



# Correcting for the free energy costs of bond or angle constraints in molecular dynamics simulations<sup>☆</sup>

Gerhard König<sup>\*</sup>, Bernard R. Brooks

Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, United States



## ARTICLE INFO

### Article history:

Received 30 June 2014

Received in revised form 28 August 2014

Accepted 1 September 2014

Available online 16 September 2014

### Keywords:

Free energy calculation

Constraint correction

Molecular dynamics simulation

Normal mode analysis

Bennett's acceptance ratio

## ABSTRACT

**Background:** Free energy simulations are an important tool in the arsenal of computational biophysics, allowing the calculation of thermodynamic properties of binding or enzymatic reactions. This paper introduces methods to increase the accuracy and precision of free energy calculations by calculating the free energy costs of constraints during post-processing. The primary purpose of employing constraints for these free energy methods is to increase the phase space overlap between ensembles, which is required for accuracy and convergence.

**Methods:** The free energy costs of applying or removing constraints are calculated as additional explicit steps in the free energy cycle. The new techniques focus on hard degrees of freedom and use both gradients and Hessian estimation. Enthalpy, vibrational entropy, and Jacobian free energy terms are considered.

**Results:** We demonstrate the utility of this method with simple classical systems involving harmonic and anharmonic oscillators, four-atomic benchmark systems, an alchemical mutation of ethane to methanol, and free energy simulations between alanine and serine. The errors for the analytical test cases are all below 0.0007 kcal/mol, and the accuracy of the free energy results of ethane to methanol is improved from 0.15 to 0.04 kcal/mol. For the alanine to serine case, the phase space overlaps of the unconstrained simulations range between 0.15 and 0.9%. The introduction of constraints increases the overlap up to 2.05%. On average, the overlap increases by 94% relative to the unconstrained value and precision is doubled.

**Conclusions:** The approach reduces errors arising from constraints by about an order of magnitude. Free energy simulations benefit from the use of constraints through enhanced convergence and higher precision.

**General significance:** The primary utility of this approach is to calculate free energies for systems with disparate energy surfaces and bonded terms, especially in multi-scale molecular mechanics/quantum mechanics simulations. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

Published by Elsevier B.V.

## 1. Introduction

Constraints are used in most molecular dynamics simulations, since the maximum length of the time step for integrating the Newtonian equations of motion is restricted by the frequency of the fastest motions in the system. Imposing constraints that remove the associated rapid vibrational modes makes those degrees of freedom rigid. Thus, it is possible to use longer time steps without losing the conservation of energy or significantly distorting the desired ensemble. Thus, bond constraints can reduce the required computer time by a factor of three [1]. Furthermore, constraints can be a very valuable tool to improve the convergence and efficiency of free energy simulations, e.g., when using the Simplified Confinement [2] or Confine-and-Release methods [3].

Several algorithms are available to impose constraints in molecular dynamics simulations [4]. Among the most widely used methods is

the simple two body constraint SHAKE [5]. SHAKE constrains bonds and angles by finding forces that maintain the right geometry with the iterative Gauss–Seidel method. However, constraining too many bond distances that are coupled to each other through SHAKE is not practical due to the recursive nature of this method (e.g., ring structures). Prominent alternative approaches include the pairwise Rattle algorithm [6], which also includes the velocities of the atoms, or the LINCS algorithm [7]. Recently, the Shape rigid body integrator [8] has been developed that combines the accuracy of SHAKE with the possibility to use higher numbers and more types of constraints (i.e., rigid bodies of three or more centers, thus including bonds, angles and dihedrals).

Notably, imposing constraints can affect the outcome of the simulation by restricting the phase space of specified degrees of freedom to a choice selected by the user. Van Gunsteren reported quenching of dihedral angle transitions if both bonds and angles are constrained [9,10]. Furthermore, the efficiency of constrained simulations is lowered by angle constraints [1]. Toxvaerd pointed out that angle constraints change the trans-gauche transition rate of decane drastically [11]. Tobias and Brooks [12] determined that constraints shift the frequencies of the normal modes in biomolecules, but only in the region between

<sup>☆</sup> This article is part of a Special Issue entitled Recent developments of molecular dynamics.

<sup>\*</sup> Corresponding author at: NIH, NHLBI, Laboratory of Computational Biology, 5635 Fishers Lane T-900 Suite, Rockville, MD 20852.

E-mail address: [gerhard.koenig@nih.gov](mailto:gerhard.koenig@nih.gov) (G. König).

100 and 1400 cm<sup>-1</sup>. Similarly, Hinsen and Kneller found that the dynamics of polypeptides are slightly perturbed by constraints [13]. Recently, Echenique et al. [14] carefully analyzed and quantified the difference between unconstrained models and several types of constraints. Moreover, deviations between 0.2 and 0.5 kcal/mol were found in solvation free energy calculations when bond length constraints were employed [15]. Deviations of such magnitude highlight the need to properly account for free energy changes due to constraints in free energy simulations.

A number of publications have focused on methods for calculating constraint corrections [16–19]; good overviews are provided by Boresch and Karplus [20] or Wang and Hermans [21]. The most prominent approach for such corrections is the use of the average constraint force [22]. However, this approach only seems to be valid for bond length constraints [20]. The potential of mean force corrections of Pearlman and Kollman [23,24] compute the contribution to the free energy for systems with constrained bond terms. However, this requires additional simulations. Furthermore, most of the currently existing constraint corrections have been developed for free energy calculations with Thermodynamic Integration [25], and, therefore, cannot be used straightforwardly in the increasingly popular Bennett's acceptance ratio (BAR) [26], multistate-BAR (MBAR) [27], or Non-Boltzmann Bennett (NBB) [28,29] methods.

For free energy simulations with BAR, the primary utility of including constraints is to increase the phase space overlap between two distinct energy surfaces. Since the variance of the free energy estimate is directly linked to the phase space overlap, constraints allow free energy simulations to be more accurate and to converge more quickly (thus saving computational time). This is of particular utility for coupling disparate energy surfaces, such as QM/MM with MM, where small changes in bond lengths and angles can significantly decrease the necessary overlap needed for an accurate result. This increased accuracy and faster convergence rate offsets the costs associated with post-processing constraints. Here, we attempt to provide a simple framework to correct for the effect of constraints in free energy calculations with BAR by postprocessing the involved trajectories. In particular, we achieve this by explicitly calculating the free energy costs of adding and removing constraints as additional steps in the free energy cycle. We achieve this by employing a gradient calculation combined with normal mode analysis [30] to approximate the contributions to the partition function associated with the constrained degrees of freedom. Notably, this approach is also compatible with most quantum-packages and QM/MM [31–37].

Our approach is inspired by Gō and Scheraga [38], who showed that bond lengths and angles can be treated as functions of the dihedral angles. This implies that the optimal position of a bond or angle at each frame of a trajectory can be determined after the simulation has been conducted with an suboptimal configuration of that particular bond or angle. In particular, they regarded proteins to be a collection of independent normal modes. Some of the normal modes can be considered hard if their force constant is high enough, while all other normal modes are considered soft. The hard normal modes tend to be insensitive to the changes in conformation, so they can be regarded as functions of the soft variables (changing the bond length and bond angles after the conformation changes). This approximation is of course only valid close to the equilibrium conformation. Using this procedure, the hard modes can be considered to be frozen (constrained) during the dynamics, while the soft modes are treated classically to approximate the partition function. Thus, to account for the changes of the hard degrees of freedom, it is possible to use a post-processing step that changes the bond lengths and angles according to their (soft) environment, and the free energy decrease associated with that change.

