Habilitation Thesis

Advances in Free-Energy Calculations Based on Classical Molecular Dynamics Simulations

Author(s):
Hansen, Niels

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010407855

Rights / License:
In Copyright - Non-Commercial Use Permitted
Advances in Free-Energy Calculations Based on Classical Molecular Dynamics Simulations

Habilitationsschrift

vorgelegt dem Departement Chemie und Angewandte Biowissenschaften der ETH Zürich

von
Dr. Niels Hansen

Date: May 3, 2014
Contents

1 Practical aspects of free-energy calculations: A review 1
  1.1 Introduction ................................................. 2
  1.2 Solvation of small neutral molecules ..................... 4
  1.3 Solvation of larger (drug-like) heteromolecules .......... 11
  1.4 Solvation of (single) monoatomic ions .................... 13
  1.5 Solvation of polyatomic ions .............................. 17
  1.6 Noncovalent binding ..................................... 18
  1.7 Conformational changes .................................. 24
  1.8 Conclusion and outlook .................................. 25

2 Calculation of derivative thermodynamic hydration and aqueous partial molar properties of ions based on atomistic simulations 29
  2.1 Introduction ................................................. 30
  2.2 Theory ..................................................... 35
    2.2.1 Derivative thermodynamic parameters ................. 35
    2.2.2 Standard states ....................................... 36
    2.2.3 Calculation of derivative thermodynamic properties .. 44
    2.2.4 Correction scheme .................................... 47
  2.3 Computational details ..................................... 52
    2.3.1 Simulation parameters ................................. 52
    2.3.2 Method I .................................................. 54
    2.3.3 Method II .................................................. 55
    2.3.4 Correction scheme .................................... 57
  2.4 Results and discussion .................................... 59
    2.4.1 Method I .................................................. 59
2.4.2 Method II ........................................... 65
2.4.3 Comparison with available experimental data ...... 72
2.5 Conclusion ............................................ 76

3 Assessment of enveloping distribution sampling to calculate relative free enthalpies of binding for eight netropsin-DNA duplex complexes in aqueous solution 81
3.1 Introduction .......................................... 82
3.2 Theory .................................................. 84
  3.2.1 Two state EDS Hamiltonian ....................... 84
  3.2.2 Parameter update scheme ......................... 85
3.3 Methods ............................................... 89
  3.3.1 Simulation protocols .............................. 89
  3.3.2 Topology .......................................... 90
  3.3.3 Trajectory analysis .............................. 91
3.4 Results and discussion ................................ 92
  3.4.1 Structural properties ............................. 92
  3.4.2 Free enthalpy differences for perturbations in neat DNA 92
  3.4.3 Free enthalpy differences for perturbations in DNA-netropsin complexes ......................... 96
  3.4.4 Relative binding free enthalpies of netropsin to different DNA binding sequences ................. 98
3.5 Conclusion and outlook .............................. 101

4 Relative free enthalpies for point mutations in two proteins with highly similar sequences but different folds 103
4.1 Introduction ........................................... 104
4.2 Theory .................................................. 106
4.3 Computational details ................................. 107
  4.3.1 Simulation parameters ............................ 107
  4.3.2 Simulated systems ............................... 108
  4.3.3 Trajectory analysis ............................. 109
4.4 Results and discussion ................................ 110
  4.4.1 Structural properties ............................ 110
  4.4.2 Free enthalpy changes ........................... 111
4.5 Conclusion ............................................ 119
5 Efficient combination of environment change and alchemical perturbation within the Enveloping Distribution Sampling (EDS) scheme: Twin-system EDS and application to the determination of octanol-water partition coefficients

5.1 Introduction .............................................................................................................. 122

5.2 Theory ...................................................................................................................... 127
  5.2.1 Calculation of partition coefficients .............................................................. 127
  5.2.2 Enveloping distribution sampling (EDS) ...................................................... 128
  5.2.3 Twin-system EDS ............................................................................................ 130
  5.2.4 Convergence assessment ............................................................................... 131

5.3 Computational details ............................................................................................. 132
  5.3.1 Simulation parameters .................................................................................... 132
  5.3.2 Simulated systems ............................................................................................ 133
  5.3.3 Trajectory analysis .......................................................................................... 135

5.4 Results and discussion ............................................................................................ 135
  5.4.1 Free enthalpies of hydration and of solvation in octanol ............................... 135
  5.4.2 Octanol-water partition coefficients .............................................................. 138
  5.4.3 Evaluation of the twin-system setup .............................................................. 143

5.5 Conclusion ................................................................................................................ 145

6 Outlook ......................................................................................................................... 149

Appendices ...................................................................................................................... 151

A Supplementary material for chapter 2 ....................................................................... 151

B Supplementary material for chapter 3 ....................................................................... 163

C Supplementary material for chapter 4 ....................................................................... 175

D Supplementary material for chapter 5 ....................................................................... 187

Bibliography .................................................................................................................. 195

Published work .............................................................................................................. 241

Lebenslauf ....................................................................................................................... 245
Practical aspects of free-energy calculations: A review

Free-energy calculations in the framework of classical molecular dynamics simulations are nowadays used in a wide range of research areas including solvation thermodynamics, molecular recognition and protein folding. The basic components of a free-energy calculation, i.e. a suitable model Hamiltonian, a sampling protocol and an estimator for the free energy are independent of the specific application. However, the attention that one has to pay to these components depends considerably on the specific application. Here, we review six different areas of application and discuss the relative importance of the three main components to provide the reader with an organigram and to make nonexperts aware of the many pitfalls present in free energy calculations.
1.1 Introduction

A wide range of fundamental chemical quantities such as binding or equilibrium constants, solubilities, partition coefficients, and adsorption coefficients are related to the difference in free energy between particular (non)physical states of a system.\(^1\) By means of statistical mechanics, free-energy differences may be expressed in terms of averages over ensembles of atomic configurations for the molecular system of interest. Such an ensemble can be generated by Monte Carlo (MC) or molecular-dynamics (MD) techniques. Much of the statistical-mechanical framework for calculating free energy differences has been developed some time ago.\(^2\)–\(^6\)

The free energy \(F\) of a system in the canonical ensemble, i.e. at constant number of particles, volume and temperature is given as

\[
F = -k_B T \ln \left\{ \left[ h^{3N} N! \right]^{-1} \int \int e^{-H(\vec{p}^N, \vec{r}^N)/k_B T} d\vec{p}^N d\vec{r}^N \right\},
\]

where \(k_B\) is Boltzmann’s constant, \(T\) the temperature, \(h\) Planck’s constant, \(N\) the number of particles or atoms in the classically treated molecular system, and \(\vec{r}^N = (\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)\) and \(\vec{p}^N = (\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_N)\) the Cartesian coordinates and conjugate momenta of all \(N\) atoms respectively. The factor \(N!\) is only present for indistinguishable particles. The Hamiltonian \(H(\vec{p}^N, \vec{r}^N)\) consists of the kinetic energy \(K(\vec{p}^N)\) and the potential energy \(V(\vec{r}^N)\) of the system. The latter describes the interactions between the various atoms and is also called a molecular force field.\(^7\) Since the integral in Eq. 1.1 is \(6N\)-dimensional and the integrand is positive definite, the absolute free energy can only be calculated in particular cases, i.e. for very simple model systems for which an analytical expression for the partition function can be obtained. In an isothermal-isobaric ensemble, the corresponding free enthalpy\(^8\) or Gibbs energy\(^9\) is given by

\[
G = -k_B T \ln \left\{ \left[ V h^{3N} N! \right]^{-1} \int \int \int e^{-[H(\vec{p}^N, \vec{r}^N) + pV]/k_B T} d\vec{p}^N d\vec{r}^N dV \right\}.
\]

In condensed phase systems such as biomolecules in aqueous solutions the pressure-volume-work term \(pV\) is usually negligible such that we use the term free energy throughout this review. For practical applications it generally suffices to calculate relative free energies,

\[
\Delta F_{BA} = F_B - F_A,
\]

where \(A\) and \(B\) denote two different systems, \(H_A(\vec{p}^N, \vec{r}^N)\) and \(H_B(\vec{p}^N, \vec{r}^N)\), or two different states that are connected by a coupling parameter \(\lambda\), \(H(\vec{p}^N, \vec{r}^N; \lambda)\).
leading to

\[ F_\lambda(\lambda') = -k_B T \ln \left\{ \left[ h^{3N} N! \right]^{-1} \int \int e^{-\frac{H(\vec{p}^N, \vec{r}^N; \lambda')}{k_B T}} d\vec{p}^N d\vec{r}^N \right\}, \quad (1.4) \]

or two different states along a phase space (reaction) coordinate \( R(\vec{r}^N) \), leading to

\[ F_R(R') = -k_B T \ln \left\{ \left[ h^{3N} N! \right]^{-1} \int \int \delta(R' - R(\vec{r}^N)) e^{-\frac{H(\vec{p}^N, \vec{r}^N)}{k_B T}} d\vec{p}^N d\vec{r}^N \right\}. \quad (1.5) \]

How to apply this framework in computer simulations in an effective fashion remains, however, a very active area of research, giving rise to over 3500 papers using the most popular free energy methods that were published in the first decade of this century with the publication rate increasing \( \sim 17\% \) per year.\(^\text{10}\) This rapid development comes at the price that it is increasingly difficult for researchers to find their way through the maze of available computational techniques. For reviews on the methodology to calculate free energy \textit{via} molecular simulation, we refer to the literature.\(^\text{11-20}\) Reviews which are more application oriented can be found in Refs 21–23. No single method for free-energy calculation can be considered as clearly superior to the others and the proper choice depends very much on the system under consideration. Christ et al.\(^\text{18}\) identified three main challenges that have to be met in free energy estimation from molecular simulation, which are (i) the choice of a suitable model Hamiltonian, (ii) the choice of a sampling protocol which allows generating a representative ensemble of configurations and (iii) the choice of an estimator for the free energy difference. The aim of the latter review was to enable the reader to classify the vast array of methods by identifying which choices have to be made for these three basic components. Here, we make an attempt to enable the reader to be aware of the various peculiarities and pitfalls that are characteristic for the various (bio)molecular systems for which free-energy calculations are performed. We do this by discussing equilibrium free-energy computation from the systems point of view. In classical molecular simulation three categories of free-energy differences can be distinguished which are \textit{conformational}, \textit{alchemical} and \textit{thermodynamic} free energy differences. Conformational free-energy differences refer to differences between two distinct conformational states, e.g. a left and a right-handed helix, of the same system. Alchemical free-energy differences refer to differences between two states differing in their Hamiltonian. Thermodynamic free-energy differences refer to differences between two thermodynamic state points, e.g. two different temperatures. In this review we focus on the first two types of calculations, their differences can be stated as follows. The free-energy difference between two conformational states is a logarithmic measure...
of the relative partition functions corresponding to a common Hamiltonian, but integrated over different regions in terms of the degrees of freedom of the system. The free-energy difference between two alchemical states is a logarithmic measure of the relative partition functions corresponding to the entire space in terms of the degrees of freedom of the system, but performed considering two different Hamiltonians (that are both functions of the same number of degrees of freedom).

Due to the explosion in the number of publications since the early 1990’s no review of the state of the art can hope to be comprehensive. However, there is broad consensus about the necessity of further methodological developments and the definition of best practices in all areas of applications in (bio)molecular modeling, of which the main ones are: (i) solvation of small neutral molecules, (ii) solvation of larger (drug-like) heteromolecules, (iii) solvation of (single) monoatomic ions, (iv) solvation of polyatomic ions, (v) noncovalent binding, (vi) conformational changes. In the following these application areas are reviewed separately with an eye to identify current obstacles and challenges.

### 1.2 Solvation of small neutral molecules

The free energy of solvation corresponds to the free energy of transferring a compound from a well-defined state (gas) to another (solution), allowing a direct comparison with experiment. The calculation of the free energy of solvation usually involves an alchemical transformation and was one of the first practical applications of the free energy perturbation and integration methodology.\(^{21,22,24}\) Today, solvation free energies remain of primary importance for developing and testing force fields, for testing new methodology, for gaining fundamental insights into the solvation process, and for specific applications such as predicting how molecular compounds will partition between different environments.

Condensed-phase or biomolecular force fields, e.g. CHARMM,\(^ {25–29}\) AMBER,\(^ {30–33}\) OPLS,\(^ {34–38}\) and GROMOS\(^ {39–49}\) mainly aim at the description of (bio)molecules in solution. They focus on the description of torsional-angle properties, nonbonded interactions, and solvation effects. However, the parameter sets used to describe most compounds have historically been based primarily on structural properties and fitting to results from quantum-mechanical calculations.\(^ {50,51}\) This is despite the fact that many properties of interest, especially in biomolecular systems, such as protein or peptide folding, depend on how compounds or their constituent moieties partition
between different environments. The primary reason why thermodynamic properties have until recently not been more generally incorporated into force-field parametrization was cost, and the difficulty in obtaining converged results.

However, solvation free energies have been used extensively for the verification of the OPLS force field, especially to rationalize the choice of partial atomic charges which are empirical parameters. For the most recent versions of the OPLS all-atom (AA) force field the partial charge assignment was based on hydration free energies calculated in explicit solvent for a set of 239 small molecules spanning diverse chemical functional groups commonly found in drugs and drug-like molecules.

When reparametrizing the GROMOS force field for the aliphatic CH\textsubscript{n} united atoms in 1998, Daura et al. fitted the repulsive van der Waals parameter of the water oxygen for interactions with nonpolar atoms to reproduce the experimental free energies of hydration of five of the alkanes studied: methane, ethane, propane, butane, and isobutane. Subsequent versions of the GROMOS force field have specifically been developed to reproduce the free enthalpy of hydration and apolar solvation in explicit solvent. The 53A5 force field was developed for an accurate description of the thermodynamic properties of pure liquids. Since it did not seem possible at that time to reconcile both pure liquid and hydration properties with a sufficient accuracy, the 53A6 force field aimed primarily at reproducing solvation properties of (neutral) amino acid side chain analogs in aqueous and nonaqueous solvents. The 53A6\textsubscript{OXY} parameter set strikes a balance between 53A5 and 53A6 versions, by providing a single set to reproduce both pure-liquid properties and solvation free energies of small oxygen-containing molecules. The 54A7 force field is based on 53A6 with a number of force field adjustments, namely, (i) the use of modified peptide \( \phi, \psi \) torsional-angle energy terms, in combination with a slightly reduced Lennard-Jones repulsion between the backbone oxygen and nitrogen atom types; (ii) the introduction of a new atom type for the methyl group in a positively charged entity, as e.g. found in the choline headgroup of phosphatidylcholine lipids; (iii) a change of the Lennard-Jones interaction parameters of the sodium and chloride ions based on calculated absolute single ion solvation free energies (see also below). In subsequent work also other functional groups such as occurring in amines, amides, thiols, sulfides, aromatics, and sulfones were (re)parametrized aiming at simultaneously improving pure-liquid and solvation properties. Recently force-field parameters describing 110 post-translationally modified (PTM) amino acids and protein termini, compatible with the GROMOS force fields 45A3 and 54A7, were derived. Validation against the hydration free energy of
side-chain analogs showed that the newly generated parameters compatible with the GROMOS 54A7 parameter set reproduce experimental data almost equally well as the original parameters reproduce hydration free energies for side-chain analogs of naturally occurring amino acids.\textsuperscript{64} A web server for automated introduction of PTMs to protein 3D structures is available.\textsuperscript{65} Parameters for post-translationally modified amino acids have also been reported for the AMBER ffd3 force field. However, the assessment was based on structural properties of proteins rather than on hydration free energies of side-chain analogs.\textsuperscript{66}

Recently, Mobley et al.\textsuperscript{67,68} have reported hydration free energies for 504 small molecules parametrized using the AMBER Antechamber program\textsuperscript{69} to assign parameters for a general AMBER force field (GAFF),\textsuperscript{70} which is compatible with the all-atom AMBER biomolecular force field. They identified systematic errors in force-field parameters for particular functional groups such as alkynes which could be fixed by using OPLS Lennard-Jones parameters for triple bonded carbons. Subsequent work focused on ethane, biphenyl and dioxin and their chlorinated derivatives.\textsuperscript{71} Except for the ethane derivatives, the agreement with experimental data was less good compared to their earlier set of small molecules. The introduction of virtual charge sites led to some improvement but the remaining error might not be reducible using fixed charge models. Hydration free energies were calculated with both implicit\textsuperscript{67} and explicit\textsuperscript{68} solvent. It was concluded that the explicit solvent simulations reliably outperform today’s implicit solvent models. Based on their extensive set of data Mobley et al.\textsuperscript{68} also proposed several remedies to improve implicit solvent models. The better performance of explicit solvent simulations over implicit ones was also noted by Shivakumar et al.\textsuperscript{72} using the GAFF\textsuperscript{70} force field and the all-atom CHARMM-MSI\textsuperscript{73} one.

More recently, Knight et al.\textsuperscript{74} calculated the hydration free energies of a set of 457 small neutral molecules using the CHARMM general force field CGenFF,\textsuperscript{75,76} which is compatible with the all-atom CHARMM biomolecular force field,\textsuperscript{77} using three different implicit solvent models. The set of molecules was based on the one used by Mobley et al.\textsuperscript{68} and parameterized automatically using the CHARMM-compatible ligand parametrization tools MATCH,\textsuperscript{78} ParamChem,\textsuperscript{75,76} and SwissParam,\textsuperscript{79} as well as by converting the GAFF parameters from Mobley et al.\textsuperscript{68} to CHARMM format. GAFF uses the restrained electrostatic potential (RESP) fitting approach\textsuperscript{80} to generate charges for the entire molecule concurrently while the MATCH, ParamChem, and SwissParam tools use a fragment-based approach, where charge distributions of a molecule are built-up from charges assigned to the component fragments of the molecule. Of the twelve combinations of solvent model
and parametrization scheme the Antechamber parameters (GAFF with semi-empirical Merek-Frosst AM1-BCC partial atomic charges$^{81,82}$) yielded the most accurate estimates.

The latter study and others$^{83-85}$ show that existing methods based on implicit solvation models are often reasonably accurate in calculating solvation free energies of a large number of compounds. However, explicit solvent simulations should still be regarded as the “gold standard” due to the aim of using these force fields for heterogeneous systems in solution.$^{68}$ It should be kept in mind that force-field parametrizations are usually carried out for one particular water model. While Shivakumar et al.$^{58}$ showed that there is little difference in overall performance of hydration free-energy predictions between the SPC, TIP3P, and TIP4P water models for a subset of 13 molecules out of the 239 molecules used in their study, Shirts and Pande showed that a modified version of the TIP3P model better reproduces hydration free energies of 15 amino acid side-chain analogs.$^{86}$ The influence of the solvent model on calculated hydration free energies of nucleobases and chloroform-to-water partition coefficients was recently studied by Wolf and Groenhof$^{87}$ using various combinations of force fields in conjunction with their native and non-native water models. It was found that the large differences in solvation free energies between different force fields are actually due to the nucleobase parameters and not the solvent models and that the difference in hydration free energy due to the use of a different water model is larger in case of aromatic amino acid analogs than in case of nucleobases. Hess and van der Vegt$^{88}$ also came to the conclusion that the choice of water model strongly influences the accuracy of calculated free energies of hydration of amino acid side-chain analogues.

Predictions of hydration free energies for compounds with multiple functional groups showed that the training sets currently used still lack sufficient coverage of chemical space.$^{89,90}$ For the class of nitroaromatic compounds, Ahmed and Sandler$^{91}$ tested 16 force-field/(charge+water) models out of which only 6 performed approximately equally well in predicting measured hydration free energies.

To assess the state of the art of a force field other properties than the solvation free energy should also be considered. Using a total of 146 molecular liquids, Caleman et al.$^{92}$ compared the ability of the OPLS-AA and GAFF force fields to reproduce key properties of neat liquids such as density, enthalpy of vaporization, surface tension, heat capacity at constant volume and pressure, isothermal compressibility, volumetric expansion coefficient, and static dielectric constant. The overall performance of the OPLS-AA force field was found to be somewhat better but both force fields have issues
with reproduction of the surface tension and the dielectric constant.

The more diverse the test sets become and the more diverse the chemical environments are in which small molecules should be simulated, the more likely it becomes that fixed charge models are reaching their limits. Mobley et al.\textsuperscript{68,90} reported an RMS deviation from experimental numbers of 5.2 kJ mol\(^{-1}\) for their test set of 504 molecules, which is comparable to the RMSD of 6.5 kJ mol\(^{-1}\) Caleman et al.\textsuperscript{92} found for the heat of vaporization. For larger molecules with multiple functional groups the error in the hydration free energy increases to up to 10 kJ mol\(^{-1}\).\textsuperscript{89} To be compatible with statistical QSPR methods\textsuperscript{93,94} to predict solubility, which have negligible computational cost, the RMS error has to decrease to about 4 kJ mol\(^{-1}\), also for complex molecules to justify the significant computational overhead of explicit solvent simulations. Additional degrees of freedom can be included in the parametrization process by modifying the form of the potential energy function for the short-range interactions, the combination rules for unlike interaction partners, and by considering explicit polarization. While the Lennard-Jones 12-6 potential energy function is the most widely used one, it is well known that increasing the repulsion exponent may improve the description of vapor-liquid equilibrium data.\textsuperscript{95,96} Repulsive exponents smaller than 12 such as 9 or 8 have also been suggested.\textsuperscript{97,98} The former is used for example in the condensed-phase force field COMPASS.\textsuperscript{99} Regarding the combination rules, the standard arithmetic or geometric mean rules often perform poorly\textsuperscript{100} and are mainly used due to the lack of systematic and transferable procedures for the derivation of improved combination rules. For including explicit polarization in empirical force fields various strategies are currently developed for many of the condensed phase force fields\textsuperscript{101} together with automated parametrization schemes that aim at reproducing hydration free energies.\textsuperscript{102} However, from disagreement between an experimental and a calculated hydration free energy alone, it cannot be concluded whether explicit polarization is needed. This usually requires the consideration of additional properties such as the dielectric permittivity. Missing polarizability has for example been demonstrated to be the reason for the underestimation of the hydration free energy and the dielectric permittivity of N-methyl acetamide by fixed charge models.\textsuperscript{103} We note, however, that the calculation of solvation free energies for small molecules is only one aspect of the complex process of (bio)molecular force-field development.

For protein simulations force-field improvements can only be evaluated on the basis of comparing simulation results to experimental data for a diverse set of protein structures. Recently Nerenberg et al.\textsuperscript{104} presented a parametrization strategy for a fixed-charge protein force field based on the AMBER
ff99SB parameters combined with the TIP4P-Ew water model, which is a re-parameterized version of the standard TIP4P water model for use with Ewald summation techniques. They calibrated the solute-solvent van der Waals interaction parameters of a set of 47 small molecules representing all of the chemical functionalities of standard protein side chains and backbone groups. To be consistent with the charge model used in AMBER ff99SB, partial charges were obtained by fitting calculated electrostatic potentials at the HF/6-31G* level using the RESP method. The RMSE in solvation free energies could be reduced from 7.3 to 2.5 kJ mol\(^{-1}\). With the combination of original AMBER ff99SB parameters and TIP4P-Ew water, nearly every chemical moiety was undersolvated. The new solute-solvent interaction parameters were evaluated based on simulations of dipeptide solutions, of short disordered peptides and of ubiquitin. For the latter case the favorable enhancement of solute-water interactions resulted in partial unfolding although the hydrophobic core remained intact. By reintroducing a 12-10 hydrogen bonding potential energy term (instead of 12-6) this problem could be remedied.

The latter study made use of reweighting for evaluating the influence of modified van der Waals parameters based on trajectories generated with the unmodified ones. This approach is very efficient for parameter studies if the perturbed ensemble has sufficient overlap with the unperturbed one. Through the use of Jacobian factors it can also be used in case of geometric changes.

We conclude that the use of solvation free energies in force-field calibration, testing, and comparison has become a standard tool. However, it is again noted that the force-field parameters are empirical. They have been derived using a specific set of conditions to reproduce a specific set of properties. As with any empirical force field, the choice of temperature, treatment of long-range nonbonded interactions, pressure coupling scheme, and so on, are implicitly incorporated into the parametrization but are in fact properties of the underlying algorithms used to integrate Newton’s equations of motion. This should be kept in mind when using force fields in conjunction with MD codes that were not used for the original parametrization or when changing the implicitly included parameters. The density of liquid octanol for example was reported to be significantly overestimated while at the same time the heat of vaporization was reported to be underestimated with the GROMOS 53A6 force field, using the GROMACS code presumably with non-standard settings compared to the recommended settings in GROMOS. This problem might become more subtle for polarizable force fields which include additional choices, e.g. how they treat intra-molecular polarization.
The current momentum in the area of solvation free energy calculations will soon lead to a further reduction of the discrepancy between experiment and simulation and it remains to be seen what the residual error will be that defines the limit of classical force fields.

In terms of the three choices we conclude that the choice of the Hamiltonian is the one which has the strongest influence on the accuracy of solvation free energies of small neutral molecules. For the other two choices the current literature demonstrates that a variety of techniques is available (see below) which can in many cases be used interchangeably to calculate a solvation free energy with a precision higher than the accuracy of the force field. However, also relatively small molecules require special attention if they show slow torsional-angle transitions such as in carboxylic acids, ibuprofen, or dimethoxyethane.

Because many methods can often be used interchangeably to calculate solvation free energies, such simulations are frequently used in method development, testing and comparison. Prompted by difficulties with complex intramolecular potential energy surfaces, expanded ensemble methods, a hybrid Monte Carlo - molecular dynamics approach, has found to be significantly more efficient than TI or BAR due to an enhanced conformational sampling. A combination of $\lambda$-dynamics and metadynamics has been suggested recently to enhance sampling on virtual variables.

Alternatives to the widely used nonlinear soft-core scaling proposed by Beutler et al. are being developed that flatten the potential energy only in a region that is energetically accessible under normal conditions. Naden and Shirts proposed a formalism based on splitting the potential energy function into a configurational and an alchemical part. This approach leads to a lower variance and can be efficiently implemented. Also methods based on sampling along a reaction coordinate, such as the adaptive biasing force method, have been used to calculate the solvation free energy directly by transferring a solute from the gas phase into the solvent across the gas/solvent interface.

Hydration free energies provide a probe of the underlying physics. Due to the asymmetry of the charge sites in water, its response to polar solutes depends on the internal charge distribution of the latter. Mobley et al. studied the hydration of various polar solutes (both fictitious ones with simple geometry and real ones) in different explicit water models. By inverting the charge distribution in the artificial solutes large differences in the hydration free energy of up to $40 \text{kJ mol}^{-1}$ were found, largely driven by the structure of water in the first hydration shell. Charge hydration asymmetry also occurs in ionic solvation and the insights gained from explicit solvent
1.3 Solvation of larger (drug-like) heteromolecules

Simulations have helped to improve implicit solvent models.\textsuperscript{131–133} Also the treatment of nonpolar contributions to the solvation free energy in implicit solvent models could be improved based on insights gained from explicit solvent simulations.\textsuperscript{134} By studying the solubility of alkanes up to \textit{n}-eicosane in water Ferguson et al.\textsuperscript{135} detected no sharp peak in the dependence of the solubility upon carbon numbers larger than twelve as suggested by different experimental sources, but rather a nearly exponential decrease. Moreover, remarkable similarities in the conformational ensemble in the gas and solvated phases were found suggesting that the effect of the solvent interaction is the appearance of a free energy barrier of order $k_B T$ separating the compact and extended free energy basins for sufficiently long chains and a destabilization of the most extended chain conformations.

Finally solvation free energy calculations are used to predict physico-chemical properties of compounds for which experimental data are scarce and structure-property relationships are uncertain such as for nitroaromatic compounds.\textsuperscript{136,137}

1.3 Solvation of larger (drug-like) heteromolecules

In many practical applications topologies and force-field parameters of heteromolecules such as substrates, inhibitors, cofactors or drug molecules\textsuperscript{138–141} are needed. These parameters are not standardly available and often have no close analogs within the desired biomolecular force field. Therefore, they have to be specifically assigned manually or by automated procedures which are available for all of the main families of force fields such as for Amber/GAFF,\textsuperscript{69,142} CHARMM/CGenFF,\textsuperscript{75,76,78,143} OPLS,\textsuperscript{144} and GROMOS\textsuperscript{145} although with different levels of sophistication. Because the parametrization is an underdetermined problem it has to be ensured that the parameters are consistent with the macromolecular force field applied to the other components of the system. For thermodynamically calibrated force fields this validation should include the calculation of the solvation free energy in polar and nonpolar environments. Unfortunately, experimental values for solvation free energies are only known for less than one percent of the millions of organic compounds prepared to date.\textsuperscript{146,147} It is of importance to develop efficient robust and reliable calculation procedures for such molecules that are often characterized by multiple conformational substates\textsuperscript{148–151} which have to be sampled sufficiently in both the liquid and the gas phase such that the calculated free energy depends only on the force-field parameters.
This makes the use of enhanced sampling techniques mandatory in many cases. Khavrutskii and Wallqvist\textsuperscript{152} combined thermodynamic integration with Hamiltonian replica exchange and showed that converged results are obtained for molecules with internal rotational barriers of up to 60 kJ mol\textsuperscript{-1} using only a few nanoseconds of simulation time.

Paluch et al.\textsuperscript{153} proposed a method to predict the solubility limit of low solubility solids based on a single experimental reference solubility for each solute and a single free energy simulation of the solute-solvent system. The latter was carried out using an expanded ensemble calculation along with a combined Wang-Landau/Bennett’s acceptance ratio method. A particularly important aspect was that the proposed method requires fewer experimental data points than the Abraham general solvation model\textsuperscript{154} used for comparison.

In the absence of any experimental data solvation free energies are often predicted by group contribution methods, e.g. to predict the hydration free energies of amino acids or proteins. However, explicit-solvent MD-based hydration free energies for 15 N-acetyl-methylamide amino acids with neutral side chains differ considerably from those based on additive group contribution methods.\textsuperscript{155}

Another example of missing experimental hydration free energies is that of barbiturates which are of great pharmaceutical interest. Using parameters from the OPLS-AA force field Garrido et al.\textsuperscript{156} tested different simulation setups for an efficient calculation of the hydration free energy including alternative ways to account for electrostatic interactions, such as the reaction-field method and the particle-mesh Ewald method. If the former is used in an automated setup, the problem of assigning atoms to charge groups may arise. Recently Canzar et al.\textsuperscript{157} showed how this problem can be solved efficiently.

A recent blind test including 63 complex drug-like molecules showed that solvation free energies are at present predicted with an RMS error of 10 – 14 kJ mol\textsuperscript{-1}.\textsuperscript{146}

In terms of the three choices we conclude that the choice of Hamiltonian for the solvation of drug-like molecules is as crucial as for the solvation of small neutral molecules but its accuracy is more difficult to assess due to scarce experimental data. Moreover the sampling protocol and free-energy estimator determine the efficiency of the calculation much more than is the case for small molecules. The choice of an appropriate technique is more case-dependent than for small molecules.

This renders the question of convergence to be more crucial for drug-like molecules compared to the case of small molecules discussed above. Approaches such as “double-wide sampling” or forward and reverse transform-
1.4 Solvation of (single) monoatomic ions

Within the realm of classical thermodynamics, only sums of thermodynamic quantities associated with neutral sets of ions can be probed. These sums may be partitioned into single-ion contributions, but only within an unknown offset constant weighted by the integer ion charge. Single-ion thermodynamics becomes accessible when spectroscopic techniques such as photoionization or laser photodetachment are combined with statistical mechanics. However, the resulting parameters, also referred to as real single-ion solvation parameters, still account for the mixture of two physically very different effects: bulk solvation and liquid-surface properties. The estimation of the surface term requires extrathermodynamic considerations, which may be either experimental or theoretical, and show considerable spread (e.g. as much as 100 kJ mol\(^{-1}\) in terms of solvation free energies for monovalent ions), making the estimation of intrinsic absolute solvation parameters, i.e. those which originate exclusively from the ion-solvent interactions, uncertain. For a comprehensive discussion of these issues, we refer to the book of Hünenberger and Reif. In the context of atomistic (explicit-solvent) simulations the single-ion solvation free energy is usually measured as the work of coupling the solute (ion) to the solvent (water) and thus reflects naturally the intrinsic absolute solvation free energy of an ion \(I^z\) (denoted by \(\Delta_s G^\text{ex} [I^z_g]\) in ref. 166), i.e. ideal solvation in a surface-free fluid at a fixed reference concentration (the same number of moles per liter in the vapor and solution) under which condition the entropy of translation in the vapor equals the entropy of liber-
ation from a fixed point in solution.

Since the earliest calculations of ionic solvation free energies in the 1980s,\textsuperscript{167–170} it has been realized that the raw results can be dramatically sensitive to the boundary conditions and treatment of electrostatic interactions used during these simulations, with typical variations on the order of 100\,kJ\,mol\textsuperscript{−1} or larger in terms of solvation free energies for monovalent ions in water. As a result the parametrization of ion-solvent interaction parameters suffers from problems related to the ambiguity of the experimental training set as well as from the need to correct the raw methodology-dependent simulation results, so that methodology-independent values are obtained. The main factors that lead to discrepancies between simulation (using fixed-charge water models) and experiment in the context of single-ion solvation are:

(1) approximate force-field representation (functional form, \textit{e.g.} absence of explicit electronic polarizability in the model)
(2) approximate force-field parameters (water and ion-water parameters)
(3) finite-sampling errors
(4) approximate-electrostatics errors
(5) finite-size errors
(6) improper-summation errors
(7) additional approximations involved in the evaluation of properties other than the solvation free energy (see below)
(8) inaccuracy or ambiguity of the experimental data

Points (1) and (2) are related to the choice of the Hamiltonian and are thus not different from other free-energy calculations. Also point (3) has to be considered carefully in every free-energy calculation. For points (4) to (6) a scheme has been developed that allows to correct raw solvation free energies \textit{ex post}, so that methodology-independent values are obtained.\textsuperscript{61,166,171–175} The corrected results are then exclusively characteristic of the underlying molecular model, as determined by the representation of the solvent, of the ion, and of the ion-solvent van der Waals interactions, and no longer depend on arbitrary simulation parameters such as the system size or the electrostatic cutoff distance. These values correspond to the idealized situation of an infinite bulk phase exempt of surface polarization, in which electrostatic interactions are exactly Coulombic. Using this correction scheme, a reparametrization of the ion-solvent Lennard-Jones interaction coefficients for Na\textsuperscript{+} and Cl\textsuperscript{−} (among other ions) with the SPC water model\textsuperscript{176} (as well as the SPC/E model\textsuperscript{177}) against experimental hydration free energies was conducted by Reif and Hünenberger.\textsuperscript{61} Three different parameter sets (L, M and H) were calibrated, corresponding to different assumed values for the absolute intrinsic hydration free energy Δ\textsubscript{hyd}$^\Phi$[H\textsuperscript{+}] of the proton at P\textsuperscript{o} = 1 bar and
1.4 Solvation of (single) monoatomic ions

\[ T^- = 298.15 \text{K}, \] which is an experimentally elusive quantity. Recently, Dahlgren et al.\textsuperscript{178} extended the correction scheme to derivative thermodynamic hydration and aqueous partial molar properties (point (7) in the above list), and showed that approximate internal consistency and qualitative agreement with experimental results can only be achieved when an appropriate correction scheme is applied, along with careful consideration of standard-state issues. As for the free energy itself the correction terms for derivative thermodynamic hydration and aqueous partial molar properties are substantial. Directions for future improvements, with the ultimate goal of reaching a consistent and quantitative description of single-ion hydration thermodynamics in atomistic simulations, are also provided in the latter work. Apart from point (8), the correction scheme permits a thermodynamically consistent calibration of ion-solvent interaction parameters. Not correcting for points (4) to (6) necessarily leads to some kind of system-size dependence in the calculated quantities, the degree of which depends on the actual electrostatic scheme used. This might be acceptable for simulations focusing on structural properties such as radial distribution functions, which show remarkable insensitivity to the detailed treatment of electrostatic forces. Simulations attempting to calculate quantities related to the energy of the system, however, lump all the overlooked errors into the non-Coulomb interaction parameters and are therefore not transferable to other system sizes and electrostatic schemes. When reviewing the most recent literature related to ionic force-field parametrization or the use of atomistic simulations to probe single-ion solvation properties,\textsuperscript{179–188} however, it is striking that correction terms are seldom taken into account. One of the reasons for not including such corrections might be that in practical simulations ions are propagated with approximate electrostatic interactions within systems of finite sizes which differs from the ideal situation of Coulombic electrostatic interactions in a macroscopic nonperiodic system underlying the interaction parameters calibrated against methodology independent hydration free energies. This might strongly affect the configurational sampling and might lead to a significant undersolvation especially for small box sizes,\textsuperscript{187} because the correction terms coining methodology-independent ion hydration free energies are predominantly negative. It would therefore be desirable to design effective electrostatic interaction schemes which correct the approximate electrostatics and finite-size effects at the level of the forces, so as to achieve a solvent polarization around ionic groups that is exempt of artifacts. Attempts to include the correction terms into the equation of motion by means of restraints are in progress.\textsuperscript{189}

We note that calibrating ion-pairs instead of single ions leads to an ap-
proximate cancellation of improper summation errors because these are linear in the ionic charge, but has the disadvantage that the underlying single-ion solvation free energies can only be determined up to an additive constant leading to a range of possible interaction parameters that all lead to the correct ion pair properties in dilute solutions. When using simulations of finite concentrations to determine these additive constants and thus the appropriate interaction parameters, one faces the problem of calibrating simultaneously ion-ion and ion-solvent interactions, even for low concentration.\textsuperscript{190,191} Note that the problem of an unknown additive constant also appears in case of single-ion solvation where all values are anchored on the elusive absolute intrinsic hydration free energy of the proton. Also in that case a range of interaction parameters consistent with what is known about pair properties in dilute solutions is obtained.\textsuperscript{61} However, the advantage of calibrating against absolute intrinsic single-ion properties is the possibility to sequentially calibrate ion-solvent (infinite dilution regime) and ion-ion (finite concentration regime) interaction parameters.

In terms of the three choices we conclude that neither the sampling protocol nor the free energy estimator constitute an obstacle in calculating the solvation free energies of ions. The choice of the Hamiltonian is classically reduced to the choice of the ion-solvent interaction parameters and the choice of the water model, possibly complicated by the need to use non-standard mixing rules for ion-ion interactions.\textsuperscript{61,180} Whether the water model is chosen to be polarizable depends on the purpose of the ion force field, i.e. whether it should be applied in combination with a biomolecular force field that is calibrated for a specific water model. The consideration of the correction scheme represents a fourth choice in the simulation of ions in solution and allows to obtain solvation free energies which are independent of the method used to treat the electrostatic interactions. However, it is currently only available as an \textit{ex post} scheme. As a result, ions parametrized in the context of this scheme might have the tendency to be undersolvated in practical applications such as using them as counterions in an MD simulation of a biomolecule. We note that while it is commonly taken for granted to choose the ionic partial charges equal to their formal (integer) charges in nonpolarizable force fields, an alternative suggestion has been made to use reduced charges of $0.7\,e$ as a first-order approximation for electronic polarizability.\textsuperscript{192–195}
1.5 Solvation of polyatomic ions

In force-field calibration for monoatomic ions, the only degree of freedom is usually the Lennard-Jones repulsion parameter, while the dispersion parameter can be derived on the basis of approximate formulas relating it to the ionic polarizability\textsuperscript{61,166,196} and the ionic charge is either set to the formal charge or scaled by a factor of approximately 0.7 to implicitly account for the effect of electronic polarization.\textsuperscript{192–195} For a given value of the standard absolute intrinsic hydration free energy of the proton, $\Delta G_{\text{hyd}}^{\Phi}[\text{H}^+]$, and a given water model there is a unique relationship between the repulsion parameter and the experimentally accessible conventional hydration free energy. In contrast, the parametrization of polyatomic ions is an underdetermined problem; it allows for multiple solutions of similar quality with respect to the reproduction of a single experimental value. This is due to the multiplicity of atomic partial charges and Lennard-Jones repulsion parameters, as well as to the loss of a direct connection between dispersion parameters and atomic polarizabilities, the latter being ill-defined when considering (united) atoms within molecules.

Another difference to free energy calculations of monoatomic ions is that the electrostatic interaction correction scheme\textsuperscript{61,173,174} requires a numerical solution of the Poisson equation to obtain a continuum-electrostatics estimate for the charging free energy of the ionic solute in a macroscopic nonperiodic system with Coulombic electrostatic interactions and based on the experimental solvent permittivity, and in a periodic system with a specific electrostatic interaction scheme and based on the model solvent permittivity. These two calculations allow to account for the combined effect of approximate-electrostatics errors, finite size errors and deviations between experimental and model solvent permittivity. In case of rigid ions these calculations can be performed on the basis of a single structure. The improper-summation errors can be accounted for in a way similar to the case of monoatomic ions.

The latter scheme was proposed by Reif et al.\textsuperscript{49} and used to recalibrate the nonbonded interaction parameters for the charged amino acid side chains in the GROMOS force field, based on ionic side-chain analogs. As for the monoatomic ions, the parametrization was based on $\Delta G_{\text{hyd}}^{\Phi}[\text{H}^+] = -1100 \text{kJ mol}^{-1}$. The resulting GROMOS 54A8 force field is the first of its kind to contain nonbonded parameters for charged amino acid side chains that are derived in such a rigorously thermodynamic fashion. Subsequently the force field was tested on structural properties of electrolyte solutions, lipid bilayers and proteins.\textsuperscript{197}

Note that the recently revised AMBER parameters\textsuperscript{198} for phosphate ions
are anchored to a different value of $\Delta G_{\text{hyd}}^{\circ}[\text{H}^+] \ (= -1052 \text{ kJ mol}^{-1})$ compared to the GROMOS force field illustrating an additional difficulty when comparing force-field parameters for ions.

In terms of the three choices we note that all the difficulties present in the context of monoatomic ions also apply to polyatomic ions. Depending on the size and flexibility of the latter, sampling problems similar to the ones discussed in the context of neutral flexible molecules might appear.

### 1.6 Noncovalent binding

Molecular recognition forms the basis of virtually all biological processes. Understanding the interactions between proteins and their ligands is key to rationalize the molecular aspect of enzymatic processes and the mechanisms by which cellular systems integrate and respond to regulatory signals. From a medicinal perspective, there is great interest in the development of computer models capable of predicting accurately the strength of protein-ligand association making the accurate computation of free energies of binding a key challenge for computer-aided drug design. The main advantage of atomistic simulations over faster, empirical scoring functions is a more realistic inclusion of all thermodynamically relevant phenomena such as protein or/and ligand flexibility and the possibility of the explicit inclusion of the solvent, which is usually necessary to account for the entropic contribution to the free energy. However, despite their potential the effectiveness of atomistic simulations as predictive tools for protein-ligand binding remains uncertain.

In contrast to the preceding sections, binding free energy calculations are usually affected by a mixture of inaccuracies in all three choices, i.e. by errors in the force field, by insufficient sampling, and by the propagation of the former two effects by the free energy estimator. For sizeable systems such as protein-ligand complexes, convergence may be hard to assess because of slow degrees of freedom or rare events such as side-chain flipping at the binding site, both introducing considerable noise into the calculated free energy. It is therefore important to ask whether the time scale characteristic of the slowest degree of freedom is crucial for the free-energy change being estimated.

The theoretical framework for calculating binding free energies in the realm of statistical mechanics is well established. However, depending on the desired accuracy and available computational resources, different ap-
Noncovalent binding approaches can be used to calculate binding affinities. The state of the art for calculating binding free energies has been discussed in several recent reviews and perspective articles,\textsuperscript{23,198,203,204,210–215} such that we limit our discussion to the three choices and some general challenges, which are independent of the methodology chosen.

Different choices for the three basic ingredients give rise to a multitude of methods each with a different trade-off between accuracy and computational efficiency. Three main classes of methods can be envisioned to calculate binding free energies, which are (i) end-point methods, (ii) methods based on the calculation of the free energy along a reaction coordinate, the potential of mean force (PMF), and (iii) methods based on the calculation of the free energy along a thermodynamic pathway employing alchemical transformations.

End-point and PMF methods are usually used to calculate the free energy change associated with the process that brings a ligand from an unbound state to bind to a receptor. This free energy change is often referred to as the \textit{absolute} free energy of binding, while it is in fact a free energy difference. Here, we prefer the term \textit{binding free energy} for this quantity and distinguish it from a \textit{relative binding free energy}, which denotes the difference in binding free energies between two compounds. The calculation of relative binding free energies is a common application of alchemical free energy methods. However, these methods can also be used for the calculation of binding free energies, e.g. in the context of the double-decoupling approach.\textsuperscript{200} Comparison to experimental data requires the consideration of standard state corrections\textsuperscript{200,216} leading to the \textit{standard binding free energy}.

End-point methods such as the molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA)\textsuperscript{217,218} and molecular mechanics/generalized Born surface area (MM/GBSA)\textsuperscript{217,219} only consider the two end states of interest, such as the protein free from and bound to a ligand, and calculate their absolute free energies. The binding free energy is then obtained by subtraction. Each of the free energies is decomposed into the mean enthalpic energy of the solute, the mean solute entropy, the polar solvation free energy, and the non-polar solvation free energy. The entropic contributions are restricted to the conformational entropy and can be estimated from normal mode analysis,\textsuperscript{220–223} quasi-harmonic analysis,\textsuperscript{222,224} the restrain and release approach\textsuperscript{222,225,226} or are simply neglected.\textsuperscript{210,227–229} Inaccuracies inherent in these methods arise from a sensitivity towards the choice of the dielectric constant for the PB and GB calculations,\textsuperscript{223} the use of implicit solvent,\textsuperscript{230–232} inaccuracies in calculating entropic contributions,\textsuperscript{226,233} but also from subtracting two large numbers (the approximate absolute free energies before and after binding), typically orders of magnitude larger than the binding
20 1 Practical aspects of free-energy calculations: A review

free energy.\textsuperscript{234} Recently, Silver et al.\textsuperscript{235} proposed a novel end-point method based on uniform, rotameric enumeration of ligand torsional degrees of freedom to map out and explicitly integrate over the potential energy landscape. The method is structured around the use of the dead-end elimination and A* algorithms, which sort configurations by their energies and explicitly computes their contribution to the Boltzmann distribution. Linear interaction energy (LIE) is another end-point method which is based on the assumption that the free energy of binding shows a linear dependence on the polar ($\beta$) and nonpolar ($\alpha$) changes in ligand-surroundings energies from MD averages.\textsuperscript{236} Because LIE has several parameters, care should be taken not to overfit the data.\textsuperscript{226} PDLD/S-LRA/\beta combines the semi-microscopic protein dipoles Langevin dipoles method, the linear response approximation (LRA) and the non-electrostatic part of LIE. This approach seems to be more effective than MM-PBSA, LRA or LIE.\textsuperscript{226} An alternative end-point method, not based on molecular dynamics or Monte Carlo simulations, is the "mining minima" method\textsuperscript{237} which estimates the partition function through a harmonic approximation to the Hessian matrix. The exploration of minima local to the starting configuration, by transformations along low-frequency eigenvectors, allows for the inclusion of multiple possibly relevant states in the partition function estimate.

Methods based on the calculation of the free energy along a reaction coordinate are based on the definition of a pathway that connects the two states of interest and are reviewed for example in refs 238 and 239. The major challenges are to determine which degrees of freedom are important and how they participate in the reaction coordinate. An accurate reaction coordinate should convey the reaction mechanism, provide kinetically meaningful free energy surfaces and facilitate calculations of the rate constant. For any reaction, the exact reaction coordinate is the committor probability,\textsuperscript{240-242} the fraction of trajectories initiated with Boltzmann distributed momenta from an atomic configuration $\vec{r}^N$ that commit to the product basin (B).\textsuperscript{243-245} Unfortunately, the committor probability $p_B(\vec{r}^N)$ is not easy to compute and usually approximated in terms of collective variables that are functions of the configuration that compress many atomistic details into variables supposed to be physically important. The key challenge is to learn which collective variables are important and how they are involved in the reaction coordinate. For ion-pair dissociation Mullen et al.\textsuperscript{246} showed that dynamic recrossing of the dividing surface at the $p_B(\vec{r}^N) = 1/2$ isosurface is an inescapable consequence of dimensionality reduction to a single coordinate. An erroneous reduction of energy barriers by one-dimensional potentials of mean force was recently reported by Kopelevich\textsuperscript{247} in the context of transport of a
1.6 Noncovalent binding

hydrophobic nanoparticle into a lipid bilayer. However, if calculation of the free energy of binding is the main purpose, one-dimensional PMFs combined with restraints, orthogonal to the direction of binding, can be an efficient approach.\(^{248}\) It has also to be kept in mind that molecular recognition can be significantly more complex than a two-state process.\(^{249}\) Recently, de Ruiter and Oostenbrink introduced the distance field (DF) as a reaction coordinate for the calculation of reversible protein-ligand binding.\(^{250}\) DF is a grid-based method in which the shortest distance between the binding site and a ligand is determined avoiding routes that pass through the protein.

Methods based on the calculation of the free energy along a thermodynamic pathway employing alchemical transformations are widely used in the calculation of relative binding free energies. Apart from the established methods such as TI and staged FEP many other approaches including EDS,\(^{251}\) BAR,\(^{4}\) MBAR,\(^{252}\) and \(\lambda\)-dynamics\(^{17}\) have been tested on relevant protein-ligand systems. In any case, errors due to insufficient sampling are more significant than the differences in the free-energy estimator.\(^{253}\) Binding free energies can also be calculated by alchemical transformation, usually in the context of the double-decoupling method,\(^{200,209,254}\) in which two separate calculations are carried out for decoupling the ligand from the solution and receptor environments. The latter process is often carried out using restraints to lock the non-interacting ligand into the binding pocket to enhance convergence of the free energy. The free-energy contribution associated with these restraints has to be taken into account in the thermodynamic cycle, for example by analytical approaches.\(^{200,224,255}\) From a computational perspective the determination of the binding site volume in the presence of restraining potential energy terms in the Hamiltonian is crucial for obtaining reliable free-energy estimates.\(^{203}\) Another alchemical method to calculate (absolute) binding free energies is the binding energy distribution analysis method (BE-DAM) which is in practice only applicable using an implicit solvent.\(^{256}\) It involves simulation of the ligand restricted to the protein binding site, but without interactions between the protein and the ligand. From this the probability distribution for the binding energy is determined.

In the course of molecular design projects, end-point methods are most useful in the early, exploratory stages, while methods involving conformational sampling are most useful in later stages, when the goal is to optimize promising lead compounds.\(^{200}\) Such a hierarchy of methods has been used recently to shed light on the molecular recognition of the coreceptor CXCR4 by the HIV-1 glycoprotein gp120.\(^{257}\) When screening drug candidates prior to more elaborate free-energy calculations it should be kept in mind, however, that MM-PBSA, in general, cannot be expected to reliably resolve differences
within 12 kJ mol\(^{-1}\).\textsuperscript{210} The availability of automated workflows for the setup and analysis of binding free energy calculations\textsuperscript{258–261} is expected to facilitate a more realistic evaluation of the different methods and how these could be combined in an efficient way to guide molecular design projects. However, to be acceptable for inclusion into workflows for lead-optimization, binding free energies have to be converged reliably with 4.2 kJ/mol variance error.\textsuperscript{210}

Current challenges which are of general nature and not bound to a particular method are:

1. Multiple binding modes: Often multiple binding modes are of importance either due to different possible ligand orientations within the binding site,\textsuperscript{206} or due to different conformations a ligand may adopt\textsuperscript{262} or due to a combination of both effects. These multiple orientations have to be sampled with the correct relative populations to avoid any bias. In the context of relative free-energy calculations the two ligands to be compared may have distinct orientations giving rise to a slowly converging reorientation step in the calculation. Such cases can be treated more efficiently using the recently suggested “separated topologies” method.\textsuperscript{263} This approach can also be used to determine the relative free energies of multiple orientations of the same ligand. However, also the characterization of the ligand’s unbound state can be of major relevance.\textsuperscript{264} If a ligand has different conformations in the solvent but only one conformer binds to the protein, the free energy of focusing the different conformers to the one that binds to the protein has to be accounted for in binding free-energy calculations.\textsuperscript{265} In the same line the protein may also adopt several metastable states that contribute to the binding free energy. Considering only one conformational state neglects the free energy associated with confining the protein to that particular configuration.\textsuperscript{266}

2. Binding-site hydration: Upon binding, the ligand may replace several water molecules in the binding site. Depending on the free-energy change these water molecules experience when leaving the site, this replacement might contribute (un)favorably to the binding free energy.\textsuperscript{267} Several studies on model cavity-ligand systems\textsuperscript{268–272} demonstrated that the role of water-mediated interactions and ligand dehydration can be far more relevant than the direct cavity-ligand interaction, owing to electrostatic screening and to entropic terms arising from solvent reorganization.\textsuperscript{273,274} For proteins with solvent-exposed binding sites these effects are difficult to capture with methods relying on implicit solvent models.\textsuperscript{275,276} For a binding site deeply buried the exchange of water molecules with the bulk region may be very slow which may
1.6 Noncovalent binding

lead to convergence problems.\textsuperscript{254} When estimating the effect of ligand modifications on binding free energies it is crucial to take into account that water molecules are maintained for one variant but may be displaced for others.\textsuperscript{277,278} To study which water molecules can favorably be replaced by a ligand, the free-energy difference between a water molecule and an apolar probe was calculated for a selection of water sites in the binding pockets of two proteins.\textsuperscript{279} Such an analysis may give valuable insights for potency optimization in drug design.

3. Definition of the bound state: The standard binding free energy depends on the definition of the bound state. If the binding is strong and specific, this does usually not pose a problem as long as the specific choice covers all important conformations of the complex.\textsuperscript{200,256,280,281} For weak and less localized binding the dependence on the binding site volume would be noticeable. As pointed out by Mihailescu and Gilson,\textsuperscript{281} the theoretical expression for the binding site depends on the experimental technique used.

