Lessons learned from multi-objective automatic optimizations of classical three-site rigid water models using microscopic and macroscopic target experimental observables

Mattia Perrone,[†] Riccardo Capelli,[‡] Charly Empereur-mot,[¶] Ali Hassanali,[§] and Giovanni M. Pavan[*,†,¶]

†*Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy*

‡*Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, I-20133 Milano, Italy*

¶*Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Polo Universitario Lugano, Campus Est, Via la Santa 1, CH-6962 Lugano-Viganello, Switzerland*

§*The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy.*

E-mail: giovanni.pavan@polito.it

Abstract

The development of accurate water models is of primary importance for molecular simulations. Despite their intrinsic approximations, three-site rigid water models are still ubiquitously used to simulate a variety of molecular systems. Automatic optimization approaches have been recently used to iteratively optimize three-site water

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

models to fit macroscopic (average) thermodynamic properties, providing "state-of-the-art" three-site models that still present some deviations from the liquid water properties. Here we show results obtained by automatically optimizing three-site rigid water models to fit a combination of microscopic and macroscopic experimental observables. We use *Swarm-CG*, a multi-objective particle-swarm-optimization algorithm, for training the models to reproduce the experimental radial distribution functions of liquid water at various temperatures (rich in microscopic-level information on, *e.g.*, the local orientation and interactions of the water molecules). We systematically analyze the agreement of these models with experimental observables and the effect of adding macroscopic information into the training-set. Our results demonstrate how adding microscopic-rich information in the training of water models allows achieving state-of-art accuracy in an efficient way. Limitations in the approach and in the approximated description of water in these three-site models are also discussed, providing a demonstrative case useful for the optimization of approximated molecular models in general.

Introduction

The development and optimization of classical molecular models is typically challenging and time-consuming.[1,2] Despite notable progresses in developing efficient methods and optimization approaches,[3–8] accurately predicting experimental observables and ensuring transferability across varying thermodynamics conditions remains in most cases a significant challenge.[9–11] A considerable example is the case of water, for which current state-of-art models struggle in matching all the relevant cases of interest at the same time,[12] *e.g.*, bulk properties,[13] free energy of hydration of compounds,[14] stabilization of lipid membranes,[15] interaction with proteins,[16] *etc.*

Although intrinsically approximated, classical three-site rigid water models are widely used in molecular dynamics (MD) simulations.[13] One key requirement is that such simplified models can capture fairly well the properties of water even relying on a reduced number of

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

parameters. In such a representation, the interaction potential is centered on three sites, each of which corresponds to one of the atoms in the water molecule (O, H, H). Early versions of these models, including, *e.g.*, TIP3P[17] and SPC,[18] were originally parameterized to accurately reproduce basic thermodynamic properties, *e.g.*, density and enthalpy of vaporization under standard conditions. Despite their age, these models continue to be extensively utilized in classical MD simulations, and most general-purpose forcefields are parametrized on them.[19–21] With the increase in computing power, it has become possible to perform high-throughput parameterization, often in an automatic fashion,[22,23] by considering a large set of experimental observables under different conditions as the reference data to fit.

Over the past decade, two notable general-purpose three-site water models that have been obtained through iterative optimization, TIP3P-FB[24] and OPC3,[25] led to a substantial improvement of the state-of-the-art. Such models were refined to accurately reproduce a set of thermodynamic properties including density, heat of vaporization, coefficient of thermal expansion, isothermal compressibility, isobaric heat capacity, and static dielectric constant. In particular, TIP3P-FB has been optimized to accurately reproduce these observables over a wide range of thermodynamic conditions, spanning a total of 40 training points at different temperatures and pressures. Such a parallel/multi-objective parametrization has a positive effect on the transferability of the optimized model,[26] *e.g.*, across different conditions. In contrast, OPC3 was optimized to match such observables under standard conditions (298 K and 1 bar), while simultaneously imposing a constraint on the geometry of the water molecule. Specifically, a fixed Hydrogen-Oxygen-Hydrogen angle value is imposed to ensure that the resulting linear quadrupole moment is equal to zero. This constraint is applied because the quadrupole moment is known to have minimal significance in the context of the model's overall performance and accuracy.[27] While both models have demonstrated similar accuracy in reproducing thermodynamic properties, they do exhibit some distinct characteristics. TIP3P-FB is characterized by a larger geometry, with a distance of 0.101 nm between the Oxygen and Hydrogen sites (d_{OH}) and 0.164 nm between the Hydrogen sites (d_{HH}).

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Furthermore, the Oxygen site in TIP3P-FB carries a partial charge of -0.848 e. In contrast, OPC3 exhibits a smaller geometry, with d_{OH} and d_{HH} equal to 0.098 nm and 0.160 nm respectively. Additionally, the Oxygen site in OPC3 has a charge of -0.895 e. The variability observed between the optimized models may be attributed to an intrinsic limitation arising from the simplified description of the system. Furthermore, as both models are trained solely on average parameters derived from a top-down approach, it becomes intriguing to explore the potential advantages of integrating additional data on microscopic target features through a bottom-up approach.

In recent works, we introduced *Swarm-CG*,[7,9] a versatile optimization software that is able to integrate bottom-up and top-down references in a multi-objective and multi-directional optimization framework for coarse-grained models. Building upon *Swarm-CG*'s capabilities, we propose a novel strategy for optimizing three-site water models by incorporating experimental data on the microscopic structure of water, particularly the radial distribution functions (RDF) of its atoms. Specifically, we utilize the Oxygen-Oxygen RDF (g_{OO}), Oxygen-Hydrogen RDF (g_{OH}), and Hydrogen-Hydrogen RDF (g_{HH}) as the primary references for deriving our model. While our main objective is not to develop the most accurate three-site rigid model, we aim to explore the capabilities of *Swarm-CG* and assess the room for improvement in what can be considered *de facto* a coarse-grained description of water. The results we obtain are significant for two main reasons. Firstly, we demonstrate that by selecting optimization targets spanning different scales (micro + macro), such as the RDFs, density, and dielectric constant, it is possible to obtain an optimized water model with comparable accuracy to state-of-the-art models like TIP3P-FB and OPC3, while maintaining computational efficiency and robustness. Secondly, our findings allow us to investigate the chemical and physical origins that control the accuracy limits (indeterminacy) of model optimization. We investigate how these limits are intrinsic and connected to the physical constraints of the model itself. The insights gained from this study hold significance not only for optimizing the specific system presented in this paper but also for any approxi-

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

mated model that relies on higher-accuracy data or incorporates top-down constraints based on experimental evidence.

Methods

The optimization work conducted herein builds on a multi-reference particle swarm optimization software that we developed recently: *Swarm-CG*.[7,9] In particular, *Swarm-CG* has been developed to optimize bonded and non-bonded parameters in molecular models to fit experimental results (top-down references) and the behaviour seen in all-atom MD trajectories (bottom-up references). *Swarm-CG* has been successfully tested to optimize a variety of molecular systems (*e.g.*, lipid models[26]). In this paper, *Swarm-CG* has been adapted for this specific case study (a dedicated variant can be found at: `https://github.com/GMPavanLab/wateropti`). The five parameters of a general three-site rigid water model that are iteratively tuned (illustrated in Figure 1**a**) are: (i) the intramolecular distance between the Oxygen and the Hydrogen sites, d_{OH}, (ii) the intramolecular distance between the two Hydrogen sites, d_{HH}, (iii) the absolute charge of the Oxygen site, q, and the two functional parameters of the Lennard-Jones potential which is centered on the Oxygen site, namely (iv) sigma σ and (v) epsilon ε. We conducted our optimizations initializing swarms composed of 15 particles in the first and third subsections of the results, and 26 particles in the second subsection. In each optimization procedure, a series of classical molecular dynamics simulations are performed and their discrepancy from the target properties is evaluated according to a scoring function (described below). Finally, the obtained optimized models are simulated at various temperature across the liquid regime, and observables of interest are computed.

Scoring function

To quantify the discrepancy between the RDFs obtained from the simulations of the models vs. the experimental ones from liquid water (at various temperatures), we introduced a

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

scoring function based on the Earth mover's distance (EMD) or Wasserstein distance.[28] The Wasserstein distance is a measure of the dissimilarity between two probability distributions, based on the concept of optimal transport.[29] It represents the minimum cost of transforming one distribution into the other, where the cost is proportional to the distance between pairs of points. In our case, we used the Wasserstein distance to compare the simulated RDFs with the experimental RDFs, with the distance matrix representing the differences between the radial distance of the bins of the distributions. In our work, we modified the standard computation of the Wasserstein distance by using the square of the distance matrix instead of the distance itself. Such a modification allowed to better account (weights more) the difference between the $g(r)$ at larger distance, which is important for capturing the long-range behavior of the water-water interactions and to avoid overfitting on short-range interactions. Preliminary tests demonstrated that this provided the best setup to compare $g(r)$ curves as a whole in the most robust way. This modification also allows mitigating potential problems emerging from the fact that classical three-site water models usually have difficulty reproducing the first peak of the RDFs (due to the fact that quantum effects are not included in the description of the system).[30]

In particular, the scoring function used in the optimization presented the first subsection of the results is:

$$S = EMD_{g_{OO}} + EMD_{g_{OH}} + EMD_{g_{HH}} \ , \tag{1}$$

where S represents the score and $EMD_{g_{OO}}$, $EMD_{g_{OH}}$, and $EMD_{g_{HH}}$ represent the Earth mover's distance measurements of the three RDFs considered, namely the Oxygen-Oxygen, Oxygen-Hydrogen, and Hydrogen-Hydrogen, respectively. In this way, the scoring function does not capture discrepancies only in terms of distances and spatial displacement of the water molecules respect to each other, but also in terms of their natural orientation. This provides us with a scoring function which is rich in microscopic structural information on the system.

The optimizations presented in the second and third subsection of the results involved not

https://doi.org/10.26434/chemrxiv-2023-88fpn ORCID: https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

only fitting of microscopic features, but also the density and static dielectric constant (macroscopic observables). The adopted score is expressed as:

$$S = w_{EMD}(EMD_{g_{OO}} + EMD_{g_{OH}} + EMD_{g_{HH}}) + w_\rho \left|\rho_{sim} - \rho_{exp}\right| + w_\varepsilon \left|\varepsilon_{sim} - \varepsilon_{exp}\right| \ , \quad (2)$$

where the first term represents the difference between the simulated and experimental RDFs for each type of particle-particle correlation. The second and third terms take into account the difference between the simulated and experimental values of density ρ and static dielectric constant ε, respectively. Each term in the score function has a weight assigned to it, which determines its relative importance in the optimization process. The weights were chosen as $w_{EMD} = 0.5$, $w_\rho = 0.3$, and $w_\varepsilon = 0.2$. Preliminary tests demonstrated that these weights ensured a balanced representation in the optimization process, allowing us to prioritize and place emphasis on fitting the RDFs over other macroscopic features of the systems. A comparison of experimental RDF with simulated $g(r)$ examples scored according to our metrics is present in Figure S1 of the Supporting Information.