The remainder of this paper is organized as follows. First, we outline how to calculate the free energy costs of bond and angle constraints (referred to as “constraint correction”), starting with a simple harmonic oscillator, and progressing to the analysis of trajectories in free energy simulations with multiple constraints (Section 2). Methodological

details of the benchmark systems and simulations are presented in Section 3. We then present the results for harmonic and anharmonic oscillators, four-atomic benchmark systems, water boxes of varying size, an alchemical mutation of ethane to methanol, and free energy simulations between alanine and serine (Section 4, see Fig. 1). We conclude with a short discussion concerning the practical advantages of using constraints in free energy calculations in Section 5. A description of the implementation of the constraint correction in CHARMM, as well as a discussion of alternative ways to approximate the Hessian can be found in the Supplementary material.

## 2. Theory

To illustrate the problem of calculating the free energy of releasing a constraint, we start with a simple classical harmonic oscillator with an internal coordinate  $q$ . For this system, the potential energy  $U$  is given by

$$U(\Delta q) = U_0 + \frac{K}{2} \Delta q^2, \quad (1)$$

where  $\Delta q$  is the deviation from the energy minimum,  $U_0$  is the zero point energy that depends on the reference point and  $K$  is the force constant of the bond or angle. Given this potential energy function, the partition function  $Z$  of the harmonic oscillator ( $Z^{h.o.}$ ) can be obtained by integrating over the single degree of freedom,  $q$ ,

$$Z^{h.o.} = \int e^{-\beta U(\Delta q)} d\Delta q = e^{-\beta U_0} \int e^{-\beta \frac{K}{2} \Delta q^2} d\Delta q. \quad (2)$$

The absolute free energy of the harmonic oscillator is then

$$G^{h.o.} = -\beta^{-1} \ln Z^{h.o.} = U_0 - \beta^{-1} \ln \int e^{-\beta \frac{K}{2} \Delta q^2} d\Delta q \quad (3)$$

$$= U_0 - \beta^{-1} \ln \sqrt{\frac{2\pi}{\beta K}}, \quad (4)$$

where we have made use of the Gaussian integral.

If we introduce a constraint in the harmonic oscillator, we completely remove the degree of freedom  $q$ , so the integral in the partition function becomes a single point, corresponding to a Dirac delta  $\delta(x)$ . Making use of the integral of the Dirac delta,  $\int \delta(x) dx = 1$ , the partition function of the constrained system becomes

$$Z_{cons}^{h.o.} = e^{-\beta U(\Delta q_{cons})}, \quad (5)$$

and, correspondingly, the free energy is reduced to

$$G_{cons}^{h.o.} = -\beta^{-1} \ln Z_{cons}^{h.o.} = U(\Delta q_{cons}). \quad (6)$$

Accordingly the free energy of imposing a constraint corresponds to

$$\Delta G_{cons}^{h.o.} = G_{cons}^{h.o.} - G^{h.o.} = U(\Delta q_{cons}) - U_0 + \beta^{-1} \ln \sqrt{\frac{2\pi}{\beta K}}. \quad (7)$$

For convenience, it is possible to divide the total energy into two contributions. First, the enthalpic contribution  $\Delta H$ ,

$$\Delta H^{h.o.} = U(\Delta q_{cons}) - U_0, \quad (8)$$

which is independent of the temperature. Notably,  $U(\Delta q_{cons})$  inherently contains  $U_0$ .

Second, the entropic contribution  $\Delta G_{\text{harm}}$  from the vibrations of the harmonic potential,

$$\Delta G_{\text{harm}}^{\text{h.o.}} = \beta^{-1} \ln \sqrt{\frac{2\pi}{\beta K}}. \quad (9)$$

The total free energy of imposing the constraint is just

$$\Delta G_{\text{cons}}^{\text{h.o.}} = \Delta H^{\text{h.o.}} + \Delta G_{\text{harm}}^{\text{h.o.}}. \quad (10)$$

## 2.1. Implementation using normal mode analysis

The above analysis is only valid if the harmonic oscillator is one-dimensional and isolated (i.e., without non-bonded interactions). In the presence of external forces (e.g., by adding Coulomb charges and Lennard Jones parameters to the oscillator and immersing it into a water box), the system becomes an anharmonic oscillator, with an unknown potential energy function. Starting from an arbitrary constrained structure ( $\Delta q_{\text{cons}}$ , where the  $\Delta$  refers to the deviation from the energy minimum) of the anharmonic oscillator, we can use a Taylor expansion to describe the potential energy function of the unconstrained system.

$$U(\Delta q) = U(\Delta q_{\text{cons}}) + U'(\Delta q_{\text{cons}})(\Delta q - \Delta q_{\text{cons}}) + \frac{U''(\Delta q_{\text{cons}})}{2}(\Delta q - \Delta q_{\text{cons}})^2 + \dots \quad (11)$$

where  $U'(\Delta q_{\text{cons}})$  and  $U''(\Delta q_{\text{cons}})$  denote the first and second derivatives of the potential energy function evaluated at the constrained conformation. Since most classical molecular simulations of biomolecules are conducted at room temperature in a stable chemical environment, the Taylor series is dominated by the harmonic term. This implies that the chemical bond is not going to break and non-bonded interactions are mostly going to shift the energy minimum, but not the frequency. We, therefore, approximate the potential energy function by ignoring the terms with powers of three or higher, i.e.

$$U(\Delta q) \approx U(\Delta q_{\text{cons}}) + U'(\Delta q_{\text{cons}})(\Delta q - \Delta q_{\text{cons}}) + \frac{U''(\Delta q_{\text{cons}})}{2}(\Delta q - \Delta q_{\text{cons}})^2. \quad (12)$$

Notably, this approximation ignores all anharmonic contributions to the potential energy.

To identify the energy minimum of the unconstrained system, we can use  $U'(\Delta q_{\text{cons}})$  and  $U''(\Delta q_{\text{cons}})$  of the constrained structure in a one-step Newton–Raphson minimization. Since the first derivative is zero at the energy minimum,

$$\Delta q_{\text{cons}} \approx \frac{U'(\Delta q_{\text{cons}})}{U''(\Delta q_{\text{cons}})}. \quad (13)$$

The Newton–Raphson step leads to an approximation of the enthalpic contribution in anharmonic oscillators, making use of Eq. (1),

$$\Delta H^{\text{a.o.}} \approx \frac{U'(\Delta q_{\text{cons}})^2}{2U''(\Delta q_{\text{cons}})}. \quad (14)$$

Correspondingly, if the constrained structure is close to the energy minimum, the harmonic entropy contribution can be approximated by

$$\Delta G_{\text{harm}}^{\text{a.o.}} \approx \beta^{-1} \ln \sqrt{\frac{2\pi}{\beta U''(\Delta q_{\text{cons}})}}. \quad (15)$$

However, if  $\Delta q_{\text{cons}}$  is relatively large, it might be advisable to reevaluate  $U''$  at the energy minimum to make sure that the frequencies are correct.