4. Standard state correction: To convert calculated binding free energies into standard binding free energies, a correction term needs to be added which can be expressed in terms of the system volume or ligand concentration in the unbound state.\textsuperscript{200,216} However, alternative approaches to estimate the required corrections in practice may differ significantly.\textsuperscript{282}

5. Finite size effects: Another underappreciated aspect are finite-size effects in binding free energy calculations. At present, finite-size effects on charging free energies are best understood in the context of the solvation of monoatomic ions.\textsuperscript{61,166,172–174,178,283} In this case a numerical correction scheme\textsuperscript{173} and a corresponding approximate analytical version\textsuperscript{174} are available. The numerical version of this scheme has recently been extended to the case of small polyatomic ions\textsuperscript{49} and to insertion of such ions into a simple model receptor, namely a functionalized C60 buckyball in water.\textsuperscript{284} However, considering the most general case of a complex polyatomic charged ligand inserted into a charged protein in solution none of the schemes available at present\textsuperscript{224,250,284–289} are sufficiently general, practical and accurate. Recently, a new method for removing finite size effects has been proposed based on a continuum-electrostatics analysis. It requires performing Poisson-Boltzmann calculations on the protein-ligand system.\textsuperscript{175} The approach introduces the concept of the residual integrated potential to account for the finite-size effect related to the solvent-excluded volume of the protein and the ligand, an effect which is absent in monoatomic-ion solvation.
6. Force-field inaccuracy: Due to the hundreds of parameters involved in empirical force fields the propagation of errors in these parameters on calculated binding free energies is a complex problem. Recently, Rocklin et al.\textsuperscript{290} investigated the sensitivity of binding free-energy calculations to the non-bonded energy parameters in force fields – atomic radii, dispersion well-depths, and partial charges – by performing tens of thousands of small parameter perturbations. They estimated that random, uncorrelated errors in force-field nonbonded parameters must be smaller than 0.02 e per charge, 0.06 Å per radius, and 0.04 kJ mol\(^{-1}\) per well depth in order to obtain 68\% confidence that a computed binding affinity for a moderately sized lead compound will fall within 4.2 kJ mol\(^{-1}\) of the true affinity, if these are the only sources of error considered. Fixed charge models of ligands, parametrized against hydration free energies, might easily have larger uncertainty in the partial charges, especially in non-polar binding sites.

1.7 Conformational changes

The importance of knowing the change in free energy associated with a change in molecular conformation was already mentioned in the context of hydration or binding free energy calculations. It relies on the definition of conformational states as well as on the ability to define a reduced set of (spatial) coordinates \(R(\vec{r}_N)\) on which the free energy is projected. Such a hypersurface is commonly called a reaction coordinate and, in configurational space, is a function of the positions of atoms in the system. Note that the term reaction coordinate is eventually associated to the minimum-free-energy pathway connecting the reference state to the target state but is commonly employed to characterize the order parameter along which the variation of the free energy is determined. The free energy as a function of the reaction coordinate \(R(\vec{r}_N)\), or the potential of mean force, is given by Eq. 1.5, where the term in curly brackets is the probability of finding the system lying on the reaction coordinate. Difficulties related to the representation of the reaction coordinate have been sketched in the previous section. A comprehensive discussion of methods to obtain reaction coordinates\textsuperscript{291–293} is beyond the scope of the present review. Another difficulty is related to the definition of conformational states. For small systems such as carbohydrates, states can be relatively well defined due to the rigid nature of the glycosidic linkage. Carbohydrates are therefore often used for testing new methodology\textsuperscript{294,295} or to calibrate force fields.\textsuperscript{296} Small peptides in solution show significant more flexibility. Dipeptides are often used to test new methodology\textsuperscript{297} or to invest-
igate the pathway dependence of the free energy. For larger (oligo)peptides the free-energy difference between different helical forms may depend on the definition of states. Relative free energies of 4, 0, and 12 kJ mol$^{-1}$ for the $\pi$, $\alpha$, and $3_{10}$ helical forms of a deca-alanine peptide in water were calculated by BS-LEUS using the sum of the two dihedral angles $\phi_{n+1} + \psi_{n}$ encompassing the successive peptide bonds to define the helical states, and values of 5, 0, and 47 kJ mol$^{-1}$ were obtained by EDS using an RMSD criterion to define the helical states.

Like the definition of optimal reaction coordinates, the structure classification of biopolymeric structures is also an active field of research. For proteins the characterisation of the unfolded state represents a major obstacle in the calculation of folding free energies. A further difficulty in conformational free-energy calculations is the need to used enhanced sampling and biasing techniques for all but a few simple systems for which simple counting of configurations from a long MD trajectory works.

In terms of the three choices we conclude that conformational free energies are among the most challenging systems due to the many choices involved in setting up these calculations. As in the case of the choice of reaction coordinate, the free-energy difference will depend on the choice of definition of the conformational states. Next, if such conformational states are of high energy, the accuracy of the Hamiltonian is difficult to assess because conformations high in free energy are difficult to sample. Recently, EDS has been applied to solve this problem. So, state definition and securing sufficient sampling are the main challenges when calculating free-energy differences between different conformational states.

## 1.8 Conclusion and outlook

Although the two fundamental problems of inaccuracies in the Hamiltonian and of insufficient sampling are still prevalent, the calculation of free energy differences has seen some consolidation through the definition of best practices and through a healthy skepticism towards the performance of computational models. One of the great challenges remains the accurate calculation of the entropic contribution in molecular processes, which may or may not be a substantial portion of the free energy. Some developments in the field that may lead to further consolidation are:

1. Evaluation and standards: Validation sets such as provided on alchemy.org in the binding database or through various blind tests will help to generate a commonly accepted set of benchmark
data useful for method and force-field development.

2. Sensitivity analysis: Because free energies may be sensitive to the force-field parameters, the use of efficient perturbation approaches may help to provide an estimate for the uncertainty due to an inaccurate Hamiltonian. The use of different free-energy methods for the same problem may give information regarding convergence and the sensitivity towards the free-energy estimator.

3. Comparison of methods: Because different free-energy calculation methods use different information from the Hamiltonian, a comparison of methods on a pure theoretical basis may not always be sufficient to provide practical recommendations. Comparison of methods in practical settings is therefore as important.

4. Inter-availability of code: The transfer of a new method developed, implemented and tested for one particular MD software package to another one is far from being trivial because today’s molecular dynamics packages are very complex pieces of software developed over decades often by a diverse group of contributors with different backgrounds and experience. However, such transfer is essential for a wider acceptance of a particular method and also for a better comparison to other methods not implemented in the original software. Plug-ins with interfaces to different MD codes may help to disseminate new methods to a wider community of users.

5. Critical use of experimental data. Experimental measurements are invariably contaminated with error which may affect the maximally possible correlation between simulation and experiment that can be achieved. Often modelers try too hard to reproduce experimental data as precisely as possible ignoring the fact that these data are also subject to uncertainty.

This thesis contributes to the further development of free-energy calculations in classical molecular dynamics simulation in various ways. In Chapter 2, the feasibility of calculating and correcting derivative thermodynamic hydration and aqueous partial molar properties for sodium and chloride ions is investigated. The main merit of that study is to set a clear framework for these types of calculations and to point towards directions for future improvements, with the ultimate goal of reaching a consistent and quantitative description of single-ion hydration thermodynamics in molecular dynamics simulations.

In Chapter 3, the performance of the enveloping distribution sampling (EDS) method to calculate free-energy differences is tested by calculating relative binding affinities of netropsin to various binding sites in the DNA
minor groove. A new approach to determine the parameters of the EDS reference state is presented and compared to a previous scheme.

The capability of a protein to change to a completely different fold as a result of minor mutagenic perturbation is studied in Chapter 4 by means of free energy calculations. Currently the structures of protein sequences at the interface between folds cannot be predicted reliably with both ab initio and knowledge based prediction algorithms. It is demonstrated that point mutations at different positions in the amino acid sequence do not contribute equally to the free-enthalpy difference between the native and non-native sequence-structure combinations. The individual free-enthalpy changes are almost sequence independent in the four-strand/one-helix structure, the stable form of GB95, while in the three-helix bundle structure, the stable form of GA95, an interplay between residues 20 and 45 is observed, supporting the experimental findings for this protein.

Chapter 5 contributes to the methodological development of the EDS approach by extending it to probe a single-simulation alternative to the thermodynamic cycle that is standardly used for measuring the effect of a modification of a chemical compound, e. g. from a given species to a chemical derivative for a ligand or solute molecule, on the free-enthalpy change associated with a change in environment, e. g. from the unbound state to the bound state for a protein ligand system of from one solvent to another one for a solute molecule.

In Chapter 6, a brief outlook for the future advancement of free-energy calculations is given.
Calculation of derivative thermodynamic hydration and aqueous partial molar properties of ions based on atomistic simulations

The raw ionic solvation free energies calculated on the basis of atomistic (explicit-solvent) simulations are extremely sensitive to the boundary conditions and treatment of electrostatic interactions used during these simulations. However, as shown recently [Kastenholz, M. A.; Hünenerberger, P. H. J. Chem. Phys. 2006, 124, 224501; Reif, M. M.; Hünenerberger, P. H. J. Chem. Phys. 2011, 134, 144104], the application of an appropriate correction scheme allows for a conversion of the methodology-dependent raw data into methodology-independent results. In this work, methodology-independent derivative thermodynamic hydration and aqueous partial molar properties are calculated for the Na\(^{+}\) and Cl\(^{-}\) ions at \(P^\circ = 1\) bar and \(T^\circ = 298\) K, based on the SPC water model and on ion-solvent Lennard-Jones interaction coefficients previously reoptimized against experimental hydration free energies. The hydration parameters considered are the hydration free energy and enthalpy. The aqueous partial molar parameters considered are the partial molar entropy, volume, heat capacity, volume-compressibility and volume-expansivity. Two alternative calculation methods are employed to access these properties. Method I relies on the difference in average volume and energy between two aqueous systems involving the same number of water molecules, either in the absence or in the presence of the ion, along with variations of these differences corresponding
to finite pressure or/and temperature changes. Method II relies on the calculation of the hydration free energy of the ion, along with variations of this free energy corresponding to finite pressure or/and temperature changes. Both methods are used considering two distinct variants in the application of the correction scheme. In variant A, the raw values from the simulations are corrected after the application of finite difference in pressure or/and temperature, based on correction terms specifically designed for derivative parameters at \(P^0\) and \(T^-\). In variant B, these raw values are corrected prior to differentiation, based on corresponding correction terms appropriate for the different simulation pressures \(P\) and temperatures \(T\). The results corresponding to the different calculation schemes show that, except for the hydration free energy itself, accurate methodological independence and quantitative agreement with even the most reliable experimental parameters (ion-pair properties) are not yet reached. Nevertheless, approximate internal consistency and qualitative agreement with experimental results can be achieved, but only when an appropriate correction scheme is applied, along with a careful consideration of standard-state issues. In this sense, the main merit of the present study is to set a clear framework for these types of calculations and to point towards directions for future improvements, with the ultimate goal of reaching a consistent and quantitative description of single-ion hydration thermodynamics in molecular dynamics simulations.

2.1 Introduction

The microscopic size of the systems considered and the approximate treatment of electrostatic interactions currently represent two of the most severe limitations in the accuracy of atomistic (explicit-solvent) simulations concerning (bio-)molecular systems in the condensed phase.\(^{328,329}\) The problem is particularly severe in the context of systems involving species or functional groups with net charges, \(e.g.\) in simulations of protein or electrolyte solutions, and even more so when processes are considered that involve a change in the net charge of a system, \(e.g.\) in calculations of solvation free energies for ionic species. Since the earliest calculations of ionic solvation free energies in the 1980s,\(^{167–170}\) it has been realized that the raw results can be dramatically sensitive to the boundary conditions and treatment of electrostatic interac-
2.1 Introduction

Simulations used during these simulations, with typical variations on the order of 100 kJ mol$^{-1}$ or larger for monovalent ions in water.

Recently, however, progress has been made towards the reliable calculation of single-ion solvation properties, by correcting the raw methodology-dependent simulation results \textit{ex post}, so that methodology-independent values are obtained.\textsuperscript{61,166,172–174} The corrected results are then exclusively characteristic of the underlying molecular model, as determined by the representation of the solvent, of the ion, and of the ion-solvent van der Waals interactions, and no longer depend on arbitrary simulation parameters such as the system size or the electrostatic cutoff distance. These values correspond to the idealized situation of an infinite bulk phase exempt of surface polarization, in which electrostatic interactions are exactly Coulombic. Using this correction scheme, a reparametrization of the ion-solvent Lennard-Jones interaction coefficients for Na$^+$ and Cl$^-$ (among other ions) with the SPC water model\textsuperscript{176} (as well as the SPC/E model\textsuperscript{177}) against experimental hydration free energies was conducted by Reif and Hünenberger.\textsuperscript{61} Three different parameter sets (L, M and H) were calibrated, corresponding to different assumed values for the absolute intrinsic hydration free energy $\Delta G_{\text{hyd}}^\circ[H^+]$ of the proton at $P^0 = 1$ bar and $T^- = 298.15$ K, which is an experimentally elusive quantity.\textsuperscript{166} An initial assessment of the reoptimized parameters based on experimental quantities other than hydration free energies,\textsuperscript{61} including first peak positions in the ion-water radial distribution functions, aqueous salt partial molar volumes and salt crystal properties, led to preliminary evidence for the appropriateness of set L, calibrated based on $\Delta G_{\text{hyd}}^\circ[H^+] = -1100$ kJ mol$^{-1}$. This value is also the one recommended in Ref. 166, based on a careful analysis of the nearly hundred estimates reported to date for this quantity, relying on various types of experiments and associated extra-thermodynamic assumptions.

The aim of the present study is to investigate the feasibility of calculating and correcting derivative thermodynamic hydration and aqueous partial molar properties for Na$^+$ and Cl$^-$ at $P^0$ and $T^-$, so as to use these quantities as additional validation data for set L with SPC water, and for the associated $\Delta G_{\text{hyd}}^\circ[H^+]$ estimate of $-1100$ kJ mol$^{-1}$. More specifically, the hydration parameters considered are the hydration free energy $\Delta G_{\text{hyd}}^\circ$ and enthalpy $\Delta H_{\text{hyd}}^\circ$, and the aqueous partial molar parameters considered are the partial molar entropy $s_{\text{aq}}^\circ$, volume $v_{\text{aq}}^\circ$, isobaric heat capacity $c_{P,\text{aq}}^\circ$, isothermal volume-compressibility $k_{T,\text{aq}}^\circ$ and isobaric volume-expansivity $a_{P,\text{aq}}^\circ$. These quantities are directly connected to the hydration free energy $\Delta G_{\text{hyd}}^\circ$ of the ion, along with its first and second pressure or/and temperature derivat-
For the aqueous partial molar parameters, the connection also involves the corresponding gas-phase molar parameters, which are known analytically for the alkali and halide ions at room temperature. Because these gas-phase parameters permit a trivial interconversion between hydration and aqueous partial molar properties, the hydration parameters $\Delta S_{\text{hyd}}^\circ$, $\Delta V_{\text{hyd}}^\circ$, $\Delta C_{P,\text{hyd}}^\circ$, $\Delta K_{T,\text{hyd}}^\circ$, and $\Delta A_{P,\text{hyd}}^\circ$ are not discussed explicitly here. In contrast, the free energy and enthalpy cannot be formulated in terms of corresponding aqueous partial molar parameters in the absence of a fixed zero point for the energy, and are therefore discussed in the form of $\Delta G_{\text{hyd}}^\circ$ and $\Delta H_{\text{hyd}}^\circ$. An alternative, not used in the present work, would be to consider the formation parameters of the aqueous ions instead.

The connection existing between the seven thermodynamic parameters considered here and the hydration free energy $\Delta G_{\text{hyd}}^\circ$, along with its first and second pressure or/and temperature derivatives, suggests one possible approach for the calculation of these properties based on atomistic simulations, via finite-difference analysis of $\Delta G_{\text{hyd}}(P,T)$ values calculated at different pressures $P$ and temperatures $T$. Alternatively, the hydration enthalpy $\Delta H_{\text{hyd}}^\circ$ and the partial molar volume $v_{aq}^\circ$ can also be directly connected to the variations of the energy and volume, respectively, upon inserting the ion into a pure-liquid sample of infinite extent. If the pressure and temperature dependences of these variations are known, this second connection applies to five of the thermodynamic parameters considered here, i.e. all except $\Delta G_{\text{hyd}}^\circ$ and $s_{aq}^\circ$. This suggests another possible approach for the calculation of these five properties based on atomistic simulations, via finite-difference analysis of $\Delta H_{\text{hyd}}(P,T)$ and $v_{aq}(P,T)$ values calculated from the above energy and volume differences at different pressures $P$ and temperatures $T$.

The two strategies outlined above for calculating (a subset of) the seven properties of interest are both considered and compared in the present study. The approach relying on energy and volume differences between aqueous systems excluding or including the ion, referred to here as Method I, is in principle simpler because it only requires two simulations per $P,T$-point. The approach relying on the hydration free energy, referred to here as Method II, is somewhat more complicated because it requires a calculation of the reversible work associated with the cavitation and charging processes, i.e. multiple simulations, at each $P,T$-point.

Additionally, for each of these two methods, two distinct variants are considered for the application of the correction scheme to the raw simulation results (energy and volume differences in Method I, hydration free energies in Method II). In Variant A, the raw values are corrected after application of finite difference in pressure or/and temperature, based on cor-
2.1 Introduction

rection terms specifically designed for derivative thermodynamic parameters at $P^o$ and $T^-$. In Variant B, these raw values are corrected prior to differentiation, based on corresponding correction terms appropriate for the different simulation pressures $P$ and temperatures $T$. The four combinations of methods and variants are thus IA, IB, IIA and IIB. For a given molecular model, these different schemes become entirely equivalent in the limit of infinitesimal finite-difference intervals and infinite sampling times. However, the corresponding results may differ in practice due to finite-difference errors and finite simulation times, and the observed discrepancies, along with the consideration of different system sizes, are expected to provide an indication on the level of precision and convergence of the calculated numbers.

It should be stressed that the calculation of derivative thermodynamic properties and their use as validation data for a given molecular model in the context of ionic solvation, although of very fundamental significance, is a highly non-trivial task. The main issues affecting these types of calculation are the following:

1. Long simulation times are required to reach estimates that are both precise and converged. This requirement results from a trade-off between the size of the finite-difference interval and the sampling time of the two compared simulations. On the one hand, the precision of a fully-converged finite-difference estimate decreases with increasing size of the interval considered, due to the increasing contribution of non-linear components. On the other hand, the error on the calculated number at finite simulation time decreases with increasing interval size, because it depends on the statistical uncertainties affecting the two compared values, divided by this interval size. As a result, precise estimates, implying a reasonably small finite-difference interval, can only be calculated to an acceptable level of convergence using relatively long simulations.

2. The simplified force-field representation is typically less adequate and accurate when considering derivative as opposed to non-derivative thermodynamic properties. This limitation results from the more limited emphasis generally placed on the reproduction of derivative properties during force-field design and parametrization. In general, the functional form and parameters of a force field are selected primarily to reproduce observables related to free energies, enthalpies and volumes, e.g. the density, enthalpy of vaporization and excess free energy for a liquid, or the solvation free energy for an ion. Their accuracies in terms of derivative properties, such as the entropy, heat capacity, compressibility and expansivity, is usually lower, due to both a limited functional flexibility and a lower weight attributed to these properties.
in the parametrization procedure.

3. The correction scheme for single-ion solvation free energies,\textsuperscript{173} recently generalized to derivative thermodynamic parameters,\textsuperscript{174} is affected by errors concerning its input parameters. This generalized scheme requires experimental estimates for the pressure and temperature dependences of the effective ionic radius \( R_I \), of the solvent density \( \rho \), dielectric permittivity \( \varepsilon \), air-liquid interfacial potential \( \chi \), and derivative of this potential with respect to the inverse curvature radius for convex \( \tilde{\chi}^+ \) or concave \( \tilde{\chi}^- \) interfaces, as well as corresponding estimates \( \rho' \), \( \varepsilon' \), \( \chi' \), \( \tilde{\chi}'^+ \) and \( \tilde{\chi}'^- \), respectively, for the solvent model employed in the simulations. Ambiguities and uncertainties in the definition of these functions result in corresponding errors on the corrected values.

4. The experimental data to compare with is ambiguous. Similarly to the proton hydration free energy \( \Delta G_{\text{hyd}}^{\circ} [\text{H}^+] \) (see above), the derivative thermodynamic hydration and aqueous partial molar parameters of this species are experimentally elusive.\textsuperscript{166} Although values have been suggested for \( \Delta H_{\text{hyd}}^{\circ} [\text{H}^+] \), \( s_{\text{aq}}^{\phi} [\text{H}^+] \), \( v_{\text{aq}}^{\phi} [\text{H}^+] \), \( c_{P,aq}^{\phi} [\text{H}^+] \), \( k_{T,aq}^{\phi} [\text{H}^+] \) and \( a_{P,aq}^{\phi} [\text{H}^+] \), see e.g. the recommended values in Ref. 166, they should be taken with great caution in view of the sparseness and indirect nature of the corresponding experimental determinations, and of their dependence on specific extra-thermodynamic assumptions. As a result, comparison of data on the basis of salt parameters (e.g. \( \text{Na}^+ + \text{Cl}^- \)) or of parameter differences between ions of identical charges (e.g. \( \text{Na}^+ \) vs. \( \text{K}^+ \)) is certainly at present more meaningful than single-ion comparisons. For the second-derivative quantities \( c_P^{\phi} \), \( k_T^{\phi} \) and \( a_P^{\phi} \), however, the experimental estimates even for these sums or differences are not highly reliable, in view of the very small number of independent determinations and of the limited consistency among the resulting estimates.\textsuperscript{166}

In view of these difficulties, the goal of the present study is to cut a first path through an unexplored jungle rather than to reach quantitative agreement with sometimes rather uncertain experimental data. Note, finally, that if many previous simulation studies have investigated ionic solvation free energies \( \Delta G_{\text{hyd}}^{\phi} \) and (approximate) enthalpies \( \Delta H_{\text{hyd}}^{\phi} \) (see Refs. 61,166 and references therein), only a few have considered ionic solvation entropies\textsuperscript{330–335} \( \Delta S_{\text{hyd}}^{\phi} \) and aqueous ion partial molar volumes\textsuperscript{174,336–338} \( v_{\text{aq}}^{\phi} \), while none to our knowledge has considered the derivative properties \( c_{P,aq}^{\phi} \), \( k_{T,aq}^{\phi} \) and \( a_{P,aq}^{\phi} \).
2.2 Theory

2.2.1 Derivative thermodynamic parameters

The thermodynamic parameters considered here are the Gibbs free energy $G$, enthalpy $H$, entropy $S$, volume $V$, isobaric heat capacity $C_P$, isothermal volume-compressibility $K_T$ and isobaric volume-expansivity $A_P$. Collectively denoting these seven variables as $Y$, they are related to the free energy by

$$Y = \hat{O}_G G,$$  \hspace{1cm} (2.1)

where $\hat{O}_Y$ is the thermodynamic-derivative operator\textsuperscript{166} associated with the parameter $Y$. These operators are given by

$$\begin{align*}
\hat{O}_G &= 1 \\
\hat{O}_H &= 1 - T \partial_T \left[ = -T^2 \partial_T T^{-1} \right] \\
\hat{O}_S &= - \partial_T \\
\hat{O}_V &= \partial_P \\
\hat{O}_{K_T} &= - \partial^2_P \\
\hat{O}_{A_P} &= \partial^2_{P,T}
\end{align*}$$  \hspace{1cm} (2.2)

where $\partial_P$ and $\partial_T$ represent partial differentiation with respect to pressure or temperature, respectively, the other variable as well as the system composition being kept constant, and $\partial^2$ indicates corresponding second (self or cross) derivatives. Main and alternative operator expressions are provided for $H$ and $C_P$, the alternative ones between square brackets. For $H$, the two expressions are related to the Gibbs and Gibbs-Helmholtz equations, respectively. For $C_P$, they correspond to the temperature derivative of the entropy and enthalpy, respectively. The two pairs of expressions are equivalent in the context of exact derivatives, but may differ numerically when translated into corresponding finite-difference equations. Unless otherwise specified, the main expressions are used in the present work for both $H$ and $C_P$. The extensive variables $K_T$ and $A_P$ are referred to as the volume-compressibility and the volume-expansivity, respectively. These names are selected to distinguish $K_T$ and $A_P$ from the intensive compressibility $\kappa_T$ and expansivity (thermal expansion coefficient) $\alpha_P$, respectively, the definition of which is ambiguous in the context of partial molar properties.\textsuperscript{174}

For a pure gas-phase species, the molar parameter corresponding to variable $Y$ will be noted $y_g$, where $y = \mu$, $h$, $s$, $v$, $c_p$, $k_T$ or $a_P$ for $Y = G$, $H$, $S$, $V$, $C_P$, $K_T$ or $A_P$. For a single solute species in aqueous solution, the corresponding partial molar parameters will be noted $y_{aq}$. The hydration parameters of the species are then defined by

$$\Delta Y_{hyd} = y_aq - y_g.$$  \hspace{1cm} (2.3)
Obviously, Eq. 2.1 also applies to gas-phase molar and aqueous partial molar properties, as well as to hydration parameters, \( i.e. \)

\[
y_g = \dot{O}_Y \mu_g , \quad y_{aq} = \dot{O}_Y \mu_{aq} \quad \text{and} \quad \Delta Y_{hyd} = \dot{O}_Y \Delta G_{hyd} . \tag{2.4}
\]

Note that all the thermodynamic parameters considered in this work are intrinsic (as opposed to real), \( i.e. \) they exclude any contribution associated with the crossing of the polarized air-water interface.\(^{166}\)

### 2.2.2 Standard states

Most commonly, although by no means systematically, experimental thermodynamic data pertaining to ionic solvation is tabulated in the literature according to the \( bbme_T \) standard-state convention.\(^{166}\) In this case, reference is made to the standard pressure \( P^o = 1 \) bar, molality \( b^o = 1 \) mol kg\(^{-1} \) and temperature \( T^- = 298.15 \) K. Since formation parameters are not considered in the present work, the definition of a standard state for the electron (ideal electron gas at temperature \( T^- \) with properties calculated according to Fermi-Dirac statistics in \( bbme_T \)) is irrelevant here.

In the \( bbme_T \) convention, the standard gas-phase molar properties of a species, noted here \( y_g^o \), refer to a pure ideal gas (no intermolecular interactions) at pressure \( P^o \) and temperature \( T^- \). Note in particular that the ideal-gas law implies

\[
v_g^o = (P^o)^{-1}RT^- , \quad k_{T,g}^o = (P^o)^{-2}RT^- \quad \text{and} \quad a_{P,g}^o = (P^o)^{-1}R , \tag{2.5}
\]

where \( R \) is the ideal gas constant. For closed-shell monoatomic species exempt of thermally-accessible electronically excited states at \( T^- \), as is the case for \( e.g. \) alkali cations and halide anions at room temperature,\(^{166} \) the Sackur-Tetrode equation\(^ {339–341} \) implies in addition

\[
s_g^o = R \left\{ (5/2) + \ln \left[ CM^{3/2}(P^o)^{-1}(T^-)^{5/2} \right] \right\} \quad \text{and} \quad c_{P,g}^o = (5/2)R \tag{2.6}
\]

for \( e.g. \) Na\(^+ \) and Cl\(^- \), where \( M \) is the molar mass of the species and

\[
C = \frac{(2\pi)^{3/2}R^{5/2}}{N_A^4h^3} = 820.512 \text{ (kg mol}^{-1})^{-3/2} \text{ bar K}^{-5/2} , \tag{2.7}
\]

\( N_A \) being the Avogadro constant and \( h \) the Planck constant. Note that Eq. 2.6 is in principle not applicable to the proton and the electron, where Fermi-Dirac statistics should be used instead of Boltzmann statistics.\(^ {166,342} \) Although the resulting values differ noticeably for the electron, Eq. 2.6 still
provides an excellent approximation in the case of the proton. In the \textit{bbmert} convention, the standard aqueous partial molar properties of a species, noted here \( y_{\text{aq}}^{\varphi} \), refer to the solute alone in an ideal aqueous solution (no direct or solvent-mediated solute-solute interactions) at pressure \( P^\circ \), molality \( b^\circ \) and temperature \( T^- \).

The direct outcome of theoretical calculations, \textit{e.g.} based on atomistic simulations, generally concerns fixed-point rather than standard parameters.\textsuperscript{166} The fixed-point gas-phase molar properties of a species, noted here \( y_{g}^{\varphi} \), refer to the translationally motionless species at temperature \( T^- \), surrounded by vacuum. In contrast to the corresponding standard parameters \( y_{\text{aq}}^{\varphi} \), these parameters no longer depend on the choice of a standard pressure \( P^\circ \). The removal of molecular translation implies in particular that in terms of fixed-point quantities, the molar volume as well as its derivatives vanish, \textit{i.e.}

\[
\varepsilon_g^{\varphi} = 0 \quad , \quad k_{T,g}^{\varphi} = 0 \quad \text{and} \quad a_{P,g}^{\varphi} = 0 . \tag{2.8}
\]

For closed-shell monoatomic species exempt of thermally-accessible electronically excited states at \( T^- \), as is the case for \textit{e.g.} alkali cations and halide anions at room temperature, one also has

\[
s_g^{\varphi} = 0 \quad \text{and} \quad c_{P,g}^{\varphi} = 0 \quad \text{for} \quad \text{e.g.} \quad \text{Na}^+, \text{Cl}^- . \tag{2.9}
\]

The fixed-point aqueous partial molar properties of a species, noted here \( y_{\text{aq}}^{\varphi} \), refer to the translationally motionless solute species at temperature \( T^- \), surrounded by pure water at pressure \( P^\circ \) and temperature \( T^- \). In contrast to the corresponding standard parameters \( y_{\text{aq}}^{\varphi} \), these parameters only depend on the choice of a standard pressure \( P^\circ \) \textit{via} its effect on the state of the solvent, and no longer rely on the definition of a standard molality \( b^\circ \).

Considering either fixed-point or standard quantities, Eq. 2.3 provides a definition for the corresponding hydration parameters, referred to as point-to-point or standard, respectively, \textit{i.e.}

\[
\Delta Y_{\text{hyd}}^{\varphi} = y_{\text{aq}}^{\varphi} - y_{g}^{\varphi} . \tag{2.10}
\]

and

\[
\Delta Y_{\text{hyd}} = y_{\text{aq}} - y_{g} . \tag{2.11}
\]

Note in particular that the notations \( \Delta Y_{\text{hyd}}^{\varphi} \) and \( y_{\text{aq}}^{\varphi} \) can be used interchangeably for \( Y = V, K_T \) and \( A_P \) (due to Eq. 2.8), as well as for \( S \) and \( C_P \) in the context of alkali and halide ions (due to Eq. 2.9). Point-to-point hydration parameters can also be viewed as resulting from an alternative standard-state convention where identical effective solute concentrations (molarities)
are selected for the gas-phase and aqueous states. Such a convention is also commonly used in the literature concerning ionic thermodynamics.\textsuperscript{343-350}

In terms of fixed-point gas-phase molar and aqueous partial molar properties, as well as point-to-point hydration parameters, Eq. 2.4 becomes

\[
y^\Phi_g = \hat{\Omega}_Y \mu^\Phi_g, \quad y^\Phi_{aq} = \hat{\Omega}_Y \mu^\Phi_{aq} \quad \text{and} \quad \Delta Y^\Phi_{\text{hyd}} = \hat{\Omega}_Y \Delta G^\Phi_{\text{hyd}},
\]

(2.12)

with the convention that \( \hat{\Omega}_Y \) applied to a function of \( P^o \) and \( T^- \) implies differentiation of the corresponding \( P,T \)-dependent quantity followed by evaluation at \( P^o \) and \( T^- \). In terms of standard gas-phase molar properties, Eq. 2.4 also implies

\[
y^\Phi_g = \hat{\Omega}_Y \mu^\Phi_g.
\]

(2.13)

However, corresponding equations for standard aqueous partial molar and hydration parameters must be considered more carefully, due to the presence of an ambiguity concerning the derivatives of the solvent density (see below).

The difference between fixed-point and standard quantities is related to the phase-space volume accessible to the species in terms of translational degrees of freedom in its standard state.\textsuperscript{166} The translational momentum-space volume is identical in the gas-phase and aqueous states, and proportional to \((MT^-)^{3/2}\). However, the translational coordinate-space volume differs between the two states. For gas-phase species, this coordinate-space volume is the ideal-gas molar volume \((P^o)^{-1}RT^-\), as determined by \( P^o \) and \( T^- \). In this case, the fixed-point molar properties can be standardized as

\[
y^\Phi_g = y^\Phi_g + \Delta y^\Phi_g \to^\Phi = y^\Phi_g - R\hat{\Omega}_Y \left\{ T^- \ln \left[ \tilde{C}(MT^-)^{3/2}RT^- \right] \right\},
\]

(2.14)

where \( \tilde{C} = R^{-1}C \). It is easily verified that Eqs. 2.5 and 2.8, as well as Eqs. 2.6 and 2.9 in the specific case of alkali and halide ions, satisfy the connection of Eq. 2.14. For aqueous species, the coordinate-space volume is the effective molar volume \((b^o\rho^\Phi)^{-1}\) accessible to the solute species, as determined by \( b^o \) and \( \rho^\Phi \), where \( \rho^\Phi \) is the pure-solvent density at \( P^o \) and \( T^- \). In this case, the fixed-point partial molar properties can be standardized as

\[
y^\Phi_{aq} = y^\Phi_{aq} + \Delta y^\Phi_{aq} \to^\Phi = y^\Phi_{aq} - R\hat{\Omega}_Y \left\{ T^- \ln \left[ \tilde{C}(MT^-)^{3/2} RT^- \right] \right\}.
\]

(2.15)

Given the definitions of Eqs 2.10 and 2.11, it follows from Eqs. 2.14 and 2.15 that the two types of hydration parameters are connected by

\[
\Delta Y^\Phi_{\text{hyd}} = \Delta Y^\Phi_{\text{hyd}} + \Delta \Delta Y^\Phi_{\text{hyd}} = \Delta Y^\Phi_{\text{hyd}} + R\hat{\Omega}_Y \left\{ T^- \ln \left[ \frac{RT^o b^o \rho^\Phi}{P^o} \right] \right\}.
\]

(2.16)
Finally, in analogy with Eq. 2.12 for fixed-point parameters, Eq. 2.13 for the standard gas-phase properties can be complemented by

$$y_{aq}^{\star} = \hat{O}_Y \rho_{aq}^{\star} \quad \text{and} \quad \Delta Y_{hyd}^{\star} = \hat{O}_Y \Delta G_{hyd}^{\star},$$

(2.17)

for the standard aqueous partial molar and hydration properties, respectively.

The strict application of Eqs. 2.15 and 2.16 implies that the operator \( \hat{O}_Y \) acting on \( \rho^{\star} \) introduces corresponding derivatives into the resulting expressions for standard derivative thermodynamic properties. This choice represents the so-called standard variant of the solute standard-state definition.166,174 Adopting this convention leads to a number of counter-intuitive results, which are most easily illustrated by considering the hypothetical case of a non-interacting and volume-free single-particle “ghost” solute. For such a species, all fixed-point and point-to-point parameters evaluate to zero. However, Eqs. 2.15 and 2.16 imply, for example, that the particle has a standard hydration enthalpy \( \Delta H_{hyd}^{\star} \) differing from the mere \(-RT^\circ \) term, although it presents no interaction with the solvent, and that it has a non-vanishing standard aqueous partial molar volume \( v_{aq}^{\star} \) (as well as derivatives \( k_{T,aq}^{\star} \) and \( a_{P,aq}^{\star} \)), although it has been assumed volume-free. The corresponding partial molar entropy \( s_{aq}^{\star} \) also differs from that expected for a free particle within an effective molar volume \((b^2 \rho^{\star})^{-1} \), while the corresponding partial molar heat-capacity \( c_{P,aq}^{\star} \) differs from the expected value of \((3/2)R \). Furthermore, the values of these unexpected additional contributions are related to the expansivity and compressibility of pure water, two properties that appear to be entirely irrelevant in terms of hydration under isobaric and isothermal conditions. Their occurrence is even more surprising considering that the particle does not “feel” the solvent, and should thus be insensitive to its nature beyond the role of \( \rho^{\star} \) in determining the effective molar volume associated with the solute standard state.

These counter-intuitive features can be alleviated by introducing a slightly different definition for the solute standard state, referred to as the density-corrected variant,166,174 in which the derivatives of \( \rho^{\star} \) are omitted in the application of \( \hat{O}_Y \). The standard parameters in this density-corrected variant will be noted with a star superscript, to distinguish them from the corresponding quantities in the standard variant. The two variants are equivalent in terms of all standard gas-phase molar parameters, of the standard chemical potential in the aqueous phase, and of the hydration free energy, i.e.

$$y_g^{\star*} = y_g^{\star}, \quad \mu_{aq}^{\star*} = \mu_{aq}^{\star} \quad \text{and} \quad \Delta G_{hyd}^{\star*} = \Delta G_{hyd}^{\star},$$

(2.18)

However, they differ in terms of the derivative hydration parameters and aqueous partial molar properties, namely \( \Delta H_{hyd}^{\star*}, v_{aq}^{\star*}, c_{P,aq}^{\star*}, k_{T,aq}^{\star*} \) and
In the density-corrected variant, Eq. 2.14 is unaltered, i.e.

\[ y_\phi^\ast = y_\phi + \Delta y_\phi \rightarrow \phi^\ast = y_\phi + \Delta y_\phi \rightarrow \phi \]  \quad (2.19)

while Eq. 2.15 is modified to

\[ y_\phi = y_\phi + \Delta y_\phi \rightarrow \phi^\ast = y_\phi + \Delta y_\phi \rightarrow \phi - R \hat{O}_Y \left\{ T^{-} \ln \left[ \tilde{C}(MT^{-})^{3/2} \frac{1}{b^\phi \rho^\ast} \right] \right\} , \]  \quad (2.20)

where $\rho^\ast$ is a strict constant (no derivatives upon application of $\hat{O}_Y$), the value of which is set to the pure-solvent density at $P^o$ and $T^-$, i.e.

\[ \rho^\ast = \rho^\phi . \]  \quad (2.21)

Accordingly, with the definition of Eq. 2.11 translated to

\[ \Delta Y_\phi^\ast = y_\phi^\ast - y_\phi^\ast = y_\phi^\ast - y_\phi \]  \quad (2.22)

Eq. 2.16 is modified to

\[ \Delta Y_{\phi^\ast} = \Delta Y_{\phi^\ast} + \Delta \Delta Y_{\phi^\ast} \rightarrow \phi^\ast = \Delta Y_{\phi^\ast} + R \hat{O}_Y \left\{ T^{-} \ln \left[ \frac{RT^{-} b^\phi \rho^\ast}{P_o} \right] \right\} . \]  \quad (2.23)

Adopting this alternative convention alleviates the counter-intuitive observations described above in the illustrative context of a single-particle “ghost” solute. In the density-corrected variant, Eqs. 2.20 and 2.23 imply, for example, that the particle has a standard hydration enthalpy $\Delta H_{\phi^\ast}$ exactly equal to $-RT^-$ and that it has a vanishing standard aqueous partial molar volume $v_\phi^\ast$ (as well as volume derivatives $k_{T,aq}^\phi$ and $a_{P,aq}^\phi$). The corresponding partial molar entropy $s_\phi^\ast$ is also the one expected for a free particle within an effective molar volume $(b^\phi \rho^\phi)^{-1}$, and the corresponding partial molar heat-capacity $c_{P,aq}^\phi$ evaluates to $(3/2)R$ as expected. Furthermore, the standard properties in the density-corrected variant no longer depend on the expansivity and compressibility of the pure solvent. These observations are much more intuitive compared to those made in the context of the standard variant. However, they come at the cost of giving up the direct relationship between free energy and derivative thermodynamic properties via the thermodynamic-derivative operators of Eq. 2.2, i.e. in the density-corrected variant, Eq. 2.13 can no longer be complemented by equations of the form of Eq. 2.17.
As explained in Ref. 166, the reason for distinguishing the two above solute standard-state variants is that different types of experiments probing derivative thermodynamic properties may lead, as their natural outcome, to standardized results corresponding to either of the two variants, both simply reported as “standard” in the literature. The differences between quantities reported in the two variants are then given by
\[ \Delta y_{g}^{\Theta_{s} \rightarrow \Theta} = \Delta y_{g}^{\Theta_{s} \rightarrow \Theta} - \Delta y_{g}^{\Theta_{s} \rightarrow \Theta_{*}} = 0, \] (2.24)
\[ \Delta y_{aq}^{\Theta_{s} \rightarrow \Theta} = \Delta y_{aq}^{\Theta_{s} \rightarrow \Theta} - \Delta y_{aq}^{\Theta_{s} \rightarrow \Theta_{*}} = R\hat{O}_{Y} \left\{ T^{-} \ln \left[ \frac{\rho^{\Theta}}{\rho^{*}} \right] \right\}, \] (2.25)
and
\[ \Delta \Delta Y_{hyd}^{\Theta_{s} \rightarrow \Theta} = \Delta \Delta Y_{hyd}^{\Theta_{s} \rightarrow \Theta} - \Delta \Delta Y_{hyd}^{\Theta_{s} \rightarrow \Theta_{*}} = R\hat{O}_{Y} \left\{ T^{-} \ln \left[ \frac{\rho^{\Theta}}{\rho^{*}} \right] \right\}. \] (2.26)

This ambiguity may represent a non-negligible source of error, especially for aqueous partial molar volumes, heat capacities and compressibilities. For example, for the Na\(^+\) and Cl\(^-\) ions, \(v_{aq}^{\Theta_{s}}\), \(c_{P,aq}^{\Theta_{s}}\) and \(k_{T,aq}^{\Theta_{s}}\) are in magnitude on the order of \(10^{-5}\) m\(^3\) mol\(^{-1}\), \(10^2\) J mol\(^{-1}\) K\(^{-1}\) and \(10^{-9}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\), respectively (Section 2.4) to be compared with \(\Delta v_{aq}^{\Theta_{s} \rightarrow \Theta_{*}}\), \(\Delta c_{P,aq}^{\Theta_{s} \rightarrow \Theta_{*}}\) and \(\Delta k_{T,aq}^{\Theta_{s} \rightarrow \Theta_{*}}\) evaluating to \(1.12\cdot10^{-6}\) m\(^3\) mol\(^{-1}\), \(8.39\) J mol\(^{-1}\) K\(^{-1}\) and \(2.88\cdot10^{-10}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\), respectively. In the present work, this difficulty is circumvented by assuming experimental data appropriate for the standard variant in the case of the free energy, enthalpy, entropy and heat capacity, and in the density-corrected variant in the case of the volume, volume-compressibility and volume-expansivity, converting it to point-to-point (hydration parameters) or fixed-point (aqueous partial molar parameters) values, and performing the comparison between simulation and experiment in terms of the latter values only. This assumption concerning the actual nature of literature data tabulated as “standard” is motivated by the type of experiments usually performed to evaluate these parameters, as detailed in Ref. 166 (see pp 171-177 therein).

For convenience, the constants \(\Delta y_{g}^{\Theta_{s} \rightarrow \Theta}\) (Eq. 2.14), \(\Delta y_{g}^{\Theta_{s} \rightarrow \Theta_{*}}\) (Eq. 2.24), \(\Delta y_{aq}^{\Theta_{s} \rightarrow \Theta}\) (Eq. 2.15), \(\Delta y_{aq}^{\Theta_{s} \rightarrow \Theta_{*}}\) (Eq. 2.25), \(\Delta \Delta Y_{hyd}^{\Theta_{s} \rightarrow \Theta}\) (Eq. 2.16) and \(\Delta \Delta Y_{hyd}^{\Theta_{s} \rightarrow \Theta_{*}}\) (Eq. 2.26), required for the interconversion between fixed-point and standard quantities in the standard or density-corrected variants are provided in Table 2.1. The analytical standard (Eqs. 2.5 and 2.6) and fixed-point (Eqs. 2.8 and 2.9) gas-phase parameters, along with experimental estimates for the corresponding hydration and aqueous partial molar parameters, are reported in Table 2.2 for the Na\(^+\) and Cl\(^-\) ions as well as the proton.
Note that the aqueous and hydration parameters reported in this table represent experimentally elusive quantities, and are thus to be considered with some caution.\textsuperscript{166}

Table 2.1: Constants required for standard-state interconversions of thermodynamic properties\textsuperscript{(a)}

| \(m_a\) | Unit | Equation\textsuperscript{(a)} | \(\Delta y_{\text{aq}}^{g \rightarrow \phi}\) | Numerical value\textsuperscript{(a)} | Equation\textsuperscript{(a)} | \(\Delta y_{\text{aq}}^{\phi 
rightarrow \phi}\) | Numerical value\textsuperscript{(a)} |
|---|---|---|---|---|---|---|---|
| \(m_a\) | [kJ mol\(^{-1}\)] | \(\Delta y_{\text{aq}}^{g \rightarrow \phi} = \Delta y_{\text{aq}}^{\phi 
rightarrow \phi} = \Delta y_{\text{aq}}^{\phi \rightarrow \phi} = 0\) | 0 | 0 | 0 | 0 |
| \(b_a\) | [kJ mol\(^{-1}\)] | \(\Delta y_{\text{aq}}^{g \rightarrow \phi} = \Delta y_{\text{aq}}^{\phi 
rightarrow \phi} = \Delta y_{\text{aq}}^{\phi \rightarrow \phi} = 0\) | 0 | 0 | 0 | 0 |
| \(c_a\) | [kJ mol\(^{-1}\) K\(^{-1}\)] | \(\Delta y_{\text{aq}}^{g \rightarrow \phi} = \Delta y_{\text{aq}}^{\phi 
rightarrow \phi} = \Delta y_{\text{aq}}^{\phi \rightarrow \phi} = 0\) | 0 | 0 | 0 | 0 |
| \(e_a\) | [kJ mol\(^{-1}\)] | \(\Delta y_{\text{aq}}^{g \rightarrow \phi} = \Delta y_{\text{aq}}^{\phi 
rightarrow \phi} = \Delta y_{\text{aq}}^{\phi \rightarrow \phi} = 0\) | 0 | 0 | 0 | 0 |
| \(k_a\) | [kJ mol\(^{-1}\) K\(^{-1}\)] | \(\Delta y_{\text{aq}}^{g \rightarrow \phi} = \Delta y_{\text{aq}}^{\phi 
rightarrow \phi} = \Delta y_{\text{aq}}^{\phi \rightarrow \phi} = 0\) | 0 | 0 | 0 | 0 |

(a) The reported quantities \(\Delta y_{\text{aq}}^{g \rightarrow \phi}\) (Eq. 2.14), \(\Delta y_{\text{aq}}^{\phi 
rightarrow \phi}\) (Eq. 2.24), \(\Delta y_{\text{aq}}^{\phi \rightarrow \phi}\) (Eq. 2.15), \(\Delta y_{\text{aq}}^{\phi \rightarrow \phi}\) (Eq. 2.25), \(\Delta y_{\text{aq}}^{\phi \rightarrow \phi}\) (Eq. 2.16) and \(\Delta y_{\text{aq}}^{\phi \rightarrow \phi}\) (Eq. 2.26) permute the interconversion between fixed-point or/and point-to-point (\(=\) superscript) and standard (\(\neq\) superscript) quantities in the standard (no star superscript) and density-corrected (star superscript) variants. Expressions and numerical values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\), volume \(V\), heat capacity \(C_P\), volume-compressibility \(K_T\) and volume-expansivity \(A_P\). The numerical values refer to the standard pressure \(P_0 = 1\) bar, molarity \(b^* = 1\) mol kg\(^{-1}\) and temperature \(T^* = 298.15\) K, and to water as a solvent; (b) Eqs. 2.14 and 2.19 with \(\tilde{C} = R^{-1}C\) and \(C\) defined by Eq. 2.7; (c) The factor \(f_M\) is defined as \(f_M = \ln(M/M^0)\), where \(M\) is the molar mass of the species and \(M^0 = 1\) kg mol\(^{-1}\); (d) Eq. 2.24; (e) Eq. 2.20 with \(\tilde{C} = R^{-1}C\) and \(C\) defined by Eq. 2.7, and \(\rho^*\) defined by Eq. 2.21 as a strict constant equal to \(\rho^0\); (f) The entries for the free energy and entropy are based on \(\rho^*\) set to the water density \(\rho^0\) as reported in Table 2.3; (g) Eq. 2.25; (h) Although \(\ln\rho^0\) should formally rather be written \(\ln[\rho^0/P_0]\), to avoid a dependence on the arbitrary unit selected for \(\rho^0\), this dependence has no influence on the derivatives of the logarithm involved in these expressions; (i) Based on the water density \(\rho^0\), as well as its pressure or/and temperature derivatives \(\partial_P\rho^0\), \(\partial_T\rho^0\), \(\partial_P\rho^0\), \(\partial_T\rho^0\) and \(\partial_P\rho^0\) as reported in Table 2.3 (the values for the volume-compressibility and volume-expansivity reported in Table IV of Ref. 174 and Table 4.1 of Ref. 166 differ slightly due to the use of different \(\rho(P,T)\) data for water); (j) Eq. 2.23 with \(\rho^*\) defined by Eq. 2.21 as a strict constant equal to \(\rho^0\); (k) Eq. 2.26.
### Table 2.2: Standard and fixed-point or point-to-point thermodynamic parameters for the Na\(^+\) and Cl\(^-\) ions as well as the proton\(^{(a)}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ion</th>
<th>(\Delta H^*_T)</th>
<th>(\Delta P^*_T)</th>
<th>(\Delta V^*_T)</th>
<th>(\Delta G^*_T)</th>
<th>(\Delta V^*_P)</th>
<th>(\Delta G^*_P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>Na(^+)</td>
<td>1513.12 (\pm) 3</td>
<td>153.10 (\pm) 4</td>
<td>109.75 (\pm) 6</td>
<td>20.79 (\pm) 7</td>
<td>4.46 (\pm) 1.0</td>
<td>4.46 (\pm) 1.0</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>Cl(^-)</td>
<td>-239.77 (\pm) 3</td>
<td>-230.08 (\pm) 3</td>
<td>153.35 (\pm) 6</td>
<td>20.79 (\pm) 7</td>
<td>4.46 (\pm) 1.0</td>
<td>4.46 (\pm) 1.0</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>H(^+)</td>
<td>0.6083 (\pm) 0.0011</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>Na(^+)</td>
<td>115.25 (\pm) 1</td>
<td>153.82 (\pm) 4</td>
<td>57.21 (\pm) 6</td>
<td>-22.07 (\pm) 7</td>
<td>-22.07 (\pm) 7</td>
<td>-22.07 (\pm) 7</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>Cl(^-)</td>
<td>-544.17 (\pm) 6</td>
<td>-561.17 (\pm) 6</td>
<td>77.94 (\pm) 6</td>
<td>-23.33 (\pm) 7</td>
<td>-23.33 (\pm) 7</td>
<td>-23.33 (\pm) 7</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>H(^+)</td>
<td>413.32 (\pm) 4</td>
<td>393.86 (\pm) 4</td>
<td>-22.04 (\pm) 6</td>
<td>-66.90 (\pm) 8</td>
<td>-66.90 (\pm) 8</td>
<td>-66.90 (\pm) 8</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>Na(^+)</td>
<td>115.21 (\pm) 1</td>
<td>149.91 (\pm) 4</td>
<td>-76.40 (\pm) 6</td>
<td>-85.87 (\pm) 8</td>
<td>-85.87 (\pm) 8</td>
<td>-85.87 (\pm) 8</td>
</tr>
<tr>
<td>(\Delta g_{aq}^*)</td>
<td>Cl(^-)</td>
<td>-512.80 (\pm) 6</td>
<td>-565.08 (\pm) 6</td>
<td>-41.09 (\pm) 6</td>
<td>-81.90 (\pm) 8</td>
<td>-81.90 (\pm) 8</td>
<td>-81.90 (\pm) 8</td>
</tr>
<tr>
<td>(\Delta Y_{aq}^\infty)</td>
<td>H(^+)</td>
<td>-1109.78 (\pm) 3</td>
<td>-1138.14 (\pm) 3</td>
<td>-96.62 (\pm) 10</td>
<td>-96.62 (\pm) 10</td>
<td>-96.62 (\pm) 10</td>
<td>-96.62 (\pm) 10</td>
</tr>
<tr>
<td>(\Delta Y_{aq}^\infty)</td>
<td>Na(^+)</td>
<td>-437.41 (\pm) 3</td>
<td>-410.29 (\pm) 3</td>
<td>-76.40 (\pm) 4</td>
<td>-85.87 (\pm) 8</td>
<td>-85.87 (\pm) 8</td>
<td>-85.87 (\pm) 8</td>
</tr>
<tr>
<td>(\Delta Y_{aq}^\infty)</td>
<td>Cl(^-)</td>
<td>-316.53 (\pm) 3</td>
<td>-328.80 (\pm) 3</td>
<td>-41.09 (\pm) 4</td>
<td>-81.90 (\pm) 8</td>
<td>-81.90 (\pm) 8</td>
<td>-81.90 (\pm) 8</td>
</tr>
</tbody>
</table>

\(\text{(a)}\) The parameters reported are the standard \((y_{aq}^*\)) and fixed-point \((y_{aq}^*\)) molar parameters of the gas-phase ions, the standard \((y_{aq}^*\)) and fixed-point \((y_{aq}^*\)) partial molar parameters of the aqueous ions, and the point-to-point \((\Delta Y_{aq}^\infty\)) hydration parameters of the ions. The aqueous ion parameters are reported in either the standard (no star superscript) or the density-corrected (star superscript) variants. Values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\), volume \(V\), heat capacity \(C_P\), volume-compressibility \(K_T\) and volume-expansivity \(A_P\). For the free energy \(G\) and enthalpy \(H\), formation parameters \((\Delta f_{aq}^*\), \(\Delta f_{aq}^*\)) are reported instead of partial molar parameters. The standard hydration parameters of the ions \((\Delta Y_{aq}^\infty\) or \(\Delta Y_{aq}^\infty\)) are not explicitly reported, and can be easily deduced as \((\Delta Y_{aq}^\infty = \Delta Y_{aq}^\infty - \Delta Y_{aq}^\infty)\). The data refers to the standard pressure \(P^0 = 1\) bar, molality \(b^0 = 1\) mol kg\(^{-1}\) and temperature \(T^0 = 298.15\) K. The formation parameters also refer to an electron standard state defined as the ideal electron gas at temperature \(T^0\) with properties calculated according to Fermi-Dirac statistics.
2.2.3 Calculation of derivative thermodynamic properties

Two alternative methods can be employed to evaluate derivative single-ion hydration and aqueous partial molar properties based on atomistic simulations, referred to here as Methods I and II. In both cases, the outcome of the calculations are raw point-to-point (hydration parameters) or fixed-point (aqueous partial molar parameters) quantities. These must be appropriately corrected for finite-system and approximate-electrostatics errors, as described in Section 2.2.4, where two variants in the application of the correction scheme, referred to as Variants A and B, are distinguished. The comparison with experiment is then performed directly in terms of these corrected point-to-point or fixed-point parameters.

