

# Hydration free energies of cyanide and hydroxide ions from molecular dynamics simulations with accurate force fields

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 20303

Myung Won Lee and Markus Meuwly\*

The evaluation of hydration free energies is a sensitive test to assess force fields used in atomistic simulations. We showed recently that the vibrational relaxation times, 1D- and 2D-infrared spectroscopies for  $\text{CN}^-$  in water can be quantitatively described from molecular dynamics (MD) simulations with multipolar force fields and slightly enlarged van der Waals radii for the C- and N-atoms. To validate such an approach, the present work investigates the solvation free energy of cyanide in water using MD simulations with accurate multipolar electrostatics. It is found that larger van der Waals radii are indeed necessary to obtain results close to the experimental values when a multipolar force field is used. For  $\text{CN}^-$ , the van der Waals ranges refined in our previous work yield hydration free energy between  $-72.0$  and  $-77.2$  kcal mol $^{-1}$ , which is in excellent agreement with the experimental data. In addition to the cyanide ion, we also study the hydroxide ion to show that the method used here is readily applicable to similar systems. Hydration free energies are found to sensitively depend on the intermolecular interactions, while bonded interactions are less important, as expected. We also investigate in the present work the possibility of applying the multipolar force field in scoring trajectories generated using computationally inexpensive methods, which should be useful in broader parametrization studies with reduced computational resources, as scoring is much faster than the generation of the trajectories.

Received 28th June 2013,  
Accepted 27th September 2013

DOI: 10.1039/c3cp52713a

[www.rsc.org/pccp](http://www.rsc.org/pccp)

## 1 Introduction

Solvation of neutral and ionic species is important for characterizing chemical processes in the liquid phase.<sup>1,2</sup> Of particular interest are the organization and dynamics of solvent molecules around solutes<sup>3</sup> or the behaviour of ionic species at interfaces.<sup>4</sup> Ionic solvation is also of considerable relevance in understanding electrolyte solutions and the behaviour of salts at surfaces. Both the structure and dynamics of the solvent surrounding ions are of fundamental interest, including *e.g.* their effects on protein solubility and stability.<sup>5</sup>

An important and experimentally accessible property of simple ions is their free energy of solvation, which can also be determined from molecular dynamics (MD) simulations. A large effort has been made over the past few decades for developing computationally feasible and reliable methods to determine such free energies.<sup>6–19</sup> The most rigorous approach is to treat the solvent explicitly and determine the solvation free energy either using thermodynamic integration (TI), free energy perturbation (FEP) or related methods. For larger systems – such as

proteins – these methods become computationally prohibitive and thus more approximate implicit schemes such as MM/PBSA and MM/GBSA<sup>20–23</sup> have been devised. In electronic structure calculations the polarized continuum model (PCM)<sup>24</sup> is commonly employed whereby the molecular electrostatic potential or the electric field at preselected points is used to take into account environmental effects of a solvent on the properties of a molecular system. Alternatively, the Langevin dipole method considers a solute surrounded by solvents represented as point dipoles placed on a grid.<sup>14</sup> While these methods can be appropriate for the calculation of the solvation free energy for a small solute molecule, more extensive sampling of the conformation would be needed for more complicated systems.

At the atomistic level, solvation free energies are also potentially sensitive to the intermolecular interactions.<sup>25</sup> Depending on the force field used, the conformations sampled and the free energies determined from them will differ. Hence, in standard force field optimizations, solvation free energies are an important thermodynamic property often included in the fitting.<sup>26–29</sup> In our previous work on cyanide in water,<sup>30</sup> it was found that the vibrational relaxation time of the solute sensitively depends on the van der Waals parameters when accurate, multipolar

Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland. E-mail: [m.meuwly@unibas.ch](mailto:m.meuwly@unibas.ch)

electrostatics is used in the simulations. Scaling the Lennard-Jones (LJ) size parameter ( $\sigma$ ) by a few percent almost quantitatively reproduces experimental vibrational relaxation times of  $\text{CN}^-$  in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$ .<sup>30</sup> Hence, it is of interest to investigate whether thermodynamic quantities such as the free energy of solvation are equally susceptible to the nonbonded interactions. In the present work, we determine the hydration free energy of the cyanide and hydroxide ions in water from atomistic simulations using a range of force fields, and investigate how specifically the LJ-ranges affect the solvation free energy of the ion for given multipolar electrostatics.

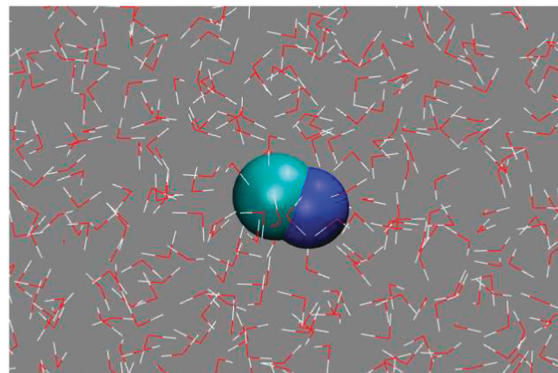
In many cases, it is considerably more expensive computationally to generate the configurations than to determine the experimental observables with which to compare. A possibly advantageous approach would be to sample configurations using less expensive simulations with a simpler force field, whereas “scoring” is carried out by using a computationally more demanding but more accurate method.<sup>31</sup> In order to quantitatively assess the influence of sampling with one model and scoring with another one, the configurations generated with computationally cheaper models have been scored with more elaborate and thus more expensive models in the present work.

While there are many different methods to carry out free energy calculations, the computationally expensive part of such an undertaking is the sampling of the configuration, which is the same for a range of methods. Therefore, it would be interesting to evaluate free energy using various methods and to compare the results.<sup>32,33</sup> In the present work, TI, FEP, and Bennett acceptance ratio (BAR) methods are applied to the same configurations and the hydration free energies of the cyanide and hydroxide ions are evaluated. In addition, we also compare the results with that obtained from the umbrella sampling (US), which needs different sampling. The results are then compared to experiments and the sensitivity of the computations to modifications in the force fields is determined.

## 2 Theory and computational methods

All molecular dynamics (MD) simulations were carried out using the CHARMM program<sup>34</sup> with provisions for multipolar interactions.<sup>35</sup> Spherical boundary conditions (SBCs) were employed with a cyanide or hydroxide ion solvated inside a water sphere of radius 19 Å containing 997 water molecules, with a water density close to 1 g cm<sup>-3</sup>. A snapshot of the simulation system is shown in Fig. 1. These simulations were carried out in the *NVT* ensemble. A Nosé–Hoover thermostat was employed with a thermal piston mass of  $Q = 50 \text{ kcal mol}^{-1} \text{ ps}^2$ .<sup>36,37</sup>

The nonbonded interactions (electrostatic and Lennard-Jones) were truncated at a distance of 12 Å and switched between 10 and 12 Å. Before configurational sampling, an equilibration simulation was made for several hundred ps at 300 K with the *NVT* ensemble using the TIP3P water model.<sup>38</sup> For simulations using the flexible KKY water model (see below), an additional 200 ps of *NVT* equilibration followed for every system.



**Fig. 1** The simulation system used in the present work. The C and N atoms of the cyanide ion are represented as cyan and blue van der Waals spheres, respectively. Water molecules are depicted as lines.

### 2.1 Intermolecular interactions

The stretching potential for the cyanide or hydroxide ion is a Morse potential

$$V(r) = D_e \{1 - \exp[-\beta(r - r_e)]\}^2. \quad (1)$$

For the cyanide ion,  $D_e = 237.5 \text{ kcal mol}^{-1}$ ,  $\beta = 2.283 \text{ \AA}^{-1}$ , and  $r_e = 1.172 \text{ \AA}$  were used.<sup>39</sup> For the hydroxide ion, DFT calculations at the B3LYP/aug-cc-pVQZ level were performed and yielded  $D_e = 96.97 \text{ kcal mol}^{-1}$ ,  $\beta = 2.437 \text{ \AA}^{-1}$ , and  $r_e = 0.963 \text{ \AA}$ .