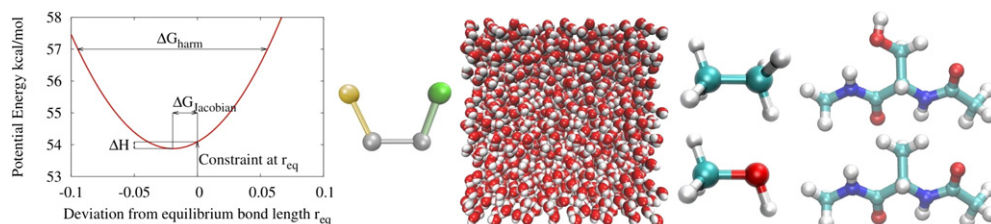
In biomolecules, multiple constraints are used simultaneously. It is, therefore, convenient to reformulate the above analysis in terms of a reduced basis harmonic analysis. This approach relies on partitioning the full Hessian matrix of the whole system ( $3n \times 3n$ , where  $n$  is the number of atoms) into relevant and irrelevant parts, approximately block diagonalizing the Hessian. This assumes that the parts can be separated without incurring too large errors from the missing coupling. However, the influence of soft degrees of freedom on hard degrees of freedom is considered to be small and most hard terms are not coupled to each other (except for the Urey–Bradley term in bond angles).

The “relevant” part in this application is the constrained degrees of freedom. All other degrees of freedom need to be fixed during this analysis to make sure that the ensemble does not get perturbed. If there are  $m$  constrained degrees of freedom that need to be corrected, they form an orthonormal (sub-)basis set  $C$  that consists of  $m$  vectors  $\mathbf{c}$ . Each vector must be normalized in root-mass-weighted Cartesian bases and must be defined in terms of Cartesian displacements. For a constrained bond  $r_i$ , the associated basis vector is defined by  $\mathbf{c}_i = \frac{\partial \mathbf{x}}{\partial r_i}$ , where  $\mathbf{x}$  are the  $3n$  Cartesian coordinates of all atoms. The basis vector corresponds to a vector from the Cartesian coordinates of the first atom to the second atom involved in the bond and vice versa.

For a constrained angle  $\theta_i$  between three atoms  $j$ ,  $k$ , and  $l$  (the constraint being imposed on the coordinates of atom  $l$ ), the vector is  $\mathbf{c}_i = \frac{\partial \mathbf{x}}{\partial \theta_i}$ , corresponding to the tangential vector (i.e. normal to the bond  $r_{kl}$ ) on the plane defined by the three atoms  $j$ ,  $k$ , and  $l$ . Notably, a displacement along such a vector changes the bond length  $r_{kl}$ . However, this can be accounted for by including the curvature of the angle in the form of the vector along  $r_{kl}$  according to  $\Delta \mathbf{x} = \sin(\Delta \theta_i) \mathbf{c}_i r_{kl} + (1 - \cos(\Delta \theta_i)) \mathbf{r}_{lk}$ , where  $\mathbf{r}_{lk}$  is the Cartesian vector between atoms  $l$  and  $k$  (this approach is implemented in the new CANG modes, which are discussed in the Supplementary Material).

The corresponding reduced Hessian  $H$  for the constrained degrees of freedom (an  $m \times m$  matrix) is generated by projecting the matrix of the mass weighted energy second derivatives  $F$  into the  $3n \times m$  reduced basis set  $C$

$$H = C^\dagger F C. \quad (16)$$



**Fig. 1.** Benchmark systems considered in this paper: Anharmonic oscillators, four-atomic benchmark systems, water boxes of varying size, an alchemical mutation of ethane to methanol, and free energy simulations between alanine and serine.

Similarly, the force vector for the reduced basis set,  $\mathbf{g}$  (for gradient), is generated from the force vector of the full system  $\mathbf{f}$  by projection into the basis set,

$$\mathbf{g} = \mathbf{C}^\dagger \mathbf{f}. \quad (17)$$

Due to the coupling between the constrained degrees of freedom (in the form of off-diagonal elements in  $H$ ), the reduced Hessian cannot be used directly in Eqs. (14) and (15). However, using an eigendecomposition with a unitary matrix  $U$ , we obtain

$$\Lambda = U^\dagger H U, \quad (18)$$

which is a diagonal matrix that contains the eigenvalues, and, correspondingly, the force constants (and frequencies) of the constrained degrees of freedom. In CHARMM, the above analysis can be conducted in two different ways. a) Using the REDUCE command of VIBRAN, which corresponds to using the full  $3N \times m$  matrix  $\mathbf{C}$  to evaluate the reduced Hessian and eigendecomposition. b) Employing the RAYLeigh command, which conducts the above analysis separately for each vector  $\mathbf{c}$  of the basis set (thus neglecting the coupling within the constrained degrees of freedom).

From here, it is straightforward to rewrite the equivalent correction terms for enthalpy and entropy of Eqs. (14) and (15) with

$$\Delta H \approx \frac{1}{2} \|\mathbf{g}^\dagger \Lambda^{-1} \mathbf{g}\|_1, \quad (19)$$

and

$$\Delta G_{\text{harm}} \approx \left\| \frac{1}{2\beta} \ln \left( \frac{2\pi}{\beta} \Lambda^{-1} \right) \right\|_1, \quad (20)$$

where  $\|\dots\|_1$  denotes a sum over all elements. Since  $\Lambda$  is a diagonal matrix, applying the logarithm is straightforward and just involves taking the logarithm of each element.

## 2.2. Jacobian factors

The above constraint correction, and thus Eq. (10), only applies to internal coordinates. However, molecular dynamics simulations are performed in Cartesian coordinates. The conversion from internal coordinates to Cartesian coordinates introduces another contribution to the free energy, the so-called Jacobian. It is entropic in nature and arises because longer bond lengths and obtuse angles have more phase space available than short bonds or straight angles. Thus, they are required for both bonds and angles, but only have to be considered for degrees of freedom that change (i.e., the degrees of freedom that are affected by the constraints) [39,20]. The free energy change related to the release of constraints in Cartesian space, therefore, becomes

$$\Delta G_{\text{rls}} = -\Delta G_{\text{cons}} = -\Delta H - \Delta G_{\text{harm}} - \Delta G_{\text{Jacobian}}. \quad (21)$$

For the general case, the change of the Jacobian contribution to the free energy as constraints are released,  $\Delta G_{\text{Jacobian}}$ , can be calculated by

$$\Delta G_{\text{Jacobian}} = -\beta^{-1} \ln \left( \prod_i \frac{J_i^{\text{after}}}{J_i^{\text{before}}} \right), \quad (22)$$

where  $J_i$  is the Jacobian factor for the transformation between internal and Cartesian coordinates of the degree of freedom corresponding to the  $i^{\text{th}}$  constraint. The superscripts “before” and “after” refer to the coordinates before and after applying the correction. The necessary factors  $J$  can be obtained analytically by using the rigid rotor

analysis developed by Herschbach et al. [40], where the Jacobian term for a bond  $r$  between two atoms  $j$  and  $k$  is

$$J_r = r_{jk}^2, \quad (23)$$

and the Jacobian for an angle  $\theta_{jkl}$  between atoms  $j$ ,  $k$ , and  $l$  is

$$J_\theta = \sin \theta_{jkl}, \quad (24)$$

if  $j$ ,  $k$  and  $l$  form a linear chain. If a terminal atom  $k$  branches off from a chain of atoms  $i$ ,  $j$ ,  $\dots$  at atom  $j$ , the Jacobian factor of the angle  $\theta'$  between the bond  $r_{jk}$  and the plane defined by the two bonds  $r_{ij}$  and  $r_{j\ell}$  is defined by

$$J_{\theta'} = \sin^{-1} \theta'_{ijk\ell}. \quad (25)$$

Thus, special care has to be taken when dealing with angle constraints of branching atoms, as the current version of the code only considers angle constraints of linear chains. (Notably, this is not a major restriction since the SHAKE algorithm is usually not applied to angles of branched structures, since they cause poor convergence.)