Results and Discussion

This part is organized as follows. The first subsection presents the results of the model optimized to reproduce the experimental RDFs (g_{OO}, g_{OH}, and g_{HH}) under standard conditions of 298 K and 1 bar. This approach focuses primarily on a pure bottom-up methodology, where the optimization is driven by the microscopic features of the water model. In the second subsection, we extend our analysis by optimizing the model to reproduce not only the RDFs but also experimental density and static dielectric constant. Furthermore, the system is trained at two additional temperatures, specifically 280 K and 343 K. This comprehensive optimization approach aims to capture a broader range of experimental observables, combining both bottom-up and top-down references. Finally, the last subsection provides a detailed investigation into the indeterminacy of the optimization problem within the context of the

https://doi.org/10.26434/chemrxiv-2023-88fpn ORCID: https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

three-site representation.

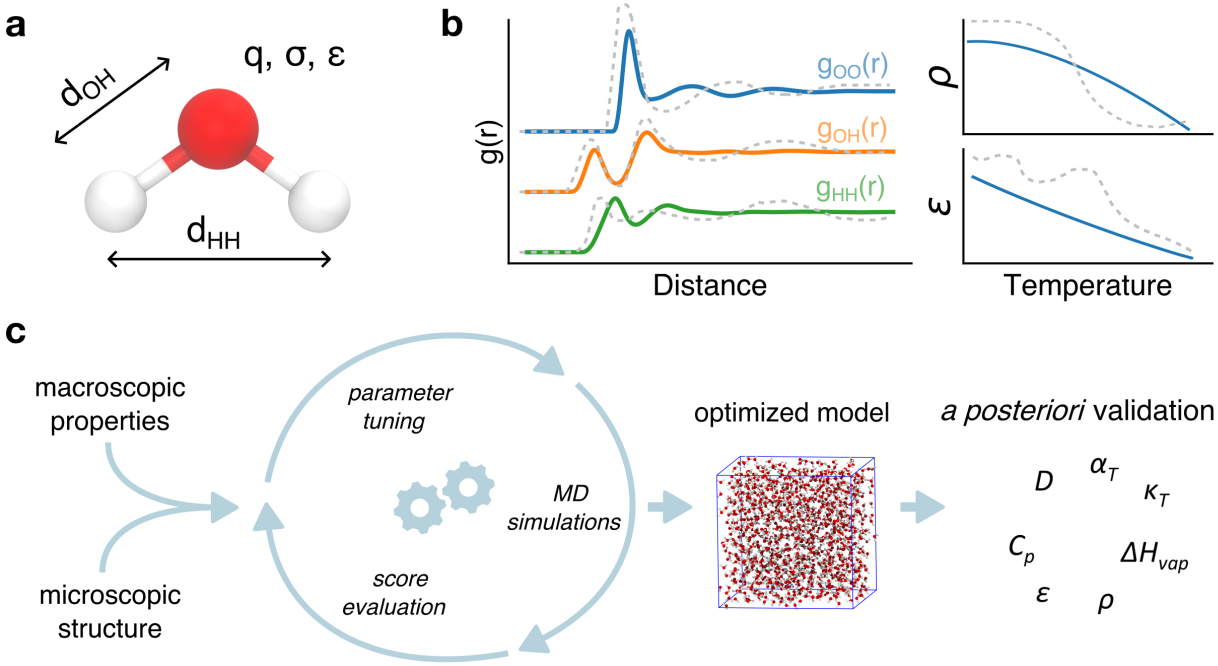

Figure 1: Overview of the study. **a** Representation of a water molecule and schematic of the five parameters that define the three-site model. **b** Experimental data used in this work: radial distribution functions $g(r)$ (bottom-up reference), liquid water density (ρ) as a function of temperature and static dielectric constant (ε) as a function of temperature (top-down references). **c** Workflow diagram illustrating the process of the study. Reference experimental data serve as the targets guiding the optimization process. Swarm-CG runs iterative MD simulations, adjusting the parameters of the water molecule to reach the best match with the reference experimental data. The resulting optimized model is then evaluated and validated *a posteriori* against a set of experimental observables at different temperatures not in the training set.

Multi-objective optimization based on microscopic system features

In a first optimization test, we trained the optimized water model according to a purely bottom-up approach to reproduce the experimental RDFs (g_{OO}, g_{OH} and g_{HH}) of water at the standard conditions of 298 K and 1 bar. At every iteration, *Swarm-CG* tests new parameters in the attempt to minimize the discrepancy between the g_{OO}, g_{OH} and g_{HH} obtained from the model and the experimental ones in standard conditions. The obtained results are presented in Figure 2. A comparison with other popular three-site water models of the same

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

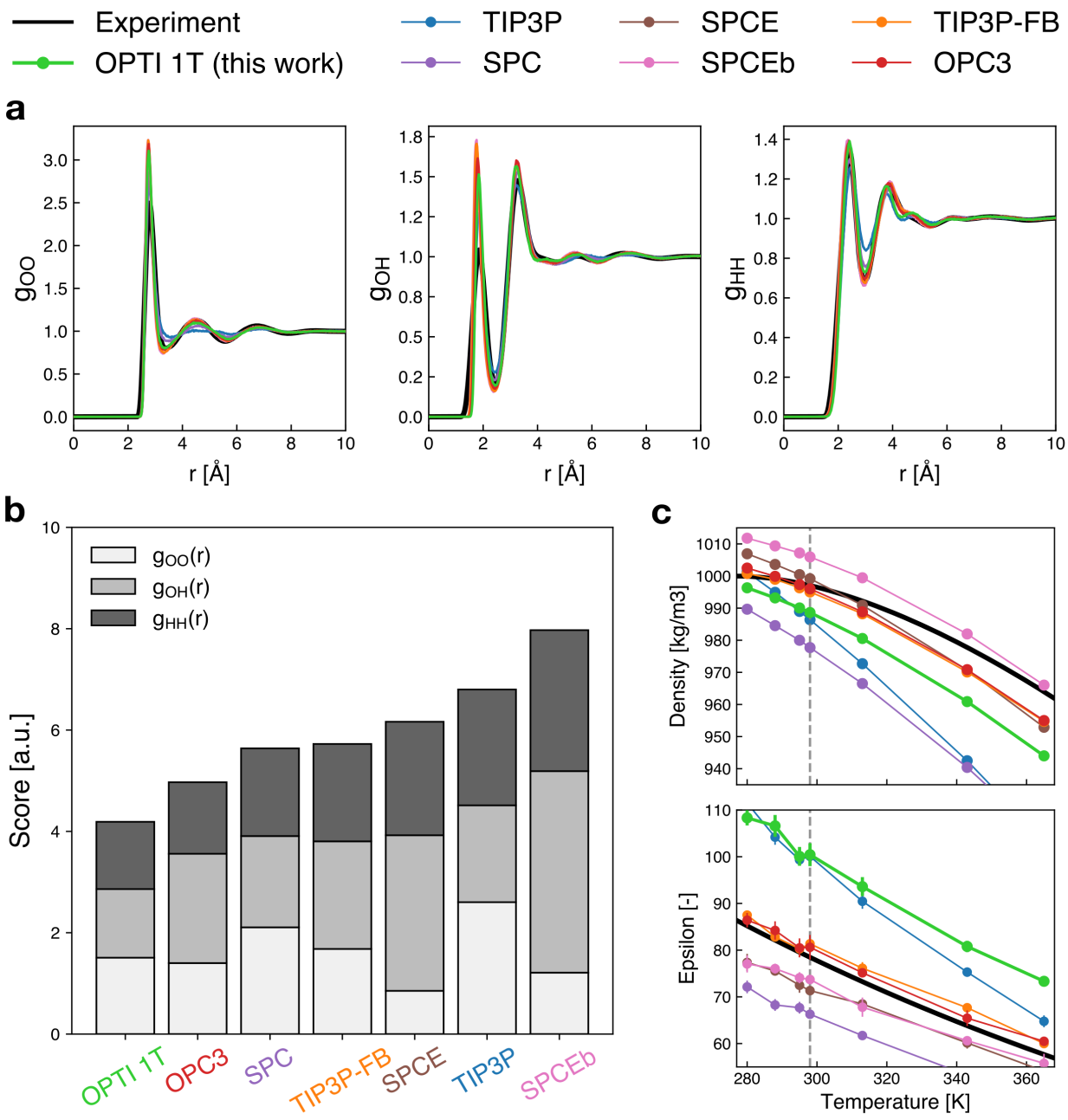

Figure 2: Results obtained from the first optimization, where the model OPTI 1T has been trained to reproduce experimental RDFs at 298 K and 1 bar. **a** RDFs reproduction and comparison with other three-site water models: complete data (without superposition of curves) are provided in the Supplementary Information (Figure S3). **b** Ranking of RDFs reproduction accuracy based on our score. **c** *A posteriori* validation of the model with respect to density and static dielectric constant. Dashed vertical gray lines indicate the temperature at which the model was trained.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

type (OPC3,[31] TIP3P-FB,[24] SPC,[18] SPCE,[32] SPCEb[33] and TIP3P[17]) is also provided. A summary of the parameters for these models can be found in Table S1 of the Supplementary Information.

According to the score that we formulated to quantify the deviation of the simulated RDFs from the experimental reference (equation (1)), our model exhibited the highest level of accuracy in replicating the experimental RDFs (Figure 2**b**). Despite the fact that a model optimized as such is found the best one, this is not surprising since our model was optimized to reproduce experimental RDFs. Nonetheless it can be noticed that also in our case the Oxygen-Hydrogen RDF is overlocalized which appears as unavoidable in such models, where the nuclear quantum effects are not explicitly included. Similarly, in the attempt to fit at best the 2^{nd} and 3^{rd} solvation shell in the Oxygen-Oxygen RDF (identified by the 2^{nd} and 3^{rd} g_{OO} peaks) produces an unavoidable overlocalization of the 1^{st} peak (an enlarged plot of the RDFs around the solvation shells is provided for clarity in Figure S2 of the Supporting Information).

