In situ parameterisation of SCC-DFTB repulsive potentials by iterative Boltzmann inversion

Manuel Doemer\textsuperscript{a}, Elisa Liberatore\textsuperscript{a}, Jan M. Knaup\textsuperscript{ab}, Ivano Tavernelli\textsuperscript{a} & Ursula Rothlisberger\textsuperscript{a}

\textsuperscript{a} Laboratoire de Chimie et Biochimie Computationelle, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

\textsuperscript{b} Bremen Center of Computational Materials Science, University of Bremen, Bremen, Germany

Accepted author version posted online: 10 Sep 2013. Published online: 23 Oct 2013.

To cite this article: Manuel Doemer, Elisa Liberatore, Jan M. Knaup, Ivano Tavernelli & Ursula Rothlisberger (2013) In situ parameterisation of SCC-DFTB repulsive potentials by iterative Boltzmann inversion, Molecular Physics: An International Journal at the Interface Between Chemistry and Physics, 111:22-23, 3595-3607, DOI: 10.1080/00268976.2013.842011

To link to this article: http://dx.doi.org/10.1080/00268976.2013.842011

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
INVITED ARTICLE

In situ parameterisation of SCC-DFTB repulsive potentials by iterative Boltzmann inversion

Manuel Doemera, Elisa Liberatorea, Jan M. Knaupa,b, Ivano Tavernellia and Ursula Rothlisbergera,*

aLaboratoire de Chimie et Biochimie Computationelle, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; bBremen Center of Computational Materials Science, University of Bremen, Bremen, Germany

(Received 2 May 2013; accepted 27 May 2013)

The parameterisation of accurate and transferable repulsive potentials is a key ingredient for the self-consistent-charge density functional tight-binding method (SCC-DFTB). In the conventional parameterisation scheme the balanced description of different chemical environments involves significant human effort and chemical intuition. In this work, we propose an in situ parameterisation method with reduced transferability but maximal accuracy for the chemical and physical environment under investigation. Starting from an initial guess, we used iterative Boltzmann inversion to successively improve the repulsive potentials. The corrections were extracted iteratively from the differences in the radial distribution functions with respect to a density functional reference calculation. With this protocol convergence was reached within a few iterations involving only minimal human input. We applied this new scheme to liquid water at ambient conditions, a particularly challenging case for conventional SCC-DFTB. The newly determined parameters lead to a clear improvement of both the structural and dynamical properties of liquid water.

Keywords: self-consistent-charge density functional tight-binding; repulsive potentials; iterative Boltzmann inversion; liquid water

1. Introduction

The choice of the underlying interaction model and its computational cost determines the accessible system size, time scale and accuracy of molecular dynamics (MD) simulations [1–3]. Classical molecular mechanics force fields (FFs) allow for sampling times up to microseconds even for large systems involved in biological and materials science applications [4,5]. On the other hand, their transferability is limited and their functional form usually does not allow the description of chemical reactions. Electronic structure-based ab initio MD (AIMD), on the other hand, provides the functional flexibility to describe instantaneous electronic rearrangements and the transferability to different chemical environments and points in the phase diagram. In particular, AIMD based on Kohn–Sham (KS) density functional theory (DFT) [6,7] has proven an efficient and flexible tool for electronic structure calculations in a variety of fields [8–10]. Many DFT methods provide a reasonable compromise between accuracy and accessible simulation time. Nevertheless, for larger system sizes sampling times are limited to a few tens to a hundred picoseconds and, in practice, often not long enough to extract structural, dynamical, thermodynamical or kinetic data at a desired accuracy from the trajectories.

Semi-empirical (SE) and tight-binding (TB) methods represent a bridge between DFT and FFs. They are capable of describing rearrangements of the electronic structure, while being typically 2-3 orders of magnitude faster than DFT approaches [11]. The limitations in sampling time and system size of the latter can, therefore, be pushed further towards the regime currently dominated by FF methods. Two large families of such methods exist: The traditional SE wave function-based methods like the popular MNDO [12], AMx [13] and PMx [14] schemes, which are derived from Hartree–Fock theory, and the KS DFT [7,15] based TB methods [16].

A variant of the TB family is the self-consistent-charge density functional-based tight binding method (SCC-DFTB) [17,18]. It has been applied to many problems in biochemistry and material science [19–21]. The method is entirely formulated in terms of parameterised two-body interactions, which are naturally divided into a band structure term of a core Hamiltonian, a charge fluctuation term and a sum of repulsive potentials. Obviously, the accuracy and transferability of the method is determined to a large extent by the parameterisation.

