Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory

*Georgios Gerogiokas, † Gaetano Calabro, † Richard H. Henchman, ‡ Michelle W. Y. Southey, § Richard J. Law, § and Julien Michel †**

† EastCHEM School of Chemistry, Joseph Black Building, The King's Buildings, Edinburgh, EH9 3JJ, United Kingdom ‡ Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom and School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom § Evotec (UK) limited, 114 Milton Park, Abingdon, Oxfordshire, OX14 4SA, UK

Keywords: hydration thermodynamics, cell theory, thermodynamic integration

Abstract: An efficient methodology has been developed to quantify water energetics by analysis of explicit solvent molecular simulations of organic and biomolecular systems. The approach, grid cell theory (GCT), relies on a discretization of the cell theory methodology on a three-dimensional grid to spatially resolve the density, enthalpy and entropy of water molecules in the vicinity of solute(s) of interest. Entropies of hydration are found to converge more efficiently than enthalpies of hydration. GCT predictions of free energies of hydration on a dataset of small molecules are strongly correlated with thermodynamic integration predictions. Agreement with experiment is comparable for both approaches. A key advantage of GCT is its ability to provide,

from a single simulation, insightful graphical analyses of spatially resolved components of the enthalpies and entropies of hydration.

1. Introduction

The accurate description of hydration and associated energetics is key for adequate molecular modeling of a broad range of fundamental chemical processes, ranging from chemical/enzymatic reactions, ligand binding to protein folding.[1] After decades of intense research, water remains a difficult liquid to model quantitatively and there is still debate about the behavior of water structure and dynamic near solutes, interfaces, or in confined environments.[2-6] Methodologies such as thermodynamic integration or free energy perturbation can account for solvent effects in computed free energy changes,[7] but enthalpic and entropic components are less easily accessible.[8]One can use a finite difference approximation to compute an entropy change from free energy changes computed at different temperatures, but the statistical uncertainties in the computed entropies is frequently much higher than for the computed free energies, to illustrate Kubo et al estimated entropies of hydration for a series of small molecules that were ca. 10 fold less precise than the free energies computed by thermodynamic integration.[9] Alternatively, neglecting a pressure-volume term, one can derive entropy changes by subtracting from a computed free energy change the difference in the average total potential energy of the systems simulated in the two thermodynamic states of interest.[6] Because fluctuations in total potential energy increase with the number of simulated particles, the procedure does not scale favorably with increased box sizes. In addition, hydration free energies computed with alchemical methods typically do not provide details about contributions from different solvent regionsNew methodologies are warranted to provide missing details that are not readily obtained from perturbation theories and to progress understanding of

molecular recognition in aqueous conditions. It is desirable to use theory and computation to help interpret for instance biophysical measurements of enthalpies and entropies.[10-11] The computation of entropies and enthalpies is also important to validate potential energy functions, as evidenced by a recent study of Baron and Molinero on the thermodynamic signature of ligand binding to model cavities.[12] Progress on this front is expected to expand the scope and opportunities for molecular modeling in pharmaceutical research,[13] or nanoscale materials.[14]

Several methodologies have been recently proposed to compute free energies, enthalpies and entropies of water molecules in different environments through for instance state probabilities,[15] or thermodynamic integration.[16-17] One such methodology, GIST (grid inhomogeneous solvation theory), has been proposed by Nguyen et al. to produce detailed analyses of hydration thermodynamics.[18] GIST is based on a discretization of the inhomogeneous fluid solvation theory (IFST) method developed by Lazaridis and co-workers.[19-21] Various implementations of IFST,[22-24] including the Watermap program of the company Schrödinger, have proven popular to compute the thermodynamic properties of water molecules in protein binding sites. Recently, Raman and Mackerell have developed a protocol that combines Grand Canonical Monte Carlo sampling and GIST analyses to compute water properties near proteins.[25] Huggins and Payne have recently reported a GIST study of the hydration thermodynamics of set small molecules using a custom implementation. The results produced by GIST were encouraging, but also indicated a number of issues, such as the need to post-process a large number of snapshots to converge reasonable entropies of hydration.[22] Further methodological work is desirable to facilitate routine analyses of hydration thermodynamics. Our alternative approach relies on the cell theory methodology. The application of cell theory to model the thermodynamic properties of water and small molecules has been pioneered by Henchman and co-workers.[26-30] These

results have suggested that the methodology could be an intriguing alternative to IFST and relative strengths and weaknesses should be further explored. This issue is addressed by the present report which describes the extension of the cell theory methodology by discretization on a three dimensional grid. Parameters for bulk water are derived, and a detailed analysis of the convergence behavior of the enthalpic and entropic components of grid cell theory (GCT) is undertaken. The computed free energies of hydration for a dataset of small molecules are then compared with thermodynamic integration calculations and experimental data. Finally, free energies, enthalpies and entropies of hydration are spatially resolved to provide insightful graphical analyses of the contributions of different water regions to hydration thermodynamics.

2. Theory and Methods

2.1 Cell Theory.

The enthalpy of hydration at temperature T for the transfer of one solute X from a gas phase concentration c^0_g such that it has an equivalent mole fraction $x^0_w = V_w c^0_x(g)$ in a solution of N_w water molecules is given by Eq 1. V_w is the mean volume of a water molecule (ca. 30 Å^3) Note that no pressure volume term is present for this equivalent mole fraction solvation process, unlike the frequently used constant concentration solvation process.[29]

(1) $\Delta H_{X+w} = \langle U_{w(X)} \rangle - \langle U_{w(l)} \rangle - \langle U_{X(g)} \rangle$

where the symbols U denote the ensemble averages of the total potential energy of the solvated solute, bulk water, or gas phase solute recorded over the course of a molecular simulation. When the potential energy is computed with an additive pairwise force field, it is convenient to rewrite this equation in terms of intramolecular and intermolecular components.

(2) $\Delta H_{X+w} = \langle U^{intra}_{X,w(X)} \rangle - \langle U^{intra}_{X(g)} \rangle + \frac{1}{2}\langle U^{inter}_{X,w(X)} \rangle + \sum_{i=1}^{N_w} \left(\frac{1}{2}\langle U^{inter}_{w_i,w(X)} \rangle - U^{inter}_{w(l)} \right)$

where the last term is the mean bulk intermolecular energy of a water molecule. The water molecules are assumed to be rigid and intramolecular water energy terms have been omitted.

The entropy of hydration shown in Eq 3 is decomposed in six contributions.

(3) $\Delta S^{\circ}_{X+w} = \Delta S^{\circ,tr}_{X} + \Delta S^{rot}_{X} + \Delta S^{int}_{X} + \Delta S^{ori}_{w,X} + \Delta S^{vib}_{w,X} + \Delta S^{lib}_{w,X}$

In cell theory, hindered translations of the solute are accounted with a three dimensional harmonic potential and the solute translational entropy is given by Eq 4.

(4) $\Delta S^{\circ,tr}_{X} = k_B \ln\left\{\frac{1}{V_w}\prod_{j=1}^{3}\frac{2k_BTe^{1/2}}{\langle F^{j}_{X(aq)}\rangle}\right\}$

e is the natural logarithm base, k_B the Boltzmann constant, V_w is the volume available to a water molecule in the gas phase with the above standard state definition. $\langle F^{j}_{X(aq)}\rangle$ is the ensemble average of half the magnitude of the average force component felt by the solute in a frame of reference defined by the three principal axes j of the solute.

A similar expression applies for the solute rotational entropy.

(5) $\Delta S^{rot}_{X} = \min(0, k_B \ln\left\{\frac{1}{V_w}\prod_{j=1}^{3} r^{j}_{X}\frac{2k_BTe^{1/2}}{\langle \tau^{j}_{X(aq)}\rangle}\right\})$

$\langle \tau^{j}_{X(aq)}\rangle$ is the ensemble average of half the magnitude of the force constants for hindered rotations (torques) about the three principal axes. r^{j}_{X} is the radius of the solute along principal axis j to the edge of the solute van der Waals surface. The ratio of the product of the principal axes radii to the mean volume of a water molecule accounts for the number of orientational minima. For some small weakly interacting solutes (e.g. methane) the harmonic approximation sometimes breaks down and adjacent orientational minima overlap, in which case Eq 5 may yield an increase in rotational entropy upon hydration. As this is not physically sound, the minimum entropy change computed with Eq 5 is set to 0 cal.mol^{-1}.K^{-1}. Previous implementations have attempted to evaluate rotational entropies only along axes that are more rotationally

hindered,[28] but this approach is not followed here because it requires a-priori knowledge about the solute of interest. The breakdown of the harmonic approximation for the evaluation of the rotational entropy change is also less likely for more complex typical drug-like small molecules.

The solute internal entropy is given by Eq 6 and is the ratio of the Gibbs entropies of the probability distribution functions of the **r** Cartesian internal coordinates that define the solute conformation in solution and in the gas phase.

(6) $\Delta S_X^{int} = -k_B \int \frac{\rho(\boldsymbol{r})_{w(X)} \ln p(\boldsymbol{r})_{w(X)} d\boldsymbol{r}}{\rho(\boldsymbol{r})_{X(g)} \ln p(\boldsymbol{r})_{X(g)} d\boldsymbol{r}}$

The orientational entropy of the N_w water molecules solvating the solute X is given by Eq 7:

(7) $\Delta S_{w,X}^{ori} = N_w k_B \ln\left\{\frac{\langle \Omega_{w(X)}^{ori} \rangle}{\Omega_{w(l)}^{ori}}\right\}$

$\Omega^{ori}{}_{w(l)}$ is the average number of orientations in bulk water and is a function of the water model and simulation conditions. $<\Omega^{ori}{}_{w(X)}>$ is the ensemble average of the average number of orientations of the N_w water molecules, *i.e.* Eq 8.

(8) $\left\langle \Omega_{w(X)}^{ori} \right\rangle = \left\langle \frac{1}{N_w} \sum_{i=1}^{N_w} \Omega_{w,i}^{ori} \right\rangle$

Following previous work, a generalized Pauling's residual ice entropy model is used to compute $\Omega^{ori}{}_{w,i}$ for each water molecule i.[29, 31] Eq 9 is used by default, unless the oxygen atom of water i is within the coordination shell of solute polar atom in which case Eq 10 applies.

(9) $\Omega_{w,i}^{ori} = \frac{N_{a,i}(N_{a,i}-1)}{2}\left(\frac{N_{a,i}-2}{N_{a,i}}\right)^2$

(10) $\Omega_{w,i}^{ori} = \frac{N_{a,i}^{eff}\left(N_{a,i}^{eff}-1\right)}{2}\left(\frac{N_w^{bulk}-2}{N_w^{bulk}}\right)^{2-p_{HB}^X}$

$N_{a,i}$ is the number of hydrogen bond acceptors (*i.e.* oxygen atoms of other water molecules and solute polar atoms) within the coordination shell of water i. For the purpose of defining such coordination shells, a radius of 3.4 Å centered on the water oxygen atom is used and solute atom

types are defined. N^{bulk}_{w} is the coordination number of bulk water. Its value depends on the water model used and the simulation conditions. $N^{eff}_{a,i}$ is the effective coordination number of water *i*. This term accounts for the different number of orientations available to a water molecule near a solute with respect to the bulk environment. To compute the effective coordination number, $N_{a,i}$ is decomposed into contributions from the number of solute acceptor atoms ($N_{X,i}$), the number of water oxygen atoms that are within the first coordination shell of any solute acceptor atom ($N_{ws,i}$), and the number of remaining water oxygen atoms ($N_{wb,i}$) .