## 2.3. Application to free energy simulations

Running a free energy simulation with constraints corresponds to exploring phase space for a confined subspace, only considering an infinitesimally slim slice of the phase space available to the degrees of freedom that have been constrained. Accounting for the energetic costs of releasing a constraint thus corresponds to the effect of adding extra dimensions to the constrained free energy result. This problem is similar to the treatment of separate parts of phase space in rotational isomers, developed by Straatsma and McCammon [41].

For this purpose, we consider that the partition function of the unconstrained system  $Z$  to consist of integrals for the constrained degrees of freedom ( $Z_{\text{cons}}$ ) and the additional degrees of freedom that arise from releasing the constraints ( $Z_{\text{add}}$ ). This leads to

$$Z = Z_{\text{cons}} \cdot Z_{\text{add}} = \int_{\text{cons}} \int_{\text{add}} e^{-\beta U(\mathbf{X})} = \int_{\text{cons}} e^{-\beta U_0(\Delta \mathbf{X}_{\text{cons}})} \int_{\text{add}} e^{-\beta U(\Delta \mathbf{X})} \quad (26)$$

$$= \int_{\text{cons}} e^{-\beta U_0(\Delta \mathbf{X}_{\text{cons}})} e^{-\beta \Delta G_{\text{cons}}(\mathbf{X}_{\text{cons}})}, \quad (27)$$

where  $U_0$  is the part of the potential energy that only depends on the unconstrained degrees of freedom, and the free energy of releasing the constraint ( $\Delta G_{\text{cons}}(\mathbf{X}_{\text{cons}})$ ) is calculated for each point in the integral over the constrained degrees of freedom.

The free energy change due to releasing all constraints in a free energy simulation is

$$\begin{aligned} \Delta G_{\text{rls}} &= -\Delta G_{\text{cons}} = \beta^{-1} \ln \frac{Z}{Z_{\text{cons}}} \\ &= \beta^{-1} \ln \frac{\int_{\text{cons}} e^{-\beta U_0(\Delta \mathbf{X}_{\text{cons}})} e^{-\beta \Delta G_{\text{cons}}(\mathbf{X}_{\text{cons}})}}{\int_{\text{cons}} e^{-\beta U_0(\Delta \mathbf{X}_{\text{cons}})}} \end{aligned} \quad (28)$$

$$= \beta^{-1} \ln \langle e^{-\beta \Delta G_{\text{cons}}(\mathbf{X}_{\text{cons}})} \rangle_{\text{cons}}, \quad (29)$$

where  $\langle \rangle_{\text{cons}}$  denotes an ensemble average using constraints. Notably, this result corresponds to the Zwanzig equation [42] (also known as Thermodynamic Perturbation or the exponential formula), which is implemented in most major simulation packages.

Thus, the harmonic analysis corresponds to performing an *a posteriori* sampling of the constrained degrees of freedom. If the target potential is harmonic, it actually performs perfect sampling of that degree of freedom, given a particular set of “soft” coordinates. Therefore, the only errors arise from anharmonicity and from incomplete sampling of the soft degrees of freedom. Those issues are discussed in Sections 4.1 and 4.2.

### 3. Methods

#### 3.1. (An-)harmonic oscillators

Our first model systems are (an-)harmonic oscillators that mimic a chemical bond between two atoms. While the first atom is fixed, the second atom is free to change its bond length and corresponds to a hydrogen atom. For such systems, quasi-analytical results can be computed very easily. We consider three cases a) an isolated harmonic oscillator in gas phase (without interaction partners), b) the harmonic oscillator in the presence of a sodium ion at a fixed distance of 2.5 Å from the initial position of the hydrogen, projected along the bond, and c) the harmonic oscillator in the presence of a chlorine ion at a fixed distance of 2.5 Å from the initial position of the hydrogen, projected along the bond. Cases b) and c) correspond to a transformation into anharmonic oscillators. Notably, the only degree of freedom is the bond length of the hydrogen atom.

Each of the three cases was evaluated with parameters from five different types of hydrogens from the CHARMM general force field [43]: HGA3 (corresponding to a hydrogen in ethane), HGA5 (corresponding to a hydrogen in ethene), HGR61 (corresponding to a hydrogen in benzene), HGP1 (corresponding to the hydrogen of the hydroxyl group in, e.g., methanol), and HGP3 (corresponding to the hydrogen of the thiol group in, e.g., ethanethiol).

Using Eq. (3), it is possible to compute the absolute free energies of each type of harmonic oscillator by numerical integration of the partition function. All integrations were carried out with the NIntegrate function of Mathematica 9.0 with AccuracyGoal set to Infinity. The singularities of the potential energy function that arise when the positions of the hydrogens and the chlorine atom overlap were avoided by integrating only within  $\pm 1.9$  Å of the equilibrium bond length (we tested integration ranges between 1 and 2 Å, but the free energy results were invariant to the choice of this cutoff since the probability of states far away from the equilibrium bond length is smaller than the finite machine precision of the computer representation). The gradients and Hessians for Eqs. (13), (14), and (15) were determined by numerical differentiation with the ND function at the equilibrium bond length specified in the CGenFF parameter set. We used a scale of variation of 0.0001 Å. Hessians were computed with the ND function at the global energy minimum (as determined by the FindMinimum function).

#### 3.2. Four-atomic benchmark systems

The calculations on the four-atomic benchmark system test whether the correction step is capable to account for changes of bond lengths and angles. This benchmark system is very sensitive to deviations from the right ensemble (e.g. by using the wrong thermostat) and has been used extensively in the literature [9,39,20,44,45]. Accurate results can be calculated both analytically and with appropriate molecular dynamics simulations.

The initial state for all corrections had equilibrium bond lengths of 2 Å for all bonds. The force constant for all bonds was 200 kcal/(mol Å<sup>2</sup>). Bond angles were set to 110°, with angle force constants of 50 kcal/(mol rad<sup>2</sup>). The dihedral force constant was 1 kcal/mol with a multiplicity 3.

We computed the free energy differences to end states with the following changes relative to this initial state: a) Changing the first bond length to  $r_{12} = 2.1, 2.25, 2.5$ , and 3 Å, b) changing the first bond angle

to  $\theta_{123} = 111, 115, 120, 135$ , and 160°, and c) changing the bond lengths and angles of the first and last atoms to  $r_{12} = 3$  Å,  $r_{34} = 1$  Å,  $\theta_{123} = 160^\circ$ , and  $\theta_{234} = 50^\circ$ . The corrections were performed with the new OPTI command in the VIBRAN code of CHARMM.

#### 3.3. Water boxes

Five different boxes with TIP3P [46,47] water were prepared. The first consisted only of a water pentamer in gas phase without periodic boundary conditions. For this system, the structure of the global energy minimum [48] was used as the initial state. Water boxes with 395, 787, 1636, and 3290 molecules were generated with CHARMM-GUI [49], using cubic boxes with side lengths of 24.01, 30.25, 38.11, and 48.02 Å, respectively. Since the VIBRAN code does not support Ewald summation, the nonbonded interactions were calculated with a cutoff between 10 and 12 Å, using force-shifting for electrostatic interactions and shifting for Van-der-Waals interactions.

In the first step, SHAKE constraints were imposed according to the equilibrium bond length in the parameter file, and each water box was minimized to the local energy minimum using 100 steps of steepest descent and 2000 steps of Adopted Basis Newton–Raphson minimization. This corresponds to the “uncorrected” structure with constraints and a potential energy  $U_{\text{uncorrected}}$ . Next, the SHAKE constraints were released and the constraint correction was calculated, using all bond lengths and bond angles as the basis set for the harmonic analysis with the CBND and CANG mode types. This leads to the “corrected” structure with a potential energy  $U_{\text{corrected}}$ . To determine the true local energy minimum, the corrected structure was further minimized with 100 steps of steepest descent and 2000 steps Adopted Basis Newton–Raphson minimization, leading to the potential energy of the minimum  $U_{\text{min}}$ .