For the electrostatic interactions, a multipolar (MTP) or a conventional two-point charge (PC) model was used for cyanide and hydroxide. The multipole moments for  $\text{CN}^-$  and  $\text{OH}^-$  are calculated from density functional theory (DFT) calculations at the B3LYP/aug-cc-pVQZ level. The Gaussian03 suite of programs<sup>40</sup> was used to obtain the SCF wavefunctions, from which the charges  $q$ , dipole  $\vec{\mu}$  and quadrupole  $\Theta$  moments on the C and N atoms were calculated through the GDMA program.<sup>41</sup> Each multipole moment  $Q$  was calculated on a grid ranging from 0.9 to 1.1  $r_e$  with a step size of 0.01  $r_e$ , with  $r_e = 1.16 \text{ \AA}$  for  $\text{CN}^-$  and 0.958 Å for  $\text{OH}^-$ , which are close to the experimental gas-phase bond lengths of 1.1765 Å for  $\text{CN}^-$  (ref. 42) and 0.954 Å for  $\text{OH}^-$ .<sup>43</sup> The values of all nonzero components up to the quadrupole are summarized in Table 1.<sup>30</sup> To describe the orientations of the atomic multipole moments with respect to the global Cartesian coordinate system in which the simulations are carried out, a reference axis system needs to be assigned to the molecule and the interactions are transformed in each step to the global axis system.<sup>44</sup> For the cyanide or hydroxide anion, the charge distribution is axially symmetric and only one axis is required. Here, the z-axis is chosen, which is parallel to the

**Table 1** Multipole parameters and CHELPG charges of  $\text{CN}^-$  and  $\text{OH}^-$ . Only nonzero MTP components are shown. MTP components are  $Q_{\ell m}$ , where  $\ell$  represents angular momentum labels (00, 10, 11c, 11s, 20, 21c, 21s, 22c, and 22s)

| Site | $Q_{00} [e]$ | $Q_{10} [ea_0]$ | $Q_{20} [ea_0^2]$ | $q_{\text{CHELPG}} [e]$ |
|------|--------------|-----------------|-------------------|-------------------------|
| C    | -0.447       | 0.973           | 0.318             | -0.500                  |
| N    | -0.553       | -0.698          | 0.963             | -0.500                  |
| O    | -1.008       | -0.516          | 2.389             | -1.183                  |
| H    | 0.008        | 0.196           | 0.437             | 0.183                   |

molecular axis. The molecular dipole and quadrupole moments from the DFT calculations are  $0.249 ea_0$  and  $-3.668 ea_0^2$  for  $CN^-$  and  $-0.539 ea_0$  and  $1.963 ea_0^2$  for  $OH^-$ .

The two-point charge model employed charges from electrostatic potentials using a grid based method (CHELPG).<sup>45</sup> The B3LYP/aug-cc-pVQZ calculations were carried out on the equilibrium structure of the ions. Table 1 shows that CHELPG charges are quite similar to the charges from the distributed multipole analysis ( $Q_{00}$ ). Because the decomposition of an electron density is not necessarily unique, higher MTPs can be interpreted as corrections to the ESP charges.<sup>31</sup> According to this, the charges are first determined to best reproduce the ESP and then higher MTPs are used to improve on this. Therefore, one can envisage using a computationally inexpensive model to sample configurational space and then obtain improved observables from “scoring” the conformations generated using a more elaborate model.

The Lennard-Jones (LJ) parameters for the cyanide and hydroxide ions were adapted from the universal force field (UFF):<sup>46</sup>  $\epsilon^C = 0.105 \text{ kcal mol}^{-1}$ ,  $\sigma^C = 1.9255 \text{ \AA}$ ,  $\epsilon^N = 0.069 \text{ kcal mol}^{-1}$ ,  $\sigma^N = 1.8300 \text{ \AA}$ ,  $\epsilon^O = 0.060 \text{ kcal mol}^{-1}$ ,  $\sigma^O = 1.7500 \text{ \AA}$ ,  $\epsilon^H = 0.044 \text{ kcal mol}^{-1}$ , and  $\sigma^H = 1.4430 \text{ \AA}$ , where  $\sigma = r_{\min}/2$ . For  $\epsilon$  the original UFF values were used without change, whereas for  $\sigma$  both original and scaled values were applied. In the previous work it was found that scaling  $\sigma$  by +7.5% almost quantitatively reproduces experimental vibrational relaxation times of  $CN^-$  in  $H_2O$  and  $D_2O$ .<sup>30</sup> The ranges of  $\sigma$  values for the C, N, O, and H atoms in the CHARMM force field are as follows:  $1.56 \leq \sigma^C \leq 2.30 \text{ \AA}$ ,  $1.79 \leq \sigma^N \leq 2.06 \text{ \AA}$ ,  $1.65 \leq \sigma^O \leq 1.77 \text{ \AA}$ , and  $0.22 \leq \sigma^H \leq 1.48 \text{ \AA}$ . Hence, the  $\sigma$  values used here are within the range of those from the CHARMM force field. We note that the value of  $\sigma^H$  varies over a wider range depending on its chemical environment. It has been shown very recently that the parameters for the van der Waals interactions need to be adapted when MTP is used.<sup>47</sup> While we did not explicitly optimize  $\sigma$  values in the present work, we tested various values of  $\sigma$  in order to see their effect on the solvation free energy.

For water, two different models are employed. One is a standard TIP3P potential<sup>38</sup> with SHAKE.<sup>48,49</sup> Furthermore, a flexible water model based on the parametrization by Kumagai, Kawamura, and Yokokawa (KKY)<sup>50</sup> is used with different electrostatic models and a smaller time step of 0.4 fs to account for the flexible  $O \cdots H$  bonds. The functional form of the KKY potential for the stretching and bending energies is

$$E_{\text{str}} = D\{1 - \exp[-\beta(r - r_0)]\}^2 \quad (2)$$

and

$$E_{\text{bend}} = 2f_k \sqrt{k_1 k_2} \sin^2(\theta - \theta_0) \quad (3)$$

where  $k_i = 1/\{\exp[g_r(r_i - r_m)] + 1\}$ , in which  $r_i$  is the distance of one O–H bond of the water molecule and  $g_r$  and  $r_m$  are parameters. All KKY parameters used in the present work are summarized in Table 2.<sup>30</sup> With KKY, electrostatic interactions are either TIP3P charges or multipole moments previously determined,<sup>51</sup> which are independent of the geometry of the

**Table 2** Parameters for water potential reported in the work of Kumagai *et al.*<sup>50</sup> and used in this work. Atomic units are used for  $D$ ,  $\beta$ ,  $r_0$ ,  $f_k$ ,  $r_m$ , and  $g_r$ .  $\theta_0$  is given in degrees

|           | $D [E_h]$ | $\beta [a_0^{-1}]$ | $r_0 [a_0]$ | $f_k [E_h]$       | $\theta_0 [\text{deg}]$ | $r_m [a_0]$ | $g_r [a_0^{-1}]$ |
|-----------|-----------|--------------------|-------------|-------------------|-------------------------|-------------|------------------|
| KKY       | 0.120     | 1.45               | 1.55        | $2.5 \times 10^6$ | 99.5                    | 2.65        | 3.7              |
| This work | 0.120     | 1.44               | 1.81        | 0.0388            | 99.5                    | 2.65        | 3.7              |

water molecule. For the Lennard-Jones parameters  $\sigma$  and  $\epsilon$ , those of the TIP3P model were used in all simulations. Because two different force fields are used for both  $CN^-$  and  $H_2O$ , they can be combined in various ways, which are summarized in Table 3.