(11) $N_{a,i} = N_{X,i} + N_{w_s,i} + N_{w_b,i}$

Next, the ratios of each type of acceptors that are hydrogen bonded to water *i* is defined.

(12) $p^{X}_{HB} = \frac{N_{XHB,i}}{N_{X,i}}$; $p^{w_s}_{HB} = \frac{N_{w_sHB,i}}{N_{w_s,i}}$; $p^{w_b}_{HB} = \frac{N_{w_bHB,i}}{N_{w_b,i}}$

And the effective coordination number is given by Eq 13:

(13) $N^{eff}_{a,i} = \frac{N_{XHB,i}+N_{w_sHB,i}+N_{w_bHB,i}}{\max(p^{X}_{HB},p^{w_s}_{HB},p^{w_b}_{HB})}$

Eqs 12-13 require the definition of hydrogen-bonding interactions. Following previous work, a force-based definition is used.[28] For classical water models such as TIP4P that do not have Lennard-Jones radii on the hydrogen atoms of water this amounts to forming a hydrogen bond with the acceptor atom A with the largest value of q_A/r^2_{AH}, where r_{AH} is the separation between atoms H and A and q_A the atomic partial charge of atom A.

The vibrational and librational water entropies are given by Eqs 14 and 15, where $\langle F^{j}_{w(X)} \rangle$, $\langle \tau^{j}_{w(X)} \rangle$, are the ensemble averages of the average half magnitudes of the forces and torques measured along and about the principal axes of the N_w water molecules solvating *X*. $F^{j}_{w(l)}$ and $\tau^{j}_{w(l)}$ are the average forces and torques of bulk water and depend on the water model and the simulation conditions.

(14) $\Delta S_{w,X}^{vib} = N_w k_B \ln\left\{\prod_{j=1}^{3} \frac{F_{w(l)}^{j}}{\langle F_{w(X)}^{j} \rangle}\right\}$

(15) $\Delta S_{w,X}^{lib} = N_w k_B \ln\left\{\prod_{j=1}^{3} \frac{\tau_{w(l)}^{j}}{\langle \tau_{w(X)}^{j} \rangle}\right\}$

2.2 Grid Cell Theory.

In grid cell theory the water components of the enthalpies and entropies of hydration are spatially resolved on a three dimensional grid positioned to cover a volume of space **s** around a solute *X*. The grid may be non-uniform and unevenly spaced, but for simplicity a cubic evenly spaced grid is used here. To speed up convergence and to facilitate the analysis of the grid properties, the solute is constrained to adopt a single conformation **r**. As an alternative, positional harmonic restraints on the solute heavy atoms can be used, in which case a narrow range of conformations similar to **r** is allowed. If multiple conformations of *X* are of interest, one could perform separate calculations on each conformation. The use of constraints or restraints for *X* has two major benefits. Firstly, it speeds up convergence of the calculations as only rigid body translations and rotations of water molecules need to be sampled. Secondly, it facilitates graphical analysis of the computed grids. It is possible to perform GCT analyses on unconstrained/unrestrained solutes, but this was not attempted in this study.

Because the solute is rigid, the solute intramolecular energy terms cancel out from Eq 2 and the contribution of a volume of space **s** to the enthalpy of hydration is given by Eq 16:

(16) $\Delta H_{X(\boldsymbol{r})+w}^{\boldsymbol{s}} = \left\langle \frac{1}{2} U_{X(\boldsymbol{r}),w(X(\boldsymbol{r}))}^{inter} \right\rangle + \left\langle \sum_{i=1}^{N_w^{\boldsymbol{s}}} \left(\frac{1}{2} U_{w_i,w(X(\boldsymbol{r}))}^{inter} - U_{w(l)}^{inter} \right) \right\rangle = \Delta H_{X(\boldsymbol{r})}^{\boldsymbol{s}} + \Delta H_{w}^{\boldsymbol{s}}$

N^{s}_{w} is the number of water molecules present within **s**. Note that the intermolecular interactions of particles in **s** with particles within and outside of **s** are taken into account.

Similarly the change in solute internal entropy cancels out and the contribution of the volume of space **s** to the entropy of hydration is given by Eqs 17a and 17b:

(17a) $\Delta S^{\boldsymbol{s},\circ}_{X(\boldsymbol{r})+w} = \Delta S^{\boldsymbol{s},\circ,tr}_{X(\boldsymbol{r})} + \Delta S^{\boldsymbol{s},rot}_{X(\boldsymbol{r})} + \Delta S^{\boldsymbol{s},w}_{w}$

(17b) $\Delta S^{\boldsymbol{s},w}_{w} = \Delta S^{\boldsymbol{s},ori}_{w} + \Delta S^{\boldsymbol{s},vib}_{w} + \Delta S^{\boldsymbol{s},lib}_{w}$

The volume **s** is discretized into N_s volume elements of volume *V(k)* and cell parameters are computed as ensemble averages for each voxel according to Eqs 18-23:

(18) $\Delta H^{\boldsymbol{s}}_{X(\boldsymbol{r})}(k) = \left\langle \frac{\sum_{i=1}^{N_w^S} \frac{1}{2} U^{inter}_{w_i,X(\boldsymbol{r})} I(k)}{\max\left(1, \sum_{i=1}^{N_w^S} I(k)\right)} \right\rangle$

(19) $\Delta H^{\boldsymbol{s}}_{w}(k) = \left\langle \frac{\sum_{i=1}^{N_w^S} \left(\frac{1}{2} U^{inter}_{w_i,w(X(\boldsymbol{r}))} - U^{inter}_{w(l)}\right) I(k)}{\max\left(1, \sum_{i=1}^{N_w^S} I(k)\right)} \right\rangle$

(20) $\Omega^{ori}_{w}(k) = \left\langle \frac{\sum_{i=1}^{N_w^S} \Omega^{ori}_{w,i} I(k)}{\max\left(1, \sum_{i=1}^{N_w^S} I(k)\right)} \right\rangle$

(21) $F^{j}_{w}(k) = \left\langle \frac{\sum_{i=1}^{N_w^S} F^{j}_{w,i} I(k)}{\max\left(1, \sum_{i=1}^{N_w^S} I(k)\right)} \right\rangle$

(22) $\tau^{j}_{w}(k) = \left\langle \frac{\sum_{i=1}^{N_w^S} \tau^{j}_{w,i} I(k)}{\max\left(1, \sum_{i=1}^{N_w^S} I(k)\right)} \right\rangle$

(23) $\rho(k) = \left\langle \sum_{i=1}^{N_w^S} \frac{I(k)}{V(k)\rho_b} \right\rangle$

where *I(k)* is an indicator function that is equal to 1 if the coordinates of the oxygen atom of water molecule *i* are within the volume covered by *k* and 0 otherwise. $\rho(k)$ is the ratio of water density in voxel *k* relative to bulk density ρ_b. ρ_b is function of the water model and the simulation conditions used. The denominator in Eqs 18-22 is used to normalise the voxel parameters to yield cell parameters per water molecule.

The average number of water molecules within voxel *k* and space **s** are, respectively:

(24a) $N_w(k) = \rho(k)V(k)$

(24b) $N_w(\boldsymbol{s}) = \sum_{k=1}^{N_s} N_w(k)$

(25) $\rho(\boldsymbol{s}) = \sum_{k=1}^{N_s} \rho(k) / N_s$

The solute and solvent hydration enthalpies of **s** are then given by Eqs 26-27. For some analyses it is convenient to use normalized quantities given by Eq 26a-27a, which yield enthalpic components per water molecule within **s**. Analysis of normalized quantities is useful to compare the properties of a water molecule in different environments.

(26) $\Delta H^{\boldsymbol{s}}_{X(\boldsymbol{r})} = \sum_{k=1}^{N_s} N_w(k) \Delta H^{\boldsymbol{s}}_{X(\boldsymbol{r})}(k)$

(26a) $\Delta H^{n,\boldsymbol{s}}_{X(\boldsymbol{r})} = \Delta H^{\boldsymbol{s}}_{X(\boldsymbol{r})} / N_w(\boldsymbol{s})$

(27) $\Delta H^{\boldsymbol{s}}_{w} = \sum_{k=1}^{N_s} N_w(k) \Delta H^{\boldsymbol{s}}_{w}(k)$

(27a) $\Delta H^{n,\boldsymbol{s}}_{w} = \Delta H^{\boldsymbol{s}}_{w} / N_w(\boldsymbol{s})$

The average water orientational number, forces and torques for **s** are given by Eqs 28-30:

(28) $\Omega^{ori}_w(\boldsymbol{s}) = \max(1, \frac{1}{N_w(\boldsymbol{s})} \sum_{k=1}^{N_s} N_w(k)\, \Omega^{ori}_w(k))$

(29) $F^j_w(\boldsymbol{s}) = \frac{1}{N_w(\boldsymbol{s})} \sum_{k=1}^{N_s} N_w(k)\, F^j_w(k)$

(30) $\tau^j_w(\boldsymbol{s}) = \frac{1}{N_w(\boldsymbol{s})} \sum_{k=1}^{N_s} N_w(k)\, \tau^j_w(k)$

Note that the orientational number of water within **s** cannot be lower than 1. The water orientational, vibrational and librational entropies of **s** are given by Eqs 31-33:

(31) $\Delta S^{\boldsymbol{s},ori}_{w,X(\boldsymbol{r})} = N_w(\boldsymbol{s}) k_B \ln\left\{\frac{\Omega^{ori}_w(\boldsymbol{s})}{\Omega^{ori}_{w(l)}}\right\}$

(32) $\Delta S^{\boldsymbol{s},vib}_{w,X(\boldsymbol{r})} = N_w(\boldsymbol{s}) k_B \ln\left\{\prod_{j=1}^{3} \frac{F^j_{w(l)}}{F^j_w(\boldsymbol{s})}\right\}$

(33) $\Delta S^{\boldsymbol{s},lib}_{w,X(\boldsymbol{r})} = N_w(\boldsymbol{s}) k_B \ln\left\{\prod_{j=1}^{3} \frac{\tau^j_{w(l)}}{\tau^j_w(\boldsymbol{s})}\right\}$

Similarly, where desirable, per water entropy changes per water molecule within **s** can be alternatively computed. It is also convenient to define with Eq 34 the contribution of water to the free energy of hydration a solute *X*.

(34) $\Delta G^{\boldsymbol{s}}_{w,X(\boldsymbol{r})} = \Delta H^{\boldsymbol{s}}_{X(\boldsymbol{r})+w} - T\Delta S^{\boldsymbol{s},w}_{w}$

The solute translational and rotational entropies are not discretized on a grid, but can be computed according to Eqs 4-5.

