2.ps^{-1}$) with D values ranging from 0.34 to 0.42 $Å^2.ps^{-1}$, *i.e.* higher but closer to the experimental value. This highlights that, similarly to $P_{OOO}(\theta)$, an improvement of the RDFs results in a slight improvement of the D value although a perfect agreement is hardly reachable in that case. This statement can be further supported by the work of Lourenço *et al.* which optimized O-O and O-H potentials lead to D values about twice lower than the experimental one at 298 K, despite $g_{OO}(r)$ and $g_{OH}(r)$ curves close to the experimental ones.[60] More importantly, in this study, considering optimized O-O and O-

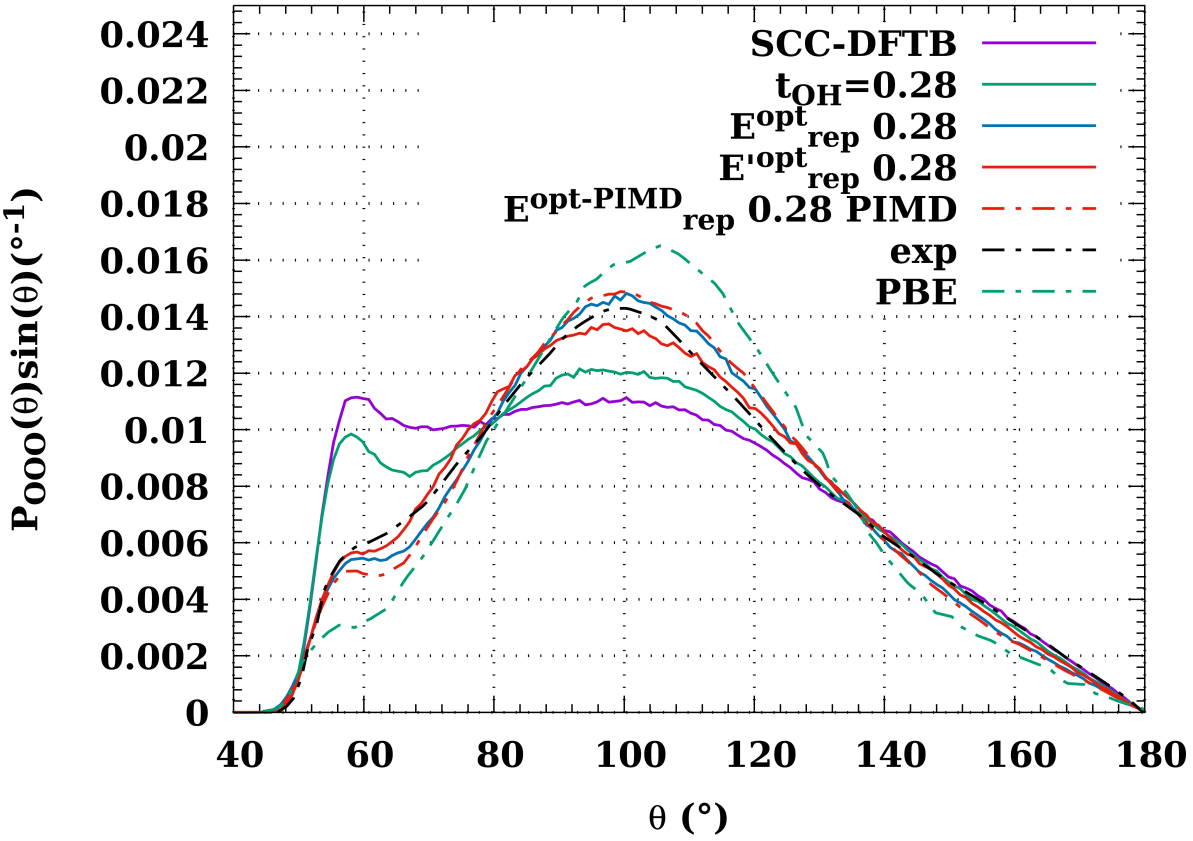

Figure 5: The oxygen-oxygen-oxygen triplet angular distribution function $P_{OOO}(\theta)$ of liquid water obtained with original SCC-DFTB (purple), WMull charges (t_{OH}=0.28) with the original repulsive O-H potential (green), WMull charges (t_{OH}=0.28) in combination with E^{opt}_{rep} (blue), E'^{opt}_{rep} (plain red) and $E^{opt-PIMD}_{rep}$ within a PIMD simulation (dashed red). DFT results using the PBE functional (dashed green)[38] and experimental results (dashed black)[80] are also reported for comparison.