##### 3.3.1. Ethane–Methanol

Solvation free energy differences between ethane and methanol were calculated as described in Reference [15]. All simulations were conducted with CHARMM [50,51], using the CHARMM22 [47] force field. Following the recommendations by Boresch and Karplus [20,52], a dual topology hybrid scheme was implemented using the MSCALE module [53] of CHARMM. Bond and angle terms were calculated in the main MSCALE process, while the dihedral and non-bonded energy terms were evaluated in two subprocesses: one corresponding to the ethane end state and one corresponding to methanol. The potential energies from the subprocesses were mixed according to  $\lambda$ .

Gas phase simulations were conducted with Langevin dynamics, using a friction coefficient of 5 ps<sup>−1</sup> on all atoms and a temperature of 300 K. The cut-off radius was set to 998 Å. In solution, 862 water molecules and an octahedral box of 32.168 Å were used. The temperature was maintained at about 300 K by a Nosé–Hoover thermostat [54]. Lennard–Jones interactions were switched off between 10 and 12 Å, while electrostatic interactions were computed with the PME method [55]. The time step was 1 fs and simulations with and without SHAKE on all hydrogen atoms were employed.

Free energy differences were calculated with Bennett's acceptance ratio method [26] based on simulations of 5 ns in gas phase and 1 ns in solution. Trajectories were written every 100 steps in gas phase and every 20 steps in solution. For the free energy calculations, five  $\lambda$  points were employed in gas phase (0.00, 0.25, 0.50, 0.75 and 1.00) and eleven in solution (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). The standard deviations of the free energy results were determined by repeating each simulation four times, starting with different initial random velocities.

##### 3.3.2. Alanine–Serine

Free energy differences between N-acetyl-methylamide alanine and serine were calculated in aqueous solution. Both amino acids were simulated using the AMBER Cornell et al. [56] and the CHARMM22 [47] force fields once with SHAKE constraints and once without SHAKE

constraints. For better comparability, the same setup was used as in Reference [57]. 243 TIP3P water molecules were present, and the temperature was maintained at about 300 K by a Nosé-Hoover thermostat [54]. Lennard–Jones interactions were switched off between 9 and 10 Å, while electrostatic interactions were computed with the Particle Mesh Ewald method [55]. The simulation box was a truncated octahedron, cut-out of a cube with side length 21.4 Å. Trajectories were written every 10 steps. The standard deviations of the free energy results were determined by repeating each simulation four times, starting from different initial random velocities. The energies required for BAR were extracted from the trajectories using the EAVG command of the BLOCK module of CHARMM. The constraint corrections were calculated using the VIBRAN module of CHARMM and the analyzed with the Zwanzig equation [42].

## 4. Results and discussion

### 4.1. Analytical results for (an-)harmonic oscillators

The magnitude of  $\Delta G_{\text{cons}}$  depends on multiple factors, including the force constant of the bond or angle, but also non-bonded interactions. To illustrate the dependency on those factors quantitatively, we calculated  $\Delta G_{\text{cons}}$  for three different kinds of oscillators. In each case, a hydrogen atom is attached to a fixed, non-interacting atom at a distance that corresponds to the equilibrium bond length in the parameter set. In the first instance (gas phase), there are no non-bonded interactions that influence the bond. In the second and third datasets, the harmonic oscillator is interacting with a fixed sodium or chloride ion at a distance of 2.5 Å. The two cases illustrate how  $\Delta G_{\text{cons}}$  can be affected by interactions with the environment. To test the influence of the bonded terms on  $\Delta G_{\text{cons}}$ , the benchmark systems were evaluated using parameters from several hydrogen atom types in the CHARMM Generalized Force Field (HGA3, HGA5, HGR61, HGP1, HGP3). The hydrogen types correspond to

**Table 1**

Examples of free energy differences from imposing a constraint for multiple hydrogen types in (an-)harmonic oscillators. The individual contributions from enthalpy, vibrational entropy, the Jacobian factor and anharmonic effects are listed in the four rightmost columns. The free energies are evaluated in gas phase and in the presence of a sodium and a chlorine ion, setting the constraint to the equilibrium bond length from the parameter file. All data are in kcal/mol.

Hydrogen type <sup>a</sup>	$\Delta G_{\text{cons}}$ <sup>b</sup>	$\Delta H^c$	$\Delta G_{\text{harm}}$ <sup>d</sup>	$\Delta G_{\text{Jacobian}}$ <sup>e</sup>	Anharm. error <sup>f</sup>
<i>In gas phase at equilibrium bond length</i>					
HGA3 (ethane)	−1.528	0	−1.528	0	0
HGA5 (ethene)	−1.565	0	−1.565	0	0
HGR61 (benzene)	−1.544	0	−1.544	0	0
HGP1 (methanol)	−1.684	0	−1.684	0	0
HGP3 (ethanethiol)	−1.481	0	−1.481	0	0
<i>Interacting with a sodium ion at 2.5 Å</i>					
HGA3 (ethane)	−1.512	0.017	−1.530	−0.002	0.0001
HGA5 (ethene)	−1.488	0.080	−1.568	−0.004	0.0001
HGR61 (benzene)	−1.519	0.027	−1.546	−0.002	0.0001
HGP1 (methanol)	−1.477	0.211	−1.687	−0.005	0.0003
HGP3 (ethanethiol)	−1.423	0.060	−1.483	−0.004	0.0005
<i>Steric clash with a chlorine ion at 2.5 Å</i>					
HGA3 (ethane)	−1.390	0.179	−1.570	−0.005	0.0000
HGA5 (ethene)	−1.576	0.019	−1.595	−0.001	0.0002
HGR61 (benzene)	−1.362	0.227	−1.589	−0.006	0.0007
HGP1 (methanol)	−1.461	0.218	−1.678	−0.005	0.0004
HGP3 (ethanethiol)	−1.428	0.052	−1.480	+0.003	−0.0001

<sup>a</sup> Atom type in the CHARMM General Force Field.

<sup>b</sup> Free energy difference of releasing the constraint.

<sup>c</sup> Enthalpic contribution from potential energy difference to the energy minimum.

<sup>d</sup> Vibrational entropy determined by force constant.

<sup>e</sup> Jacobian factor that accounts for transformation from internal to Cartesian coordinates.

<sup>f</sup> Anharmonic error that was not accounted for in the harmonic analysis.

hydrogens found in ethane, ethene, benzene, methanol and ethanethiol, thus covering both apolar, aromatic and polar groups that are encountered in biomolecular simulation. Since the only degree of freedom is the bond length of the hydrogen, the free energy can be calculated analytically for those systems. Table 1 lists their  $\Delta G_{\text{cons}}$ , as well as their enthalpic ( $\Delta H$ ) and entropic components ( $\Delta G_{\text{harm}}$ ,  $\Delta G_{\text{Jacobian}}$ ).

In gas phase (top third of Table 1), there are no interactions that might influence the free energy except for the bond stretching term of the harmonic oscillator itself. Since the constraint was introduced at the equilibrium bond length, the system resides at the energy minimum. Therefore there are neither enthalpic contributions nor Jacobian contributions, as the bond length does not change when the constraint is released. The only contributions to  $\Delta G_{\text{cons}}$  arise from the vibrational entropy (fourth column), which is directly related to the force constant of the bond. The weakest force constant is found for the thiol-hydrogen type HGP3 (last row) which, correspondingly, exhibits the lowest entropic penalty of introducing a free bond with 1.481 kcal/mol. The system with the highest vibrational contribution is HGP1, which corresponds to a hydroxyl group. Due to its strong force constant, the potential energy rises faster as the bond length changes, thus limiting the available phase space for the bond more than all the other atom types. Correspondingly,  $\Delta G_{\text{harm}}$  is also an indicator whether external forces are limiting the available phase space of the bond (a larger absolute value of  $\Delta G_{\text{harm}}$  indicating more confinement).