It is worth noting that, even at the training temperature, the value of the dielectric constant predicted by our best model deviated from the target by a substantial amount. Such a lack of accuracy can be attributed to the fact that training the water model on RDFs alone does not provide sufficient information on the interactions between atoms. As a result, quantities such as the static dielectric constant, which depends on dipole fluctuations and is sensitive to the charges on the water model, are not reproduced accurately enough if the model is not trained to do so. This also means that, although the RDFs are well reproduced, this is not a sufficient condition for macroscopic properties to emerge spontaneously in the system. These considerations motivated us to incorporate also additional top-down experimental targets into the scoring function. The results of this integration are illustrated in the following section.

In terms of computational time, the refinement of 5 parameters of the water model at single temperature condition required 4 days (wall-clock time) to reach convergence (500 swarm

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

iterations) using 26 particles in the swarm and using 36 CPU cores, each simulation running on 9 CPU cores equipped with a GPU.

Multi-objective multi-temperature optimization based on microscopic and macroscopic observables

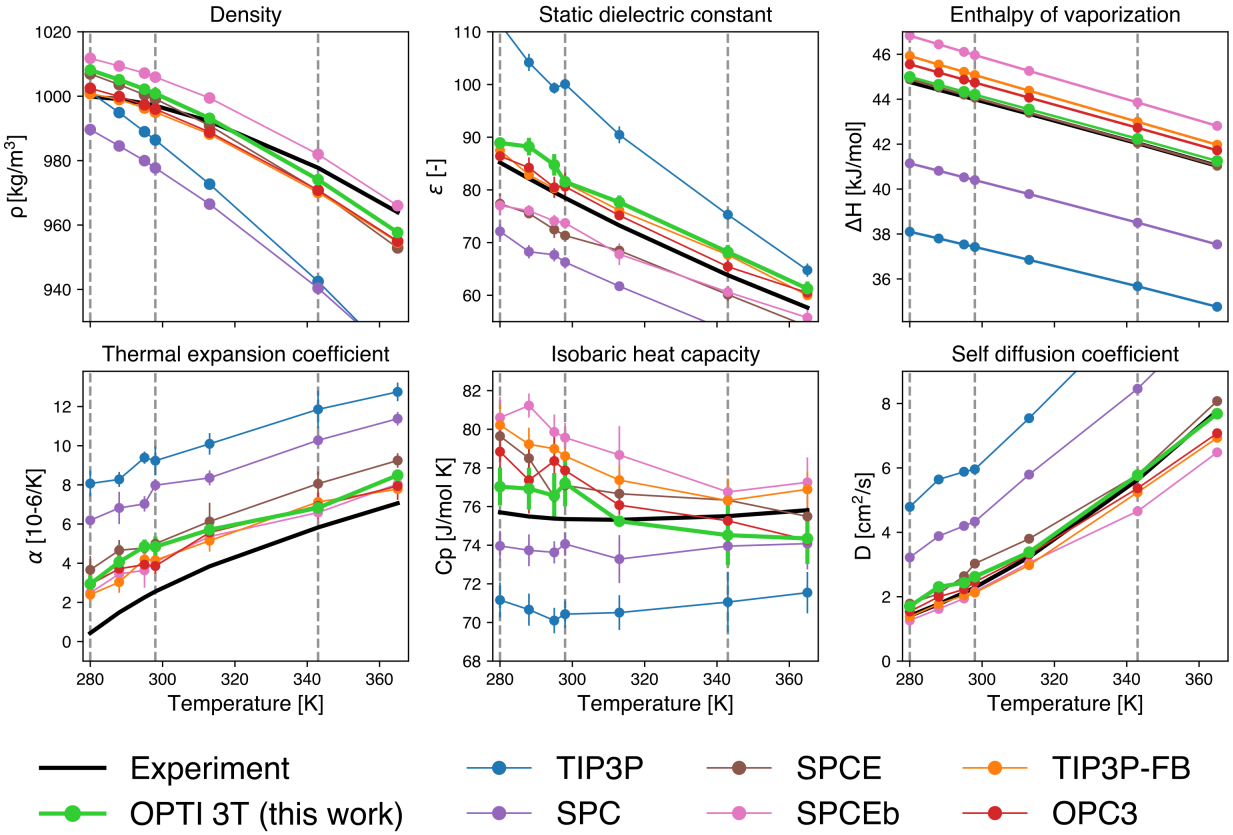

Figure 3: Validation of the model OPTI 3T, trained at the temperatures of 280 K, 298 K, and 343 K. Reproduction of RDFs, isothermal compressibility, surface tension, and quantification of the deviation of simulated observables with respect to experimental data are provided in the Supporting Information (Figures S3-S6)

In a second test, we trained our water model using a hybrid approach, incorporating both top-down and bottom-up references, to attain an accurate reproduction of the experimental radial distribution functions (RDFs), density, and static dielectric constant at three distinct

https://doi.org/10.26434/chemrxiv-2023-88fpn ORCID: https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

temperatures: 280K, 298K, and 343K. This is thus a multi-temperature multi-objective optimization combining top-down (microscopic) and bottom-up (macroscopic) target observables. Figure 3 illustrates the model's performance in reproducing observables of interest in the liquid regime, namely density, static dielectric constant, enthalpy of vaporization, thermal expansion coefficient, isobaric heat capacity and self-diffusion coefficient. The plots containing RDFs, the values of isothermal compressibility and surface tension are present in the Supporting Information (Figures S3, S4 and S5). The overall accuracy of the model that we obtained can be compared with the most advanced state-of-the-art data-driven trained models, such as TIP3P-FB[24] and OPC3.[31] In particular, it is worth noting the agreement of our model (OPTI-3T, in green) with the experimental enthalpy of vaporization and self-diffusion coefficient at all explored temperatures. The enthalpy of vaporization reflects the strength of interactions between water molecules in the liquid state, representing the energy required to transition a molecule from the liquid to the vapor phase. On the other hand, the self-diffusion coefficient characterizes the dynamics of individual molecule diffusion within the liquid, indicating the ease of movement in a medium comprised of other water molecules. Notably, our optimized model demonstrates remarkable agreement with experimental results for these two parameters, despite not being explicitly targeted during training. This agreement underscores the significance of training the model on the radial distribution function and these two additional macroscopic targets, as they provide essential information for accurately reproducing fundamental thermodynamic and kinetic properties at the local level. A quantification of the accuracy of our OPTI-3T model by means of average deviation from the various experimental observables is present in Figure S6 of the Supporting Information. Overall, these results show the striking positive effect of training the water model based on microscopic information rich observables (*e.g.*, the RDFs), and how microscopic characteristics of the model significantly influence most of its properties.

In terms of computational time, the refinement of 5 parameters of the water model at 3 levels of temperature required 8 days (wall-clock time) to reach convergence (300 swarm iterations)

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

using 15 particles in the swarm and using 36 CPU cores, each simulation running on 9 CPU cores equipped with a GPU.

These results give rise to several important considerations. Firstly, the results obtained with our method demonstrate that, despite the fact that our model reproduces globally well the explored thermodynamic properties across the different conditions, the performances of OPTI-3T are not distant from those of *e.g.*, OPC3 and TIP3P-FB. This demonstrates that, substantially, there is a limited room for radically improving the performances of three-site rigid water models. All our results suggest that there is an intrinsic limit in the accuracy that is achievable with models where the representation of the water molecule is so simplified. This leads us to fundamental questions. What are the key factors underpinning such limits? Are these imputable, *e.g.*, to limitations in the optimization method itself, or to intrinsic limits of the model? In the next section, we will deeper investigate these questions, obtaining interesting insights.

Intrinsic physical limits and indeterminate optimizations of rigid three-site water models

Recently, Izadi *et al.*[31] suggested that three-site water models somehow possess inherent accuracy limitations due to their oversimplified nature, which hinders their ability to achieve a complete and experimentally consistent reproduction of observables across the liquid phase. Nevertheless, an relevant question that remains unanswered is the precise reason behind this instrinsic limitation.

The performance of an automatic optimization procedure may be significantly influenced by *a priori* choices concerning the methodology and training variables. As a result, the model's ability to accurately fit different observables may vary to some extent. In the case of OPC3, for example, it was considered crucial to impose constraints on the geometry of the molecules in order to ensure a quadrupole moment of zero.[31] In the case of TIP3P-FB, a predominant emphasis was placed on a top-down approach, involving the simultaneous fitting to multiple