It is important to reflect that the above analysis is motivated by the wish to spatially resolve thermodynamic components of the hydration process. The usefulness of components analysis has been debated on several occasions. Formally, there is no unique way to partition a free energy into a sum of individual contributions because of coupling between individual contributions.[32] Nevertheless components analysis of individual contributions has been shown in several cases to yield meaningful interpretation of thermodynamic processes.[33-34] Huggins has analysed the excess entropy of liquid water using an entropy expansion based on the inhomogeneous fluid solvation theory formalism and concluded that a 2nd order expansion reproduced the full excess entropy of liquid water to within 20%.[35] By contrast, thermodynamic quantities are additive in the mean-field approximation made in cell theory which assumes that there exists an effective potential that only depends on the coordinate of that potential. The curvature of the potential is inferred from the force along that coordinate which explicitly represents the interactions of each molecule with all its neighbours. Because of the connection with the force, cell theory parameters can be directly linked to specific solvent motions and have thus a physical interpretation (e.g. how hindered are translations/rotations of a water molecule in a given region of space). Previous work has shown that this decomposition yields the excess entropy of liquid

water to within ca. 10% of experimental data.[27] While the present spatial decomposition is not unique, as demonstrated in the results section, the ability of the model to provide from summation of its components thermodynamic parameters that correlate well with thermodynamic integration and with experimental data suggests that a component analysis within the framework of cell theory is meaningful.

2.4 Preparation of molecular models

The methodology was tested with a range of systems of increasing complexity including a water box of 804 water molecules, and a dataset of small molecules (neon, xenon, chloride, sodium, methane, ethane, n-butane, isobutene, benzene, methanol, acetamide and N-methylacetamide). The TIP4P-Ew water model was used throughout.[36] The noble gas parameters were taken from Bondi,[37] and Guillot and Guissani,[38] the ion parameters from Joung and Cheatham.[39] The other small molecules were parameterized using the GAFF force field,[40] and AM1-BCC charges,[41] as implemented in the AMBER11 software suite.[42]

Unless otherwise mentioned, each small molecule was solvated in the waterbox of 804 TIP4P-Ew water molecules using the program leap. The resulting models were energy minimized and equilibrated under NPT conditions at 1 atm and 298 K. A velocity-verlet integrator and a time step of 2 fs was used. Temperature control was achieved with a Langevin thermostat and coupling constant of 5 ps^{-1},[43] whereas pressure control was achieved with a Berendsen barostat and coupling constant of 2 ps^{-1}.[44] The SHAKE method with a tolerance of 0.00001 Å was applied to constrain intramolecular degrees of freedoms in water molecules and bonds involving hydrogen atoms in solutes.[45-46] Electrostatic interactions were handled with the particle mesh Ewald method,[47-48] and a cutoff of 10 Å. Lennard-Jones interactions were truncated after 10 Å.

Simulations were run with the program sander until the box density was judged stable, which typically required ca. 100 ps. Harmonic positional restraints of force constant 10 kcal.mol^{-1}.Å^2 were applied to the solute heavy atoms throughout so that each solute retained its initial conformation and remained centered in the box.

2.5 Molecular Dynamics simulations.

Production molecular simulations were performed with the software Sire/OpenMM. In the present study, this program results from the runtime linking of the general purpose molecular simulation package Sire revision 1786,[49] with the GPU Molecular dynamics library OpenMM revision 3537.[50] Simulations were performed at 1 atm and 298 K using an atom-based Generalised Reaction Field non-bonded cutoff of 10 Å for the electrostatic interactions,[51] and an atom based non-bonded cutoff of 10 Å for the Lennard-Jones interactions. Small molecules or model ligands were restrained to their input conformation using harmonic positional restraints with a force constant of 10 kcal.mol^{-1}.Å^2 on the solute heavy atoms. A velocity-verlet integrator was used with a timestep of 2 fs. Temperature control was achieved with an Andersen thermostat and a coupling constant of 10 ps^{-1}.[52] Pressure control was achieved by attempting isotropic box edge scaling Monte Carlo moves every 25 time-steps. The intramolecular degrees of freedom of water molecules and bonds involving hydrogen atoms in the solutes were constrained using the OpenMM default error tolerance settings.

Unless otherwise stated, each system was simulated three times for 50 ns using the same starting conformation but a different random velocity assignment. Snapshots were stored every 1 ps in a DCD file format for subsequent analyses. Snapshots collected during the first 1 ns of each simulation were not considered in subsequent analyses.

2.6 Nautilus analyses.

The grid cell theory method described in section 2.2 was implemented in the trajectory analysis software Nautilus. Nautilus presently consists in a set of python scripts that link with the python libraries of Sire, Numpy[53] and MDanalysis.[54] A Nautilus analysis typically consists in the following workflow:

● *Conversion of a molecular dynamics trajectory into a collection of cell files.* Each solvent cell file stores the forces, torques, orientational number, energies and oxygen atom coordinates for every water molecule present within a defined volume of space **s1.** Each solute's forces, torques, radii and host-guest interaction energies are stored in separate cell files.

● *Conversion of the solvent cell files into grid cell files.* The data from a user-defined set of cell files is discretized on a volume of space **s2** at a desired density. Note that **s2** can be a subset of **s1** if desired. The main output is a grid cell file that lists the average forces, torques, orientational number, energies and densities of water for each voxel *k*. Additionally, a set of grid cell files listing components of the solvent entropies and enthalpies for each voxel *k* are output in pdb and dx file formats for visualization of the results using the software VMD.[55]

● *Evaluation of the thermodynamic properties of water*. The properties of a user-defined arbitrary volume of space **s3** (that can be a subset of **s2**) are computed by parsing the grid forces file.

● *Evaluation of the thermodynamic properties of the restrained solute*. If enthalpies and entropies of hydration are desired, solute cell files are processed to compute solutes translational and orientational entropies.

2.7 Thermodynamic integration calculations

Absolute free energies of hydration were computed using a single topology coupling paradigm,[56] as implemented in the simulation package Sire/OpenMM (Sire revision 1994, OpenMM 5.1). To facilitate comparison with GCT results, the same potential energy function, including solute restraints, was used throughout. Hydration free energies were computed using a two stage protocol. Each step was performed using 21 evenly spaced values of the coupling parameter λ (0.00, 0.05, … , 0.95, 1.00). Free energy gradients were evaluated using a finite-difference thermodynamic integration (FDTI) approach with Δλ set to 0.001.[57] Free energy changes were obtained by numerical integration of the free energy gradients using a polynomial regression scheme.[58] The free energy change for turning off the solutes atomic partial charges in vacuum was first computed ($\Delta G_{vac\_coul\_off}$). The free energy change for turning off the Lennard-Jones parameters of the discharged solutes in vacuum was then computed ($\Delta G_{vac\_LJ\_off}$). The solutes were then transferred to a waterbox of 804 TIP4P-Ew water molecules and the free energy change for turning on the Lennard-Jones interactions was computed ($\Delta G_{solv\_LJ\_on}$). Finally, the free energy change for restoring the solutes atomic partial charges in solution was computed ($\Delta G_{solv\_coul\_on}$). The hydration free energy was then given as:

(33) $\Delta G_{hyd} = \Delta G_{vac\_coul\_off} + \Delta G_{vac\_LJ\_off} + \Delta G_{solv\_LJ\_on} + \Delta G_{solv\_coul\_on}$

In order to avoid numerical instabilities, soft-core potential energy functions were used for transformations of the Lennard-Jones parameters.[59-60] An implementation identical to Michel et al. was used.[61] The softening parameter was set to δ=3.0 except for waterbox simulations of ethane, benzene, isobutane, n-butane and N-methylacetamide where δ was set to 4.0 to avoid abrupt changes in free energy gradients when the Lennard-Jones interactions of the solutes is restored in the waterbox. The waterbox FDTI simulations used the same input files as for the

GCT simulations. The vacuum simulations used the same input solute conformation as in the waterbox simulations. Each λ value was simulated for 1 ns (waterbox) or 100 ps (vacuum). To enable solvent re-equilibration upon changes in λ, statistics from the first 100 ps of the waterbox simulations were discarded. The overall sampling time for the FDTI protocol was thus 42 ns waterbox and 4.2 ns vacuum. This is broadly comparable to the sampling time of a 50 ns GCT simulation as in the current Sire/OpenMM implementation, the FDTI energy function is ca. 40% more expensive to evaluate than the default MD energy function (the vacuum simulations have a negligible computing time). The FDTI simulations can however be run in parallel. Each hydration free energy calculation was repeated three times using different random velocity assignments and the mean and standard errors were computed. All calculations were performed on a cluster of Tesla M2090 nodes using the OpenMM OpenCL platform in mixed precision mode.

3. Results and Discussion

3.1 Bulk water

Nautilus analyses require the evaluation of reference parameters for water in bulk conditions. Table 1 lists the reference bulk parameters used in this study. These were derived from the average forces, torques, orientational numbers, density and intermolecular energy per water molecule recorded over three separate 50 ns simulations of the 804 water molecules box. The results compare well to previously reported cell theory parameters for TIP4P-Ew,[27] though small differences are seen due to the use of a reaction field instead of particle mesh Ewald for the treatment of long-range electrostatic interactions. Excess enthalpies were corrected from the intermolecular energy with the dielectric depolarisation correction term.[36] Excess entropies were

computed with Eq 7 of ref[27] and compared with experimental data in Table 1. Agreement with the density and enthalpy of bulk water is good, whereas the entropy is overestimated by 10%.

A major goal of the present study is the ability to spatially resolve water properties. Grids made of voxels with sub-angstrom spacing enable fine spatial resolution but also require a larger number of snapshots to converge properties to an acceptable degree of precision because each trajectory snapshot will contain on average fewer water molecules within each voxel *k*. Figure 1 shows the convergence of the enthalpies and entropies of water for evenly distributed regions **s** of space that altogether define a cube of volume of 4096 $Å^3$ centered at the center of the box. Because the simulated system is isotropic, in the limit of sufficient sampling, the cell parameters of each region **s** should match the reference bulk parameters and the excess enthalpies and entropies should all converge to zero. In practice, deviations are observed due to sampling errors. The distributions are approximately Gaussian and become more sharply peaked as the number of snapshots used for averaging increases. As expected decreasing the grid resolution decreases the spread of the distributions. The enthalpy distribution is not exactly centered on zero because the mean water intermolecular energy from the specific run used to produce Figure 1A deviates slightly from the reference bulk value. The entropy distribution on the other hand is almost exactly centered on zero. No systematic discretization errors are apparent. As long as sufficient statistics have been collected for each region, the results will converge to bulk properties, regardless of the grid resolution. The entropy is consistently better converged than the enthalpy, for instance with a 1 Å spacing and 50 ns averaging time, the standard deviation of for ΔH^w_s is 30% larger than for $-T\Delta S^w$ ($\sigma = 0.045$ kcal.mol^{-1} and $\sigma = 0.035$ kcal.mol^{-1} respectively).

Figure 2 illustrates the enthalpy and entropy of water within cubic regions **s** of increasing edge lengths. Again, all results should converge to zero, but deviations will occur in practice owing to

finite sampling errors. As expected, the uncertainty in the computed thermodynamic properties increases with the volume monitored by **s**. The enthalpy diverges more rapidly than the entropy components. The convergence behavior of the enthalpy and entropy components can be rationalized by inspecting Table 1 and Eqs 26 and 30-32. As the volume of space covered by **s** increases, the number of water molecules N_w contributing to the enthalpy/entropy increases, and small random deviations of the averaged forces, torques and energies will contribute an increasingly significant enthalpy/entropy change. The enthalpy diverges more rapidly because more sampling is required to converge the mean per water intermolecular energy and there is greater uncertainty in the value of the reference bulk parameter. For bulk water, a cube of edge length 12 Å and three simulations of 50 ns yield converged predictions to within ±0.2 and ±0.05 kcal.mol^{-1} for the enthalpy and entropy respectively. The convergence behavior is likely to be system dependent; protein binding sites will typically cover a smaller volume and contain fewer water molecules, which should facilitate convergence. On the other hand greater correlation times are expected for water molecules in the vicinity of biomolecular surfaces. Nevertheless, the present results indicate that for quantitative studies, sampling errors introduce a practical upper limit to the volume of space **s** that can be reliably monitored with GCT.