Method I

In Method I, the point-to-point hydration enthalpy \( \Delta H_{\text{hyd}}^{\text{pt}} \) and the fixed-point aqueous partial molar volume \( \nu_{\text{aq}}^{\text{fix}} \) are calculated based on the variations of the average potential energy and box volume, respectively, upon inserting the ion into a pure-liquid sample. Finite-difference analysis of the corresponding values calculated at different pressures \( P \) and temperatures \( T \) also permits the determination of \( c_{\text{P,aq}}^{\text{pt}} \), \( k_{\text{T,aq}}^{\text{pt}} \), \( a_{\text{P,aq}}^{\text{pt}} \). However, the method does not give access to \( \Delta G_{\text{hyd}}^{\text{pt}} \) and \( s_{\text{aq}}^{\text{fix}} \). This approach requires two isobaric-isothermal simulations per \( P,T \)-point, both involving the same number of water molecules and differing only by the presence or absence of the ion.

Given the variation \( \Delta U \) of the average potential energy of the system and the corresponding variation \( \Delta V \) of the average box volume, both assumed to be expressed on a per mole basis, \( \Delta H_{\text{hyd}}^{\text{pt}} \) can be evaluated as

\[
\Delta H_{\text{hyd}}^{\text{pt}} = \Delta U + P \Delta V \approx \Delta U .
\]  

(2.27)

This equation is only valid for monoatomic species, for which the internal potential energy is zero within the classical model used in the simulations. For polyatomic species, it would have to be corrected by the average intramolecular potential energy of the species in the gas phase. The equation holds irrespective of whether the species is free or positionally constrained during the simulation. However, if the variation \( \Delta E \) of the total energy is considered instead of the variation \( \Delta U \) of the potential energy, a term \( -(3/2)RT \) must be added to compensate for the average translational kinetic energy of the monoatomic species, unless it is positionally constrained.
2.2 Theory

A calculation based on $\Delta U$ was preferred in the present work, because it removes the unnecessary noise associated with the kinetic energy fluctuations of the system. The connection between $v_{aq}^+$ and $\Delta V$ is given by

$$v_{aq}^+ = \Delta V .$$  \hspace{1cm} (2.28)

This equation is generally valid, i.e. for monoatomic as well as polyatomic species. The neglect of the term $P\Delta V$ in Eq. 2.27 is justified by the very small partial molar volumes of the species considered here. For example, noting that $v_{aq}^+$ is on the order of $\pm 3 \cdot 10^{-5}$ m$^3$ mol$^{-1}$ for Na$^+$ and Cl$^-$ in water (Section 2.4), $P\Delta V$ is on the order of $\pm 3 \cdot 10^{-3}$ kJ mol$^{-1}$ at $P^o$ and $T^-$, which is entirely negligible given the statistical error of the calculations.

Finite-difference expressions for $c_{P,aq}^+$, $k_{T,aq}^+$ and $a_{P,aq}^+$ can be deduced from Eq. 2.12 given the forms of the thermodynamic-derivative operators $\hat{O}_Y$ in Eq. 2.2, namely

$$c_{P,aq}^+ \approx \tilde{\partial}_{\Delta T} \Delta H_{hyd}^+ , \quad k_{T,aq}^+ \approx -\tilde{\partial}_{\Delta P} v_{aq}^+ \quad \text{and} \quad a_{P,aq}^+ \approx \tilde{\partial}_{\Delta T} v_{aq}^+ ,$$  \hspace{1cm} (2.29)

where $\tilde{\partial}_{\Delta Q}$ represents a first-order finite-difference operator for property $Q$ based on an interval $\Delta Q$. For example, for a centered finite difference, one has

$$\tilde{\partial}_{\Delta Q} f(Q,R) = (\Delta Q)^{-1} \left[ f(Q + \Delta Q/2,R) - f(Q - \Delta Q/2,R) \right].$$  \hspace{1cm} (2.30)

Note that the expression for $c_{P,aq}^+$ in Eq. 2.29 relies on the Gibbs-Helmholtz alternative of $\hat{O}_H$ in Eq. 2.2. This expression is also restricted to monoatomic species (or fully-rigid polyatomic species) because it assumes that $c_{P,g}^+$ = 0, while those for $k_{T,aq}^+$ and $a_{P,aq}^+$ are generally valid, i.e. for monoatomic as well as rigid or flexible polyatomic species. The neglect of the temperature derivative of the term $P\Delta V$ in Eq. 2.27 when evaluating $c_{P,aq}^+$ is justified by the very small partial molar volume-expansivity of the species considered here. For example, noting that $a_{P,aq}^+$ is on the order of $\pm 2 \cdot 10^{-7}$ m$^3$ mol$^{-1}$ K$^{-1}$ for Na$^+$ and Cl$^-$ in water (Section 2.4), $P\partial_T \Delta V$ is on the order of $\pm 2 \cdot 10^{-2}$ J mol$^{-1}$ K$^{-1}$ at $P^o$ and $T^-$, which is entirely negligible given the statistical error of the calculations.

Eqs. 2.27 and 2.28 are only exact in the limit of infinite dilution, i.e. considering a system including the entire solvation range where the influence of the ion on the potential energy (solvent polarization) and volume (excluded volume, electrostriction) is significant. In practice, this condition is not satisfied for the system sizes considered here (Section 2.3), and corrections for finite-system (as well as approximate-electrostatics) errors must be applied to
the raw results, as described in Section 2.2.4. In Variant A (Section 2.2.4), Eq. 2.29 is immediately applied to the raw results of Eqs. 2.27 and 2.28, and the five types of raw variables are corrected subsequently. In Variant B (Section 2.2.4), the raw results of Eqs. 2.27 and 2.28 are corrected prior to application of Eq. 2.29, directly leading to the five types of variables in a corrected form.

**Method II**

In Method II, the point-to-point hydration free energy $\Delta G_{\text{hyd}}$ is calculated based on the reversible work associated with the insertion of the ion into a pure-liquid sample. Finite-difference analysis of the corresponding values calculated at different pressures $P$ and temperatures $T$ also permits in principle the determination of $\Delta H_{\text{hyd}}$, $s_{aq}$, $v_{aq}$, $c_{P,aq}$, $k_{T,aq}$ and $a_{P,aq}$, i.e. in contrast to method I, the free energy and entropy can also be evaluated. This approach requires two sets of isobaric-isothermal simulations per $P,T$-point, accounting for the progressive growth of a neutral ion-sized cavity, referred to as the cavitation process, and the progressive increase of the cavity charge from zero to the full ionic charge, referred to as the charging process.

Given the reversible work $\Delta G_{\text{cav}}$ of cavitation and the reversible work $\Delta G_{\text{chg}}$ of charging, both assumed to be expressed on a per mole basis, $\Delta G_{\text{hyd}}$ can be evaluated as

$$\Delta G_{\text{hyd}} = \Delta G_{\text{cav}} + \Delta G_{\text{chg}} \quad .$$

(2.31)

This equation is only valid for monoatomic species, for which the internal potential energy is zero within the classical model used in the simulations. For polyatomic species, it would have to be corrected by the corresponding reversible work in the gas phase. The equation holds irrespective of whether the species is free or positionally constrained during the simulation because the solvation process involves no change in the atomic mass.

Finite-difference expressions for $\Delta H_{\text{hyd}}$, $s_{aq}$, $v_{aq}$, $c_{P,aq}$, $k_{T,aq}$ and $a_{P,aq}$ can be deduced from Eq. 2.12 given the forms of the thermodynamic-derivative operators $\hat{O}_Y$ in Eq. 2.2, namely

$$\Delta H_{\text{hyd}} \approx (1 - T\tilde{\partial}_T)\Delta G_{\text{hyd}} \quad s_{aq} \approx -\tilde{\partial}_T \Delta G_{\text{hyd}}$$

$$v_{aq} \approx \tilde{\partial}_P \Delta G_{\text{hyd}} \quad c_{P,aq} \approx -T\tilde{\partial}_T^2 \Delta G_{\text{hyd}}$$

$$k_{T,aq} \approx -\tilde{\partial}_T^2 \Delta G_{\text{hyd}} \quad a_{P,aq} \approx \tilde{\partial}_P^2 \Delta G_{\text{hyd}}$$

(2.32)

where $\tilde{\partial}_Q$ represents a first-order finite-difference operator for property $Q$ based on an interval $\Delta Q$ (e.g. Eq. 2.30), while $\tilde{\partial}_Q^2$ and $\tilde{\partial}_Q^2 R$ represent corresponding second-order (self or cross) operators. For example, for centered
finite differences, one has
\[ \partial^2_{\Delta Q} f(Q, R) = \left( \frac{\Delta Q}{2} \right)^{-2} \left[ f(Q - \Delta Q/2, R) - 2f(Q, R) + f(Q + \Delta Q/2, R) \right] \]
and
\[ \partial^2_{\Delta Q, \Delta R} f(Q, R) = \left( \frac{\Delta Q\Delta R}{4} \right)^{-1} \left[ f(Q - \Delta Q/2, R - \Delta R/2) + f(Q + \Delta Q/2, R + \Delta R/2) 
- f(Q + \Delta Q/2, R - \Delta R/2) - f(Q - \Delta Q/2, R + \Delta R/2) \right]. \]

The expressions for \( \Delta H_{\text{hyd}}^\text{aq}, v_{\text{aq}}^\text{aq}, k_{T,aq}^\text{aq} \) and \( a_{P,aq}^\text{aq} \) in Eq. 2.32 are generally valid, i.e. for monoatomic as well as rigid or flexible polyatomic species, the omission of the terms \( v_g^\text{aq}, k_{T,g}^\text{aq} \) and \( a_{P,g}^\text{aq} \) being justified by Eq. 2.8, although the calculation of \( \Delta G_{\text{hyd}}^\text{aq} \) itself based on Eq. 2.31 is restricted to monoatomic species. The expressions for \( s_{aq}^\text{aq} \) and \( c_{P,aq}^\text{aq} \), however, assume a closed-shell monoatomic species exempt of thermally-accessible electronically excited states at \( T^- \), in which case of the omission of the terms \( s_g^\text{aq} \) and \( c_{P,g}^\text{aq} \) is justified by Eq. 2.9.

Eq. 2.31 is only exact in the limit of infinite dilution, i.e. considering a system including the entire solvation range where the influence of the ion on the potential energy (solvent polarization) and volume (excluded volume, electrostriction) is significant. In practice, this condition is not satisfied for system sizes considered here (Section 2.3), and corrections for finite-system effects (as well as approximate-electrostatics) errors must applied to the raw results, as described in Section 2.2.4. In Variant A (Section 2.2.4), Eq. 2.32 is immediately applied to the raw results of Eq. 2.31, and the seven types of raw variables are corrected subsequently. In Variant B (Section 2.2.4), the raw results of Eq. 2.31 are corrected prior to application of Eq. 2.32, directly leading to the seven types of variables in a corrected form.

### 2.2.4 Correction scheme

The raw thermodynamic parameters calculated based on atomistic simulations using Methods I or II (Section 2.2.3) are affected by errors related to the finite size of the simulated systems and to the approximate treatment of electrostatic interactions in the simulations. However, the application of an appropriate correction scheme permits to convert the methodology-dependent raw data into methodology-independent results.\(^\text{61,166,172–174}\)
corrected values are then exclusively characteristic of the underlying molecular model, and correspond to the idealized situation of an infinite bulk phase exempt of surface polarization, in which electrostatic interactions are exactly Coulombic. In terms of the point-to-point hydration free energy, the application of this correction scheme can be written

\[ \Delta G_{\text{hyd}}^{\text{cor}} = \Delta G_{\text{hyd, raw}}^{\text{cor}} + \Delta G_{\text{cor}}^{\text{cor}}. \] (2.35)

The correction term \( \Delta G_{\text{cor}}^{\text{cor}} \) must account for (see Refs. 61, 166, 173 and references therein):

A. The deviation of the solvent polarization around the ion relative to the polarization in an ideal Coulombic system and the incomplete or/and inexact interaction of the ion with the polarized solvent, consequence of possible approximations made in the representation of electrostatic interactions during the simulation (e.g. non-Coulombic interactions involving cutoff truncation).

B. The deviation of the solvent polarization around the ion relative to the polarization in an ideal macroscopic system, consequence of the use of a finite (microscopic and possibly periodic) system during the simulation (e.g. computational box simulated under periodic boundary conditions).

C. The deviation of the solvent-generated electrostatic potential at the ion site as calculated from the simulated trajectory relative to the “correct” electrostatic potential, consequence of the possible application of an inappropriate summation scheme (along with a possibly non-Coulombic potential function) for the contributions of individual solvent atomic charges to this potential, as well as of the possible presence of a constant offset in this potential (e.g. related to the existence of an interfacial potential at the ionic surface along with a constraint of vanishing average potential over the computational box in lattice-sum schemes).

D. An inaccurate dielectric permittivity of the employed solvent model.

In the context of monoatomic ions, the quantity \( \Delta G_{\text{cor}}^{\text{cor}} \) in Eq. 2.35 generally consists of six terms,\(^{174} \) which can be evaluated (either analytically, numerically, or using empirically fitted equations) based on continuum-electrostatics or analytical models. All simulations reported in the present work are performed under periodic boundary conditions based on cubic computational boxes. Except those involved in the calculation of \( \Delta G_{\text{cav}} \) (Section 2.3), for which no correction is required,\(^{61} \) they rely on a lattice-summation (LS) electrostatic scheme.\(^{352} \) In the specific case of the LS scheme with a cubic box, and considering a water model such as the SPC model\(^{176} \) characterized by a
single van der Waals interaction site, $\Delta G_{cor}^{\text{vdW}}$ can be written based on Ref. 174 (see Eq. (39) therein)

$$
\Delta G_{cor}^{\text{vdW}} = \Delta G_B (|q_I|, R_I, M_W, N_W, \rho', \epsilon') + \Delta G_{C_1} (|q_I|, R_I, M_W, N_W, \rho', \gamma') + \Delta G_{C_2} (|q_I|, R_I, M_W, N_W, \rho', \chi', \tilde{\chi}') + \Delta G_D (|q_I|, R_I, \epsilon', \epsilon),
$$

(2.36)

Note the absence of type-A correction term for the LS scheme, which is formally Coulombic in the limit of infinite box size. The parameters involved in Eq. 2.36 are: the full ionic charge $q_I$, the effective ionic radius $R_I$, the molar mass $M_W$ of water, the number $N_W$ of water molecules in the computational box, the density $\rho'$ of the employed water model, the permittivity $\epsilon'$ of the water model, the quadrupole-moment trace $\gamma'$ of the water model relative to its single van der Waals interaction site, the air-liquid interfacial potential $\chi'$ (air-to-liquid direction) of the water model and its dependence $\tilde{\chi}'$ on the inverse curvature radius for a concave interface, as well as the experimental solvent permittivity $\epsilon$. The values for the parameters $R_I, \rho', \epsilon', \chi', \tilde{\chi}'$ and $\epsilon$ must be selected as appropriate for $P^0 = 1$ bar and $T^- = 298.15$ K, which define the reference state point for $\Delta G_{hyd}^{\text{vdW}}$. Note that the correction is to be applied to raw charging free energies excluding the contribution of the LS self-energy term.

The parameter dependences of the terms in Eq. 2.36 have been slightly altered compared to Ref. 174, namely, the box-edge length $L$ has been expanded in terms of the equilibrium density $\rho'$ of the water model (neglecting the effect of electrostriction), the molar mass $M_W$ of water, the number of water molecules $N_W$ in the box and the effective ionic radius $R_I$, as

$$
L = \left( \frac{N_W M_W}{N_A \rho'} + \frac{4\pi}{3} R_I^3 \right)^{1/3}.
$$

(2.37)

The exact functional form and the physical interpretation of the terms occurring in Eq. 2.36, as well as the procedure to evaluate them, are described in details elsewhere, and this information will not be repeated here. The equations used in the present work are merely reported below, for the sake of reproducibility.

The term $\Delta G_B$ is given by

$$
\Delta G_B = (8\pi \epsilon_o)^{-1} N_A q_I^2 \left[ 1 - (\epsilon')^{-1} \right] L^{-1} \left[ \alpha_{LS} + \frac{4\pi}{3} \left( \frac{R_I}{L} \right)^2 - \frac{16\pi^2}{45} \left( \frac{R_I}{L} \right)^5 \right],
$$

(2.38)
where $\alpha_{LS} \approx -2.837297$ is the LS self-term constant for a cubic box, and $L$ is given by Eq. 2.37. The term $\Delta G_C$ is given by

$$\Delta G_C = -N_A q_l \left( 1 - \frac{4\pi R_l^3}{3L^3} \right) \xi',$$

(2.39)

where $L$ is given by Eq. 2.37 and $\xi'$ is the exclusion potential of the solvent model at density $\rho'$, defined as

$$\xi' = (6\epsilon_o)^{-1} N_A M_W^{-1} \rho' \gamma'',$$

(2.40)

$\epsilon_o$ being the permittivity of vacuum. Inserting Eqs. 2.37 and 2.40, Eq. 2.39 can be simplified to

$$\Delta G_C = -N_A (6\epsilon_o)^{-1} N_W \gamma' q_l L^{-3}.$$

(2.41)

The term $\Delta G_C^2$ is given by

$$\Delta G_C^2 = -N_A q_l \frac{4\pi R_l^3}{3L^3} \left( \chi' + \tilde{\chi}' - R_l^{-1} \right),$$

(2.42)

where $L$ is given by Eq. 2.37. Finally, the term $\Delta G_D$ is given by

$$\Delta G_D = (8\pi\epsilon_o)^{-1} N_A q_l^2 \left[ \epsilon^{-1} - (\epsilon')^{-1} \right] R_l^{-1}.$$

(2.43)

The correction scheme of Eq. 2.35 can be generalized to derivative hydration parameters and aqueous partial molar properties, as

$$\Delta Y_{hyd}^{\text{cor}} = \Delta Y_{hyd,raw}^{\text{cor}} + \Delta Y_{cor}^{\text{hyd}}$$

(2.44)

and

$$y_{aq}^{\text{cor}} = y_{aq,raw}^{\text{cor}} + \Delta Y_{cor}^{\text{aq}},$$

(2.45)

respectively. The correction term $\Delta Y_{cor}^{\text{aq}}$ is the same in both cases considering that there is no error on the gas-phase properties. These generalized correction terms are given by

$$\Delta Y_{cor}^{\text{cor}} = \hat{O}_Y \Delta G_{cor}^{\text{cor}}.$$ 

(2.46)

In principle, $\Delta Y_{cor}^{\text{cor}}$ can be formulated analytically for all thermodynamic parameters considered here, based on explicit differentiation of the terms contributing to Eq. 2.36 as defined by Eqs. 2.37, 2.38, 2.41, 2.42 and 2.43. The resulting expressions are, however, relatively cumbersome, especially for second derivatives. For this reason, in the present work, the quantities
\( \Delta Y_{\text{cor}} \) are determined in practice by finite-difference analysis of these terms based on very small variations of \( P \) and \( T \), the corresponding intervals being sufficiently small for the resulting estimates to be considered indistinguishable from the analytical ones, yet sufficiently large to prevent the occurrence of numerical errors (Section 2.3.4). Irrespective whether an analytical or a numerical approach is selected, the evaluation of the correction terms \( \Delta Y_{\text{cor}} \) requires knowledge of the first and second (self and cross) derivatives of the involved parameters \( (R_I, \rho', \epsilon', \chi', \tilde{\chi}_\cdot \text{ and } \epsilon) \) with respect to temperature and pressure at \( P^0 \) and \( T^- \). Note that \( M_W \) and \( \gamma' \) are constants, independent of pressure and temperature.

For both Methods I and II (Section 2.2.3), two alternative variants are considered for the application of the correction scheme to derivative thermodynamic properties, referred to here as Variants A and B.

**Variant A**

In Variant A, the raw values from the simulations are corrected after application of finite difference in pressure or/and temperature, based on correction terms specifically designed for derivative parameters at \( P^0 \) and \( T^- \) according to Eq. 2.46. For Method I (Section 2.2.3), this implies that the five raw quantities \( \Delta H_{\text{hyd,raw}}^\pm, \, v_{\text{hyd,raw}}^\pm, \, c_p_{\text{hyd,raw}}^\pm, \, k_T_{\text{hyd,raw}}^\pm \) and \( a_{P_{\text{hyd,raw}}}^\pm \) are calculated by means of Eqs. 2.27-2.29, and subsequently corrected using Eqs. 2.44 or 2.45. For Method II (Section 2.2.3), this implies that the seven raw quantities \( \Delta G_{\text{hyd,raw}}^\pm, \Delta H_{\text{hyd,raw}}^\pm, \, s_{\text{hyd,raw}}^\pm, \, v_{\text{hyd,raw}}^\pm, \, c_p_{\text{hyd,raw}}^\pm, \, k_T_{\text{hyd,raw}}^\pm \) and \( a_{P_{\text{hyd,raw}}}^\pm \) are calculated by means of Eqs. 2.31 and 2.32, and subsequently corrected using Eqs. 2.44 or 2.45. Because only the correction terms \( \Delta Y_{\text{cor}}^\pm \) appropriate for \( P^0 \) and \( T^- \) are employed, Variant A only requires knowledge of the first and second (self and cross) derivatives of the involved parameters \( (R_I, \rho', \epsilon', \chi', \tilde{\chi}_\cdot \text{ and } \epsilon) \) with respect to temperature and pressure at \( P^0 \) and \( T^- \), but not of their full functional dependence on \( P \) and \( T \). Arguments in favor of Variant A are that it requires less information on the input parameters of the correction scheme (derivatives at \( P^0 \) and \( T^- \), rather than full functional dependence on \( P \) and \( T \)), and that it involves an exact evaluation of the correction terms at \( P^0 \) and \( T^- \) (rather than a finite-difference approximation to these terms).

**Variant B**

In Variant B, the raw values from the simulations are corrected prior to application of finite difference in pressure or/and temperature, based on cor-
responding correction terms appropriate for the different simulation pressures and temperatures $P$ and $T$ considered, *i.e.* according to analogs of Eq. 2.46 at $P, T$-points differing from $P^\circ$ and $T^-$. For Method I (Section 2.2.3), this implies that the raw quantities $\Delta H_{\text{hyd, raw}}(P, T)$ and $v_{aq, \text{raw}}(P, T)$ are calculated by means of equations analogous to Eqs. 2.27 and 2.28 at all required $P, T$-points, corrected to $\Delta H_{\text{hyd}}(P, T)$ and $v_{aq}(P, T)$ by equations analogous to Eqs. 2.44 or 2.45 at the corresponding $P, T$-point, and subsequently entered into the finite-difference scheme of Eq. 2.29. For Method II (Section 2.2.3), this implies that the raw quantity $\Delta G_{\text{hyd, raw}}(P, T)$ is calculated by means of an equation analogous to Eq. 2.31 at all required $P, T$-points, corrected to $\Delta G_{\text{hyd}}(P, T)$ by an equation analogous to Eq. 2.35 at the corresponding $P, T$-point, and subsequently entered into the finite-difference scheme of Eq. 2.32. Because correction terms $\Delta Y_{\text{cor}}(P, T)$ at different $P, T$-points are employed, Variant B requires knowledge of the full functional dependence of the involved parameters ($R_I$, $\rho'$, $\epsilon'$, $\chi'$, $\tilde{\chi}'$ and $\epsilon$) with respect to temperature and pressure. Compared to Variant A, this variant thus requires more information on the input parameters of the correction scheme (full functional dependence on $P$ and $T$, rather than only derivatives at $P^\circ$ and $T^-$), and involves an approximate evaluation of the correction terms at $P^\circ$ and $T^-$ (finite-difference approximation to these terms). However, an argument in favor of Variant B is that the finite-difference errors (effect of non-linear contributions) committed in the evaluation of the raw derivative parameters may be partially cancelled out by corresponding finite-difference errors in the evaluation of the correction terms.

### 2.3 Computational details

#### 2.3.1 Simulation parameters

All MD simulations required for the application of Method I (Section 2.2.3) and for the evaluation of the charging component in Method II (Section 2.2.3) were carried out using a modified version of the GROMOS96 program$^{39,353}$ implementing the particle-particle particle-mesh (P$^3$M) LS method$^{352,354}$ for calculating the electrostatic interactions. The MD simulations required for the evaluation of the cavitation component in Method II were carried out using the GROMOS05 program$^{355}$ and relied on a molecule-based cutoff-truncation scheme with a Barker-Watts reaction-field (RF) correction$^{356}$ for calculating the electrostatic interactions. The use of the RF scheme in the latter case is motivated by a lower computational cost and by the limited sensitivity of the cavitation free energy on the selected electrostatic scheme.
observed in Ref. 61 (see e.g. Supplementary Material Table S2.Ib therein).

The simulations were performed under periodic boundary conditions based on cubic computational boxes containing 512 (Method I only), 724 or 1024 simple point charge (SPC) water molecules. Depending on the calculation, the box possibly also contained a single Na\(^+\) ion or a single Cl\(^-\) ion. The corresponding ion-solvent Lennard-Jones interaction parameters were taken from set L in Ref. 61 (see Tables II and III therein). The equations of motion were integrated using the leap-frog scheme\(^{357}\) with a timestep of 2 fs. The rigidity of the water molecules was enforced by application of the SHAKE procedure\(^{358}\) with a relative geometric tolerance of 10\(^{-4}\). The center of mass translation of the computational box was removed every 2 ps. All simulations were performed at constant pressure and temperature. The pressure was maintained close to its reference value \(P\) by weak coupling\(^{359}\) to a pressure bath (isotropic coordinate scaling) using a value of 7.513 \(\times\) 10\(^{-4}\) kJ\(^{-1}\) mol nm\(^3\) for the isothermal compressibility of water\(^{360}\) and a coupling time of 0.5 ps. The temperature was maintained close to its reference value \(T\) by weak coupling\(^{359}\) to a temperature bath using a coupling time of 0.1 ps.

The LS scheme (Method I and charging component of Method II) was applied with tinfoil boundary conditions using\(^{352,354,361,362}\) a spherical hat charge-shaping function of width 1.0 nm, a triangular shaped cloud assignment function, a finite-difference scheme of order two and grid-spacings of about 0.08, 0.09 or 0.10 nm for the systems involving 512, 724 and 1024 water molecules, respectively. This choice of parameters led to a ratio of the root-mean square error in the atomic forces to the average norm of these forces\(^{352,362–364}\) of about 0.02-0.05% for the different systems. The LS self-energy term\(^{352,365–368}\) was removed in the calculation of the total potential energy and of the electrostatic potential at the ion site. A pressure correction was applied so as to remove the effect of this self-energy term on the virial, as described by Bogusz et al.\(^{369}\). Although the application of the latter correction is of limited importance when calculating ionic solvation free energies,\(^{61}\) it turns out to be essential for the accurate evaluation of ionic partial molar volumes and their derivatives (data not shown). The non-bonded pairlist was updated every timestep and the Lennard-Jones interactions were truncated at a distance of 1.0 nm.

The RF scheme (cavitation component of Method II) was applied using a molecule-based cutoff distance of 1.4 nm and a reaction-field permittivity of 66.6 as appropriate\(^{370}\) for the SPC water model. The non-bonded pairlist was updated every timestep and the Lennard-Jones interactions were truncated at a distance of 1.4 nm. Only computational boxes containing 724 or 1024 water molecules were considered here, the box containing 512 molecules being
too small for the selected cutoff distance.

For the estimation of derivative properties by finite difference, pressure increments of 5000 bar and temperature increments of 25 K (Method I) or 12.5 K (Method II), respectively, were employed. The choice of this relatively large pressure increments is motivated by the magnitude of the ionic partial molar volumes, on the order of $10^{-5}$ m$^3$ mol$^{-1}$ (Section 4), requiring a pressure change of 1000 bar = $10^8$ J m$^{-3}$ for a change in free energy of about 1 kJ mol$^{-1}$, i.e. measurable with a sufficient precision in the present calculations.

The experimental data against which the calculation results were compared (in terms of point-to-point or fixed-point parameters) is summarized in Table 2.2. The three following sections provide further details on the application of Method I, of Method II, and of the correction scheme within the two methods.

### 2.3.2 Method I

The application of Method I relied on five sets of nine calculations, each set involving simulations of the water box (512, 724 or 1024 molecules) either in the absence or in the presence of one ion (Na$^+$ or Cl$^-$). The five sets corresponded to a pressure $P = 1$ bar along with temperatures $T$ of 273.15, 298.15 or 323.15 K, and to a temperature $T = 298.15$ K along with pressures $P$ of 1, 5000 or 10000 bar. The 45 simulations, relying on LS electrostatics, were carried out for a period of 30 ns (at 5000 bar and 10000 bar) or 60 ns (at 1 bar) after 0.1 ns equilibration. The total potential energy (exempt of LS self-energy term) was written to file every 0.2 ps and the box volume every 2 fs.

The raw point-to-point hydration enthalpy $\Delta H_{\text{hyd}}$ and fixed-point aqueous partial molar volume $v_{\text{aq}}$ were calculated according to Eqs. 2.27 and 2.28, respectively, based on the simulations at 1 bar and 298.15 K. Errors in the individual ensemble averages for the total potential energy and box volume were calculated by block-averaging, and propagated into an error estimate for $\Delta H_{\text{hyd}}$ and $v_{\text{aq}}$, respectively, by taking the square-root of the sum of the squared errors corresponding to the simulations in the absence or presence of the ion. The fixed-point aqueous partial molar heat capacity $c_{P,aq}$ and volume-expansivity $a_{P,aq}$ were calculated according to Eq. 2.29 using a centered finite difference (Eq. 2.30) based on the simulations at 1 bar and 273.15 or 323.15 K. The fixed-point aqueous partial molar volume-compressibility $k_{P,aq}$ was calculated according to Eq. 2.29 based on the simulations at 298.15 K and 1, 5000 or 10000 bar. Since the use of a negative
pressure of −5000 bar led to cavitation in the simulated systems (data not shown), the application of a centered finite difference (Eq. 2.30) was not possible here, and \( \tilde{\partial}_{\Delta P}v_{aq}^{\pm} \) in Eq. 2.29 was evaluated instead as the slope of a straight line fitted to the pressure dependence of the partial molar volume. For the determination of corrected \( c_{P,aq}^{\pm}, k_{T,aq}^{\pm} \) and \( a_{P,aq}^{\pm} \) values, the correction scheme was applied either before or after differentiation in Variants A and B, respectively (there is no such distinction for \( \Delta H_{hyd}^{\pm} \) and \( v_{aq}^{\pm} \)). Errors on \( c_{P,aq}^{\pm} \) and \( a_{P,aq}^{\pm} \) were estimated by taking the square-root of the sum of the squared errors in \( \Delta H_{hyd}^{\pm} \) and \( v_{aq}^{\pm} \), corresponding to the simulations at 273.15 and 323.15 K, and dividing by the temperature interval of 50 K. Errors on \( k_{T,aq}^{\pm} \) were estimated by propagating analytically the individual errors in the partial molar volumes onto the slope of the fitted line representing \( \tilde{\partial}_{\Delta P}v_{aq}^{\pm} \).

2.3.3 Method II

The application of Method II relied on seven sets of eight (or four) calculations, each set involving the evaluation of two free-energy differences (charging and cavitation) for a given ion (Na\(^+\) or Cl\(^-\)) and considering systems involving either 724 or 1024 (or only 1024) water molecules. The seven sets corresponded to a pressure \( P = 1 \) bar along with temperatures \( T \) of 273.15, 285.65, 298.15, 310.65 or 323.15 K, and to a temperature \( T = 298.15 \) K along with pressures \( P \) of 1, 5000 or 10000 bar, the two latter sets only considering systems involving 1024 water molecules.

The 24 calculations of cavitation free energies \( \Delta G_{cav} \), carried out using RF electrostatics, relied on thermodynamic integration\(^{119} \) over the derivative of the Hamiltonian with respect to a scaling parameter \( \lambda \) applied to the ion-water Lennard-Jones interaction of the uncharged ion, from no interaction at \( \lambda = 0 \) to full interaction at \( \lambda = 1 \). The \( \lambda \)-coupling was chosen to be quadratic and involved a soft-core scaling\(^{39} \) with \( \alpha_{LJ} = 0.5 \). The calculation relied on 28 \( \lambda \)-points (spacing of 0.05, decreased to 0.01 between 0.45 and 0.55 due to a sharper variation of the Hamiltonian derivative near 0.5) and the integration was performed using the trapezoidal rule. At each \( \lambda \)-point, the simulation was carried out for a period of 2 ns after 0.1 ns equilibration, and the Hamiltonian derivative was written to file every 0.2 ps. Errors in the individual ensemble averages for the Hamiltonian derivative were calculated by block-averaging,\(^{371} \) and propagated into an error estimate for \( \Delta G_{cav} \) according to the trapezoidal rule.

The 24 calculations of raw charging free energies \( \Delta G_{chg} \), carried out using LS electrostatics, relied on a scheme introduced by Hummer \textit{et al.}\(^{367} \) In this
this scheme, closely related to thermodynamic integration,\textsuperscript{119} the averages and fluctuations of the solvent-generated electrostatic potential at the ion site are recorded for different ionic charge states $q'_I$, from zero to the full ionic charge $q_I$. This information is used to construct an optimal free-energy profile, from which the free-energy difference is then extracted. The calculation relied on three charge states ($q'_I/q_I = 0.0, 0.5, \text{ and } 1.0$), which is sufficient considering that the free-energy profiles are very close to quadratic for monovalent ions in water (Born model\textsuperscript{373}). At each $q'_I$-point, the simulation was carried out for a period of 1 ns after 0.1 ns equilibration, and the instantaneous solvent-generated potential (exempt of LS self-energy term contribution) was written to file every 2 fs. Errors in the individual ensemble averages for the solvent-generated potential were calculated by block-averaging,\textsuperscript{371} and propagated into an error estimate for $\Delta G_{chg}$ according to the trapezoidal rule.

The raw point-to-point hydration free energy $\Delta G_{h yd}^{\text{raw}}$ was calculated according to Eq. 2.31 based on the simulations at 1 bar and 298.15 K. The point-to-point hydration enthalpy $\Delta H_{h yd}^{\text{raw}}$ and fixed-point aqueous partial molar entropy $s_{aq}^{\text{raw}}$ were calculated according to Eq. 2.32 based on the simulations at 1 bar and 273.15, 285.65, 298.15, 310.65 or 323.15 K. For increased accuracy, $\partial_{\Delta T} \Delta G_{h yd}^{\text{raw}}$ in Eq. 2.32 was not evaluated by centered finite difference (Eq. 2.30), but as the slope of a straight line fitted to the temperature dependence of the hydration free energy (separate fits for cavitation and raw (Variant A) or corrected (Variant B) charging contributions, respectively). The fixed-point aqueous partial molar volume $\nu_{aq}^{\text{raw}}$ was calculated according to Eq. 2.32 based on the simulations at 298.15 K and 1, 5000 or 10000 bar. Here also, $\partial_{\Delta P} \Delta G_{h yd}^{\text{raw}}$ in Eq. 2.32 was evaluated as the slope of a straight line fitted to the pressure dependence of the hydration free energy (separate fits for cavitation and raw (Variant A) or corrected (Variant B) charging contributions, respectively). The evaluation of the quantities $\partial_{\Delta P} \Delta G_{h yd}^{\text{raw}}$, $\partial_{\Delta T} \Delta G_{h yd}^{\text{raw}}$ and $\partial_{\Delta P,\Delta T} \Delta G_{h yd}^{\text{raw}}$ in Eq. 2.32 could not be performed with sufficient accuracy for an evaluation of the fixed-point aqueous partial molar heat capacity $c_{P,aq}^{\text{raw}}$, volume-compressibility $\kappa_{T,aq}^{\text{raw}}$ and volume-expansivity $\alpha_{P,aq}^{\text{raw}}$ via Method II. It was nevertheless checked that the estimates for these quantities obtained via Method I are compatible with the curvature of $\Delta G_{h yd}^{\text{raw}}$ as a function of pressure or/and temperature, given the associated uncertainty (Section 2.4.2). For the determination of corrected $\Delta H_{h yd}^{\text{raw}}$, $s_{aq}^{\text{raw}}$ and $\nu_{aq}^{\text{raw}}$ values, the correction scheme was applied either before or after differentiation in Variants A and B, respectively (there is no such distinction for $\Delta G_{h yd}^{\text{raw}}$).

The square root of the sum of squared errors on $\Delta G_{c a v}$ and $\Delta G_{c h g}$ (see
above) was used as estimate for the corresponding error on the hydration free energy. Errors on $s_{aq}$ and $v_{aq}$ were estimated by propagating analytically the individual errors in the free energies onto the slope of the fitted lines representing $\partial_{\Delta T} \Delta G_{hyd}^{\text{aq}}$ or $\partial_{\Delta P} \Delta G_{hyd}^{\text{aq}}$, respectively. Errors on $\Delta H_{hyd}^{\text{aq}}$ were estimated by propagating the free energy errors and the error on $\partial_{\Delta T} \Delta G_{hyd}^{\text{aq}}$ through Eq. 2.32.

2.3.4 Correction scheme

The correction term $\Delta G_{cor}^{\text{aq}}$ defined by Eq. 2.36 is to be applied to raw charging free energies excluding the contribution of the LS self-energy term, and consists of four terms that are functions of nine parameters, based on the expressions of Eqs. 2.37, 2.38, 2.41, 2.42 and 2.43. For those parameters depending on $P$ and $T$, the corresponding values must be selected as appropriate for $P^o = 1$ bar and $T^- = 298.15$ K. The values selected in the present work are: $q_I = +1.0 \text{ or } -1.0$ e for Na$^+$ and Cl$^-$, respectively, $R_I = 0.168$ or 0.246 nm for Na$^+$ and Cl$^-$, respectively (effective radius $R_{\text{mean}}$, based on Ref. 61), $N_W = 512, 724$ or 1024 (depending on the simulated system), $\rho' = 968.2$ kg m$^{-3}$ (SPC water model, based on the simulations carried out in the present work), $\epsilon' = 66.6$ (SPC water model, based on Ref. 370), $\gamma' = 8.2 \cdot 10^{-3} \text{ e-nm}^2$ (SPC water model, analytical, see e.g. Refs. 166,173,174), $\chi' = 0.73 \text{ V}$ (SPC water model, based on Ref. 173), $\chi_-' = -0.11 \text{ V nm}$ (SPC water model, based on Ref. 173), and $\epsilon = 78.4$ (experimental, based on Ref. 374).

The box-edge length $L$ of the system does not enter explicitly into the correction scheme, but is derived here from $R_I$, $N_W$ and $\rho'$ according to Eq. 2.37 (using $M_W = 0.018015$ kg mol$^{-1}$). The resulting values are $L = 2.51, 2.82$ or 3.16 nm for $N_W = 512, 724$ and 1024, respectively. Similarly, the exclusion potential $\xi'$ of the solvent model is not an explicit parameter of the scheme. It is defined by Eq. 2.40 and evaluates to $\xi' = 0.80 \text{ V}$ for the SPC water model at $P^o$ and $T^-$ with an “accurate” electrostatic scheme (the scheme may have a slight influence on $\xi'$ via $\rho'$).

The corresponding correction terms $\Delta Y_{cor}^{\text{aq}}$ for derivative properties at $P^o$ and $T^-$ are defined by Eq. 2.46. As mentioned in Section 2.2.4, the resulting analytical expressions are relatively cumbersome. For this reason, in the present work, the quantities $\Delta Y_{cor}^{\text{aq}}$ were determined by finite-difference analysis of Eq. 2.36 based on very small variations of $P$ and $T$. In practice, a grid-based finite-difference formula was applied considering 121 values of $\Delta G_{cor}$ corresponding to 11 equispaced temperatures ($\Delta T = 3$ K) and pressures ($\Delta P = 0.1$ bar) around $P^o$ and $T^-$. This evaluation requires knowledge of the first and second (self and cross) derivatives of the nine parameters of
the scheme with respect to temperature and pressure at \( P^o \) and \( T^- \) (see further below).

The quantities \( \Delta Y_{cor}^{+} \) at \( P^o \) and \( T^- \), including \( \Delta G_{cor}^{+} \), are sufficient to calculate all correction terms when applying Variant A of the correction scheme (Section 2.2.4). However, when applying Variant B (Section 2.2.4), it is necessary to evaluate \( \Delta G_{cor}(P,T) \) in Eq. 2.36 and \( \Delta Y_{cor}(P,T) \) in Eq. 2.46 at pressures \( P \) and temperatures \( T \) differing from \( P^o \) and \( T^- \). Here also, \( \Delta Y_{cor}(P,T) \) is obtained from a grid-based finite-difference formula,\(^{375}\) applied considering 121 values of \( \Delta G_{cor} \) around the selected \( P,T \)-point. Clearly, however, the evaluation of the quantities \( \Delta Y_{cor}(P,T) \), including \( \Delta G_{cor}(P,T) \), requires knowledge of the full pressure and temperature dependences of the nine parameters of the scheme.

The parameters \( q_I, N_W \) and \( \gamma' \) are independent of pressure and temperature. The parameters \( R_I, \chi' \) and \( \tilde{\chi}' \) were assumed linearly dependent of temperature (slopes from Ref. 174) and independent of pressure. The relative dependences of the density \( \rho' \) and permittivity \( \varepsilon' \) of the solvent model on pressure and temperature were taken from experimentally parametrized curves for water. More specifically, \( \rho'(P,T) \) was calculated as

\[
\rho'(P,T) = \frac{\rho'^{o}}{\rho^o} \rho(P,T), \tag{2.47}
\]

where \( \rho(P,T) \) denotes the experimental density of water at \( P \) and \( T \) according to the IAPWS-95 parametrization,\(^{376}\) and \( \varepsilon'(P,T) \) was calculated as

\[
\varepsilon'(P,T) = \frac{\varepsilon'^{o}}{\varepsilon^o} \varepsilon(P,T), \tag{2.48}
\]

where \( \varepsilon(P,T) \) denotes the experimental permittivity of water at \( P \) and \( T \) according to the Bradley and Pitzer parametrization.\(^{374}\) The same parametrization was employed for the experimental solvent permittivity \( \varepsilon(P,T) \), which is also a parameter of the correction scheme. The expression for \( \varepsilon(P,T) \) is provided explicitly in the Supplementary Material. The calculation of \( \rho(P,T) \) was performed using available IAPWS software.\(^{377}\) For all the other parameters, function values and derivatives at \( P^o \) and \( T^- \) are sufficient to characterize the correction scheme for both Variants A and B. These quantities are reported in Table 2.3.
Table 2.3: Parameters used in the application of the correction scheme.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit (u)</th>
<th>(f)</th>
<th>(\partial f / [\text{u K}^{-1}])</th>
<th>(\partial^2 f / [\text{u K}^2])</th>
<th>(\partial f / [\text{u bar}^{-1}])</th>
<th>(\partial^2 f / [\text{u bar}^2 K^{-1}])</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R^c_i) (Na(^+)) (^{(a,b)})</td>
<td>m</td>
<td>1.68·10^{-10}</td>
<td>-4·10^{-14}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(R^c_i) (Cl(^-)) (^{(a)})</td>
<td>m</td>
<td>2.46·10^{-10}</td>
<td>-4·10^{-14}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\rho) (^{(a)})</td>
<td>kg m(^{-3})</td>
<td>9.9705·10^{2}</td>
<td>-2.57·10^{-1}</td>
<td>-9.53·10^{-2}</td>
<td>4.51·10^{-2}</td>
<td>-9.56·10^{-6}</td>
</tr>
<tr>
<td>(\rho') (^{(a)})</td>
<td>kg m(^{-3})</td>
<td>9.682·10^{2}</td>
<td>-2.49·10^{-1}</td>
<td>-9.26·10^{-2}</td>
<td>4.35·10^{-2}</td>
<td>-9.28·10^{-6}</td>
</tr>
<tr>
<td>(\epsilon) (^{(a)})</td>
<td>1</td>
<td>7.8384·10^{4}</td>
<td>-3.5939·10^{-1}</td>
<td>1.5526·10^{-2}</td>
<td>3.7055·10^{-2}</td>
<td>-5.4964·10^{-7}</td>
</tr>
<tr>
<td>(\epsilon') (^{(a)})</td>
<td>1</td>
<td>6.66·10^{4}</td>
<td>-3.6522·10^{-1}</td>
<td>1.3196·10^{-2}</td>
<td>3.1478·10^{-2}</td>
<td>-4.6691·10^{-7}</td>
</tr>
<tr>
<td>(\chi') (^{(a)})</td>
<td>V</td>
<td>7.3·10^{-1}</td>
<td>-4.8·10^{-4}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\chi'') (^{(a)})</td>
<td>Vm</td>
<td>-1.1·10^{-10}</td>
<td>-6.72·10^{-14}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\chi'') (^{(a)})</td>
<td>Vm</td>
<td>-1.11·10^{-9}</td>
<td>-7.3·10^{-13}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^{(a)}\) These parameters enter into the evaluation of \(\Delta C^\infty_{\text{corr}}\) (Eq. 2.36) and are the effective ionic radius \(R_i\), the density \(\rho\) of the employed water model, the permittivity \(\epsilon\) of the water model, the air-liquid interfacial potential \(\chi'\) (air-to-liquid direction) of the water model and its dependence \(\tilde{\chi}'\) on the inverse curvature radius for a concave interface, as well as the experimental solvent permittivity \(\epsilon\). Additional entries for the experimental solvent density \(\rho\) are also included. The corresponding first and second (self and cross) derivatives \((\partial P, \partial T, \partial^2 P, \partial^2 T, \partial^2 P, \partial^2 T)\) with respect to pressure \(P\) or/and temperature \(T\) are also reported. The data refers to the SPC water model and the Na\(^+\) and Cl\(^-\) ions of set L in Ref. 61, as well as to the standard pressure \(P_0 = 1\) bar and temperature \(T = 298.15\) K. This information entirely defines the correction scheme for Variant A. It must be complemented by the functional dependence of \(\rho\) and \(\epsilon\) for Variant B. (b) Value from Ref. 61 (entry \(R_{\text{mon}}\) in Tables II and III therein), and estimated temperature derivative from Ref. 174 (see Table VI therein); (c) Second temperature derivative as well as pressure dependence neglected due to lack of data, as in Ref. 174; (d) Diffs slightly from the parameters of Ref. 174 (see Table VI therein), due to the use of different source data or approximations; (e) Based on the IAPWS-95 parametrization;\(^{376}\) (f) Density based on the present simulations, derivatives approximated by means of Eq. 2.47, \(i.e.\) calculated from those of \(\rho\) using a scaling factor of 0.971 = 968.2/997.05; (g) Based on the Bradley and Pitzer parametrization\(^{374}\) (see Supplementary Material); (h) Permittivity based on Ref. 370, derivatives approximated by means of Eq. 2.48, \(i.e.\) calculated from those of \(\epsilon\) using a scaling factor of 0.849 = 66.678.384; (i) Value and estimated temperature derivative from Ref. 174 (see Table VI therein), based on Ref. 172; (j) Reported for completeness, not necessary in the case of LS simulations (present work).

### 2.4 Results and discussion

#### 2.4.1 Method I

The \(\Delta H^p_{\text{hyd}}, \psi^p_{\text{aq}}, c^p_{\text{aq}}, k^p_{\text{aq}}\) and \(\epsilon^p_{\text{aq}}\) estimates obtained using Method I (Sections 2.2.3 and 2.3.2) considering three different system sizes (512, 724 or 1024 water molecules) for the Na\(^+\) or Cl\(^-\) ions are reported in Tables 2.4 and 2.5, respectively.
Table 2.4: Raw simulation data, correction terms and corrected results corresponding to hydration and aqueous partial molar parameters for Na⁺ calculated via Method I, along with Variants A or B.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(N_W)</th>
<th>(L) [nm]</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta Y_{\text{aq}})</td>
<td>(\Delta Y_{\text{bi}})</td>
<td>(\Delta Y_{\text{C1}})</td>
<td>(\Delta Y_{\text{C2}})</td>
<td>(\Delta Y_{\text{D}})</td>
</tr>
<tr>
<td>(\Delta H_{\text{aq}}) (\text{[kJ mol}^{-1})</td>
<td>512 2.51</td>
<td>0.0</td>
<td>-80.3</td>
<td>-31.0</td>
</tr>
<tr>
<td>724 2.82</td>
<td>0.0</td>
<td>-71.6</td>
<td>-31.1</td>
<td>0.0</td>
</tr>
<tr>
<td>1024 3.16</td>
<td>0.0</td>
<td>-63.9</td>
<td>-31.1</td>
<td>0.0</td>
</tr>
<tr>
<td>(\Delta P) (\text{[10}^{-5}\text{ m}^{3}\text{ mol}^{-1})</td>
<td>512 2.51</td>
<td>0.0</td>
<td>-12.0</td>
<td>-31.1</td>
</tr>
<tr>
<td>724 2.82</td>
<td>0.0</td>
<td>-10.7</td>
<td>-31.1</td>
<td>0.0</td>
</tr>
<tr>
<td>1024 3.16</td>
<td>0.0</td>
<td>-9.6</td>
<td>-31.1</td>
<td>0.0</td>
</tr>
<tr>
<td>(\Delta v) (\text{[10}^{-5}\text{ m}^{3}\text{ mol}^{-1})</td>
<td>512 2.51</td>
<td>0.0</td>
<td>-79.8</td>
<td>-219.9</td>
</tr>
<tr>
<td>724 2.82</td>
<td>0.0</td>
<td>-71.4</td>
<td>-219.9</td>
<td>0.0</td>
</tr>
<tr>
<td>1024 3.16</td>
<td>0.0</td>
<td>-63.8</td>
<td>-219.9</td>
<td>0.0</td>
</tr>
<tr>
<td>(k_{P}) (\text{[10}^{-10}\text{ m}^{3}\text{ mol}^{-1}\text{ bar}^{-1})</td>
<td>512 2.51</td>
<td>0.0</td>
<td>-28.9</td>
<td>-73.9</td>
</tr>
<tr>
<td>724 2.82</td>
<td>0.0</td>
<td>-25.8</td>
<td>-73.9</td>
<td>0.0</td>
</tr>
<tr>
<td>1024 3.16</td>
<td>0.0</td>
<td>-23.1</td>
<td>-74.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(\alpha_{P}) (\text{[10}^{-8}\text{ m}^{3}\text{ mol}^{-1}\text{ K}^{-1})</td>
<td>512 2.51</td>
<td>0.0</td>
<td>2.6</td>
<td>9.5</td>
</tr>
<tr>
<td>724 2.82</td>
<td>0.0</td>
<td>2.3</td>
<td>9.5</td>
<td>0.0</td>
</tr>
<tr>
<td>1024 3.16</td>
<td>0.0</td>
<td>2.1</td>
<td>9.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter \(Y\) (or \(y\)) considered, the number of water molecules \(N_W\) in the system, the (average) box-edge length \(L\), the correction terms \(\Delta Y_{\text{A}}\), \(\Delta Y_{\text{B}}\), \(\Delta Y_{\text{C1}}\), \(\Delta Y_{\text{C2}}\) and \(\Delta Y_{\text{D}}\), the corresponding sum \(\Delta Y_{\text{corr}}\), the raw simulation result \(\Delta Y_{\text{raw}}\) and the corrected result \(\Delta Y_{\text{rot}}\). Values are provided for the enthalpy \(H\), volume \(V\), heat capacity \(C_P\), volume-compressibility \(K_T\) and volume-expansivity \(A_P\). The data refers to the standard pressure \(P^0\) = 1 bar and temperature \(T^-\) = 298.15 K. Additional details on the calculations can be found in Supplementary Material (Table A.2 and Figure A.1).