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

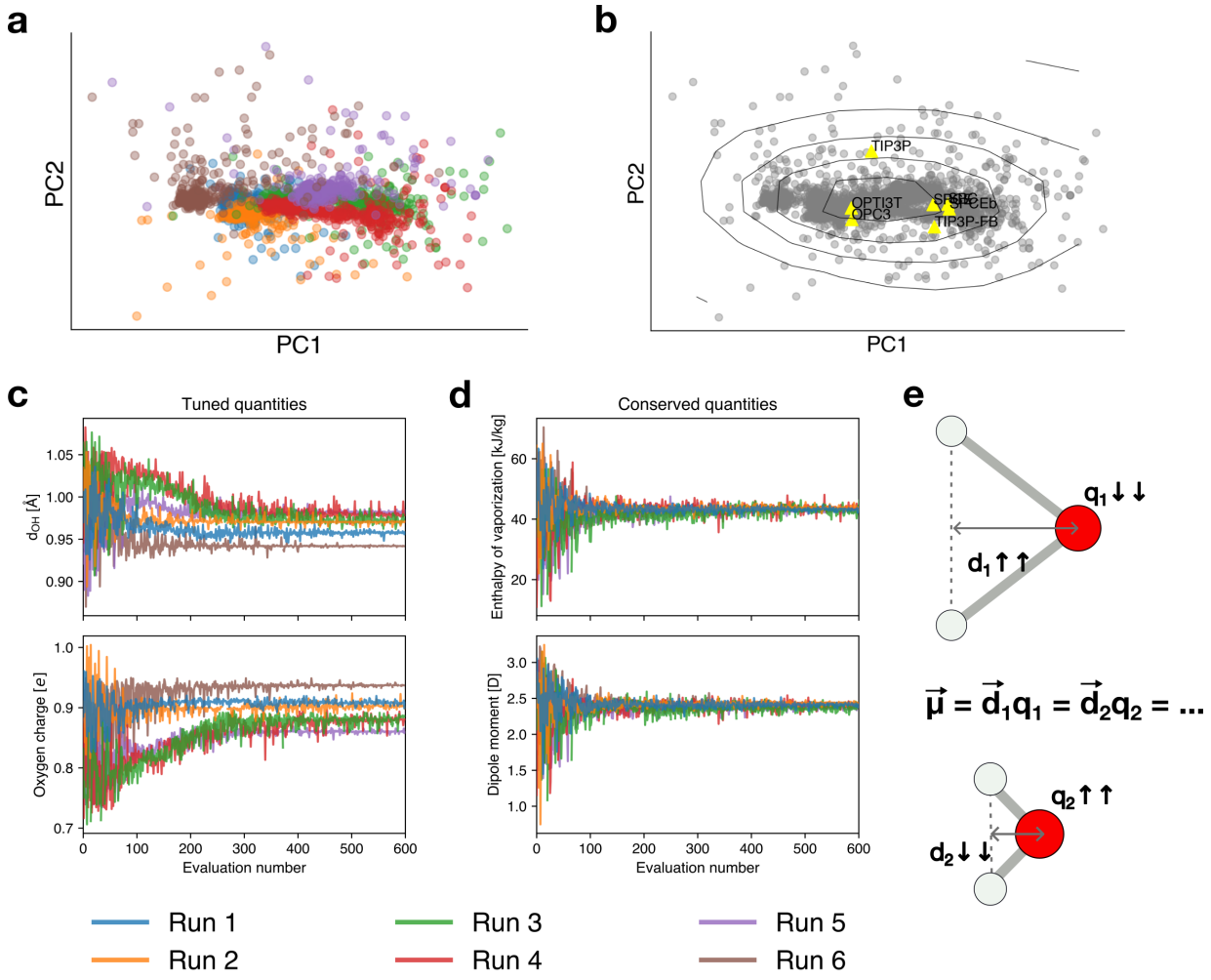

Figure 4: Results obtained by running a series of identical optimizations, initialized in different points of the parameters space. **a** Models obtained as solutions, displayed with principal component analysis. Each color represents a different optimization run. **b** Density isolines. **c** Parameters that are tuned during the optimization - d_{HH} and Oxygen charge - as a function of the number of iterations. **d** Example of quantities calculated *a posteriori* that are related to the energy of interaction between molecules - *e.g.*, enthalpy of vaporization and dipole moment - as a function of the number of iterations.**e** Schematic representation of the solutions obtained from the series of optimizations: inter-dependence between size of the molecule and and partial charges.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

thermodynamic observables across the liquid regime of water.[24] To investigate the impact of these initial conditions and gain a comprehensive understanding of the optimization process, we conducted a series of six optimizations under identical constraints. Specifically, we minimized the discrepancy of radial distribution functions (RDFs), density, and static dielectric constant, at the standard conditions of 298 K and 1 bar. Since the results obtained with such optimization cycles vary to some extent, this setup did not produce a single solution (identical in all six runs) but rather a group of solutions. Noteworthy, the obtained solutions demonstrate a comparable score (as illustrated in Figure S7 of the Supporting Information). Moreover, we conducted a Principal Component Analysis (PCA) of all explored solutions achieved through the use of *Swarm-CG*. Figure 4a shows the high density regions (*i.e.*, the solutions projected on the first two principal components) that represent the optimal solution obtained from our optimization cycles. The different colors represent the different runs. The contour lines illustrated in Figure 4b represent density isolines, which enable us to identify the regions of higher density points containing a set of optimal water models (according to our scoring function), which are characterized by slightly different set of parameters. These data demonstrate how *Swarm-CG* brings the model systematically not to a to a specific solution (*i.e.*, to a specific optimal model), but to a region of the space which contains "equally optimal, although slightly different solutions". Interesting questions are, for example, why the method behaves in this way, and specifically why slightly different solutions are "equally optimal".

A deeper inspection of the "optimal solutions" provided by *Swarm-CG* revealed interesting patterns. In particular, it is interesting to observe that the dipole moment of all the models belonging to this minimum is identical. Namely, despite the fact that their geometry or partial charges can be slightly larger/smaller (Figure 4c) in the various solutions, these change in such a way that the dipole moment the molecule is conserved. In such a way, the enthalpy of vaporization is also conserved across the various solutions (Figure 4d).

It is worth noting that these two properties (the dipole moment and the enthalpy of vapor-

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

ization) are evaluated *a posteriori* and are not explicitly used to train the models during the optimization process. Moreover, both such observables are related to the extent of the intermolecular interactions, as the enthalpy of vaporization is proportional to the potential energy in the system, and the interaction between dipoles of the molecules plays a key role in it. Figure 4d presents a visual representation of the variability in charge and geometry of the water models generated by the different optimization runs. In qualitative terms, an increase in charge corresponds to a reduction in the size of the molecule, while a decrease in charge results in an increase in size (Figure 4c,e). Additionally, we observed changes in the Lennard-Jones interactions with variations in sigma and epsilon values (Figure S8 in the Supporting Information). The obtained results and considerations illustrate how the collective properties of these models are largely controlled by the interplay of molecular dipoles and their interactions with each other.

Since in such a simplified three-site models the majority of water-water intermolecular interaction are largely governed by the dipole moment, this introduces an intrinsic level of indeterminacy in the optimization. Recently, we have observed similar results also in the framework of the automatic optimization of, *e.g.*, lipid models using *Swarm-CG*, where a certain level of model accuracy can be achieved, although accompanied by an inherent uncertainty. While uncertainties can arise from various sources in automatic approaches, such as the number of objectives, parameter selection, and optimization methods, it is worth noting that the uncertainty we are referring to in this context originates elsewhere. Here, due to the simplified physical description of the three-site rigid water model, the optimization problem becomes inherently undetermined as it seeks to find an optimized dipole, which is a composite variable represented by the product of charge and geometry ($\mu = qd$). This leads to a degeneracy characterized by different optimal solutions with varying combinations of charge and geometry. To overcome this limitation, a potential improvement could involve incorporating additional parameters that decouple the geometric and electrostatic characteristics of the water model during the training process. For example, one approach

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

could be training the model to reproduce a specific geometry obtained with higher accuracy from quantum mechanical (QM) approaches or by calculating the electrostatic potential. However, implementing such an approach encounters challenges due to the substantial differences between the geometric and electrostatic descriptions at the QM level compared to the all-atom models, as demonstrated by recent research on OPC3.[25]

An interesting outcome of these considerations is that while our approach can achieve optimal solutions at least as good as the state-of-the-art models in a very efficient way, it also underlines how such all atom water models are *de facto* a coarse grained description of the real water molecule features. Like other coarse-grained models, they encounter a degeneracy due to simplified representations of system degrees of freedom, resulting in a certain level of precision combined with inherent indeterminacy. This indeterminacy implies that different parametrizations can lead to similar behaviors. Notably, in our case, this degeneracy yields a set of non-identical solutions that belong to a minimum that is identified by the PCA.

These considerations suggest that similar principles used to parameterize other three-site rigid water models generally encounter similar limitations (*i.e.*, the challenges and constraints faced in developing and optimizing water models based on similar principles are likely to be shared). To reduce the uncertainty in the model outcomes, another possible solution is to employ extra sites in the model, such as in the four-site models like TIP4P[17] or TIP4P-ICE.[34] Such an addition permits to expand the degrees of freedom to tune, allowing for a better fit of some properties , *e.g.*, the curve of density across different temperatures. *De facto*, this underlines how, to improve substantially the performances of three-site water models, it is necessary to increase the resolution of the model accounting for more degrees of freedom.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Conclusions

In this work, we explored the effect of combining microscopic-rich and macroscopic-rich information into a training set of experimental observables used to automatically optimize classical three-site water models. In particular, as the microscopic target observables, we use the experimental g_{OO}, g_{OH}, and g_{HH} radial distribution functions of liquid water at various temperatures that, altogether, contain information not only of how strongly the water molecules interact but also on how the molecules are organized in space respect to each other. A first optimization of the water model under standard conditions (298 K and 1 bar) using only such a bottom-up (microscopic) reference demonstrated how such microscopic information alone is insufficient to obtain an experimentally consistent reproduction of all other screened macroscopic observables for the liquid phase of water, especially for what pertains to the density and dielectric constant of liquid water.

Including in the training-set and in the score, in a second test, the density and static dielectric constant of liquid water, was then seen to provide considerable improvements. The obtained model showed a remarkable improvement in reproducing macroscopic properties, especially with respect to self-diffusion coefficient and enthalpy of vaporization. This also suggests that these properties – density and dielectric constant – are not strongly dependent on the $g(r)$ of water. Overall, we found that our optimized water model (called OPTI-3T herein) exhibits a comparable level of accuracy as two models, OPC3 and TIP3P-FB, which were also obtained through automatic optimization approaches and are considered state-of-the-art models in the realm of three-site rigid water models. Nonetheless, it is worth noting how, in our case, combining microscopic and macroscopic target properties allows achieving such a level of accuracy in an efficient way, and with a relatively reduced computational time (*e.g.*, TIP3P-FB is trained on a large amount of thermodynamic properties at various temperatures in a computationally intensive process). At the same time, our tests show that there is little room for further improvement in these models by, *e.g.*, adding more experimen-

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

tal observables in the training-set, etc., which suggested that all these models are somewhat very similar and possibly nearly consistent with each other, considered the precision that it is reasonable to expect from them.

The series of optimizations that we conducted herein under identical constraints (namely, fitting the experimental RDFs, density and static dielectric constant at standard conditions) shows that these models are somewhat intrinsically limited in their accuracy. The same is true in some sense concerning the determinacy of their optimization cycles. The results shown in Figure 4 show how many of the screened thermodynamic observables are controlled in these simplified water models by the water dipole (μ), which is a composite variable that depends on both the charge (q) and geometry (d) of the water model ($\mu = qd$). This leads to an inherent indeterminacy in the solutions that are systematically obtained. This means that different combinations of charge and size can correspond to equally optimal solutions toward the fitting of the targeted properties. A set of optimal solutions is thus typically obtained in such automatic optimizations instead of a single specific one. The PCA data of Figure 4a,b show how all such obtained "optimal" solutions belong, in our case, to the same global high-density minimum. In Figures 4c,d it is demonstrated how all the slightly different solutions belonging to such minimum represents share nearly identical molecular dipole and enthalpy of vaporization. While the broadness of such minima could be interpreted, *e.g.*, as to be imputable to some kind of statistical error/limit in the particle swarm optimization method used herein, these results suggest that this is most likely related to an intrinsic indeterminacy in how the problem is posed. In particular, the degrees of freedom in such "coarse-grained" atomistic description of the water models are so limited that the optimization process degenerates, providing equally optimal solutions that are nonetheless different from each other.