In the traditional parameterisation scheme, the repulsive potentials in SCC-DFTB have to be individually constructed for every possible pair of elements by fitting to chosen reference data. In addition, they have to be tested case by case and therefore a considerable amount of human effort and chemical intuition is needed to achieve accurate and transferable results for a new element [21]. Attempts towards automated parameterisation schemes have been

*Corresponding author. Email: ursula.roethlisberger@epfl.ch
developed based on the solution of a linear inverse problem [22,23] and an approach similar to the ‘Learn on the Fly’ [24] method involving a genetic algorithm for parameter optimisation [25].

The present work focuses on an in situ parameterisation scheme for the generation of system-specific repulsive potentials from higher level computational reference data. In this exploratory study, we especially probe the flexibility of the repulsive potentials to account for subtle effects of the intermolecular interactions in the condensed phase.

We employed the so-called iterative Boltzmann inversion (IBI) method, which has been developed previously in the context of the parameterisation of empirical potentials from diffraction data [26] and extended to deriving coarse-grained models based on atomistic simulations [27]. IBI represents a standard framework to parameterise interaction potentials of a given model in order to reproduce reference radial pair distribution functions (RDFs) in condensed-phase systems with a minimal amount of manual interference.

Here, we adopted a similar approach to parameterise the SCC-DFTB repulsive potentials in order to reproduce the RDFs from DFT calculations, while keeping the electronic parameters unchanged. In this way, we obtained system-specific parameters that allow for simulations with an accuracy comparable to DFT at the computational cost of SCC-DFTB. Since the RDFs, as the target quantities in the parameterisation protocol, are extracted from finite temperature MD trajectories in the condensed phase, effects of the chemical environment, temperature and pressure are automatically taken into account.

As a challenging test case, we applied this scheme to the case of liquid water at ambient conditions. Water is an ubiquitous solvent in biochemical processes [28,29] and the accurate description of its structural and dynamical properties is essential for the modelling of chemical reactions in aqueous solution [30,31]. Unfortunately, water has been notoriously difficult to model with SE and TB methods. Only very recently improvements could be achieved by extending NDDO type of methods via self-consistent-charge density functional tight-binding (SCC-DFT) [18] and an approach similar to the ‘Learn on the Fly’ [24,25] method involving a genetic algorithm for parameter optimisation [25].

SCC-DFTB [18] is the self-consistent extension of the original DFTB [17] method, an approximate electronic structure method based on KS DFT. The derivation can be found in the seminal papers [17,18] and has been revisited in numerous reviews [19,38,39].

2. Methods

2.1. The self-consistent-charge density functional tight-binding method

Throughout this paper, we use \( r \) for the spatial coordinate vectors of electrons, while nuclear coordinates of an atom \( \alpha \) of element \( a \) are denoted by \( \mathbf{R}_\alpha \). The internuclear distance between two atoms becomes \( R_{\alpha\beta} = |\mathbf{R}_\beta - \mathbf{R}_\alpha| \).

SCC-DFTB [18] is the self-consistent extension of the original DFTB [17] method, an approximate electronic structure method based on KS DFT. The derivation can be found in the seminal papers [17,18] and has been revisited in numerous reviews [19,38,39]. We therefore start our discussion immediately with the total energy expression of SCC-DFT, which can be written as:

\[
E_{\text{SCC-DFTB}} = E_{\text{core}} + E_{\text{dq}} + E_{\text{rep}},
\]

(1)

where \( E_{\text{core}} \) represents the contribution due to a core Hamiltonian \( \hat{H}_0[\rho_0] \):

\[
E_{\text{core}} = \sum_{\text{occ}} n_i \int \psi_i^* \hat{H}_0[\rho_0] \psi_i \, d\mathbf{r},
\]

(2)

with the corresponding KS orbitals \( \{ \psi_i \} \) and occupation numbers \( \{ n_i \} \). The core Hamiltonian \( \hat{H}_0 \) is constructed based on a reference density \( \rho_0(\mathbf{r}) = \sum_{\alpha} \rho_0(\alpha)(\mathbf{r} - \mathbf{R}_\alpha) \). The latter is a superposition of confined atomic reference densities, \( \rho_0(\alpha) \), which are pre-determined specific for each element \( a \).