Footnote (c) provides a comparison with previous results from Ref. 61 (see Supplementary Material Table S2.1a therein), corresponding to the same force-field parameters, system, and calculation method IA (the small differences arise from a longer simulation time and the use of slightly different parameters for the correction scheme). (b) For \(v\), \(c_P\), \(k_T\) and \(A_P\), these entries refer to the quantities usually written \(\Delta Y_{\text{raw}}\) and \(\Delta Y_{\text{rot}}\) (equivalent to \(\Delta Y_{\text{raw}}\) and \(\Delta Y_{\text{rot}}\) in the present context owing to Eqs. 2.8 and 2.9); (c) To be compared with \(\Delta Y_{\text{raw}} = \sum_{\text{aq}} Y - \sum_{\text{aq}} Y = (9.89 \pm 2.17) \cdot 10^{-6}\) and \(\Delta Y_{\text{rot}} = (33.25 \pm 2.17) \cdot 10^{-6}\) m³ mol⁻¹ in Ref. 61; (d) The corresponding values when applying a two-point forward finite difference (based on the results at 1 and 5000 bar) instead of the three-point least-squares fit are 0.7 ± 1.4, −1.1 ± 1.5 and −2.0 ± 1.9 \(10^{-10}\) m³ mol⁻¹ bar⁻¹ for 512, 724 and 1024 water molecules, respectively; (e) The entries for \(\Delta H\) and \(v\) are not repeated for Variant B (identical to the corresponding entries for Variant A); (f) The individual correction terms were calculated by a centered finite difference in temperature, based on the data reported in Supplementary Material Table A.2; (g) The individual correction terms were estimated by least-squares fitting in pressure, based on the values reported in Supplementary Material Table A.2, while their sum \(\Delta K_{T,\text{raw}}\) was back-calculated from \(k_{P,\text{aq,raw}}\) and \(k_{P,\text{aq,raw}}\) (as a result, the terms may not sum up exactly to \(\Delta K_{T,\text{raw}}\), reason for which they are reported between parentheses).
2.4 Results and discussion

Table 2.5: Raw simulation data, correction terms and corrected results corresponding to hydration and aqueous partial molar parameters for Cl− calculated via Method I, along with Variants A or B.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variant A</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N_{\text{w}})</td>
<td>(L) [nm]</td>
<td>(\Delta Y_A)</td>
</tr>
<tr>
<td>(\Delta H) [kJ mol(^{-1})]</td>
<td>512 2.51 0.0</td>
<td>-79.5 82.8 0.2</td>
<td>0.3 3.8</td>
</tr>
<tr>
<td></td>
<td>724 2.82 0.0</td>
<td>-71.1 82.9 0.2</td>
<td>0.3 12.2</td>
</tr>
<tr>
<td></td>
<td>1024 3.16 0.0</td>
<td>-63.5 83.0 0.1</td>
<td>0.3 19.8</td>
</tr>
<tr>
<td>(v) (10^{-6}) m(^3) mol(^{-1})</td>
<td>512 2.51 0.0</td>
<td>-11.7 34.7 0.0</td>
<td>0.3 23.4</td>
</tr>
<tr>
<td></td>
<td>724 2.82 0.0</td>
<td>-10.5 34.7 0.0</td>
<td>0.3 24.6</td>
</tr>
<tr>
<td></td>
<td>1024 3.16 0.0</td>
<td>-9.4 34.8 0.0</td>
<td>0.3 25.7</td>
</tr>
<tr>
<td>(c_p) (\text{J mol}^{-1})K(^{-1})</td>
<td>512 2.51 0.0</td>
<td>-77.9 215.5 0.2</td>
<td>4.5 145.3</td>
</tr>
<tr>
<td></td>
<td>724 2.82 0.0</td>
<td>-70.0 218.9 0.1</td>
<td>4.5 152.5</td>
</tr>
<tr>
<td></td>
<td>1024 3.16 0.0</td>
<td>-62.8 219.3 0.1</td>
<td>4.5 161.0</td>
</tr>
<tr>
<td>(k_T) (10^{-10}) m(^3) mol(^{-1}) bar(^{-1})</td>
<td>512 2.51 0.0</td>
<td>-28.2 73.6 0.1</td>
<td>0.7 46.2</td>
</tr>
<tr>
<td></td>
<td>724 2.82 0.0</td>
<td>-25.4 73.7 0.1</td>
<td>0.7 49.1</td>
</tr>
<tr>
<td></td>
<td>1024 3.16 0.0</td>
<td>-22.8 73.8 0.1</td>
<td>0.7 51.8</td>
</tr>
<tr>
<td>(a_T) (10^{-8}) m(^3) mol(^{-1}) K(^{-1})</td>
<td>512 2.51 0.0</td>
<td>2.5 -9.4 0.0</td>
<td>0.2 -6.7</td>
</tr>
<tr>
<td></td>
<td>724 2.82 0.0</td>
<td>2.3 -9.4 0.0</td>
<td>0.2 -7.0</td>
</tr>
<tr>
<td></td>
<td>1024 3.16 0.0</td>
<td>2.1 -9.5 0.0</td>
<td>0.2 -7.2</td>
</tr>
</tbody>
</table>

\(a\) The quantities reported are the parameter \(Y\) (or \(y\)) considered, the number of water molecules \(N_{\text{w}}\) in the system, the (average) box-edge length \(L\), the correction terms \(\Delta Y_A\), \(\Delta Y_B\), \(\Delta Y_{C_1}\), \(\Delta Y_{C_2}\) and \(\Delta Y_{\text{cor}}\), the corresponding sum \(\Delta Y_{\text{cor}}^\text{raw}\), the raw simulation result \(\Delta Y_{\text{cor}}^\text{hyd,raw}\), and the corrected result \(\Delta Y_{\text{cor}}^\text{hyd}\). Values are provided for the enthalpy \(H\), volume \(V\), heat capacity \(C_p\), volume-compressibility \(K_T\) and volume-expansivity \(A_T\). The data refers to the standard pressure \(P^\circ = 1\) bar and temperature \(T^\circ = 298.15\) K. Additional details on the calculations can be found in Supplementary Material (Table A.3 and Figure A.1).

Footnote (c) provides a comparison with previous results from Ref. 61 (see Supplementary Material Table S2.Va therein), corresponding to the same force-field parameters, system, and calculation method IA (the small differences arises from a longer simulation time and the use of slightly different parameters for the correction scheme). (b) For \(v\), \(c_p\), \(k_T\) and \(a_T\), these entries refer to the quantities usually written \(y_{aq,raw}\) and \(y_{aq}^\text{raw}\) (equivalent to \(\Delta Y_{\text{cor}}^\text{hyd,raw}\) and \(\Delta Y_{\text{cor}}^\text{hyd}\)) in the present context owing to Eqs. 2.8 and 2.9; (c) To be compared with \(y_{aq,raw} = (14.97 ± 2.14) \cdot 10^{-6}\) and \(y_{aq} = (40.63 ± 2.14) \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\) in Ref. 61; (d) The corresponding values when applying a two-point forward finite difference (based on the results at 1 and 5000 bar) instead of the three-point least-squares fit are \(-10.3 ± 1.4\), \(-9.2 ± 1.6\) and \(-11.6 ± 1.8 \cdot 10^{-10}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\) for 512, 724 and 1024 water molecules, respectively; (e) The entries for \(\Delta H\) and \(v\) are not repeated for Variant B (identical to the corresponding entries for Variant A); (f) The individual correction terms were calculated by a centered finite difference in temperature, based on the data reported in Supplementary Material Table A.3; (g) The individual correction terms were estimated by least-squares fitting in pressure, based on the values reported in Supplementary Material Table A.3, while their sum \(\Delta K_{T,cor}\) was back-calculated from \(k_{T,aq,raw}\) and \(k_{T,aq}^\text{raw}\) (as a result, the terms may not sum up exactly to \(\Delta K_{T,cor}\)) for reason which they are reported between parentheses.
More details on the calculations can also be found in Tables A.2 and A.3 as well as Figure A.1 of the Supplementary Material. The results for Variants A (Section 2.2.4) and B (Section 2.2.4) are discussed separately in the two following sections.

**Variant A**

For both Na\(^+\) and Cl\(^-\), the raw hydration enthalpy is negative and increases in magnitude with increasing system size, which indicates that the systems considered fail to encompass a sufficient volume of correctly polarized solvent around the ion. For both ions, this finite-size effect is slightly over-compensated by the type-B correction, leading to corrected hydration enthalpies $\Delta H_{\text{hyd}}$ that slightly decrease in magnitude with increasing system size, evaluating to about $-476 \text{ kJ mol}^{-1}$ and $-365 \text{ kJ mol}^{-1}$ for Na\(^+\) and Cl\(^-\), respectively, considering the largest systems. In both cases, the type-B correction for finite-size effects (on the order of $\approx -60 \text{ to } -80 \text{ kJ mol}^{-1}$) is very large for the truly microscopic system sizes considered, which remain very far from a bulk (infinite-dilution) situation. Note also the important contribution of the type-C\(_1\) correction for improper potential summation to these estimates, of comparable magnitude (about 83 kJ mol\(^{-1}\)) but opposite sign for the two ions.

The raw partial molar volumes are positive for the two ions, and decrease with increasing system size. This trend is also easily understood considering that the positive excluded-volume contribution to this parameter is short-ranged, while the negative electrostrictive contribution is longer-ranged. This finite-size effect is appropriately removed by the type-B correction, leading to corrected partial molar volumes $v_{\text{aq}}$ that are essentially identical within the statistical uncertainty (about $1 \cdot 10^{-6} \text{ m}^3 \text{ mol}^{-1}$) for the three system sizes considered, evaluating to about $-33 \cdot 10^{-6}$ or $+39 \cdot 10^{-6} \text{ m}^3 \text{ mol}^{-1}$, for Na\(^+\) and Cl\(^-\), respectively. The type-B correction (on the order of $\approx 10 \cdot 10^{-6} \text{ m}^3 \text{ mol}^{-1}$) is again large for both ions, but comparatively less important than for the hydration enthalpy. Note again the very large contribution of the type-C\(_1\) correction to these estimates, of comparable magnitude (about $35\cdot 10^{-6} \text{ m}^3 \text{ mol}^{-1}$) but opposite sign for the two ions.

The raw partial molar heat capacities are affected by very large statistical uncertainties, suggesting that much longer ($\gg 60 \text{ ns}$) simulations would be required to achieve proper convergence. This is easily understood considering that Method I involves in this case the comparison of two very large numbers, namely the total potential energies of two aqueous systems, including or excluding the ion. Although an uncertainty of about 1 kJ mol\(^{-1}\) on these two
numbers is perfectly acceptable in terms of the calculated hydration enthalpy, for a heat capacity evaluated by finite difference over a temperature interval of 50 K, it already leads to an error of about 40 J mol\(^{-1}\) K\(^{-1}\). For Na\(^+\) and Cl\(^-\), the corrected partial molar heat capacities \(c^\infty_{P,aq}\) obtained for the systems involving 512 and 724 water molecules are very close. For this reason, and because a better convergence of the total potential energy is likely to be achieved for a smaller system, the results obtained using 512 water molecules are considered here to be the most reliable. The suggested \(c^\infty_{P,aq}\) values are thus about \(-339\) or \(+184\) J mol\(^{-1}\) K\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively, with a large statistical uncertainty (about 30 J mol\(^{-1}\) K\(^{-1}\)). The type-B correction (on the order of \(-60\) to \(-80\) J mol\(^{-1}\) K\(^{-1}\)) is very large for both ions, as was the case for the hydration enthalpy. However, the type-C\(_1\) correction is even more important here, and again of comparable magnitude (about 220 J mol\(^{-1}\) K\(^{-1}\)) but opposite sign for the two ions.

The raw partial molar volume-compressibilities are of opposite signs for the two ions, and tendentially decrease with increasing system size. Note that slightly lower raw values are obtained if a two-point forward difference is applied for the calculation instead of the three-point least-squares fit, see footnote (c) in Tables 2.4 and 2.5. The trend in the raw values suggests that the residual long-range polarization of the solvent is correlated with a reduction of its volume response to pressure, presumably through a corresponding alteration of the hydrogen-bonding network. This finite-size effect is reduced by the type-B correction, leading to corrected partial molar volume-compressibilities \(k^\infty_{T,aq}\) that are similar within the statistical uncertainty (about \(1 \cdot 10^{-10}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\)) for the three system sizes considered, evaluating to about \(-95 \cdot 10^{-10}\) or \(+45 \cdot 10^{-10}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\) respectively, for Na\(^+\) and Cl\(^-\) considering the largest system. As will be discussed later (Section 2.4.2), however, these estimates are not considered to be accurate. The type-B correction (on the order of \(-20 \cdot 10^{-10}\) to \(-30 \cdot 10^{-10}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\)) is again large for both ions, but comparatively less important than for the hydration enthalpy. Note again the important contribution of the type-C\(_1\) correction to these estimates, of comparable magnitude (about 75 \(10^{-10}\) m\(^3\) mol\(^{-1}\) bar\(^{-1}\)) but opposite sign for the two ions.

Finally, the raw partial molar volume-expansivities are of opposite signs for the two ions, and decrease with increasing system size. This trend suggests that the residual long-range polarization of the solvent is correlated with a reduction of its volume response to temperature, presumably through a corresponding alteration of the hydrogen-bonding network. For this parameter, however, the type-B correction also decreases with increasing system size, and seems unable to capture the finite-size effect appropriately. For this
reason, the results obtained using 512 water molecules are considered here to be the most reliable. The corrected partial molar volume-compressibilities $a_{P,aq}^{\pm}$ are nevertheless similar within the statistical uncertainty (about $2 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$) for the three system sizes considered, evaluating to about $+26 \cdot 10^{-8}$ or $-17 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$, for Na$^+$ or Cl$^-$, respectively, considering the smallest system. The type-B correction (on the order of $3 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$) is again large for both ions, but comparatively less important than for the hydration enthalpy. Note again the important contribution of the type-C$_1$ correction to these estimates, of comparable magnitude (about $9.5 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$) but opposite sign for the two ions.

**Variant B**

In terms of the hydration enthalpies and partial molar volumes, there is no distinction between Variants A and B. In terms of the partial molar heat capacities, volume-expansivities and volume-compressibilities, the raw values are also identical in the two variants.

Considering the values obtained using 512 water molecules to be the most reliable (Section 2.4.1), the corrected partial molar heat capacities $c_{P,aq}^{\pm}$ calculated using Variant B, evaluating to about $-362$ or $+196$ J mol$^{-1}$ K$^{-1}$, for Na$^+$ and Cl$^-$, respectively, differ noticeably from those calculated using Variant A (about $-339$ or $+184$ J mol$^{-1}$ K$^{-1}$, respectively). The difference is due to a larger magnitudes of the type-B and type-C$_1$ correction terms, mainly related to the non-linear dependence of the density of water as a function of temperature. Considering the very large statistical uncertainty (about 30 J mol$^{-1}$ K$^{-1}$) affecting the results, the discrepancy between the two variants remains nevertheless acceptable.

The corrected partial molar volume-compressibilities $k_{T,aq}^{\pm}$ calculated using Variant B are similar within the statistical uncertainty for the three system sizes considered, evaluating to about $-29 \cdot 10^{-10}$ or $+10 \cdot 10^{-10}$ m$^3$ mol$^{-1}$ bar$^{-1}$, for Na$^+$ and Cl$^-$, respectively, considering the largest system. These values differ considerably from those obtained using Variant A (about $-95 \cdot 10^{-10}$ or $+45 \cdot 10^{-10}$ m$^3$ mol$^{-1}$ bar$^{-1}$ respectively). Here, the difference is due to much smaller magnitudes of the type-B and type-C$_1$ correction terms, mainly related to the non-linear dependence of the density of water as a function of pressure. The effect of this non-linearity is clearly visible in panels (c,d) of Figure A.1 in the Supplementary Material, which shows that the corrected values of $v_{aq}^{\pm}$ at the three pressures considered are not aligned. As a result, the calculated values are quite sensitive to the calculation procedure. As will be discussed later (Section 2.4.2), these estimates
are considered to be more accurate than those of method IA.

The corrected partial molar volume-expansivities $\tilde{a}_{P,aq}$ calculated using Variant B also decrease slightly with increasing system size, but are nevertheless similar within the statistical uncertainty for the three system sizes considered, evaluating to about $+28 \cdot 10^{-8}$ or $-18 \cdot 10^{-8} \text{ m}^3 \text{ mol}^{-1} \text{ K}^{-1}$ for Na$^+$ and Cl$^-$, respectively, considering the values obtained using 512 water molecules to be the most reliable (Section 2.4.1). These values are in excellent agreement with those obtained using Variant A (about $+26 \cdot 10^{-8}$ or $-17 \cdot 10^{-8} \text{ m}^3 \text{ mol}^{-1} \text{ K}^{-1}$, respectively), considering the same system size.

### 2.4.2 Method II

The $\Delta G_{\text{hyd}}^\text{aq}$, $\Delta H_{\text{hyd}}^\text{aq}$, $s_{\text{aq}}^\text{aq}$ and $v_{\text{aq}}^\text{aq}$ estimates obtained using Method II (Sections 2.2.3 and 2.3.3) considering two different system sizes (724 or 1024 water molecules; only 1024 for $v_{\text{aq}}^\text{aq}$) for the Na$^+$ or Cl$^-$ ions are reported in Tables 2.6 and 2.7, respectively. More details on the calculations can also be found in Tables A.4 and A.5 as well as Figure A.2 of the Supplementary Material document. The results for Variants A (Section 2.2.4) and B (Section 2.2.4) are discussed separately in the two following sections.
Table 2.6: Raw simulation data, correction terms and corrected results corresponding to hydration and aqueous partial molar parameters for Na⁺ calculated via Method II, along with Variants A or B.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(N_W)</th>
<th>(L_\text{[mm]})</th>
<th>Correction terms</th>
<th>Result</th>
<th>(\Delta Y_{\text{hyd}})</th>
<th>(\Delta Y_{\text{ch}})</th>
<th>(\Delta Y_{\text{cav}})</th>
<th>(\Delta Y_{\text{D}})</th>
<th>(\Delta Y_{\text{corr}})</th>
<th>(\Delta Y_{\text{corr}})(^{(3)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variant A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta \Omega)</td>
<td>724</td>
<td>2.22</td>
<td>0.0</td>
<td>-68.5</td>
<td>-77.2</td>
<td>0.0</td>
<td>-0.9</td>
<td>-146.6</td>
<td>10.5 ± 0.2</td>
<td>-295.0 ± 0.7</td>
</tr>
<tr>
<td>(\text{[kJ mol}^{-1})]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-61.1</td>
<td>-77.2</td>
<td>0.0</td>
<td>-0.9</td>
<td>-139.2(^{(4)})</td>
<td>11.7 ± 0.2(^{(3)})</td>
<td>-303.4 ± 0.6(^{(3)})</td>
</tr>
<tr>
<td>(\Delta \Omega)</td>
<td>724</td>
<td>2.22</td>
<td>0.0</td>
<td>-71.6</td>
<td>-81.1</td>
<td>0.0</td>
<td>0.4</td>
<td>-154.3</td>
<td>5.7 ± 1.4</td>
<td>-314.4 ± 5.3</td>
</tr>
<tr>
<td>(\text{[kJ mol}^{-1})]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-63.9</td>
<td>-83.1</td>
<td>0.0</td>
<td>0.4</td>
<td>-146.6(^{(7)})</td>
<td>7.5 ± 1.4(^{(6)})</td>
<td>-333.2 ± 5.2(^{(6)})</td>
</tr>
<tr>
<td>(\text{[kJ mol}^{-1} \cdot \text{K}^{-1})]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-9.3</td>
<td>-19.8</td>
<td>-0.1</td>
<td>4.5</td>
<td>-247.1</td>
<td>-14.3 ± 4.8(^{(8)})</td>
<td>-100.0 ± 17.3(^{(8)})</td>
</tr>
<tr>
<td>(s)</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-9.6</td>
<td>-34.9</td>
<td>0.0</td>
<td>0.4</td>
<td>-44.9</td>
<td>13.9 ± 0.6</td>
<td>0.6 ± 0.8</td>
</tr>
<tr>
<td>(\text{[10}^{-3}\text{mol}^2\text{mol}^{-1}])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variant B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta \Omega)</td>
<td>724</td>
<td>2.22</td>
<td>0.0</td>
<td>(-71.5)</td>
<td>(-82.7)</td>
<td>(0.0)</td>
<td>(0.4)</td>
<td>-153.8</td>
<td>5.7 ± 1.4</td>
<td>-314.4 ± 5.3</td>
</tr>
<tr>
<td>(\text{[kJ mol}^{-1})]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>(-63.8)</td>
<td>(-82.7)</td>
<td>(0.0)</td>
<td>(0.4)</td>
<td>-146.1</td>
<td>7.5 ± 1.4(^{(4)})</td>
<td>-333.2 ± 5.2(^{(4)})</td>
</tr>
<tr>
<td>(\text{[kJ mol}^{-1} \cdot \text{K}^{-1})]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>(-9.0)</td>
<td>(-18.5)</td>
<td>(-0.1)</td>
<td>(4.5)</td>
<td>-241</td>
<td>-15.8 ± 4.8</td>
<td>-65.0 ± 17.5</td>
</tr>
<tr>
<td>(s)</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>(-9.9)</td>
<td>(-18.5)</td>
<td>(-0.1)</td>
<td>(4.5)</td>
<td>-231</td>
<td>-14.3 ± 4.8(^{(8)})</td>
<td>-100.0 ± 17.3(^{(8)})</td>
</tr>
<tr>
<td>(\text{[10}^{-3}\text{mol}^2\text{mol}^{-1}])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter \(Y\) (or \(y\)) considered, the number of water molecules \(N_W\) in the system, the (average) box-edge length \(L\), the correction terms \(\Delta Y_A, \Delta Y_B, \Delta Y_{C_1}, \Delta Y_{C_2}\) and \(\Delta Y_D\), the corresponding sum \(\Delta Y_{\text{corr}}\), the cavitation term \(\Delta Y_{\text{cav}}\), the raw charging term \(\Delta Y_{\text{chg},\text{raw}}\), and the corrected result \(\Delta Y_{\text{chg}}\). Values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\) and volume \(V\). The data refers to the standard pressure \(P\) = 1 bar and temperature \(T\) = 298.15 K. Additional details on the calculations can be found in Supplementary Material (Table A.4 and Figure A.2). Footnotes (c)-(k) provide a comparison with previous results from Ref. 61 (see Supplementary Material Table S2.IVa therein) and Ref. 333 (see Table 5.7 therein) corresponding to the same force-field parameters, system, and calculation method IIA (the small differences arise from a longer simulation time and the use of slightly different parameters for the correction scheme). (b) For \(s\) and \(v\), this entry refers to the quantity usually written \(y_{aq}\) (equivalent to \(\Delta Y_{\text{chg}}\) in the present context owing to Eqs. 2.8 and 2.9); (c) To be compared with \(-138.65\text{kJ mol}^{-1}\) in Ref. 61; (d) To be compared with \(12.49 ± 0.27\text{kJ mol}^{-1}\) in Ref. 61; (e) To be compared with \(-302.29 ± 0.64\text{kJ mol}^{-1}\) in Ref. 61; (f) To be compared with \(-161.9\text{kJ mol}^{-1}\) based on Refs. 61,333 using the Gibbs equation; (g) To be compared with \(+8.45 ± 3.81\text{kJ mol}^{-1}\) based on Refs. 61,333 using the Gibbs equation; (h) To be compared with \(+327.33: 7.77\text{kJ mol}^{-1}\) based on Refs. 61,333 using the Gibbs equation; (i) To be compared with \(-78.12\text{kJ mol}^{-1} \cdot \text{K}^{-1}\) in Ref. 333 (the large difference arises from the use of a different value of \(\partial \phi / \partial T\) in the two studies); (j) To be compared with \(-13.25 ± 12.75\text{kJ mol}^{-1} \cdot \text{K}^{-1}\) in Ref. 333; (k) To be compared with \(-80.93 ± 25.98\text{kJ mol}^{-1} \cdot \text{K}^{-1}\) in Ref. 333; (l) The entry for \(\Delta G\) is not repeated for Variant B (identical to the corresponding entry for Variant A); (m) The individual correction terms were estimated by least-squares fitting in temperature or pressure, based on the values reported in Supplementary Material Table A.4, while their sum \(\Delta Y_{\text{corr}}\) was back-calculated from \(\Delta Y_{\text{chg},\text{raw}}\) (as obtained by Variant A) and \(\Delta Y_{\text{chg}}\) (as a result, the terms may not sum up exactly to \(\Delta Y_{\text{corr}}\), reason for which they are reported between parentheses).
Table 2.7: Raw simulation data, correction terms and corrected results corresponding to hydration and aqueous partial molar parameters for Cl− calculated via Method II, along with Variants A or B. (a)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Nw</th>
<th>L [nm]</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ΔY_A</td>
<td>ΔY_B</td>
</tr>
<tr>
<td>Variant A</td>
<td></td>
<td></td>
<td>12.8</td>
<td>18.4</td>
</tr>
<tr>
<td>ΔG</td>
<td>724</td>
<td>2.82</td>
<td>0.0</td>
<td>68.1</td>
</tr>
<tr>
<td>[kJ mol⁻¹]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>60.8</td>
</tr>
<tr>
<td>ΔH</td>
<td>724</td>
<td>2.82</td>
<td>0.0</td>
<td>-71.1</td>
</tr>
<tr>
<td>[kJ mol⁻¹]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>63.5</td>
</tr>
<tr>
<td>s</td>
<td>724</td>
<td>2.82</td>
<td>0.0</td>
<td>-10.2</td>
</tr>
<tr>
<td>[mol⁻¹ K⁻¹]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-9.2</td>
</tr>
<tr>
<td>v</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-9.4</td>
</tr>
<tr>
<td>[10⁴ m³ mol⁻¹]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variant B (a)</td>
<td></td>
<td></td>
<td>12.8</td>
<td>18.4</td>
</tr>
<tr>
<td>ΔH</td>
<td>724</td>
<td>2.82</td>
<td>0.0</td>
<td>-71.0</td>
</tr>
<tr>
<td>[kJ mol⁻¹]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-63.4</td>
</tr>
<tr>
<td>s</td>
<td>724</td>
<td>2.82</td>
<td>0.0</td>
<td>-10.2</td>
</tr>
<tr>
<td>[mol⁻¹ K⁻¹]</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-9.2</td>
</tr>
<tr>
<td>v</td>
<td>1024</td>
<td>3.16</td>
<td>0.0</td>
<td>-4.7</td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter Y (or y) considered, the number of water molecules Nw in the system, the (average) box-edge length L, the correction terms ΔY_A, ΔY_B, ΔY_C1, ΔY_C2 and ΔY_D, the corresponding sum ΔY_cor, the cavitation term ΔY_cav, the raw charging term ΔY_chg,raw, and the corrected result ΔY_hyd. Values are provided for the free energy G, enthalpy H, entropy S and volume V. The data refers to the standard pressure P° = 1 bar and temperature T° = 298.15 K. Additional details on the calculations can be found in Supplementary Material (Table A.5 and Figure A.2). Footnotes (c)-(k) provide a comparison with previous results from Ref. 61 (see Supplementary Material Table S2.1Va therein) and Ref. 333 (see Table 5.7 therein) corresponding to the same force-field parameters, system, and calculation method IIa (the small differences arise from a longer simulation time and the use of slightly different parameters for the correction scheme). (b) For s and v, this entry refers to the quantity usually written y_ab (equivalent to ΔY_hyd in the present context owing to Eqs. 2.8 and 2.9); (c) To be compared with 15.34 kJ mol⁻¹ in Ref. 61; (d) To be compared with 14.97 ± 0.46 kJ mol⁻¹ in Ref. 61; (e) To be compared with -348.39 ± 0.61 kJ mol⁻¹ in Ref. 61; (f) To be compared with 26.3 kJ mol⁻¹ based on Refs. 61,333 using the Gibbs equation; (g) To be compared with 6.41 ± 7.10 kJ mol⁻¹ based on Refs. 61,333 using the Gibbs equation; (h) To be compared with -381.77 ± 7.90 kJ mol⁻¹ based on Refs. 61,333 using the Gibbs equation; (i) To be compared with 36.86 J mol⁻¹ K⁻¹ in Ref. 333 (the large difference arises from the use of a different value of ∂G/∂T in the two studies); (j) To be compared with -27.75 ± 23.75 J mol⁻¹ K⁻¹ in Ref. 333; (k) To be compared with -114.43 ± 26.40 J mol⁻¹ K⁻¹ in Ref. 333; (l) The entry for ΔG is not repeated for Variant B (identical to the corresponding entry for Variant A); (m) The individual correction terms were estimated by least-squares fitting in temperature or pressure, based on the values reported in Supplementary Material Table A.5, while their sum ΔY_cor was back-calculated from ΔY_hyd,raw (as obtained by Variant A) and ΔY_hyd (as a result, the terms may not sum up exactly to ΔY_cor, reason for which they are reported between parentheses).
Variant A

For both Na\(^+\) and Cl\(^-\), the raw hydration free energy is negative and increases in magnitude with increasing system size, which indicates that the systems considered fail to encompass a sufficient volume of correctly polarized solvent around the ion. For Na\(^+\), this finite-size effect is appropriately compensated by the type-B correction, leading to a corrected hydration free energy \(\Delta G_{\text{hyd}}\) of about \(-431\) kJ mol\(^{-1}\). For Cl\(^-\), the finite-size effect is slightly over-compensated, leading to a corrected hydration free energy that slightly decreases in magnitude with increasing system size, evaluating to about \(-318\) kJ mol\(^{-1}\) for the largest system. These values can be compared to the values of \(-428.4 \pm 0.8\) or \(-318.1 \pm 1.1\) kJ mol\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively, obtained previously\(^6\) with the same force-field parameters, system (1024 water molecules) and calculation method. The small difference for Na\(^+\) arises from a longer simulation time and the use of slightly different parameters for the cavitation free energy calculation (different set of \(\lambda\)-points) and for the correction scheme. In both cases, the type-B correction for finite-size effects (on the order of \(-60\) to \(-70\) kJ mol\(^{-1}\)) is very large for the truly microscopic system sizes considered, which remain very far from a bulk (infinite-dilution) situation. The cavitation contribution (about 10-15 kJ mol\(^{-1}\)) is small compared to the charging term. Note also the important contribution of the type-C\(_1\) correction term for improper potential summation to these estimates, of comparable magnitude (about 77 kJ mol\(^{-1}\)) but opposite sign for the two ions.

For the hydration enthalpy, similar observations hold as those made in the context of Method I (Section 2.4.1). However, here, the system-size dependence of the raw values appears to be slightly more pronounced for both ions, the corrected results present a larger residual system-size dependence for Na\(^+\) compared to Cl\(^-\), and this residual dependence for Na\(^+\) corresponds to a more negative solvation enthalpy in the largest system. In spite of these small differences, the corrected hydration enthalpies \(\Delta H_{\text{hyd}}\) calculated using Method II, evaluating to about \(-472\) or \(-350\) kJ mol\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively, considering the largest system, are in acceptable agreement with those obtained using Method I (about \(-476\) or \(-365\) kJ mol\(^{-1}\), respectively), noting that these values are affected by larger statistical uncertainties (about 6 kJ mol\(^{-1}\)). Here also, the cavitation contribution is very small compared to the charging term.

The charging component of the raw partial molar entropies is negative for the two ions, and increases in magnitude with increasing system size. This trend is easily understood considering that the residual long-range polarization of the solvent is correlated with an intrinsic ordering of the solvent.
molecules. The corresponding cavitation component is also negative for the two ions. For Na\(^+\), it is relatively small and system-size independent. For Cl\(^-\), it is significantly larger and decreases in magnitude with increasing system size. After application of the correction terms, and considering the largest system for Na\(^+\), the corrected partial molar entropies \(s_{aq}^{\pm}\) evaluate to about \(-139\) or \(-106\) J mol\(^{-1}\) K\(^{-1}\), for Na\(^+\) and Cl\(^-\), respectively, with relatively large statistical uncertainties (about 18 J mol\(^{-1}\) K\(^{-1}\)). Observing that the error on \(\Delta H_{hyd}^{\pm}\) is lower in Method I than in Method II, possibly more accurate estimates for \(s_{aq}^{\pm}\) can be deduced from the hydration free energies (about \(-431\) or \(-318\) kJ mol\(^{-1}\), respectively) along with the hydration enthalpies of Method I (about \(-476\) or \(-365\) kJ mol\(^{-1}\), respectively) using the Gibbs equation, resulting in about \(-153\) J mol\(^{-1}\) K\(^{-1}\) for Na\(^+\) and \(-155\) J mol\(^{-1}\) K\(^{-1}\) for Cl\(^-\). For the two ions, the type-B correction term (on the order of \(-10\) J mol\(^{-1}\) K\(^{-1}\)), the type-C\(_1\) correction term (magnitude on the order of 20 J mol\(^{-1}\) K\(^{-1}\)), and the cavitation term (on the order of \(-15\) or \(-80\) J mol\(^{-1}\) K\(^{-1}\) for Na\(^+\) and Cl\(^-\)) all provide contributions of large magnitudes. In practice, the accurate calculation of ionic solvation entropies is notoriously difficult,\(^{330-335}\) and the results also turn out to be very sensitive to the detailed structural properties of the employed water model, as well as its ability to reproduce the experimental temperature dependence of the density and permittivity (see e.g. Table 5.7 in Ref. 333).

For the partial molar volume, only one system size was considered. For this parameter, the cavitation component, accounting for the excluded-volume of the ion, is expectedly positive and becomes the leading term. For Na\(^+\), the charging component, accounting for electrostrictive effects, is negligible. In contrast, for Cl\(^-\), this term is important and expectedly negative. The corrected partial molar volumes \(v_{aq}^{\pm}\) calculated using Method II, evaluating to about \(-30 \cdot 10^{-6}\) or \(+43 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively, are in good agreement with those obtained using Method I (about \(-33 \cdot 10^{-6}\) or \(+39 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\), respectively), the values being affected by similar statistical uncertainties (about \(1 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\)).

**Variant B**

In terms of the hydration free energy, there is no distinction between Variants A and B. In terms of the hydration enthalpy as well as the partial molar volume and entropy, the raw values are also identical in the two variants.

The corrected hydration enthalpies \(\Delta H_{hyd}^{\pm}\) and partial molar entropies \(s_{aq}^{\pm}\) calculated using Variant B are very close to the corresponding values calculated using Variant A, with maximal differences on the order of 1 kJ
mol\(^{-1}\) and 2 J mol\(^{-1}\) K\(^{-1}\), respectively.

In contrast, the corrected partial molar volumes \(v_{aq}^{\text{\(\square\)}}\) calculated using Variant B, evaluating to about \(-9 \cdot 10^{-6}\) or \(+31 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively, differ significantly from those obtained using Variant A (about \(-30 \cdot 10^{-6}\) or \(+43 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\), respectively) or using Method I (about \(-33 \cdot 10^{-6}\) or \(+39 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\), respectively), the values being affected by similar statistical uncertainties (about \(1 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\)). The difference is mainly due to a smaller magnitudes of the type-C\(_1\) correction term, related to the non-linear dependence of the density of water as a function of pressure. For Na\(^+\), for which the difference is most pronounced, the effect of this non-linearity is clearly visible in panel (c) of Figure A.2 in the Supplementary Material, which shows that the corrected values of the charging component of \(\Delta G_{\text{hyd}}^{\text{\(\square\)}}\) at the three pressures considered are not aligned. As a result, the calculated values are quite sensitive to the calculation procedure.

**Second-order derivatives**

The evaluation of the hydration free energy as a function of pressure or/and temperature could not be performed with a sufficient accuracy to permit a reliable determination of the second-order derivatives \(c_{P,aq}^{\text{\(\square\)}}\), \(k_{T,aq}^{\text{\(\square\)}}\) and \(a_{P,aq}^{\text{\(\square\)}}\) via Method II. This is clear from the free-energy curves displayed in Figure A.2 of the Supplementary Material document, which suggest that obtaining accurate curvature information is essentially impossible. However, given the important differences in the estimates of \(k_{T,aq}^{\text{\(\square\)}}\) and, to a lesser extent, \(c_{P,aq}^{\text{\(\square\)}}\) via the two variants of Method I (Section 2.4.1), these estimates were checked for compatibility with the curvature of the solvation free-energy curves as a function of pressure or/and temperature, given the associated uncertainty. This comparison is performed in Figure 2.1, based on the model functions

\[
\Delta G_{\text{hyd}}(T) = \Delta H_{\text{hyd}}^{\text{\(\square\)}} + c_{P,aq}^{\text{\(\square\)}} (T - T^-) \tag{2.49}
\]

\[
-T \left( \frac{\Delta H_{\text{hyd}}^{\text{\(\square\)}} - \Delta G_{\text{hyd}}^{\text{\(\square\)}}}{T^-} + c_{P,aq}^{\text{\(\square\)}} \ln \frac{T}{T^-} \right)
\]

for the temperature dependence of \(\Delta G_{\text{hyd}}\) at \(P = P^\circ\), and

\[
\Delta G_{\text{hyd}}(P) = \Delta G_{\text{hyd}}^{\text{\(\square\)}} + v_{aq}^{\text{\(\square\)}} (P - P^\circ) - (1/2)k_{T,aq}^{\text{\(\square\)}} (P - P^\circ)^2 \tag{2.50}
\]

for the pressure dependence of \(\Delta G_{\text{hyd}}\) at \(T = T^-\). These are the appropriate thermodynamic equations for an extrapolation of \(\Delta G_{\text{hyd}}\) assuming a temperature-independent \(c_{P,aq}^{\text{\(\square\)}}\) (equal to \(c_{P,aq}^{\text{\(\square\)}}\)) and a pressure-independent
Figure 2.1: Assessment of the aqueous partial molar heat capacities and compressibilities calculated for Na\(^+\) or Cl\(^-\) via Methods IA or IB against the pressure or temperature dependence of hydration free energies calculated via Method II. The top panels (a,b) show the temperature dependence at \(P = P^0\) and the bottom panels (c,d) the pressure dependence at \(T = T^0\) of the corrected hydration free energy \(\Delta G_{hyd}\). The left panels (a,c) refer to Na\(^+\) and the right panels (b,d) to Cl\(^-\). The system considered is the one involving 1024 water molecules. Curves corresponding to the model functions of Eq. 2.49 for the temperature dependence and Eq. 2.50 for the pressure dependence, along with various combinations of parameters, are also displayed. For panels (a) and (b), the different curves rely on \(\Delta G_{hyd}^{\square}\) from method II, \(\Delta H_{hyd}^{\square}\) from either Method I (\(\Delta H_{hyd,I}^{\square}\)) or Method IIA (\(\Delta H_{hyd,IIA}^{\square}\)), the latter essentially identical to IIB, and \(c_{P,aq}^{\square}\) from either Method IA (\(c_{P,aq,IA}^{\square}\)) or Method IB (\(c_{P,aq,IB}^{\square}\)). Note that the effect of \(c_{P,aq}^{\square}\) is very small so that the corresponding curves are essentially indistinguishable. For panels (c) and (d), the different curves rely on \(\Delta G_{hyd}^{\square}\) from method II, \(v_{aq}^{\square}\) from either Method I (\(v_{aq,I}^{\square}\)), Method IIA (\(v_{aq,IIA}^{\square}\)) or Method IIB (\(v_{aq,IIB}^{\square}\)), and \(k_{T,aq}^{\square}\) from either Method IA (\(k_{T,aq,IA}^{\square}\)) or Method IB (\(k_{T,aq,IB}^{\square}\)). An additional curve corresponds to \(k_{T,aq,IB}^{\square}\) along with \(v_{aq,fit}^{\square}\) values leading to the best fit of the model equation to the hydration free energy. These values are \(v_{aq,fit}^{\square} = -23.3 \cdot 10^{-6}\) or \(v_{aq,fit}^{\square} = +36.4 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively.
I are in good agreement (Section 2.4.1).

Considering the temperature dependence of the hydration free energy for Na\(^+\), all combinations of \(\Delta H_{\text{hyd}}\) (I or IIA, the latter essentially equal to IIB) and \(c_{P,aq}\) (IA and IB) estimates are compatible with the \(\Delta G_{\text{hyd}}\) curve, given the the associated uncertainty. For the Cl\(^-\) ion, the model function obtained using the \(\Delta H_{\text{hyd}}\) estimate from Method IIA (essentially equal to IIB) is clearly more accurate than that obtained using the corresponding estimate from Method I, the effect of the selected \(c_{P,aq}\) (IA and IB) estimate being negligible. These observations illustrate that the slope of the \(\Delta G_{\text{hyd}}\) curve provides sufficient information on \(\Delta H_{\text{hyd}}\) (similar values from I and IIA for Na\(^+\), larger difference for Cl\(^-\)), but that the corresponding curvature is too small to provide reliable information on \(c_{P,aq}\), given the uncertainty affecting the data.

Considering the pressure dependence of the hydration free energy for both ions, the model functions obtained using the \(k_{T,aq}\) estimates from Method IA clearly evidence too pronounced curvatures. Using the \(k_{T,aq}\) estimates from Method IB, the curve relying on the \(v_{aq}\) estimate from Method IIA is the most accurate for Na\(^+\), and that relying on the \(v_{aq}\) estimate from Method I is the most accurate for Cl\(^-\). Reoptimizing \(v_{aq}\) for best agreement results in values of about \(-23 \cdot 10^{-6}\) or \(+36 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\) for Na\(^+\) and Cl\(^-\), respectively, acceptably close to the values for method IIA (\(-30 \cdot 10^{-6}\) or \(+43 \cdot 10^{-6}\) m\(^3\) mol\(^{-1}\), respectively). Following from this analysis, the \(v_{aq}\) estimates from Method IIA and the \(k_{T,aq}\) estimates from Method IB were retained as the most reliable.

### 2.4.3 Comparison with available experimental data

A detailed summary of the calculation results for the four combinations of methods and variants, and for all considered system sizes, is provided in Tables A.6-A.8 of the Supplementary Material document. Based on the discussion of Sections 2.4.1 and 2.4.2, values considered by the authors of this article as the most reliable were selected for each parameter and for both ions. These values are reported in Table 2.8 and compared with available experimental or recommended data, the latter based on or derived from the numbers reported in Table 2.2.
Table 2.8: Corrected results corresponding to hydration and aqueous partial molar parameters for Na\(^+\) and Cl\(^-\), considered as the most reliable estimates based on the present simulations.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Property</th>
<th>Sel.</th>
<th>Na(^+)</th>
<th>Cl(^-)</th>
<th>Na(^+)+Cl(^-)</th>
<th>H(^+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta G_{\text{hyd}}) ([\text{kJ mol}^{-1}])</td>
<td>II</td>
<td>-430.9</td>
<td>-427.41</td>
<td>-318.4</td>
<td>-316.55</td>
</tr>
<tr>
<td>(\Delta H_{\text{hyd}}) ([\text{kJ mol}^{-1}])</td>
<td>I</td>
<td>-476.3</td>
<td>-450.20</td>
<td>-364.6</td>
<td>-328.80</td>
</tr>
<tr>
<td>(s) (\left[\text{J mol}^{-1} \text{K}^{-1}\right])</td>
<td>1+II (^{(c)})</td>
<td>-152.5</td>
<td>-76.40</td>
<td>-154.8</td>
<td>-41.09</td>
</tr>
<tr>
<td>(v) (\left[10^{-3} \text{m}^{3} \text{mol}^{-1}\right])</td>
<td>IA (^{(d)})</td>
<td>-29.5</td>
<td>-5.86</td>
<td>42.7</td>
<td>22.21</td>
</tr>
<tr>
<td>(k_{T}) (\left[\text{J mol}^{-1} \text{K}^{-1}\right])</td>
<td>IA (^{(f)})</td>
<td>-339.0</td>
<td>-43.84</td>
<td>184.4</td>
<td>-81.90</td>
</tr>
<tr>
<td>(a_{P}) (\left[10^{-3} \text{m}^{3} \text{mol}^{-1} \text{bar}^{-1}\right])</td>
<td>IB (^{(g)})</td>
<td>-29.3</td>
<td>-27.45</td>
<td>10.1</td>
<td>-21.29</td>
</tr>
<tr>
<td>(a_{P}) (\left[10^{-3} \text{m}^{3} \text{mol}^{-1} \text{K}^{-1}\right])</td>
<td>IA (^{(e)})</td>
<td>26.0</td>
<td>2.42</td>
<td>-17.3</td>
<td>4.62</td>
</tr>
</tbody>
</table>

\(^{(a)}\) The selection was made based on the discussion presented in Sections 2.4.1 and 2.4.2, and the selected (Sel.) calculation method (Method I or II), variant (Variant A or B) and system size (by default, system with 1024 water molecules) are indicated. A detailed summary of the calculation results for the four combinations of methods and variants, and for all considered system sizes, can also be found in Tables A.6-A.8 of the Supplementary Material. The quantities reported are the parameter \(Y\) (or \(y\)) considered, and the corrected simulation estimate \(\Delta Y_{\text{hyd}}\). Values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\), volume \(V\), heat capacity \(C_P\), volume-compressibility \(K_T\) and volume-expansivity \(A_P\). The reported values for Na\(^+\)+Cl\(^-\) correspond to sum of parameters over the individual ions Na\(^+\) and Cl\(^-\). The data refers to the standard pressure \(P^\circ\) = 1 bar and temperature \(T^\circ\) = 298.15 K. The simulation values are reported from the corresponding entries of Tables 2.4-2.7. The simulation (Sim.) results are compared with recommended (Rec.) or experimental (Exp.) values, taken from the corresponding entry in Table 2.2. For the individual ions, the recommended values are inferred from experiment, but correspond to experimental elusive quantities, and are thus to be considered with some caution.\(^{166}\) For the ion pair, the experimental values are in principle experimentally accessible. However, the entries for \(c_{P,aq}^a\), \(k_{T,aq}^a\) and \(a_{P,aq}^a\) should still be taken with some caution in view of the sparseness and indirect nature of the corresponding experimental determinations.\(^{166}\) Values for the proton inferred from the simulations are also provided, and compared to corresponding recommended data. (b) For \(s, v, c_P, k_T\) and \(a_P\), the entries of this table refer to the quantities usually written \(\gamma_{aq}^a\) (equivalent to \(\Delta Y_{\text{hyd}}^a\) in the present context owing to Eqs. 2.8 and 2.9); (c) Value for the proton deduced from the simulation results for Na\(^+\) as \(\Delta Y_{\text{hyd, rec}}^a[\text{Na}^+]-\Delta Y_{\text{hyd, rec}}^a[\text{Na}^+]+\Delta Y_{\text{hyd, rec}}^a[H^+]\), “rec” indicating the recommended value of Table 2.2 and “sim” the value from the simulation; (d) Value for the proton deduced from the simulation results for Cl\(^-\) as \(-\Delta Y_{\text{hyd, rec}}^a[\text{Cl}^-]+\Delta Y_{\text{hyd, rec}}^a[\text{Cl}^-]+\Delta Y_{\text{hyd, rec}}^a[H^+]\), “rec” indicating the recommended value of Table 2.2 and “sim” the value from the simulation; (e) Value calculated based on the Gibbs equation using \(\Delta H\) from method I and \(\Delta G\) from method II; (f) System with 512 water molecules; (g) Value supported by compatibility analysis.
For the Na\(^+\) and Cl\(^-\) ions taken individually, the comparison between the calculated numbers and the recommended values from Table 2.2 should be preformed with caution, considering that the latter estimates depend on a specific choice for the experimentally elusive proton parameters \(\Delta G_{\text{hyd}}\)\(^+[\text{H}^+]\), \(\Delta H_{\text{hyd}}\)\(^+[\text{H}^+]\), \(s_{\text{aq}}\)\(^+[\text{H}^+]\), \(v_{\text{aq}}\)\(^+[\text{H}^+]\), \(c_{P,\text{aq}}\)\(^+[\text{H}^+]\), \(k_{T,\text{aq}}\)\(^+[\text{H}^+]\) and \(a_{P,\text{aq}}\)\(^+[\text{H}^+]\). For both ions, the calculated and recommended hydration free energies are in excellent agreement. This is no surprise considering that the ion-solvent Lennard-Jones interaction parameters used here were explicitly calibrated against these recommended values\(^61\) (Set L, assuming \(\Delta G_{\text{hyd}}\)\(^+[\text{H}^+]\) = -1100 kJ mol\(^-1\) or, equivalently, \(\Delta G_{\text{hyd}}\)\(^+[\text{H}^+]\) = -1107.95 kJ mol\(^-1\)). For all the other parameters, the agreement is qualitative at best, which may result either from inaccuracies in the computational model and calculation procedure, from inappropriate values for the recommended parameters of the proton, or from a combination of both. For \(\Delta H_{\text{hyd}}\), qualitative agreement is achieved, i.e. the calculated and recommended values have the same orders of magnitudes for the two ions. For \(s_{\text{aq}}\), the two ions are characterized by large negative values, the calculated ones being considerably larger in magnitude. For \(v_{\text{aq}}\), the two ions are characterized by values of opposite signs. Na\(^+\) has a negative partial molar volume, suggesting a dominance of electrostrictive over excluded-volume effects, the opposite being true for Cl\(^-\). For \(c_{P,\text{aq}}\), \(k_{T,\text{aq}}\) and \(a_{P,\text{aq}}\), the calculated and recommended values are of the same sign for Na\(^+\), but of opposite signs for Cl\(^-\). For \(c_{P,\text{aq}}\) and \(a_{P,\text{aq}}\) the calculated values are also considerably larger in magnitude, whereas the orders of magnitude are comparable for \(k_{T,\text{aq}}\).

In contrast to the single-ion parameters, the sums of these parameters for the Na\(^+\)+Cl\(^-\) pair are in principle experimentally accessible, and are thus referred to here as experimental rather than recommended. Note, however, that for the second-derivative quantities \(c_{P,\text{aq}}\), \(k_{T,\text{aq}}\) and \(a_{P,\text{aq}}\), the experimental estimates even for these sums are not highly reliable, in view of the very small number of independent determinations and of the limited consistency among the resulting estimates.\(^{166}\) Expectedly, the calculated and experimental ion-pair hydration free energies are in excellent agreement. On the other hand, the enthalpy-entropy partitioning is not accurate, both \(\Delta H_{\text{hyd}}\) and \(s_{\text{aq}}\) being overestimated in magnitude in the present calculations. This failure is likely due at least in part to the use of an implicit-polarization model for the ion-water interactions, which has already been suggested to result in an overstructuring of the hydration shells.\(^{181,378,379}\) Note, however, that other authors report comparable ion-water radial distribution functions for implicit- and explicit-polarization models, while density functional calculations em-
ploying a plane-wave basis set resulted in less structured hydration shells. As observed previously, the calculated ion-pair partial molar volume $v_{\text{aq}}^{\text{□}}$ is in good agreement with the corresponding experimental estimate, although the partitioning in terms of single-ion volumes differs significantly from that suggested by the recommended value $v_{\text{aq}}^{\text{□}}[\text{H}^+] = -4.5 \cdot 10^{-6}$ m$^3$ mol$^{-1}$. A similar observation applies to the partitioning of the partial molar heat capacity, for which the calculated ion-pair $c_{\text{P,aq}}^{\text{□}}$ value is also in reasonable agreement with the corresponding experimental estimate. The calculated and experimental $k_{\text{T,aq}}^{\text{□}}$ values all have a negative sign, but the magnitude of the experimental numbers is about twice as large, which may also result at least in part from an overstructuring of the solvation shells. In contrast, the calculated and experimental ion-pair $a_{\text{P,aq}}^{\text{□}}$ values agree very well. Here also, however, the partitioning in terms of single-ion volume-expansivities differs significantly from that suggested by the recommended value $a_{\text{P,aq}}^{\text{□}}[\text{H}^+] = -3.0 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$.

Finally, on the basis of the single-ion parameters obtained from the calculations, it is possible to suggest alternative values for the corresponding proton parameters, possibly differing from the recommended ones. Clearly, these suggested alternatives will only be similar for Na$^+$ and Cl$^-$, and thus acceptable, in cases where the ion-pair parameters are compatible with the corresponding experimental numbers. Expectedly, for the hydration free energy, the two alternative proton values are similar to each other and to the recommended value $\Delta G_{\text{hyd}}^{\text{□}} = -1107.95$ kJ mol$^{-1}$. For $\Delta H_{\text{hyd}}^{\text{□}}$, $s_{\text{aq}}^{\text{□}}$ and $k_{\text{T,aq}}^{\text{□}}$, the estimates for the two ions are very different, and thus unreliable. For $v_{\text{aq}}^{\text{□}}$, the two values are very close, and an estimate of about $-26.5 \cdot 10^{-6}$ m$^3$ mol$^{-1}$ can be formulated for $v_{\text{aq}}^{\text{□}}[\text{H}^+]$ on the basis of the simulations, as an alternative to the recommended value of $-4.5 \cdot 10^{-6}$ m$^3$ mol$^{-1}$. For $c_{\text{P,aq}}^{\text{□}}$, the two values are acceptably close given the large uncertainty on the calculated numbers, and an estimate of about $-366$ J K$^{-1}$ mol$^{-1}$ can be formulated for $c_{\text{P,aq}}^{\text{□}}[\text{H}^+]$ on the basis of the simulations, as an alternative to the recommended (but highly uncertain) value of $-86$ J K$^{-1}$ mol$^{-1}$. This estimate remains nevertheless uncertain in view of the overestimation of the magnitude of the corresponding ion-pair parameter. For $a_{\text{P,aq}}^{\text{□}}$, the two values are very close, and an estimate of about $+20.0 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$ can be formulated for $a_{\text{P,aq}}^{\text{□}}[\text{H}^+]$ on the basis of the simulations, as an alternative to the recommended value of $-3.0 \cdot 10^{-8}$ m$^3$ mol$^{-1}$ K$^{-1}$. 
2.5 Conclusion

The aim of the present study was to investigate the feasibility of calculating and correcting derivative thermodynamic hydration and aqueous partial molar properties for Na\(^+\) and Cl\(^-\) at \(P^0\) and \(T^\circ\). As stated in Section 2.1, this type of calculations is notoriously difficult, and the goal of this work was to cut a first path through an unexplored jungle rather than to reach quantitative agreement with sometimes rather uncertain experimental data. The results corresponding to different calculation schemes (Methods I or II in conjunction with Variants A or B) as well as the final results presented in Table 2.8 clearly show that, except for the hydration free energy itself, accurate methodological independence and quantitative agreement with even the most reliable experimental parameters (ion-pair properties) are not reached.