These results are interesting because they demonstrate that, when dealing with the optimization of approximated models, there will be inevitably intrinsic limits due to degeneration

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

of the optimal set of parameters that satisfy the conditions that are posed. In such a case, further improvements cannot be achieved without introducing additional degrees of freedom that can decouple in some way such composite variables into the fundamental ones allowing to fine-tune the model. One way could be, *e.g.*, to add some higher (quantum) level additional constraint that allows to decouple the dependence on the charge (q) from that of geometry (d) in the solution. However, the geometry and electrostatics of QM water molecules are so different from those of these AA models, that tests in this sense proved inefficient. In the case of the classical water models studied herein, reaching higher precision thus requires expanding the model's representation by adding additional "classical" degrees of freedom, for example, allowing for a more flexible and accurate description of the system. This is exactly the case of the higher precision that can be achieved by, *e.g.*, 4- or 5-site models.[35,36] Moreover, altering the degrees of freedom can have important effects for example on subtle dynamical mechanisms associated with water reorientational dynamics as recently shown.[37] However, these results are also interesting for the development of approximated molecular models in general. Recently, we have observed similar intrinsic limitations also in the optimization of, *e.g.*, coarse-grained models of a variety of other molecular systems.[7,9,26] This observation serves as a valuable lesson for developing models of all kinds, not just in the context of water simulations. Such inherent limitations and these challenges encountered in optimizing approximated models demonstrate the importance of considering the complexity of the system being studied, and the type of information lost with approximated molecular models. The integration of multiple references, and in particular combining bottom-up and top-down microscopic/macroscopic-level information in the training-set can improve the efficiency and robustness in the models' optimization. Nonetheless, the results discussed herein also offer an unambiguous example of how, understanding the physical limits of approximated models, can provide a precious knowledge for guiding future research towards more robust and reliable modeling approaches.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Computational details

MD Simulations

All the simulations have been conducted using GROMACS version 2021.5 [38] with the following protocol. The starting systems' configuration is a cubic box containing 1024 water molecules, arranged in initial random configurations, using Packmol. [39] After a preliminary energy minimization via steepest descent algorithm (for $2 \cdot 10^3$ steps), the system is then equilibrated for 5ns in NpT ensemble and simulated for 10ns in the same ensemble. Both these equilibration and production phases are simulated with a 2 fs timestep. We kept temperature constant with velocity rescale thermostat [40] (with a time constant of 0.2 ps), and pressure constant to 1 bar with cell rescale barostat [41] (with a coupling constant of 1 ps and compressibility of $4.5 \cdot 10^{-5}$ bar). A cutoff distance of 1 nm was used for short-range electrostatic and van der Waals interactions, and the long-range interactions were computed with the particle-mesh Ewald summation method. [42] Corrections to long-range pressure and potential energy were considered. [43]

Observables

Density. The mass density of water ρ is calculated as follows:

$$\rho = \frac{N \cdot m_{H_2O}}{\mathcal{N}_A \cdot V_{\text{box}}} \ , \tag{3}$$

where N is the number of water molecules (1024 in our case), m_{H_2O} is the mass of water molecules in u.a., $\mathcal{N}_A$ is Avogadro's number, and V_{box} is the volume of the simulation box. Experimental reference data of ρ are taken from Ref. [44]

Static dielectric constant. We calculate the static dielectric constant from the fluctu-

https://doi.org/10.26434/chemrxiv-2023-88fpn ORCID: https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

ations of the total dipole moment M of the simulation box, *i.e.*, as:

$$\varepsilon = 1 + \frac{\langle M^2 \rangle - \langle M \rangle^2}{3\varepsilon_0 V k_B T} \ , \tag{4}$$

where ε_0 is the permittivity of the vacuum, V the volume of the simulation box, k_B the Boltzmann constant, T the temperature of the system, and $\langle \cdot \rangle$ represents the thermodynamic average. We calculated this observable using the routine `gmx dipoles` of GROMACS suite. Experimental reference data of ε are taken from Ref.[44]

Radial distribution functions. We calculated the radial distribution functions of Oxygen-Oxygen, Oxygen-Hydrogen and Hydrogen-Hydrogen ($g_{OO}(r)$, $g_{OH}(r)$, $g_{HH}(r)$) pairs with MDAnalysis 2.0.0.[45] We considered a cutoff of 10Å and 500 equally spaced bins. Experimental reference data of radial distribution functions are taken from Ref.[46]

Self diffusion coefficient. The self-diffusion coefficient D is calculated using Einstein's relation for a diffusive particle as:

$$D = \lim_{t \to \infty} \frac{\langle |\vec{r}\,(t) - \vec{r}\,(0)\,|^2 \rangle}{6t} \ , \tag{5}$$

where the quantity in the numerator is the mean square displacement (MSD), averaged over the trajectories of individual particles. Diffusion coefficients calculated with MD simulation are often referred to as D_{PBC}, because they contain systematic errors due to the finite box size.[47] Following Ref.,[47] it is possible to correct this artifact obtaining the theoretical value of self-diffusion coefficient of water in an infinite box (D_0). To this end, we calculated D_{PBC} in cubic simulation cells with N = 512, 1024, 2048, 4096 and 8192 water molecules. The protocol used for these simulations is identical to the one described in section , except for a different production time, *i.e.*, 20 ns (N = 512), 15 ns (N = 1024, 2048), 10 ns (N = 4096, 8192). We calculated the various D_{PBC} values using the `gmx msd` routine of GROMACS,[38] and D_0 with linear interpolation. Experimental reference data are taken from Ref.[48]

Enthalpy of vaporization. The enthalpy of vaporization ΔH_{vap} of one mole of liquid

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

water in the gas phase can be approximated as:[49]

$$\Delta H_{vap} \approx -U + RT - p_{sat}V - E_{pol} + C \ , \tag{6}$$

where U and V are respectively the average potential energy and the volume of one mole of water molecules at pressure p and bath temperature T. p_{sat} is the value of saturation pressure at temperature T. The term E_{pol} represents the depolarization energy of one mole of water molecules when it is transferred from the liquid to the gas phase.[32] It can be expressed as:

$$E_{pol} = \frac{(\mu - \mu_{gas})^2}{2\alpha_{gas}} \tag{7}$$

where μ is the dipole moment of the simulated model, μ_{gas} and α_{gas} are the dipole moment and average polarizability of a water molecule in the gas phase,[49] respectively. The last term in equation (6) contains corrections that account for the vibrational effects of water molecules and non-ideality of the gas phase. These corrections are reported in Ref.[49] for different temperatures. Experimental reference data are taken from Ref.[49]

Specific heat capacity . We computed the isobaric heat capacity c_p using the enthalpy fluctuation formula, namely:

$$c_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{N k_B \langle T \rangle^2} \ , \tag{8}$$

We computed this observable by using the `gmx energy` routine of the GROMACS[38] suite. The value obtained was then corrected to account for quantum effects that are not considered in the classically computed heat capacity in eq.(8). Specifically, these corrections include estimation of intra-molecular vibrational energies (due to the fact that our model is rigid) and inter-molecular high frequency modes. The values of these correction are reported in the Ref.[49] of Horn et al. The experimental reference data of c_p are taken from Ref.[50]

Thermal expansion coefficient. We calculated the thermal expansion coefficient α_T

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

using the enthalpy-volume fluctuation formula:

$$\alpha_T = \frac{\langle VH \rangle - \langle V \rangle \langle H \rangle}{k_B \langle T \rangle^2 \langle V \rangle} \ , \tag{9}$$

We computed this observable by using the `gmx energy` routine of the GROMACS[38] suite. The experimental reference data of α_T are taken from Ref.[50]

Isothermal compressibility. We calculated the thermal expansion coefficient κ_T using the volume fluctuation formula:

$$\kappa_T = \frac{\langle V \rangle^2 - \langle V^2 \rangle}{k_B \langle T \rangle \langle V \rangle} \ , \tag{10}$$

We computed this observable using the `gmx energy` routine of GROMACS[38] suite. The experimental reference data of κ_T were taken from Ref.[50]

Surface tension. The interface between water and void was prepared and simulated following the good practices outlined in Ref.[51] Firstly, a cubic box containing 1024 water molecules was equilibrated in the NPT ensemble. To represent the void phase, the z-axis of the simulation box was elongated by a factor of 4. The resulting biphasic system was then simulated for 50 ns in the NVT ensemble. The surface tension of the water-void interface was calculated using the mechanical or pressure approach,[52] which involves evaluating the inhomogeneity of the pressure tensor as follows:

$$\gamma(t) = \frac{L_z}{2} \left(P_{zz}(t) - \frac{P_{xx}(t) + P_{yy}(t)}{2} \right) \ , \tag{11}$$

where L_z is the elongation of the z-axis, and P_{xx}, P_{yy}, and P_{zz} are the diagonal components of the pressure tensor. To perform the analysis, we used the `gmx energy` routine of the GROMACS suite.[38] Experimental reference data for the surface tension are obtained from Ref.[53]

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Acknowledgements

GMP acknowledges the funding received by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 818776 - DYNAPOL).

Data Availability

The input files, relevant analysis script, and the results of the optimizations and simulations conducted herein are available at `https://github.com/GMPavanLab/wateropti` (temporary link that will be replaced with a definitive Zenodo archive upon acceptance of the final version of this paper). Other information needed is available from the corresponding author upon reasonable request.

Notes

The authors declare no competing financial interests.

Supporting Information

The Supporting Information is available free of charge online. Additional details regarding the automatic optimizations and molecular simulations conducted (PDF).

References

(1) Paton, R. S.; Goodman, J. M. Hydrogen bonding and π-stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions. *J. Chem. Inf. Model.* **2009**, *49*, 944–955, DOI: 10.1021/ci900009f.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(2) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604, DOI: 10.1002/jcc.20082.

(3) Köfinger, J.; Hummer, G. Empirical optimization of molecular simulation force fields by Bayesian inference. *Eur. Phys. J. B* **2021**, *94*, 245, DOI: 10.1140/epjb/s10051-021-00234-4.