## 2.2 Free energy simulations

**Thermodynamic integration.** In thermodynamic integration (TI), the free energy difference between two given states, A and B, is determined from an interpolating Hamiltonian.<sup>52,53</sup> For a reaction,  $A \rightarrow B$ , a coupling parameter  $\lambda$  is introduced with  $\lambda = 0$  for A (reactant) and  $\lambda = 1$  for B (product). Then the potential energy of the system can be written as

$$U(\mathbf{r}_1, \dots, \mathbf{r}_N, \lambda) = (1 - \lambda)U_A(\mathbf{r}_1, \dots, \mathbf{r}_N) + \lambda U_B(\mathbf{r}_1, \dots, \mathbf{r}_N). \quad (4)$$

The free energy change for the reaction  $A \rightarrow B$  can be computed from

$$\Delta F(A \rightarrow B) = \int_0^1 d\lambda \frac{\partial F(\lambda)}{\partial \lambda} = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda, \quad (5)$$

where  $U$  is the potential energy and  $\langle \dots \rangle_\lambda$  denotes an ensemble average over the distribution  $\exp[-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N, \lambda)]$  with  $\lambda$  fixed at a particular value. From eqn (4), we obtain

$$\frac{\partial U(\lambda)}{\partial \lambda} = U_B - U_A. \quad (6)$$

Decomposing  $U$  into water–water (ww) and ion–water (iw) interactions yields  $U_A = U_{\text{ww}}$  and  $U_B = U_{\text{ww}} + U_{\text{iw}}$ , and thus

$$\frac{\partial U(\lambda)}{\partial \lambda} = U_{\text{iw}}. \quad (7)$$

The ion–water interaction energies  $U_{\text{iw}}$  are obtained from the van der Waals (vdW) and electrostatic (es) interactions,

$$U_{\text{iw}} = U_{\text{iw}}^{\text{vdW}} + U_{\text{iw}}^{\text{es}}. \quad (8)$$

For TI it is necessary to calculate  $\langle U_{\text{iw}} \rangle_\lambda$  for different  $\lambda$  values between 0 and 1. In the present work, we first carried out MD

**Table 3** Simulation models M1 to M5, used in this work.  $V_{\text{bond}}$  refers to bonded interactions and  $V_{\text{es}}$  to electrostatic interactions.  $q$ ,  $\mu$ , and  $\theta$  refer to charges, dipole and quadrupole moments

| Model | SHAKE | $V_{\text{bond}}^{\text{water}}$ | $V_{\text{es}}^{\text{water}}$ | $V_{\text{bond}}^{\text{ion}}$ | $V_{\text{es}}^{\text{ion}}$ |
|-------|-------|----------------------------------|--------------------------------|--------------------------------|------------------------------|
| M1    | Yes   | TIP3P                            | $q_{\text{TIP3P}}$             | Harmonic                       | $q_{\text{CHELPG}}$          |
| M2    | Yes   | TIP3P                            | $q_{\text{TIP3P}}$             | Harmonic                       | $q, \mu, \theta$             |
| M3    | Yes   | TIP3P                            | $q_{\text{TIP3P}}$             | Morse                          | $q, \mu, \theta$             |
| M4    | No    | KKY                              | $q_{\text{TIP3P}}$             | Morse                          | $q, \mu, \theta$             |
| M5    | No    | KKY                              | $q, \mu, \theta$               | Morse                          | $q, \mu, \theta$             |

simulations for 15  $\lambda$ -values ( $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$ , 0.05, 0.1, 0.2, ..., 1.0) from which  $\langle U_{iw} \rangle_\lambda$  was determined and the free energy difference was then determined *via* numerical integration.

**Free energy perturbation.** As the free energy of system A can be obtained from the canonical partition function  $Q_A(N, V, T)$  *via*

$$F_A = -k_B T \ln Q_A(N, V, T) \quad (9)$$

and similarly for B, the free energy difference for the change from A to B is

$$\Delta F(A \rightarrow B) = -k_B T \ln \left( \frac{Q_B}{Q_A} \right) = -k_B T \ln \left( \frac{Z_B}{Z_A} \right), \quad (10)$$

where  $Z_A$  and  $Z_B$  are the configurational integrals for states A and B, respectively. This approach is termed free energy perturbation (FEP). As the ratio of  $Z_B$  and  $Z_A$  is

$$\frac{Z_B}{Z_A} = \frac{1}{Z_A} \int d^N \mathbf{r} e^{-\beta U_A} e^{-\beta(U_B - U_A)} = \left\langle e^{-\beta(U_B - U_A)} \right\rangle_A, \quad (11)$$

the free energy difference between A and B can be computed using ref. 54

$$\Delta F(A \rightarrow B) = -k_B T \ln \langle e^{-\beta(U_B - U_A)} \rangle_A. \quad (12)$$

Eqn (12) can only be applied directly if states A and B are very similar. For the present case, states A and B correspond to a system without and with solvent-ion interactions, respectively. Hence, states A and B are very different. To reduce the error, the path from A to B is divided into multiple steps, and the free energy difference for the change from A to B is obtained by summing the free energy differences between all consecutive steps. The same  $\lambda$  values as in TI are used for the FEP simulations in the present work, which allows us to reuse the trajectories.

**Bennett acceptance ratio.** The Bennett acceptance ratio (BAR) method is an algorithm to calculate the free energy difference between two systems (A and B).<sup>55</sup> For any function  $f(x)$  satisfying  $f(x)/f(-x) = e^{-x}$  and for any value of  $C$ , the following relationship holds:

$$e^{-\beta(\Delta F - C)} = \frac{\langle f(\beta(U_B - U_A - C)) \rangle_A}{\langle f(\beta(U_A - U_B + C)) \rangle_B}, \quad (13)$$

where  $\Delta F$  is the free energy difference and  $U_A$  is the potential energy of system A. The optimal choice of  $f(x)$  and  $C$  has been shown to be  $f(x) = 1/(1 + e^x)$  and  $C \approx \Delta F$ .<sup>55</sup> As  $\Delta F$  is the value to be evaluated,  $C$  is found self-consistently by iteration starting with an initial guess. In the present work, the value obtained from FEP was given as an initial value of  $C$ . Similar to FEP, the path from A to B is divided into multiple steps, and eqn (13) is applied to each consecutive step. The free energy change from A to B is obtained by summing free energy differences between all consecutive steps.

**Umbrella sampling.** In principle, the free energy difference between two states A and B can be obtained by

$$F(\delta_B) - F(\delta_A) = -k_B T \ln [P(\delta_B)/P(\delta_A)], \quad (14)$$

where  $P(\delta)$  is the probability of finding the system at reaction coordinate  $\delta$ .  $\delta_A$  and  $\delta_B$  correspond to the states A and B, respectively, in the trajectory. However, only low energy regions would be explored sufficiently during an MD simulation unless any artificial potential is applied. For a trajectory obtained with a suitable umbrella potential  $U^{(i)}(\delta) = K^{(i)}(\delta - \delta_0^{(i)})^2$  added to the original potential energy surface, the free energy can be obtained from

$$F^{(i)}(\delta) = -k_B T \ln [P^{(i)}(\delta)] - U^{(i)}(\delta) + C^{(i)} \quad (15)$$

where  $C^{(i)}$  is a constant for the  $i$ -th window. This constant is in general different for each window and can be determined using the Weighted Histogram Analysis Method (WHAM).<sup>56,57</sup> In the present work,  $\delta$  is defined as the distance of the center of the ion from the center of the water sphere and 23 windows were used with the  $\delta_0$  values ranging from 10 to 32 Å, each displaced by 1 Å. Force constants  $K$  were in the range between 3 and 4 kcal mol<sup>-1</sup> Å<sup>-2</sup>.

**Soft-core Lennard-Jones potential.** While we did not use soft-core potentials<sup>58</sup> in most of the present work, one TI simulation was carried out for comparison. Only the solute-water van der Waals (vdW) interactions were treated with a soft-core potential of the following form

$$U_{iw}(\lambda) = \frac{A}{[r^2 + \delta_v(1 - \lambda)]^6} - \frac{B}{[r^2 + \delta_v(1 - \lambda)]^3}, \quad (16)$$

where  $\delta_v$  is the separation parameter for vdW interactions, which allows a smooth transition from the original LJ potential. In the present test,  $\delta_v = 5$  Å<sup>2</sup> was used. It is noted that  $U_{iw}(\lambda)$  corresponds to the original LJ potential when  $\lambda = 1$ , and that the potential energy of the system,  $U(\lambda) = U_{ww} + \lambda U_{iw}(\lambda)$ , is also the same as that without soft-core potential when  $\lambda = 0$ . In other words, while the integration path is changed by using a soft-core potential, the two end points are not affected. Because the free energy is a state function, its value should be path-independent, which is verified here. For TI an extra term needs to be added to the first derivative of  $U(\lambda)$  with respect to  $\lambda$ ,

$$\frac{\partial U(\lambda)}{\partial \lambda} = U_{iw}(\lambda) + \lambda \frac{\partial U_{iw}(\lambda)}{\partial \lambda}, \quad (17)$$

as  $U_{iw}(\lambda)$  depends explicitly on  $\lambda$  for soft-core potentials.