H potentials do not significantly modify the D value obtained with the original SCC-DFTB formulation. Considering E'^{opt}_{rep} leads to even smaller modifications of D with values ranging from 0.42 to 0.56 $Å^2.ps^{-1}$. The D value obtained with $E^{opt-PIMD}_{rep}$ in combination with a classical MD simulation falls in the range of values obtained with E^{opt}_{rep}. Overall, considering the standard deviations we obtained, all three optimized potentials lead to similar D values which highlights the difficulty to reach perfect agreement with the experiment using our optimization procedure only. Furthermore, finite size effects have not been taken into account here, although they can be important for the computation of self-diffusion coefficients. A more detailed study on this property would thus necessitate simulations on larger unit cells.

Table 1: Self-diffusion coefficients in $Å^2.ps^{-1}$ determined for liquid water at 300 K. Mean values D and standard deviations σ obtained for t_{OH} values between 0.0 and 0.5 in combination with E^{opt}_{rep} (bold) and E'^{opt}_{rep} (italic). Values obtained with $E^{opt-PIMD}_{rep}$ in combination with WMull charges (t_{OH}=0.28) within a classical MD simulation is also provided[a] as well as values for WMull charges (t_{OH}=0.28) only. Original SCC-DFTB[b],[58] and experimental[c],[79] values are also reported for comparison.

Methods	D	σ
0.0	**0.40**/*0.51*	**0.13**/*0.15*
0.1	**0.58**/*0.56*	**0.11**/*0.10*
0.2	**0.42**/*0.42*	**0.11**/*0.08*
0.28	**0.39**/*0.49*	**0.11**/*0.07*
0.3	**0.34**/*0.53*	**0.09**/*0.14*
0.4	**0.35**/*0.44*	**0.07**/*0.06*
0.5	**0.40**/*0.43*	**0.09**/*0.08*
$E^{opt-PIMD}_{rep}$ [a]	0.39	0.06
SCC-DFTB/0.28	0.50	0.07
SCC-DFTB[b]	1.1	0.04
Exp.[c]	0.23	

Heat of Vaporization. Obtaining an accurate value for the heat of vaporization, ΔH_{vap}, of liquid water remains an issue at the SCC-DFTB level as previously mentioned by Goyal *et al.*.[65] As can be seen in Table 2, although the experimental value was measured to be 10.50 $kcal.mol^{-1}$, the SCC-DFTB value computed with the original hamiltonian amounts to only 4.09 $kcal.mol^{-1}$. Considering the WMull charges with t_{OH}=0.28 leads to an increase of

ΔH_{vap} up to 8.13 kcal.mol^{-1}. Optimizing in addition the O-H repulsive potential systematically leads to a decrease of ΔH_{vap}. For instance, considering E_{rep}^{opt} in combination with t_{OH}=0.28, ΔH_{vap} falls to 6.12 kcal.mol^{-1}. Overall, when using both WMull charges and E_{rep}^{opt}, the increase of t_{OH} leads to an increase of ΔH_{vap} even though for the highest considered t_{OH} value, t_{OH}=0.5, ΔH_{vap} remains lower than 8.13 kcal.mol^{-1}. Considering $E_{rep}^{opt-PIMD}$ instead of E_{rep}^{opt} only slightly improves the agreement with the experiment as ΔH_{vap} increases by only 0.51 kcal.mol^{-1}. ΔH_{vap} values computed using E'^{opt}_{rep} are much too low, even leading to a non cohesive liquid for t_{OH} values smaller than 0.2. The decrease of ΔH_{vap} optimizing the repulsive O-H potential can be explained by the shape of the latter. Indeed, E_{rep}^{opt} is much more repulsive than the original potential (see Figure 3) which thus promotes vaporization. This effect is even more pronounced for E'^{opt}_{rep} that is more repulsive than E_{rep}^{opt}. The trends obtained for ΔH_{vap} can be rationalized when looking at the evolution of the water dimer intermolecular binding energy, D_0, computed using the same set of parameters (see Table S1 in the SI). These values are found lower than the experimental value of 3.16 kcal.mol^{-1},[82] and, as expected, increase with t_{OH} due to stronger electrostatic interactions. Similarly, the dipole moment of the isolated water molecule is found to be too low for the lowest t_{OH} value, i.e. 1.58 D, and increases up to 2.21 D when increasing t_{OH}, the experimental value being 1.85 D.[83] More interestingly, the values are the same whatever the considered O-H potential, E_{rep}^{opt}, E'^{opt}_{rep} or $E_{rep}^{opt-PIMD}$. This demonstrates that the dipole moment of the isolated water molecule is only determined by the t_{OH} value while the intermolecular binding energy results from a balance between the actual shape of the O-H repulsive potential and the t_{OH} value. The former being too repulsive in all cases, it can only lead to underestimated values of ΔH_{vap} as was previously highlighted by Goyal and co-workers.[65] In contrast to the other considered properties, D and $P_{OOO}(\theta)$, this shows that the improvement of structural and dynamics properties does not necessarily lead to satisfactory energetics. However, the combination of WMull and E_{rep}^{opt} still improves the results with respect to the original SCC-DFTB hamiltonian.