Expectedly, the calculated solvation free energies agree well with experimental ion-pair data and recommended single-ion values, due to appropriate calibration of the selected ion-solvent Lennard-Jones interaction parameters.\(^{61}\) However, the calculated ion-pair solvation enthalpies and partial molar entropies are too negative. This is probably due at least in part to the use of an implicit-polarization model for the ion-solvent interactions, resulting in an overstructuring of the hydration shells.\(^{181,378,379}\) The calculated partial molar heat capacities are affected by very large statistical uncertainties, and much longer (\(\gg 60\) ns) simulations would be required to achieve proper convergence. Given these uncertainties, the ion-pair results for this property are in acceptable agreement with experimental data. However, the single-ion results suggest a partitioning compatible with \(c_{P,aq}^\text{\(\cdot\)}[\text{H}\(^+\)] = -366\ \text{J K}^{-1}\ \text{mol}^{-1}\) for the proton, rather than with the recommended\(^{166}\) (but highly uncertain) value of \(-86\ \text{J K}^{-1}\ \text{mol}^{-1}\). In particular, the former choice leads to a large and positive partial molar heat capacity for Cl\(^-\), while the latter choice suggests a negative value instead. The calculated ion-pair partial molar volumes agree well with experimental ion-pair data, as is the case for the other alkali-halide ion pairs.\(^{61}\) However, here also, the single-ion results suggest a partitioning compatible with \(v_{aq}^\text{\(\cdot\)}[\text{H}\(^+\)] = -26.5 \cdot 10^{-6}\ \text{m}^3\ \text{mol}^{-1}\) for the proton, rather than with the recommended\(^{166}\) value of \(-4.5 \cdot 10^{-6}\ \text{m}^3\ \text{mol}^{-1}\). The calculated ion-pair partial molar volume-expansivities agree well with experimental ion-pair data. However, here again, the single-ion results suggest a partitioning compatible with \(a_{P,aq}^\text{\(\cdot\)}[\text{H}\(^+\)] = +20.0 \cdot 10^{-8}\ \text{m}^3\ \text{mol}^{-1}\ \text{K}^{-1}\) for the proton, rather than with the recommended\(^{166}\) value of \(-3.0 \cdot 10^{-8}\ \text{m}^3\ \text{mol}^{-1}\ \text{K}^{-1}\). In particular, the former choice leads to a large and negative partial molar volume-expansivity for Cl\(^-\), while the latter choice suggests a small positive value instead. Finally, the calculated partial molar volume-
compressibilities are affected by a strong sensitivity to the calculation protocol, and even the estimates considered most reliable in this work are not in good agreement with experimental ion-pair data. It should be kept in mind, however, that the experimental ion-pair data for the second-derivative quantities $c_P^\infty$, $k_T^\infty$, and $a_P^\infty$ relies on a handful of indirect and often relatively ancient determinations. The uncertainty on the corresponding recommended proton values is even higher, considering that the tentative experimental determinations of these parameters are sparser, depend on specific extrathermodynamic assumptions, and often present a limited consistency among each other. For these second-derivative parameters, it is thus not entirely clear whether these recommended values or the results of the present calculations are to be considered most trustworthy. Finally, it should be kept in mind that the ion-pair data quoted here as experimental, as well as the proton data quoted here as recommended, are affected by an ambiguity related to the solute standard-state variant associated with the experimental numbers (Section 2.2.2). Based on the most common types of experimental determinations, a specific variant (standard or density-corrected) has been assumed here for each type of quantity, but an incorrect choice may lead to non-negligible errors, especially for the partial molar volume, heat capacity and compressibility (see $\Delta y_{aq}^{\infty \rightleftharpoons \infty}$ and $\Delta\Delta Y_{hyd}^{\infty \rightleftharpoons \infty}$ values in Table 2.1).

Even if methodological independence and quantitative agreement with available experimental data is not yet reached, the present study underlines two important points: (i) the need to consider standard-state issues very carefully; (ii) the requirement to correct the calculated raw numbers appropriately. With respect to the latter point, the type-B correction term for finite-size effects and the type-C$_1$ correction term for improper potential summation are typically important, and their omission would lead to extremely large errors for all the calculated parameters. In this sense, the main merit of the present study is to set a clear framework for these types of calculations and to point towards directions for future improvements, with the ultimate goal of reaching a consistent and quantitative description of single-ion hydration thermodynamics in molecular dynamics simulations.

The residual dependence of the estimated parameters on the simulation protocol and calculation scheme results from three main problems: too short sampling times, too large finite-difference intervals, and inaccurate correction-scheme parameters. While the first two problems can in principle be remedied by investing more computing resources, a significant improvement in the correction scheme could be achieved by an explicit determination of the pressure and temperature dependences of the density and dielectric permittivity of the water model employed in the simulations. At present, the
main reason for the discrepancies between the different calculation schemes (Methods I or II in conjunction with variants A or B) probably results from using the approximation of Eqs. 2.47 and 2.48 for these functions. Three other lines of further methodological developments could involve: (i) the use of generalized-ensemble techniques\textsuperscript{381–385} for the direct sampling of system configurations over a finite pressure-temperature range within a single simulation, as an alternative to the multiple-point approach employed here; (ii) the use of appropriate fluctuation formulae\textsuperscript{371,386–397} for the direct evaluation of the pressure and temperature derivatives of the quantities considered, as another alternative to the multiple-point approach; (iii) the use of multiple-parameter fitting formulae such as Eqs. 2.49 and 2.50, as an alternative to the finite-difference or least-squares fitting approach employed here. The application of these alternative schemes would render the calculation procedure more automatic, i.e. considerably reduce the human time involved in the (tedious) set-up and post-processing of the simulations. Note that the use of fluctuation formulae rests on the ability of the simulation procedure to accurately reproduce not only ensemble averages but also fluctuations appropriate for the isobaric-isothermal ensemble. For example, the Berendsen barostat and thermostat employed here\textsuperscript{359} do not fulfill this condition, even in principle.\textsuperscript{392,398–400} Other formally isobaric-isothermal algorithms e.g. the Andersen barostat\textsuperscript{401} and the Nosé-Hoover thermostat\textsuperscript{402} may also fail to satisfy this criterion in practice, due to numerical noise and ergodicity violations\textsuperscript{403–407} (steady-state rather than equilibrium situation). This issue introduces another factor of uncertainty in the simulation results, reason for which this approach was not considered in the present study. The comparison of finite-difference, generalized-ensemble and fluctuation approaches will be the scope of future work.

Provided that the residual methodological dependence of the estimated parameters can be reduced to an acceptable level, the remaining sources of discrepancies between calculation and experiment will reside in the inaccuracy of the molecular model employed and that of the experimental data to compare with. For the hydration enthalpy and the partial molar entropy, the experimental ion-pair parameters and, to a slightly lesser extent, the recommended single-ion parameters can probably be trusted. As mentioned above, the too negative values of the calculated ion-pair solvation enthalpies and partial molar entropies are probably due at least in part to the use of an implicit-polarization model for the ion-solvent interactions. The application of explicitly polarizable models\textsuperscript{378,379} or even of quantum-mechanical models (e.g. Car-Parrinello\textsuperscript{380,408–411} or quantum-mechanics molecular mechanics\textsuperscript{412–414}), would probably enable a more accurate representation of the
enthalpy-entropy partitioning. Note in this respect that the present work considers relatively large monovalent ions, for which electronic polarization of the solvent and the partially covalent character of ion-solvent interactions are expected to have limited effects. However, an appropriate treatment of these effects will become essential when small or/and polyvalent ions are considered. For the second-order derivatives, where the uncertainty in experimental ion-pair data and, to an even higher extent, recommended proton values is very large, it is not impossible that more accurate calculations with an improved molecular model ultimately provide results that are more trustworthy than experimentally inferred ones.

Finally, another important aspect of this research remains to be mentioned. If the calculation of derivative thermodynamic hydration and aqueous partial molar properties is highly relevant in the context of force-field parametrization and validation, it is also of interest per se in terms of the characterization of ionic hydration properties. For example, the recommended values as well as the results of the present calculations suggest that Na$^+$ promotes the orientational polarization (enthalpy), the orientational correlation (entropy) and the contraction (volume) of the surrounding solvent, along with a decrease of the energetic thermal response (heat capacity), volumic pressure response (volume-compressibility) and volumic temperature response (volume-expansivity). While the first three effects are well known, understanding the nature of the latter four in terms of a corresponding structural alteration of the hydrogen-bonding network of water by the ion remains challenging, and of great theoretical interest.
Assessment of enveloping distribution sampling to calculate relative free enthalpies of binding for eight netropsin-DNA duplex complexes in aqueous solution

The performance of enveloping distribution sampling (EDS) simulations to estimate free enthalpy differences associated with seven alchemical transformations of A-T into G-C base pairs at the netropsin binding site in the minor groove of a 13-base pair DNA duplex in aqueous solution is evaluated. It is demonstrated that sufficient sampling can be achieved with a two-state EDS Hamiltonian even for large perturbations such as the simultaneous transformation of up to three A-T into three G-C base pairs. The two parameters required to define the EDS reference state Hamiltonian are obtained automatically using a modified version of a scheme presented in earlier work. The sensitivity of the configurational sampling to a variation of these parameters is investigated in detail. Although for relatively small perturbations, that is, one base pair, the free enthalpy estimate depends only weakly on the EDS parameters, the sensitivity is stronger for the largest perturbation. Yet, EDS offers various convenient measures to evaluate the degree of sampling and thus the reliability of the free enthalpy estimate and appears to be an efficient alternative to the conventional thermodynamic integration methodology to obtain free energy differences for molecular systems.
3.1 Introduction

The knowledge of free energy changes associated with biomolecular processes such as protein-ligand binding or molecular recognition is essential for their understanding and possible utilisation in drug design. Therefore, the calculation of relative free energies using molecular simulation techniques has been an active field of research for more than three decades. An essential part of this ongoing effort is the development of efficient, reliable, and robust methods for free energy calculation in the framework of molecular dynamics. A frequently used approach is thermodynamic integration (TI) and extensions thereof that are based on a model Hamiltonian \( H_R(p, r; \lambda) \) which connects a pair of states \((i, j)\) by a coupling parameter \(\lambda\). With \(r\) and \(p\) we denote the \(3N\)-dimensional vector of spatial coordinates and the \(3N\)-dimensional vector of conjugate momenta, respectively, where \(N\) is the number of atoms in the system. Practical issues with this method are related to (i) the choice of the \(\lambda\)-dependence of the Hamiltonian, (ii) the choice of the number of \(\lambda\)-points between two states, (iii) the length of the sampling period at each \(\lambda\)-point, and (iv) the choice of the numerical integration method. A technique that avoids this \(\lambda\)-dependence is one-step perturbation which allows for the evaluation of relative free energies between many states from just a single simulation of a reference state \(R\), given that in state \(R\) conformations relevant to all desired end states are sampled. Therefore, the choice of the reference state Hamiltonian is crucial for the convergence of the one-step perturbation approach and is strongly dependent on the type of molecular system. Moreover, no easy way to evaluate the degree of convergence is available. A promising solution to these problems is offered by the recently proposed implementation of the umbrella sampling approach termed enveloping distribution sampling (EDS). The functional form of the EDS reference state Hamiltonian \(H_R(p, r)\) is chosen such that it envelopes those of states \(i\),

\[
H_R(p, r) = -\frac{1}{\beta s} \ln \sum_{i=1}^{N_H} e^{-\beta s (H_i(p, r) - E^R_i)}, \tag{3.1}
\]

The free energy difference between a pair of states \((i, j)\) follows from the expression

\[
F_i - F_j = -\frac{1}{\beta} \ln \frac{\langle e^{-\beta (H_i - H_R)} \rangle_R}{\langle e^{-\beta (H_j - H_R)} \rangle_R}. \tag{3.2}
\]

In this approach the problem of choosing a priori a pathway that connects two states or a particular reference Hamiltonian that has sufficient overlap with
the end states is converted to the determination of a smoothness parameter $s$ ($0 < s \leq 1$), which smoothens transitions between the different states $i$ for smaller $s$ values, and energy offset parameters $E_i^R$ of the states $i$, that bring the energies of the states to the same level. In equations (3.1) and (3.2), $\beta$ has the meaning of $1/k_B T$ where $k_B$ is the Boltzmann constant and $T$ the absolute temperature. The benefit of this formulation compared to the other methods discussed is that it allows for an automatic update procedure for these parameters during an MD simulation. For a number of model systems it was shown that these automatically determined parameters lead to very efficient sampling of the relevant conformations of states $i$. Moreover, it was demonstrated how $H_R(p, r)$ in equation (3.1) can be generalised to one with multiple smoothness parameters $s_{ij}$ smoothing transitions between states $i$ and $j$.

However, a recent application of the EDS methodology to a more complex system, calculation of binding free energies of phenylethanolamine N-methyltransferase (PNMT) inhibitors, showed that the automatically derived EDS parameters cause less sufficient sampling of the relevant phase space the more the end states differ in potential energy. The reason for this sampling problem was that the smoothness parameter was calculated too small due to a reciprocal relation between $s$ and $E_i^R$ in the original update algorithm. While it could be demonstrated that this is not a problem of the EDS Hamiltonian itself, it turned out that a new update scheme had to be invented that is based on counting how often a state is visited during a simulation. To decide to which state the conformations sampled in a simulation belong at a particular time an energy criterion was used. With the newly determined parameters it was possible to sample the relevant phase space and to reach agreement with results from thermodynamic integration.

In the present work we present a slightly modified parameter update scheme, which relies on the energy difference between two states rather than on the absolute energies. We apply the EDS methodology to a complex problem, prediction of sequence specific recognition between a 13-base pair DNA duplex and the minor groove binding ligand netropsin in aqueous solution. Specifically we have analysed netropsin binding to a set of $2^3 = 8$ mixed AT/GC-containing binding sequences which differ in the composition of three central base pairs. The results are compared to results from TI and experimental data where available. The current limitations of the methodology are discussed and a set of criteria is presented that allow to assess the degree of sampling obtained in an EDS simulation.
3.2 Theory

3.2.1 Two state EDS Hamiltonian

In force fields typically used for molecular simulations the potential energy function is independent of velocities. As a result the kinetic contribution to the free energy can be calculated analytically. For a two-state system the potential energy term of the reference Hamiltonian then reads

\[
V_R(r; s, \Delta E_{BA}^R) = -\frac{1}{\beta s} \ln \left[ e^{-\beta s (V_A(r) - E_A^R)} + e^{-\beta s (V_B(r) - E_B^R)} \right],
\]  

(3.3)

where \(V_X\) denotes the potential energy of state \(X\). The corresponding equations of motion depend only on the relative energy offset, \(\Delta E_{BA}^R = E_B^R - E_A^R\).

Therefore, \(E_A^R\) can be set to zero such that the problem of defining the reference state Hamiltonian reduces to the determination of \(s\) and \(E_B^R\).

The free energy difference between state \(B\) and state \(A\) can be estimated from an MD simulation producing a reference state ensemble using

\[
\Delta F_{BA} = -\frac{1}{\beta} \ln \frac{\langle e^{-\beta (V_B - V_A)} \rangle_R}{\langle e^{-\beta (V_A - V_B)} \rangle_R},
\]

(3.4)

The free energy difference can also be obtained from

\[
\ln \frac{\rho_A(\Delta V; \Delta V_{BA})}{\rho_B(\Delta V; \Delta V_{BA})} = +\beta \Delta V - \beta \Delta F_{BA},
\]

(3.5)

where \(\rho_A(\Delta V; \Delta V_{BA})\) and \(\rho_B(\Delta V; \Delta V_{BA})\) are the energy difference distributions sampled in the \(A\) and \(B\) ensembles, respectively, describing the probability that a particular energy difference \(V_B - V_A\) is sampled when performing simulations in state \(A\) and state \(B\), respectively. According to eq. (3.5) the free energy difference is equal to the potential energy difference \(\Delta V\) at the intersection of \(\rho_A\) and \(\rho_B\). To obtain these distributions from an EDS reference state simulation which yields only the energy difference distribution \(\rho_R\),

\[
\rho_R(\Delta V; \Delta V_{BA}) = \langle \delta [\Delta V - (V_B - V_A)] \rangle_R,
\]

(3.6)

rewetting of the distribution \(\rho_R\) to the \(A\) or \(B\) ensemble can be used,

\[
\rho_X(\Delta V; \Delta V_{BA}) = \frac{\langle \delta [\Delta V - \Delta V_{BA}] e^{-\beta (V_X - V_R)} \rangle_R}{\langle e^{-\beta (V_X - V_R)} \rangle_R}
\]

(3.7)

with \(X = A\) or \(B\), respectively. Equations (3.4) and (3.5) provide insight into the sampling requirements during an EDS simulation. While eq. (3.4)
Figure 3.1: The basic principle of EDS illustrated by two offset harmonic 1-dimensional potential energy functions with different curvatures: $V_A(x) = 3x^2$, $V_B(x) = (x - 3)^2$. (a) Potential energy curves, (b) energy difference distributions, (c) energy difference $\Delta V_{BA}(x) = V_B(x) - V_A(x)$. The reference potential energy function values $V_R(x)$ were computed with $\beta = 1$ and $E_{R}^i = 0$ (see eq. (3.3)). The four smoothness parameter values used were $s_1 = 1.0$, $s_2 = 0.23$, $s_3 = 0.15$, $s_4 = 0.08$.

points to the importance of sampling the relevant high energy tails which contribute most to the exponentials, eq. (3.5) relies on the accuracy of the distributions near their intersection region. The intersection region can only be sampled if the relevant parts of the configurational spaces of systems A and B exhibit sufficient overlap, which can be achieved by lowering the $s$-parameter in eq. (3.3). However, as the $s$-parameter gets smaller, sampling of the high energy tails becomes increasingly unlikely, as is illustrated using an analytical example in Section 3.2.2. The challenge for an automatic parameter update scheme is now to determine $s$ and $E_{B}^R$ such that both sampling requirements are met as much as possible.

3.2.2 Parameter update scheme

The parameter update scheme makes use of the energy difference $\Delta V_{BA}$ as a descriptor to identify in which state the simulation is sampling at a particular time. Given the potential energy curves for the harmonic model system shown in Fig. 3.1a, $\Delta V_{BA}$ becomes positive if state A is sampled while $\Delta V_{BA}$ becomes negative if state B is sampled (see Fig. 3.1c). Since equation 3.5 points out the importance of sampling the region where the energy difference distributions intersect, it is useful to define a transition region between the two end states. Because in general applications an analytical description of the potential energy as function of the spatial coordinates is not available,
this transition region can be estimated by performing a short simulation with a small $s$-parameter. Then the potential energy curve of the reference state looses its property of having two minima separated by an $s$-dependent barrier. The system is trapped in a global minimum and sampling is restricted to a narrow subset of the energy differences (case 4 in Fig. 3.1) that could also be sampled with a larger $s$-parameter. This _undersampling_ is very easy to detect in practice because the corresponding energy difference distributions are lacking any high energy tails. In practical applications the exact position and the width of the transition region were found to be less important because the energy differences sampled with good EDS parameters are orders of magnitude larger than those sampled with too small $s$-parameters. Once the position and the width of the transition region are determined, an iterative procedure can be started to obtain EDS parameters that lead to good sampling of the end states as well as the transition region. A flow-chart of the procedure is presented in Fig. 3.2. It is assumed that the procedure is started at undersampling conditions. Parameters to be specified are the energy difference, $E_{\text{sep}}$, that separates sampling of state A from sampling of state B as well as the width of the transition region $T$ which is specified by a parameter $\Delta E_{\text{tr}}$ such that every $\Delta V_{BA}$ that fulfills $E_{\text{sep}} - \Delta E_{\text{tr}} < \Delta V_{BA} < E_{\text{sep}} + \Delta E_{\text{tr}}$ is considered to belong to the transition region. Moreover the minimum number of transitions, $N_{\text{tr, min}}$, that should occur during a sampling period has to be specified. Next a short MD simulation is performed using initial EDS parameters that lead to undersampling. After this simulation the number of visits to states $X$ ($X = A, B, \text{or} T$), $N_X$, as well as the number of transitions, $N_{\text{tr}}$, is evaluated. Depending on which of the conditions - specified in the rhombic symbols in Fig. 3.2 - is fulfilled the $s$-parameter will be increased, decreased or kept constant. The factors determining the $s$-parameter changes were obtained empirically and might not be optimal for every system. The new relative energy offset $\Delta E_{BA}^{R}$ is obtained after calculating increments of the individual energy offsets according to \[\begin{align*}
E_A^{R}(\text{new}) &= -\beta^{-1} \ln \left( e^{-\beta(\Delta V_{BA} - \Delta E_{BA}^{R, \text{new}})} + 1 \right)_{s(\text{new}), E_i^{R}(\text{old})}^{-1} + E_A^{R}(\text{old}) \\
E_B^{R}(\text{new}) &= -\beta^{-1} \ln \left( e^{+\beta(\Delta V_{BA} - \Delta E_{BA}^{R, \text{new}})} + 1 \right)_{s(\text{new}), E_i^{R}(\text{old})}^{-1} + E_B^{R}(\text{old})
\end{align*}\] In these equations the ensemble average is defined such that the new $s$-parameter and the old energy offsets $E_i^{R}$ are used, where $E_A^{R}(\text{old}) = 0$. Next a short MD simulation is started using the updated EDS parameters. The length of the MD simulations between two updates depends on the phase of the update scheme. To this end four phases are distinguished. Phase 1
3.2 Theory

Figure 3.2: Flowchart of the update procedure to determine the smoothness parameter $s$ and the energy offsets $E_i^R$. Parameters to be specified in addition to initial EDS parameters are (i) $E_{\text{sep}}$, the energy separating sampling of state A from sampling of state B, (ii) $\Delta E_{\text{tr}}$, the width of the transition region, and (iii) $N_{\text{trmin}}$, the minimum number of transitions that should occur during a sampling period. After a short sampling period the numbers of configurations belonging to state X, $N_X$ ($X = A$, B, or T), are counted. Depending on which of the conditions, specified in the rhombic symbols, is fulfilled, the $s$-parameter will be increased, decreased or kept constant. New energy offsets are then calculated according to eq. (3.8). The factors determining the $s$-parameter changes and the occurrence of slight undersampling are empirical and may depend on the system studied.

corresponds to strong undersampling, i.e. the system spends more time in the transition region than in the two end states. In that case the $s$-parameter is increased by 10% at every update step. If the first condition is not fulfilled any more the next phase is entered in which the time between two updates is doubled. A longer sampling time is necessary because the $s$-parameter is now sufficiently large to sample at least one of the two end states. Moreover sampling becomes more sensitive to the energy offsets which requires bet-
Figure 3.3: Evolution of smoothness parameters $s$ (solid lines) and energy offsets $E = E_B^R$ (dashed lines) during EDS update simulations for the perturbations AAAAA → AAAAG (left) and AAAAA → AAAGG (right) in neat DNA. The free enthalpy differences $\Delta G$ calculated at each update point are displayed by dotted lines. The four update phases are separated by vertical lines. Initial values and update parameters were $s = 0.02$, $E_B^R = -158.5 \text{ kJ mol}^{-1}$, $E_{\text{sep}} = -158.0 \text{ kJ mol}^{-1}$, and $\Delta E_{\text{tr}} = 30 \text{ kJ mol}^{-1}$ for AAAAA → AAAAG and $s = 0.01$, $E_B^R = -317.2 \text{ kJ mol}^{-1}$, $E_{\text{sep}} = -316.9 \text{ kJ mol}^{-1}$, and $\Delta E_{\text{tr}} = 30 \text{ kJ mol}^{-1}$ for AAAAA → AAAGG.

The update scheme was tested and calibrated on a small set of tetrahydroisoquinoline derivatives solvated in water for which free energy changes related to alchemical perturbations were known from earlier work.\cite{320} A typical evolution of the EDS parameters during the parameter update procedure for one particular perturbation is shown in Fig. B.1 of the Supplementary Material. Without further changes the update scheme was then applied to the DNA-netropsin system studied in the present work. Typical evolutions of the EDS parameters during update simulations for the latter system are shown in Fig. 3.3. Compared to the update scheme proposed recently\cite{320} the present procedure requires also two parameters to identify undersampling conditions which are $E_{\text{sep}}$ and $\Delta E_{\text{tr}}$ for the present scheme and two threshold energies for the former. All other parameters such as the empirical factors appearing in the conditional blocks in Fig. 3.2 or the maximum fluctuation $\Delta E^R$ allowed for consecutive energy offsets are meant...
3.3 Methods

3.3.1 Simulation protocols

The DNA duplex consists of 13 base pairs, \textit{d(GCGCAAXXXGCGC)}\textit{•d(GCGCYYYTTGCGC)}, with X-Y representing either A-T or G-C base pairs, respectively. Initial coordinates for the neat DNA duplex and its complex with netropsin, for which X = A and Y = T, were taken from previous work.\textsuperscript{437} The DNA and DNA-netropsin complexes were solvated in cubic boxes using 10960 and 11097 simple point charge (SPC) water molecules,\textsuperscript{176} respectively, such that the minimal solute-to-wall distance was 1.1 nm. To neutralize the negative charge on the phosphate groups of the DNA backbone and to mimic the experimental conditions in drug-DNA binding experiments, the system containing DNA oligomers was neutralized with 44 Na\textsuperscript{+} and 20 Cl\textsuperscript{−} ions, while for the DNA-netropsin complex 42 Na\textsuperscript{+} and 20 Cl\textsuperscript{−} ions were employed, resulting in a salt concentration of approximately 100 mM NaCl. The initial coordinates of the counterions were obtained by replacing water molecules with the lowest or highest electrostatic potential by Na\textsuperscript{+} or Cl\textsuperscript{−} ions while preserving a minimal interionic distance of 0.35 nm. The AAAAA and AAAAA(NET) systems were relaxed by performing a steepest descent energy minimization with harmonic positional restraints, using a force constant of $2.5 \times 10^4 \text{kJ mol}^{-1} \text{nm}^{-2}$ on all solute atoms, followed by an equilibration period of 120 ps in which the strength of the positional restraints was gradually released from $2.5 \times 10^4$ to $0.0 \text{kJ mol}^{-1} \text{nm}^{-2}$ and the temperature was raised from 50 to 298 K.

All MD simulations were performed under periodic boundary conditions using a modified version of the GROMOS05 program package\textsuperscript{355} and the GROMOS45A4 force field.\textsuperscript{45} The equations of motion were integrated using the leapfrog algorithm with a time step of 2 fs. Bond lengths and the bond angle of water molecules were constrained by applying the SHAKE
algorithm\textsuperscript{358} with a relative geometric tolerance of $10^{-4}$. The temperature was maintained at 298 K by weak coupling to an external bath,\textsuperscript{359} with a relaxation time of 0.1 ps. The pressure was calculated with a molecular virial and held constant at 1 atm using the weak coupling method with a relaxation time of 0.5 ps\textsuperscript{359} and an isothermal compressibility of $4.575 \times 10^{-4} \text{(kJ mol}^{-1} \text{nm}^{-3})^{-1}$. The nonbonded van der Waals and electrostatic interactions were calculated using a triple-range cutoff scheme. The short-range interactions were calculated every time step by updating the molecular or charge-group pair list for distances smaller than the first cutoff radius of 0.8 nm. Interactions between molecules or charge-groups with distances between 0.8 nm and 1.4 nm were reevaluated every fifth time step and kept constant between reevaluations. The long-range electrostatic interactions beyond the outer cutoff of 1.4 nm were represented by a reaction field\textsuperscript{356,446} with $\varepsilon_{RF} = 61.447$ The motion of the center of mass was removed every 2 ps.

End states were simulated for 11 ns starting from the AAAAA and AAAAAA(NET) configurations at the end of the equilibration simulation after changing the appropriate atoms corresponding to the desired end state. The last 10 ns were used for further analysis.

EDS reference state simulations were conducted using different methods to determine the required smoothness parameter and energy offset. In a first attempt a series of 127 simulations each of 100 ps duration was performed, using the algorithm proposed in ref. 434 to update the parameters after 1, 3, 7, 15, 63, and 127 simulation periods.\textsuperscript{435} Initial parameters were $s = 1.0$ and $E_i^R = 0.0 \text{kJ mol}^{-1}$. In a second attempt the parameters were determined from the new update scheme proposed in the present work. Production simulations with fixed EDS parameters were run for 11 ns. The last 10 ns were used for further analysis.

### 3.3.2 Topology

EDS reference state simulations were conducted within the single topology approach.\textsuperscript{448} All end and reference state simulations were based on a common topology containing three hybrid building blocks in each DNA strand (DDAG and DDTC, see Fig. 3.4) that include all atoms needed for the simulation of A-T or G-C base pairs, respectively. In contrast to the TI calculation reported earlier\textsuperscript{437} all bond lengths in the 6-membered rings of the hybrid building blocks were kept fixed to those of the adenine-thymine values when mutating one base into the other. Because the change of bond lengths during an alchemical transformation modifies the solute-solvent and
3.3 Methods

Figure 3.4: Single topology representation of the hybrid building blocks DDAG and DDTC consisting of a common core (black) and all atoms needed to simulate A-T (red) and G-C (green) base pairs.

solvent-solvent interactions\textsuperscript{12} an additional TI calculation for the perturbation AAAAA\textsuperscript{(NET)} → AAGAA\textsuperscript{(NET)} was performed using a simulation setup similar to the one used before\textsuperscript{437} but keeping all bond lengths constant (see Supplementary Material for additional technical details). The difference in free energy or rather free enthalphy\textsuperscript{8} or Gibbs energy\textsuperscript{9} between A-T and G-C, $\Delta G = -155.8 \pm 3.4 \text{ kJ mol}^{-1}$, calculated with constant bond lengths agrees very well with the value, $\Delta G = -156.2 \pm 3.4 \text{ kJ mol}^{-1}$, reported earlier,\textsuperscript{437} showing that the contribution from the change of bond lengths to the free enthalpy difference between A-T and G-C is negligible in the present case.

3.3.3 Trajectory analysis

Atom coordinates and energies were written to file every 2 ps. The trajectories for all end states were analysed in terms of atom-positional root-mean-square deviation (rmsd) from the starting configuration of each production simulation using the backbone atoms to perform the superposition of centres of mass and rotational least-squares fit superposition.\textsuperscript{449} The rmsd was calculated based on (i) all of the heavy atoms, (ii) only the heavy atoms of the DNA backbone, (iii) only the heavy atoms of the bases, and (iv) only the heavy atoms of netropsin. Moreover, all end and reference state trajectories were analysed in terms of intrasolute Watson-Crick (WC) hydrogen bonds. These were assumed to exist between the base pairs if the distance between
the hydrogen and the acceptor atom was smaller than 0.25 nm and the angle between the donor hydrogen and the acceptor hydrogen vectors was larger than 135°. Distributions of the solute-solute plus solute-solvent nonbonded energies were calculated for all end state trajectories in order to compare them with reweighted end state energy distributions extracted from EDS reference state simulations. Since the perturbations only involve the solute, solvent-solvent potential energy contributions can safely be omitted.

3.4 Results and discussion

3.4.1 Structural properties

Time series of the atom-positional rmsd of the DNA and netropsin-DNA heavy atoms with respect to the first structure of each production simulation are displayed in Fig. B.2 of the Supplementary Material. For the neat DNA duplex the rmsd is mostly below 0.5 nm. More importantly no drift is displayed during the 10 ns of sampling time. The DNA-netropsin complexes show on average a smaller rmsd than the neat DNA duplexes, an observation that was also made in previous work.437 In some cases the rmsd of netropsin shows relatively large values. This can be ascribed to rotations of the terminal groups of netropsin (see Supplementary Materials for details).

The high occurrence of WC-hydrogen bonds between the bases in DNA duplexes and DNA-netropsin complexes presented in Figs B.3 and B.4 of the Supplementary Material provide additional evidence for the stability of the model structures over the sampling period. A disruption of the terminal WC base pairing which is rather common in MD simulations45,453 occurs in some cases (see Supplementary Material) but is not expected to affect the outcome of this study because the binding site is located in the center of the structure.

3.4.2 Free enthalpy differences for perturbations in neat DNA

The seven transformations of A-T into G-C pairs for neat DNA and DNA-netropsin are summarized in Table 3.1. The corresponding free enthalpy differences calculated by TI and EDS are listed in Table 3.2. For each transformation two EDS results are reported. The first set (EDS1) was calculated using reference state parameters obtained with the originally proposed update scheme434,435 while the second set (EDS2) was obtained with the parameter update scheme proposed in this work. The s and $E_R^B$ parameters are
### Table 3.1: Summary of the A-T/G-C base pair perturbations in the d(GCGCAAAAAGCGC)•d(GCGCTTTTTTGCGC) DNA duplex free or complexed with netropsin (NET) in aqueous solution.

<table>
<thead>
<tr>
<th>Perturbation</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAA → AAAAG</td>
<td>1</td>
</tr>
<tr>
<td>AAAAA → AAAGA</td>
<td>2</td>
</tr>
<tr>
<td>AAAAA → AAGAA</td>
<td>3</td>
</tr>
<tr>
<td>AAAAA → AAAGG</td>
<td>4</td>
</tr>
<tr>
<td>AAAAA → AAGAG</td>
<td>5</td>
</tr>
<tr>
<td>AAAAA → AAGGA</td>
<td>6</td>
</tr>
<tr>
<td>AAAAA → AAGGG</td>
<td>7</td>
</tr>
<tr>
<td>AAAAA(NET) → AAAAG(NET)</td>
<td>1n</td>
</tr>
<tr>
<td>AAAAA(NET) → AAAGA(NET)</td>
<td>2n</td>
</tr>
<tr>
<td>AAAAA(NET) → AAGAA(NET)</td>
<td>3n</td>
</tr>
<tr>
<td>AAAAA(NET) → AAAGG(NET)</td>
<td>4n</td>
</tr>
<tr>
<td>AAAAA(NET) → AAGAG(NET)</td>
<td>5n</td>
</tr>
<tr>
<td>AAAAA(NET) → AAGGA(NET)</td>
<td>6n</td>
</tr>
<tr>
<td>AAAAA(NET) → AAGGG(NET)</td>
<td>7n</td>
</tr>
</tbody>
</table>

listed in Table 3.3. The EDS1 results are generally less negative than the TI ones with differences of up to 20 kJ mol\(^{-1}\) while the EDS2 results are in good agreement with TI. The largest difference occurs for perturbations 4 and 6 (4.8 kJ mol\(^{-1}\)). However, this value is still within the error estimates from the TI calculation. Repeated calculations with other EDS parameters for perturbation 4 confirmed the free enthalpy reported in Table 3.2. A TI simulation longer than the one reported in ref. 437 was hampered by instabilities of the DNA model structure for intermediate \(\lambda\)-values. From Table 3.3 it becomes evident that the original update scheme yields systematically lower values for the \(s\)-parameter. As a consequence the high energy tails of the energy difference distributions are less well sampled which results in poorer estimates of the free enthalpy because the energy difference distributions are asymmetric and, hence, the missing parts of the high energy tails do not cancel each other. Figure 3.5 shows the time series of \(\Delta V_{BA}\), the difference of the end state energies in the seven reference state simulations. The EDS1 parameters result in significant undersampling (red curve) while the EDS2 parameters lead to extensive sampling of the high energy tails. The time
Table 3.2: Comparison of free enthalpy differences for seven base pair perturbations in neat DNA and DNA-netropsin complexes calculated by TI and EDS using the EDS parameter update scheme EDS1 of refs. 434,435 and using the update scheme EDS2 proposed here.

<table>
<thead>
<tr>
<th>Code</th>
<th>TI</th>
<th>EDS1</th>
<th>EDS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−166.9 ± 3.7</td>
<td>−161.7 ± 1.7</td>
<td>−167.2 ± 0.4</td>
</tr>
<tr>
<td>2</td>
<td>−171.8 ± 3.5</td>
<td>−157.1 ± 1.9</td>
<td>−168.3 ± 0.4</td>
</tr>
<tr>
<td>3</td>
<td>−170.5 ± 3.8</td>
<td>−162.6 ± 1.3</td>
<td>−171.5 ± 0.4</td>
</tr>
<tr>
<td>4</td>
<td>−330.3 ± 5.8</td>
<td>−322.7 ± 2.0</td>
<td>−335.1 ± 0.7</td>
</tr>
<tr>
<td>5</td>
<td>−336.5 ± 5.5</td>
<td>−318.6 ± 1.6</td>
<td>−336.0 ± 0.7</td>
</tr>
<tr>
<td>6</td>
<td>−341.9 ± 6.1</td>
<td>−320.3 ± 2.5</td>
<td>−337.1 ± 0.7</td>
</tr>
<tr>
<td>7</td>
<td>−471.5 ± 2.7</td>
<td>−502.1 ± 1.2</td>
<td>−502.1 ± 1.2</td>
</tr>
<tr>
<td>1n</td>
<td>−169.8 ± 3.1</td>
<td>−161.2 ± 1.1</td>
<td>−165.6 ± 0.6</td>
</tr>
<tr>
<td>2n</td>
<td>−157.5 ± 4.1</td>
<td>−156.0 ± 1.0</td>
<td>−158.5 ± 1.3</td>
</tr>
<tr>
<td>3n</td>
<td>−156.2 ± 3.4</td>
<td>−158.6 ± 1.5</td>
<td>−161.9 ± 1.1</td>
</tr>
<tr>
<td>4n</td>
<td>−324.0 ± 5.9</td>
<td>−309.9 ± 2.4</td>
<td>−321.5 ± 1.2</td>
</tr>
<tr>
<td>5n</td>
<td>−322.2 ± 5.8</td>
<td>−308.5 ± 2.2</td>
<td>−326.1 ± 1.2</td>
</tr>
<tr>
<td>6n</td>
<td>−298.3 ± 6.0</td>
<td>−304.5 ± 2.6</td>
<td>−312.5 ± 1.1</td>
</tr>
<tr>
<td>7n</td>
<td>−453.9 ± 1.7</td>
<td>−475.0 ± 1.9</td>
<td>−475.0 ± 1.9</td>
</tr>
</tbody>
</table>

series of $\Delta V_{BA}$ indicate that the two end state potential energy functions possess different widths, because $\Delta V_{BA}$ has a smaller range of positive values than of negative values, which will be the case when $V_A$ (state AAAAA) is narrower than $V_B$ (states involving G-C pairs), see Fig. 3.1c. This asymmetry is also reflected in the energy difference distributions shown in the left column of Fig. 3.6. In all cases the reweighted distributions exhibit a long tail in the negative regime showing that the free enthalpy estimate contains high energy contributions. The energy difference distributions corresponding to the EDS1 parameters are shown in Fig. B.5 of the Supplementary Material and display the typical shape for undersampling conditions, i.e. a lack of high energy tails. Also shown in Fig. 3.6 (right column) are the distributions of the solute-solute plus solute-solvent nonbonded energies obtained from the EDS reference state trajectories through reweighting and of those from non-EDS simulations of the end states. In all cases the EDS and non-EDS end state distributions overlap providing additional evidence that those parts of the phase space that are relevant for the end states were properly sampled in the EDS reference state simulations.
### Table 3.3: EDS reference state parameters $s$ and $E^R_B$ (kJ mol$^{-1}$) obtained from two different parameter update schemes, EDS1 and EDS2. $E^R_A$ is set to zero.

<table>
<thead>
<tr>
<th>Code</th>
<th>EDS1, ref. 434,435</th>
<th>EDS2, this work</th>
<th>EDS2, this work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0102</td>
<td>−158.5</td>
<td>0.0665</td>
</tr>
<tr>
<td>2</td>
<td>0.0106</td>
<td>−157.1</td>
<td>0.0640</td>
</tr>
<tr>
<td>3</td>
<td>0.0109</td>
<td>−156.1</td>
<td>0.0584</td>
</tr>
<tr>
<td>4</td>
<td>0.0048</td>
<td>−317.2</td>
<td>0.0291</td>
</tr>
<tr>
<td>5</td>
<td>0.0048</td>
<td>−314.4</td>
<td>0.0310</td>
</tr>
<tr>
<td>6</td>
<td>0.0049</td>
<td>−314.7</td>
<td>0.0291</td>
</tr>
<tr>
<td>7</td>
<td>0.0031</td>
<td>−475.8</td>
<td>0.0199</td>
</tr>
<tr>
<td>1n</td>
<td>0.0103</td>
<td>−157.0</td>
<td>0.0974</td>
</tr>
<tr>
<td>2n</td>
<td>0.0116</td>
<td>−154.4</td>
<td>0.0640</td>
</tr>
<tr>
<td>3n</td>
<td>0.0116</td>
<td>−152.0</td>
<td>0.0608</td>
</tr>
<tr>
<td>4n</td>
<td>0.0054</td>
<td>−305.0</td>
<td>0.0320</td>
</tr>
<tr>
<td>5n</td>
<td>0.0051</td>
<td>−304.3</td>
<td>0.0320</td>
</tr>
<tr>
<td>6n</td>
<td>0.0051</td>
<td>−304.0</td>
<td>0.0259</td>
</tr>
<tr>
<td>7n</td>
<td>0.0033</td>
<td>−450.1</td>
<td>0.0200</td>
</tr>
</tbody>
</table>

The energy difference distributions of the reference states, $\rho_R(\Delta V; \Delta V_{BA})$ all have peaks close to the intersection point of the end state energy difference distributions, a property suggested by eq. (3.5). However, it seems that this is not essential for a reliable free enthalpy estimate because simulations with other sets of EDS parameters lead to completely different shapes of $\rho_R(\Delta V; \Delta V_{BA})$ without changing the free enthalpy estimates significantly. For the perturbations AAAAA → AAAGA, AAAAA → AAAGG, and AAAAA → AAGAG we show in the Supporting Information that the free enthalpy estimate is insensitive to a change of the $s$ and $E^R_B$ parameters over a wide range of values. However, for the perturbation AAAAA → AAGGG a change of the energy offset by 1 kJ mol$^{-1}$ already has a strong effect on the sampling (see Supporting Information) showing that appropriate EDS parameters are more difficult to find the larger the perturbation becomes. By analysing the time series of $\Delta V_{BA}$ together with the energy difference and end state energy distributions insufficient sampling can be detected easily.
Figure 3.5: Time series of the energy difference $\Delta V_{BA} = V_B - V_A$ in an EDS reference state simulation of neat DNA, in which B represents a G-C base pair at position X and A represents an A-T base pair at position X, for the seven different reference states. Upper row: AAAAX, AAAXA, AAXAA. Middle: AAAXX, AAXAX, AAXXA. Bottom: AAXXX.

### 3.4.3 Free enthalpy differences for perturbations in DNA-netropsin complexes

The free enthalpy differences for the seven transformations of A-T into G-C pairs in DNA-netropsin complexes are reported in Table 3.2. The agreement between TI and EDS2 results is good except for the perturbation AAAAAA(NET) $\rightarrow$ AAGGA(NET) where EDS2 predicts a free enthalpy change that is more negative by $14.2 \text{kJ mol}^{-1}$. Although the sampling of state AAAAAA was not as good as that of AAGGA in the EDS reference state simulation as suggested by the smaller range of positive $\Delta V_{BA}$ values in Fig. 3.7 and the more rugged appearance of the reweighted end state energy distribution shown in Figure 3.8, it is not expected that this accounts for all of the discrepancy. This is supported by additional simulations employing a variety of $s$ and $E_R$ parameters all resulting in free enthalpy differences that are close to the one reported in Table 3.2 (data not shown). A TI simulation longer than the one reported in ref. 437 was not successful due to the tendency...
of netropsin to drift away from the AAGGA binding site for intermediate \( \lambda \)-values. However, additional support for the free enthalpy difference calculated in the present work is obtained from considering the relative binding free enthalpies (see next section). The \( s \) and \( E_R^B \) parameters employed are listed in Table 3.3. As was the case for the neat DNA simulations the original update scheme yields smaller \( s \)-parameters. In two cases (2n and 3n) the free enthalpy differences obtained with EDS1 and EDS2 parameters are in good agreement. However, this results entirely from error compensation and does not mean sufficient sampling with the EDS1 parameters. They cause clear undersampling as can be seen from the time series of \( \Delta V_{BA} \) shown in Figure 3.7 (red curves) and the corresponding energy difference distributions shown in Fig. B.5 of the Supplementary Material. Figure 3.7 also shows that

**Figure 3.6:** Left: Energy difference distributions for the reference state, \( \rho_R(\Delta V; \Delta V_{BA}) \) (black), and the two end states, \( \rho_A(\Delta V; \Delta V_{BA}) \) (red), \( \rho_B(\Delta V; \Delta V_{BA}) \) (green), for each of the seven perturbations in neat DNA as obtained from the EDS2 reference state simulations. The energy difference distributions of the end states were determined by reweighting. Right: Non-bonded solute-solute plus solute-solvent energy distributions of the eight EDS end states obtained through reweighting (full lines) from the seven EDS2 reference state simulations and from independent MD simulations of the eight end states (dashed lines). The red colour corresponds to end state AAAAA.
3.4.4 Relative binding free enthalpies of netropsin to different DNA binding sequences

The computed relative free enthalpies of binding from both EDS and TI are listed in Table 3.4. As described in the methods section, the 5’-AAAAA-3’ binding sequence, to which netropsin binds in a 1:1 high affinity mode, was taken as reference in all free enthalpy calculations. Both approaches predict that the sequences AAAGA, AAGAA, and AAGAG do not differ in their binding affinity towards netropsin with the EDS results being consistently
3.4 Results and discussion

Figure 3.8: Left: Energy difference distributions for the reference state, \( \rho_R(\Delta V; \Delta V_{BA}) \) (black), and the two end states, \( \rho_A(\Delta V; \Delta V_{BA}) \) (red), \( \rho_B(\Delta V; \Delta V_{BA}) \) (green), for each of the seven perturbations in DNA-netropsin as obtained from the EDS2 reference state simulations. The energy difference distributions of the end states were determined by reweighting. Right: Non-bonded solute-solute plus solute-solvent energy distributions of the eight EDS end states obtained through reweighting (full lines) from the seven EDS2 reference state simulations and from independent MD simulations of the eight end states (dashed lines). The red colour corresponds to end state AAAAA.

smaller than the TI data. Moreover, TI and EDS confirm that the exchange of the A-T base pair at the end of the binding site for a G-C pair does not affect the binding affinity. The largest discrepancy between TI and EDS is obtained for netropsin binding to AAGGA. Here the TI result is more positive by 19 kJ mol\(^{-1}\). Support for the EDS result is provided by the result for AAGGG which is similar to that for AAGGA as expected due to the small effect of changing an A-T into a G-C base pair at the end of the binding site on the binding free enthalpy of netropsin.

A comparison with experimental data is not the main purpose of this work but yields additional support for the computed relative free enthalpies of binding and for the force field used. The general observation that binding of DNA to netropsin becomes more unfavourable if the substitution
Table 3.4: Change in binding free enthalpies for netropsin-DNA binding for seven base pair changes as obtained from TI and EDS at 298 K and 1 atm. Errors are estimated from the square root of the sum of the squares of the fractional errors.

<table>
<thead>
<tr>
<th>AAAAAA →</th>
<th>ΔΔG^{binding} [kJ mol(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAG</td>
<td>-2.9 ± 4.8</td>
</tr>
<tr>
<td>AAAGA</td>
<td>14.3 ± 5.4</td>
</tr>
<tr>
<td>AAGAA</td>
<td>14.3 ± 5.1</td>
</tr>
<tr>
<td>AAAGG</td>
<td>6.3 ± 8.3</td>
</tr>
<tr>
<td>AAGAG</td>
<td>14.3 ± 8.0</td>
</tr>
<tr>
<td>AAGGA</td>
<td>43.6 ± 8.6</td>
</tr>
<tr>
<td>AAGGG</td>
<td>17.6 ± 3.2</td>
</tr>
<tr>
<td>AAAAG</td>
<td>0.5 ± 2.0</td>
</tr>
<tr>
<td>AAGAG</td>
<td>1.1 ± 2.1</td>
</tr>
<tr>
<td>AAGGA</td>
<td>4.0 ± 2.0</td>
</tr>
<tr>
<td>AAAGG</td>
<td>12.8 ± 3.1</td>
</tr>
<tr>
<td>AAGAG</td>
<td>10.1 ± 2.7</td>
</tr>
<tr>
<td>AAGGA</td>
<td>15.8 ± 3.6</td>
</tr>
<tr>
<td>AAGGG</td>
<td>24.6 ± 1.3</td>
</tr>
<tr>
<td>EDS1</td>
<td>1.6 ± 0.7</td>
</tr>
<tr>
<td>EDS2</td>
<td>9.8 ± 1.4</td>
</tr>
<tr>
<td></td>
<td>9.6 ± 1.2</td>
</tr>
<tr>
<td></td>
<td>13.6 ± 1.4</td>
</tr>
<tr>
<td></td>
<td>9.9 ± 1.4</td>
</tr>
<tr>
<td></td>
<td>24.6 ± 1.3</td>
</tr>
<tr>
<td></td>
<td>27.1 ± 2.2</td>
</tr>
</tbody>
</table>

Table 3.5: Experimental binding affinities for the binding of netropsin to different sites in a DNA duplex.

<table>
<thead>
<tr>
<th>binding site</th>
<th>ΔG^0_B (kJ mol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAAA/TTTTT</td>
<td>-54.0 [455]</td>
</tr>
<tr>
<td>AAAAAA/TTTTT</td>
<td>-51.5 [457]</td>
</tr>
<tr>
<td>AAAAAA/TTTTT</td>
<td>-51.0 [458]</td>
</tr>
<tr>
<td>ATATA/TATAT</td>
<td>-54.8 [457]</td>
</tr>
<tr>
<td>ATATA/TATAT</td>
<td>-53.1 [458]</td>
</tr>
<tr>
<td>AAGTT/AACTT</td>
<td>-42.3 [455]</td>
</tr>
<tr>
<td>GCGCG/CGCGC</td>
<td>-29.7 [458]</td>
</tr>
<tr>
<td>GCGCG/CGCGC</td>
<td>-28.9 [459]</td>
</tr>
</tbody>
</table>

of A-T into G-C takes place in the middle of the binding site rather than at the end is in agreement with the results of Boger et al.,\(^\text{456}\) who report the binding affinity and sequence selectivity of 512 base pair combinations. From their data it can also be confirmed that AAAGA and AAGAA show essentially the same loss of binding affinity relative to AAAAA and that the loss of binding affinity when going from AAAAG to AAAGA is significantly higher than when going from AAAAA to AAAAG. We guess that the reason for this observation lies in the structural differences between the rigid body and the flexible tails of the netropsin molecule.\(^\text{460,461}\) For the combinations AAAGG and AAGAG Boger et al.,\(^\text{456}\) report a lower binding affinity for
the latter while the EDS results suggest that AAGAG has a slightly higher affinity towards netropsin than AAAGG. For AAGGA the binding affinity decreases significantly compared to AAAGG, a result that is in agreement with the simulations. Moreover AAGGA and AAGGG show very similar binding affinities in line with our results. A rough quantitative comparison with experimental data can be made based on the reported binding free enthalpies listed in Table 3.5. These data suggest that the substitution of an A-T into a G-C pair in the middle of the binding site leads to a decrease in the binding affinity of about 10 kJ mol$^{-1}$, assuming that AAAAA and AAATT have a similar affinity, which is in good agreement with our calculations. The difference between the affinity for the GCGCG/CGCGC and AAAAA/TTTTT sites of about 25 kJ mol$^{-1}$ can be used as an upper bound for the loss in binding affinity, which is also in reasonable agreement with our calculated data.

3.5 Conclusion and outlook

For the estimation of free enthalpy differences by molecular simulation it is well known that convergence and the amount of sampling are strictly connected. Using the EDS methodology the extent of sampling can be conveniently evaluated making the identification of unconverged or pseudoconverged estimates relatively easy. Moreover parts of the analysis can be carried out “on the fly” in a simulation by keeping track of the time series of $\Delta V_{BA}$ which allows to detect insufficient sampling in an early stage of the simulation. The proposed parameter update scheme facilitates an automatic parametrization of the EDS reference state Hamiltonian. For large perturbations it might be necessary to test different energy offsets $E_B^R$ in addition to the one suggested by the update scheme, while the determination of the smoothness parameter $s$ was found to be quite robust. The computational effort of EDS compared to TI for a two state system is expected to be comparable. A precise estimate of the difference in computational efficiency will depend on the degree of difference between the two end-state Hamiltonians, upon the choice of $\lambda$-dependence of the Hamiltonian in TI, and upon the required precision of the resulting free energy differences. However, EDS offers better control of the sampling quality and needs no choice of pathway which is expected to be advantageous if the conventional $\lambda$-dependent pathways connecting the two states contain energy barriers that inhibit proper sampling. Since EDS is a relatively recently proposed method, its potential is not yet fully exploited. However, from the present study together with previous work.$^{251,320,434-436}$
some practical hints can be established that help the user to set up an EDS calculation and to assess the reliability of the obtained free enthalpy estimates. For the parameter update scheme our recommendations are:

1. Start the parameter update scheme at an undersampling condition: Undersampling is easy to detect due to the absence of fluctuations in the energy difference time series. Initial parameters for the update scheme can therefore be generated from a short EDS simulation with a small s-parameter and $E_{RB}^R = 0$. The mean value of $\Delta V_{BA}$ can be used as estimate for the $E_{sep}$-parameter while the root-mean-square fluctuation of $\Delta V_{BA}$ can be used as estimate for the parameter $\Delta E_{tr}$. The initial energy offset can be set to $E_{sep}$.

2. Extract a range of energy offsets for the production simulation: Convergence of the parameter update scheme towards an optimal set of s and $E_{RB}^R$ parameters becomes more difficult the bigger the perturbations are. In such cases it is advisable to run several production simulations sharing the same s-parameter but differing slightly in the energy offset $E_{RB}^R$. For the free energy estimate the simulation with the most balanced sampling should be chosen.

At the postprocessing state of an EDS production simulation the following points should be addressed:

1. Monitoring of $\Delta V_{BA}(t)$. The time series of $\Delta V_{BA}$ provides information about the frequency of transitions between the end states and about the time the system spends in each state. For balanced sampling the system should switch states on a regular basis.

2. Appearance of $\rho_X(\Delta V; \Delta V_{BA})$. The energy difference distribution $\rho_R$ should exhibit high energy tails. It may have quite different shapes depending on the character of the system and perturbation.