3.2 Small molecules

Enthalpies, entropies and free energies of hydration were first computed for a dataset of four monoatomic solutes. For such simple solutes, the properties of water molecules can be conveniently monitored as a function of the distance of the water oxygen atom to the center of the solute. In addition, the lack of intramolecular degrees of freedom and therefore contributions from solute conformational changes facilitates comparison with experimental data. Because there

is debate about the extent of solute perturbations on water structure,[5, 62] a detailed analysis of the computed enthalpies and entropies of hydration as a function of the volume monitored was undertaken to establish how far away from a solute one should monitor solvent properties to observe convergence.

Figure 3 presents the value of the water enthalpy and entropy components as the radius of the spherical region of space **s** centered on the center of the solute is increased from 3 to 14 Å. For these systems the box edge lengths fluctuate around 30 Å, so larger region radii would include some water molecules twice. The solute-solvent enthalpy term $\Delta H^{s}_{X(r)}$ is well converged and fully accounted for radii greater than the 10 Å non-bonded cutoff used in the simulations. On the other hand, the solvent-solvent enthalpy term ΔH^{s}_{w} is noisier, and the standard error of the mean is ± 0.3 kcal.mol^{-1} for a radius of 10 Å, and increases rapidly beyond this value. For the neutral solutes this term appears to be reasonably flat beyond 10 Å, but it drifts upwards or downwards for chloride and sodium respectively. Similar trends are observed for the entropy components though they are typically more reproducible. The orientational entropy appears noisier than the other terms. In addition the vibrational and librational entropy components show also a systematic, albeit less pronounced, drift beyond 10 Å radii. Comparison with IFST studies of small molecule hydration is of interest. Huggins and Payne reported that, with a monitoring region of a 12 Å radius centered on the solutes of interest, about 5 million snapshots were needed to converge entropies to within one decimal, whereas only about 20,000 snapshots were needed to converge enthalpies to the same level of precision.[22]. Alternative IFST implementations that use a nearest-neighbor method instead of histograms to estimate entropies may be more efficient.[22] A more rigorous comparison would require analysis of the same trajectories of

identical systems. Nevertheless, it appears that the GCT water entropy estimates, obtained here by post-processing 150,000 snapshots, converge faster than the IFST entropy estimates.

The origin of the systematic drifts in the enthalpy or entropy components was explored further by repeating simulations of sodium in boxes of larger size. The results depicted in Figure 4 show that the computed properties for radii up to 6-8 Å are well reproducible. Beyond that the components diverge with no trends with respect to box size, but systematic drifts remain. This indicates that the drifts cannot be explained by finite-size effects. Instead, their origin can be linked to uncertainties in the values of the reference bulk parameters. Figure 5 depicts the sensitivity of the GCT results to variations of one standard error in the computed reference parameters. As the number of water molecules increases with the cube of the radial distance to the solute, any small systematic error in the reference properties of a bulk water molecule will eventually cause large variations in the computed enthalpy/entropy components. The water-water enthalpy is particularly sensitive, which corroborates the results depicted in Figure 3. With an 8 Å radius there is an uncertainty of ca. 1 kcal.mol^{-1} in ΔH_w, whereas the variability of the entropy components is only about 0.2 kcal.mol^{-1} with a 10 Å radius. Thus, if a large volume of space **s** is monitored, systematic errors should be expected for computed absolute enthalpies and entropies of hydration but less so for relative enthalpies and entropies of hydration evaluated over the same volumes where this effect largely cancels out. As an additional test, the sensitivity of the computed properties to the grid density was assessed by performing analyses of the sodium simulations using a grid spacing of 0.5 Å or 2 Å. Little dependence on the grid spacing was observed (Figure S2).Thus, for Nautilus analyses selection of an adequate voxel size should be primarily dictated by the desired trade-off between spatial resolution and trajectory size. For studies of biomolecular systems, it appears reasonable to expect that a grid spacing of 0.5 or 1 Å

will be adequate, with a spacing of 0.5 Å each voxel covers a volume of 0.125 $Å^3$, which approximately amounts to $1/80^{th}$ of the volume of a water molecule.

The present results indicate that simulations on the order of 50 ns enable reasonably well reproducible predictions of relative enthalpies and entropies of hydration by considering spherical regions centered on the solute of radius ca. 8 – 10 Å. If greater volume must be considered longer simulations should be performed. Note that the sampling errors grow with the number of water molecules considered rather than the volume monitored. A sphere of 10 Å radius centered on sodium includes about 140 water molecules which is much greater than the number of water molecules within a typical protein binding site. Thus converged analyses of water properties within a typical protein binding site do not appear unpractical.

Inspection of the GCT enthalpy and entropy components in Figure 3 provides insights into hydration of the monoatomic solutes. The enthalpy of hydration of xenon (Fig 3B) is more negative than neon (Fig 3A) because the former exhibit stronger Lennard-Jones interactions with water, which is only partly offset by a loss in water interactions. Water in the vicinity of the two solutes show negligible changes in vibrational and librational entropies, but there is a significant difference in orientational entropy because the larger, more hydrophobic xenon atom reduces more hydrogen-bonding arrangements of first and second shell water molecules. Chloride (Fig 3C) has a slightly less negative solute-water enthalpy component than sodium (Fig 3D) but also a less negative positive water-water enthalpy component, so overall the enthalpy of hydration of sodium is more negative than chloride. For both ions the water orientational entropy component differ the most for first-shell water molecules because sodium cannot accept hydrogen-bonds, but the changes in orientational entropy over larger volumes converge to similar values for the two ions. The water vibrational entropy component is higher for sodium than chloride, whereas

the librational entropy component is small in both cases. The difference in vibrational entropy likely reflects the stronger interactions of water with sodium for this force field as evidenced by the enthalpy components in Fig 3C/3D.

Next, Nautilus analyses were performed for an additional set of 8 neutral small molecules of varying polarity. To maintain a reasonable statistical error and obtain reproducible results, water molecules within a 10 Å radius of the center of the solutes were considered to compute water enthalpy and entropy components. The full solute-water enthalpy term was used, however, as it is well converged and reproducible. In addition, solute translational and rotational entropies were evaluated using Eqs 4 and 5. FDTI was also used to compute the hydration free energy of this set of small molecules using the same software and energy function to minimize variations due to differences in protocols. Figure 6A shows the correlation between free energies of hydration computed with FDTI and experimental data for the set of neutral molecules. The results, with a correlation coefficient of 0.99 and mean unsigned error of 0.98 kcal.mol^{-1}, are similar to those reported in the literature using similar force fields and methodologies.[63-64] Note that the two ions, chloride and sodium, are not plotted in Fig 6A. Their computed free energies of hydration are –63.3±0.1 and –63.7±0.1 kcal.mol^{-1} respectively, which differ considerably from the experimental data of Schmid et al.[65]. This is expected, as the results here have not been corrected for systematic errors due to the use of a periodic-boundary conditions and a reaction-field treatment of long-ranged electrostatic interactions.

Figure 6B shows the correlation between the free energies of hydration computed with the two methodologies. The GCT results are strongly correlated with the FDTI results, with a correlation coefficient of 0.97 and a mean unsigned error of 0.92 kcal.mol^{-1}. No systematic drifts are apparent. The statistical error of the GCT calculations is higher than the FDTI predictions, as

evidenced by their larger error bars in Fig 6B. The reasons for this difference is that FDTI directly yields free energies of hydration by evaluation of solute-solvent potential energy ensemble averages, whereas GCT free energies are obtained by summing enthalpies and entropies of hydration that each depend on slowly converging water-water energy terms. Note that it is possible to evaluate entropies and enthalpies of hydration by thermodynamic integration, but the required ensemble averages also contain noisy water-water terms of opposite magnitude that exactly compensate when the free energy of hydration is evaluated.[8]

Fig 7A shows the correlation of the GCT computed free energies of hydration with experimental data for the neutral molecules. The correlation coefficient is 0.98 and the mean unsigned error is 0.82 kcal.mol^{-1}. These results are slightly more accurate than the FDTI predictions, although the differences are likely not significant given the size of the dataset. As noted previously, the statistical error is larger, and FDTI appears better suited for the evaluation of free energies of hydration for this dataset. However, the Nautilus analyses also provide enthalpies and entropies of hydration, which are plotted against experimental data in Figure 7B and 7C. Some compensation in systematic errors is apparent because the mean unsigned error for both quantities is ca. 1.3 kcal.mol^{-1}, higher than for the free energies of hydration. The correlation between enthalpies of hydration and experiment is high (R^2 = 0.98), but noticeably lower for the entropies of hydration (R^2 = 0.66). The decreased correlation with entropies of hydration is partly due to the smaller energetic range of the measured entropies (ca. 7 kcal.mol^{-1} versus 18 kcal.mol^{-1} for the enthalpies), but some outliers are apparent. The accuracy of the results can be compared to a recent GIST study of Huggins and Payne on the hydration thermodynamics of 6 small molecules. Although the dataset differs and contained fewer small molecules, similar agreement with enthalpies of hydration were reported by Huggins and

Payne,[22] whereas the GIST entropies of hydration were better correlated with experiment (R^2 0.77).

Table 2 compares calculated and measured enthalpies, entropies and free energies of hydration for the dataset of neutral and charged solutes. The free energies of hydration of neon and xenon are slightly underestimated, and greater discrepancies are noted for the enthalpies and entropies of hydration. This is in agreement with previous observations and suggests that the temperature dependence of the hydration of non-polar solutes is not well captured with the present force field.[29] The effect is, however, systematic and the relative free energies, entropies and enthalpies of hydration are in good agreement with experiment. Assessment of the computed properties for chloride and sodium is more challenging because finite-size and cutoff errors are large for charged compounds. This issue has been explored in great details by Hunenberger and co-workers.[66-69] Correction terms are available for hydration free energies computed using alchemical methods, but more work is needed to derive suitable correction terms for enthalpies and entropies of hydration computed using GCT. With the present protocol, sodium is better hydrated than chloride by about 4 kcal mol^{-1}. This is not in agreement with the experimental data of Schmid et al. listed in the table,[65] although there is considerable uncertainty in the experimental data owing to the difficulty of measuring the hydration thermodynamics of the proton.[68]

As for the other non-polar solutes ethane, isobutane and n-butane, their free energies of hydration appear to be systematically slightly underestimated. The magnitudes of the enthalpies/entropies are also underestimated, suggesting again that the force field does not capture well the temperature dependence of hydrophobic hydration. Additionally for isobutane, the largest outlier in the neutral molecules dataset, it is likely that the solute rotational entropy in

solution is overestimated as it interacts weakly with water and, because of low computed torques, Eq 5 yields a null change in librational entropy.