By adding a positively charged sodium ion at a distance of 2.5 Å (middle third of Table 1), the hydrogens are affected by repulsive electrostatic forces, and, depending on the Van-der-Waals radius of they hydrogen type, slightly attractive or repulsive nonpolar interactions. Due to the external forces, the energy minimum of the bond is changed, which results in a change of bond length if the constraint is released. Correspondingly, the enthalpy changes according to the potential energy difference relative to the energy minimum, and a Jacobian term arises due to the change of the bond length. The vibrational entropy  $\Delta G_{\text{harm}}$  is the dominant contribution to  $\Delta G_{\text{cons}}$ . While the results for  $\Delta G_{\text{cons}}$  lie between −1.423 and −1.519 kcal/mol,  $\Delta G_{\text{harm}}$  ranges between −1.48 and −1.68 kcal/mol. The enthalpic contributions from  $\Delta H$  amount to up to 0.211 kcal/mol and the Jacobian factors are below −0.005 kcal/mol.

Based on the values for  $\Delta G_{\text{cons}}$  and its contributions, it might appear tempting to neglect the small contributions from  $\Delta H$  and  $\Delta G_{\text{Jacobian}}$ , since they only represent a few percent of the total free energy change. However, if the  $\Delta G_{\text{cons}}$  in gas phase and in the presence of sodium were to be used in a thermodynamic cycle to calculate the binding affinity of sodium to the harmonic oscillator, the importance of  $\Delta H$  and  $\Delta G_{\text{Jacobian}}$  becomes clearer. In such a situation, the contributions of  $\Delta G_{\text{harm}}$  mostly cancel between gas phase and in the presence of sodium, leaving differences below 0.003 kcal/mol. This is orders of magnitude lower than the contributions from  $\Delta H$  (0.017–0.211 kcal/mol), and roughly in the same range as the Jacobian factor (−0.002 to −0.005 kcal/mol). Therefore, all three factors are relevant for most applications involving free energy simulations, especially when considering that their contributions are going to add up if multiple constraints are present in the simulation.

Placing a chlorine at a distance of 2.5 Å from the hydrogen (bottom third of Table 1) causes a steric clash for most hydrogen types. This leads to larger values of  $\Delta H$  (between 0.052 and 0.383 kcal/mol) since changing the bond length allows the system to relax. Compared to gas phase,  $\Delta G_{\text{harm}}$  increases by about 0.03 to 0.05 kcal/mol, since the repulsion from the chlorine limits the available phase space of the bonds. Markedly, the polar hydrogen types HGP1 and HGP3 are an exception to this trend as they have very small Van-der-Waals radii due to polarization. Since they also have relatively large positive charges, they are attracted by the negatively charged chlorine. This leads to a stretching of the bond, which results in positive Jacobian factors. Notably, the Jacobian contribution  $\Delta G_{\text{Jacobian}}$  can be positive or negative, while  $\Delta H$  and is greater or equal to zero (the potential energy of the constrained

conformation cannot be lower than the energy minimum) and  $\Delta G_{\text{harm}}$  is always negative as no entropy can be gained by introducing a bond.

One of the main assumptions of the method described here is that bond and angle terms can be sufficiently approximated by a harmonic potential. Any anharmonicity will lead directly to errors in the estimate of  $\Delta G_{\text{cons}}$ . For the analytic case studies in this section, it is straightforward to evaluate the error arising from anharmonicity by comparing the results for  $\Delta G_{\text{cons}}$  with the analytical free energies of the constrained and unconstrained systems. The deviations of the harmonic approximation from the analytical results are shown in the rightmost column of Table 1 for the systems in gas phase and in presence of sodium or chlorine. In gas phase, there is no anharmonicity, since the potential energy is fully determined by the harmonic potential of the bond of the harmonic oscillator. In the presence of non-bonded forces, some anharmonicity arises since the Lennard Jones and Coulomb potentials are not harmonic. However, its effect on  $\Delta G_{\text{cons}}$  is negligible, since the changes of the bond length are small. Even in the worst case encountered here (HGR61 with chlorine) the anharmonic contribution is 0.0007 kcal/mol, which is three orders of magnitude smaller than the effect of  $\Delta H$ , and also one order of magnitude smaller than the Jacobian factor. Notably, the anharmonic contribution can be both positive and negative, as demonstrated by the results for HGP3 with chlorine. This indicates the possibility of cancelation of errors if multiple constraints are corrected at the same time. Moreover, since SHAKE constraints are usually applied to hydrogens in biomolecules at room temperature, the errors from the harmonic approximation in the proposed correction will, therefore, not be a limiting factor for the accuracy of free energy calculations.

#### 4.1.1. Four atomic-benchmark systems

In addition to corrections of the bond length, as illustrated in the previous section, the suggested algorithm also allows the correction of

**Table 2**

Mutations of a four atomic benchmark system based on a single structure. The first four columns indicate the magnitude of the change and which bond or angle was perturbed (first bond, last bond, first angle, last angle). The three rightmost columns present the deviation from analytical reference free energies for the uncorrected system, and two different implementations of the correction: one with the standard normal modes in CHARMM (sixth column) and the other the new CBND and CANG normal modes (rightmost column).

Deviation from equilibrium				Deviation from analytical free energy (kcal/mol)		
$\Delta r_{12}^a$	$\Delta r_{34}^b$	$\Delta \phi_{123}^c$	$\Delta \phi_{234}^d$	Uncorrected <sup>e</sup>	NMD <sup>f</sup>	C-NMD <sup>g</sup>
<i>Changes of bond length (Å)</i>						
0.10				2.00E + 00	6.39E − 05	6.91E − 19
0.25				1.25E + 01	2.38E − 03	1.66E − 18
0.50				5.00E + 01	3.50E − 02	5.38E − 19
1.00				2.00E + 02	4.71E − 01	6.04E − 18
<i>Changes of bond angle (°)</i>						
	1			1.52E − 02	6.74E − 06	9.86E − 19
	5			3.81E − 01	3.98E − 03	3.34E − 16
	10			1.52E + 00	5.95E − 02	3.34E − 16
	25			9.52E + 00	1.93E + 00	3.37E − 16
	50			3.81E + 01	2.40E + 01	4.05E − 19
<i>Changing both bond lengths and angles of terminal atoms</i>						
1.00	−1.00	50	−50	4.76E + 02	8.53E + 01	7.37E − 18

<sup>a</sup> Deviation of bond between atoms 1 and 2 from equilibrium bond length in Å.

<sup>b</sup> Deviation of bond between atoms 3 and 4 (Å).

<sup>c</sup> Deviation of bond angle between atoms 1, 2, and 3 from equilibrium in degrees (°).

<sup>d</sup> Deviation of bond angle between atoms 2, 3, and 4 (°).

<sup>e</sup> Original deviation from analytical free energy result.

<sup>f</sup> Deviation from analytical free energy result after a correction step based on normal modes and vibrational entropy.