3. Comparison of end state distributions from EDS and non-EDS simulations. Ideally, the reweighted end state energy distributions from the EDS reference state simulation should match the energy distributions from the non-EDS end state simulations. In cases of insufficient sampling the distributions reduce to one or a few sharp peaks. For complex systems such as ligand binding the distributions from the non-EDS simulation may have the same shape as the EDS distributions but may be slightly offset. This can happen if a ligand is able to bind at different positions in a large binding pocket, or in cases where the sampled structure displays high flexibility associated with slow degrees of freedom. In such cases a careful analysis of the reasons for the discrepancies in the distributions is mandatory.
Relative free enthalpies for point mutations in two proteins with highly similar sequences but different folds

Enveloping distribution sampling was used to calculate free enthalpy changes associated with single amino acid mutations for a pair of proteins, $G_A$95 and $G_B$95, that show 95% sequence identity yet fold into topologically different structures. Of the $L \rightarrow A$, $I \rightarrow F$, and $L \rightarrow Y$ mutations at positions 20, 30, and 45, respectively, of the 56-residue sequence, the first and the last contribute the most to the free-enthalpy difference between the native and non-native sequence-structure combinations, in agreement with the experimental findings for this protein pair. The individual free-enthalpy changes are almost sequence-independent in the fourstrnad/one-helix structure, the stable form of $G_B$95, while in the three-helix bundle structure, the stable form of $G_A$95, an interplay between residues 20 and 45 is observed.
4.1 Introduction

The three-dimensional structure a protein adopts in solution is a complex function of the amino acid sequence and the properties of the surrounding solvents, making de novo structure prediction a very challenging problem. Although on-going refinement of NMR structure prediction tools will continuously enhance the automation of the sequence to structure mapping and will push the size and complexity limit of these methods forward, failures in automatic fold recognition are reported occasionally, showing that some fundamental principles underlying the way in which the protein sequence determines structure are not entirely understood. The question of how much of the sequence is required to specify the final fold has recently been addressed experimentally by systematic modifications of two 56-residue proteins derived from the binding domains of Streptococcus protein G with increasing levels of sequence identity (SI). It was demonstrated that one single residue may act as a switch to change the conformation from a three-helix bundle (3α-structure) into a four-strand β-sheet with a single α-helix (4β+α-structure). Eighty-five percent of residues change their secondary structure, with only eight residues in the central α-helix plus two turn residues retaining the same conformation in both forms. Clearly, this system poses a challenge to automatic structure prediction methods. Two pairs of proteins in the series, GA95 and GB95, differing only in three amino acids were used as targets in the CASP8 structure prediction competition. GA95 has amino acids Leu, Ile and Leu at positions 20, 30 and 45 and adopts the 3α-structure while GB95 has Ala, Phe and Tyr and adopts the 4β+α-structure. Only four out of 159 participants recognized the two different folds. In particular, methods that rely on comparative modeling approaches failed for this set of proteins, while structure determination methods based on chemical shifts were shown to be more successful. Due to the nonlocal nature of the physical processes which direct the folding of proteins, and the strong dependence of the backbone conformation on solvent affinities of amino acid side chains, atomistic simulations in explicit solvent are a promising tool to study these phenomena. However, the direct sampling of protein folding by molecular dynamics (MD) is in general not feasible, except for small proteins (less than 50 residues). The long MD runs necessary to study folding directly require special purpose hardware that is not widely available. Moreover, for a reliable estimate of the free energy or rather free enthalpy or Gibbs energy of folding, many folding-unfolding events have to be sampled which is far from being feasible. Another issue in this context arises from the short timescales usually used in force-field parameter-
4.1 Introduction

Figure 4.1: Thermodynamic cycle used for the calculation of relative free enthalpy differences $\Delta \Delta G_{BA}^{\beta \alpha} = \Delta G_{BA}^{\beta} - \Delta G_{BA}^{\alpha}$, where A and B denote two different amino acid sequences while $\alpha$ and $\beta$ denote two different folds. The vertical branches of the cycle can be computed using EDS.

In the present work we make use of the fact that in a relatively short MD simulation (10-20 ns) of protein G no unfolding is observed, even if the non-native sequence-structure combinations are simulated. This allows the use of the thermodynamic cycle shown in Fig. 4.1, in which the vertical branches denote a change in the amino acid sequence within a given tertiary structure. In a previous study we investigated the two proteins G$_A$95 and G$_B$95 by a combination of structural and energetic analyses. However, unequivocal support of the experimental findings by means of the simulation data was not obtained. In the present work we study in detail the changes in free enthalpy associated with single amino acid mutations in the two folds. For this purpose we use the enveloping distribution sampling (EDS) methodology, which belongs to the class of expanded ensemble methods that find widespread use in Monte Carlo and molecular dynamics simulations in cases where non-overlapping phase spaces need to be sampled. EDS has been applied successfully to a number of model systems and, after some refinement of the parameter determination procedure, to more complex alchemical perturbations such as occur in the calculation of the relative binding free enthalpies of phenylethanolamine N-methyltransferase (PNMT) inhibitors, and of netropsin to various base pair combinations in the DNA minor groove. It was also shown to be useful for the calculation of relative free enthalpies between different conformations of helices. However, as
EDS is a relatively recently proposed method its potential is not yet fully exploited. To evaluate the performance of EDS as an alternative to thermodynamic integration (TI), a question of particular relevance is the allowed degree of non-overlap of the important phase space regions of the EDS end states. By applying the method to different systems its limits can be explored. The second aim of the present work is a further step towards this goal.

4.2 Theory

The EDS scheme is a special form of the umbrella sampling approach that focuses the sampling on the relevant phase space of a system presenting $N_H$ different states $n$ using a reference Hamiltonian $H_R$ defined as

$$H_R(r, p) = -\frac{1}{\beta s} \ln \sum_{n=1}^{N_H} e^{-\beta s (H_n(r,p) - E_{Rn})}, \quad (4.1)$$

where $r$ and $p$ denote the $3N$-dimensional vectors of spatial coordinates and conjugate momenta, respectively, $N$ being the number of atoms in the system, and $\beta = 1/k_B T$, $k_B$ being the Boltzmann constant and $T$ the absolute temperature. The smoothness parameter $s (0 < s \leq 1)$ controls the barriers determining the rate of transitions between the different states $n$. The energy offsets $E_{Rn}$ of the different states serve to bring their energies to the same level for even sampling of the states $n$. These parameters have to be determined iteratively prior to the EDS production simulation.

The free-enthalpy difference between a pair of states $(n, m)$ is calculated based on this production simulation using

$$G_n - G_m = -\frac{1}{\beta} \ln \frac{\langle e^{-\beta (H_n(r,p) - H_R(r,p))} \rangle_R}{\langle e^{-\beta (H_m(r,p) - H_R(r,p))} \rangle_R}, \quad (4.2)$$

where $\langle ... \rangle_R$ denotes ensemble averaging over a simulation generated using the reference-state Hamiltonian $H_R$ (Eq. 4.1).

In force fields typically used for molecular simulations, the potential energy function is independent of velocities. As a result, the kinetic contribution to the free enthalpy can be calculated analytically. For a two-state system, the potential energy term of the reference-state Hamiltonian then reads

$$V_R(r; s, \Delta E_{BA}^R) = -\frac{1}{\beta s} \ln \left[ e^{-\beta s (V_A(r) - E_{AR})} + e^{-\beta s (V_B(r) - E_{BR})} \right] \quad (4.3)$$
In the reference-state ensemble R, the distribution of the energy difference \( \Delta V_{BA}(t) = V_B(t) - V_A(t) \) between the two end states can be calculated as \(^{439,440}\)

\[
\rho_R(\Delta V_{BA}) = \langle \delta [\Delta V_{BA} - (V_B(r) - V_A(r))] \rangle_R , \tag{4.4}
\]

where \( \delta \) is the Dirac delta function (approximated in practice by a histogram-binning function). This distribution can be converted to predicted distributions, denoted by a tilde, corresponding to the end-state ensembles A or B via reweighting, \(^{441,442}\)

\[
\tilde{\rho}_n(\Delta V_{BA}) = \frac{\langle \delta [\Delta V_{BA} - (V_B(r) - V_A(r))] e^{-\beta(V_n(r) - V_R(r))} \rangle_R}{\langle e^{-\beta(V_n(r) - V_R(r))} \rangle_R}, \tag{4.5}
\]

with \( n = A \) or \( B \), respectively. The distributions given by Eqs 4.4 and 4.5 are used to gain insight into the quality of the sampling obtained in the EDS simulation.\(^5\)

In a similar manner, distributions of the potential energies \( V_A \) and \( V_B \) sampled in the reference-state ensemble R, \( \rho_R(V_n) \), can be converted to predicted distributions corresponding to the end-state ensembles A or B, \( \tilde{\rho}_n(V_n) \). These predicted distributions can be compared to actual distributions \( \rho_n(V_n) \) based on independent simulations at the two end states, and thus used for further assessment of the quality of the sampling obtained in the EDS simulation.\(^6\)

### 4.3 Computational details

#### 4.3.1 Simulation parameters

All MD simulations were carried out using a modified version of the GROMOS05 program package\(^3\) and the 54A7 force field.\(^4\) The simulations were performed under periodic boundary conditions based on cubic boxes. Initial coordinates for the two proteins in their native sequence and for the corresponding homology models solvated in 9162 (3α-structure) and 7797 (4β+α-structure) simple point charge (SPC) water molecules\(^7\) were taken from previous work.\(^9\)

The equations of motion were integrated using the leapfrog algorithm\(^3\) with a timestep of 2 fs. Bond lengths and the bond angle of water molecules were constrained by applying the SHAKE algorithm\(^3\) with a relative geometric tolerance of \(10^{-4} \). The center of mass motion of the computational box was removed every 2 ps. All simulations were performed at constant
pressure and temperature. The temperature was maintained at 298 K by weak coupling to an external bath\(^{359}\) with a relaxation time of 0.1 ps. Solute and solvent were coupled to separate heat baths. The pressure was calculated using a group-based virial and held constant at 1 atm using the weak coupling method with a relaxation time of 0.5 ps\(^{359}\) and an isothermal compressibility of \(4.575 \times 10^{-4} \text{ (kJ mol}^{-1} \text{ nm}^{-3})^{-1}\).

The nonbonded van der Waals and electrostatic interactions were calculated using a twin-range cut-off scheme,\(^{507}\) with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively. The short-range interactions were calculated every timestep using a group-based pairlist updated every fifth timestep. The intermediate-range interactions were reevaluated at each pairlist update and assumed constant in between. A reaction-field correction\(^{356,446}\) was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity \(\varepsilon_{RF}\) of 61.\(^{447}\) To be consistent with previous work on this system\(^{491}\) the reaction field self-term and excluded-atom-term contributions were not included in the calculation of the total electrostatic energy.\(^{39}\)

### 4.3.2 Simulated systems

End-state simulations, \textit{i.e.} standard (non-EDS) MD simulations of the eight sequences considered for each of the two folds, were conducted for structural analysis and to determine the distributions of the solute-solute plus solute-solvent non-bonded energies. These distributions can be compared to those obtained from EDS reference-state simulations \textit{via} reweighting, which allows for an assessment of the quality of sampling in the EDS reference-state simulations. The end-state simulations were carried out for 10 ns after 1 ns equilibration. Atom coordinates and energies were saved for analysis every 5 ps.

EDS reference-state simulations were performed to calculate free-enthalpy changes for each of the twelve alchemical transformations defined in Fig. 4.2 separately for each of the two distinct protein folds. These simulations, involving fixed smoothness parameter and energy offset values, were carried out for a period of 10 to 20 ns after 1 ns equilibration. Atom coordinates and energies were saved for analysis every 2 ps. The EDS parameters were determined independently for each of the 24 calculations prior to production simulations using the parameter update procedure described earlier.\(^{505}\) The simulation time necessary to obtain converged parameters amounted to 10 to 20 ns, depending on the perturbation. The optimized values are listed in Table C.1 of the Supplementary Material. All end- and reference-state
4.3 Computational details

Figure 4.2: Single mutation pathways from L20I30L45 to A20F30Y45.

simulations were based on a common topology containing three hybrid building blocks for residues 20, 30, and 45, respectively, that contain all atoms to simulate Leu/Ala, Ile/Phe, and Leu/Tyr amino acids, respectively. The single topology representations of each hybrid building block are shown in Fig. C.1 of the Supplementary Material. Atom types and partial charges are specified in Tables C.2 – C.4 of the Supplementary Material.

4.3.3 Trajectory analysis

All analyses were carried out using the GROMOS++ suite of programs. The trajectories of all end states were analysed in terms of atom-positional root-mean-square deviation (rmsd) from the starting configuration of each production simulation. The rmsd was calculated based on the Cα atoms. For the trajectories of all end and reference-states simulations secondary structure elements were assigned according to the rules defined by Kabsch and Sander.

The EDS reference-state trajectories were analyzed in terms of the time series of the potential energy difference between state B and state A, $\Delta V_{BA}(t) = V_B(t) - V_A(t)$, the distribution of this potential energy difference in the reference-state ensemble $\rho_R(\Delta V_{BA})$, the potential energy difference distributions predicted for the end-state ensembles, $\tilde{\rho}_A(\Delta V_{BA})$ and $\tilde{\rho}_B(\Delta V_{BA})$, and the solute-solute plus solute-solvent potential energy distributions, $\tilde{\rho}_A(V_A)$.
and $\tilde{\rho}_B(V_B)$, predicted for the end-state ensembles.

Distributions $\rho_A(V_A)$ and $\rho_B(V_B)$ of solute-solute plus solute-solvent non-bonded energies were calculated for all standard MD (non-EDS) end-state trajectories in order to compare them with corresponding reweighted end-state potential energy distributions $\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$ extracted from the EDS reference-state simulations. As the perturbations only involve the solute, solvent-solvent potential energy and entropic contributions exactly cancel out each other and can be omitted.\textsuperscript{450–452}

4.4 Results and discussion

4.4.1 Structural properties

As a simple estimate of the structural stability during the simulations, the atom-positional root-mean-square deviations (rmsds) of the $C_\alpha$ atoms of each sequence structure combination from the first structure of the trajectory were calculated for all end and reference-state simulations and are displayed in Figures C.2 and C.3 of the Supplementary Material. The nature of the structural changes corresponding to an increase in rmsd can be seen in the secondary structure analysis of the different sequence-structure combinations (see Figures C.4 and C.5 of the Supplementary Material). In the end-state simulations, the 3$\alpha$-structure exhibits high structural stability for all sequence combinations except for the native sequence of the 4$\beta$+\$alpha$-structure (AFY), for which the rmsd increases to about 0.4 nm during 11 ns of simulation. This increase in rmsd is accompanied by a loss in the helical character of the first few residues of the first $\alpha$-helix. For the 4$\beta$+\$alpha$-structure the rmsd values are in most cases slightly larger than for the 3$\alpha$-structure except for the native sequence which shows the lowest rmsd. One sequence (AFL) shows a slowly increasing rmsd that is accompanied by a loss in the helical character of the last few residues of the $\alpha$-helix.

In the reference-state simulations the 3$\alpha$-structure exhibits high structural stability for all perturbations studied while for the 4$\beta$+\$alpha$-structure some cases (1, 8 and 12) show some loss in the helical character of the last few residues of the $\alpha$-helix.

However, the general picture that arises from this analysis is that all sequence-structure combinations retained their initial secondary structure during the sampling period of the MD simulations, justifying the use of the thermodynamic cycle shown in Figure 4.1.
4.4 Results and discussion

4.4.2 Free enthalpy changes

Table 4.1 summarizes all point mutations studied and the corresponding changes in free enthalpy, \( \Delta G^\alpha_{BA} \) and \( \Delta G^\beta_{BA} \), respectively. The subscripts A and B denote the two end states of the mutation while the superscripts \( \alpha \) and \( \beta \) denote the 3\( \alpha \) or 4\( \beta + \alpha \)-structure, respectively. Figure 4.2 shows the single-mutation pathways resulting from the different amino acid substitutions. The numbers used to label single-mutation pathways in the text refer to this figure. The results for the perturbations in the 3\( \alpha \) and 4\( \beta + \alpha \)-structure, respectively, are discussed separately in the two following sections.

For each mutation we report the free enthalpy difference directly obtained from the EDS reference-state simulation using eq. 4.2 and a statistically optimal estimate calculated from all free enthalpy differences within a particular fold and including cycle closure (see Supplementary Material for additional details).

If not otherwise stated, the values referred to in the following sections are those obtained directly from the EDS reference state simulations because both data sets do not differ significantly.

Amino acid mutations in the 3\( \alpha \)-structure

The single amino acid mutations 1, 6, 8, and 11 involve the substitution L20 \( \rightarrow \) A20 with all possible combinations for residues 30 and 45. If 45=I, the change in free enthalpy is positive by approximately 7.5 kJ mol\(^{-1} \) independent of whether 30=I or 30=F. However, if 45=Y the change in free enthalpy is slightly negative by approximately \(-2 \) kJ mol\(^{-1} \) (\( \Delta G^\alpha_8 = -3.1 \pm 1.8 \) kJ mol\(^{-1} \) for LIY \( \rightarrow \) AIY and \( \Delta G^\alpha_{11} = -1 \pm 1.5 \) kJ mol\(^{-1} \) for LFY \( \rightarrow \) AFY), again independent of whether 30=I or 30=F. The weak effect amino acid 30 has on residues 20 and 45 is also seen from perturbations 2, 4, 9, and 12 that involve the perturbations I30 \( \rightarrow \) F30 with all possible combinations for residues 20 and 45. All free enthalpy changes are within \(-1.8 \) to \(-4.7 \) kJ mol\(^{-1} \), virtually independent of the type of the other residues. Perturbations 3, 5, 7, and 10 involve the substitution L45 \( \rightarrow \) Y45. The free enthalpy changes for perturbations 5 and 10 are consistent with the results discussed so far in that they do not depend on the nature of residue 30. A comparison between perturbation 3 and 5 shows that the change in free enthalpy does depend on residue 20 consistent with the results discussed above. For perturbation 7, LFL \( \rightarrow \) LFY, the calculated free enthalpy difference of \(-77.4 \pm 2.4 \) kJ mol\(^{-1} \) relies on rather poor sampling, as can be seen from the time series of the energy difference \( \Delta V_{BA}(t) = V_B(t) - V_A(t) \), shown in Figure C.6 of the Supplementary Material. However, \( \Delta G^\alpha_7 \) is very similar to \( \Delta G^\alpha_3 \) (LIL \( \rightarrow \) LIY) as expected due to the weak effect of the nature of residue 30.
Table 4.1: Free enthalpy differences, $\Delta G_{\text{BA}}^{\alpha}$ and $\Delta G_{\text{BA}}^{\beta}$, and their differences, $\Delta \Delta G_{\text{BA}}^{\beta\alpha}$, upon mutating residues 20, 30, and 45.

<table>
<thead>
<tr>
<th>code</th>
<th>mutation (A → B)</th>
<th>$\Delta G_{\text{BA}}^{\alpha/\beta}$</th>
<th>$\Delta G_{\text{BA}}^{\beta}$</th>
<th>$\Delta \Delta G_{\text{BA}}^{\beta\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3-structure</td>
<td>4-structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>direct  optimized</td>
<td>direct  optimized</td>
<td>direct  optimized</td>
</tr>
<tr>
<td>1</td>
<td>LIL → AIL</td>
<td>+7.7 ± 2.0 +6.5 ± 1.8</td>
<td>−11.7 ± 1.6 −9.0 ± 4.6</td>
<td>−19.4 ± 2.6 −15.5 ± 4.9</td>
</tr>
<tr>
<td>2</td>
<td>LIL → LFL</td>
<td>−18.3 ± 1.6 −17.2 ± 2.0</td>
<td>−18.7 ± 1.1 −21.0 ± 4.6</td>
<td>−0.4 ± 1.9 −3.9 ± 5.0</td>
</tr>
<tr>
<td>3</td>
<td>LIL → LIY</td>
<td>−77.0 ± 0.8 −76.9 ± 2.0</td>
<td>−92.2 ± 1.4 −92.6 ± 2.4</td>
<td>−15.2 ± 1.6 −15.6 ± 3.1</td>
</tr>
<tr>
<td>4</td>
<td>AIL → AFL</td>
<td>−16.6 ± 1.2 −17.0 ± 1.1</td>
<td>−23.6 ± 1.9 −21.6 ± 4.6</td>
<td>−7.0 ± 2.2 −4.6 ± 4.7</td>
</tr>
<tr>
<td>5</td>
<td>AIL → AIY</td>
<td>−84.3 ± 2.1 −85.1 ± 1.8</td>
<td>−92.6 ± 1.1 −91.9 ± 2.4</td>
<td>−8.3 ± 2.4 −6.8 ± 3.0</td>
</tr>
<tr>
<td>6</td>
<td>LFL → AFL</td>
<td>+7.2 ± 1.0 +6.7 ± 1.7</td>
<td>−7.5 ± 0.7 −9.5 ± 4.6</td>
<td>−14.7 ± 1.2 −16.2 ± 4.9</td>
</tr>
<tr>
<td>7</td>
<td>LFL → LFY</td>
<td>−77.4 ± 2.4 −75.8 ± 2.0</td>
<td>−90.7 ± 1.4 −91.0 ± 1.5</td>
<td>−13.3 ± 2.8 −15.2 ± 2.5</td>
</tr>
<tr>
<td>8</td>
<td>LIY → AIY</td>
<td>−3.1 ± 1.8 −1.7 ± 1.8</td>
<td>−7.3 ± 1.7 −8.4 ± 2.4</td>
<td>−4.2 ± 2.5 −6.7 ± 3.0</td>
</tr>
<tr>
<td>9</td>
<td>LIY → LFY</td>
<td>−14.7 ± 1.1 −16.0 ± 2.0</td>
<td>−20.2 ± 2.2 −19.5 ± 1.5</td>
<td>−5.5 ± 2.4 −3.4 ± 2.5</td>
</tr>
<tr>
<td>10</td>
<td>AFL → AFY</td>
<td>−82.2 ± 2.0 −83.1 ± 1.7</td>
<td>−92.2 ± 2.0 −92.2 ± 1.1</td>
<td>−10.0 ± 2.8 −9.1 ± 2.0</td>
</tr>
<tr>
<td>11</td>
<td>LFY → AFY</td>
<td>−1.0 ± 1.5 −0.7 ± 1.7</td>
<td>−11.2 ± 1.2 −10.8 ± 1.3</td>
<td>−10.2 ± 1.9 −10.1 ± 2.1</td>
</tr>
<tr>
<td>12</td>
<td>AIY → AFY</td>
<td>−15.6 ± 1.5 −15.0 ± 1.5</td>
<td>−21.4 ± 2.0 −21.8 ± 1.3</td>
<td>−5.8 ± 2.5 −6.8 ± 2.0</td>
</tr>
</tbody>
</table>

* All free enthalpies are reported in kJ mol$^{-1}$. The errors in $\Delta \Delta G_{\text{BA}}^{\beta\alpha}$ were estimated from the square root of the sum of squares of the individual errors. The values referred to as “direct” are directly obtained from the energy trajectories from the EDS reference-state simulations using Eq. 4.2. The values referred to as “optimized” are statistically optimal estimates according to ref. 510.
Table 4.2: Free enthalpy differences for the overall mutation LIL → AFY calculated along different pathways involving single amino acid mutations in each step, see Figure 4.2a

<table>
<thead>
<tr>
<th>path</th>
<th>$\Delta G^{\alpha/\beta}_{\alpha\beta}$</th>
<th>$\Delta G^{\alpha/\beta}_{4\beta+\alpha\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4-10</td>
<td>$-91.1 \pm 3.1$</td>
<td>$-127.5 \pm 3.2$</td>
</tr>
<tr>
<td>1-5-12</td>
<td>$-92.2 \pm 3.3$</td>
<td>$-125.7 \pm 2.8$</td>
</tr>
<tr>
<td>2-6-10</td>
<td>$-93.3 \pm 2.7$</td>
<td>$-118.4 \pm 2.4$</td>
</tr>
<tr>
<td>2-7-11</td>
<td>$-96.7 \pm 3.3$</td>
<td>$-120.6 \pm 2.1$</td>
</tr>
<tr>
<td>3-8-12</td>
<td>$-95.7 \pm 2.5$</td>
<td>$-120.9 \pm 3.0$</td>
</tr>
<tr>
<td>3-9-11</td>
<td>$-92.7 \pm 2.0$</td>
<td>$-123.6 \pm 2.8$</td>
</tr>
<tr>
<td>average</td>
<td>$-93.6 \pm 2.8(3.1)^b$</td>
<td>$-122.8 \pm 2.7(5.1)^b$</td>
</tr>
</tbody>
</table>

a All free enthalpies are reported in kJ mol$^{-1}$. The errors in $\Delta G^{\alpha/\beta}$ were estimated from the square root of the sum of squares of the individual errors.

b Average error obtained using statistically optimal estimates for each path

The twelve studied perturbations allow for six different pathways to estimate the change in free enthalpy for the process LIL → AFY. The free enthalpy changes along these pathways obtained from Table 4.2 are between $-91$ and $-96$ kJ mol$^{-1}$ with a mean value of $-94$ kJ mol$^{-1}$.

We also tried to obtain direct estimates of free enthalpy changes for simultaneous perturbations of more than one amino acid. However, except for LIL → AFL, no pair of smoothness parameter and energy offset values could be identified that resulted in sufficient sampling of both end states in a single EDS reference-state simulation. For the perturbation LIL → AFL the computed value of $\Delta G^\alpha = -12.6 \pm 1.5$ kJ mol$^{-1}$ is in agreement with the free enthalpies along the paths 2-6 and 1-4.

For the single amino acid perturbations the quality of sampling was evaluated by means of the energy-difference distribution, the end-state energy distribution functions and the time series of the energy difference $\Delta V_{BA}(t)$. Figure 4.3 shows the energy difference and end-state energy distributions for all perturbations studied in the $3\alpha$-structure. In all cases the reweighted energy-difference distributions exhibit high-energy tails, showing that two distinct states are sampled in the EDS production simulations. This is also evident from the time series of $\Delta V_{BA}$, shown in Figure C.6 of the Supplementary Material. However, the rather long residence times in each state, in some
The statistically optimal estimates for all free enthalpy differences agree well with the direct results, with a maximum difference of 1.6 kJ mol\(^{-1}\), fur-

![Figure 4.3](image-url)
4.4 Results and discussion

Amino acid mutations in the \(4\beta+\alpha\)-structure

In the \(4\beta+\alpha\)-structure the substitution \(L20 \rightarrow A20\) (1, 6, 8, and 11) was found to depend weakly on residue 30 in addition to a weak dependence on residue 45. If \(30=I\) and \(45=L\) the change in free enthalpy is \(\Delta G_1^\beta = -11.7 \pm 1.6 \text{ kJ mol}^{-1}\) while it is \(\Delta G_6^\beta = -7.5 \pm 0.7 \text{ kJ mol}^{-1}\) if \(30=F\). If \(45=Y\) the free enthalpy change of \(L20 \rightarrow A20\) is \(\Delta G_8^\beta = -7.3 \pm 1.7 \text{ kJ mol}^{-1}\) if \(30=I\) and \(\Delta G_{11}^\beta = -11.2 \pm 1.2 \text{ kJ mol}^{-1}\) if \(30=F\). Compared to mutations in the \(3\alpha\)-structure the changes in free enthalpy for \(L20 \rightarrow A20\) differ less strongly along the possible sequence combinations in the \(4\beta+\alpha\)-structure. For the perturbations 2, 4, 9, and 12 that involve the substitution \(I30 \rightarrow F30\) the variation in free enthalpies is slightly enhanced compared to mutations in the \(3\alpha\)-structure. If \(20=A\), \(\Delta G_4^\beta = -23.6 \pm 1.9 \text{ kJ mol}^{-1}\) for \(AIL \rightarrow AFL\) and \(\Delta G_{12}^\beta = -21.4 \pm 2.0 \text{ kJ mol}^{-1}\) for \(AIY \rightarrow AFY\). If \(20=L\), \(\Delta G_2^\beta = -18.7 \pm 1.1 \text{ kJ mol}^{-1}\) for \(LIL \rightarrow LFL\) and \(\Delta G_9^\beta = -20.2 \pm 2.2 \text{ kJ mol}^{-1}\) for \(LIY \rightarrow LFY\). The latter result relies on rather poor sampling as can be seen from the time series of the energy difference \(\Delta V_{BA}(t) = V_B(t) - V_A(t)\) shown in Figure C.6 of the Supplementary Material. However, the free enthalpy change along the path 3-9-11 is in good agreement with those along the other paths.

As was the case for the \(L20 \rightarrow A20\) mutation, the changes in free enthalpy associated with the perturbation \(L45 \rightarrow Y45\) in the \(4\beta+\alpha\)-structure show fewer differences among the possible combinations compared to mutations in the \(3\alpha\)-structure. Therefore, we conclude that in the \(4\beta+\alpha\) structure, the effects of changes in the amino acids are more independent of one another than in the \(3\alpha\)-structure. The free enthalpy changes along the six different pathways are in the range from \(-118\) to \(-128 \text{ kJ mol}^{-1}\) with a mean value of \(-123 \text{ kJ mol}^{-1}\).

Direct estimation of free enthalpy changes for the simultaneous perturbation of more than one amino acid was not successful because no pair of smoothness parameter and energy offset values could be identified that resulted in sufficient sampling of both end states in a single EDS reference-state simulation.

For the single amino acid perturbations the quality of sampling was evaluated by means of the energy-difference distribution, the end-state energy distribution functions and the time series of the energy difference \(\Delta V_{BA}(t) = V_B(t) - V_A(t)\). Figure 4.4 shows the energy-difference and end-state energy distributions for all perturbations studied in the \(4\beta+\alpha\)-structure. In all cases the reweighted energy-difference distributions exhibit high-energy tails, show-
Figure 4.4: Left: Energy-difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\hat{\rho}_A(\Delta V_A)$ (red) and $\hat{\rho}_B(\Delta V_B)$ (green), for each of the twelve perturbations in protein G as obtained from the EDS reference-state simulations of the $4\beta+\alpha$-structure. The 12 mutations from sequence A to B are defined in Table 4.1. The energy difference distributions of the end states were determined by reweighting. Right: Non-bonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained through reweighting (red and dark green) from the twelve EDS reference-state simulations and from independent MD simulations of the eight end states (orange and light green).

Figure C.6 of the Supplementary Material. Again, the rather long residence times in each state are unfavorable for an efficient use of EDS. The result reported for perturbation 12 relied on the combined analysis of two trajectories sharing the same smoothness parameter value but slightly different energy offsets of $-21.0$ and $-22.0$ kJ mol$^{-1}$ because the sampling turned out to be quite sensitive to the energy offset in this case. The solute-solute plus solute-solvent nonbonded energy distributions obtained from the EDS reference state simulations through reweighting all have the same shape as those obtained from independent MD simulations of the end states. In some cases (4, 5, 6, 7) the distributions are not exactly on top of each other showing that EDS reference-state and end-state simulations have sampled different conformations of the protein. For the
most severe case (perturbation 5, AIL → AIY) we have estimated the effect on the free enthalpy change by performing two simulations, the first starting from the final configuration of the parameter update simulation (the procedure standardly used in this work), and the second from the same coordinates and velocities as used for the end-state simulation of the AIY sequence. The obtained end-state energy distributions, shown in Figure 4.5, differ significantly while the calculated free enthalpy changes differ only by 4.3 kJ mol⁻¹. The statistically optimal estimates for the free enthalpy differences agree well with the direct results, with a maximum difference of 1.1 kJ mol⁻¹, except for those involved in the cycle 1-4-6-2. This cycle shows a comparatively large hysteresis of 9.1 kJ mol⁻¹ giving rise to larger uncertainties for the involved free enthalpy differences. The largest difference 2.7 kJ mol⁻¹ between direct calculation and optimized results occurs for perturbation 1, LIL → AIL.
Relative preference of amino acids between the 3$\alpha$ and 4$\beta$ + $\alpha$-structures

The computed relative free-enthalpy differences for all single amino acid mutations are also listed in Table 1. A negative sign indicates a preference of the state B sequence for the 4$\beta$+α-environment relative to the state A sequence. Perturbations 1, 3, 6, 7, 10 and 11, i.e. those involving the mutations L20 → A20 or L45 → Y45, all show negative $\Delta\Delta G$-values which are at least twice as large as the maximum uncertainty of 5 kJ mol$^{-1}$ and thus show a clear preference for the 4$\beta$+α fold. The six perturbations can be grouped into three pairs. For perturbations 1 and 3 state A represents the native sequence for the 3$\alpha$-structure which leads to the largest magnitude of $\Delta\Delta G$. Perturbations 6 and 7 have the sequence LFL in state A which differs from the native one by the amino acid at position 30. The effect of this substitution is marginal as expected and the $\Delta\Delta G$ is comparable to that of perturbations 1 and 3. Finally, for perturbations 10 and 11, two of the three residues that distinguish the native G$_B$95-sequence from the G$_A$95-sequence are already present and the effect of the last perturbation is smaller than for the former cases illustrating the importance of the context in which a mutation is performed.

For perturbations 2, 4, 5, 8, 9 and 12 no or only weak preferences for the 4$\beta$+α-structure are visible. Four cases of the latter set (2, 4, 9, 12) involve the transformation I30 → F30 which is not affected by a change in the secondary structure of amino acid 30. Perturbation 5 (AIL → AIY) shows a preference for the 4$\beta$+α-environment that is only slightly smaller in magnitude than $-10$ kJ mol$^{-1}$. Perturbation 8 does not show a clear preference for the 4$\beta$+α-environment, although it involves the mutation L20A. The latter result is supported by experimental data that show that the variants are unfolded when 45 = Y and 30 = I, meaning that neither of the two structural environments is preferred in this case. This shows again the importance of the context in which a mutation is performed.

The calculated relative free-enthalpy differences are compatible with the experimental findings that the amino acid mutations L20A and L45Y are responsible for fold switching. Moreover, they also show reasonable quantitative agreement. For the perturbation 7 (LFL → LFY) the calculations predict a relative preference of the LFY sequence for the 4$\beta$+α-fold by $-13.3 \pm 2.8$ kJ mol$^{-1}$. Experimentally it was found that the LF sequence prefers to fold into the 3$\alpha$-structure (> 90% populated at 20°C) while the relative gain in stability of the 4$\beta$+α-structure introduced by the L → Y mutation at position 45 is enough to shift the equilibrium to this state almost completely ($\approx 99\%$). The corresponding relative free enthalpy difference can
4.5 Conclusion

be calculated as $\Delta \Delta G^{\beta \alpha}_{\text{LFY} \rightarrow \text{LFY}} = -RT(\ln 99/1 + \ln 90/10) = -16.6 \text{ kJ mol}^{-1}$, in good agreement with the simulated result. For perturbation 11 (LFY → AFY) He et al.\textsuperscript{511} report switching from a predominantly ($\approx 95\%$) $\alpha$-fold to a $4\beta+\alpha$-fold ($>95\%$ populated). The corresponding relative free enthalpy difference is $\Delta \Delta G^{\beta \alpha}_{\text{LFY} \rightarrow \text{AFY}} = -RT(\ln 95/5 + \ln 95/5) = -14.4 \text{ kJ mol}^{-1}$, while our simulated result is $-10.2 \pm 1.9 \text{ kJ mol}^{-1}$. Note that for this particular case the experimental result corresponds to a sequence with isoleucine at position 25, while the simulated sequence has threonine at this position.

4.5 Conclusion

The present study served two purposes. First, it complements earlier work on the same system focussing on structural and energetic analysis with an extensive thermodynamic analysis in terms of free enthalpies. Second, the feasibility of calculating free enthalpy differences for point mutations using the method of Enveloping Distribution Sampling (EDS) was evaluated. With respect to the first purpose, the twelve calculated free enthalpy changes for each fold showed that the three investigated residues, 20, 30, and 45, do not act independently but that the free enthalpy change for the L20A mutation depends on the nature of residue 45 while the free enthalpy change for the L45Y mutation depends on the nature of residue 20. These interdependencies are stronger in the $3\alpha$-structure than in the $4\beta+\alpha$-structure. The changes in free enthalpy associated with the I30F mutation are essentially independent of other residue changes. The calculation results support the experimental findings that residues 20 and 45 are the main conformational switches for this protein pair.

With respect to the second purpose, we conclude that the investigated point mutations constitute a borderline case for the efficient use of the EDS methodology without any intermediate states between the two end states. The reasons are (i) the significant computational overhead caused by the EDS parameter optimization, a process which requires sufficient transitions between the two end states, and (ii) the strong sensitivity of the sampling quality on the EDS parameters, in particular the energy offset. A strong sensitivity requires a careful parameter optimization which again increases the computational cost. For future studies, involving large perturbations, it might therefore be useful to include a limited number of intermediate states (one or two) between the end states to increase the phase space overlap and make the parameter optimization more efficient.
Efficient combination of environment change and alchemical perturbation within the Enveloping Distribution Sampling (EDS) scheme: Twin-system EDS and application to the determination of octanol-water partition coefficients

The methodology of Enveloping Distribution Sampling (EDS) is extended to probe a single-simulation alternative to the thermodynamic cycle that is standardly used for measuring the effect of a modification of a chemical compound, e.g. from a given species to a chemical derivative for a ligand or solute molecule, on the free-enthalpy change associated with a change in environment, e.g. from the unbound state to the bound state for a protein-ligand system or from one solvent to another one for a solute molecule. This alternative approach relies on the coupled simulation of two systems (computational boxes) 1 and 2, and the method is therefore referred to as twin-system EDS. Systems 1 and 2 account for the two choices of environment. The end states of the alchemical perturbation for the twin-system associate the two alternative forms X and Y of the molecule to systems 1 and 2 or 2 and 1, respectively. In this way, the processes of transforming one molecule into the other are carried out simultaneously in opposite directions in the two environments, leading to a change in free enthalpy that is smaller than for the two individual processes and to an energy-difference distribution that is more symmetric.
an illustration, the method is applied to the calculation of octanol-water partition coefficients for $C_4$ to $C_8$ alkanes, 1-hexanol and 1,2-dimethoxyethane. It is shown in particular that the consideration of the residual hydration of octanol leads to calculated partition coefficients that are in better agreement with reported experimental numbers.

5.1 Introduction

The thermodynamic determinant of physicochemical processes such as protein-ligand binding or solute partitioning between different solvents is the corresponding change in free energy or rather free enthalpy, or Gibbs energy. In the context of atomistic simulations, e.g. molecular dynamics (MD), a wide variety of methods to estimate free-enthalpy differences is nowadays available, out of which the appropriate method has to be selected for the question at hand. A frequently occurring problem in (bio)chemistry is the estimation of the effect of a modification of a chemical compound, e.g. from a given species X to a modified species Y, on the change in free enthalpy associated with a change in environment, e.g. from the unbound state to the bound state for a protein-ligand system or from one solvent to another one for a solute molecule.

The usual way to address this problem computationally is displayed in the upper part of Fig. 5.1. It involves the separate estimation of two free-enthalpy differences for the perturbation from X to Y in environments 1 and 2, respectively, using an appropriate free-enthalpy calculation method. If the method relies on a Hamiltonian coupling scheme defining a $\lambda$-parametrized path connecting X and Y, as is the case for thermodynamic integration (TI) or staged free-energy perturbation (FEP), the free enthalpy along this pathway is a function of $\lambda$, as illustrated by the green and blue curves in the upper right graph of Fig. 5.1. The desired quantity $\Delta \Delta G_{X \rightarrow Y}^{1 \rightarrow 2} = \Delta G_{Y}^{2} - \Delta G_{X}^{1}$ is then obtained by closure of a thermodynamic cycle, in the form of the calculated difference $\Delta G_{X}^{2 \rightarrow Y} - \Delta G_{X}^{1 \rightarrow Y}$. In many cases, e.g. when creating or annihilating partial charges, this difference involves two numbers of comparable and large magnitudes, while the value of $\Delta \Delta G_{X \rightarrow Y}^{1 \rightarrow 2}$ itself is of significantly smaller magnitude. This is analogous to the situation where the weight of the captain is estimated by weighting the boat, with or without the captain on board. As a result, unnecessarily long simulation times may be required to reach sufficiently small uncertainties in the two individual free-
Figure 5.1: Illustration of two computational approaches to calculate a relative free enthalpy of transfer $\Delta \Delta G_{1 \rightarrow 2}^{X \rightarrow Y} = \Delta G_2^{X \rightarrow Y} - \Delta G_1^{X \rightarrow Y}$, where $X \rightarrow Y$ represents an alchemical perturbation from species $X$ to species $Y$ and $1 \rightarrow 2$ represents a transfer from environment 1 to environment 2. In both cases, the difference is calculated by closure of a thermodynamic cycle, in the form $\Delta \Delta G_{1 \rightarrow 2}^{X \rightarrow Y} = \Delta G_2^{X \rightarrow Y} - \Delta G_1^{X \rightarrow Y}$. The usual computational solution, illustrated in the upper part, involves two separate calculations of $\Delta G_1^{X \rightarrow Y}$ and $\Delta G_2^{X \rightarrow Y}$ using an appropriate free-enthalpy calculation method, typically resulting in two numbers of large and comparable magnitudes that are to be subtracted. The twin-system approach proposed in the present work, illustrated in the lower part, relies on the simultaneous simulation of two systems (computational boxes) corresponding to the two environments and couples the $X \rightarrow Y$ mutation in system 2 to the $Y \rightarrow X$ mutation in system 1. The overall perturbation leads directly to $\Delta \Delta G_{1 \rightarrow 2}^{X \rightarrow Y}$, with a large component of $\Delta G_1^{X \rightarrow Y}$ and $\Delta G_2^{X \rightarrow Y}$ cancelling out.
system 1. The overall perturbation leads directly to $\Delta\Delta G_{X \rightarrow Y}^{1 \rightarrow 2}$, with large components of $\Delta G_{X \rightarrow Y}^{1}$ and $\Delta G_{X \rightarrow Y}^{2}$ cancelling out, the cancellation being most effective when the two environments are similar. This approach, referred to here as the twin-system approach, is illustrated in the lower part of Fig. 5.1. The twin-system free enthalpy as function of $\lambda$, represented by the black curve in the lower right graph of Fig. 5.1, is much flatter than the curves representing the individual processes, and therefore in principle easier to calculate accurately. The working principle of the twin-system approach is analogous to that of a cable car: the work required is reduced by coupling a car that is going uphill (X to Y) with a car that is going downhill (Y to X), the net work being formally determined by the weights of the passengers in the two cars, analogous to the environment difference in the present case. In the special situation where the two environments 1 and 2 are the same, one expects $\Delta\Delta G_{X \rightarrow Y}^{1 \rightarrow 2} = 0$. This does not imply, however, that the corresponding free-enthalpy profile as a function of $\lambda$ is entirely flat, which will only be the case if the corresponding single-mutation profile is linear.

The twin-system approach is not bound to a particular method to compute free-enthalpy differences. Here, it is realized in the framework of the Enveloping Distribution Sampling (EDS) methodology. Note that this specific method does not involve any coupling parameter $\lambda$. The approach is illustrated here by calculating partition coefficients for a number of small solute molecules in the octanol-water system.

Partition coefficients of organic solutes in the octanol-water system are widely employed in drug design, environmental risk assessment and medicinal chemistry to characterize the lipophilicity of a given compound. Due to their importance, significant effort has been devoted to the development of simple methods for their prediction, or that of solvation free energies, based on additive group-contribution schemes, on approximate linear correlations to solute molecular properties as in quantitative structure-property models (QSPR), or to experimental properties as in property-property relationship models (PPR), or on continuum-solvation models. Molecular simulations based on classical force fields are also commonly used to calculate parameters related to solvation thermodynamics, such as partition or activity coefficients. Although this approach cannot compete with the simpler methods listed above in terms of computational cost, it becomes the only viable alternative for complex liquid-liquid equilibrium problems where simple methods are reaching their limits, either due to a lack of available parameters or because information on the conformational distribution of the solute becomes important.
The partition coefficient $P_{12}^X$ of a species X between two phases 1 and 2 is defined by the ratio $c_1/c_2$ of the corresponding equilibrium concentrations and is related to the associated free enthalpy of transfer $\Delta G_{1\rightarrow 2}^X$ as\(^{548}\)

$$\Delta G_{1\rightarrow 2}^X = RT \ln P_{12}^X,$$

(5.1)

where $R$ is the ideal-gas constant and $T$ the absolute temperature. Both quantities are in principle functions of pressure, temperature and concentration ($c_1$ or $c_2$), and typically reported for atmospheric pressure, room temperature and in the limit of infinite dilution.\(^{549}\) It is important to stress that $P_{12}^X$ refers to a ratio of molar concentrations (mole X per unit volume) irrespective of the nature of the two phases. Most commonly, these phases will consist of two essentially immiscible solvents. However, even limited mutual solubility of the solvents induces an ambiguity concerning the nature of the quantities involved, i.e. whether they refer to pure 1 and pure 2 solvents or to 2-saturated 1 and 1-saturated 2 phases. For the partitioning between 1-octanol and water, the partition coefficient of a given solute X, usually expressed in the form of log $P_{OW}^X$, where log is a decadic logarithm, is given by

$$\log P_{OW}^X = \frac{\Delta G_W^X - \Delta G_O^X}{RT \ln 10},$$

(5.2)

where $\Delta G_W^X$ denotes the hydration free enthalpy of the solute, i.e. its solvation free enthalpy in water, and $\Delta G_O^X$ its solvation free enthalpy in 1-octanol.

In the context of classical molecular simulations, partition coefficients can be calculated by Monte Carlo (MC) or MD simulations. A single-simulation approach of the former type is the Gibbs ensemble Monte Carlo (GEMC) method,\(^{550}\) usually in its constant pressure formulation\(^{551}\) utilizing two (or more) computational boxes which are in thermodynamic contact but do not have an explicit interface. By performing particle-exchange and switch moves, usually carried out by the configurational-bias Monte Carlo (CBMC) technique,\(^{552-554}\) the chemical potentials of each species in the two phases are equalized.\(^{555,556}\) The composition of the two solvent phases does not need to be specified in advance because the distribution of solvent molecules is also sampled via particle-exchange moves. The free enthalpy of transfer is directly determined from the ratio of solute number densities in the two phases.\(^{557,558}\) Alternatively, if the configurational ensemble is generated by MD, partition coefficients are often calculated from solvation free enthalpies in the two different solvents, determined e.g. by TI at fixed compositions of the two phases.\(^{112,559,560}\) While this methodology is well established, it is computationally expensive and may require a considerable amount of human
time to optimize the TI protocol, including the choice of the $\lambda$-dependence of the Hamiltonian,\textsuperscript{12} the $\lambda$-point distribution\textsuperscript{128} as well as equilibration and sampling times at the different points, so as to reach converged free-enthalpy estimates.\textsuperscript{112,559–561} Given a set of solutes of similar sizes, an alternative route is to calculate relative solvation free enthalpies \textit{via} the alchemical transformation of one solute into another in both solvents by means of \textit{e.g.} FEP.\textsuperscript{562,563} Solvation free enthalpies, and thus partition coefficients, can then be derived provided that the solvation free enthalpies of a single reference compound in the two solvents are known \textit{e.g.} as obtained by the above TI method. The latter approach is computationally more efficient provided that the FEP calculations converge on a reasonable timescale. In practice, this limits the degree by which the solutes of interest may deviate from the reference compound.\textsuperscript{564} If the alchemical perturbation induces a considerable reorganization of the solvent molecules, the FEP method is subject to convergence problems,\textsuperscript{532} necessitating the use of staging approaches.\textsuperscript{565} It is therefore of interest to study alternative free-enthalpy methods\textsuperscript{128,152,555,566,567} that are applicable to larger perturbations but have the potential to be more efficient than additional TI calculations for the estimation of a relative partition coefficient.

The EDS method was shown previously to be a powerful alternative to TI in a variety of applications, ranging from the study of protein-ligand binding\textsuperscript{320} to folding free-enthalpy predictions.\textsuperscript{300} Its strengths are that: \textit{(i)} no pathway has to be defined between two end states; \textit{(ii)} no intermediate states are required; \textit{(iii)} several measures to assess the quality of sampling can be employed.\textsuperscript{505} Limitations of the method are imposed by the magnitude of the perturbation, which affects the computational overhead associated with the precalculation of EDS-parameters that guarantee sufficient sampling of both end states during the EDS reference-state simulation.

In the present work, the twin-system EDS approach is used to calculate octanol-water partition coefficients of short (C$_4$ – C$_8$) alkanes, hexanol and 1,2-dimethoxyethane. Results obtained using different versions of the GROMOS force field\textsuperscript{47,48,60,116} are compared, as well as partition coefficients calculated considering dry and wet octanol phases, and related to other simulation studies from the literature.
5.2 Theory

5.2.1 Calculation of partition coefficients

As illustrated in Fig. 5.2, the octanol-water partition coefficient of a solute X (Eq. 5.2) can be expressed as

\[
\log P_{X}^{OW} = \frac{\Delta G_{W}^{X} - \Delta G_{O}^{X}}{RT \ln 10}
\]
\[
= \frac{\Delta G_{W}^{M' \rightarrow M} + \Delta G_{W}^{M \rightarrow X} - (\Delta G_{O}^{M' \rightarrow M} + \Delta G_{O}^{M \rightarrow X})}{RT \ln 10}
\]
\[
= \log P_{X}^{M} + \frac{\Delta G_{W}^{M \rightarrow X} - \Delta G_{O}^{M \rightarrow X}}{RT \ln 10}
\]
\[
= \log P_{X}^{M} + \frac{\Delta \Delta G_{W}^{M \rightarrow X}}{RT \ln 10},
\]

where M represents a solute used as reference compound and M' the corresponding molecule possessing full intramolecular interactions but exempt of interactions with the solvent. The superscript/subscript notations indicate: the solvation free enthalpy of X in phase 1 ($\Delta G_{1}^{X}$), the mutation free enthalpy of X to Y in phase 1 ($\Delta G_{1}^{X \rightarrow Y}$), the transfer free enthalpy of X from phase 1 to phase 2 ($\Delta G_{1 \rightarrow 2}^{X}$), and the free enthalpy difference $\Delta G_{1 \rightarrow 2}^{X} - \Delta G_{1 \rightarrow 2}^{Y} = \Delta G_{2}^{X \rightarrow Y} - \Delta G_{2}^{X \rightarrow Y}$ ($\Delta \Delta G_{1 \rightarrow 2}^{X \rightarrow Y}$). In particular, $\Delta G_{W}^{X}$ denotes the hydration free enthalpy of the solute X and $\Delta G_{O}^{X}$ its solvation free enthalpy in octanol. In Eq. 5.3 these free enthalpies are decomposed into contributions from activating the solute-solvent interactions of the non-interacting reference solute M' to the fully interacting reference solute M, $\Delta G_{S}^{M' \rightarrow M}$ (S=W,O), and the free enthalpy changes associated with an alchemical perturbation of M to X performed in water and octanol, $\Delta G_{S}^{M \rightarrow X}$ (S=W,O), using the fact that $\Delta G_{W}^{M' \rightarrow W} = 0$. The last term in the right hand side of Eq. 5.3 is a relative partition coefficient that can be calculated directly using the twin-system approach illustrated in Fig. 5.3, which involves the simultaneous simulation of two distinct systems, i.e. two distinct computational boxes. In this coupled perturbation, the initial state A combines the system containing the reference solute solvated in octanol (system 2) with the system containing the target solute solvated in water (system 1), while the final state B combines the system containing the target solute solvated in octanol (system 2) with the system containing the reference solute solvated in water (system 1). The overall process involves the two sub-processes of changing solute M to solute X in the two solvents, carried out simultaneously.
Figure 5.2: Thermodynamic cycles for the calculation of the octanol – water partition coefficient of a reference solute M and a target solute X, where M' denotes the reference solute possessing full intramolecular interactions but exempt of interactions with the solvent. The vertical branches labelled $\Delta G_{W}^{M'\rightarrow M}$ and $\Delta G_{O}^{M'\rightarrow M}$ represent the solvation free enthalpy of the reference solute in water and octanol, respectively. The vertical branches labelled $\Delta G_{W}^{M\rightarrow X}$ and $\Delta G_{O}^{M\rightarrow X}$ represent the changes in free enthalpy associated with the alchemical perturbation of the reference solute M to the target solute X in water and octanol, respectively. The horizontal branches represent the transfer free enthalpies for the process water $\rightarrow$ octanol. Note that $\Delta G_{W}^{M'\rightarrow O} = 0$.

in opposite directions, yielding directly $\Delta G_{O}^{M\rightarrow X} - \Delta G_{W}^{M\rightarrow X} = -\Delta \Delta G_{O\rightarrow W}^{M\rightarrow X}$.

This twin-system approach can conveniently be realized using the EDS methodology as outlined below.

5.2.2 Enveloping distribution sampling (EDS)

The EDS scheme is a special form of the umbrella sampling approach\cite{5} that focuses the sampling on the relevant phase space of a system presenting $N_H$ different states $n$ using a reference Hamiltonian $H_R$ defined as\cite{434,493,498,502}

$$H_R(r,p) = -\frac{1}{\beta s} \ln \sum_{n=1}^{N_H} e^{-\beta s (H_n(r,p) - E^R_n)},$$

(5.4)
Figure 5.3: Illustration of the twin-system approach for the calculation of 
the relative octanol – water partition coefficient associated with the alchemi-
cal perturbation from solute M to solute X. The indices 1 and 2 denote the 
two separate systems (computational boxes) while the two different states of 
the twin-system are labelled A and B.

where \( \mathbf{r} \) and \( \mathbf{p} \) denote the \( 3N \)-dimensional vector of spatial coordinates and 
the \( 3N \)-dimensional vector of conjugate momenta, respectively, \( N \) being the 
number of atoms in the system, and \( \beta = 1/k_B T \), \( k_B \) being the Boltzmann 
constant and \( T \) the absolute temperature. The smoothness parameter \( s (0 < s \leq 1) \) controls the barriers determining the rate of transitions between 
different states \( n \). The energy offsets \( E^R_{n} \) of the different states serve 
to bring their energies to the same level for homogeneous sampling. These 
parameters have to be determined iteratively prior to the EDS production 
simulation.\(^{320,435,505}\) The free-enthalpy difference between a pair of states 
\( (n, m) \) is calculated based on this production simulation using

\[
G_n - G_m = -\frac{1}{\beta} \ln \frac{\langle e^{-\beta(H_n(r,p)-H_R(r,p))}\rangle_R}{\langle e^{-\beta(H_m(r,p)-H_R(r,p))}\rangle_R},
\]

where \( \langle ... \rangle_R \) denotes ensemble averaging over a simulation generated using 
the reference-state Hamiltonian \( H_R \) (Eq. 5.4).