The enthalpies and entropies of hydration of benzene are well reproduced by GCT. The entropy of hydration of methanol is in good agreement with experiment, but the enthalpy is underestimated and overall methanol is not sufficiently well hydrated. This is likely an issue with the AM1-BCC GAFF forcefield parameters used in the present study. The free energy of hydration of methanol computed with this forcefield and using different alchemical protocols has been found to be too positive by ca. 1-1.5 kcal.mol^{-1}.[63-64] This discrepancy with experimental data could be addressed by further polarizing the –OH bond. Repeating the calculations after increasing the charge on the hydroxyl hydrogen atom by 0.05 e and decreasing the charge on the hydroxyl oxygen atom by 0.05 e yields $\Delta H^{s,0}_{X(\boldsymbol{r})+w}$ = -10.1 kcal.mol^{-1}, $-T\Delta S^{s,0}_{X(\boldsymbol{r})+w}$= 5.3 kcal.mol^{-1}, and $\Delta G^{s,0}_{X(\boldsymbol{r})+w}$= -4.8 kcal.mol^{-1}, all in excellent agreement with experiment. Similarly, the entropy of hydration of the more polar solutes, acetamide and N-methylacetamide, is well reproduced by GCT. The free energy of hydration is higher than experiment by 0.7-1.5 kcal/mol-1 because the enthalpies of hydration are not sufficiently negative. As for methanol, further polarization of the C-O or N-H bonds could address this discrepancy.

Overall, the GCT predicted enthalpies and entropies of hydration are reasonably well correlated with experimental data, although shortcomings in the force field used or the methodology itself are apparent, suggesting there is room for further improvement of the methodology. However, a major advantage of GCT is arguably in the ability to visualize the contribution of different solvent regions to the thermodynamics of hydration. Spatial decomposition enables the identification of solvent regions that contribute favourably/unfavourably to the free energy of hydration. Additionally, the breakdown into

enthalpic and entropic components provides insights into the nature of the intermolecular interactions of water molecules. Figure 8 depicts water's components to the free energy, enthalpy and entropy of hydration of three small molecules, N-methylacetamide (NMA), methanol and benzene. To finely resolve small energetic differences between neighboring voxels 5,000,000 snapshots were collected from ten independent 50 ns simulations to produce grid cell files with cubic voxels of edge length 0.5 Å. The resulting contours are generally smooth, except when drawn at values close to zero -or one for the relative density- due to sampling errors. Additionally, harmonic restraints were applied to every solute atom to avoid rotations of methyl hydrogen atoms, or the polar hydrogen atom of methanol, which would blur the properties of nearby water molecules. For these small solutes the resulting computed properties are similar to within statistical error to the results reported in Table 2.

For NMA (Fig 8A), the region that contributes the most to the free energy of hydration is located in a hemisphere around the amide oxygen atom. Water molecules in this region donate a hydrogen-bond to the amide oxygen atom. Interestingly, configurations with a linear angle between a water molecule oxygen atom, the amide oxygen atom, and the amide carbon atom, are slightly less favored. A smaller but significant contribution arises from a separate region where water molecules can accept a hydrogen-bond from the amide polar hydrogen. A third weakly stabilizing region is apparent in the vicinity of the methyl group bonded to the amide nitrogen atom. Water molecules in this region can favorably orient hydrogen atoms near the amide oxygen and nitrogen atoms, whilst minimizing electrostatic repulsion with the amide polar hydrogen. Regions located above and below the amide bond plane contribute unfavorably to the free energy of hydration because of unfavorable enthalpic and entropic contributions. Similar unstable regions have been identified with a GIST analysis by Huggins and Payne.[22] Water

structuring around high density regions is also seen to cause depletions in water density further away from the solute, particularly near the region where water molecules can accept a hydrogen-bond from the solute. Water in high-density regions contributes the most to solvent entropy loss. Lesser entropy loss is observed near the solute methyl groups, and small gains in entropy are observed in the second hydration shell for water molecules located close to the first shell water molecules that are accepting or donating hydrogen bonds to the solute.

For methanol (Fig 8B) two distinct regions that contribute favorably to the free energy of hydration are apparent, covering positions where water molecules can donate or accept hydrogen bonds. The water hydrogen-bond donating region covers a larger volume than the water hydrogen bond accepting region because methanol accepts more hydrogen bonds than it donates, but stronger enthalpic interactions are observed in the smaller hydrogen-bond accepting region. This is partly compensated by greater vibrational and orientational entropy loss within the water hydrogen-bond accepting region. As for NMA, both regions contribute favorably to the enthalpy of hydration and unfavorably to the entropy of hydration, but the isocontours drawn at the same isovalues are smaller. Small favorable entropic contributions and unfavorable enthalpic contributions in the second shell are apparent. The situation differs for regions near the methyl group and perpendicular to the C-O-H plane where enthalpy and entropy both contribute unfavorably to the hydration free energy. The region immediately above the hydroxyl group has a small favorable entropic component and an unfavorable enthalpic component. All together this contributes largely to the approximately 5 kcal.mol^{-1} more negative hydration free energy of NMA over methanol. As for NMA, the location of high water density regions ($\rho(\mathbf{s}) > 2.7$) correlate well with regions that contribute favorably to the free energy of hydration. A complete match is not fully apparent in Fig 8 because the contours have been drawn at different values.

High local water density is not necessarily associated with a favorable contribution to the free energy of hydration, for instance regions located near the methyl group of of methanol show significant density deviations from bulk ($\rho(\mathbf{s}) > 1.5$) but contribute unfavorably to the free energy of hydration (Figure S3).

For benzene (Fig 8C) it is apparent that the favorable free energy of hydration arises from two small regions above and below the π-cloud. Water molecules in these regions are weakly donating hydrogen-bonds to the solute. In addition, two secondary doughnut shaped regions provide additional stabilizing enthalpic contributions. Water molecules in these regions tend to donate hydrogen-bonds to the water molecule(s) interacting with the π-cloud. Water in the plane of the ring contributes unfavorably to hydration, with greater loss in free energy in regions between two neighboring solute hydrogen atoms. This arises because of an unfavorable enthalpic contribution, as well as a greater orientational entropy loss because water in these regions has fewer hydrogen bond acceptors within its coordination shell. Second hydration shell water molecules located above and below the plane of the benzene ring contribute slightly unfavorably and favorably to the enthalpy and entropy of hydration respectively, with the entropy gain due to a favorable orientational entropy component. The effect largely cancels out and there is no significant contribution to the free energy of hydration from this region.

Overall, the results depicted in Fig 8 suggest that GCT provides a rich graphical description of hydration thermodynamics that yield insights into the location and nature of stabilizing/destabilizing solvent interactions.

4. Conclusions

Overall, grid cell theory shows promising potential for molecular modeling studies of the hydration of organic and biomolecules. If free energies of hydration are of sole interest, alternative methodologies such as thermodynamic integration appear more competitive owing to the slower convergence of the water-water terms in GCT. However a Nautilus analysis yields enthalpies and entropies of hydration, and makes it possible to visualize water's components for graphical analyses of hydration thermodynamics. For quantitative analyses, relative enthalpies and entropies of hydration appear easier to predict as uncertainties in the values of the reference bulk parameters will have less impact on the computed properties.

The most similar alternative to GCT is the GIST method proposed by Nguyen et al.[18] Both approaches should provide similar enthalpies of hydration because they each rely on averaged interaction energies. The most significant differences are found in the evaluation of entropies. GIST relies on an entropy expansion with multi-particles correlation functions truncated at first or second order for computational tractability.[18] Errors arise because of the neglect of entropic contributions from higher-order correlations. Cell theory implicitly captures higher-order correlations in the cell parameters, but makes a harmonic approximation to describe vibrational and librational entropies, whereas orientational entropy is captured with a generalized Pauling's residual entropy model. Both approaches appear to give results of comparable accuracy for hydration free energies. However GCT may have a number of practical advantages: the methodology appears less sensitive to the resolution of the grid and the solvent entropic components converge more rapidly than the enthalpic components. This may be a consequence of the functional form of the cell theory equations, i.e. entropies are derived from logarithms of ratios, whereas enthalpies from differences of interaction energies. This should facilitate routine applications owing to the decreased number of snapshots that must be post-processed after a MD

simulation. The computing time needed to perform a Nautilus analysis is function of the size of the system simulated and the grid region. To illustrate, a typical analysis of 50,000 snapshots of a solvated small molecule in a cubic volume of 21,952 $Å^3$ currently requires ca. 40 CPU hours with the present implementation of the trajectory analysis software Nautilus. These can be trivially parallelized over hundreds of processors by processing trajectory snapshots concurrently. Further optimization is possible by rewriting the software in a low-level programming language. Alternatively, grid properties could be computed "on the fly" during a MD simulation.

Further work is desirable to assess the robustness and accuracy of GCT predictions with other solute/solvent force fields combinations. Huggins has shown that IFST predictions of the free energy of hydration sites can vary by up to 4 $kcal.mol^{-1}$ between TIP4P-2005 and TIP5P-Ewald.[70] Rigorous comparison of GCT and IFST by analysis of the same computed trajectories with identical energy functions on larger datasets should also prove instructive to further assess the merits of each approach and devise improved approaches to estimate solvent entropies. Henchman and Cochram have recently reported a detailed analysis of different coordination environments for bulk water.[8] Their results suggest that more detailed consideration of distributions of hydrogen-bonding patterns may lead to improve estimations of solvent orientational entropies. The present results have been obtained for solutes restrained in a given conformation. In principle GCT analyses can be performed on flexible solutes, but in cases where significant flexibility is observed, the simple grid averaging procedure used here may not yield components that can be easily visualized. If different conformational states have already been established, one could always perform separate analyses on conformation-specific grids.[71] Alternatively, protocols that combine dynamical updating of the grid properties with a sliding

time-window may prove useful. Additional research is desirable to explore efficient protocols to resolve coupling between solvent and solute degrees of freedom.

Overall the major strength of the approach arguably lies in its ability to spatially resolve enthalpy and entropy into physically intuitive components, either as a stand-alone analysis or, for instance, to guide the interpretation of free energy changes computed by alchemical methods. The present results warrant further application of Grid Cell Theory to host/guest recognition problems and to solvent mapping studies of organic and biomolecular systems.

ASSOCIATED CONTENT

Supporting Information. Convergence of the entropy components for simulations of bulk water. Effect of grid spacing on the computed free energy of hydration of sodium. Additional graphical analysis of ΔG_w^s and $\rho(\boldsymbol{s})$ for methanol. Table of the free energies of hydration for the small molecule dataset computed with the FDTI protocol. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* E-mail: mail@julienmichel.net

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

JM is supported by a University Research Fellowship from the Royal Society. RHH is supported by BBSRC grant BB/K001558/1. This research was also supported by EPSRC through award of a CASE studentship to GG, and by a Royal Society research equipment grant (RG110450).

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

Gratitude is expressed to Dr Christopher Woods for assistance with scientific computations and for his continuing efforts to develop the software Sire. The authors would like to thank Dr David Huggins for helpful comments on an earlier version of this manuscript.