<sup>g</sup> Deviation from analytical free energy result after a correction step based on CBND and CANG normal modes that account for enthalpy, vibrational entropy, Jacobian contributions and include curvature of bond angles.

bond angles. In order to test the performance of both bond and angle corrections, we employ changes of the two terminal atoms of a linear four-atomic benchmark system. Since no non-bonded interactions are calculated for this benchmark system, it is possible to calculate the free energy analytically using the rigid-rotor assumption. Furthermore, precise reference results are also available from computer simulations reported in the literature [39,45]. The four-atomic benchmark system has proven itself very valuable in the past, as it is very sensitive to even small errors of the thermostat.

Table 2 lists results for three kinds of mutations. In the first section, mutations of the bond length between atoms 1 and 2 are considered. This serves to test the implementation of the correction in CHARMM and also illustrates the dependency on the magnitude of the bond length change. Similarly, the second section tests the accuracy of the bond angle correction for the angle between atoms 1, 2 and 3. It highlights the dependence of the accuracy on the bond angle change. In the last row, we evaluate the performance of the correction if two bond lengths (between atoms 1–2 and 3–4) and two bond angles (between atoms 1–2–3 and 2–3–4) are changed simultaneously to check for possible errors when multiple coupled degrees of freedom are changed.

For each mutation, we consider three cases: a) not using a correction (Uncorrected, fifth column), and b) applying a correction with the OPTI command, but using the standard implementation of normal modes (NMD, sixth column). In this case, the Jacobian term and the curvature of bond angles are not accounted for c) employing the dedicated CBND and CANG normal modes, which save the data required for the correction of the Jacobian factor and for bond angles (C-NMD, rightmost column).

Turning the attention to the first section of Table 2 (changes of bond length), it is possible to see that four different bond length changes are considered: 0.10, 0.25, 0.50, and 1.00 Å. All of these changes are larger than the expected bond length change when releasing SHAKE constraints. They primarily serve to detect even small errors in the implementation. If uncorrected, the resulting errors range between 2 and 200 kcal/mol. By using the OPTI command with conventional normal modes in CHARMM, this error is reduced significantly to between 6.39E−05 and 0.47 kcal/mol. The residual error arises from the missing correction of the Jacobian term. With the dedicated CBND normal modes, the required data for the computation of the Jacobian term are saved. Since there are no non-bonded terms in this benchmark system, there is no anharmonicity, so the accuracy is only limited by the finite machine precision of the computer, leading to errors below 6.04E−18 kcal/mol.

In the second section of Table 2, multiple mutations of the bond angle between the first three atoms of the benchmark system are considered. The angle is changed by between 1 and 50°, leading to errors of up to 38 kcal/mol. When using the standard type of normal modes, only the tangential vector of the bond angle is used for the correction. In the case of large displacements of the angle, the use of only the tangential vector will change the bond length of the terminal atom. Thus, errors of up to 24 kcal/mol can remain after the correction step with conventional normal modes. By using the dedicated CANG normal modes, the curvature of the bond angle is considered when conducting the correction. Thus the bond length is maintained during the correction step, leading to errors below 3.34E−16 kcal/mol. However, in order to be able to obtain such high accuracy it is necessary to correct the bond lengths with CBND before the bond angles are corrected with CANG, otherwise the vectors of the bond lengths are not up to date. Moreover, the correction of bond angles employs trigonometric functions, which can lead to errors if the angle has to be moved to a different quadrant, but such drastic changes are not very likely.

To test if the simultaneous correction of multiple constraints has a negative impact on accuracy, all bond lengths and angles of the terminal atoms were changed by  $\pm 1$  Å and  $\pm 50^\circ$ . This leads to an error of 476 kcal/mol. With the conventional normal modes without Jacobian

and curvature corrections, the error is reduced to 85 kcal/mol. With the CBND and CANG normal mode types, the error is just  $7.37E - 18$ , which arises from finite machine precision. This demonstrates that constraints on bonds and angles can be fully corrected if there is no coupling by non-bonded interactions.

#### 4.2. Impact of soft degrees of freedom

The described correction is only applicable to hard degrees of freedom (bond and angle terms). However, to reach the local energy minimum, usually also a rearrangement of soft degrees of freedom (dihedral angles, translations, rotations) is necessary after the bond and angles terms have been corrected for. Unfortunately, the constraint correction cannot be applied to soft degrees of freedom, since they are not harmonic and strongly depend on external forces. There is no straightforward way to connect them to the original constrained ensemble. Moreover, changing the soft degrees of freedom would introduce a bias, since the resulting conformations would not be generated according to their correct Boltzmann weight. However, since relatively low energies are required to change soft degrees of freedom, the relevant conformations can be sampled even in the presence of constraints, given enough time.

Since the soft degrees cannot be corrected for, it is not possible to fully connect the constrained ensemble with the unconstrained ensemble in one step. There will be a residual error if the correction is applied to systems where a rearrangement is required. Given this fundamental restriction, a correction step can only aim to minimize this residual error. To evaluate the magnitude of the residual error, we apply the constraint correction to water boxes of varying size and compute the difference between the potential energy of the conformation after the correction from the true local energy minimum. Water is a very polar molecule, so its dipole moment is significantly affected by the bond lengths and angles. In turn, the dipole moment affects the orientation of water relative to the other water molecules around it. Since water clusters are arguably among the most flexible complexes imaginable in biochemistry, this benchmark system can be considered very strongly dependent on soft degrees of freedom.

Table 3 lists the errors introduced by stretching the bonds by 0.05 Å and angles by 1° for all TIP3P water molecules in the box. Of course, this benchmark is of no of practical relevance since TIP3P waters are built to be completely rigid, but it is a valuable test since its electrostatic interactions are extremely sensitive to the geometry. The number of water molecules in the box is shown in the first column ( $N_{H_2O}$ ), ranging between 5 and 3290. The errors are presented for the uncorrected

structures (second column) and after the correction (third column). The fourth column (% reduction) indicates the percentage of the error that was corrected for in a single step. The errors before the correction range between 11.2 kcal/mol for a water pentamer and 2256.4 kcal/mol for a box of 3290 water molecules. Given the relatively small perturbations that were employed, this is a rather large change of the potential energy, highlighting the sensitivity of this benchmark. If divided by the number of water molecules, the average error per water molecule is about 0.6 kcal/mol for all box sizes. After one step of correction, those errors are reduced to between 0.2 and 205.9 kcal/mol. As indicated in the fourth column, the percentage of error reduction is consistently around 90% for all cases. This corresponds to a residual error of approximately 0.05 kcal/mol per water molecule, which is an order of magnitude smaller than the uncorrected case, demonstrating that the correction is not perfect, but allows a considerable reduction of the error.

One drawback of the correction approach is the time-consuming computation of the Hessians. The required CPU time for one correction step is indicated in the rightmost column of Table 3 (Time). For a small number of constrained atoms, the computational costs are quite acceptable, but, unfortunately, the costs rise cubically with the number of atoms. Thus, the costs quickly become sizable. While a computation for 395 water molecules (1185 degrees of freedom that have to be corrected) requires 29 s on a single Q9300 2.50 GHz processor, doubling the number of molecules (787 molecules, 2361 degrees of freedom) takes almost 4 min (227 s). Doubling the number of atoms again (1636 molecules, 4908 degrees of freedom) takes half an hour. For more than 3239 molecules, the computation quickly becomes intractable (more than 4.5 h for a single computation, high memory demand).

In the Supplementary material we discuss several alternative approaches to approximate the Hessians. If the Hessian is approximated by the force constant in the parameter set, the correction of 3290 water molecules can be performed in 0.56 s on a single core (instead of 16494.8 s with the calculation of all Hessians). To put this number into perspective, a single energy call on a single core for the same system requires 1.87 s, so increasing the accuracy by employing the correction during a post-processing analysis of a trajectory only takes 30% longer. Furthermore, this process can be easily parallelized, since each frame of a trajectory can be treated independently. If the RAYLeigh command is used to compute the Hessians, also the modes are independent of each other. Thus the proposed correction can make optimal use of modern hardware architectures.