Note that, although the word “reference” is used both for the reference 
compound M for the evaluation of relative partition coefficients and the EDS 
reference state R the two concepts are entirely distinct and should not be 
confused.
5.2.3 Twin-system EDS

In the twin-system EDS approach, considering the situation depicted in Fig. 5.3, both end states A and B of the twin-system are composed of two separate systems (1 and 2) in opposite alchemical states. The Hamiltonian of state \( n (n=A,B) \) is then obtained as the sum of the Hamiltonians of the two systems,

\[
H_n = H_{n1} + H_{n2}. \tag{5.6}
\]

Note that these Hamiltonians are functions of vectors \( \mathbf{r} \) and \( \mathbf{p} \) that account for the coordinates and momenta of the particles in the two systems. As a result, the potential energy term of the EDS reference-state Hamiltonian (Eq. 5.4) reads

\[
V^R_{\text{R}}(\mathbf{r}; s, \Delta E^R_{BA}) = -\frac{1}{\beta s} \ln \left[ e^{-\beta s(V_{A1}(r)+V_{A2}(r)-E^R_A)} + e^{-\beta s(V_{B1}(r)+V_{B2}(r)-E^R_B)} \right],
\]

where \( \Delta E^R_{BA} = E^R_B - E^R_A \) denotes the relative energy offset. Accordingly, the force on an atom \( i \) at coordinate \( \mathbf{r}_i \) is given by

\[
f_i(t) = \left( -\frac{\partial V^R_{\text{R}}(\mathbf{r})}{\partial \mathbf{r}_i} \right) = \left[ 1 + e^{-\beta s(V_{B1}(r)+V_{B2}(r)-(V_{A1}(r)+V_{A2}(r))-\Delta E^R_{BA})} \right]^{-1} \left( -\frac{\partial V_{A1}(r) + V_{A2}(r)}{\partial \mathbf{r}_i} \right) + \left[ 1 + e^{+\beta s(V_{B1}(r)+V_{B2}(r)-(V_{A1}(r)+V_{A2}(r))-\Delta E^R_{BA})} \right]^{-1} \left( -\frac{\partial V_{B1}(r) + V_{B2}(r)}{\partial \mathbf{r}_i} \right). \tag{5.8}
\]

Because the potential energy of system 1 does not depend on the configuration of system 2 and vice versa, one has

\[
\frac{\partial V_{A2}}{\partial \mathbf{r}_i} = \frac{\partial V_{B2}}{\partial \mathbf{r}_i} = 0 \quad \text{for atoms } i \text{ in system 1}
\]

\[
\frac{\partial V_{A1}}{\partial \mathbf{r}_i} = \frac{\partial V_{B1}}{\partial \mathbf{r}_i} = 0 \quad \text{for atoms } i \text{ in system 2}, \tag{5.9}
\]

and Eq. 5.8 simplifies to

\[
f^I_i(t) = \left( -\frac{\partial V^R_{\text{R}}(\mathbf{r})}{\partial \mathbf{r}_i} \right) = \left[ 1 + e^{-\beta s(V_{B1}(r)+V_{B2}(r)-(V_{A1}(r)+V_{A2}(r))-\Delta E^R_{BA})} \right]^{-1} \left( -\frac{\partial V_{A1}(r)}{\partial \mathbf{r}_i} \right) + \left[ 1 + e^{+\beta s(V_{B1}(r)+V_{B2}(r)-(V_{A1}(r)+V_{A2}(r))-\Delta E^R_{BA})} \right]^{-1} \left( -\frac{\partial V_{B1}(r)}{\partial \mathbf{r}_i} \right). \tag{5.10}
\]
for the force on an atom $i$ in system 1 and to

$$\mathbf{f}_i^2(t) = \left(-\frac{\partial V_R(r)}{\partial r_i}\right)$$

$$= \left[1 + e^{-\beta s(V_{B1}(r)+V_{B2}(r)-(V_{A1}(r)+V_{A2}(r)) - \Delta E_{BA}^R)}\right]^{-1}\left(-\frac{\partial V_{A2}(r)}{\partial r_i}\right)$$

$$+ \left[1 + e^{+\beta s(V_{B1}(r)+V_{B2}(r)-(V_{A1}(r)+V_{A2}(r)) - \Delta E_{BA}^R)}\right]^{-1}\left(-\frac{\partial V_{B2}(r)}{\partial r_i}\right)$$

for the force on an atom $i$ in system 2. Therefore, the coupling between the two systems only occurs via the prefactors, and does not affect the evaluation of the derivative of the potential energy functions $V_{n1}$ and $V_{n2}$.

### 5.2.4 Convergence assessment

The time series of the potential energies of states A and B, $V_A(t)$ and $V_B(t)$, sampled in the reference-state ensemble $R$ are obtained as the sum of the contributions from each system. In the reference-state ensemble $R$, the corresponding distribution of the energy difference $\Delta V_{BA}$ between the two end states can be calculated as

$$\rho_R(\Delta V_{BA}) = \langle \delta [\Delta V_{BA} - (V_B(r) - V_A(r))] \rangle_R,$$  \hfill (5.12)

where $\delta$ is the Dirac delta function (approximated in practice by a histogram-binning function). This distribution can be converted to predicted distributions, denoted by a tilde, corresponding to the end-state ensembles A or B via reweighting,

$$\tilde{\rho}_n(\Delta V_{BA}) = \frac{\langle \delta [\Delta V_{BA} - (V_B(r) - V_A(r))] e^{-\beta(V_n(r)-V_R(r))} \rangle_R}{\langle e^{-\beta(V_n(r)-V_R(r))} \rangle_R}$$  \hfill (5.13)

with $n = A$ or B, respectively. The distributions given by Eqs 5.12 and 5.13 are used to gain insight into the quality of the sampling obtained in the EDS simulation.

In a similar manner, distributions of the potential energies $V_A$ and $V_B$ sampled in the reference-state ensemble $R$, $\rho_R(V_n)$, can be converted to predicted distributions corresponding to the end-state ensembles A or B, $\tilde{\rho}_n(V_n)$. These predicted distributions can be compared to actual distributions $\rho_n(V_n)$ based on independent simulations at the two end states, and thus used for further assessment of the quality of the sampling obtained in the EDS simulation.
5.3 Computational details

5.3.1 Simulation parameters

All MD simulations were carried out using a modified version of the GROMOS11 program package. The solute molecules considered were butane, pentane, hexane, heptane, octane, hexanol and 1,2-dimethoxyethane. The GROMOS force-field parameter sets employed were 54A7, 53A6OXY and 53A6OXY+D. For the compounds considered here, there is no difference between the 54A7 version and the former major revision of the GROMOS force field, 53A6. The 53A6OXY version differs from 54A7 for the alcohols and ethers, owing to a reparametrization of oxygen-containing chemical functions. The 53A6OXY+D version is equivalent to 53A6OXY, the only difference involving improved torsional-energy parameters for the vicinal diether function, relevant here for 1,2-dimethoxyethane only. The covalent and non-bonded interaction parameters used in the present work are summarized in Fig. D.1 and Table D.1 of the Supplementary Material, respectively.

The simulations were performed under periodic boundary conditions based on cubic computational boxes containing one solute molecule solvated in 999 simple point charge (SPC) water molecules, in 511 octanol molecules (dry octanol), or in a mixture consisting of 511 octanol and 100 water molecules (wet octanol). The latter situation is meant to account for the residual hydration of octanol in an experimental setup. The reason for using a water mole fraction of 0.16, below the experimental saturation concentration of about 0.27, is that the simulated value of the saturation concentration depends on the quality of the force field and was reported to be lower than the experimental value for the TIP4P/TraPPE-UA (0.21), TIP4P/OPLS (0.09) and a modified 45A3 version of the GROMOS force field (0.16). By using a water mole fraction of 0.16, the simulation of an oversaturated mixture is avoided, while the effect of water on the calculated partition coefficients is still observable. For 1,2-dimethoxyethane, simulations in a mixture of 511 octanol and 190 water molecules, resembling the experimental water mole fraction of 0.27 at saturation, were also carried out for comparison. However, the effect on the calculated partition coefficient compared to the corresponding simulation at water mole fraction of 0.16 was found to be small.

The equations of motion were integrated using the leap-frog scheme with a timestep of 2 fs. Bond lengths and the bond angles of water molecules were constrained by application of the SHAKE procedure with a relative geometric tolerance of $10^{-4}$. The center of mass translation of the compu-
5.3 Computational details

The computational box was removed every 2 ps. All simulations were performed at constant pressure and temperature. The temperature was maintained at 298 K by weak coupling to an external bath,\textsuperscript{359} with a relaxation time of 0.1 ps. Solute and solvent were coupled to separate heat baths. The pressure was calculated using a group-based virial and held constant at 1 atm using the weak coupling method with a relaxation time of 0.5 ps,\textsuperscript{359} and an isothermal compressibility $\kappa_T$ of $7.513 \times 10^{-4}$ (kJ mol\(^{-1}\) nm\(^{-3}\))\(^{-1}\) for water,\textsuperscript{360} $7.64 \times 10^{-4}$ (kJ mol\(^{-1}\) nm\(^{-3}\))\(^{-1}\) for dry octanol,\textsuperscript{573} and $7.62 \times 10^{-4}$ (kJ mol\(^{-1}\) nm\(^{-3}\))\(^{-1}\) for wet octanol (composition-weighted average of the two former values).

The non-bonded van der Waals and electrostatic interactions were calculated using a twin-range cut-off scheme,\textsuperscript{507} with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively. The short-range interactions were calculated every timestep using a group-based pairlist updated every fifth timestep. The intermediate-range interactions were reevaluated at each pairlist update and assumed constant in between. A reaction-field correction\textsuperscript{356,446} was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance, using the experimental relative dielectric permittivities $\varepsilon_{\text{RF}}$ of 78.5 for water,\textsuperscript{360} 10.1 for dry octanol,\textsuperscript{574} and 9.0 for wet octanol, appropriate for a water mole fraction of 0.16.\textsuperscript{575–577} The reaction-field self-term and excluded-atom-term contributions to the energy, forces, and virial were included as described previously.\textsuperscript{355}

5.3.2 Simulated systems

End-state simulations, \textit{i.e.} plain MD simulations of the seven compounds considered, solvated in water, dry octanol or wet octanol, were conducted to determine the distributions of the solute-solute plus solute-solvent non-bonded energies. These distributions can be compared to those obtained from the EDS reference-state simulations \textit{via} reweighting, which allows for an assessment of the quality of the sampling in the EDS reference-state simulations. The end-state simulations were carried out for 10 ns (30 ns for 1,2-dimethoxyethane) after 1 ns equilibration and energies were written to file every 4 ps.

In the present work the twin-system EDS approach is used to evaluate the relative effect caused by the perturbation of a reference compound M to a target compound X on the free-enthalpy change associated with a change of environment from O to W, $\Delta \Delta G_{M,O \rightarrow W}^{M \rightarrow X}$. To obtain partition coefficients \textit{via} Eq. 5.3, the hydration free enthalpy of the reference compound M has to be known as well as its solvation free enthalpy in octanol, \textit{i.e.} the change in free enthalpy for the process $M' \rightarrow M$ (vertical branches in the upper part
of Fig. 5.2). In this work, hexane was chosen as a reference compound as it represents best the average number of atoms in the seven solute molecules considered. The calculation of the hydration free enthalpy of hexane and of its solvation free enthalpy in octanol relied on thermodynamic integration\(^\text{119}\) over the average of the derivative of the Hamiltonian with respect to a scaling parameter \(\lambda\) applied to the solute-solvent interactions, from no interaction at \(\lambda = 0\) to full interaction at \(\lambda = 1\). The \(\lambda\)-coupling was chosen to be quadratic and involves a soft-core scaling\(^\text{124}\) with \(\alpha_{\text{LJ}} = 0.5\) and \(\alpha_{\text{CRF}} = 0.5\,\text{nm}^2\) for the Lennard-Jones and electrostatic interactions, respectively. The solute-solvent interactions were gradually activated using 21 equispaced \(\lambda\)-points with a sampling period of 1.8 ns after 0.2 ns equilibration. The Hamiltonian derivatives were written to file every 0.5 ps. The integration was performed using the trapezoidal rule. Errors in the individual ensemble averages for the Hamiltonian derivative were calculated by block-averaging,\(^\text{371}\) and propagated into an error estimate for the free enthalpy according to the trapezoidal rule.

The hydration free enthalpy of hexanol and its solvation free enthalpy in octanol were calculated using a similar protocol for validation purposes, \textit{i.e.} in order to have a set of independent free enthalpies to be compared with the results from the twin-system EDS approach. Due to a sharper variation the Hamiltonian derivative between \(\lambda\)-values of 0.2 and 0.3 and between \(\lambda\)-values of 0.9 and 1.0 in the case of the octanol solvent, the number of \(\lambda\)-points was increased to 28 and the sampling period at each \(\lambda\)-point was increased to 3.6 ns after 0.4 ns equilibration. Graphs, showing the average of the Hamiltonian derivative as function of \(\lambda\) are provided in Fig. D.2 of the Supplementary Material.

EDS reference-state simulations were performed to calculate each of the six \(\Delta \Delta G_{\text{M} \rightarrow \text{X}}^{\text{O} \rightarrow \text{W}}\) values according to Eq. 5.5. These simulations, involving fixed smoothness parameters and energy offsets, were carried out for a period of 10 ns (butane, pentane and heptane), 15 ns (octane) and 30 ns (hexanol and 1,2-dimethoxyethane), respectively, after 1 ns equilibration. Energies were written to file every 2 ps. The EDS parameters were determined independently for each of the six calculations prior to the production simulations using the parameter update procedure described earlier.\(^\text{505}\) The optimized values are listed in Table D.2 of the Supplementary Material.

The EDS reference-state simulations were conducted within a single-topology approach,\(^\text{448}\) \textit{i.e.} the change from the reference compound \(\text{M}\) to the target compound \(\text{X}\) is performed by changing the types of some of the atoms constituting a common core structure. For alchemical perturbations of hexane to other alkanes or to hexanol, the common core shared by the
two end states corresponds to an aliphatic chain with the covalent interaction parameters specified in Fig. D.1 of the Supplementary Material. The alchemical transformation of hexane to 1,2-dimethoxyethane involves contributions from a change of covalent interaction parameters. Because the current implementation of EDS does not allow for a perturbation of these parameters, a modified aliphatic chain was used for the alchemical perturbation of hexane to 1,2-dimethoxyethane, with covalent interaction parameters appropriate for the 1,2-dimethoxyethane molecule. The hydration free enthalpy of this modified chain and its solvation free enthalpy in octanol were calculated with the same computational protocol as used for hexane (see above). It turns out, however, that the contributions from changing covalent interaction parameters essentially cancel out in the calculation of $\Delta \Delta G_{M \rightarrow W}^{M \rightarrow X}$. Graphs, showing the average of the Hamiltonian derivative as function of $\lambda$ are provided in Fig. D.3 of the Supplementary Material.

5.3.3 Trajectory analysis

The EDS reference-state trajectories were analyzed in terms of the time series of the potential energy difference between state B and state A, $\Delta V_{BA}(t) = V_B(t) - V_A(t)$, the distribution of this potential energy difference in the reference-state ensemble $\rho_R(\Delta V_{BA})$, the predicted potential energy difference distributions in the end-state ensembles, $\tilde{\rho}_A(\Delta V_{BA})$ and $\tilde{\rho}_B(\Delta V_{BA})$, and the predicted end-state solute-solute plus solute-solvent potential energy distribution $\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$.

Distributions $\rho_A(V_A)$ and $\rho_B(V_B)$ of solute-solute plus solute-solvent non-bonded energies were calculated for all end-state trajectories to compare them with corresponding reweighted end-state potential energy distributions $\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$ extracted from the EDS reference-state simulations. As the perturbations only involve the solute, solvent-solvent potential energy and entropic contributions exactly cancel and can be omitted. This analysis was carried out using the GROMOS++ suite of programs.

5.4 Results and discussion

5.4.1 Free enthalpies of hydration and of solvation in octanol

In the context of partition coefficients, the free enthalpies appearing in Eq. 5.2 usually refer to mutually saturated octanol and water phases. While the
properties of the water phase are not altered significantly by the very small equilibrium concentration of octanol, the magnitude of the solvation free-enthalpy difference between dry and wet octanol has been matter of controversy among different experimental studies. It is consensual, however, that the free enthalpy of transfer from dry to wet octanol, ∆G_{dO→wO}, is favorable for polar alcohols and unfavorable for nonpolar molecules such as alkanes. The experimental data selected in the present work for comparison with the calculated free enthalpies is presented in Table 5.1.

The calculated hydration free enthalpies of hexane and hexanol and their solvation free enthalpies in dry and wet octanol are also reported in Table 5.1. For hexane in water, good agreement is achieved between the calculated value of 10.5 ± 0.8 kJ mol\(^{-1}\) and the experimental value of 10.4 kJ mol\(^{-1}\) as well as with the calculated value of 11.5±0.5 kJ mol\(^{-1}\) reported by Schuler et al. using the 45A3 force field, that has introduced the non-bonded interaction parameters for the aliphatic carbons still in use in the current force-field versions. For hexanol, the calculated values of −17.3±0.9 (54A7) and −16.7±0.8 kJ mol\(^{-1}\) (53A6OXY) for the free enthalpy of hydration are also in good agreement with the experimental value of −18.3 kJ mol\(^{-1}\) and with the earlier simulations of Horta et al. using the 53A6 force field (identical to 54A7 for this compound) resulting in −16.6 ± 1.7 kJ mol\(^{-1}\). The differences between 54A7 and 53A6OXY concerning the −CH2−O−H head group of hexanol have only a weak effect on the free enthalpy of hydration.

The free enthalpy of solvation of −14.6 ± 1.0 kJ mol\(^{-1}\) (54A7) for hexane calculated in dry octanol is in agreement with the experimental value of −13.9 kJ mol\(^{-1}\) reported by Berti et al. for dry octanol. For the transfer from dry to wet octanol, the simulation results suggest ∆G_{dO→wO} = +1.4 kJ mol\(^{-1}\) (54A7) in reasonable agreement with the experimental estimate of +2.0 kJ mol\(^{-1}\) reported by Bernazzani et al. Note, however, that another study reports ∆G_{dO→wO} values that are smaller than 1.0 kJ mol\(^{-1}\). Other simulation studies show that the results of ∆G_{dO→wO} calculations are sensitive to the force field used. The TIP4P/OPLS combination yields +0.1 to +0.2 kJ mol\(^{-1}\) along the series of solutes methane to butane, while the SPC/TraPPE-UA combination leads to values of +0.8 to +1.6 kJ mol\(^{-1}\) for the series butane to octane. The difference between 54A7 and the 53A6OXY force field for the free enthalpy of solvation of hexane in wet octanol only concerns the representation of the solvent hydroxyl group and is found to be negligible. For hexanol, the calculated free enthalpy of solvation of −33.1 ± 1.6 kJ mol\(^{-1}\) (54A7) in dry octanol is significantly more negative than the estimated experimental value of −28.6 kJ mol\(^{-1}\) for dry octanol. For the transfer of hexanol from dry to wet octanol, the simulation results
Table 5.1: Free enthalpies of hydration, $\Delta G_W$, and solvation free enthalpies in dry and wet octanol, $\Delta G_{dO}$ and $\Delta G_{wO}$, respectively, from MD simulations and experimental measurements\textsuperscript{a}

<table>
<thead>
<tr>
<th>solute</th>
<th>54A7</th>
<th>53A6OXY(+D)</th>
<th>experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta G_W$</td>
<td>$\Delta G_{dO}$</td>
<td>$\Delta G_{wO}$</td>
</tr>
<tr>
<td>hexane</td>
<td>10.5 ± 0.8</td>
<td>-14.6 ± 1.0</td>
<td>-13.2 ± 1.2</td>
</tr>
<tr>
<td>hexanol</td>
<td>-17.3 ± 0.9</td>
<td>-33.1 ± 1.6</td>
<td>-32.5 ± 1.7</td>
</tr>
<tr>
<td>hexane\textsubscript{D}</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Free enthalpies (in kJ mol$^{-1}$) are reported for the solute molecules hexane, hexanol and hexane with modified torsional-energy parameters (hexane\textsubscript{D}) based on a unit molarity (1 mol dm$^{-3}$) standard state for the gas and solution phases.\textsuperscript{586} The simulated values are obtained using thermodynamic integration and rely on the 54A7\textsuperscript{47,48} or 53A6OXY\textsuperscript{60}/53A6OXY\textsubscript{D}\textsuperscript{116} versions of the GROMOS force field.

\textsuperscript{b} Ref 518.
\textsuperscript{c} Ref 583.
\textsuperscript{d} Calculated from the value reported for dry octanol and the transfer free enthalpy $\Delta G_{dO\rightarrow wO} = +2.03$ kJ mol$^{-1}$ reported in Ref 582.
\textsuperscript{e} Calculated from the value reported for wet octanol and an estimated transfer free enthalpy $\Delta G_{dO\rightarrow wO} = -1.50$ kJ mol$^{-1}$ based on reported\textsuperscript{584} values of -1.76 and -1.36 kJ mol$^{-1}$ for butanol and octanol, respectively.
\textsuperscript{f} Ref 585.
\textsuperscript{g} For this compound, an additional TI calculation considering an octanol-water mixture with a water mole fraction of 0.27 (experimental saturation concentration), instead of 0.16, resulted in the same (within two digits) $\Delta G_{wO}$ value.
suggest $\Delta G_{dO\rightarrow Wo} = +0.6 \pm 2.3 \text{kJ mol}^{-1}$ (54A7). Other simulation studies report slightly negative values of about $-1 \text{kJ mol}^{-1}$ along the series of solutes methanol to butanol.\textsuperscript{556} Experimental estimates for the free enthalpy of transfer are somewhat ambiguous. Dallas and Carr report values from $-0.7$ to $-0.4 \text{kJ mol}^{-1}$ for the series ethanol to pentanol,\textsuperscript{580} while Cabani et al.\textsuperscript{584} measured $-1.76 \text{kJ mol}^{-1}$ and $-1.36 \text{kJ mol}^{-1}$ for the solutes butanol and octanol, respectively. The difference between the 54A7 and 53A6\textsubscript{OXY} force fields for the solvation of hexanol in wet octanol concern the representation of the solute and solvent hydroxyl groups, and amounts to $-1.5 \text{kJ mol}^{-1}$. The value of $-31.0 \pm 1.6 \text{kJ mol}^{-1}$ for the free enthalpy of solvation of hexane in wet octanol calculated using the 53A6\textsubscript{OXY} force field is in very good agreement with the experimental value of $-30.1 \text{kJ mol}^{-1}$ reported by Riebesehl and Tomlinson\textsuperscript{585} for wet octanol. Expectedly, the agreement is less good for the 54A7 force field.

Also reported in Table 5.1 is the free enthalpy of hydration of hexane with modified torsional-energy parameters appropriate for the 1,2-dimethoxyethane molecule, and its solvation free enthalpy in wet octanol. As expected, the results differ from those for normal hexane. However, when taking the difference between $\Delta G^M_W$ and $\Delta G^M_O$ (M = hexane\textsubscript{+D}), the results for normal and modified hexane are essentially identical, showing that the normal hexane molecule could have been used as the reference solute for 1,2-dimethoxyethane as well.

### 5.4.2 Octanol-water partition coefficients

The differences in the free enthalpy of transfer between the considered solutes and hexane for the process water $\rightarrow$ octanol, $-\Delta \Delta G^M_{dO\rightarrow XW}$, are reported in Table 5.2 for the 54A7 force field, evaluated in dry and wet octanol, and for the 53A6\textsubscript{OXY} force field, evaluated in wet octanol only. The reported numbers are calculated directly from the twin-system simulations using the EDS parameters listed in Table D.2 of the Supplementary Material. The difference between dry and wet octanol within the 54A7 force field are small and within the statistical uncertainties of the calculated numbers. The same holds for the comparison between the 54A7 and 53A6\textsubscript{OXY} versions applied to the wet octanol system. Figure 5.4 shows the energy-difference distributions obtained with the 53A6\textsubscript{OXY} force field in wet octanol for all perturbations except that involving 1,2-dimethoxyethane. For the latter system, the distributions are displayed in Fig. 5.5. Corresponding graphs for the 54A7 force field in the dry and wet octanol systems are provided in Figs. D.4 and D.5 of the Supplementary Material.
5.4 Results and discussion

Table 5.2: Differences in transfer free enthalpies between solute X and reference solute M

<table>
<thead>
<tr>
<th>Compound X</th>
<th>54A7</th>
<th>53A6OXY(+D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta G_{\text{dO}\rightarrow\text{W}}^{M\rightarrow X}$</td>
<td>$\Delta G_{\text{wO}\rightarrow\text{W}}^{M\rightarrow X}$</td>
</tr>
<tr>
<td>butane</td>
<td>$7.6 \pm 0.3$</td>
<td>$7.2 \pm 0.3$</td>
</tr>
<tr>
<td>pentane</td>
<td>$3.6 \pm 0.2$</td>
<td>$3.8 \pm 0.2$</td>
</tr>
<tr>
<td>heptane</td>
<td>$-3.4 \pm 0.2$</td>
<td>$-3.4 \pm 0.2$</td>
</tr>
<tr>
<td>octane</td>
<td>$-7.6 \pm 0.3$</td>
<td>$-6.9 \pm 0.2$</td>
</tr>
<tr>
<td>hexanol</td>
<td>$9.3 \pm 0.6$</td>
<td>$9.0 \pm 0.3$</td>
</tr>
<tr>
<td>1,2-dimethoxyethane</td>
<td>-</td>
<td>$26.7 \pm 0.6$</td>
</tr>
</tbody>
</table>

\[\text{a The values (in kJ mol}^{-1}\text{), based on a unit molarity (1 mol dm}^{-3}\text{) standard state for the two phases, are directly obtained from EDS simulations in the twin-system setup, using the 54A7 or 53A6OXY versions of the GROMOS force field. They account for the transfer process water} \rightarrow \text{octanol (} -\Delta \Delta G_{\text{wO}\rightarrow\text{W}}^{M\rightarrow X} = \Delta G_{\text{wO}\rightarrow\text{W}}^{X} - \Delta G_{\text{wO}\rightarrow\text{W}}^{M}\text{). For the systems studied here, the reference solute M is hexane(+D), where the +D applies only to simulations involving 1,2-dimethoxyethane.}\]

\[\text{b For this compound, an additional calculation considering an octanol-water mixture with a water mole fraction of 0.27 (experimental saturation concentration), instead of 0.16, resulted in } -\Delta \Delta G_{\text{wO}\rightarrow\text{W}}^{M\rightarrow X} = 26.0 \pm 0.4 \text{ kJ mol}^{-1}\text{.}\]

In all cases, the energy-difference distributions exhibit a high-energy tail, essential for an accurate estimation of the free-enthalpy difference according to Eq. 5.5. Also shown in Figs. 5.4 and 5.5 are the distributions of the solute-solute plus solute-solvent non-bonded energies obtained from the EDS reference-state trajectories via reweighting and from plain MD simulations of the various end states. In all cases, the EDS and plain MD end-state energy distributions are nearly identical, providing additional evidence that the parts of the phase space sampled in the end-state simulations were properly sampled in the EDS reference-state simulations. Figure 5.6 shows the time series of $\Delta V_{\text{BA}}$, the difference of the end-state energies, in the six reference-state simulations. In all cases frequent transitions between the values corresponding to the two end states are observed. However, for the transformations of hexane to hexanol and hexane to 1,2-dimethoxyethane, the residence times in each state are significantly longer than for the transformations of hexane to the other alkanes. The relative free enthalpies of transfer, $-\Delta \Delta G_{\text{O}\rightarrow\text{W}}^{M\rightarrow X}$, listed in Table 5.2 need to be subtracted from the difference of $\Delta G_{\text{wO}\rightarrow\text{W}}^{M}$ and $\Delta G_{\text{wO}\rightarrow\text{W}}^{0}$ obtained for the reference compound M (hex-
Figure 5.4: Left: Energy difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\tilde{\rho}_A(\Delta V_{BA})$ (red), $\tilde{\rho}_B(\Delta V_{BA})$ (green), as obtained from twin-system EDS simulations. State B represents the combined Hamiltonians of solute X solvated in wet octanol (X = butane, pentane, heptane, octane, hexanol; from top to bottom) and solute M solvated in water (M = hexanol). State A represents the combined Hamiltonians of solute X solvated in water and solute M solvated in wet octanol. The simulations were performed with the 53A6OXY force field. The energy difference distributions of the end states were determined by reweighting. Right: Non-bonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained via reweighting from EDS reference-state simulations ($\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$, red and green) and from independent MD simulations of the end states ($\rho_A(V_A)$ and $\rho_B(V_B)$, blue and orange).

A positive value of $\log P_{X{}^{\text{OW}}}$ reflects a higher affinity for the octanol phase compared to the water phase. The experimental data reported in Table 5.3 refer to mutually saturated phases. The partition coefficients for hexane and hexanol from Table 5.3 (4.00 ± 0.25 and 2.03 ± 0.03, respectively) are compatible with the experimental free enthalpies of hydration of hexane and hexanol and their solvation free enthalpies in wet octanol selected for Table 5.1, which result in partition coefficients of 3.90 and 2.08 according to Eq. 5.2. The results obtained with the 54A7 force field show that taking into account the residual hydration of octanol, systematically improves
5.4 Results and discussion

Figure 5.5: Left: Energy difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\rho_A(\Delta V_{BA})$ (red), $\rho_B(\Delta V_{BA})$ (green), as obtained from twin-system EDS simulations. State B represents the combined Hamiltonians of solute 1,2-dimethoxyethane solvated in wet octanol and solute hexane solvated in water. State A represents the combined Hamiltonians of solute 1,2-dimethoxyethane solvated in water and solute hexane solvated in wet octanol. The simulations were performed with the 53A6$_{OXY}$ force field. The energy difference distributions of the end states were determined by reweighting. Right: Nonbonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained via reweighting from EDS reference-state simulations ($\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$, red and green) and from independent MD simulations of the end states ($\rho_A(V_A)$ and $\rho_B(V_B)$, blue and orange).

The agreement with experimental data. Note that this improvement results mainly from the less negative free enthalpy of solvation of the reference compound hexane in wet octanol compared to dry octanol (Table 5.1), leading to a shift of all partition coefficients to lower values (decrease of log $P_{OW}$ by about 0.2 to 0.3). This is illustrated graphically in Fig. 5.7, showing that quantitative agreement with experimental data is only achieved for the simulation results obtained in wet octanol.

The performances of the 54A7 and 53A6$_{OXY}$ force fields are comparable for predicting the partition coefficients of the alkanes. For hexanol both force fields overpredict the experimental partition coefficient of $2.03 \pm 0.03$ by 0.5 log P units. To exclude methodological reasons, the partition coefficient for hexanol was also calculated from the free enthalpies of hydration reported in Table 5.1. The results of $2.66 \pm 0.33$ for the 54A7 force field and $2.50 \pm 0.33$ for the 53A6$_{OXY}$ force field also overpredict the experimental partition coefficient. This suggests that the observed discrepancy relative to experiment is due to force-field inaccuracies rather than to sampling problems. For 1,2-dimethoxyethane and wet octanol, the calculated value of $-0.40 \pm 0.25$ is in good agreement with the experimental value of $-0.21 \pm 0.02$, and also improved compared to the value of $-0.68$ previously calculated by Fuchs.
Figure 5.6: Time series of the energy difference $\Delta V_{BA} = V_B - V_A$ in twin-system EDS reference-state simulations coupling two simulation boxes containing one solute molecule solvated in water and wet octanol, respectively, using the 53A6OXY force field. State B represents the combined Hamiltonians of solute X solvated in wet octanol ($X = \text{butane, pentane, heptane, octane, hexanol; from top to bottom}$) and solute M solvated in water ($M = \text{hexane}$). State A represents the combined Hamiltonians of solute X solvated in water and solute M solvated in wet octanol. 1,2-dimethoxyethane was simulated using the 53A6OXY+D extension of the 53A6OXY force field.

et al. in dry octanol using the 53A6OXY+D force field. For a water mole fraction of 0.27 in the octanol phase, the calculated partition coefficient is $-0.26 \pm 0.22$, very close to the experimental number.

For the alkanes, the present results show a slight overprediction of the increase in $\log P_{OW}$ with increasing number of carbon atoms, for both the simulations in dry and in wet octanol. A similar trend was also reported for other force fields such as OPLS-AA and TraPPE, while simulated $\log P_{OW}$ values in dry octanol using a combination of different force fields, GROMOS 53A6 for water and OPLS-AA/TraPPE for the organic phase, were found to be in good agreement with experimental data, that were, however, determined for wet octanol. In the case of alkanes, reported estimates for partition coefficients in dry octanol differ from those measured in wet octanol only by
5.4 Results and discussion

Table 5.3: Octanol-water partition coefficients of C₄ to C₈ n-alkanes, hexanol and 1,2-dimethoxyethane

<table>
<thead>
<tr>
<th>Compound X</th>
<th>log ( P_{OW}^{X} )</th>
<th>Exp. (wet octanol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>54A7 dry octanol</td>
<td>53A6OXY(+D) b</td>
</tr>
<tr>
<td>butane</td>
<td>3.06 ± 0.22</td>
<td>2.89 ± 0.25</td>
</tr>
<tr>
<td>pentane</td>
<td>3.76 ± 0.22</td>
<td>3.50 ± 0.24</td>
</tr>
<tr>
<td>hexane^e,f</td>
<td>4.39 ± 0.22</td>
<td>4.15 ± 0.24</td>
</tr>
<tr>
<td>heptane</td>
<td>5.00 ± 0.22</td>
<td>4.75 ± 0.24</td>
</tr>
<tr>
<td>octane</td>
<td>5.72 ± 0.22</td>
<td>5.36 ± 0.24</td>
</tr>
<tr>
<td>hexanol</td>
<td>2.76 ± 0.25</td>
<td>2.58 ± 0.22</td>
</tr>
<tr>
<td>hexanol^f</td>
<td>2.76 ± 0.33</td>
<td>2.66 ± 0.33</td>
</tr>
<tr>
<td>1,2-dimethoxyethane</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

^a The values are obtained from EDS simulations in the twin-system setup using the 54A7^47,48 force field for dry and wet octanol, the 53A6OXY^60/53A6OXY+D^116 force field for wet octanol, and from experiment (Exp.).
^b the +D version applies only for 1,2-dimethoxyethane
^c Ref. 515, experimental uncertainty assumed to be equal to that for pentane.
^d Ref. 588
^e Reference compound.
^f Calculated directly from the TI-results reported in Table 5.1.
^g For this compound, an additional calculation considering an octanol-water mixture with a water mole fraction of 0.27 (experimental saturation concentration), instead of 0.16, resulted in a partition coefficient of −0.26 ± 0.22.
^h Ref. 587

about 0.2-0.3 log P units,^584 which is often within the error affecting the calculated values. Nevertheless, the present results suggest that the appropriate comparison between simulation results and experimental data should be based on similar conditions in terms of the water content of the octanol phase.

5.4.3 Evaluation of the twin-system setup

In general, the computational effort of EDS compared to TI will depend on the degree of difference between the end-state Hamiltonians, on the choice of a \( \lambda \)-dependence for the Hamiltonian in TI, and on the required precision for the resulting free-enthalpy differences. For example, in the present case, the computational cost of calculating the partition coefficient for hexanol using
Figure 5.7: Comparison of simulated octanol-water partition coefficients using different force field versions or/and dry and wet octanol phases against experimental data.

TI amounted to 42 ns simulation time for a box of water molecules and 112 ns for a box of octanol molecules, similar to simulation times used by others for similar systems. The combination of TI and twin-system EDS proposed in this work required two times 42 ns simulation time for determining the hydration free enthalpy of hexane and its solvation free enthalpy in octanol, and about 40 ns simulation time for the perturbation of hexane to hexanol using the twin-system EDS approach, including 10 ns for the EDS-parameter optimization. Considering that a twin-system simulation is about twice as expensive as a single-system one, the two approaches require similar computational effort for the calculation of the partition coefficient of a single arbitrary compound. However, for a series of closely related compounds for which a common model compound is adequate, the twin-system EDS approach becomes more efficient because no additional TI calculations are required. Future applications of this new methodology will show how large the differences between reference and target compounds can be for an efficient use of the twin-system setup.

A general advantage of EDS over other methods such as TI or FEP is the absence of any directionality of the process. For TI or FEP, the alchemical transformation can be carried out in a forward and a reverse direction.
that usually have different convergence properties leading to hysteresis.\textsuperscript{589,590} Such an asymmetry is not present in EDS where the system can alternate smoothly between the two end states during the reference-state simulation. However, the extents of the relevant parts of the phase space that need to be sampled will usually differ for the two end states. This is reflected in an asymmetric energy-difference distribution as is illustrated in Fig. 5.8(a) for the transformation of hexanol to hexane in water and in Fig. 5.8(b) for the transformation of hexane to hexanol in wet octanol. In both cases, the sampling of the hexane end state allows the system to adopt a variety of high-energy configurations leading to high-energy tails in the distributions, while that of the hexanol end state shows a narrower range of possible energy differences. The magnitude of the finite-sampling error\textsuperscript{591} affecting the estimate of $\Delta G$ depends on whether the two ensemble averages in Eq. 5.5 have been sampled with the correct ratio. Frequent transitions between the two end states increase the likelihood that when sampling one end state, configurations important for the other one are encountered. However, in case of asymmetric energy-difference distributions, the system with the more narrow distribution will be less likely to encounter configurations important to the other system, thus increasing the bias asymmetrically.\textsuperscript{444,592} With the twin-system setup (Fig. 5.1, bottom part), the two energy-difference distributions are more symmetric by construction compared to those of the individual perturbations, as is illustrated in Fig. 5.8(c). As a result, the free-enthalpy estimates are expected to be less susceptible to statistical inaccuracies introduced by finite sampling.

\section*{5.5 Conclusion}

The Enveloping Distribution Sampling (EDS) method has been extended to probe a single-simulation alternative to the thermodynamic cycle that is standardly used for measuring the effect of a modification of a chemical compound, \textit{e.g.} from a given species to a chemical derivative for a ligand or solute molecule, on the free-enthalpy change associated with a change in environment, \textit{e.g.} from the unbound state to the bound state for a protein-ligand system or from one solvent to another one for a solute molecule. This alternative approach relies on the coupled simulation of two systems (computational boxes) 1 and 2, and the method is therefore referred to as twin-system EDS. Systems 1 and 2 account for the two choices of environment. The end states of the alchemical perturbation for the twin-system associate the two alternative forms X and Y of the molecule to systems 1 and 2 or 2 and 1, respectively. In this way the processes of transforming one molecule
Figure 5.8: (a) Energy difference distributions for the reference state, \( \rho_R(\Delta V_{BA}) \) (black), and the two end states, \( \tilde{\rho}_A(\Delta V_{BA}) \) (blue), \( \tilde{\rho}_B(\Delta V_{BA}) \) (green) associated with the alchemical transformation of hexanol (state A) to hexane (state B) in water as obtained from a single-system EDS simulation, using a smoothness parameter \( s = 0.256 \) and a relative energy offset \( E_{BA}^R = +30.1 \text{kJ mol}^{-1} \). (b) Energy difference distributions for the reference state, \( \rho_R(\Delta V_{BA}) \) (black), and the two end states, \( \tilde{\rho}_A(\Delta V_{BA}) \) (red), \( \tilde{\rho}_B(\Delta V_{BA}) \) (blue) associated with the alchemical transformation of hexane (state A) to hexanol (state B) in wet octanol as obtained from a single-system EDS simulation, using a smoothness parameter \( s = 0.188 \) and a relative energy offset \( E_{BA}^R = -18.2 \text{kJ mol}^{-1} \). (c) Energy difference distributions for the reference state, \( \rho_R(\Delta V_{BA}) \) (black), and the two end states, \( \tilde{\rho}_A(\Delta V_{BA}) \) (magenta), \( \tilde{\rho}_B(\Delta V_{BA}) \) (cyan), as obtained from a twin-system EDS simulation, using a smoothness parameter \( s = 0.085 \) and a relative energy offset \( E_{BA}^R = +11.0 \text{kJ mol}^{-1} \). State A represents the combined Hamiltonians of hexanol solvated in water and hexane solvated in wet octanol. State B represents the combined Hamiltonians of hexanol solvated in wet octanol and hexane solvated in water. All simulations were performed with the 54A7 force field. All energy difference distributions of the end states were determined by reweighting.

into the other are carried out simultaneously in opposite directions in the two environments, leading to a change in free enthalpy that is smaller than for the two individual processes and to an energy-difference distribution that is more symmetric.

The method was introduced and applied to the calculation of octanol-water partition coefficients for short alkanes, hexanol and 1,2-dimethoxyethane. It is observed in particular that for a comparison between experimental and calculated partition coefficients, the water content of the octanol phase has to be taken into account in the simulations.

Concerning the methodology it is shown that the proposed twin-system strategy offers advantages over thermodynamic integration in terms of com-
putational efficiency for applications such as screening of a set of compounds that can be related to one or a few common model compounds, thereby allowing for larger differences in the end states than could be sampled by free-energy perturbation.

For more biologically relevant applications such as host-guest binding, the advantage of this method is to be less susceptible to statistical inaccuracies of free-enthalpy estimates that are obtained as differences of two large numbers of comparable magnitudes calculated from separate simulations along the branches of the standardly used thermodynamic cycle. Future work concerning the twin-system EDS approach will focus on these types of systems.
Outlook

Free-energy calculations have emerged as an important tool to tackle physicochemical problems that experiment alone has left unresolved. However, simulations are limited by the required computational effort as well as the model and methods they are based on. Progress in free-energy calculations therefore requires not only advances on the hardware front but also better force fields and efficient algorithms.

In Chapter 2, it is demonstrated that a physically rigorous treatment of ion solvation is more a methodological problem than one that could be solved using more computing power. Combined with previous\textsuperscript{61,172–174} and subsequent work\textsuperscript{49,197} the major difficulties are documented in great detail in the literature. Recently some of the correction terms have been included into the calculation of interatomic forces.\textsuperscript{284}

In Chapter 3, the method of enveloping distribution sampling (EDS) to calculate free energy differences between two states was used to investigate sequence-specific binding affinities of netropsin to the DNA minor groove. A new approach to determine the two parameters required to optimize the sampling in EDS was introduced and tested. Currently this approach is limited to two states, whereas the EDS methodology allows in principle the calculation of multiple free energies from a single simulation. Considering only two states the EDS method requires a similar computational effort as thermodynamic integration which is the most widely used approach today. However, even for two-state applications EDS has the advantage of possessing various convenient means to control the sampling quality. Generalization of the parameter determination approach to multiple states is needed to increase the computational efficiency of EDS.

The application of EDS to study the effect of point mutations in two proteins on their stability, reported in Chapter 4, turned out to touch the limit of EDS regarding the magnitude of the perturbations. The calculated free-energy differences for the protein pair GA95/GB95 that shows
95% sequence identity yet fold into topologically different structures show that point mutations at different positions in the amino acid sequence do not contribute equally to the free enthalpy difference between the native and non-native sequence-structure combinations. Interestingly, the relative free energy differences for side-chain mutations calculated in the present work are in quantitative agreement with recent work of Roy et al. using a completely different method, namely a ‘confine-convert-release’ approach with implicit solvent and the Amber ff99SB force field. Neighborhood and environment effects need to be studied further in order to elucidate why only certain amino acids have a strong influence on the fold. Also additivity principles need to be studied.

In addition to the generalization of the parameter update scheme of EDS to multiple states other methodological developments will be useful. First, alternative forms of the Hamiltonian might be more efficient for certain problems than the one in use today. Second, the current implementation is restricted to handle perturbations in nonbonded interactions only. Attempts for a generalization to bonded interactions have been started recently. Third, the method is only implemented in one particular molecular dynamics software package (GROMOS) which hampers the further dissemination of the method.

In Chapter 5, a standard problem of free-energy calculations has been addressed, namely that in many cases two large, slightly different numbers from two independent simulations are to be subtracted to obtain a relative free energy difference. A single-simulation alternative has been introduced and implemented within the EDS framework. The method has been tested for the calculation of octanol-water partition coefficients. Applications to more complex biomolecules are currently being performed.
Supplementary material for chapter 2

Details regarding the functional dependence of density and permittivity on pressure and temperature are provided as well as additional calculation details for methods I and II.
Parameters for the correction scheme

As described in Section 2.2.3 of the main text, two alternative methods (Methods I and II) can be employed to evaluate derivative single-ion hydration and aqueous partial molar properties based on atomistic simulations. In both cases, the outcome of the calculations are raw point-to-point (hydration parameters) or fixed-point (aqueous partial molar parameters) quantities. These must be appropriately corrected for finite-system and approximate-electrostatics errors according to two variants of the correction scheme (Variants A and B), as described in Section 2.2.4 of the main text. The computational details of these calculations are described in Section 2.3 and, in particular, the parameters for the correction scheme are detailed in Section 2.3.4. For most parameters, function values and derivatives at $P^0$ and $T^-$ are sufficient to characterize the correction scheme for both Variants A and B. These quantities are reported in Table 2.3. For Variant B, however, experimental estimates for the density $\rho(P,T)$ and the permittivity $\epsilon(P,T)$ of water at pressure $P$ and temperature $T$ must be available. These also define corresponding estimates for the employed water model via Eqs. 2.47 and 2.48. For $\rho(P,T)$, the present work relied on the IAPWS-95 parametrization and the calculation was performed using available IAPWS software. For $\epsilon(P,T)$, the present work relied on the Bradley and Pitzer parametrization. For completeness, the corresponding equations are provided below.

Bradley and Pitzer suggested the following numerical expression to calculate the permittivity $\epsilon$ of water as function of pressure $P$ and temperature $T$

$$\epsilon(P,T) = A + C \cdot \ln \left( \frac{B + P}{B + P_{\text{ref}}} \right), \quad (A.1)$$

where $A = \epsilon(P_{\text{ref}},T)$, $B$ and $C$ are temperature-dependent parameters, and the choice of $P_{\text{ref}} = 1000\text{bar}$ was made for the reference pressure. The temperature dependence of the three above parameters is given by

$$A = U_1 \exp \left( U_2 \cdot T + U_3 \cdot T^2 \right) \quad (A.2)$$
$$C = U_4 + \frac{U_5}{U_6 + T} \quad (A.3)$$
$$B = U_7 + \frac{U_8}{T} + U_9 \cdot T, \quad (A.4)$$

with the constants $U_1 - U_9$ provided in Table A.1.
**Table A.1**: Constants used to calculate the permittivity of water according to Bradley and Pitzer.\textsuperscript{374}

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_1$ [-]</td>
<td>$3.4279 \cdot 10^2$</td>
<td>$U_6$ [K]</td>
<td>$-1.8289 \cdot 10^2$</td>
</tr>
<tr>
<td>$U_2$ [K(^{-1})]</td>
<td>$-5.0866 \cdot 10^{-3}$</td>
<td>$U_7$ [bar]</td>
<td>$-8.0325 \cdot 10^3$</td>
</tr>
<tr>
<td>$U_3$ [K(^{-2})]</td>
<td>$9.4690 \cdot 10^{-7}$</td>
<td>$U_8$ [K bar]</td>
<td>$4.2142 \cdot 10^6$</td>
</tr>
<tr>
<td>$U_4$ [-]</td>
<td>$-2.0525$</td>
<td>$U_9$ [K(^{-1}) bar]</td>
<td>$2.1417$</td>
</tr>
<tr>
<td>$U_5$ [K]</td>
<td>$3.1159 \cdot 10^3$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Additional details for method I

Table A.2: Raw simulation data used to calculate hydration and aqueous partial molar parameters for Na⁺ via Method I, along with correction terms and corrected results for Variant B. (a)

<table>
<thead>
<tr>
<th>Parameter Y (→ Y')</th>
<th>N_W</th>
<th>T [K]</th>
<th>P [bar]</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH (→ ΔH) [kJ mol⁻¹]</td>
<td>512 298.15</td>
<td>1.0</td>
<td>-96.0</td>
<td>ΔV₆₇</td>
<td>ΔV₃₄</td>
</tr>
<tr>
<td>v (→ c) [10⁻⁸ m³ mol⁻¹]</td>
<td>512 298.15</td>
<td>1.0</td>
<td>-82.0</td>
<td>-75.9</td>
<td>0.0</td>
</tr>
<tr>
<td>ΔH (→ cP) [kJ mol⁻¹]</td>
<td>512 298.15</td>
<td>1.0</td>
<td>-69.4</td>
<td>-75.9</td>
<td>0.9</td>
</tr>
<tr>
<td>v (→ kP) [10⁻⁸ m³ mol⁻¹]</td>
<td>512 298.15</td>
<td>1.0</td>
<td>-82.0</td>
<td>-75.9</td>
<td>0.9</td>
</tr>
<tr>
<td>ΔV₆₇</td>
<td>ΔV₃₄</td>
<td>ΔV₂₃</td>
<td>ΔV₁₀</td>
<td>ΔV₉₀</td>
<td>ΔV₉₀⁺ (°)</td>
</tr>
<tr>
<td>v (→ kP) [10⁻⁸ m³ mol⁻¹]</td>
<td>512 298.15</td>
<td>1.0</td>
<td>-82.0</td>
<td>-75.9</td>
<td>0.9</td>
</tr>
<tr>
<td>ΔV₆₇</td>
<td>ΔV₃₄</td>
<td>ΔV₂₃</td>
<td>ΔV₁₀</td>
<td>ΔV₉₀</td>
<td>ΔV₉₀⁺ (°)</td>
</tr>
<tr>
<td>v (→ aP) [10⁻⁸ m³ mol⁻¹]</td>
<td>512 298.15</td>
<td>1.0</td>
<td>-82.0</td>
<td>-75.9</td>
<td>0.9</td>
</tr>
<tr>
<td>ΔV₆₇</td>
<td>ΔV₃₄</td>
<td>ΔV₂₃</td>
<td>ΔV₁₀</td>
<td>ΔV₉₀</td>
<td>ΔV₉₀⁺ (°)</td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter Y (or y') considered (ΔH or v in Method I), the parameter Y' (or y'; between parentheses) to be evaluated using the simulation set, the number of water molecules N_W in the system, the simulation temperature T and pressure P, the correction terms ΔY₆₇, ΔY₃₄, ΔY₂₃, ΔY₁₀, ΔY₉₀⁺, ΔY₉₀⁻ for Variant B, the corresponding sum ΔY_correct, the raw simulation result ΔY_raw and the corrected result ΔY_correct. The quantities Y_correct to be evaluated correspond to the hydration enthalpy ΔH and the aqueous ion partial molar volume v, heat capacity cP, volume-compressibility kP and volume-expansivity aP. For the sets ΔH (→ ΔH) and v (→ v), the reported results directly correspond to the ΔY_correct for both variants (no distinction between A and B for these quantities). For the sets ΔH (→ cP), v (→ kP) and v (→ aP), the finite-difference or least-squares fitting procedure applied for extracting ΔY_correct from the reported ΔY_correct values are illustrated in Figure A.1 for the systems involving 1024 water molecules. The resulting estimates for ΔY_correct in Variant B can be found in Table 2.4 of the main text, along with corresponding estimates and correction terms for Variant A. (b) For cP, kP and aP, these entries refer to the quantities usually written y₆₇, y₃₄ and y₂₃ (equivalent to ΔY_correct and ΔY_correct in the present context owing to Eqs. 2.8 and 2.9 in the main text).
Table A.3: Raw simulation data used to calculate hydration and aqueous partial molar parameters for $\text{Cl}^-$ via Method I, along with correction terms and corrected results for Variant B. (a)

<table>
<thead>
<tr>
<th>Parameter Y ($\rightarrow$ Y')</th>
<th>$N_W$</th>
<th>$T$ [K]</th>
<th>$P$ [bar]</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\Delta Y_A$</td>
<td>$\Delta Y_B$</td>
</tr>
<tr>
<td>$\Delta H$ ($\rightarrow \Delta H$)</td>
<td>512</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-79.5</td>
</tr>
<tr>
<td>[kJ mol$^{-1}$]</td>
<td>724</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-71.1</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-63.5</td>
</tr>
<tr>
<td>$v$ ($\rightarrow c$)</td>
<td>512</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-11.7</td>
</tr>
<tr>
<td>[10$^{-4}$ m$^3$ mol$^{-1}$]</td>
<td>724</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-10.5</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-9.4</td>
</tr>
<tr>
<td>$\Delta H$ ($\rightarrow cP$)</td>
<td>512</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-77.0</td>
</tr>
<tr>
<td>[kJ mol$^{-1}$]</td>
<td>724</td>
<td>273.15</td>
<td>1</td>
<td>0.0</td>
<td>-68.9</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-72.6</td>
</tr>
<tr>
<td>$v$ ($\rightarrow kP$)</td>
<td>512</td>
<td>298.15</td>
<td>5000</td>
<td>0.0</td>
<td>-5.1</td>
</tr>
<tr>
<td>[10$^{-4}$ m$^3$ mol$^{-1}$]</td>
<td>724</td>
<td>298.15</td>
<td>10 000</td>
<td>0.0</td>
<td>-3.3</td>
</tr>
<tr>
<td></td>
<td>724</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-10.5</td>
</tr>
<tr>
<td></td>
<td>724</td>
<td>298.15</td>
<td>5000</td>
<td>0.0</td>
<td>-4.6</td>
</tr>
<tr>
<td></td>
<td>724</td>
<td>298.15</td>
<td>10 000</td>
<td>0.0</td>
<td>-3.0</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>298.15</td>
<td>1</td>
<td>0.0</td>
<td>-9.4</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>298.15</td>
<td>5000</td>
<td>0.0</td>
<td>-4.1</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>298.15</td>
<td>10 000</td>
<td>0.0</td>
<td>-2.7</td>
</tr>
<tr>
<td>$v$ ($\rightarrow aP$)</td>
<td>512</td>
<td>274.15</td>
<td>1</td>
<td>0.0</td>
<td>-13.0</td>
</tr>
<tr>
<td>[10$^{-4}$ m$^3$ mol$^{-1}$]</td>
<td>724</td>
<td>274.15</td>
<td>1</td>
<td>0.0</td>
<td>-11.5</td>
</tr>
<tr>
<td></td>
<td>724</td>
<td>274.15</td>
<td>1</td>
<td>0.0</td>
<td>-11.7</td>
</tr>
<tr>
<td></td>
<td>724</td>
<td>274.15</td>
<td>1</td>
<td>0.0</td>
<td>-10.3</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>274.15</td>
<td>1</td>
<td>0.0</td>
<td>-10.5</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>274.15</td>
<td>1</td>
<td>0.0</td>
<td>-9.3</td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter Y (or y') considered ($\Delta H$ or $v$ in Method I), the parameter Y' (or y'; between parentheses) to be evaluated using the simulation set, the number of water molecules $N_W$ in the system, the simulation temperature $T$ and pressure $P$, the correction terms $\Delta Y_A$, $\Delta Y_B$, $\Delta Y_C^A$, $\Delta Y_C^B$ and $\Delta Y_C^C$ for Variant B, the corresponding sum $\Delta Y_{\text{cor}}$, the raw simulation result $\Delta Y_{\text{hyd,raw}}$ and the corrected result $\Delta Y_{\text{hyd}}$. The quantities Y' to be evaluated correspond to the hydration enthalpy $\Delta H$ and the aqueous ion partial molar volume $v$, heat capacity $c_P$, volume-compressibility $k_T$ and volume-expansivity $a_P$. For the sets $\Delta H(\rightarrow \Delta H)$ and $v(\rightarrow v)$, the reported results directly correspond to the $\Delta Y_{\text{hyd,raw}}$ estimates for both variants (no distinction between A and B for these quantities). For the sets $\Delta H(\rightarrow cP)$, $v(\rightarrow kP)$ and $v(\rightarrow aP)$, the finite-difference or least-squares fitting procedure applied for extracting $\Delta Y_{\text{hyd,raw}}$ from the reported $\Delta Y_{\text{hyd}}$ values are illustrated in Figure A.1 for the systems involving 1024 water molecules. The resulting estimates for $\Delta Y_{\text{hyd}}$ in Variant B can be found in Table 2.5 of the main text, along with corresponding estimates and correction terms for Variant A. (b) For $v$, $c_P$, $k_T$ and $a_P$, these entries refer to the quantities usually written $\theta_{\text{hyd,raw}}$ and $\theta_{\text{hyd}}$ (equivalent to $\Delta Y_{\text{hyd,raw}}$ and $\Delta Y_{\text{hyd}}$ in the present context owing to Eqs. 2.8 and 2.9 in the main text).
Figure A.1: Finite-difference or least-squares fitting procedure used to calculate hydration and aqueous partial molar parameters for Na$^+$ or Cl$^-$ via Method IB. The raw simulation data is reported in Tables A.2 and A.3. The figure illustrates the calculation of $c_P$ based on the $\Delta H(\rightarrow c_P)$ set involving different temperatures $T$ (a,b), of $k_T$ based on the $v(\rightarrow k_T)$ set involving different pressures $P$ (c,d), and of $a_P$ based on the $v(\rightarrow a_P)$ set involving different temperatures $T$ (e,f). The system considered involves 1024 water molecules. The left panels (a,c,e) refer to Na$^+$ and the right panels (b,d,f) to Cl$^-$. The results are reported in Tables 2.4 and 2.5 of the main text, along with $\Delta H$ and $v$ from the $\Delta H(\rightarrow \Delta H)$ and $v(\rightarrow v)$ sets, respectively, for both variants (no distinction between A and B for these quantities) and with corresponding estimates and correction terms for Variant A.
Additional details for method II