Table 2: Heat of vaporization in kcal.mol^{-1} **determined at 300 K. Mean values** ΔH **and standard deviations** σ **obtained for** t_{OH} **values between 0.0 and 0.5 in combination with** E_{rep}^{opt} **(bold) and** E'^{opt}_{rep} **(italic). Values obtained with** $E_{rep}^{opt-PIMD}$ **in combination with WMull charges (**t_{OH}**=0.28) within a classical MD simulation is also provided**[a] **as well as values for WMull charges (**t_{OH}**=0.28) only. Original SCC-DFTB**[b]**,**[58] **and experimental**[c]**,**[84] **values are also reported for comparison.**

Methods	ΔH_{vap}	σ
0.0	**5.07**/*-0.33*	**0.12**/*0.12*
0.1	**4.99**/*-3.73*	**0.12**/*0.08*
0.2	**6.09**/*0.39*	**0.12**/*0.10*
0.28	**6.12**/*0.18*	**0.14**/*0.06*
0.3	**7.09**/*1.98*	**0.12**/*0.11*
0.4	**7.24**/*3.30*	**0.10**/*0.09*
0.5	**7.81**/*4.34*	**0.08**/*0.09*
$E_{rep}^{opt-PIMD}$ [a]	6.63	0.05
SCC-DFTB/0.28	8.13	0.07
SCC-DFTB[b]	4.09	
Exp.[c]	10.50	

Proton Transfer Energy Barrier. Finally, a static property that needs to be well described as soon as reactivity in liquid water is considered is the proton transfer energy barrier (PTEB). Figure 6 shows that the PTEB obtained with the Mulliken description of charges is highly underestimated compared to MP2/aug-cc-pVTZ with respective values 0.16 and 1.90 kcal.mol^{-1}. Moreover, the Mulliken curve displays a plateau-like shape which is not the case at the MP2 level. Using improved WMull charges was shown to increase this barrier with respect to original SCC-DFTB.[61] It rises to 1.00 kcal.mol^{-1} but the plateau-like shape remains. Considering E_{rep}^{opt} leads to the absence of barrier. This can be explained because the O-H distances involved are between 1.0 and 1.6 Å which corresponds to the distances involved in the strongly attractive part of the optimized potential. Other considered properties: $P_{OOO}(\theta)$, D and ΔH_{vap}, were not impacted by this spurious behaviour of E_{rep}^{opt}, but it strongly influence the PTEB. This highlights the difficulty to optimize the potential to fit structural properties while preserving local properties such as the PTEB. In contrast, $E_{rep}^{opt-PIMD}$ leads to the best PTEB. Indeed, it does not only re-establish the PTEB as compared to E_{rep}^{opt},

but also provides the highest value and best estimate: 1.13 kcal.mol^{-1}. It also corrects the plateau-like shape. This better behaviour comes from the strongly attenuated attractive part of the $E_{rep}^{opt-PIMD}$ potential compared to E_{rep}^{opt}. Finally, the E'^{opt}_{rep} curve is similar to the one obtained with WMull only (t_{OH}=0.28) as, by construction, both repulsive potentials are identical at short distances involved in the proton transfer.