The second term in Equation (1) takes into account the interaction of density fluctuations with respect to the atomic reference densities. This is modelled by a damped Coulomb potential \( \gamma_{\alpha\beta} \), which depends on the atomic Hubbard parameters of the elements \( a \) and \( b \) for the interaction of fluctuating atomic charges (\( \Delta q_\alpha \)) with respect to the reference charge distribution \( \rho_0(\alpha) \):

\[
E_{\text{dq}} = \frac{1}{2} \sum_{\alpha \neq \beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta}(R_{\alpha\beta}).
\]

(3)

The atomic charges are calculated from the instantaneous Mulliken population which is dependent on the orbitals \( \{ \psi_i \} \) and, therefore, a self-consistent procedure has to be adopted.

In order to solve for the minimal energy numerically the \( \psi_i \) are expanded in a minimal set of atom-centred basis functions \( \psi_i = \sum_{\nu} C_{i\nu} \phi_\nu \) and the total energy expression is recast in matrix form. The resulting core Hamiltonian matrix elements are constructed so as to retain only two-body contributions which can be parameterised based on the corresponding atomic reference states as functions of internuclear distances \( R_{ab} \) solely [40–42]. The parameters for the electronic terms are typically obtained from DFT calculations with the PBE [43] functional. Further extensions, such as spin polarisation [44,45], inclusion of dispersion...
effects [46,47] or orbital-dependent correlation have been proposed [48].

Finally, a set of additive pairwise 'repulsive potentials' is used to model the internuclear repulsion and to compensate for approximations made in the electronic terms. They are specifically parameterised for each pair of elements as functions of the internuclear distance:

$$E_{\text{rep}} = \frac{1}{2} \sum_{a \neq b} U_{ab}(R_{ab}).$$

(4)

In the traditional approach to parameterise repulsive potentials, $E_{\text{rep}}$ is defined with respect to a DFT reference. A suitable reference system is chosen to parameterise a specific repulsive potential, for example $H_2O$ for the O–H potential. The total energy at the DFT level is then computed along one O–H bond stretching coordinate. The corresponding repulsive potential is then fitted to match the DFTB total energies to the DFT references: $E_{\text{rep}}(ROH) = E_{\text{DFT}}(ROH) - [E_{\text{core}}(ROH) + E_{\text{qz}}(ROH)]$. In the case of the mio-0-1 parameter set for biological systems, which we used throughout this work for the electronic parameters and whose repulsive potentials served as an initial guess for our fitting scheme, the B3LYP [49,50] functional has been used to compute the DFT reference energies. In order to describe an $H_2O$ molecule two more potentials (O–O and H–H) are needed. The oxygen molecule served as a reference system to parameterise the O–O potential and $H_2$ to parameterise the H–H repulsive potential. This example illustrates the human effort to produce accurate and at the same time transferable repulsive potentials for each individual pair of chemical species. In the case of carbon, for instance, different hybridisation states have to be taken into account, since the typical range of the repulsive potentials can extend over various bond types. In the case of the C–C repulsive potential, energy curves for single, double and triple bonds have to be considered and matched in a way to obtain a balanced description [18,51]. A similar approach has been followed in an effort to fit the repulsive potentials to specific vibrational modes [52,53]. First attempts towards an automated parameterisation of the repulsive potentials have been undertaken following a genetic algorithm-based approach [25] to fit total energies along a pre-defined reaction path and based on the solution of a linear inverse problem [22]. In this work, we apply the iterative Boltzmann inversion protocol based on DFT reference trajectories to obtain in situ repulsive potentials for a given system, while keeping the electronic parameters fixed at their original SCC-DFTB values.

### 2.2. Iterative Boltzmann-inversion applied to repulsive SCC-DFTB potentials

The central quantity for the following discussion is the radial distribution function (RDF) $g_{ab}(R)$, which measures in a system of $N$ particles of different kinds $a, b, \ldots$ in a volume $\Omega$ the density of particles $b$ at a distance $R$ from a reference particle of type $a$ relative to the corresponding bulk density [54]:

$$g_{ab}(R) = \frac{\Omega}{4\pi r^2} p_{ab}(R),$$

(5)

with the probability density $p_{ab}(R)$ of distances $R$ between particle types $b$ and $a$, which can be readily extracted from the configurations of an MD trajectory. The pre-factor comes from the spherical average. Henderson’s theorem states that the potential energy function which gives rise to a given RDF is unique up to a constant [55]. Furthermore, it has been demonstrated that in the case of particles interacting by potentials including up to $n$-body terms, the system is completely described by all $n$-body correlation functions [56]. Therefore, in order to derive effective pair-potentials, such as the repulsive potentials $U_{ab}(R)$ considered here, we use the corresponding two-body correlation functions $g_{ab}(R)$ as target quantities. All remaining many-body effects are instead described by the electronic terms, which we do not aim to re-parameterise in this work.