Table A.4: Raw simulation data used to calculate hydration and aqueous partial molar parameters for Na$^+$ via Method II, along with correction terms and corrected results for Variant B.$^{(a)}$

<table>
<thead>
<tr>
<th>Parameter $G$ ($\rightarrow Y$)</th>
<th>$N_W$</th>
<th>$T$ [K]</th>
<th>$P$ [bar]</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta G$ ($\rightarrow \Delta G$) [kJ mol$^{-1}$]</td>
<td></td>
<td></td>
<td></td>
<td>$\Delta G_A$</td>
<td>$\Delta G_B$</td>
</tr>
<tr>
<td>724 298.15 1</td>
<td>0.0</td>
<td>-68.5</td>
<td>-77.2</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>1024 298.15 1</td>
<td>0.0</td>
<td>-61.1</td>
<td>-77.2</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>$\Delta G$ ($\rightarrow \Delta H, s$) [kJ mol$^{-1}$]</td>
<td></td>
<td></td>
<td></td>
<td>$\Delta G_A$</td>
<td>$\Delta G_B$</td>
</tr>
<tr>
<td>724 298.15 1</td>
<td>0.0</td>
<td>-68.7</td>
<td>-77.4</td>
<td>0.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>724 298.15 1</td>
<td>0.0</td>
<td>-68.6</td>
<td>-77.3</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>724 298.15 1</td>
<td>0.0</td>
<td>-68.5</td>
<td>-77.2</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>724 310.65 1</td>
<td>0.0</td>
<td>-68.4</td>
<td>-76.9</td>
<td>0.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>1024 273.15 1</td>
<td>0.0</td>
<td>-61.3</td>
<td>-77.4</td>
<td>0.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>1024 285.15 1</td>
<td>0.0</td>
<td>-61.2</td>
<td>-77.4</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>1024 298.15 1</td>
<td>0.0</td>
<td>-61.1</td>
<td>-77.2</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>1024 310.65 1</td>
<td>0.0</td>
<td>-61.0</td>
<td>-76.9</td>
<td>0.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>1024 323.15 1</td>
<td>0.0</td>
<td>-60.8</td>
<td>-76.5</td>
<td>0.0</td>
<td>-1.1</td>
</tr>
</tbody>
</table>

$^{(a)}$ The quantities reported are the parameter $Y$ (or $y$) considered (always $\Delta G$ in Method II), the parameter $Y'$ (or $y'$; between parentheses) to be evaluated using the simulation set, the number of water molecules $N_W$ in the system, the simulation temperature $T$ and pressure $P$, the correction terms $\Delta G_A$, $\Delta G_B$, $\Delta G_{Cl}$, $\Delta G_{D}$ and $\Delta G_{total}$ for Variant B, the corresponding sum $\Delta G_{cor}$, the cavitation free energy $\Delta G_{cav}$, the raw charging free energy $\Delta G_{h,q,raw}$, and the corrected hydration free energy $\Delta G_{h,cor}$. The quantities $Y'$ to be evaluated correspond to the hydration free energy $\Delta G$ and enthalpy $\Delta H$, as well as the aqueous ion partial molar entropy $s$ and volume $v$. For the set $\Delta G$ ($\rightarrow \Delta G$), the reported results directly correspond to the $\Delta G_{h,cor}$ estimates for both variants (no distinction between A and B for this quantity). For the sets $\Delta G$ ($\rightarrow \Delta H, s$) and $\Delta G$ ($\rightarrow v$), the finite-difference or least-squares fitting procedure applied for extracting $\Delta Y_{h,cor}$ from the reported $\Delta G_{h,cor}$ values are illustrated in Figure A.2 for the systems involving 1024 water molecules. The resulting estimates for $\Delta Y_{h,cor}$ in Variant B can be found in Table 2.6 of the main text, along with corresponding estimates and correction terms for Variant A.
Table A.5: Raw simulation data used to calculate hydration and aqueous partial molar parameters for Cl$^-$ via Method II, along with correction terms and corrected results for Variant B.$^{(a)}$

<table>
<thead>
<tr>
<th>Parameter $G$ ($\rightarrow Y'$)</th>
<th>$N_W$</th>
<th>$T$ [K]</th>
<th>$P$ [bar]</th>
<th>Correction terms</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\Delta G$</td>
<td>$\Delta G_{ch,raw}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\Delta G_A$</td>
<td>$\Delta G_B$</td>
</tr>
<tr>
<td>$\Delta G$ ($\rightarrow \Delta G$)</td>
<td></td>
<td></td>
<td></td>
<td>[kJ mol$^{-1}$]</td>
<td>[kJ mol$^{-1}$]</td>
</tr>
<tr>
<td>724</td>
<td>298.15</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1024</td>
<td>298.15</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>724</td>
<td>273.15</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1024</td>
<td>273.15</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>724</td>
<td>285.65</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1024</td>
<td>285.65</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>724</td>
<td>310.65</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1024</td>
<td>310.65</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\Delta G$ ($\rightarrow \Delta H, s$)</td>
<td></td>
<td></td>
<td></td>
<td>[kJ mol$^{-1}$]</td>
<td>[kJ mol$^{-1}$]</td>
</tr>
<tr>
<td>724</td>
<td>298.15</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1024</td>
<td>298.15</td>
<td>1</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>724</td>
<td>298.15</td>
<td>5000</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1024</td>
<td>298.15</td>
<td>10000</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter $Y$ (or $y$) considered (always $\Delta G$ in Method II), the parameter $Y'$ (or $y'$; between parentheses) to be evaluated using the simulation set, the number of water molecules $N_W$ in the system, the simulation temperature $T$ and pressure $P$, the correction terms $\Delta G_A$, $\Delta G_B$, $\Delta G_C$, $\Delta G_D$, and $\Delta G_C^{\text{v}}$ for Variant B, the corresponding sum $\Delta G^{\text{v}}$, the cavitation free energy $\Delta G_{\text{cav}}$, the raw charging free energy $\Delta G_{\text{ch}, \text{raw}}$, and the corrected hydration free energy $\Delta G_{\text{hyd}}$. The quantities $Y'$ to be evaluated correspond to the hydration free energy $\Delta G$ and enthalpy $\Delta H$, as well as the aqueous ion partial molar entropy $s$ and volume $v$. For the set $\Delta G(\rightarrow \Delta G)$, the reported results directly correspond to the $\Delta G_{\text{hyd}}^{\text{v}}$ estimates for both variants (no distinction between A and B for this quantity). For the sets $\Delta G(\rightarrow \Delta H, s)$ and $\Delta G(\rightarrow v)$, the finite-difference or least-squares fitting procedure applied for extracting $\Delta Y_{\text{hyd}}^{\text{v}}$ from the reported $\Delta G_{\text{hyd}}^{\text{v}}$ values are illustrated in Figure A.2 for the systems involving 1024 water molecules. The resulting estimates for $\Delta Y_{\text{hyd}}^{\text{v}}$ in Variant B can be found in Table 2.7 of the main text, along with corresponding estimates and correction terms for Variant A.
Figure A.2: Least-squares fitting procedure used to calculate hydration and aqueous partial molar parameters for Na$^+$ or Cl$^-$ via Method IIB. The raw simulation data is reported in Tables A.4 and A.5. The figure illustrates the calculation of $\Delta H$ and $s$ based on the $\Delta G$($\rightarrow \Delta H, s$) set involving different temperatures $T$ (a,b) and of $v$ based on the $\Delta G$($\rightarrow v$) set involving different pressures $P$ (c,d). Each panel consists of two graphs illustrating the separate fitting of $\Delta G_{cav}$ (top) and $\Delta G_{chg}$ (bottom). Note that $\Delta G_{chg}$ refers to the corrected charging free energy in Method IIB. The system considered involves 1024 water molecules. The left panels (a,c) refer to Na$^+$ and the right panels (b,d) to Cl$^-$. The results are reported in Tables 2.6 and 2.7 of the main text, along with $\Delta G$ from the $\Delta G$($\rightarrow \Delta G$) set for both variants (no distinction between A and B for this quantity), and with corresponding estimates and correction terms for Variant A.
Summary of the results

Table A.6: Corrected results corresponding to hydration and aqueous partial molar parameters for Na\(^+\) calculated via Methods I or II along with Variants A or B\(^{(a)}\)

<table>
<thead>
<tr>
<th>Property</th>
<th>(N_W)</th>
<th>Method/Variant</th>
<th>Recommended (\Delta Y_{\text{hyd}}) ((^\circ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta G) ([\text{kJ mol}^{-1}])</td>
<td>724</td>
<td>-</td>
<td>(-431.2 \pm 0.7)</td>
</tr>
<tr>
<td>(\Delta H) ([\text{kJ mol}^{-1}])</td>
<td>1024</td>
<td>-</td>
<td>(-430.9 \pm 0.6)</td>
</tr>
<tr>
<td>(s) ([\text{J mol}^{-1}\text{K}^{-1}])</td>
<td>512</td>
<td>(-481.4 \pm 0.9)</td>
<td>(-481.4 \pm 0.9)</td>
</tr>
<tr>
<td>(v) ([10^{-6} \text{m}^3\text{mol}^{-1}])</td>
<td>724</td>
<td>(-478.4 \pm 1.0)</td>
<td>(-478.4 \pm 1.0)</td>
</tr>
<tr>
<td>(c_P) ([\text{J mol}^{-1}\text{K}^{-1}])</td>
<td>1024</td>
<td>(-476.3 \pm 1.1)</td>
<td>(-476.3 \pm 1.1)</td>
</tr>
<tr>
<td>(k_T) ([10^{-10} \text{m}^3\text{mol}^{-1}\text{bar}^{-1}])</td>
<td>724</td>
<td>(-33.2 \pm 0.6)</td>
<td>(-33.2 \pm 0.6)</td>
</tr>
<tr>
<td>(a_P) ([10^{-6} \text{m}^3\text{mol}^{-1}\text{K}^{-1}])</td>
<td>1024</td>
<td>(-33.2 \pm 0.8)</td>
<td>(-33.2 \pm 0.8)</td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter \(Y\) (or \(y\)) considered, the number of water molecules \(N_W\) in the system, and the corrected simulation estimate \(\Delta Y_{\text{hyd}}\). Values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\), volume \(V\), heat capacity \(C_P\), volume-compressibility \(K_T\) and volume-expansivity \(A_P\). The data refers to the standard pressure \(P^0 = 1\) bar and temperature \(T^\circ = 298.15\) K. The simulation values are reported from the corresponding entries of Tables 2.4 and 2.6 of the main text. The recommended value is reported from the corresponding entry in Table 2.2. These recommended values are inferred from experiment, but correspond to experimentally elusive quantities, and are thus to be considered with some caution.\(^{166}\) (b) For \(s, v, c_P, k_T\) and \(a_P\), the entries of this table refer to the quantities usually written \(y_{\text{hyd}}\) (equivalent to \(\Delta Y_{\text{hyd}}\) in the present context owing to Eqs. 2.8 and 2.9 in the main text).
Table A.7: Corrected results corresponding to hydration and aqueous partial molar parameters for Cl\(^-\) calculated \textit{via} Methods I or II along with Variants A or B.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Property</th>
<th>(N_W)</th>
<th>\text{Method/Variant}</th>
<th>\text{Recommended}</th>
<th>(\Delta Y_{\text{hyd}}^\text{w}^{(1)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta G) [kJ mol(^{-1})]</td>
<td>724</td>
<td>IIA</td>
<td>-321.4 ± 0.7</td>
<td>-321.4 ± 0.7</td>
</tr>
<tr>
<td>1024</td>
<td>IB</td>
<td>-318.4 ± 0.7</td>
<td>-318.4 ± 0.7</td>
<td></td>
</tr>
<tr>
<td>(\Delta H) [kJ mol(^{-1})]</td>
<td>512</td>
<td>-370.3 ± 0.9</td>
<td>-370.3 ± 0.9</td>
<td>-353.2 ± 5.2</td>
</tr>
<tr>
<td>724</td>
<td>-366.9 ± 1.0</td>
<td>-366.9 ± 1.0</td>
<td>-350.0 ± 5.5</td>
<td>-350.3 ± 5.5</td>
</tr>
<tr>
<td>1024</td>
<td>-364.6 ± 1.1</td>
<td>-364.6 ± 1.1</td>
<td>-350.0 ± 5.5</td>
<td>-350.3 ± 5.5</td>
</tr>
<tr>
<td>(s) [J mol(^{-1}) K(^{-1})]</td>
<td>724</td>
<td>-106.8 ± 17.4</td>
<td>-106.7 ± 17.4</td>
<td>-41.09</td>
</tr>
<tr>
<td>1024</td>
<td>-106.1 ± 18.2</td>
<td>-106.1 ± 18.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>(v) ([10^{-6} \text{m}^3 \text{mol}^{-1}])</td>
<td>512</td>
<td>38.0 ± 0.6</td>
<td>38.0 ± 0.6</td>
<td>22.21</td>
</tr>
<tr>
<td>724</td>
<td>38.0 ± 0.6</td>
<td>38.0 ± 0.6</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td>39.1 ± 0.8</td>
<td>39.1 ± 0.8</td>
<td>42.7 ± 1.0</td>
<td>31.1 ± 1.0</td>
</tr>
<tr>
<td>(c_P) [J mol(^{-1}) K(^{-1})]</td>
<td>512</td>
<td>184.4 ± 28.5</td>
<td>195.7 ± 28.5</td>
<td>-</td>
</tr>
<tr>
<td>724</td>
<td>180.1 ± 30.8</td>
<td>192.1 ± 30.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1024</td>
<td>125.8 ± 37.4</td>
<td>138.3 ± 37.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(k_f) ([10^{-10} \text{m}^3 \text{mol}^{-1} \text{bar}^{-1}])</td>
<td>512</td>
<td>8.6 ± 0.6</td>
<td>-</td>
<td>-21.29</td>
</tr>
<tr>
<td>724</td>
<td>4.6 ± 0.7</td>
<td>8.5 ± 0.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1024</td>
<td>4.5 ± 0.8</td>
<td>10.1 ± 0.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(a_P) ([10^{-8} \text{m}^3 \text{mol}^{-1} \text{K}^{-1}])</td>
<td>512</td>
<td>-18.5 ± 2.1</td>
<td>-19.7 ± 2.1</td>
<td>4.62</td>
</tr>
<tr>
<td>724</td>
<td>-18.5 ± 2.1</td>
<td>-19.7 ± 2.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1024</td>
<td>-20.4 ± 2.4</td>
<td>-21.7 ± 2.4</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^{(a)}\) The quantities reported are the parameter \(Y\) (or \(y\)) considered, the number of water molecules \(N_W\) in the system, and the corrected simulation estimate \(\Delta Y_{\text{hyd}}^\text{w}\). Values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\), volume \(V\), heat capacity \(C_P\), volume-compressibility \(K_f\) and volume-expansivity \(A_P\). The data refers to the standard pressure \(P^* = 1\) bar and temperature \(T^* = 298.15\) K. The simulation values are reported from the corresponding entries of Tables 2.5 and 2.7 of the main text. The recommended value is reported from the corresponding entry in Table 2.2. These recommended values are inferred from experiment, but correspond to experimentally elusive quantities, and are thus to be considered with some caution.\(^{166}\) (b) For \(s\), \(v\), \(c_P\), \(k_f\) and \(a_P\), the entries of this table refer to the quantities usually written \(\gamma_{\text{aq}}\) (equivalent to \(\Delta Y_{\text{hyd}}^\text{w}\) in the present context owing to Eqs. 2.8 and 2.9 in the main text).
Table A.8: Corrected results corresponding to hydration and aqueous partial molar parameters for the Na\(^+\)+Cl\(^-\) pair calculated via Methods I or II along with Variants A or B.(a)

<table>
<thead>
<tr>
<th>Property</th>
<th>N(_{W})</th>
<th>I A</th>
<th>I B</th>
<th>IIA</th>
<th>IIB</th>
<th>Experimental (\Delta Y_{\text{hyd}}) (^{(1)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta G) [kJ mol(^{-1})]</td>
<td>724</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-749.3 ± 0.9</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-749.3 ± 0.9</td>
</tr>
<tr>
<td>(\Delta H) [kJ mol(^{-1})]</td>
<td>724</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-749.3 ± 0.9</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-749.3 ± 0.9</td>
</tr>
<tr>
<td>(s) [J mol(^{-1}) K(^{-1})]</td>
<td>724</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-117.49</td>
</tr>
<tr>
<td>(v) [10(^{-6}) m(^3) mol(^{-1})]</td>
<td>724</td>
<td>4.5 ± 0.8</td>
<td>4.5 ± 0.8</td>
<td>-</td>
<td>-</td>
<td>16.35</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>5.9 ± 1.1</td>
<td>5.9 ± 1.1</td>
<td>13.2 ± 1.4</td>
<td>22.4 ± 1.4</td>
<td></td>
</tr>
<tr>
<td>(c_p) [J mol(^{-1}) K(^{-1})]</td>
<td>724</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-125.74</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>(k_T) [10(^{-10}) m(^3) mol(^{-1}) bar(^{-1})]</td>
<td>724</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-48.74</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>(a_P) [10(^{-8}) m(^3) mol(^{-1}) K(^{-1})]</td>
<td>724</td>
<td>7.0 ± 2.9</td>
<td>8.1 ± 2.9</td>
<td>-</td>
<td>-</td>
<td>7.04</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>0.4 ± 3.8</td>
<td>1.3 ± 3.8</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

(a) The quantities reported are the parameter \(Y\) (or \(y\)) considered, the number of water molecules \(N_{W}\) in the system, and the corrected simulation estimate \(\Delta Y_{\text{hyd}}\). Values are provided for the free energy \(G\), enthalpy \(H\), entropy \(S\), volume \(V\), heat capacity \(C_p\), volume-compressibility \(K_T\) and volume-expansivity \(a_P\). The reported values correspond to sum of parameters over the individual ions Na\(^+\) and Cl\(^-\). The data refers to the standard pressure \(P^\circ = 1\) bar and temperature \(T^\circ = 298.15\) K. The simulation values are summed from the corresponding entries of Tables A.6 and A.7. The experimental value is summed from the corresponding entries in Table 2.2. These experimental values over the ion pair are in principle experimentally accessible. However, the entries for \(c_{P,\text{aq}}\), \(k_{T,\text{aq}}\) and \(a_{P,\text{aq}}\) should still be taken with some caution in view of the sparseness and indirect nature of the corresponding experimental determinations.\(^{166}\)

(a) For \(s\), \(v\), \(c_p\), \(k_T\) and \(a_P\), the entries of this table refer to the quantities usually written \(y_{\text{aq}}\) (equivalent to \(\Delta Y_{\text{hyd}}\) in the present context owing to Eqs. 2.8 and 2.9 in the main text).
Supplementary material for chapter 3

Additional tests of the parameter update scheme, structural properties of neat DNA and DNA-netropsin complexes, details of TI calculations, sensitivity of EDS results with respect to EDS parameters.
Parameter update scheme for alchemical perturbations of tetrahydroisoquinoline derivatives in water

The parameter update scheme presented in Chapter 3 was tested and calibrated by performing alchemical perturbations on a small set of tetrahydroisoquinoline derivatives in water. The free-enthalpy changes associated with these perturbations as well as a target set of EDS parameters that allow sufficient sampling of the end states during an EDS reference state simulation, were available from previous work. Figure B.1 shows a typical update profile for one of the studied perturbations.

**Figure B.1:** Typical evolution of smoothness parameter $s$ and energy offset $E = E_B^R$ during an EDS parameter update simulation for the alchemical perturbation of tetrahydroisoquinoline derivatives in water. The present case represents the perturbation 3(R) – 8(S) (see Fig. 4 of reference [320] for the nomenclature). The free enthalpy difference $\Delta G$ calculated at each update point is displayed by the red line. The four update phases are separated by vertical lines. Initial values and update parameters were $s_{\text{init}} = 0.01$, $E_B^{R, \text{init}} = -405.5 \text{ kJ mol}^{-1}$, $E_{\text{sep}} = -404.1 \text{ kJ mol}^{-1}$, and $\Delta E_{\text{tr}} = 24.0 \text{ kJ mol}^{-1}$. The initial simulation time between two updates in phase 1 was 200 ps. The optimized EDS parameters are $s_{\text{opt}} = 0.0433$ and $E_B^{R, \text{opt}} = -390.1 \text{ kJ mol}^{-1}$. 
Structural properties of neat DNA and DNA-netropsin

Figure B.2 displays the time series of the atom-positional rmsd from the starting configuration for each end state simulation. The DNA double helix diverges from the starting structure on average by 0.5 nm. For some DNA-netropsin complexes (AAAGA, AAAGG, AAGGG) the rmsd of netropsin is larger than the rmsd of the DNA double helix due to conformational changes in the tails of netropsin. However, in all cases netropsin remains in the binding site during the 10 ns production run. Figures B.3 and B.4 show the high occurrence of WC-hydrogen bonds between the bases in the neat DNA duplexes and the DNA-netropsin complexes.

Figure B.2: Time-evolution of the atom-positional rmsd from the starting structure of each simulation, calculated based on all heavy atoms (black), heavy atoms of the DNA bases (red), heavy atoms of the DNA backbone (green), and netropsin heavy atoms (blue).
Figure B.3: Occurrence of Watson-Crick hydrogen bonds in simulation of DNA duplexes (10 ns sampling time). Each of the 7 investigated perturbations is represent by one column. The first row represents the common end state AAAAAA, the second row represents the reference state, and the third row the end state of the perturbation. Hydrogen bonds between dummy atoms are shown in grey, hydrogen bonds of the perturbed base pair (A-T $\rightarrow$ G-C) are displayed in red. The sequence of the primary DNA strand is given along the x-axis.
Figure B.4: Occurrence of Watson-Crick hydrogen bonds in simulation of netropsin-DNA complexes (10 ns sampling time). Each of the 7 investigated perturbations is represented by one column. The first row represents the common end state AAAAA, the second row represents the reference state, and the third row the end state of the perturbation. Hydrogen bonds between dummy atoms are shown in grey, hydrogen bonds of the perturbed base pair (A-T → G-C) are displayed in red. The sequence of the primary DNA strand is given along the x-axis.
Energy-difference distributions in case of undersampling

Figure B.5: Energy difference distributions for the reference state, \( \rho_R(\Delta V; \Delta V_{BA}) \) (black), and the two end states, \( \rho_A(\Delta V; \Delta V_{BA}) \) (red), \( \rho_B(\Delta V; \Delta V_{BA}) \) (green), for each of the seven perturbations in neat DNA and DNA-netropsin as obtained from the EDS1 reference state simulations showing undersampling. The energy difference distributions of the end states were obtained by reweighting.

TI calculations

In the TI calculations of Dolenc et al.\(^{437}\) three bond lengths in the 6-membered ring were altered when transforming adenine to guanine and two bond lengths when transforming thymine to cytosine. In the EDS calculations reported in the present work all bond lengths in the 6-membered rings were fixed to the values of adenine/thymine. To estimate whether this simplification affects free enthalpy differences we performed a TI calculation for the transformation \( \text{AAAAA(NET)} \rightarrow \text{AAGAA(NET)} \) without changing the bond lengths using the same computational box as was used in the EDS simulations for this system. The free enthalpy difference \( \Delta G \) between the two states was calculated using a \( \lambda \)-dependent Hamiltonian \( H(\lambda) \)^{39,119} and
evaluating

\[
\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda \, d\lambda
\]  

where \( \langle \rangle_\lambda \) denotes an ensemble average at a given \( \lambda \)-value. In the current work, the integral was evaluated using the trapezoidal rule based on simulations at 21 \( \lambda \)-values spaced equidistantly between 0 and 1. A 0.4 ns production simulation, preceded by a 0.1 ns equilibration period was performed at each intermediate \( \lambda \)-value using the final coordinates and velocities from the previous \( \lambda \)-value as initial values. At the end points (\( \lambda = 0 \), 1) a 0.9 ns production simulation was performed, preceded by a 0.1 ns equilibration period. To prevent instabilities at the end points the soft core approach was employed with the softness parameter \( \alpha_{LJ} = 0.5 \) for the Lennard-Jones interactions and \( \alpha_{C}^{(2)} = 0.5 \text{nm}^2 \) for the electrostatic interactions.

The free enthalpy difference calculated with constant bond lengths, \( \Delta G = -155.8 \pm 3.4 \text{kJ mol}^{-1} \), agrees very well with the value, \( \Delta G = -156.2 \pm 3.4 \text{kJ mol}^{-1} \), reported earlier, showing that the contribution from the change of bond lengths to the free enthalpy difference is negligible in the cases studied here. Statistical errors for the individual thermodynamic integration \( \lambda \)-values were computed according to the block averaging method. The statistical error of the integral was obtained by summing the errors of the \( n \) individual simulations (with weights 1/2 for the two end points; trapezoidal rule) and dividing the result by \( n - 1 \).

The averaged derivative \( \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \) as function of \( \lambda \) is shown in Figure B.6. The red curve was obtained by doubling the sampling time at each \( \lambda \)-value. The free enthalpy change for AAAAA(NET) \( \rightarrow \) AAGAA(NET) with this longer sampling time is \( \Delta G = -154.9 \pm 3.4 \text{kJ mol}^{-1} \) showing that no convergence problem was present for this perturbation.

Sensitivity of free-enthalpy estimates with respect to the smoothness parameter and energy offset

For the perturbation of one base pair, AAAAA \( \rightarrow \) AAAGA, we have studied the influence of the \( s \) and \( E^R_B \) parameters on the estimate of the free enthalpy difference obtained in a 10 ns production simulation. For EDS parameters in the range of \( 0.055 < s < 0.097 \) and \( -166.7 < E^R_B < -161.0 \text{kJ mol}^{-1} \) the free enthalpy difference is essentially independent of \( s \) and \( E^R_B \). For all parameter combinations studied, the resulting change in free enthalpy is smaller than the change of the energy offset (see Table B.1). However, the degree of sampling obtained with the various parameter sets differs as can be seen from the energy difference distributions and the reweighted end state energy distributions that can be directly compared to the energy distribu-
Figure B.6: Free enthalpy derivative as function of $\lambda$ for the change AAAAAA(NET), $\lambda=0 \rightarrow$ AAGAAA(NET), $\lambda=1$. Black curve: 0.4 ns sampling time at intermediate points and 0.9 ns at end points. Red curve: 0.9 ns sampling time at intermediate points and 1.9 ns at end points.

The reweighted end state energy distributions are in very good agreement with those from non-EDS simulations. Similar observations were made for the perturbations of two base pairs, AAAAA $\rightarrow$ AAAGG and AAAAAA $\rightarrow$ AAGAG, data not shown. For the largest perturbation studied, AAAAA $\rightarrow$ AAGGG, the sensitivity towards the EDS parameters is larger. While it was still possible to identify EDS parameters that lead to sufficient sampling, a change in the energy offset results in a change of free enthalpy that is larger in magnitude, see Table B.2. Figure B.8 shows how the sampling is shifted from state AAGGG to AAAAA when the energy offset for state B, i.e. AAGGG, becomes more negative. However, by monitoring the time series of $\Delta V_{BA}$ as well as the energy difference and end state energy distributions insufficient sampling is easy to detect.
Table B.1: Free enthalpy change $\Delta G$ for the change AAAAA → AAAGA in the DNA duplex and EDS smoothness parameters and energy offsets for the reference state AAAXA

<table>
<thead>
<tr>
<th>code</th>
<th>$s$</th>
<th>$E^R_B$ (kJ mol$^{-1}$)</th>
<th>$\Delta G$ (kJ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.055</td>
<td>$-161.0$</td>
<td>$-170.0 \pm 0.6$</td>
</tr>
<tr>
<td>P2</td>
<td>0.058</td>
<td>$-161.0$</td>
<td>$-169.0 \pm 0.5$</td>
</tr>
<tr>
<td>P3</td>
<td>0.061</td>
<td>$-161.0$</td>
<td>$-169.0 \pm 0.4$</td>
</tr>
<tr>
<td>P4</td>
<td>0.064</td>
<td>$-161.0$</td>
<td>$-169.9 \pm 0.6$</td>
</tr>
<tr>
<td>P5</td>
<td>0.055</td>
<td>$-166.0$</td>
<td>$-169.3 \pm 0.6$</td>
</tr>
<tr>
<td>P6</td>
<td>0.058</td>
<td>$-166.0$</td>
<td>$-168.6 \pm 0.5$</td>
</tr>
<tr>
<td>P7</td>
<td>0.061</td>
<td>$-166.0$</td>
<td>$-169.5 \pm 0.4$</td>
</tr>
<tr>
<td>P8</td>
<td>0.064</td>
<td>$-166.0$</td>
<td>$-168.3 \pm 0.4$</td>
</tr>
<tr>
<td>P9</td>
<td>0.0974</td>
<td>$-166.7$</td>
<td>$-169.0 \pm 0.9$</td>
</tr>
</tbody>
</table>

Table B.2: Free enthalpy change $\Delta G$ for the change AAAAA → AAGGG in the DNA duplex and EDS smoothness parameters and energy offsets for the reference state AAXXX

<table>
<thead>
<tr>
<th>code</th>
<th>$s$</th>
<th>$E^R_B$ (kJ mol$^{-1}$)</th>
<th>$\Delta G$ (kJ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.0199</td>
<td>$-496.0$</td>
<td>$-505.7 \pm 1.3$</td>
</tr>
<tr>
<td>P2</td>
<td>0.0199</td>
<td>$-497.0$</td>
<td>$-502.1 \pm 1.2$</td>
</tr>
<tr>
<td>P3</td>
<td>0.0199</td>
<td>$-498.0$</td>
<td>$-500.2 \pm 1.5$</td>
</tr>
</tbody>
</table>
Figure B.7: Left: Energy difference distributions for the reference state, \( \rho_R(\Delta V; \Delta V_{BA}) \) (black), and the two end states, \( \rho_A(\Delta V; \Delta V_{BA}) \) (red), \( \rho_B(\Delta V; \Delta V_{BA}) \) (green), for AAAAA \( \rightarrow \) AAAGA obtained using different combinations of smoothness parameter and energy offset as specified in Table B.1. The energy difference distributions of the end states were determined by reweighting. Right: Nonbonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained through reweighting (full lines) from the different reference state simulations and from independent MD simulations of the two end states (dashed lines). The red colour corresponds to end state AAAAA.
Figure B.8: Left column: Time series of the energy difference $\Delta V_{BA} = V_B - V_A$ in an EDS reference state simulation of neat DNA, in which B represents a G-C base pair at position X and A represents an A-T base pair at position X, for the different combinations of smoothness parameter and energy offset as specified in Table B.2. Middle column: Energy difference distributions for the reference state, $\rho_R(\Delta V; \Delta V_{BA})$ (black), and the two end states, $\rho_A(\Delta V; \Delta V_{BA})$ (red), $\rho_B(\Delta V; \Delta V_{BA})$ (green), for AAAAA $\rightarrow$ AAGGG obtained using different combinations of smoothness parameter and energy offset as specified in Table B.2. The energy difference distributions of the end states were determined by reweighting. Right column: Nonbonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained through reweighting (full lines) from the different reference state simulations and from independent MD simulations of the two end states (dashed lines). The red colour corresponds to end state AAAAA.
Supplementary material for chapter 4

EDS parameters, details on the topologies, plots showing the atom-positional rmsd and secondary structures during the MD simulations, plots showing the time series of the energy difference between the two end states during the EDS reference-state simulations, and details regarding the computation of statistically optimal free-enthalpy differences.
Table C.1: EDS reference state parameters $s$ and $E^R_B$ (kJ mol$^{-1}$) used in the EDS production simulations.

<table>
<thead>
<tr>
<th>code</th>
<th>perturbation (A $\rightarrow$ B)</th>
<th>3$\alpha$ structure</th>
<th>$E^R_B$</th>
<th>4$\beta+\alpha$ structure</th>
<th>$E^R_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LIL $\rightarrow$ AIL</td>
<td>0.1035</td>
<td>24.12</td>
<td>0.0530</td>
<td>19.00</td>
</tr>
<tr>
<td>2</td>
<td>LIL $\rightarrow$ LFL</td>
<td>0.0782</td>
<td>−22.60</td>
<td>0.0788</td>
<td>−20.19</td>
</tr>
<tr>
<td>3</td>
<td>LIL $\rightarrow$ LIY</td>
<td>0.0653</td>
<td>−77.00</td>
<td>0.0586</td>
<td>−98.45</td>
</tr>
<tr>
<td>4</td>
<td>AIL $\rightarrow$ AFL</td>
<td>0.0700</td>
<td>−15.00</td>
<td>0.0700</td>
<td>−22.00</td>
</tr>
<tr>
<td>5</td>
<td>AIL $\rightarrow$ AIY</td>
<td>0.0657</td>
<td>−86.00</td>
<td>0.0599</td>
<td>−94.00</td>
</tr>
<tr>
<td>6</td>
<td>LFL $\rightarrow$ AFL</td>
<td>0.0900</td>
<td>23.00</td>
<td>0.0530</td>
<td>21.00</td>
</tr>
<tr>
<td>7</td>
<td>LFL $\rightarrow$ LFY</td>
<td>0.1050</td>
<td>−84.72</td>
<td>0.0673</td>
<td>−94.00</td>
</tr>
<tr>
<td>8</td>
<td>LIY $\rightarrow$ AIY</td>
<td>0.0600</td>
<td>22.00</td>
<td>0.0600</td>
<td>18.00</td>
</tr>
<tr>
<td>9</td>
<td>LIY $\rightarrow$ LFY</td>
<td>0.0720</td>
<td>−15.00</td>
<td>0.0590</td>
<td>−16.00</td>
</tr>
<tr>
<td>10</td>
<td>AFL $\rightarrow$ AFY</td>
<td>0.0610</td>
<td>−82.00</td>
<td>0.0600</td>
<td>−94.00</td>
</tr>
<tr>
<td>11</td>
<td>LFY $\rightarrow$ AFY</td>
<td>0.0600</td>
<td>22.00</td>
<td>0.0518</td>
<td>19.60</td>
</tr>
<tr>
<td>12</td>
<td>AIY $\rightarrow$ AFY</td>
<td>0.0630</td>
<td>−22.00</td>
<td>0.0700</td>
<td>−21.50</td>
</tr>
</tbody>
</table>

$^a$Combined analysis of two trajectories sharing the same $s$-parameter but different energy offsets of $E^R_{B,1} = -21.0$ kJ mol$^{-1}$ and $E^R_{B,2} = -22.0$ kJ mol$^{-1}$

Figure C.1: Single topologies used for the perturbations L $\leftrightarrow$ A (a), I $\leftrightarrow$ F (b), and L $\leftrightarrow$ Y (c). The GROMOS atom types and charges corresponding to each atom are provided in Tables C.2, C.3, and C.4. Charge groups are marked with different colors.
Table C.2: GROMOS atom types and charges (e) used in the single topology for the perturbation leucine (state A) ↔ alanine (state B). Atom type codes refer to the 54A7 version of the GROMOS force field. The atom type code 22 refers to a dummy atom. Values for unperturbed atoms are only provided for state A. The sequential numbers correspond to Figure C.1(a).

<table>
<thead>
<tr>
<th>Atom</th>
<th>state A</th>
<th></th>
<th>state B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>atom type</td>
<td>charge</td>
<td>atom type</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>−0.310</td>
<td>id</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>+0.310</td>
<td>id</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>0.000</td>
<td>id</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>0.000</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>0.000</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>0.000</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>0.000</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>+0.450</td>
<td>id</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>−0.450</td>
<td>id</td>
</tr>
</tbody>
</table>
Table C.3: GROMOS atom types and charges (e) used in the single topology for the perturbation isoleucine (state A) ↔ phenylalanine (state B). Atom type codes refer to the 54A7 version of the GROMOS force field.\textsuperscript{47,48} The atom type code 22 refers to a dummy atom. Values for unperturbed atoms are only provided for state A. The sequential numbers correspond to Figure C.1(b).

<table>
<thead>
<tr>
<th>Atom</th>
<th>state A</th>
<th>state B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>atom type</td>
<td>charge</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>−0.310</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>+0.310</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>14</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>17</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>18</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>19</td>
<td>12</td>
<td>+0.450</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>−0.450</td>
</tr>
</tbody>
</table>
Table C.4: GROMOS atom types and charges (e) used in the single topology for the perturbation leucine (state A) ↔ tyrosine (state B). Atom type codes refer to the 54A7 version of the GROMOS force field. The atom type code 22 refers to a dummy atom. Values for unperturbed atoms are only provided for state A. The sequential numbers correspond to Figure C.1(c).

<table>
<thead>
<tr>
<th>Atom</th>
<th>state A</th>
<th>state B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>atom type</td>
<td>charge</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>-0.310</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>+0.310</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>14</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>17</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>18</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>19</td>
<td>22</td>
<td>0.000</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>+0.450</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>-0.450</td>
</tr>
</tbody>
</table>
Figure C.2: Time series of the atom-positional RMSD of the $C_\alpha$ atoms of residues 10-50 ($3\alpha$ structure) and of all residues ($4\beta + \alpha$ structure), respectively, with respect to the first structure of each trajectory during 11 ns of MD simulations of the various sequence-structure combinations in water at 298 K. The rotational least-squares fit was performed with respect to the backbone atoms of residues 10-50 ($3\alpha$ structure) and of all residues ($4\beta + \alpha$ structure), respectively.
Figure C.3: Time series of the atom-positional RMSD of the Co atoms of residues 10-50 (3α structure) and of all residues (4β + α structure), respectively, with respect to the first structure of each trajectory during 11 ns of EDS reference state simulations in water at 298 K. The rotational least-squares fit was performed with respect to the backbone atoms of residues 10-50 (3α structure) and of all residues (4β + α structure), respectively. In case of the 4β + α structure the free enthalpy difference for perturbation 12 was obtained from two trajectories sharing the same s-parameter value but with slightly different energy offsets (see Table C.1).
Figure C.4: Time series of the secondary structure of the various sequence-structure combinations in water at 298 K during 11 ns of MD simulations. Color codes: $\alpha$-helix (red), bend (blue), $\beta$-strand (green).
Figure C.5: Time series of the secondary structure during 11 ns of the various EDS reference state simulations in water at 298 K. Color codes: \( \alpha \)-helix (red), bend (blue), \( \beta \)-strand (green). In case of the \( 4\beta + \alpha \) structure the free enthalpy difference for perturbation 12 was obtained from two trajectories sharing the same \( s \)-parameter value but with slightly different energy offsets (see Table C.1).
Figure C.6: Time series of the energy difference $\Delta V_{BA} = V_B - V_A$ of the various EDS reference state simulations in water at 298 K. The states A and B are defined in Table C.1.
Obtaining a statistically optimal set of free-enthalpy differences

Under the assumption that the calculated free-enthalpy differences $\Delta G_i$ as well as the hysteresis $\Delta$ of cycle closure are Gaussian distributed, we calculated a set of statistically optimal free-enthalpy differences $\hat{\Delta} G_i$ that fulfill the necessary condition of cycle closure by solving

$$\min \sum_i \frac{(\Delta G_i - \hat{\Delta} G_i)^2}{2\sigma_i^2} \quad \text{with} \quad i = 1 \ldots 12 \quad (C.1)$$

subject to

$$\begin{align*}
\Delta \hat{G}_1 + \Delta \hat{G}_4 - \Delta \hat{G}_6 - \Delta \hat{G}_2 &= 0 \\
\Delta \hat{G}_2 + \Delta \hat{G}_7 - \Delta \hat{G}_9 - \Delta \hat{G}_3 &= 0 \\
\Delta \hat{G}_1 + \Delta \hat{G}_5 - \Delta \hat{G}_8 - \Delta \hat{G}_3 &= 0 \\
\Delta \hat{G}_8 + \Delta \hat{G}_{12} - \Delta \hat{G}_{11} - \Delta \hat{G}_9 &= 0 \\
\Delta \hat{G}_7 + \Delta \hat{G}_{11} - \Delta \hat{G}_{10} - \Delta \hat{G}_6 &= 0 \quad (C.2)
\end{align*}$$

where the indices refer to Figure 4.2 of the main text. With $\sigma_i$ we denote the (unknown) standard deviation associated with each free-enthalpy difference. The optimization problem was solved under the assumption of one identical $\sigma_i$-value for all free-enthalpy differences in Figure 4.2. Then an estimate $\hat{\sigma}$ for a cycle can be obtained from

$$\hat{\sigma} = \sum_{i \in \text{cycle}} \frac{\Delta G_i}{2}. \quad (C.3)$$

In case several cycles share an edge the largest $\hat{\sigma}$ is reported in Table 4.1 of the main text.
Supplementary material for chapter 5

Relevant nonbonded interaction parameters, covalent interaction types, EDS reference-state parameters, average of the derivatives of the Hamiltonian, energy difference distributions, and nonbonded solute-solute plus solute-solvent energy distributions of the EDS end states.
Table D.1: Relevant non-bonded interaction parameters\(^a\) used for the $-\text{CH}_2-\text{O}-\text{H}$ headgroup in alcohols (A), the CH\(_x\) groups in aliphatic hydrocarbon chains (B) and the ether function in 1,2-dimethoxyethane (C), in the 54A7\(^{47,48}\) and 53A\(_6\text{OXY}\)\(^{60}/53A\(_6\text{OXY+D}\)\(^{116}\) versions of the GROMOS force field. For each atom type, denoted by the atom type code IAC, they include the Lennard-Jones (LJ) dispersion parameter, \(C_6^{1/2}\), repulsion parameter for non-hydrogen-bonding interactions, \(C_{12}^{1/2}\), and repulsion parameter for hydrogen-bonding interactions, \(C_{12,II}^{1/2}\), as well as the atomic partial charges \(q\). The functional forms of the corresponding terms and the applied (geometric-mean) combination rule for LJ parameters are described in Refs 39,571. For alcohols, 54A7 and 53A\(_6\text{OXY}\) differ in the LJ parameters and charges for the atom type OA and in the charges for the atom types H and CH\(_2\), connected to OA. For alkanes the two versions are identical. For ethers, 54A7 and 53A\(_6\text{OXY}\) differ in the LJ parameters and charges for the atom type OE and in the charges for the atom types CH\(_3\) and CH\(_2\) connected to OE. 53A\(_6\text{OXY}\) and 53A\(_6\text{OXY+D}\) differ only in the torsional-energy parameters for the dihedral types OCCO and CCOC, which are reported in Figure D.1.

<table>
<thead>
<tr>
<th>group</th>
<th>IAC</th>
<th>type</th>
<th>(C_6^{1/2})</th>
<th>(C_{12}^{1/2})</th>
<th>(q)</th>
<th>(C_6^{1/2})</th>
<th>(C_{12}^{1/2})</th>
<th>(q)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21</td>
<td>H</td>
<td>–</td>
<td>–</td>
<td>0.408</td>
<td>–</td>
<td>–</td>
<td>0.410</td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>OA</td>
<td>0.04756</td>
<td>1.100</td>
<td>1.227</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>CH(_2)</td>
<td>0.08642</td>
<td>5.828</td>
<td>–</td>
<td>0.266</td>
<td>0.08642</td>
<td>5.828</td>
</tr>
<tr>
<td>B</td>
<td>16</td>
<td>CH(_3)</td>
<td>0.09805</td>
<td>5.162</td>
<td>–</td>
<td>0.000</td>
<td>0.09805</td>
<td>5.162</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>CH(_2)</td>
<td>0.08642</td>
<td>5.828</td>
<td>–</td>
<td>0.000</td>
<td>0.08642</td>
<td>5.828</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>CH(_3)</td>
<td>0.09805</td>
<td>5.162</td>
<td>–</td>
<td>0.162</td>
<td>0.09805</td>
<td>5.162</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>OE</td>
<td>0.04756</td>
<td>1.100</td>
<td>1.227</td>
<td>–0.324</td>
<td>0.04123</td>
<td>1.5297</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>CH(_2)</td>
<td>0.08642</td>
<td>5.828</td>
<td>–</td>
<td>0.162</td>
<td>0.08642</td>
<td>5.828</td>
</tr>
</tbody>
</table>

\(^a\)The units of \(C_6^{1/2}\), \(C_{12}^{1/2}\) and \(q\) are (kJ mol\(^{-1}\)nm\(^6\))\(^{1/2}\), \((10^{-3}\text{[kJ mol}^{-1}\text{nm}^{12}\text{]}^{1/2}\) and (\(e\)).
Figure D.1: Covalent interaction types (bond stretching, bond-angle bending and dihedral-angle torsion) used in the present work. The numbers refer to the type codes of the 54A7 parameter set, except for the dihedral-angle types of 1,2-dimethoxyethane which are defined in the 53A6$_{OXY+D}$ parameter set (see Table IV in Ref. 116).
Table D.2: EDS reference-state parameters $s$ and $E_B^R$ (kJ mol$^{-1}$) used in the twin-system EDS production simulations to obtain free enthalpy differences between hexane$^a_{(+D)}$ and compounds X in dry octanol or wet octanol and water. $E_A^R$ was fixed to zero in all cases.

<table>
<thead>
<tr>
<th>Compound X</th>
<th>$s$</th>
<th>$E_B^R$</th>
<th>$s$</th>
<th>$E_B^R$</th>
<th>$s$</th>
<th>$E_B^R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>butane</td>
<td>0.25</td>
<td>7.0</td>
<td>0.25</td>
<td>8.0</td>
<td>0.25</td>
<td>8.0</td>
</tr>
<tr>
<td>pentane</td>
<td>0.40</td>
<td>3.0</td>
<td>0.40</td>
<td>3.0</td>
<td>0.40</td>
<td>2.0</td>
</tr>
<tr>
<td>heptane</td>
<td>0.60</td>
<td>−3.0</td>
<td>0.60</td>
<td>−3.0</td>
<td>0.60</td>
<td>−4.0</td>
</tr>
<tr>
<td>octane</td>
<td>0.25</td>
<td>−7.0</td>
<td>0.25</td>
<td>−7.0</td>
<td>0.25</td>
<td>−7.0</td>
</tr>
<tr>
<td>hexanol</td>
<td>0.08</td>
<td>11.0</td>
<td>0.085</td>
<td>11.0</td>
<td>0.085</td>
<td>14.0</td>
</tr>
<tr>
<td>1,2-dimethoxyethane$^b$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.128</td>
<td>26.6</td>
</tr>
</tbody>
</table>

$^a$ the +D version applies only for 1,2-dimethoxyethane.

$^b$ for this compound an additional calculation considering an octanol-water mixture with a water mole fraction of 0.27 (experimental saturation concentration) instead of 0.16 was carried out using EDS parameters of $s = 0.128$ and $E_B^R = 26.0$ kJ mol$^{-1}$. 
Figure D.2: Average of the derivative of the Hamiltonian with respect to the coupling parameter $\lambda$ as a function of $\lambda$ for the process of gradually activating the solute-solvent interactions from no interactions at $\lambda = 0$ to full interactions at $\lambda = 1$. The upper panel summarizes the results obtained for hexane in different solvents, the lower panel summarizes the results obtained for hexanol.
Figure D.3: Average of the derivative of the Hamiltonian with respect to the coupling parameter $\lambda$ as a function of $\lambda$ for the process of gradually activating the solute-solvent interactions of hexane with modified torsional-energy parameters$^{116}$ from no interactions at $\lambda = 0$ to full interactions at $\lambda = 1$. The solvents employed were SPC water and wet octanol.
Figure D.4: Left: Energy difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\tilde{\rho}_A(\Delta V_{BA})$ (red), $\tilde{\rho}_B(\Delta V_{BA})$ (green), as obtained from twin-system EDS simulations. State B represents the combined Hamiltonians of solute X solvated in dry octanol ($X = \text{butane, pentane, heptane, octane, hexanol}$; from top to bottom) and solute M solvated in water ($M = \text{hexane}$). State A represents the combined Hamiltonians of solute X solvated in water and solute M solvated in dry octanol. The simulations were performed with the 54A7\textsuperscript{47,48} force field. The energy difference distributions of the end states were determined by reweighting. Right: Nonbonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained \textit{via} reweighting from EDS reference-state simulations ($\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$, red and green) and from independent MD simulations of the end states ($\rho_A(V_A)$ and $\rho_B(V_B)$, blue and orange).
Figure D.5: Energy difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\tilde{\rho}_A(\Delta V_{BA})$ (red), $\tilde{\rho}_B(\Delta V_{BA})$ (green), as obtained from twin-system EDS simulations. State B represents the combined Hamiltonians of solute X solvated in wet octanol ($X = \text{butane, pentane, heptane, octane, hexanol;}$ from top to bottom) and solute M solvated in water ($M = \text{hexane}$). State A represents the combined Hamiltonians of solute X solvated in water and solute M solvated in wet octanol. The simulations were performed with the 54A7$^{47,48}$ force field. The energy difference distributions of the end states were determined by reweighting. Right: Nonbonded solute-solute plus solute-solvent energy distributions of the EDS end states obtained via reweighting from EDS reference-state simulations ($\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$, red and green) and from independent MD simulations of the end states ($\rho_A(V_A)$ and $\rho_B(V_B)$, blue and orange).
Bibliography


198 Bibliography


Bibliography

1117–1125.
[60] B. A. C. Horta, P. F. J. Fuchs, W. F. van Gunsteren, P. H. Hünenberger, New interaction parameters for oxygen compounds in the


[97] A. Warshel, Consistent force field calculations. II. Crystal structures,


[109] S. J. Bachmann, J. Dolenc, W. F. van Gunsteren, On the use of one-step perturbation to investigate the dependence of different properties


[122] L. Yang, A. Ahmed, S. I. Sandler, Comparison of two simulation methods to compute solvation free energies and partition coefficients, *J.
[136] N. Bhatnagar, G. Kamath, J. J. Potoff, Prediction of 1-octanol-water and air-water partition coefficients for nitro-aromatic compounds from


[174] M. M. Reif, P. H. Hünenberger, Computation of methodology-independent single-ion solvation properties from molecular simulations.


[186] A. Saxena, D. Sept, Multisite ion models that improve coordination and


[189] M. M. Reif, C. Oostenbrink, Personal communication.


[201] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular re-
[215] Y. Westermaier, R. E. Hubbard, Free energy methods in ligand design, in De novo Molecular Design, G. Schneider, ed., Wiley-VCH, Wein-


[228] R. Hu, F. Barbault, F. Maurel, M. Delamar, R. Zhang, Molecular dy-


[252] M. R. Shirts, J. D. Chodera, Statistically optimal analysis of samples


[266] D. L. Mobley, J. D. Chodera, K. A. Dill, Confine-and-release method: Obtaining correct binding free energies in the presence of protein con-


[280] H. Luo, K. Sharp, On the calculation of absolute macromolecular bind-


1081–1095.
[334] S. J. Irudayam, R. H. Henchman, Prediction and interpretation of


[378] S. J. Stuart, B. J. Berne, Effects of polarizability on the hydration of


[422] C. J. Woods, M. A. King, J. W. Essex, Replica-Exchange-Based Free-


[447] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, Comparison of four methods to compute the dielectric permittivity of liquids from mo-


[460] J. Dolenc, R. Baron, C. Oostenbrink, J. Koller, W. F. van Gunsteren, Configurational entropy change of netropsin and distamycin upon DNA


H. Li, M. Fajer, W. Yang, Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: A general method for molecular mechanical, quantum mech-


1977.


[543] J. Gmehling, Present status of group-contribution methods for the syn-


Published work


Lebenslauf

Persönliche Daten
Name Niels Hansen
Geburtsdatum 22.07.1977
Geburtsort Hamburg

Schulausbildung
1984 – 1988 Gorch-Fock Grundschule Blankenese
1988 – 1997 Gymnasium Blankenese
06/1997 Abitur

Zivildienst

Studium
01/2002 – 06/2002 Chemical Engineering an Royal Institute of Technology, Stockholm
04/2004 Abschluss: Diplom-Ingenieur

Wissenschaftliche Tätigkeit
seit 01/2014 Juniorprofessor an der Universität Stuttgart
03/2010 – 12/2013 Postdoktorand an der ETH Zürich, Laboratorium für Physikalische Chemie
06/2004 – 01/2010 Wissenschaftlicher Mitarbeiter an der TU Hamburg-Harburg, Institut für Chemische Reaktionstechnik
08/2006 – 12/2006 & Drei Forschungsaufenthalte an der University of California, Berkeley
08/2007 – 12/2007 &
09/2008 – 12/2008