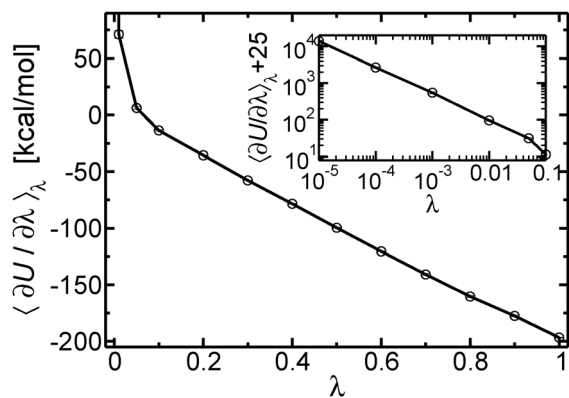
## 3 Results and discussion

After establishing that different methods give very similar results, the sensitivity of the solvation free energy to different descriptions of the intermolecular interactions (electrostatics and van der Waals) is assessed. In a separate subsection, the results are quantitatively compared with experimental data.

### 3.1 Solvation free energies from different methods

Solvation free energies of CN<sup>-</sup> and OH<sup>-</sup> in H<sub>2</sub>O were first determined using TI, FEP, and BAR. This helps to verify whether a particular choice of method affects the computed values appreciably. First, TI was tested for 3 × 50-ps trajectories of CN<sup>-</sup> in H<sub>2</sub>O, generated with model 4 (M4) using  $r_{\min} = r_{\text{UFF}}$  for CN<sup>-</sup>.





**Fig. 2** The plot of  $\langle \partial U / \partial \lambda \rangle_\lambda$  as a function of  $\lambda$  for  $3 \times 50$ -ps simulations of  $\text{CN}^-$  in  $\text{H}_2\text{O}$  using M4 for thermodynamic integration. The inset shows a small lambda range ( $10^{-5}$ –0.1) on a logarithmic scale.

**Table 4** Solvation free energies of  $\text{CN}^-$  in  $\text{H}_2\text{O}$  in  $\text{kcal mol}^{-1}$  computed with TI, FEP, and BAR at each interval of  $\lambda$  ( $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$ , 0.05, 0.1, 0.2, ..., 1.0) from the same trajectories obtained with M4 using  $r_{\text{min}} = r_{\text{UFF}}$  for  $\text{CN}^-$ . The total free energy change for  $A \rightarrow B$  was obtained by summing the free energy changes at all intervals. While column TI was obtained by fitting functions to the data set, TI (seg) was obtained by the trapezoidal rule at each interval. FEP ( $\lambda_-$ ) was obtained from the trajectory of the smaller  $\lambda$  at a given interval, while FEP ( $\lambda_+$ ) from that of the larger  $\lambda$ . FEP is the average of FEP ( $\lambda_-$ ) and FEP ( $\lambda_+$ ). For comparison, the total free energy change from US is also shown, which was obtained from separate simulations with umbrella potential

| $\lambda$         | TI    | TI (seg) | FEP   | FEP ( $\lambda_-$ ) | FEP ( $\lambda_+$ ) | BAR   | US              |
|-------------------|-------|----------|-------|---------------------|---------------------|-------|-----------------|
| $0-10^{-5}$       | 0.15  | 0.15     | 0.29  | 0.29                | 0.29                | 0.29  |                 |
| $10^{-5}-10^{-4}$ | 0.45  | 0.74     | 0.44  | 0.47                | 0.42                | 0.46  |                 |
| $10^{-4}-10^{-3}$ | 0.85  | 1.43     | 0.84  | 0.87                | 0.81                | 0.88  |                 |
| $10^{-3}-10^{-2}$ | 1.46  | 2.68     | 1.33  | 1.53                | 1.13                | 1.48  |                 |
| 0.01–0.05         | 1.03  | 1.55     | 1.04  | 1.00                | 1.08                | 1.03  |                 |
| 0.05–0.1          | -0.28 | -0.19    | -0.25 | -0.19               | -0.30               | -0.24 |                 |
| 0.1–0.2           | -2.66 | -2.46    | -2.49 | -2.56               | -2.42               | -2.50 |                 |
| 0.2–0.3           | -4.70 | -4.66    | -4.66 | -4.62               | -4.71               | -4.67 |                 |
| 0.3–0.4           | -6.73 | -6.80    | -6.83 | -6.88               | -6.78               | -6.81 |                 |
| 0.4–0.5           | -8.77 | -8.90    | -8.90 | -8.91               | -8.89               | -8.89 |                 |
| 0.5–0.6           | -10.8 | -11.0    | -11.0 | -11.0               | -11.0               | -11.0 |                 |
| 0.6–0.7           | -12.8 | -13.1    | -13.1 | -13.0               | -13.1               | -13.1 |                 |
| 0.7–0.8           | -14.9 | -15.0    | -15.0 | -15.0               | -15.0               | -15.0 |                 |
| 0.8–0.9           | -16.9 | -16.9    | -16.9 | -17.0               | -16.8               | -16.9 |                 |
| 0.9–1.0           | -18.9 | -18.7    | -18.7 | -18.7               | -18.7               | -18.7 |                 |
| Total             | -93.6 | -91.2    | -93.9 | -93.8               | -94.1               | -93.7 | $\approx -90.3$ |

The function  $\langle \partial U / \partial \lambda \rangle_\lambda$  is shown in Fig. 2. For  $0.1 \leq \lambda \leq 1.0$ ,  $\langle \partial U / \partial \lambda \rangle_\lambda$  is close to a linear function of  $\lambda$ , while this is not the case for  $\lambda < 0.1$ . In the column “TI (seg)” in Table 4, the values of  $\int_{\lambda_1}^{\lambda_2} d\lambda \langle \partial U / \partial \lambda \rangle_\lambda$  for each  $\lambda$  interval and the sum are given. The integral for each interval was obtained by drawing a line connecting both ends of the interval. While this procedure looks reasonable in the range where  $\lambda$  is larger than 0.1, the error may be large when the  $\lambda$  is smaller than 0.1.

To estimate the integral more accurately, we empirically divided the entire range of  $\lambda$  into three parts. The first range is  $\lambda = 0.1$  to 1.0, and a straight line was fitted to compute the integral in this range. In the range of  $\lambda = 10^{-5}$  to 0.05, a straight line was fitted to  $\ln(\langle \partial U / \partial \lambda \rangle_\lambda + \delta U)$  vs.  $\ln \lambda$ . Two procedures were used. In one of them, the parameter  $\delta U$  was fitted and in the other one  $\delta U = 25 \text{ kcal mol}^{-1}$  was used. In both cases the

correlation coefficient of the fit was very close to  $-1$ . An example for the fit is shown in the inset of Fig. 2. In the range 0.01–0.1, a straight line was fitted to  $\ln(\langle \partial U / \partial \lambda \rangle_\lambda + \delta U)$  vs.  $\lambda^{1/8}$ . Again,  $\delta U$  was allowed to change in a fit or a constant value  $\delta U = 50 \text{ kcal mol}^{-1}$  was used.