REFERENCES

1. van Gunsteren, W.; Berendsen, H. *Angew. Chem. Int. Ed.* **1990,** *29* (9), 992-1023.
2. Ball, P. *Nature* **2008,** *452* (7185), 291-292.
3. Stirnemann, G.; Wernersson, E.; Jungwirth, P.; Laage, D. *J. Am. Chem. Soc.* **2013,** *135* (32), 11824-11831.
4. Kyakuno, H.; Matsuda, K.; Yahiro, H.; Inami, Y.; Fukuoka, T.; Miyata, Y.; Yanagi, K.; Maniwa, Y.; Kataura, H.; Saito, T.; Yumura, M.; Iijima, S. *J. Chem. Phys.* **2011,** *134* (24).
5. Irudayam, S. J.; Henchman, R. H. *J. Chem. Phys.* **2012,** *137* (3).
6. Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006,** *110* (35), 17616-17626.
7. Rogers, K.; Ortiz-Sánchez, J.; Baron, R.; Fajer, M.; de Oliveira, C.; McCammon, A. *J. Chem. Theory Comput.* **2012,** *9* (1), 46-53.
8. Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. *J. Chem. Phys.* **2004,** *120* (6), 2652-2661.
9. Kubo, M. M.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **1997,** *101* (49), 10527-10534.
10. Fenley, A. T.; Muddana, H. S.; Gilson, M. K. *Proc. Natl. Acad. Sci. USA* **2012,** *109* (49), 20006-20011.
11. Chodera, J. D.; Mobley, D. L., Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. In *Annual Review of Biophysics, Vol 42*, Dill, K. A., Ed. 2013; Vol. 42, pp 121-142.
12. Baron, R.; Molinero, V. *J. Chem. Theory Comput.* **2012,** *8* (10), 3696-3704.
13. Michel, J.; Foloppe, N.; Essex, J. W. *Mol. Inf.* **2010,** *29* (8-9), 570-578.
14. Schatz, G. C. *Proc. Natl. Acad. Sci. USA* **2007,** *104* (17), 6885-6892.
15. Huber, R.; Fuchs, J.; von Grafenstein, S.; Laner, M.; Wallnoefer, H.; Abdelkader, N.; Kroemer, R.; Liedl, K. *J. Phys. Chem. B* **2013,** *117* (21), 6466-6472.

16. Olano, R.; Rick, S. *J. Am. Chem. Soc.* **2004,** *126* (25), 7991-8000.
17. Yu, H.; Rick, S. *J. Phys. Chem. B* **2010,** *114* (35), 11552-11560.
18. Nguyen, C.; Young, T. K.; Gilson, M. *J. Chem. Phys.* **2012,** *137* (4), 044101.
19. Lazaridis, T. *J. Phys. Chem. B* **1998,** *102* (18), 3531-3541.
20. Lazaridis, T. *J. Phys. Chem. B* **1998,** *102* (18), 3542-3550.
21. Lazaridis, T. *J. Phys. Chem. B* **2000,** *104*, 4964-4979.
22. Huggins, D.; Payne, M. *J. Phys. Chem. B* **2013,** *117* (27), 8232-8244.
23. Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. *Proteins: Struct. Funct. Bioinform.* **2012,** *80* (3), 871-883.
24. Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2004,** *109* (1), 662-670.
25. Raman, P.; Mackerell Jr., A. D. *J. Chem. Phys.* **2013,** *139* (5), 055105.
26. Irudayam, S.; Henchman, R. *J. Phys. Chem. B* **2009,** *113* (17), 5871-5884.
27. Henchman, R. *J. Chem. Phys.* **2007,** *126* (6), 064504.
28. Irudayam, S.; Plumb, R.; Henchman, R. *Faraday Discuss.* **2010,** *145* (0), 467-485.
29. Irudayam, S. J.; Henchman, R. H. *J. Phys. Condens-Mat* **2010,** *22* (28).
30. Irudayam, S. J.; Henchman, R. H. *Mol. Phys.* **2011,** *109* (1), 37-48.
31. Pauling, L. *J. Am. Chem. Soc.* **1935,** *57* (12), 2680-2684.
32. Mark, A. E.; Vangunsteren, W. F. *J. Mol. Biol.* **1994,** *240* (2), 167-176.
33. Boresch, S.; Karplus, M. *J. Mol. Biol.* **1995,** *254* (5), 801-807.
34. Brady, G. P.; Sharp, K. A. *J. Mol. Biol.* **1995,** *254* (1), 77-85.
35. Huggins, D. J. *J. Chem. Phys.* **2012,** *136* (6).
36. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004,** *120* (20), 9665-9678.
37. Bondi, A. *J. Phys. Chem.* **1964,** *68* (3), 441-&.
38. Guillot, B.; Guissani, Y. *J. Chem. Phys.* **1993,** *99* (10), 8075-8094.
39. Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008,** *112* (30), 9020-9041.
40. Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
41. Jakalian, A.; Jack, D.; Bayly, C. *J. Comput. Chem.* **2002,** *23* (16), 1623-1641.
42. Case, D.; Darden, T. A.; Cheatham, T. E.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P., *Amber 11*. 2010.
43. Uberuaga, B.; Anghel, M.; Voter, A. *J. Chem. Phys.* **2004,** *120* (14), 6363-6374.
44. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
45. Miyamoto, S.; Kollman, P. *J. Comput. Chem.* **1992,** *13* (8), 952-962.
46. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977,** *23* (3), 327-341.
47. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
48. Sagui, C.; Darden, T.; Simulation and Theory of Electrostatic Interactions in Solution: Computational Chemistry, Biophysics and Aqueous Solutions, first edition; AIP Conference Proceedings (Book 492); American Institute of Physics: Woodbury, NY, 1999; p 64-
49. Woods, C. J.; Michel, J. *Sire Molecular Simulation Framework*, Revision 1786; UK, 2013

50. Eastman, P.; Friedrichs, M.; Chodera, J.; Radmer, R.; Bruns, C.; Ku, J.; Beauchamp, K.; Lane, T.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M.; Pande, V. *J. Chem. Theory Comput.* **2012,** *9* (1), 461-469.
51. Tironi, I.; Sperb, R.; Smith, P.; van Gunsteren, W. *J. Chem. Phys.* **1995,** *102* (13), 5451-5459.
52. Andersen, H. *J. Chem. Phys.* **1980,** *72* (4), 2384-2393.
53. Oliphant, T. *Comput. Sci. Eng.* **2007,** *9* (3), 10-20.
54. Michaud-Agrawal, N.; Denning, E.; Woolf, T.; Beckstein, O. *J. Comput. Chem.* **2011,** *32* (10), 2319-2327.
55. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996,** *14* (1).
56. Jorgensen, W.; Ravimohan, C. *J. Chem. Phys.* **1985,** *83* (6), 3050-3054.
57. Zwanzig, R. *J. Chem. Phys.* **1954,** *22* (8), 1420-1426.
58. Shyu, C.; Ytreberg, M. *J. Comput. Chem.* **2009,** *30* (14), 2297-2304.
59. Beutler, T.; Mark, A.; van Schaik, R.; Gerber, P.; van Gunsteren, W. *Chem. Phys. Lett.* **1994,** *222* (6), 529-539.
60. Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994,** *100* (12), 9025-9031.
61. Michel, J.; Verdonk, M.; Essex, J. *J. Chem. Theory Comput.* **2007,** *3* (5), 1645-1655.
62. Shelton, D. P. *J. Chem. Phys.* **2012,** *136* (4).
63. Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007,** *111* (9), 2242-2254.
64. Michel, J.; Orsi, M.; Essex, J. W. *J. Phys. Chem. B* **2008,** *112* (3), 657-660.
65. Schmid, R.; Miah, A.; Sapunov, V. *PCCP* **2000,** *2* (1), 97-102.
66. Kastenholz, M. A.; Hunenberger, P. H. *J. Chem. Phys.* **2006,** *124* (22).
67. Kastenholz, M. A.; Hunenberger, P. H. *J. Chem. Phys.* **2006,** *124* (12).
68. Reif, M. M.; Hunenberger, P. H. *J. Chem. Phys.* **2011,** *134* (14).
69. Reif, M. M.; Hunenbergera, P. H. *J. Chem. Phys.* **2011,** *134* (14).
70. Huggins, D. *PCCP* **2012,** *14* (43), 15106-15117.
71. Henchman, R. H.; McCammon, J. A. *J. Comput. Chem.* **2002,** *23* (9), 861-869.
72. Wagner, W.; Pruss, A. *J. Phys. Chem. Ref. Data* **2002,** *31* (2), 387-535.
73. Naim, B.; Marcus, Y. *J. Chem. Phys.* **1984,** *81* (4), 2016-2027.
74. Gatta; Barone, G.; Elia, V. *J. Solution Chem.* **1986,** *15* (2), 157-167.
75. Wolfenden, R. *Biochemistry* **1978,** *17* (1), 201-204.
76. Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981,** *10* (8), 563-595.

Table 1. The reference bulk parameters and computed thermodynamic values of liquid water at 298 K.

Parameters and properties	TIP4P-Ew	Experiment[a]
$U_{w(l)}^{inter}$(kcal mol^{-1})	-11.025(9)	-
$F_{w(l)}^{1}$(10^{-10} N)	1.587(1)	-
$F_{w(l)}^{2}$(10^{-10} N)	1.735(1)	-
$F_{w(l)}^{3}$(10^{-10} N)	1.334(1)	-
$\tau_{w(l)}^{1}$(10^{-20} N m^{-1})	1.061(1)	-
$\tau_{w(l)}^{2}$(10^{-20} N m^{-1})	1.194(1)	-
$\tau_{w(l)}^{3}$(10^{-20} N m^{-1})	1.453(1)	-
$\Omega_{w(l)}^{ori}$	3.305(4)	-
N_{w}^{bulk}	4.883(2)	-
ρ_b (kg m^{-3})	995.4(2)	997.1
ΔH (kcal mol^{-1})[b]	-9.98(1)	-9.92
ΔS (cal K^{-1} mol^{-1})[c]	-15.42(5)	-14.05

a) Ref[72] for ρb, ΔH, and ΔS b) Includes a dielectric depolarisation correction term of 1.044 kcal mol^{-1}.[36] c) Computed using Eq 7 from ref [27]. The dash (-) signifies not available.

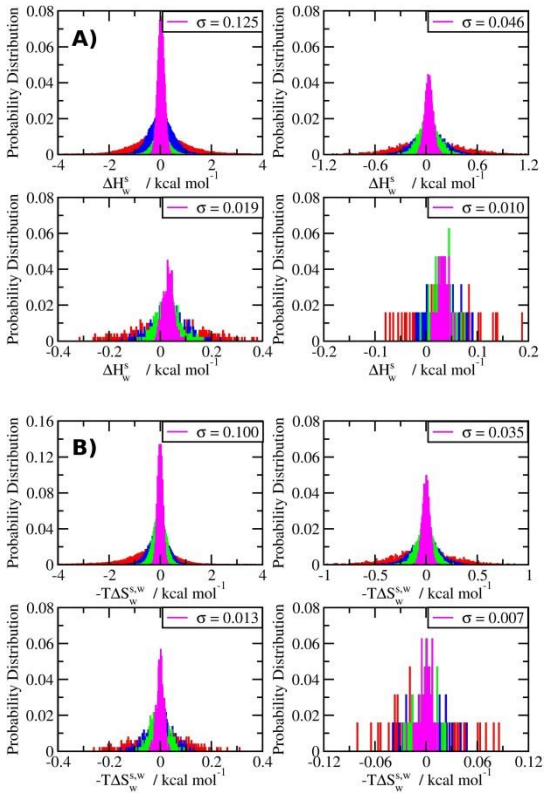

Figure 1. The distribution of A) water enthalpies ΔH^{s}_{w} and B) water entropies $-T\Delta S^{s,w}_{w}$ in bulk water. Each distribution was computed by dividing a cubic volume 4096 $Å^3$ centered at the center of a box of 804 TIP4P-Ew molecules into evenly distributed regions of space **s** covering each 0.125 $Å^3$ (top left), 1 $Å^3$ (top right), 8 $Å^3$ (bottom left) and 64 $Å^3$ (bottom right) respectively. The red, blue, green and magenta color indicate distributions computed from a simulation of 2 ns, 5 ns, 10 ns and 50 ns duration respectively. The first ns was discarded and snapshots were analysed every 1 ps. The legend indicates the estimated standard deviation for each distribution using the full dataset.