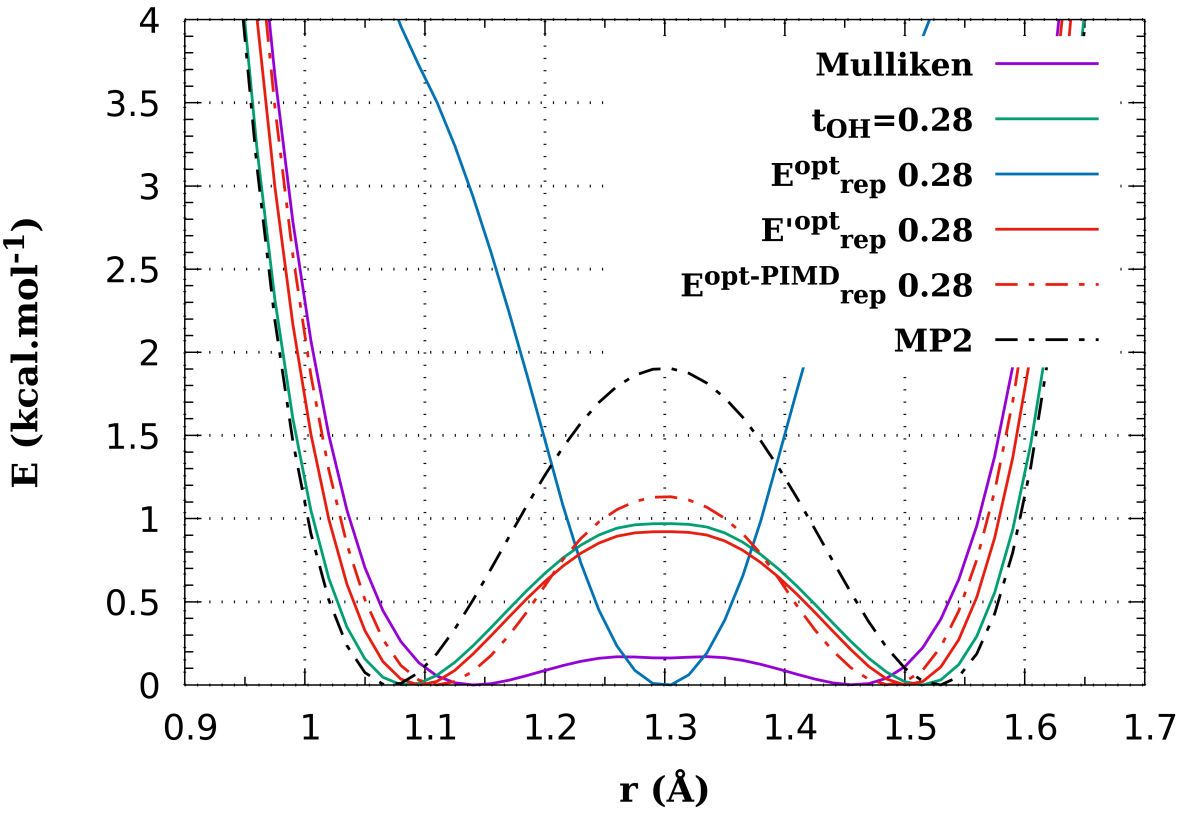

Figure 6: Proton transfer energy barrier obtained with WMull scheme t_{OH}=0.28 only (green), WMull t_{OH}=0.28 in combination with E_{rep}^{opt} (blue), $E_{rep}^{opt-PIMD}$ (dashed red) and E'^{opt}_{rep} (red). The curves obtained at the MP2/aug-cc-pVTZ (dashed black) and original SCC-DFTB (violet) level are provided for comparison.

4 Conclusion

In this work, we combined the use of improved atomic charges and optimized O-H repulsive potential through the Iterative Boltzmann Inversion process in order to improve the

description of liquid water using the SCC-DFTB formalism. The use of WMull charges enables a better description of bond polarisation and its main macroscopic effect is to increase the structuring of the $g_{OO}(r)$ RDF, leading to a better agreement with experimental data. However, this also leads to an over-structuring of the $g_{OO}(r)$ RDF at short distances. This is counteracted combining WMull charges with the IBI process to optimize the O-H repulsive potential, leading to several sets of parameters. The O-H repulsive potential was first optimized including short O-H distances (WMull, E_{rep}^{opt}). This led to very accurate RDFs as well as improvements on the self-diffusion coefficient, oxygen-oxygen-oxygen triplet angular distribution function and heat of vaporization. However it significantly deteriorates of the proton transfer energy barrier due to a spurious attractive contribution in the potential at short distances. Considering the optimization of the O-H repulsive potential at longer distances (WMull, E'^{opt}_{rep}) enables to recover the proton transfer energy barrier but deteriorates the heat of vaporization while $g_{OH}(r)$ is less accurate at short distances. In spite of this, these two sets of potentials improve significantly the overall description of liquid water even though in both cases one property is deteriorated. This highlights the limitation of the SCC-DFTB method to describe liquid water and demonstrates that an optimisation of the O-H repulsive potential only can not fix all the shortcuts of this formalism. Indeed, the shape of the O-H repulsive potentials we obtained displays an attractive part at short distances which highlights that some physical contributions are not properly taken into account, such as atomic polarization, or are taken into account in a fortuitous way, such as nuclear quantum effects.