(4) Paesani, F. Getting the Right Answers for the Right Reasons: Toward Predictive Molecular Simulations of Water with Many-Body Potential Energy Functions. *Acc. Chem. Res.* **2016**, *49*, 1844–1851, DOI: 10.1021/acs.accounts.6b00285.

(5) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355–360, DOI: 10.1126/science.aat4010.

(6) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian inference method for heterogeneous systems. *Sci. Adv.* **2016**, *2*, e1501177, DOI: 10.1126/sciadv.1501177.

(7) Empereur-Mot, C.; Pesce, L.; Doni, G.; Bochicchio, D.; Capelli, R.; Perego, C.; Pavan, G. M. Swarm-CG: Automatic Parametrization of Bonded Terms in MARTINI-Based Coarse-Grained Models of Simple to Complex Molecules via Fuzzy Self-Tuning Particle Swarm Optimization. *ACS Omega* **2020**, *5*, 32823–32843, DOI: 10.1021/acsomega.0c05469.

(8) Daru, J.; Forbert, H.; Behler, J.; Marx, D. Coupled Cluster Molecular Dynamics of Condensed Phase Systems Enabled by Machine Learning Potentials: Liquid Water Benchmark. *Phys. Rev. Lett.* **2022**, *129*, 226001, DOI: 10.1103/physrevlett.129.226001.

(9) Empereur-mot, C.; Capelli, R.; Perrone, M.; Caruso, C.; Doni, G.; Pavan, G. M. Automatic multi-objective optimization of coarse-grained lipid force fields using SwarmCG. *J. Chem. Phys.* **2022**, *156*, 024801, DOI: 10.1063/5.0079044.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(10) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2011**, *8*, 61–74, DOI: 10.1021/ct200731v.

(11) Gong, Z.; Sun, H.; Eichinger, B. E. Temperature Transferability of Force Field Parameters for Dispersion Interactions. *J. Chem. Theory Comput.* **2018**, *14*, 3595–3602, DOI: 10.1021/acs.jctc.8b00104.

(12) Wallqvist, A.; Mountain, R. D. *Reviews in Computational Chemistry*; John Wiley and Sons, Ltd, 1999; pp 183–247, DOI: 10.1002/9780470125908.ch4.

(13) Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **2002**, *101*, 219–260, DOI: 10.1016/s0167-7322(02)00094-6.

(14) Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254, DOI: 10.1021/jp0667442.

(15) Tempra, C.; Ollila, O. H. S.; Javanainen, M. Accurate Simulations of Lipid Monolayers Require a Water Model with Correct Surface Tension. *J. Chem. Theory Comput.* **2022**, *18*, 1862–1869, DOI: 10.1021/acs.jctc.1c00951.

(16) Emperador, A.; Crehuet, R.; Guàrdia, E. Effect of the Water Model in Simulations of Protein–Protein Recognition and Association. *Polymers* **2021**, *13*, 176, DOI: 10.3390/polym13020176.

(17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935, DOI: 10.1063/1.445869.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(18) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. Intermolecular forces. 1981; pp 331–342, DOI: 10.1007/978-94-015-7658-1_21.

(19) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174, DOI: 10.1002/jcc.20035.

(20) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236, DOI: 10.1021/ja9621760.

(21) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751, DOI: 10.1002/jcc.20303.

(22) Wang, X.; Tse, Y.-L. S. Flexible Polarizable Water Model Parameterized via Gaussian Process Regression. *J. Chem. Theory Comput.* **2022**, *18*, 7155–7165, DOI: 10.1021/acs.jctc.2c00529.

(23) Chan, H.; Cherukara, M. J.; Narayanan, B.; Loeffler, T. D.; Benmore, C.; Gray, S. K.; Sankaranarayanan, S. K. R. S. Machine learning coarse grained models for water. *Nat. Commun.* **2019**, *10*, 379, DOI: 10.1038/s41467-018-08222-6.

(24) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891, DOI: 10.1021/jz500737m.

(25) Izadi, S.; Onufriev, A. V. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **2016**, *145*, 074501, DOI: 10.1063/1.4960175.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(26) Empereur-mot, C.; Pedersen, K. B.; Capelli, R.; Crippa, M.; Caruso, C.; Perrone, M.; Souza, P. C. T.; Marrink, S. J.; Pavan, G. M. Automatic Optimization of Lipid Models in the Martini Force Field Using SwarmCG. *J. Chem. Inf. Model.* **2023**, *63*, 3827–3838, DOI: 10.1021/acs.jcim.3c00530.

(27) Rick, S. W. A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums. *J. Chem. Phys.* **2004**, *120*, 6085–6093, DOI: 10.1063/1.1652434.

(28) Pele, O.; Werman, M. Fast and robust Earth Mover's Distances. 2009 IEEE 12th International Conference on Computer Vision. 2009; pp 460–467, DOI: 10.1109/ICCV.2009.5459199.

(29) Peyré, G.; Cuturi, M., et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning* **2019**, *11*, 355–607, DOI: 10.1561/2200000073.

(30) Giberti, F.; Hassanali, A. A.; Ceriotti, M.; Parrinello, M. The Role of Quantum Effects on Structural and Electronic Fluctuations in Neat and Charged Water. *J. Phys. Chem. B* **2014**, *118*, 13226–13235, DOI: 10.1021/jp507752e.

(31) Izadi, S.; Onufriev, A. V. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **2016**, *145*, 074501.

(32) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271, DOI: 10.1021/j100308a038.

(33) Takemura, K.; Kitao, A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. *J. Phys. Chem. B* **2012**, *116*, 6279–6287, DOI: 10.1021/jp301100g.

(34) Abascal, J.; Sanz, E.; García Fernández, R.; Vega, C. A potential model for the study

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, 234511, DOI: 10.1063/1.1931662.

(35) Pathirannahalage, S. P. K.; Meftahi, N.; Elbourne, A.; Weiss, A. C. G.; McConville, C. F.; Padua, A.; Winkler, D. A.; Gomes, M. C.; Greaves, T. L.; Le, T. C.; Besford, Q. A.; Christofferson, A. J. Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2021**, *61*, 4521–4536, DOI: 10.1021/acs.jcim.1c00794.

(36) Capelli, R.; Muniz-Miranda, F.; Pavan, G. M. Ephemeral ice-like local environments in classical rigid models of liquid water. *J. Chem. Phys.* **2022**, *156*, 214503, DOI: 10.1063/5.0088599.

(37) Offei-Danso, A.; Morzan, U. N.; Rodriguez, A.; Hassanali, A.; Jelic, A. The collective burst mechanism of angular jumps in liquid water. *Nat. Commun.* **2023**, *14*, DOI: 10.1038/s41467-023-37069-9.

(38) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25, DOI: 10.1016/j.softx.2015.06.001.

(39) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164, DOI: 10.1002/jcc.21224.

(40) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101, DOI: 10.1063/1.2408420.

(41) Bernetti, M.; Bussi, G. Pressure control using stochastic cell rescaling. *J. Chem. Phys.* **2020**, *153*, 114107, DOI: 10.1063/5.0020514.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(42) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593, DOI: 10.1063/1.470117.

(43) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B* **2007**, *111*, 13052–13063, DOI: 10.1021/jp0735987.

(44) Lide, D. R. *CRC handbook of chemistry and physics*; CRC press, 2004; Vol. 85.

(45) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327, DOI: 10.1002/jcc.21787.

(46) Soper, A. K. Water and ice structure in the range 220 - 365K from radiation total scattering experiments. **2014**, DOI: 10.48550/arXiv.1411.1322.

(47) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879, DOI: 10.1021/jp0477147.

(48) Oelkers, E. H. Calculation of diffusion coefficients for aqueous organic species at temperatures from 0 to 350 °C. *Geochim. Cosmochim. Acta* **1991**, *55*, 3515–3529, DOI: 10.1016/0016-7037(91)90052-7.

(49) 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, DOI: 10.1063/1.1683075.

(50) Wagner, W.; Pruß, A. The IAPWS formulation 1995 for the thermodynamic properties

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

of ordinary water substance for general and scientific use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–535, DOI: 10.1063/1.1461829.

(51) Muller, E. A.; Ervik, Å.; Mejía, A. A Guide to Computing Interfacial Properties of Fluids from Molecular Simulations. *Living J. Comp. Mol. Sci.* **2020**, *2*, 21385, DOI: 10.33011/livecoms.2.1.21385.

(52) Hulshof, H. Ueber die Oberflächenspannung. *Ann. Phys. (Berl.)* **1901**, *309*, 165–186, DOI: 10.1002/andp.19013090110.

(53) Vargaftik, N. B.; Volkov, B. N.; Voljak, L. D. International Tables of the Surface Tension of Water. *J. Phys. Chem. Ref. Data* **1983**, *12*, 817–820, DOI: 10.1063/1.555688.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Lessons learned from multi-objective automatic optimizations of classical three-site rigid water models using microscopic and macroscopic target experimental observables

Mattia Perrone[1], Riccardo Capelli[2], Charly Empereur-mot[3], Ali Hassanali[4], and Giovanni M. Pavan[1,3,*]

[1]Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy
[2]Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, I-20133 Milano, Italy
[3]Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Polo Universitario Lugano, Campus Est, Via la Santa 1, CH-6962 Lugano-Viganello, Switzerland
[4]The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy
[*]giovanni.pavan@polito.it

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Model	Refs.	Sigma [nm]	Epsilon [kJ/mol]	O charge (e)	O-H distance [nm]	H-H distance [nm]
SPC	ref.[1]	0.31655700	0.65062900	-0.820000	0.1000000	0.1633000
SPCE	ref.[2]	0.31655700	0.65062900	-0.847600	0.1000000	0.1633000
SPCEb	ref.[3]	0.31657195	0.64977520	-0.847600	0.1010000	0.1649300
TIP3P	ref.[4]	0.31506100	0.63638600	-0.834000	0.0957200	0.1513900
TIP3P-FB	ref.[5]	0.31779646	0.65214334	-0.848448	0.1011811	0.1638684
OPC3	ref.[6]	0.31742704	0.68369070	-0.895170	0.0978882	0.1598507
OPTI 1T		0.31921898	0.61796336	-0.935014	0.0940835	0.1518014
OPTI 3T		0.31657153	0.69750474	-0.889239	0.0977138	0.1590416

Table S1: Summary of the classical three-site rigid water models compared in this work.