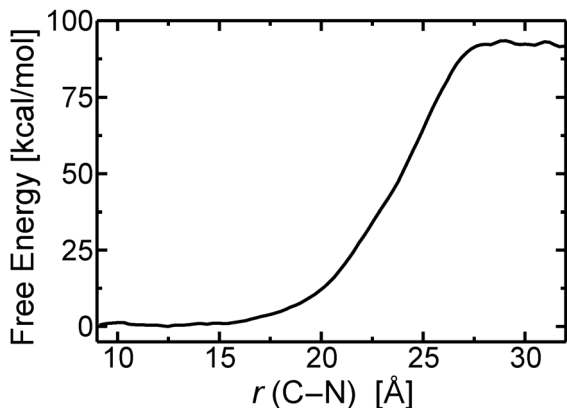
Alternatively, it is possible to use a soft-core potential for the van der Waals interactions<sup>58</sup> in order to prevent the diverging behaviour of  $\langle \partial U / \partial \lambda \rangle_\lambda$  at very small  $\lambda$ . We computed the solvation free energy of  $\text{CN}^-$  using TI for M1 with  $r_{\text{min}} = 1.075 r_{\text{UFF}}$  for  $\text{CN}^-$ , applying a soft-core LJ potential in the range of  $\lambda$  between 0 and 0.1, for comparison. The solvation free energy obtained using the soft-core potential is  $-83.2 \text{ kcal mol}^{-1}$ , which compares favourably with the value obtained by fitting  $\langle \partial U / \partial \lambda \rangle_\lambda$  and integrating, which is  $-83.6 \text{ kcal mol}^{-1}$ . This validates the procedures used throughout the remainder of the present work.

The result of the thermodynamic integration performed with the fitted functions is shown in the column “TI” in Table 4. As can be seen, the TI values show appreciable differences in the range of  $\lambda$  that is smaller than 0.1, and therefore the total free energy change differs by a few  $\text{kcal mol}^{-1}$ . The fitting of the empirical equations seems to give a much better agreement with FEP and BAR. Therefore, we used the empirical fitting method described above for the evaluation of the integrals in all TI calculations described in the following.

For TI it is straightforward to decompose the solvation free energy into contributions from electrostatic and van der Waals interactions, although the decomposed values can show some dependence on the path. The decomposition for M4 with  $r_{\text{min}} = r_{\text{UFF}}$  yields  $-105.6 \text{ kcal mol}^{-1}$  and  $12.3 \text{ kcal mol}^{-1}$  for the electrostatic and van der Waals interactions, respectively. The sum of these two terms ( $-93.3 \text{ kcal mol}^{-1}$ ) slightly differs from the value from TI ( $-93.6 \text{ kcal mol}^{-1}$ ) because the electrostatic and van der Waals contributions to the total solvation energy were fitted separately. Nevertheless, the two results agree with each other to within the error bars. As can be seen clearly, electrostatic interactions are mainly responsible for the solvation free energies, while van der Waals interactions contribute  $\approx 10\%$  of unfavorable desolvation.

For the calculation with FEP, the same trajectories as generated for TI were used. With FEP it is possible to use the trajectories generated at  $\lambda_-$  or  $\lambda_+$  to determine the free energy differences. For every interval except the 1st, we used both trajectories and took the average value of the two free energy changes computed with trajectories at  $\lambda_-$  and  $\lambda_+$ , so as to minimize the error. In the present simulations, the difference in free energy change for the same interval, computed using two different trajectories at  $\lambda_-$  and  $\lambda_+$ , amounted to 0.1–0.2  $\text{kcal mol}^{-1}$ . The columns “FEP ( $\lambda_-$ )” and “FEP ( $\lambda_+$ )” in Table 4 summarize these results, and the column “FEP” gives their average. The total free energy change from FEP is very close to that obtained from TI, as can be seen in Table 4.

Analysis of the solvation free energy using BAR also used the trajectories generated for TI. In computing the free energy change using BAR, the initial  $C$  value of eqn (13) was taken as that obtained from FEP. Then,  $C$  was updated self-consistently at each interval, until the difference between  $C$  and the free



**Fig. 3** The free energy profile of  $\text{CN}^-$  in  $\text{H}_2\text{O}$  from umbrella sampling computed by WHAM. For this simulation 23 windows were used, with a simulation time of 40 ps for each window.

energy change was  $< 10^{-4}$  kcal mol $^{-1}$ . Typically, 1–3 iterations were necessary to obtain a converged result. The free energy change computed using BAR is also very close to the values obtained from TI or FEP, as shown in Table 4.

Finally, the total free energy change was also computed using umbrella sampling (US) together with WHAM. As this method requires trajectories generated at different conditions from those for TI, FEP, and BAR, separate trajectories needed to be run. They included 23 individual simulations with an umbrella potential applied to the center of mass of  $\text{CN}^-$ . A harmonic potential centered around different displacements ( $\delta_0$ ) of  $\text{CN}^-$  away from the center of the water sphere was used for the umbrella potential in each window. While the radius of the water sphere is about 19 Å,  $\delta_0$  values considerably larger than this radius were necessary for US as  $\text{CN}^-$  pulls water molecules with it when  $\delta_0$  is slightly larger than this radius. The free energy as a function of  $\delta_0$  is shown in Fig. 3. As can be seen, the free energy is quite constant if  $\text{CN}^-$  is completely inside or outside the water sphere. The free energy change computed using US is about  $-90.3$  kcal mol $^{-1}$ , which is close to the values computed using the other methods.

Table 5 reports hydration free energies computed using TI, FEP, and BAR from trajectories generated using various interaction

**Table 5** Solvation free energies of  $\text{CN}^-$  in  $\text{H}_2\text{O}$  in kcal mol $^{-1}$  computed with TI, FEP, and BAR from trajectories generated with various models. In parentheses are shown the estimated errors, which are standard deviations of partial averages computed over the data divided into 10 subgroups. Previous work yielded solvation free energy values ranging from  $-72$  to  $-77$  kcal mol $^{-1}$  from experiments. The ratios  $r_{\text{min}}/r_{\text{UFF}}$  used for  $\text{CN}^-$  are 1.0, 1.075, 1.15, and 1.20, and are denoted by  $-$ ,  $+$ ,  $++$ , and  $+++$ , respectively

| Model | TI          | FEP         | BAR         |
|-------|-------------|-------------|-------------|
| M1–   | –91.1(0.09) | –91.6(0.16) | –91.3(0.07) |
| M1+   | –83.6(0.04) | –83.6(0.35) | –83.7(0.29) |
| M1++  | –77.2(0.04) | –77.5(0.21) | –77.5(0.06) |
| M1+++ | –72.8(0.04) | –73.0(0.24) | –73.0(0.06) |
| M4–   | –93.6(0.14) | –93.9(0.31) | –93.7(0.18) |
| M4+   | –84.6(0.04) | –84.9(0.16) | –84.7(0.05) |
| M4++  | –77.2(0.07) | –77.5(0.18) | –77.3(0.09) |
| M4+++ | –72.4(0.06) | –72.9(0.12) | –72.6(0.06) |

**Table 6** Solvation free energies of  $\text{CN}^-$  and  $\text{OH}^-$  in  $\text{H}_2\text{O}$  computed with TI. Experimental values from the literature<sup>14,60,61</sup> are  $-72$  to  $-77$  kcal mol $^{-1}$  for  $\text{CN}^-$  and  $-105$  to  $-110$  kcal mol $^{-1}$  for  $\text{OH}^-$ . Models with  $\alpha = r_{\text{min}}/r_{\text{UFF}} = 1.0$  for the ions are labelled by ‘–’,  $\alpha = 1.075$  by ‘+’,  $\alpha = 1.15$  by ‘++’, and  $\alpha = 1.20$  by ‘+++’

| Ion           | Model | –      | +      | ++     | +++    |
|---------------|-------|--------|--------|--------|--------|
| $\text{CN}^-$ | M1    | –91.1  | –83.6  | –77.2  | –72.8  |
|               | M2    | –93.6  | –84.4  | –77.0  | –72.2  |
|               | M4    | –93.6  | –84.6  | –77.2  | –72.4  |
|               | M5    | –86.2  | –78.7  | –72.0  | –68.2  |
| $\text{OH}^-$ | M1    | –131.0 | –118.6 | –107.8 | –101.8 |
|               | M2    | –171.5 | –149.7 | –131.3 | –120.5 |
|               | M4    | –175.6 | –150.8 | –131.5 | –121.1 |
|               | M5    | –147.6 | –131.2 | –114.7 | –107.3 |

models. The error associated with the computed free energy for each case has been estimated by splitting the available data into 10 subgroups, computing the free energy from each subgroup, and then calculating the standard deviation of the free energies from all the subgroups. The error thus estimated is shown in parentheses in Table 5. As can be seen, the estimated errors are around 0.1 kcal mol $^{-1}$  in many cases, and even the largest estimated error does not exceed 0.4 kcal mol $^{-1}$ . We note that the computed hydration free energies depend little on the method for all cases studied, which also supports the details in the procedures employed. Therefore, any of the methods can be used for calculating hydration free energies in the present case. We use only TI in the remainder of this work.