The Boltzmann inversion (BI) [57] relates the radial distribution function with the Helmholtz free energy up to an arbitrary additive constant $C$:

$$F_{ab}(R) = -k_B T \ln g_{ab}(R) + C.$$

(6)

It represents the effective potential of mean force (PMF) between two particles of types $a$ and $b$. $T$ is the temperature of the system and $k_B$ is the Boltzmann constant. In the general case of finite density of the species $a, b$ the PMF is a free energy and only an approximation for the underlying interaction potential can be obtained from Equation (6), which can then be refined iteratively. Within the restrictions of the functional form of the chosen potentials and the simulation techniques, this procedure is likely to converge [26,27].

In this work, the goal was to derive repulsive potentials $U_{ab}(R)$ that reproduce the reference RDFs $g_{ab}(R)$ at the PBE level. This choice is consistent with the original parameterisation of the electronic terms. The same approach could be applied to target more accurate methods or even experimental data. Our practical implementation of the iterative BI protocol looks as follows: From the reference trajectory we extract the RDFs $g_{ab}(R)$ between pairs of particle types. We use the mio-0-1 parameter set as initial guess for the corresponding repulsive potentials $U_{ab}^{(0)}(R)$ and extract the corresponding RDFs $g_{ab}^{(0)}(R)$ from MD. We then correct the guess potentials by the difference in the PMFs with respect to the reference:

$$\Delta E_{ab}^{(0)}(R) = k_B T \ln \left[ \frac{g_{ab}^{(0)}(R)}{g_{ab}(R)} \right].$$

(7)
This step is iterated:

\[ U_{ab}^{i+1}(R) = U_{ab}^{i}(R) + \Delta F_{ab}^{i}(R) \]  

(8)

with

\[ \Delta F_{ab}^{i}(R) = k_B T \ln \left( \frac{g_{ab}(R)}{g_{ab}(R)} \right) \]  

(9)

until the \( g_{ab}(R) \) are reproduced.

The internal geometries of the individual water molecules in liquid bulk water are very well reproduced by the standard mio-0-1 parameters. It can be expected that the dominant contributions to the intermolecular interactions are due to the O–O and O–H potentials with only minor contributions from the H–H potentials. In order to restrict our re-parameterisation to a minimum, we modify only the O–O and O–H repulsive potentials, while keeping the H–H repulsive potential fixed at the mio-0-1 parameter set.

### 2.3. Computational details

For computational efficiency, the IBI protocol was performed for a minimal system of 32 water molecules in a periodic box of 9.85\(^3\) Å\(^3\), which is sufficient to capture the main structural features. For a correct description of finer structural features and dynamical properties, finite size effects can, however, be important [58]. The results of the IBI are presented in Section 3.1.

The Car–Parrinello molecular dynamics reference trajectory of a periodic system containing 32 light water molecules was generated with the CPMD code [59] using the PBE [43] functional. Valence electrons were described using a plane-wave expansion up to a kinetic energy cut-off of 125 Ry, and soft norm-conserving Martins-Troullier pseudopotentials [60] were employed to represent the inter-atomic interactions. The propagation of the equations of motion was performed within the Car–Parrinello scheme [61] using a fictitious electron mass of 500 a.u. and a time step of 0.1 fs. The average temperature of this reference trajectory was 314 K and the total length 117 ps with configurations taken for analysis every 10 steps.

For the geometry optimisations on the isolated water dimer we used PBE with the aug-cc-pVQZ basis set, as implemented in Gaussian G09 [68] with tight convergence criteria.