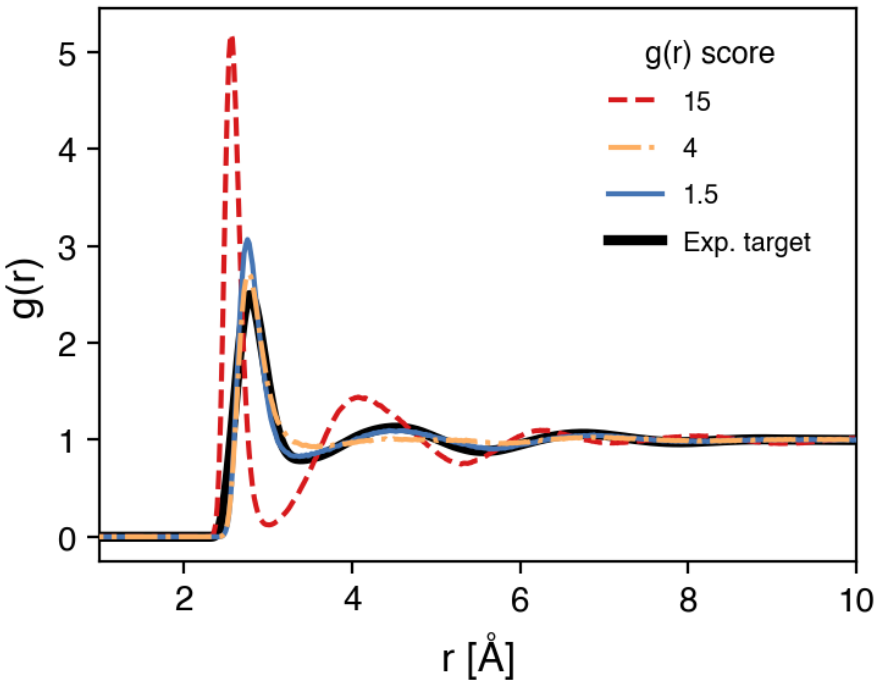

Figure S1: Comparison of experimental RDF and simulated RFDs examples scored according to our metrics. Notably, the orange curve receives a higher score (i.e., worse performance) compared to the blue one, despite capturing the first peak more accurately. The difference in scores arises from the orange curve's poorer reproduction of the second and third peaks. This observation highlights the scoring function preference for favoring long-range reproduction of RDFs in the evaluation of simulated curves.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

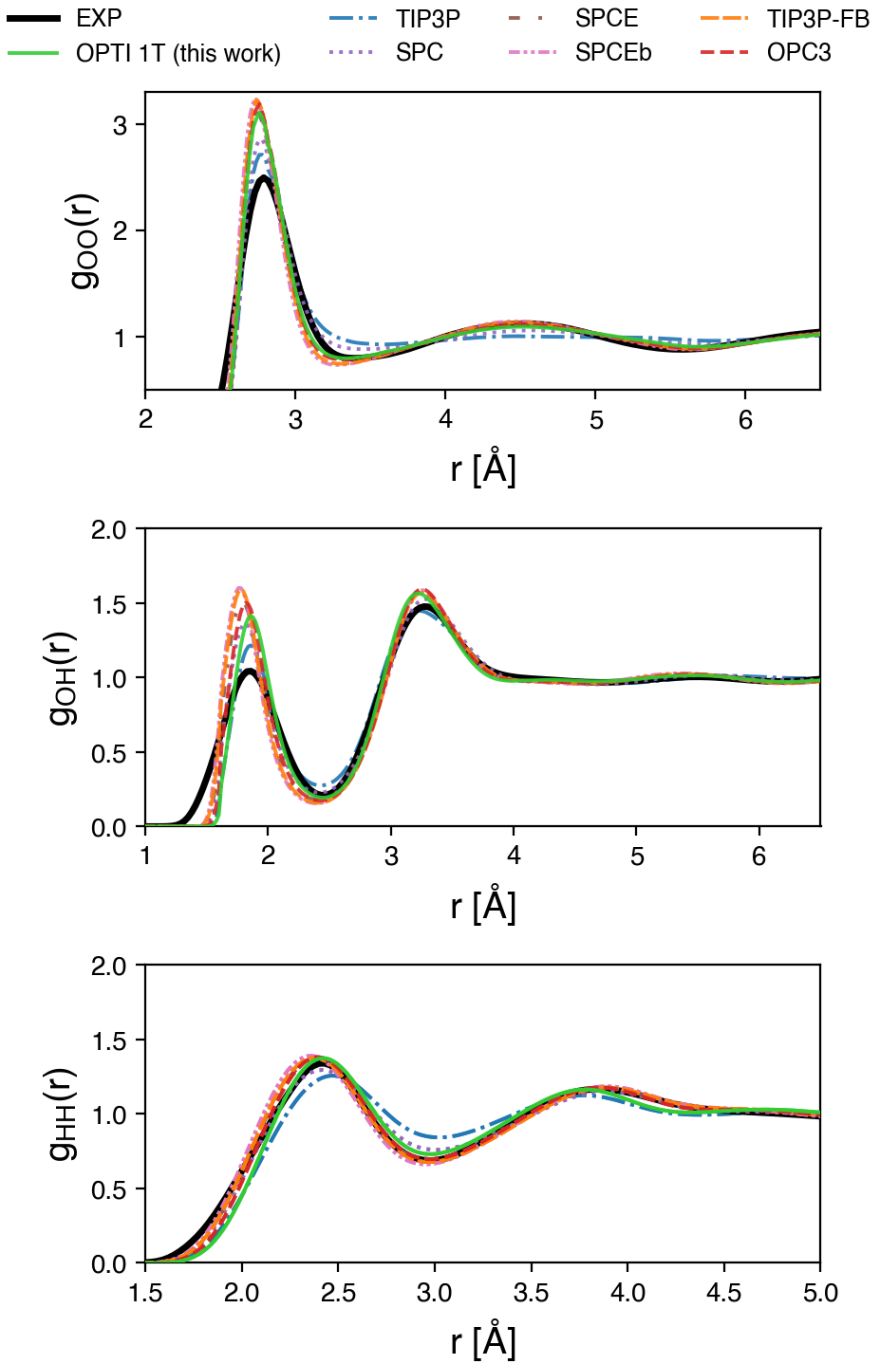

Figure S2: Enlarged plot of the reproduction of RDFs contained in Fig.2**a**

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

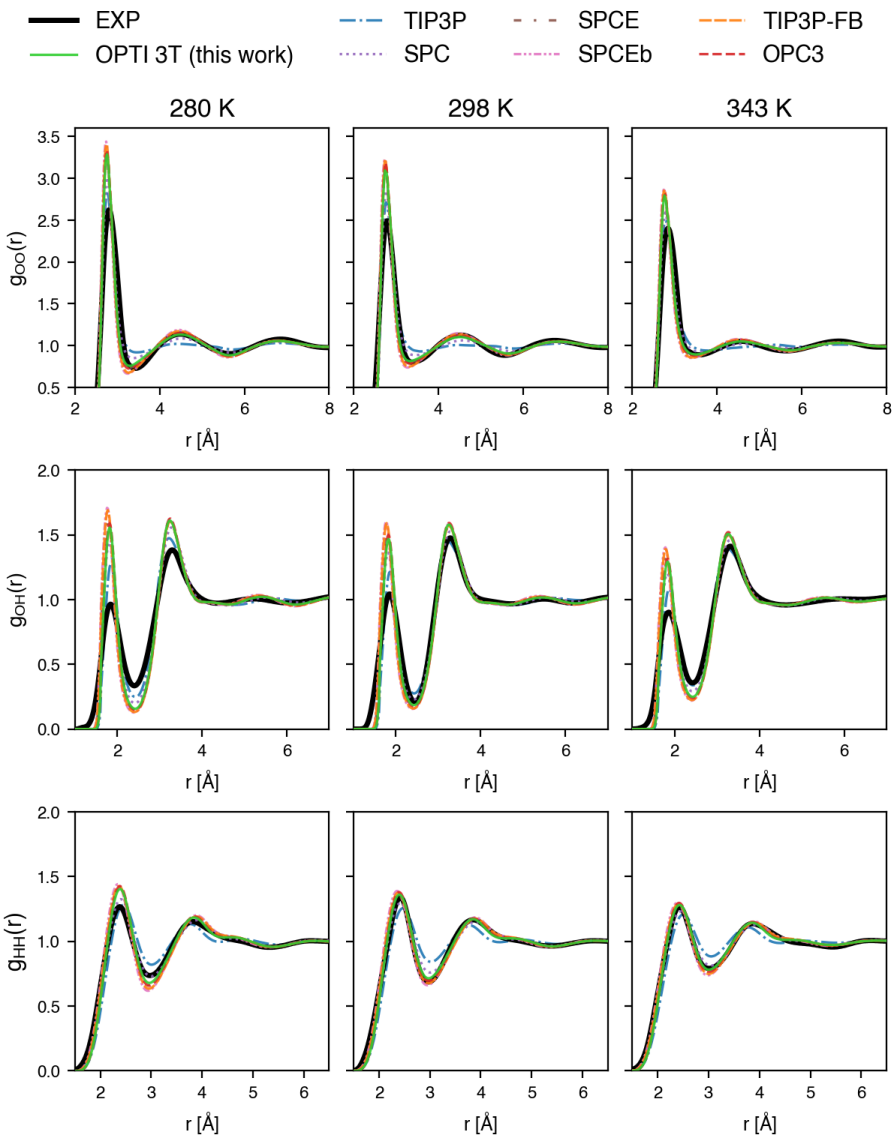

Figure S3: Radial distribution functions obtained of the model OPTI-3T along with a comparison with other models.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