Regarding the latter, an alternative was to optimize the O-H repulsive potential within the PIMD scheme to exclude the NQEs from the potential. While facilitating the optimisation at short distances, this procedure also greatly limits the appearance of an attractive contribution at short distances. Consequently, in conjunction with the WMull charges (WMull, $E_{rep}^{opt-PIMD}$), this potential provides similar results as the first set of parameters (WMull, E_{rep}^{opt}) for structural, dynamic and thermodynamic properties while recovering a

large part of the proton transfer energy barrier. It therefore offers a good compromise to study both structural and dynamical properties. Further testing of $E_{rep}^{opt-PIMD}$, as well as E_{rep}^{opt} and E'^{opt}_{rep}, would require, among others, simulations in the isothermal–isobaric (NPT) ensemble to determine their ability to reproduce the density of liquid water at ambient and higher temperatures. Again, this study highlights the tremendous difficulty to find a unique potential to describe at the same time structural, energetic, dynamic and thermodynamic properties of liquid water but satisfying compromises can be achieved as in the case of $E_{rep}^{opt-PIMD}$. Future studies using these newly developed potentials will involve the interface between liquid water and an inorganic solvent such as toluene which is of interest in petroleum industry.

Supporting Information Available

Supporting Information contains $g_{OO}(r)$, $g_{OH}(r)$, and $g_{HH}(r)$ RDFs, oxygen-oxygen-oxygen triplet angular distribution functions $P_{OOO}(\theta)$, and proton transfer energy barriers obtained with E_{rep}^{opt} and E'^{opt}_{rep} potentials in combination with varying values of t_{OH} (0.0, 0.1, 0.2, 0.28, 0.3, 0.4, 0.5). It also contains the intermolecular binding energies and zero-point energy corrections for the water dimer in the gas phase as well as the dipole moment of the water monomer obtained with E_{rep}^{opt} and E'^{opt}_{rep} potentials in combination with varying values of t_{OH} (0.0, 0.1, 0.2, 0.28, 0.3, 0.4, 0.5). Finally, it contains $g_{OH}(r)$ RDFs obtained with $E_{rep}^{opt-PIMD}$ potential within classical and path-integral molecular dynamics simulations.

Supporting Information is available free of charge.

Acknowledgement

The authors acknowledge the computing mesocenter CALMIP ("CALcul en MIdi Pyrénées", UAR3667 of CNRS) for generous allocation of computer resources (projects P17002 and P0059). This work has been funded by the French Agence Nationale de la Recherche (ANR)

project MUSCOFI ANR-18-CE05-0028. For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (version arising from this submission).

References

(1) Gillan, M. J.; Alfè, D.; Bartók, A. P.; Csányi, G. First-principles energetics of water clusters and ice: A many-body analysis. J. Chem. Phys. **2013**, 139, 244504.

(2) Pettersson, L. G. M.; Henchman, R. H.; Nilsson, A. Water—The Most Anomalous Liquid. Chem. Rev. **2016**, 116, 7459–7462.

(3) Gallo, P.; Amann-Winkel, K.; Angell, C. A.; Anisimov, M. A.; Caupin, F.; Chakravarty, C.; Lascaris, E.; Loerting, T.; Panagiotopoulos, A. Z.; Russo, J. et al. Water: A Tale of Two Liquids. Chem. Rev. **2016**, 116, 7463–7500.

(4) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. Chem. Rev. **2016**, 116, 7501–7528.

(5) Fransson, T.; Harada, Y.; Kosugi, N.; Besley, N. A.; Winter, B.; Rehr, J. J.; Pettersson, L. G. M.; Nilsson, A. X-ray and Electron Spectroscopy of Water. Chem. Rev. **2016**, 116, 7551–7569.

(6) Amann-Winkel, K.; Bellissent-Funel, M.-C.; Bove, L. E.; Loerting, T.; Nilsson, A.; Paciaroni, A.; Schlesinger, D.; Skinner, L. X-ray and Neutron Scattering of Water. Chem. Rev. **2016**, 116, 7570–7589.

(7) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear Quantum Effects in Water and Aqueous

Systems: Experiment, Theory, and Current Challenges. Chem. Rev. **2016**, 116, 7529–7550.

(8) Cho, C. H.; Singh, S.; Robinson, G. W. Understanding all of water's anomalies with a nonlocal potential. J. Chem. Phys. **1997**, 107, 7979–7988.

(9) Robinson, G. W.; Cho, C. H.; Urquidi, J. Isosbestic points in liquid water: Further strong evidence for the two-state mixture model. J. Chem. Phys. **1999**, 111, 698–702.

(10) Stanley, H.; Buldyrev, S.; Canpolat, M.; Mishima, O.; Sadr-Lahijany, M.; Scala, A.; Starr, F. The puzzling behavior of water at very low temperature. Phys. Chem. Chem. Phys. **2000**, 2, 1551–1558.