### 2.4. Analysis methods

To evaluate the radial distribution functions \( g_{ab}(R) \), we computed the probability densities of distances \( R \) in Equation (5) within periodic boundary conditions from a finite number of \( M \) configurations extracted from an MD trajectory by means of histogramming as described in [69,70]. In order to obtain the average coordination number of atoms of type \( b \) around atoms of type \( a \) within a shell of radius \( R \), we integrate the corresponding radial distribution function:

\[ N_{ab}(R) = 4\pi \frac{N_{\text{pairs}}}{\Omega} \int_0^R R^2 g_{ab}(R) dR, \]  

(10)

where \( N_{\text{pairs}} \) is the number of unique \( a, b \) pairs. As a measure for the flexibility of the hydrogen bonds we analysed the probability distributions for the hydrogen-bond angles \( \delta = \angle_{OH-OH-OH} \), \( \epsilon = \angle_{OH-OH-OH} \), and \( \eta = \angle_{OH-OH-OH} \), as depicted in Figure 1. We considered angles within the first solvation shell with \( R_{OH} \leq 2.5 \) Å, corresponding to the first intermolecular minimum in \( g_{OH} \).

The self-diffusion coefficient \( D \) was evaluated using the Einstein relation from plotting the mean-square displacement (MSD) of the oxygen atoms (after removing the centre of mass translation) versus time:

\[ \langle |\mathbf{R}(t) - \mathbf{R}(0)|^2 \rangle = 6t D. \]  

(11)
The MSD curve was averaged over several overlapping data blocks with a length of half of the total simulation time. For each data block a different time origin $t_0$ ranging from 0 to half of the total simulation time was chosen. $D$ was then determined from a least-square fit to the MSD curve in the linear regime.

In order to analyse the time scale for molecular reorientations, we calculated the orientational auto-correlation functions:

$$C_{l=1,2} = \frac{1}{N} \sum_i \langle P_l[\cos \theta_i(t)] \rangle$$

(12)

in which the instantaneous molecular orientation was calculated as the axis connecting the midpoint of the two hydrogen atoms with the oxygen atom. $\theta_i(t)$ is the angle between the direction of the water molecule $i$ at times $t$ and $t_0$. $P_l$ is the Legendre polynomial of order $l$. Again we averaged over an ensemble of data blocks of 20 ps length with different time origins $t_0$ up to the total simulation length minus 20 ps. In order to extract time constants $\tau_l$ for this relaxation process we fitted the functions $C_l$ (for $l = 1, 2$) to the functional form $A_l \exp[-(t/\tau_l)^2]$. To account for the different time scales among the methods and relaxation processes we chose the fit ranges for $C_1$ 0.1-20 ps for all methods and for $C_2$, 0.5-20 ps for PBE and mio-0-1+BI and 0.1-20 ps for mio-0-1.

3. Results and discussion

3.1. Parameterisation of the repulsive potentials

During the iterative Boltzmann inversion we improve the O–O and O–H repulsive potentials simultaneously, while the H–H potential and all the electronic parameters are kept fixed at the values of the mio-0-1 set. Figures 2 and 3 show the radial distribution functions (bottom), the corresponding potentials of mean force (top) and the resulting repulsive potentials (middle) over the course of the iterative Boltzmann inversion for the O–O and O–H interactions. The initial radial distribution functions, $g_{OO}^{(0)}$ and $g_{OH}^{(0)}$ show the DFTB deficiencies in describing the structure of liquid water. The peak corresponding to the first solvation shell in the initial $g_{OO}^{(0)}$ is positioned at $\sim 0.1$ Å too long a distance with respect to the PBE reference $g_{OO}^{\text{PBE}}$. Furthermore, the structure of the second solvation shell is barely visible at this scale. There is only a shallow minimum at $\sim 3.8$ Å. The structure of $g_{OH}^{(0)}$ in Figure 3 looks qualitatively better, but the hydrogen-bonding peak at $\sim 1.8$ Å is slightly shifted towards longer distances, and the second intermolecular peak at $\sim 3.5$ Å is not structured enough. This can be linked to the form of the initial repulsive potential $U_{OH}^{(0)}$, which goes to zero at 1.8 Å, in the middle of the hydrogen-bonding peak in $g_{OH}^{(0)}$. It also features a minimum at 1.3 Å and a local maximum at 1.4 Å. The latter can directly be linked with the shift of the hydrogen-bond peak. This can be seen in the corresponding difference in the potential of mean force with respect to the PBE reference $\Delta F_{OH}^{(0)}$. The spike in the positive range centred at 1.4 Å indicates that the potential should be more repulsive in this range, while at 1.6 Å the potential is too repulsive. Furthermore, the maximum at 2.5 Å indicates that a repulsive contribution is needed in order to enhance the structure between the first and second intermolecular peak of the RDF.