### 3.2 Force field dependence of solvation free energies

The hydration free energies of  $\text{CN}^-$  and  $\text{OH}^-$  computed for various models by TI in the present work are summarized in Table 6. As is clear, standard van der Waals radii ( $r_{\text{min}} = r_{\text{UFF}}$ ) give larger free energies of solvation by about 15 kcal mol $^{-1}$  for  $\text{CN}^-$  and even larger for  $\text{OH}^-$  compared with the experimental values for all models investigated (for corrections due to cut-offs, see below). Increasing  $r_{\text{min}}$  decreases the magnitude of the solvation free energies due to “shielding” of the electrostatic interactions. Without discussing the numbers in too much detail compared to experiment, this finding is consistent with previous observations that in order to quantitatively reproduce spectroscopic properties or vibrational relaxation for  $\text{CN}^-$  the  $r_{\text{min}}$  values for the LJ potential need to be somewhat increased.<sup>30,59</sup>

For both anions, simulations with CHELPG point charges (M1) show the same dependence on  $r_{\text{min}}$  as also found for multipolar models. For the cyanide anion, M4 (MTPs on the C and N atoms) gives results similar to M1 (CHELPG on the C and N atoms), while this is not true for hydroxide. This may be because the charges in the MTP model are similar to CHELPG for  $\text{CN}^-$  but not for  $\text{OH}^-$ . We also find that the dependence of the solvation free energy on the value of  $r_{\text{min}}$  is more pronounced for  $\text{OH}^-$  compared to  $\text{CN}^-$ . Increasing  $r_{\text{min}}$  by 20% for  $\text{CN}^-$  reduces the solvation free energy by 23% (for M4) whereas the change is 32% for  $\text{OH}^-$ . This effect is less pronounced for a point charge model (M1) where the changes are 20% and 23%, respectively.

The contribution of each of the multipole orders was determined by keeping 0th, 1st, and 2nd order terms, *i.e.*, (1) only  $q$ , (2)  $q$  and  $\mu$ , and (3)  $q$ ,  $\mu$ , and  $\theta$  on  $\text{H}_2\text{O}$ , in the analysis of the M5 trajectories generated with  $q$ ,  $\mu$ , and  $\theta$  on  $\text{H}_2\text{O}$ . The electrostatic model of the anion was unmodified (M5) in all cases. This gives free energies of (1)  $-59.1$ , (2)  $-62.7$ , and (3)  $-68.2$  kcal mol $^{-1}$  for  $\text{CN}^-$  and (1)  $-91.0$ , (2)  $-96.7$ , and (3)  $-107.3$  kcal mol $^{-1}$  for  $\text{OH}^-$ , where  $r_{\text{min}}/r_{\text{UFF}} = 1.20$  was used. Thus the interactions between the anion and the point charges on water molecules are responsible for more than 80% of the solvation free energies. We also note that dipoles and quadrupoles on water molecules give an appreciable contribution to the solvation free energies.

### 3.3 Comparison with experiment

For simulations with periodic boundary conditions (PBCs), it was recently shown that the absolute values of the solvation free energy can be sensitive to surface effects.<sup>62</sup> They arise because the relevant reference state – the ion in vacuum – is never explicitly sampled in PBC simulations for TI or FEP. In order to account for this, the standard internal Galvani potential (or the interfacial potential) for the particular water model needs to be evaluated and taken into account when comparing with experimental solvation free energies.<sup>62,63</sup> This contribution, amounting to  $\approx 10$  kcal mol $^{-1}$  for monovalent sodium in methane, was shown to be absent if simulations are carried out for the ion in a water droplet and the interfacial potential model was developed to correct solvation studies in PBC.<sup>62</sup>

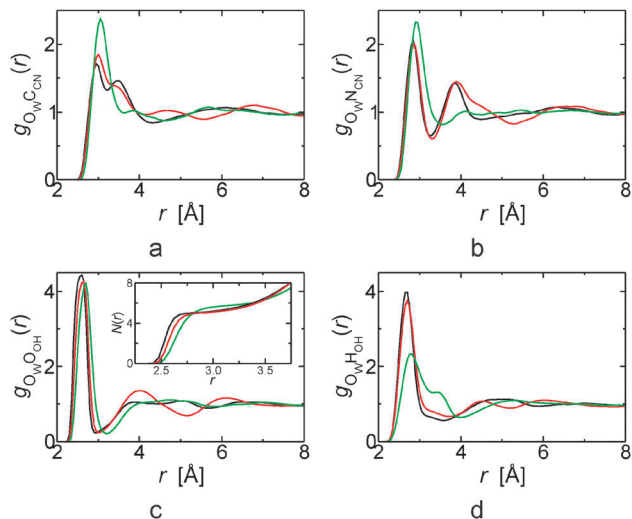
Nevertheless, if direct comparison from simulations with spherical boundaries with experiment is sought, it is necessary to assess the convergence of the simulations with respect to the size of the water sphere and the cutoff used. This was done in the present case for the two models M1 and M4. Additional simulations were run with two considerably larger water spheres, containing 2000 and 4000 water molecules, respectively. For these and the sphere with 997 water molecules – used throughout the rest of the present work – the solvation free energies with TI, FEP and BAR were computed by using an “infinite cutoff” (*i.e.* including all nonbonded interactions between the ion and the surrounding water molecules) in both, sampling the conformations and the calculations to determine solvation free energies. Again, the three methods yield very similar results and only those from TI are further analyzed: for the water sphere with 997 water molecules and “infinite” cutoff the solvation free energies for M1+ are  $-75.5$ ,  $-75.9$ , and  $-75.8$  kcal mol $^{-1}$  for TI, FEP and BAR, respectively. If the size of the water sphere is increased to 2000 and 4000 water molecules, and simulations and evaluation of free energies are again done with infinite cutoff, the solvation free energies change to  $-74.4$  to  $-76.1$  (TI),  $-74.3$  to  $-76.7$  (FEP) and  $-74.6$  to  $-76.4$  (BAR) kcal mol $^{-1}$ . With point charge and multipolar models for the ions, the differences between simulations with finite and “infinite” cutoff are found to be approximately constant for a given size of the water sphere, independent of the method used. They amount to  $\approx 8$  kcal mol $^{-1}$  for the water sphere with 997 water molecules if point charges are used and somewhat

less ( $\approx 7.5$  kcal mol $^{-1}$ ) for simulations with M4 for  $\text{CN}^-$ . For  $\text{OH}^-$ , the corrections were also determined explicitly for M1–, M1+, M4–, and M4+. They were found to be slightly larger – between 9.6 and 10.0 kcal mol $^{-1}$ . Hence, for a direct comparison with experiment all data in Tables 5 and 6 can be corrected by a constant destabilizing contribution which is  $\approx 8.0$  kcal mol $^{-1}$  for  $\text{CN}^-$  and  $\approx 10.0$  kcal mol $^{-1}$  for  $\text{OH}^-$ . We note in passing that the magnitude and sign of the corrections are similar to those found in the work by Roux and coworkers.<sup>62</sup> Using a shorter cutoff, such as 12 Å, is therefore suitable to study *relative changes* in solvation free energies – *e.g.* for different interaction models – but is clearly insufficient for converged results and a much larger cutoff is required if the solvation free energies are to be compared with experimental data.