(11) Errington, J. R.; Debenedetti, P. G. Relationship between structural order and the anomalies of liquid water. Nature **2001**, 409, 318–321.

(12) Angell, C. A. Insights into phases of liquid water from study of its unusual glass-forming properties. Science **2008**, 319, 582–587.

(13) Pallares, G.; El Mekki Azouzi, M.; González, M. A.; Aragones, J. L.; Abascal, J. L. F.; Valeriani, C.; Caupin, F. Anomalies in bulk supercooled water at negative pressure. Proc. Natl. Acad. Sci. USA **2014**, 111, 7936–7941.

(14) Perakis, F.; De Marco, L.; Shalit, A.; Tang, F.; Kann, Z. R.; Kühne, T. D.; Torre, R.; Bonn, M.; Nagata, Y. Vibrational Spectroscopy and Dynamics of Water. Chem. Rev. **2016**, 116, 7590–7607.

(15) Dasgupta, S.; Lambros, E.; Perdew, J. P.; Paesani, F. Elevating density functional theory to chemical accuracy for water simulations through a density-corrected many-body formalism. Nat. Commun. **2021**, 12, 6359.

(16) Spura, T.; John, C.; Habershon, S.; Kühne, T. D. Nuclear quantum effects in liquid

water from path-integral simulations using an ab initio force-matching approach. Mol. Phys. **2015**, 113, 808–822.

(17) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. **1983**, 79, 926–935.

(18) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J. Chem. Phys. **2000**, 112, 8910–8922.

(19) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. **2004**, 120, 9665–9678.

(20) Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. Phys. Chem. Chem. Phys. **2011**, 13, 19663.

(21) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. J. Phys. Chem. **1987**, 91, 6269–6271.

(22) Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters (n=2–21) and the phonon spectrum of ice Ih. J. Chem. Phys. **1999**, 110, 4566–4581.

(23) Burnham, C.; Xantheas, S. Development of Transferable Interaction Models for Water: III. Reparameterization of an All-Atom Polarizable Rigid Model (TTM2-R) from First Principles. J. Chem. Phys. **2002**, 116, 1500–1510.

(24) Fanourgakis, G. S.; Xantheas, S. S. The flexible, polarizable, thole-type interaction potential for water (TTM2-F) revisited. J. Phys. Chem. A **2006**, 110, 4100–4106.

(25) Fanourgakis, G. S.; Xantheas, S. S. Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable, Thole-type model potential (TTM3-F , v. 3.0) to describe the vibrational spectra of water clusters and liquid water. J. Chem. Phys. **2008**, 128, 074506.

(26) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. J. Phys. Chem. B **2003**, 107, 5933–5947.

(27) Ren, P.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. J. Phys. Chem. B **2004**, 108, 13427–13437.

(28) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. J. Phys. Chem. B **2013**, 117, 9956–9972.

(29) Xiong, Y.; Izadi, S.; Onufriev, A. V. Fast Polarizable Water Model for Atomistic Simulations. J. Chem. Theory Comput. **2022**, 18, 6324–6333.

(30) Lambros, E.; Paesani, F. How Good are Polarizable and Flexible Models for Water: Insights from a Many-Body Perspective. J. Chem. Phys. **2020**, 153, 060901.

(31) Lamoureux, G.; MacKerell, J., Alexander D.; Roux, B. A Simple Polarizable Model of Water based on Classical Drude Oscillators. J. Chem. Phys. **2003**, 119, 5185–5197.

(32) Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. J. Chem. Theory Comput. **2013**, 9, 5395–5403, PMID: 26592277.

(33) Babin, V.; Medders, G. R.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. J. Chem. Theory Comput. **2014**, 10, 1599–1607, PMID: 26580372.

(34) Medders, G. R.; Babin, V.; Paesani, F. Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. J. Chem. Theory Comput. **2014**, 10, 2906–2910, PMID: 26588266.

(35) Bore, S. L.; Paesani, F. Realistic Phase Diagram of Water from "first principles" Data-Driven Quantum Simulations. Nat. Commun. **2023**, 14, 3349.

(36) Jonchiere, R.; Seitsonen, A. P.; Ferlat, G.; Saitta, A. M.; Vuilleumier, R. Van der Waals effects in ab initio water at ambient and supercritical conditions. J. Chem. Phys. **2011**, 135, 154503.

(37) Zhang, C.; Wu, J.; Galli, G.; Gygi, F. Structural and Vibrational Properties of Liquid Water from van der Waals Density Functionals. J. Chem. Theory Comput. **2011**, 7, 3054–3061.