Compared to the mio-0-1 range, we extended the repulsive O–H potential from 1.8 Å to 2.9 Å, with an additional 0.5 Å to allow the first two derivatives to go to zero smoothly. The region over which the repulsive potential was adjusted over the course of the IBI is indicated with black dotted vertical lines in the top section. Over the iterations the $U_{OH}$ becomes more repulsive in the range below 1.4 Å, attractive in the range 1.4–1.9 Å and repulsive between 1.9 Å and 2.9 Å. Along this series the features in the PMF differences decrease. Since the O–H and O–O potentials are adjusted simultaneously, unphysical features appear at the early stages, like the double peak in the $g_{OH}^{(0)}$ at iteration 1. This is smoothed out quickly and already at the third iteration $g_{OH}^{(3)}$ is in good agreement with the PBE reference.

The tail of the original mio-0-1 potential for the O–O interaction is plotted as $U_{OO}^{(0)}$ in the middle panel of Figure 2. It goes to zero at 2.2 Å. This is shorter than the first intermolecular peak in $g_{OO}$ and any modification limited to the range $\leq 2.2$ Å within a reasonable energy range will not sufficiently influence the intermolecular interactions at larger distances. We therefore extended the range of the potential up to 5.0 Å. The initial PMF difference with respect to the PBE reference, $\Delta F_{OO}^{(0)}$, shows a binding feature with a pronounced tip at 2.5 Å in order to shift the first peak in the $g_{OO}^{(0)}$ towards smaller distance, a repulsive feature centred at 3.2 Å to account for the first intermolecular
Figure 2. Variations of the PMF difference with respect to the PBE reference (top), the O–O repulsive potential (middle) and the O–O radial distribution function (bottom) during the iterative Boltzmann inversion. The dotted lines indicate the range of the repulsive potential subject to optimisation.

minimum and, finally, an attractive feature at 4.3 Å to shape the second solvation shell. This is directly reflected in the improvements of $U_{OO}$ over the iterations, while, at the same time, the PMF differences decrease. Again, the RDF shows a peculiar feature at 2.6 Å in the first iteration $g_{OO}^{(1)}$, which is caused by the simultaneous correction scheme, but then improves quickly and already $g_{OO}^{(3)}$ corresponds well to the reference $g_{OO}^{PBE}$. 
Two critical consequences for the structure of the repulsive potentials arise from the results of this fitting scheme, which could both be seen as a breaking with the traditional definitions and derivations of parameters in TB methods in general and DFTB in particular. First of all, the range of the potentials extends by far beyond the first neighbour shell. But the restriction to the first neighbours in the parameterisation of the DFTB repulsive potentials is mainly...
a consequence of the fact that only isolated molecules were used as reference systems in the traditional parameterisation schemes, where only relatively short-range interactions can be parameterised. In an *in situ* parameterisation scheme as applied in this work, however, the effect of the surrounding can directly be taken into account and the transferability is not a primary goal. We do not aim for a new set of universal SCC-DFTB parameters, but rather a special set for water–water interactions at ambient conditions. Furthermore, already $U^{(0)}_{OH}$ reaches up to the first intermolecular peak in $g_{OH}$ and therefore extends over the first neighbour interactions. Second, parts of the ‘repulsive potentials’ become attractive, since at longer ranges the core–core repulsion is not the dominant interaction in the correction terms. Note as well, that also the $U^{(0)}_{OH}$ is not monotonically decreasing, but shows a local minimum at 1.3 Å. Moreover, the standard mio-0-1 H–H repulsive potential has an attractive part as well.

In the following sections, we call the IBI converged parameter set mio-0-1+BI.

As a first performance test of the potentials for larger system sizes, the new parameterisation was applied to a periodic system of 64 deuterated water molecules and its performance was evaluated by detailed analysis of the structural and dynamical properties.

### 3.2. Structural properties

Figure 4 compares the mio-0-1+BI RDFs for O–O, O–H and H–H with the PBE reference, standard mio-0-1 and experiment. The PBE curves are over-structured compared to experiment and, consequently, the mio-0-1+BI curves as well. Interestingly, although the $U_{HH}$ was not refitted, the $g_{HH}$ of mio-0-1+BI matches closely the PBE curve, while the one of the original mio-0-1 is actually closer to the experimental data. Note that the intramolecular peak in the $g_{HH}$ at 1.5 Å is influenced by nuclear quantum effects, which are not taken into account neither by the PBE, nor any SCC-DFTB MD simulation.