The solvation free energies of  $\text{CN}^-$  and  $\text{OH}^-$  obtained by Florián and Warshel<sup>14</sup> based on experimental data are  $-75 \pm 5$  and  $-110 \pm 5$  kcal mol $^{-1}$ , respectively, while those by Pearson<sup>60</sup> are  $-77$  and  $-106$  kcal mol $^{-1}$ . Pliego and Riveros report slightly smaller values, *i.e.*,  $-72.0 \pm 0.7$  and  $-105.0 \pm 0.5$  kcal mol $^{-1}$  for  $\text{CN}^-$  and  $\text{OH}^-$ , respectively.<sup>61</sup> For  $\text{CN}^-$ , accounting for the correction due to finite cutoffs, the model best reproducing the experimental free energy of solvation has  $r_{\text{min}}$  increased by 7.5%. For models M1+ and M4+ the corrected solvation free energies are between  $-75.6$  and  $-76.6$  kcal mol $^{-1}$  (see Table 6). This is in excellent agreement with our previous studies, including the vibrational relaxation time<sup>30</sup> and the 1D- and 2D-infrared spectroscopies.<sup>59</sup> This finding applies to models M1, M2, and M4, but less for M5, which is a somewhat imbalanced model as it employs MTPs on the water but with van der Waals parameters from the TIP3P parametrization. For  $\text{OH}^-$  similar effects are found – *i.e.* increasing  $r_{\text{min}}$  from standard UFF values yields better agreement with experimental solvation free energies. From the data in Table 6, increasing  $r_{\text{min}}$  by 7.5% (M1+) and 15% to 20% (M2+++ , M4+++ and M5+++), respectively, yields free energies of solvation commensurate with experiment.

### 3.4 Scoring

To sample configurational space of a system, some models are much more expensive computationally than others. As configurational sampling is the most time-consuming step in the free energy calculations, it would be quite beneficial if this could be done using a computationally cheaper model. With the present setup, M4 takes twice as long and M5 about 6 times longer compared to M1, M2, or M3, which are the cheapest models. Once the trajectories are generated, the choice of model for analysis is less critical as the analysis step is computationally far less demanding than configurational sampling. To explore this possibility, simulations were carried out using M1++ (which is the cheapest model) for  $\text{CN}^-$  in  $\text{H}_2\text{O}$  for conformational sampling. The free energy calculations were then carried out with M1, M4 and M5 – denoted *e.g.* as (M1/M4) for sampling and scoring with M1 and M4, respectively – employing the same LJ parameters that were used in the configurational sampling. The computed solvation free energies are  $-74.6$ ,  $-56.0$ , and  $-77.2$  kcal mol $^{-1}$ , for (M1/M4), (M1/M5), and (M1/M1), and those from M1+++ ( $r_{\text{min}}/r_{\text{UFF}} = 1.20$ ) trajectories



**Fig. 4** The radial distribution functions from the trajectories generated with various models for  $r_{\min} = 1.075 r_{\text{UFF}}$ . M1 is shown in green, M4 in black, and M5 in red. (a)  $g_{\text{O}_w\text{C}_{\text{CN}}}(r)$ , (b)  $g_{\text{O}_w\text{N}_{\text{CN}}}(r)$ , (c)  $g_{\text{O}_w\text{O}_{\text{OH}}}(r)$ , and (d)  $g_{\text{O}_w\text{H}_{\text{OH}}}(r)$ . The inset of (c) shows  $N(r)$  for  $\text{O}_w$  around  $\text{O}_{\text{OH}}$  (see text).

are  $-70.6$ ,  $-52.5$ , and  $-72.8$  kcal mol $^{-1}$ . (M4/M4)++ and (M4/M4)+++ yield solvation free energies of  $-77.2$  and  $-72.4$  kcal mol $^{-1}$ , compared to  $-74.6$  and  $-70.6$  kcal mol $^{-1}$  from (M1/M4), which are 2 to 3 kcal mol $^{-1}$  lower in magnitude than those from (M4/M4). Hence, there are some differences from the scoring approach while the overall magnitude of the solvation free energy is retained. Still, most of the solvation free energies are within the experimental errors. On the other hand, analyses using M5 give much reduced solvation free energies. In order to find the cause for this, the solvent structure around the ion was considered next.

The radial distribution functions (RDFs) of the water oxygen atoms around a selected atom of the two anions were computed from the trajectories generated using M1, M4, or M5 for  $r_{\min}/r_{\text{UFF}} = 1.075$ . As can be seen in Fig. 4, the distribution of the water oxygen atoms depends on the type of model used in generating the trajectories. There is a significant difference between the RDFs from trajectories run with M1 and those from using M4 or M5, specifically for the first solvation shell. We also note that the RDFs for M5 have a different long-range behaviour compared with those for M1 and M4, specifically for the location of the maxima and minima of the RDFs. The different water distribution is the main reason why solvation free energies from trajectories generated with M1 and scored with M5 agree poorly with experiment. Conversely, if the phase space sampled is comparable, scoring may be a convenient means to somewhat improve the computation of an observable at reduced computational cost (here a factor of two – the difference in computational time between M4 and M1).

Finally, the coordination number for  $\text{OH}^-$  was determined from the RDFs according to

$$N(r) = 4\pi \int_0^{r_s} r'^2 g(r') \rho dr', \quad (18)$$

which is reported in the inset of Fig. 4(c). The coordination numbers for water around the hydroxide ion obtained from  $g_{\text{O}_w\text{O}_{\text{OH}}}(r)$  are 5.8 for M1, 5.0 for M4, and 5.1 for M5 for  $r_s$  corresponding to the first minimum of  $g_{\text{O}_w\text{O}_{\text{OH}}}(r)$ . For M4 and M5 this agrees with the experimentally determined value where it has been found that at low  $\text{OH}^-$  concentrations, clusters with 4 and 5 surrounding water molecules are prevalent.<sup>64,65</sup> The coordination numbers computed from AIMD studies with PW91, BLYP and HCTH functionals are 4.2, 4.8 and 4.7, respectively.<sup>66</sup>

## 4 Conclusions

In the present work, the solvation free energies of two simple anions,  $\text{CN}^-$  and  $\text{OH}^-$ , were computed using a number of different interaction models. Specifically, electrostatics at the level of usual point charge models, and more elaborate multipolar representations for both the solute and the solvent, were employed in sampling configurational space and computing the solvation free energy. Previous work along similar lines showed that differences between experimental observations and computations in simulations with accurate multipolar electrostatics can be used to refine other force field terms.<sup>30,59</sup> In the present work, computed solvation free energies are in close agreement with experimental estimates if the van der Waals ranges are enlarged by 7.5% (for  $\text{CN}^-$ ) and 10% to 20% (for  $\text{OH}^-$ ) relative to their UFF-values. For  $\text{CN}^-$  in  $\text{H}_2\text{O}$ , such a modification is on par with the findings from vibrational relaxation of  $\text{CN}^-$  in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  and the 2D-infrared spectroscopy in  $\text{D}_2\text{O}$ , where the use of the *same* multipolar representation and increased van der Waals ranges gave almost quantitative agreement with experiment.<sup>30,59</sup> It is quite remarkable that *one* single physically motivated force field parametrization can capture spectroscopy, nuclear dynamics and thermodynamics quantitatively.

As expected, solvation free energies are sensitive to the nonbonded interactions whereas bonded interactions seem to be much less important in the systems studied here. It is interesting to note that charges from electrostatic potentials using a grid-based method (CHELPG) on the solute ions perform well for the present applications while the computational effort is reduced compared to multipolar electrostatics. Anharmonic potentials and multipolar force fields are more expensive computationally and their influence on the hydration free energy appears to be small, which is in contrast to vibrational energy relaxation, where anharmonic potentials and MTPs were essential for the correct description of the process.<sup>30</sup>

For more efficient simulations, the possibility was explored to sample conformational space using a “cheap” model and “score” the trajectory with a more expensive model. In this scoring only the electrostatics was changed whereas the van der Waals ranges were identical to those used in the conformational sampling. We found that (M1, M4) – *i.e.* sampling with M1 and scoring with M4 – gives results close to (M4, M4) although M1 is a point charge model and M4 is a multipolar model. It was also found that pronounced differences in the



solvation structure around the ion make “scoring” almost impossible, such as for (M1, M5). In other words, if the conformational ensemble sampled by the “cheap” model differs too much from that corresponding to the “expensive” one, “scoring” is not a viable alternative.