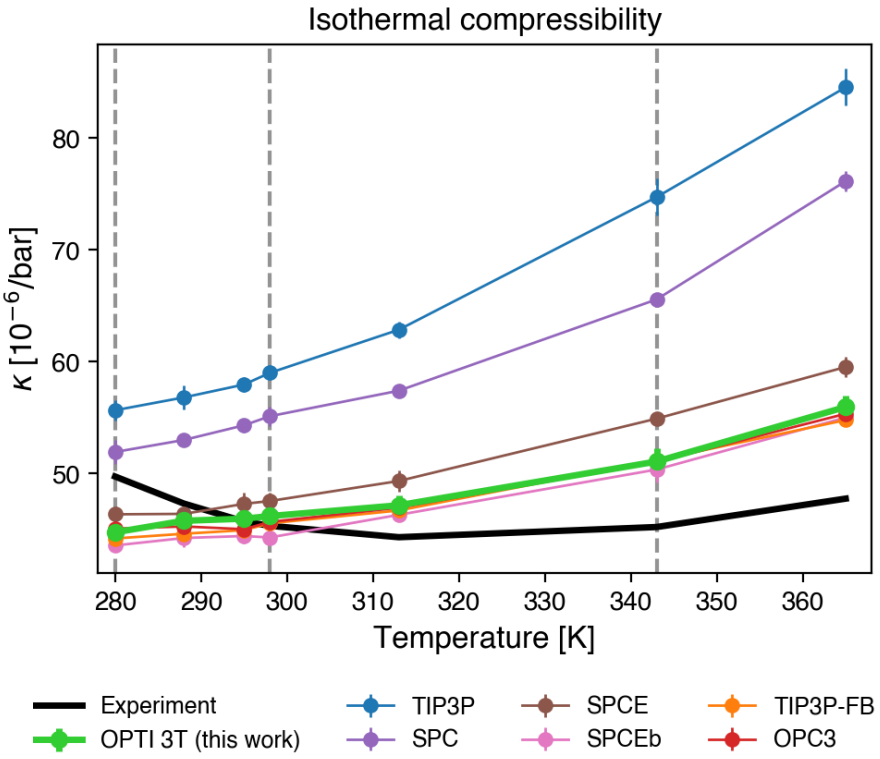

Figure S4: Isothermal compressibility as a function of temperature. Dashed vertical gray lines indicate the temperature at which the model was trained.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

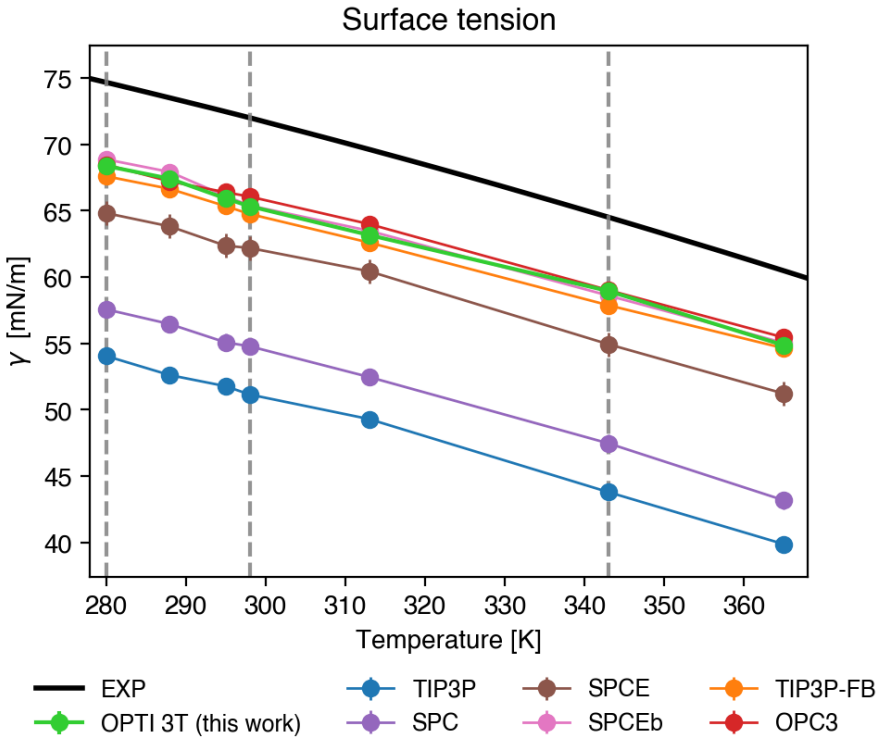

Figure S5: Surface tension as a function of temperature. Dashed vertical gray lines indicate the temperature at which the model was trained.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

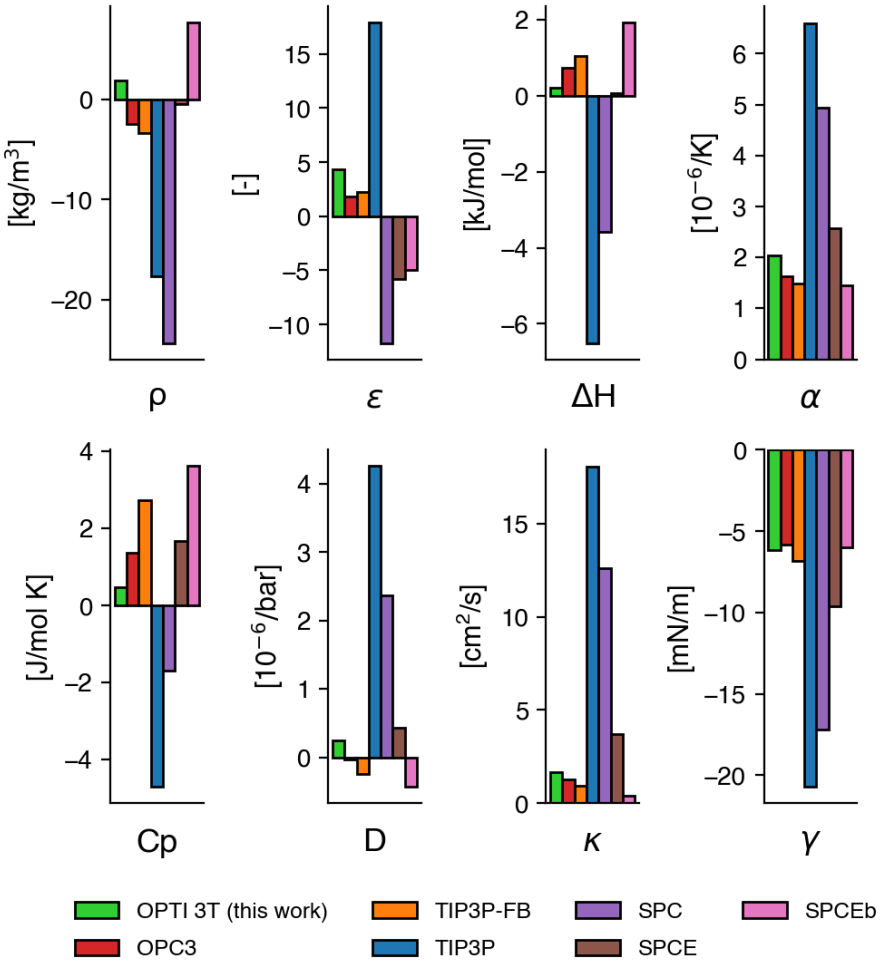

Figure S6: Average deviation of simulated observables with respect to experimental data in the liquid regime. The plots show water density, static dielectric constant, enthalpy of vaporization, thermal expansion coefficient, isobaric heat capacity, diffusion coefficient, adiabatic bulk modulus and surface tension.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

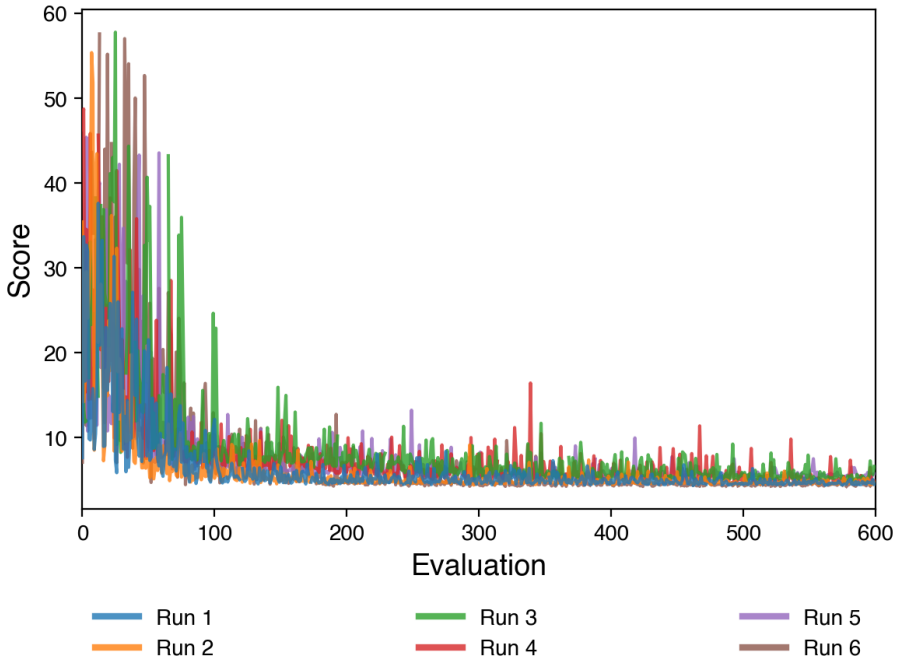

Figure S7: Score as a function of iteration number of the six optimizations. Each color represent an independent optimization run.

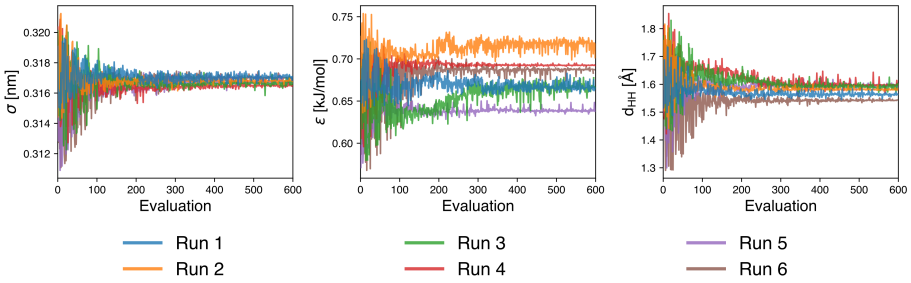

Figure S8: Values of sigma, epsilon and distance between the Hydrogens atoms as a function of iteration number. Each color represent an independent optimization run.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

References

[1] Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. Intermolecular forces. 1981; pp 331–342.

[2] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

[3] Takemura, K.; Kitao, A. *J. Phys. Chem. B* **2012**, *116*, 6279–6287.

[4] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

[5] Wang, L.-P.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.

[6] Izadi, S.; Onufriev, A. V. *J. Chem. Phys.* **2016**, *145*, 074501.

https://doi.org/10.26434/chemrxiv-2023-88fpn **ORCID:** https://orcid.org/0000-0002-3473-8471 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0