Average coordination numbers were determined according to Equation (10) by choosing the outer integration limit as the minimum between first and second intermolecular peaks in the RDFs for O–O and O–H (intersections of dotted horizontal/vertical lines in Figure 4). This definition is somewhat ill-defined in the under structured mio-0-1 O–O RDF, where the coordination number amounts to double the experimental value when integrating up to the shallow minimum between first and second peak [36]. The mio-0-1+BI RDF, on the other hand, follows closely the PBE reference with a pronounced minimum between the two intermolecular peaks and, therefore, the coordination number can be determined as slightly below 4, as for PBE.

Figure 5 gives the probability distributions for the hydrogen-bonding angles $\epsilon$ (a), $\eta$ (b) and $\delta$ (c), as defined in Figure 1. Experimental data [71], based on NMR measurements, are only available for $\epsilon$. The red curve was reproduced from data recorded at 300 K. $P(\epsilon)$ for PBE is slightly too narrow and peaked at a slightly smaller angle compared to the experimental curve, in line with the general tendency for over structured water of PBE. Mio-0-1, on the other hand, produces a too broad distribution with a maximum at larger $\epsilon$ than the experimental reference. Also $P(\eta)$ and $P(\delta)$ of mio-0-1 are significantly broadened compared to the PBE reference, indicating an overall too high flexibility of the hydrogen bonds. The optimised repulsive potentials in mio-0-1+BI improve the angular structure of the hydrogen-bonding network in liquid water and the corresponding probability distributions follow closely the PBE curves.

### 3.3. Dynamical properties

Since the finite size effects can to a certain extent affect dynamical properties [58], our analysis on a periodic system of 64 water molecules represents a compromise between accuracy and computational costs. The self-diffusion coefficients are given in Table 1. In accord with the over structuring of PBE water, the dynamical properties reflect a too small fluidity. The self-diffusion coefficient is underestimated by a factor of 8 with respect to the experimental value. In contrast, the mio-0-1 water is too fluid, with a 3.5 times higher diffusion coefficient than real (light) water. This is significantly reduced by the improved repulsive potentials mio-0-1+BI. Still the diffusion coefficient of 0.15 [Å²/ps] is 5 times higher than the PBE reference value but, by chance, it is very close to the experimental value.

Figure 6 shows the orientational autocorrelation functions $C_i(t)$ for $l = 1$ (a) and $l = 2$ (b) and the corresponding exponential fits. The fitting parameters are given in Table 2 and the time constants $t_1$ and $t_2$ can be compared to the experimental values. The orientational correlation times are much too long at the PBE level compared to the experimental data. Mio-0-1, on the other hand, predicts far too short correlation times, while the estimates obtained from mio-0-1+BI are in fortuitously good agreement with the experimental values.

### 3.4. Water dimer

Although the set of repulsive potentials was optimised for the condensed phase, the isolated water dimer should also be described accurately. The geometrical parameters reported in Table 3 for mio-0-1+BI do not differ substantially from the standard mio-0-1, being in generally good agreement with the experimental and the PBE references. The binding energy, however, changes more drastically from 3.5 kcal/mol to 4.5 kcal/mol, now being much closer to the reference value.
Figure 4. Comparison of computed RDFs (solid lines) using PBE, SCC-DFTB/mio-0-1 and SCC-DFTB/mio-0-1+BI with experimental data: O–O (a), O–H (b) and H–H (c). Integrated coordination numbers are printed as dotted lines for the O–O and O–H interactions only. Dotted vertical lines mark the minimum between first and second intermolecular peaks in the RDFs for O–O and O–H.
Figure 5. Probability distributions of the hydrogen-bonding angles $\epsilon$ (a), $\eta$ (b) and $\delta$ (c) as defined in Figure 1. Experimental values for $\epsilon$ were determined from NMR data at 300K [71].

Table 1. Experimental diffusion coefficient and calculated values using PBE, SCC-DFTB/mio-0-1 and SCC-DFTB/mio-0-1+BI.