(38) DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. J. Chem. Phys. **2014**, 141, 084502.

(39) Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. Molecular Dynamics Simulation of Liquid Water Hybrid Density Functionals. J. Phys. Chem. B **2006**, 110, 3685–3691.

(40) Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, J. Ab initio molecular dynamics using hybrid density functionals. J. Chem. Phys. **2008**, 128, 214104.

(41) Chen, B.; Ivanov, I.; Klein, M. L.; Parrinello, M. Hydrogen Bonding in Water. Phys. Rev. Lett. **2003**, 91, 215503.

(42) Morrone, J. A.; Car, R. Nuclear Quantum Effects in Water. Phys. Rev. Lett. **2008**, 101, 017801.

(43) Fritsch, S.; Potestio, R.; Donadio, D.; Kremer, K. Nuclear Quantum Effects in Water: A Multiscale Study. J. Chem. Theory Comput. **2014**, 10, 816–824.

(44) Marsalek, O.; Markland, T. E. Ab initio molecular dynamics with nuclear quantum effects at classical cost: Ring polymer contraction for density functional theory. J. Chem. Phys. **2016**, 144, 054112.

(45) Marx, D.; Parrinello, M. Ab initio path-integral molecular dynamics. Z. Phys. B: Condens. Matter **1994**, 95, 143–144.

(46) Marx, D.; Parrinello, M. Ab initio path integral molecular dynamics: Basic ideas. J. Chem. Phys. **1996**, 104, 4077–4082.

(47) Habershon, S.; Braams, B. J.; Manolopoulos, D. E. Quantum mechanical correlation functions, maximum entropy analytic continuation, and ring polymer molecular dynamics. J. Chem. Phys. **2007**, 127, 174108.

(48) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. J. Chem. Phys. **2009**, 131, 024501.

(49) Zheng, L.; Chen, M.; Sun, Z.; Ko, H.-Y.; Santra, B.; Dhuvad, P.; Wu, X. Structural, electronic, and dynamical properties of liquid water by *ab initio* molecular dynamics based on SCAN functional within the canonical ensemble. J. Chem. Phys. **2018**, 148, 164505.

(50) Zhang, C.; Tang, F.; Chen, M.; Xu, J.; Zhang, L.; Qiu, D. Y.; Perdew, J. P.; Klein, M. L.; Wu, X. Modeling Liquid Water by Climbing up Jacob's Ladder in Density Functional Theory Facilitated by Using Deep Neural Network Potentials. J. Phys. Chem. B **2021**, 125, 11444–11456.

(51) Liu, J.; Lan, J.; He, X. Toward High-level Machine Learning Potential for Water Based

on Quantum Fragmentation and Neural Networks. J. Phys. Chem. A **2022**, 126, 3926–3936.

(52) Cheng, B.; Engel, E. A.; Behler, J.; Dellago, C.; Ceriotti, M. Ab initio thermodynamics of liquid and solid water. Proc. Natl. Acad. Sci. USA **2019**, 116, 1110–1115.

(53) Gartner, T. E.; Zhang, L.; Piaggi, P. M.; Car, R.; Panagiotopoulos, A. Z.; Debenedetti, P. G. Signatures of a liquid–liquid transition in an ab initio deep neural network model for water. Proc. Natl. Acad. Sci. USA **2020**, 117, 26040–26046.

(54) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys. Rev. B **1998**, 58, 7260–7268.

(55) Frauenheim, T.; Seifert, G.; Elstern, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. A Self-Consistent Charge Density-Functional Based Tight-Binding Method for Predictive Materials Simulations in Physics, Chemistry and Biology. Phys. Status Solidi (b) **2000**, 217, 41–62.

(56) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. Phys. Rev. B **1995**, 51, 12947–12957.

(57) Elstner, M. The SCC-DFTB method and its application to biological systems. Theor. Chem. Acc. **2006**, 116, 316–325.

(58) Maupin, C. M.; Aradi, B.; Voth, G. A. The Self-Consistent Charge Density Functional Tight Binding Method Applied to Liquid Water and the Hydrated Excess Proton: Benchmark Simulations. J. Phys. Chem. B **2010**, 114, 6922–6931.

(59) Goyal, P.; Elstner, M.; Cui, Q. Application of the SCC-DFTB Method to Neutral and Protonated Water Clusters and Bulk Water. J. Phys. Chem. B **2011**, 115, 6790–6805.

(60) Lourenço, M. P.; dos Santos, E. C.; Pettersson, L. G. M.; Duarte, H. A. Accurate SCC-DFTB Parametrization for Bulk Water. J. Chem. Theory Comput. **2020**, 16, 1768–1778.