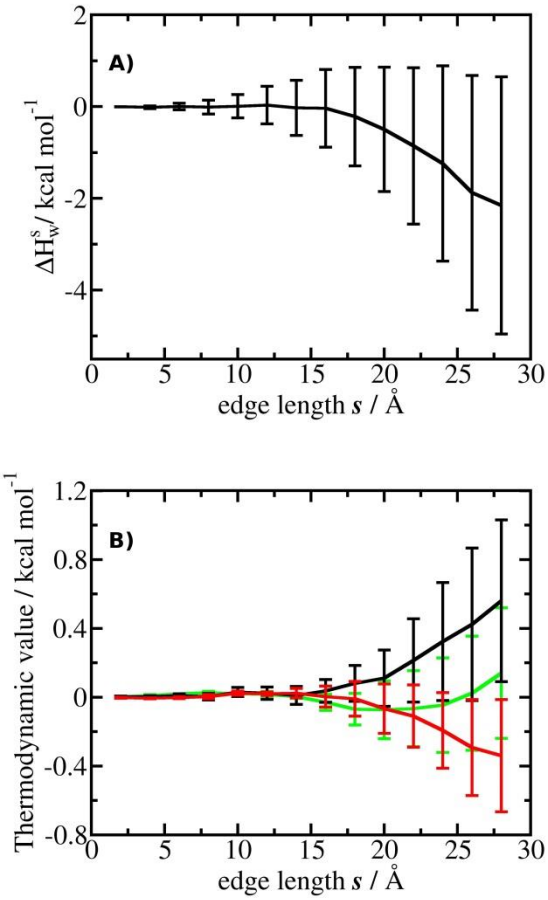

Figure 2. The convergence of the water A) enthalpies ΔH^{s}_{w} and B) entropy components red $-T\Delta S^{s,or}_{w,X(r)}$ (red), $-T\Delta S^{s,vib}_{w,X(r)}$ (black), $-T\Delta S^{s,lib}_{w,X(r)}$ (green) in bulk water as a function of the size of the monitored region **s**. Each data point is the mean of three 50 ns simulations and the error bars show the standard error of the mean.

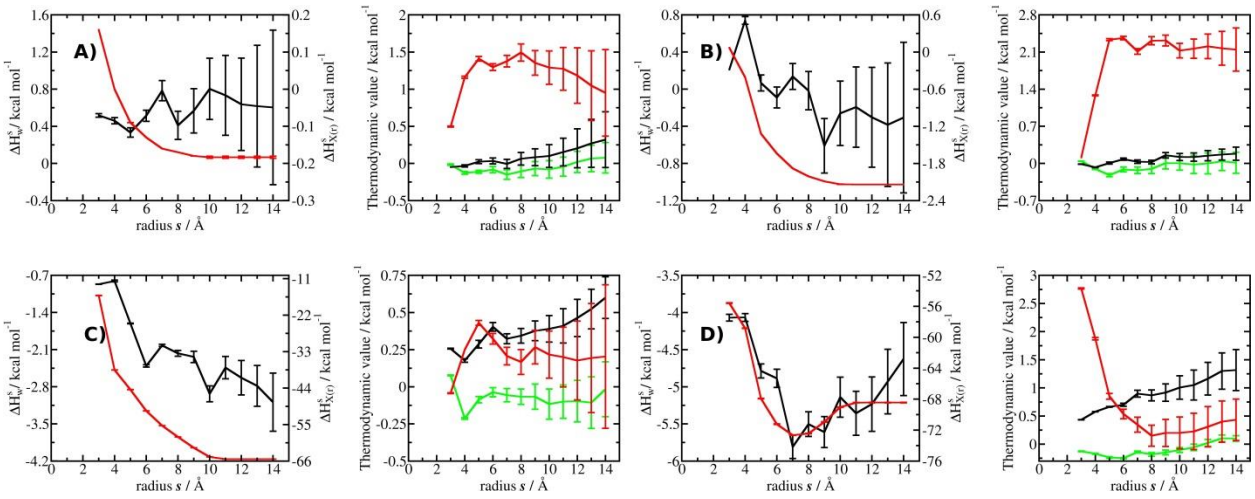

Figure 3. The convergence of hydration enthalpy and entropy components as a function of the size of **s** for (A) Neon, (B) Xenon, (C) Chloride, (D) Sodium. All solutes were solvated in a waterbox of 804 water molecules. The x-axis depicts the radius of a spherical region **s** centered on the solute. The y-axis indicates components of the enthalpy and entropy of hydration. Left panel: black ($\Delta H_w^{\boldsymbol{s}}$), red ($\Delta H_{X(r)}^{\boldsymbol{s}}$).Right panel: red ($\Delta S_{w,X(\boldsymbol{r})}^{\boldsymbol{s},or}$), black ($\Delta S_{w,X(\boldsymbol{r})}^{\boldsymbol{s},vib}$), green ($\Delta S_{w,X(\boldsymbol{r})}^{\boldsymbol{s},lib}$). The error bars indicate the standard error of the mean obtained from three independent simulations.

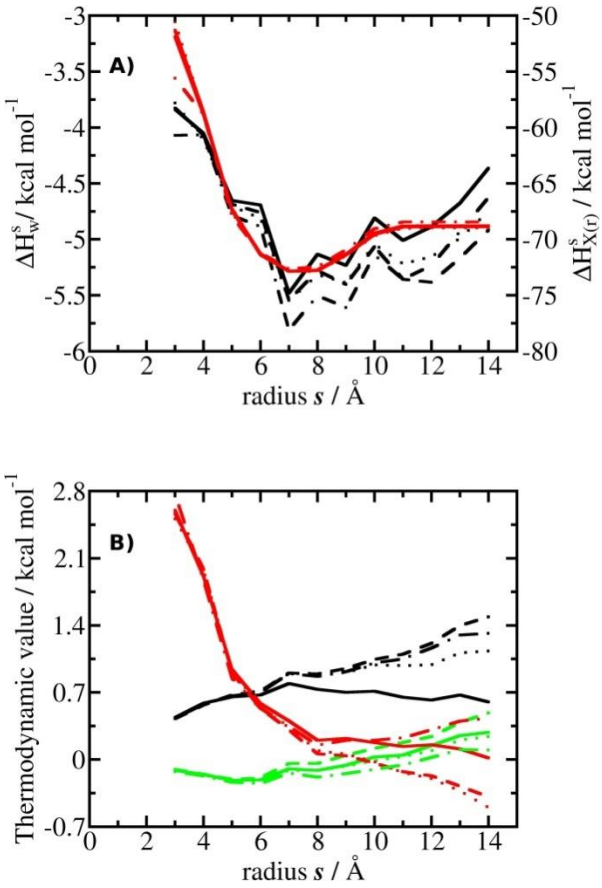

Figure 4. The effect of box-size on the enthalpy and entropy components of the free energy of hydration of sodium. The solid; dashed and dotted; dashed; and dotted lines depict results for a box with approximate average half-edge lengths of 15 Å, 17.5 Å, 20 Å and 25 Å respectively. Error bars, which are comparable to those seen in Figure 4 have been omitted for clarity. Other symbols are as in Figure 3.

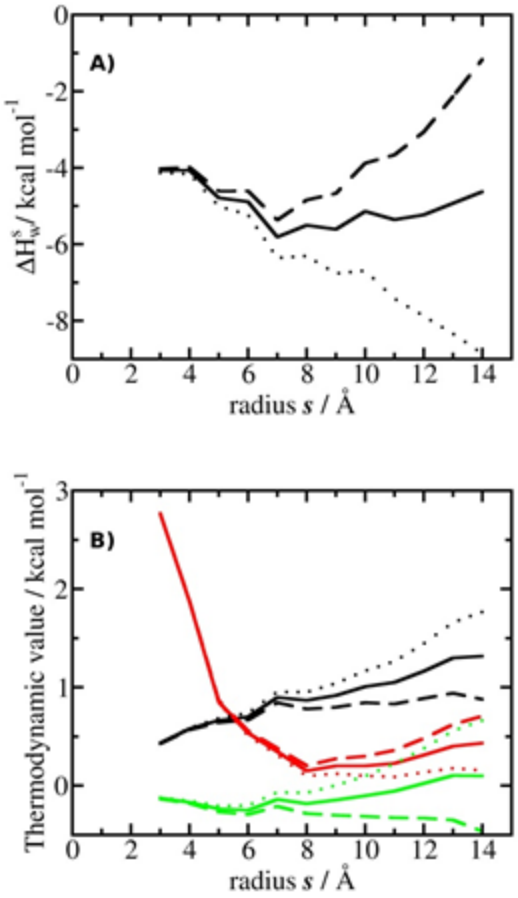

Figure 5. The effect of uncertainties in the reference bulk parameters on the enthalpy and entropy components of the free energy of hydration of sodium. The solid lines depict results obtained using the parameters listed in Table 1, and the dashed and dotted lines with parameters modified by ± 1 standard error respectively. Error bars, which are comparable to those seen in Figure 3 have been omitted for clarity. Other symbols are as in Figure 3.

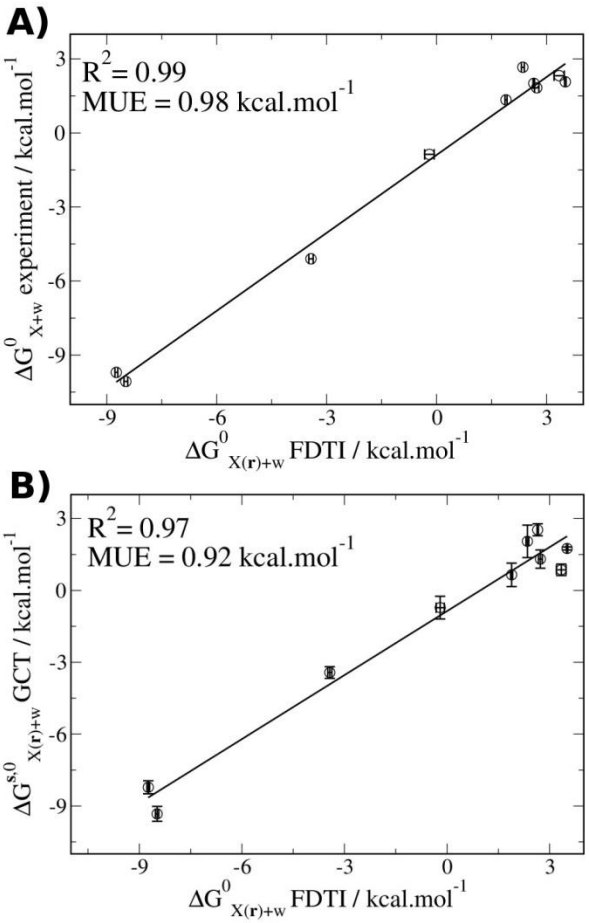

Figure 6. The accuracy of the thermodynamic integration predictions and correlation with Grid Cell Theory results. A) Correlation between Finite Difference Thermodynamic integration free energies and experimental data. B) Correlation between Grid Cell Theory and Finite Difference Thermodynamic integration free energies. The error bars indicate the standard error of the mean obtained from three independent simulations.