<table>
<thead>
<tr>
<th>Method</th>
<th>Exp.</th>
<th>PBE</th>
<th>mio-0-1</th>
<th>mio-0-1+BI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.19</td>
<td>0.03</td>
<td>0.81</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>units: ($\text{Å}^2$/ps)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Prefactors $A_i$, time constants $\tau_i$ (in ps) and exponents $\zeta_i$ of the orientational auto-correlation functions $C_{i=1,2}$. Experimental values were measured at 300 K[73].

<table>
<thead>
<tr>
<th>Method</th>
<th>$A_{i=1}$</th>
<th>$\tau_{i=1}$</th>
<th>$\zeta_{i=1}$</th>
<th>$A_{i=2}$</th>
<th>$\tau_{i=2}$</th>
<th>$\zeta_{i=2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>–</td>
<td>4.76</td>
<td>–</td>
<td>–</td>
<td>1.92</td>
<td>–</td>
</tr>
<tr>
<td>PBE</td>
<td>0.92</td>
<td>35.63</td>
<td>0.79</td>
<td>0.78</td>
<td>14.55</td>
<td>0.73</td>
</tr>
<tr>
<td>mio-0-1</td>
<td>0.83</td>
<td>0.48</td>
<td>0.96</td>
<td>0.64</td>
<td>0.19</td>
<td>0.87</td>
</tr>
<tr>
<td>mio-0-1+BI</td>
<td>0.86</td>
<td>4.59</td>
<td>0.90</td>
<td>0.88</td>
<td>1.33</td>
<td>0.71</td>
</tr>
</tbody>
</table>
4. Conclusions

We have applied the iterative Boltzmann inversion method to parameterise repulsive SCC-DFTB potentials for the oxygen–oxygen and oxygen–hydrogen interactions in liquid water at ambient conditions. The hydrogen–hydrogen repulsive potential and all electronic parameters were kept unchanged with respect to the original parameter set mio-0-1. We started with initial guesses from the standard mio-0-1 parameter set derived for biological systems, which has been shown to perform unsatisfactory in describing the properties of liquid water [35–37]. We used a reference trajectory of a periodic system of 32 water molecules at the PBE level of theory. Along the course of the IBI we simultaneously improved the oxygen–oxygen and oxygen–hydrogen RDFs with respect to the PBE references by the corresponding PMF differences. Convergence was reached within three iterations, resulting in the final parameter set mio-0-1+BI.

As a first performance test of the new potentials for larger system sizes, structural and dynamical properties were analysed on periodic systems of 64 deuterated water molecules. Besides the O–O and O–H RDFs, which were used as target quantities during the fitting protocol...
employing the 32 water system, also the final H–H RDF matches the PBE reference well. The probability distributions of the hydrogen bonding angles are significantly broadened with the mio-0-1 parameter set with respect to the PBE and experimental references. The improved parameters mio-0-1+BI yield angular distributions close to the PBE references. In addition, the overestimated mobility in terms of self-diffusion coefficient of mio-0-1 water could be corrected by the newly derived repulsive potentials. Possibly due to a fortunate cancellation of errors, the diffusion coefficient obtained for the new parameter set is even closer to the experimental value than the PBE estimate. Furthermore, the orientational correlation times of the mio-0-1+BI simulations are in good agreement with the experimental references.

Besides improving the structural and dynamical properties of bulk water, the mio-0-1+BI set also yields an improved binding energy of the water dimer, while at the same time preserving the good performance in the geometrical parameters of the standard mio-0-1 parameter set with respect to the PBE reference.

In order to assess the transferability of the new water parameters, we have performed first tests on the structural properties of water–methanol mixtures at different methanol concentrations. Our general conclusion is that the new repulsive potentials improve the water–water RDFs with respect to experiment even at higher methanol concentrations, but they are not able to fully account for the subtle hydrogen-bonding effects modulating the water–methanol and methanol–methanol RDFs over some concentration range [76]. Such effects could be addressed by additional adjustments in the repulsive potentials employing a similar IBI approach for the corresponding repulsive pair potentials involving carbon, by extension of the SCC formulation to multipole moments [77] or by further developments in the electronic part of the SCC-DFTB Hamiltonian.

Acknowledgements
The authors would like to thank Ari Seitsonen for useful discussions. The CADMOS project, EPFL central computing facilities and the Swiss Super Computing Center (CSCS) are acknowledged for providing computer time.

Funding
This work was funded by the Swiss National Science Foundation [grant number 200020-130082]; the NCCR-MUST interdisciplinary research programme.

References

Molecular Physics
3607