(61) Cuny, J.; Cerda Calatayud, J.; Ansari, N.; Hassanali, A. A.; Rapacioli, M.; Simon, A. Simulation of Liquids with the Tight-Binding Density-Functional Approach and Improved Atomic Charges. J. Phys. Chem. B **2020**, 124, 7421–7432.

(62) Doemer, M.; Liberatore, E.; Knaup, J. M.; Tavernelli, I.; Rothlisberger, U. *In situ* parameterisation of SCC-DFTB repulsive potentials by iterative Boltzmann inversion. Mol. Phys. **2013**, 111, 3595–3607.

(63) Michoulier, E.; Ben Amor, N.; Rapacioli, M.; Noble, J. A.; Mascetti, J.; Toubin, C.; Simon, A. Theoretical determination of adsorption and ionisation energies of polycyclic aromatic hydrocarbons on water ice. Phys. Chem. Chem. Phys. **2018**, 20, 11941–11953.

(64) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). J. Chem. Theory Comput. **2011**, 7, 931–948.

(65) Goyal, P.; Qian, H.-J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. Molecular Simulation of Water and Hydration Effects in Different Environments: Challenges and Developments for DFTB Based Models. J. Phys. Chem. B **2014**, 118, 11007–11027.

(66) Rapacioli, M.; Spiegelman, F.; Talbi, D.; Mineva, T.; Goursot, A.; Heine, T.; Seifert, G. Correction for dispersion and Coulombic interactions in molecular clusters with density functional derived methods: Application to polycyclic aromatic hydrocarbon clusters. J. Chem. Phys. **2009**, 130, 244304.

(67) Simon, A.; Rapacioli, M.; Michoulier, E.; Zheng, L.; Korchagina, K.; Cuny, J. Contribution of the density-functional-based tight-binding scheme to the description of water

clusters: methods, applications and extension to bulk systems. Mol. Simul. **2019**, 45, 249–268.

(68) Wu, Q.; Yang, W. Empirical correction to density functional theory for van der Waals interactions. J. Chem. Phys. **2002**, 116, 515–524.

(69) Soper, A. K. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? ISRN Physical Chemistry **2013**, 2013, 1–67.

(70) McQuarrie, Donald A. Statistical Mechanics; University Science Books: Sausalito, CA, 2000; Vol. 1.

(71) Rapacioli, M.; Heine, T.; Dontot, L.; Yusef Buey, M.; Louisnard, F.; Cuny, J.; Morinière, M.; Dubosq, C.; Patchkovskii, J., S.and Frenzel; Michoulier, E. et al. 2023 deMonNano experiment. `http://demon-nano.ups-tlse.fr/`, 2023/07/04.

(72) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. J. Chem. Phys. **1992**, 97, 2635–2643.

(73) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A **1985**, 31, 1695–1697.

(74) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. J. Chem. Phys. **1984**, 81, 511–519.

(75) Ceriotti, M.; More, J.; Manolopoulos, D. E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. Comput. Phys. Commun. **2014**, 185, 1019–1026.

(76) Ceriotti, M.; Bussi, G.; Parrinello, M. Nuclear Quantum Effects in Solids Using a Colored-Noise Thermostat. Phys. Rev. Lett. **2009**, 103, 030603.

(77) Ceriotti, M.; Manolopoulos, D. E. Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei. Phys. Rev. Lett. **2012**, 109, 100604.

(78) GLE4MD Project Homepage: colored noise thermostats for molecular dynamics. `http://gle4md.org/`, 2023/03/01.

(79) Mills, R. Self-diffusion in normal and heavy water in the range 1-45.deg. J. Phys. Chem. **1973**, 77, 685–688.

(80) Soper, A. K.; Benmore, C. J. Quantum Differences between Heavy and Light Water. Phys. Rev. Lett. **2008**, 101, 065502.

(81) Sit, P. H.-L.; Marzari, N. Static and dynamical properties of heavy water at ambient conditions from first-principles molecular dynamics. J. Chem. Phys. **2005**, 122, 204510.

(82) Rocher-Casterline, B. E.; Ch'ng, L. C.; Mollner, A. K.; Reisler, H. Communication: Determination of the bond dissociation energy (D_0) of the water dimer, $(H_2O)_2$, by velocity map imaging. J. Chem. Phys. **2011**, 134, 211101.

(83) Dyke, T. R.; Muenter, J. S. Electric dipole moments of low J states of H2O and D2O. J. Chem. Phys. **2003**, 59, 3125.

(84) Marsh, K. N., Ed. Recommended reference materials for the realization of physicochemical properties; Blackwell: Oxford, 1987.

TOC Graphic