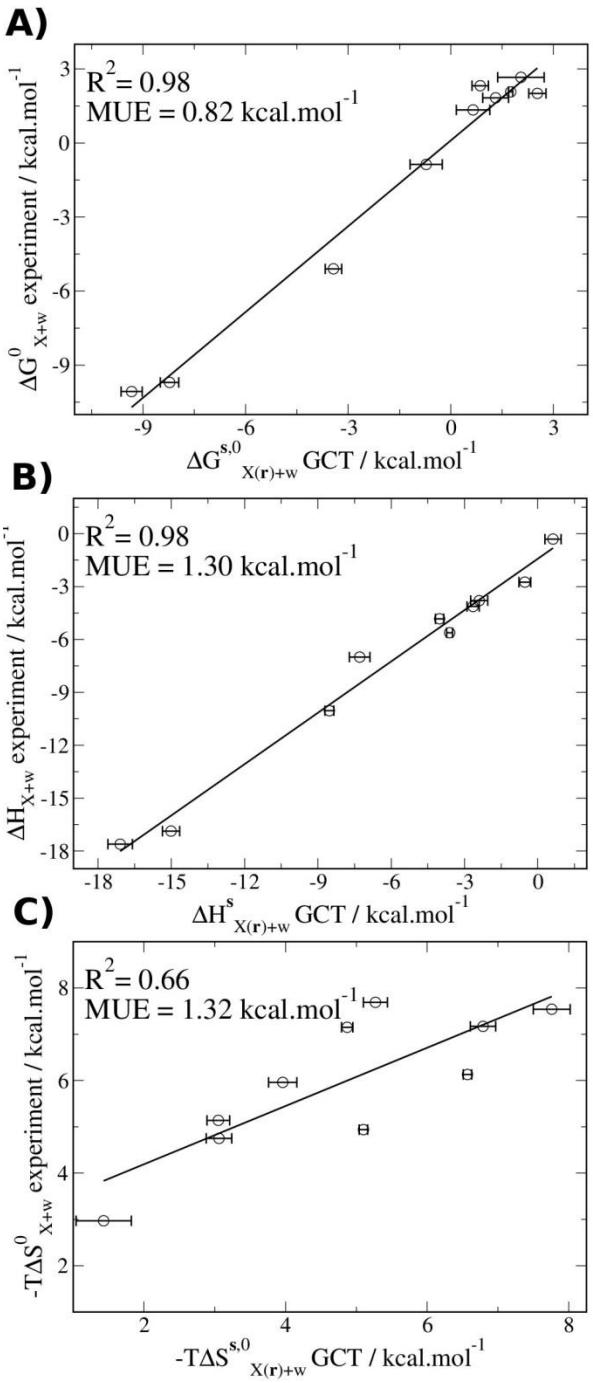

Figure 7. The correlation of the Grid Cell Theory computed hydration thermodynamics with experimental data. A) Correlation between computed free energies of hydration and experimental data. B) Correlation between computed enthalpies of hydration and experimental data. C) Correlation between computed entropies of hydration and experimental data. The error bars indicate the standard error of the mean obtained from three independent simulations.

Table 2. The computed and measured hydration enthalpies, entropies and free energies of a set of small molecules.[a]

	Computed			Experimental[b]		
	$\Delta H^{s}_{X(r)+w}$	$-T\Delta S^{s,\circ}_{X(r)+w}$	$\Delta G^{s,\circ}_{X(r)+w}$	ΔH_{X+w}	$-T\Delta S^{\circ}_{X+w}$	ΔG°_{X+w}
Neon	0.62(33)	1.43(39)	2.05(68)	-0.31	2.97	2.66
Xenon	-2.40(35)	3.05(16)	0.65(49)	-3.80	5.14	1.34
Chloride	-68.36(15)	3.10(17)	-65.26(04)	-93.69	4.59	-89.10
Sodium	-73.61(29)	4.27(40)	-69.34(31)	-93.45	4.86	-88.59
Methane	-0.53(24)	3.06(18)	2.53(25)	-2.75	4.75	2.01
Ethane	-2.65(25)	3.96(20)	1.31(38)	-4.13	5.96	1.83
Isobutane	-4.02(17)	4.87(08)	0.86(24)	-4.83	7.15	2.32
N-butane	-3.61(12)	5.27(17)	1.75(06)	-5.62	7.69	2.07
Benzene	-7.29(42)	6.57(06)	-0.72(47)	-7.00	6.13	-0.87
Methanol	-8.53(18)	5.10(06)	-3.43(24)	-10.04	4.94	-5.10
Acetamide	-15.01(35)	6.79(18)	-8.22(27)	-16.87	7.17	-9.70
N-methylacetamide	-17.09(05)	7.76(26)	-9.33(31)	-17.61	7.54	-10.07

a) All data is in kcal mol^{-1} b) Experimental data from ref[73] for neon, xenon, Ref[65] for chloride, sodium. Ref[74] (ΔH_{X+w}) and Ref[75] (ΔG°_{X+w}) for acetamide and N-methylacetamide, Ref[76] for other solutes. Ref[76] gives enthalpies for a constant pressure solvation process px and these were converted using $\Delta H_{X+w} = \Delta H^{px}_{X+w} + k_B T$. When missing, entropies were derived from the difference of the Gibbs free energies with the enthalpies.

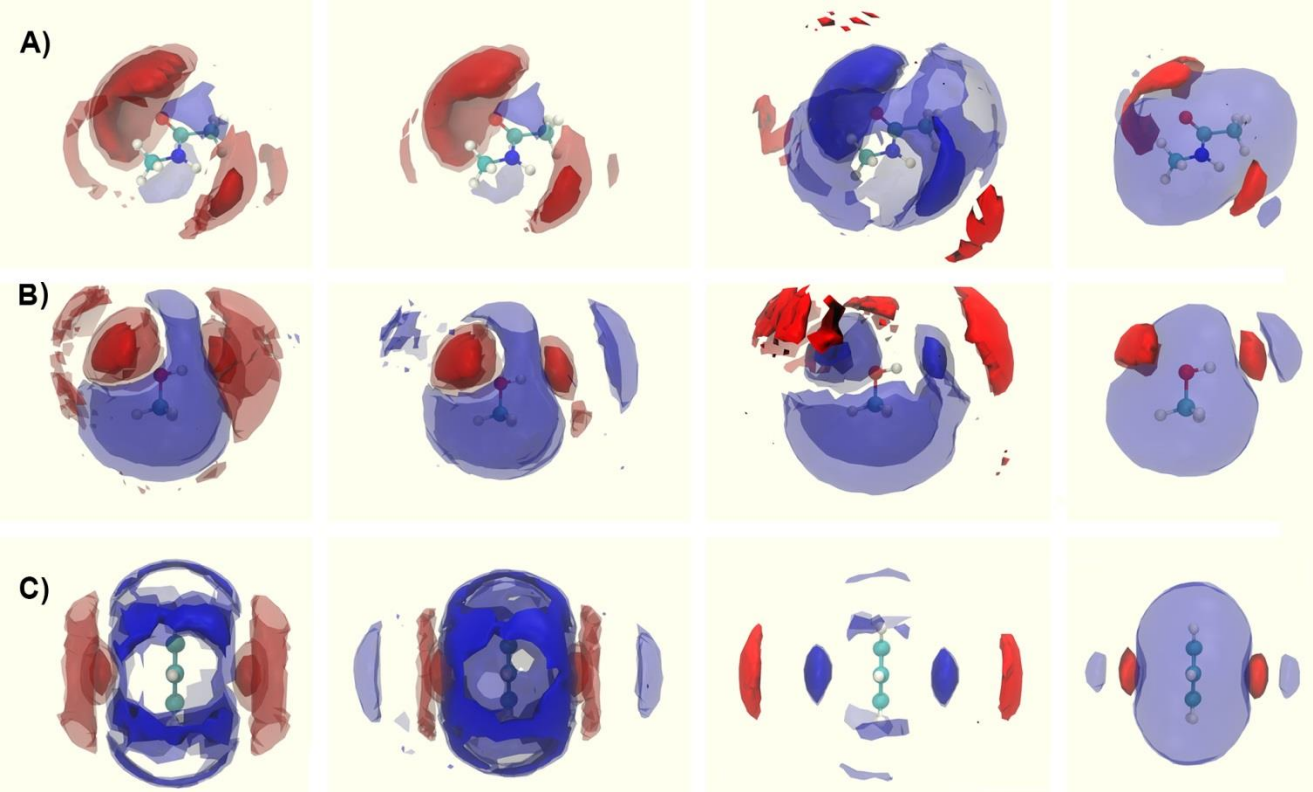

Figure 8. The spatial resolution of hydration thermodynamics around A) N-methylacetamide, B) methanol, C) benzene. For each solute, voxel contributions to $\Delta G_w^{\boldsymbol{s}}$, $\Delta H_{X(\boldsymbol{r})+w}^{\boldsymbol{s}}$, $-T\Delta S_w^{\boldsymbol{s},w}$ and $\rho(\boldsymbol{s})$ are shown from left to right. The blue isocontours indicate regions where water is less stable or has a lower density than in bulk. The red isocontours indicate regions where water is more stable or has a higher density than in bulk. The isocontours units are in kcal.mol^{-1}.Å^{-3} except relative density which is unitless. All density isocontours were drawn with the same isovalues $\rho(\boldsymbol{s})$: 2.7 (dark red); 0.4 (light blue). Other isovalues were: A) N-methylacetamide. $\Delta G_w^{\boldsymbol{s}}$: -0.1 (dark red); -0.0016 (light red); 0.02 (light blue). $\Delta H_{X(\boldsymbol{r})+w}^{\boldsymbol{s}}$: -0.1 (dark red); -0.03 (light red); 0.01 (light blue). $-T\Delta S_w^{\boldsymbol{s},w}$: -0.003 (dark red); 0.01 (light blue); 0.02 (dark blue);. B) Methanol. $\Delta G_w^{\boldsymbol{s}}$: -0.1 (dark red); -0.007 (light red); 0.05 (light blue). $\Delta H_{X(\boldsymbol{r})+w}^{\boldsymbol{s}}$: -0.1 (dark red); -0.02 (light red); 0.0025 (light blue). $-T\Delta S_w^{\boldsymbol{s},w}$: -0.002 (dark red); 0.006 (light blue); 0.05 (dark

blue). C) Benzene. ΔG_w^s: -0.1 (dark red); -0.009 (light red); 0.015 (light blue); 0.02 (dark blue). $\Delta H_{X(r)+w}^s$: -0.1 (dark red); -0.019 (light red); 0.0025 (light blue); 0.008 (dark blue). $-T\Delta S_w^{s,w}$: -0.0027 (dark red); 0.015 (light blue); 0.02 (dark blue).

Graphical TOC