One a priori possibility to determine how meaningful such an approach is likely to be is to reevaluate energies from the “cheap” trajectory using an “expensive” model. For several thousand configurations sampled by M1, the energies of the snapshots were reevaluated using M4 and M5 and compared with those from M1. The correlation coefficients between M1 and M4/M5, respectively, of a linear fit are 0.95 and 0.88, which suggests that “scoring” with M5 is likely to be meaningless. This may be an efficient means to select suitable simulation conditions for broader refinement studies involving larger molecules for which sampling conformational space with rigorous multipolar electrostatics will be considerably more time-consuming.<sup>31,44,47</sup> Such a fine-tuning of the nonbonded parameters is relevant for more quantitative applications of force fields in atomistic simulations, particularly for protein–ligand binding studies.<sup>31</sup>

## Acknowledgements

The authors gratefully acknowledge financial support from the Swiss National Science Foundation through grant 200021-117810 and the NCCR MUST. The authors thank Dr Tristan Bereau for helpful discussions.

## References

- J. T. Hynes, *Annu. Rev. Phys. Chem.*, 1985, **36**, 573–597.
- B. Bagchi, *Annu. Rev. Phys. Chem.*, 1989, **40**, 115–141.
- D. Laage, G. Stirnemann, F. Sterpone, R. Rey and J. T. Hynes, *Annu. Rev. Phys. Chem.*, 2011, **62**, 395–416.
- P. Jungwirth and D. Tobias, *Chem. Rev.*, 2006, **106**, 1259–1281.
- A. Savelyev and G. A. Papoian, *J. Phys. Chem. B*, 2008, **112**, 9135–9145.
- W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, 1985, **83**, 3050–3054.
- P. A. Bash, U. C. Singh, R. Langridge and P. A. Kollman, *Science*, 1987, **236**, 564–568.
- D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, 1989, **18**, 431–492.
- J. Åqvist, *J. Phys. Chem.*, 1990, **94**, 8021–8024.
- M. W. Wong, M. J. Frisch and K. B. Wiberg, *J. Am. Chem. Soc.*, 1991, **113**, 4776–4782.
- P. Kollman, *Chem. Rev.*, 1993, **93**, 2395–2417.
- D. Sitkoff, K. A. Sharp and B. Honig, *J. Phys. Chem.*, 1994, **98**, 1978–1988.
- G. Hummer, L. R. Pratt and A. E. García, *J. Phys. Chem.*, 1996, **100**, 1206–1215.
- J. Florián and A. Warshel, *J. Phys. Chem. B*, 1997, **101**, 5583–5595.
- D. J. Giesen, G. D. Hawkins, D. A. Liotard, C. J. Cramer and D. G. Truhlar, *Theor. Chem. Acc.*, 1997, **98**, 85–109.
- M. Sprik and G. Ciccotti, *J. Chem. Phys.*, 1998, **109**, 7737–7744.
- S. C. L. Kamerlin, M. Haranczyk and A. Warshel, *J. Phys. Chem. B*, 2009, **113**, 1253–1272.
- C. M. Baker, P. E. M. Lopes, X. Zhu, B. Roux and A. D. MacKerell Jr, *J. Chem. Theor. Comput.*, 2010, **6**, 1181–1198.
- Y. Shi, C. Wu, J. W. Ponder and P. Ren, *J. Comput. Chem.*, 2011, **32**, 967–977.
- W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
- W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.
- P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
- S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
- F. Sedlmeier and R. R. Netz, *J. Chem. Phys.*, 2013, **138**, 115101.
- G. Kaminski, E. M. Duffy, T. Matsui and W. L. Jorgensen, *J. Phys. Chem.*, 1994, **98**, 13077–13082.
- C. Oostenbrink, A. Villa, A. E. Mark and W. F. Van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656–1676.
- D. L. Mobley, E. Dumont, J. D. Chodera and K. A. Dill, *J. Phys. Chem. B*, 2007, **111**, 2242–2254.
- D. Horinek, S. I. Mamatkulov and R. R. Netz, *J. Chem. Phys.*, 2009, **130**, 124507.
- M. W. Lee and M. Meuwly, *J. Phys. Chem. A*, 2011, **115**, 5053–5061.
- T. Bereau, C. Kramer, F. W. Monnard, E. S. Nogueira, T. R. Ward and M. Meuwly, *J. Phys. Chem. B*, 2013, **117**, 5460–5471.
- R. Radmer and P. A. Kollman, *J. Comput. Chem.*, 1997, **18**, 902–919.
- F. M. Ytreberg, R. H. Swendsen and D. M. Zuckerman, *J. Chem. Phys.*, 2006, **125**, 184114.
- B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- N. Plattner and M. Meuwly, *Biophys. J.*, 2008, **94**, 2505–2515.
- S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.
- W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
- W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- S. Midda and A. K. Das, *Int. J. Quantum Chem.*, 2004, **98**, 447–455.
- M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara,

- K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revision B.01*, Pittsburgh, PA, 2003.
- 41 A. J. Stone, *J. Chem. Theor. Comput.*, 2005, **1**, 1128–1132.
- 42 S. E. Bradforth, E. H. Kim, D. W. Arnold and D. M. Neumark, *J. Chem. Phys.*, 1993, **98**, 800–810.
- 43 E. M. Patrino and P. Paredes-Olivera, *Surf. Sci.*, 2003, **527**, 149–162.
- 44 C. Kramer, P. Gedeck and M. Meuwly, *J. Comput. Chem.*, 2012, **33**, 1673–1688.
- 45 C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361–373.
- 46 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
- 47 C. Kramer, P. Gedeck and M. Meuwly, *J. Chem. Theor. Comput.*, 2013, **9**, 1499–1511.
- 48 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 49 M. Yoneya, H. J. C. Berendsen and K. Hirasawa, *Mol. Simul.*, 1994, **13**, 395–405.
- 50 N. Kumagai, K. Kawamura and T. Yokokawa, *Mol. Simul.*, 1994, **12**, 177–186.
- 51 N. Plattner, M. W. Lee and M. Meuwly, *Faraday Discuss.*, 2010, **147**, 217–230.
- 52 J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.
- 53 T. P. Straatsma and J. A. McCammon, *J. Chem. Phys.*, 1991, **95**, 1175–1188.
- 54 R. W. Zwanzig, *J. Chem. Phys.*, 1954, **22**, 1420–1426.
- 55 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.
- 56 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
- 57 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1339–1350.
- 58 M. Zacharias, T. P. Straatsma and J. A. McCammon, *J. Chem. Phys.*, 1994, **100**, 9025–9031.
- 59 M. W. Lee, J. K. Carr, M. Göllner, P. Hamm and M. Meuwly, *J. Chem. Phys.*, 2013, **139**, 054506.
- 60 R. G. Pearson, *J. Am. Chem. Soc.*, 1986, **108**, 6109–6114.
- 61 R. Pliego Jr. and J. M. Riveros, *Chem. Phys. Lett.*, 2000, **332**, 597–602.
- 62 E. Harder and B. Roux, *J. Chem. Phys.*, 2008, **129**, 234706.
- 63 G. Lamoureux and B. Roux, *J. Phys. Chem. B*, 2006, **110**, 3308–3322.
- 64 A. Botti, F. Bruni, S. Imberti, M. Ricci and A. Soper, *J. Chem. Phys.*, 2003, **119**, 5001–5004.
- 65 A. Botti, F. Bruni, S. Imberti, M. Ricci and A. Soper, *J. Chem. Phys.*, 2004, **120**, 10154–10162.
- 66 M. E. Tuckerman, A. Chandra and D. Marx, *Acc. Chem. Res.*, 2006, **39**, 151–158.