#### 4.3. Application to free energy calculations

We use an alchemical transformation from ethane to methanol to evaluate the performance of the constraint correction in a simple free energy calculation. Very high accuracy can be achieved for this system, and proper conformational sampling can be performed within nanoseconds of simulation time. Thus it has become an informal standard benchmark for free energy methods [58,59,52,60,61,57,15]. The mutation is performed both in gas phase and in explicit solvent, using the standard thermodynamic cycle to calculate solvation free energy differences. The corresponding relative free energies in gas phase ( $\Delta A_{gas}$ ), aqueous solution ( $\Delta A_{H_2O}$ ) and the solvation free energy difference ( $\Delta \Delta A_{solv}$ ) are shown in Table 4. The alchemical mutation was carried out once without SHAKE constraints, using a time step of 1 fs (1.0 fs, first column) and once with SHAKE constraints on all hydrogens in the dual topology setup (also including dummy-hydrogens), using a time step of 1 fs (1.0 fs/SHAKE, second column). The corresponding free energy result that incorporates the constraint correction for all hydrogens in the dual topology setup is shown in the third column (1.0 fs/corr).

While the unconstrained simulations yield a deviation of only 0.04 kcal/mol from experiment ( $-6.93$  kcal/mol) [62], the introduction of SHAKE constraints increases the error to about 0.19 kcal/mol.

**Table 3**

Deviations from the local potential energy minimum of unconstrained water structures with increasing size. All SHAKE constraints were released and only the enthalpy correction ( $\Delta H$ ) was applied. The first column indicates the number of water molecules in the box ( $N_{H_2O}$ ). The second and third columns highlight the deviation from the minimal potential energy before and after the optimization. The error reduction in percent and the required computer time are shown in the two rightmost columns.

$N_{H_2O}$ <sup>a</sup>	Uncorrected <sup>b</sup>	Corrected <sup>c</sup>	% Reduction <sup>d</sup>	Time <sup>e</sup>
5	11.2	0.2	98.2	<0.01
395	263.6	20.0	92.4	29.0
787	538.7	53.9	90.0	227.4
1636	1111.0	105.0	90.5	2038.7
3290	2256.4	205.9	90.9	16494.8

<sup>a</sup> Number of water molecules.

<sup>b</sup> Deviation of the potential energy of the perturbed structure from the potential energy minimum (kcal/mol).

<sup>c</sup> Deviation of the potential energy of the corrected structure from the potential energy minimum after OPTI correction of bonds and angles (kcal/mol).

<sup>d</sup> Error reduction of correction in percent.

<sup>e</sup> Required CPU time for the calculation of Hessians and OPTI command in seconds on a single Q9300 2.50 GHz processor.

**Table 4**

Relative solvation free energy for the mutation of ethane to methanol. Free energy simulations are conducted once without SHAKE, using a time step of 1 fs (1.0 fs, first column) and once with SHAKE, using a time step of 1 fs. The errors from using constraints are corrected for in the third column. All energies are reported in kcal/mol.

	1.0 fs <sup>a</sup>	1.0 fs/SHAKE <sup>b</sup>	1.0 fs/corr <sup>c</sup>	$\Delta A_{\text{cons}}^{\text{Ethane}}$	$\Delta A_{\text{cons}}^{\text{Methanol}}$
$\Delta A_{\text{H}_2\text{O}}^{\text{d}}$	$-0.86 \pm 0.02$	$-0.70 \pm 0.03$	$-0.90 \pm 0.03$	10.972	10.772
$\Delta A_{\text{gas}}^{\text{e}}$	$6.024 \pm 0.006$	$6.039 \pm 0.002$	$6.027 \pm 0.002$	10.96404	10.95180
$\Delta \Delta A_{\text{solv}}^{\text{f}}$	$-6.89 \pm 0.02$	$-6.74 \pm 0.03$	$-6.93 \pm 0.03$	0.008	-0.180
Deviation <sup>g</sup>		0.15	0.04		

<sup>a</sup> Unconstrained simulations with  $\delta t = 1.0$  fs.

<sup>b</sup> Constrained simulations with SHAKE applied to all hydrogens and  $\delta t = 1.0$  fs.

<sup>c</sup> Results from column b after applying the correction for constraints on solute at both end points.

<sup>d</sup> Free energy difference in aqueous solution.

<sup>e</sup> Free energy difference in gas phase.

<sup>f</sup> Solvation free energy difference ( $\Delta A_{\text{H}_2\text{O}} - \Delta A_{\text{gas}}$ ), experimental reference  $-6.93$  kcal/mol – Reference [62].

<sup>g</sup> Deviation from the unconstrained simulation results.

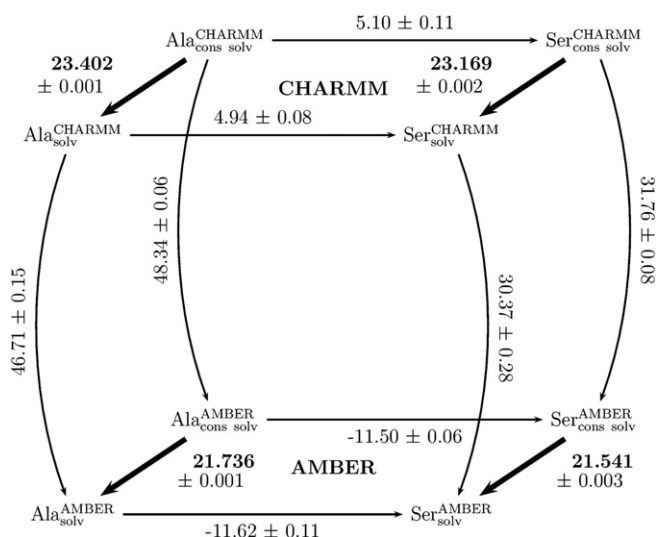
Most of this error (0.18 kcal/mol) arises on the methanol side of the mutation ( $\Delta A_{\text{cons}}^{\text{Methanol}}$  in rightmost column of Table 4) and can probably be attributed to the missing stretching of the hydroxyl bond to accommodate the electrostatic interactions and hydrogen bonds with water. After the constraint correction is applied, the deviation from experiment is reduced, and the deviation from the unconstrained results is only 0.04 kcal/mol. Notably, the difference between unconstrained and corrected results is not statistically significant. This demonstrates that the correction can be a useful tool in free energy calculations, especially if strong non-bonded interactions that might change the bond length during the simulation are present.

#### 4.4. Free energy difference between alanine and serine

Previous work [57] has shown that the transformation of alanine to serine is among the largest alchemical mutations that can be conducted in a single step (i.e., without  $\lambda$  intermediate states). However, the phase space overlap between the end states is only 0.15%, which is critically low. Since the phase space overlap is directly connected to the variance of the free energy estimate in BAR (c.f., equation 11 in Reference [26]), low overlap can lead to poor convergence and imprecise results. As a rule of thumb, it is recommended to ensure a phase space overlap of at least 1% [63,64]. Interestingly, when conducting the same mutation of alanine to serine using the AMBER Cornell et al. force field [56], one obtains a slightly higher phase space overlap of 0.55%. Using BAR, it is possible to compute the free energy differences between CHARMM and AMBER, yielding a phase space overlap of 0.90% for alanine and 0.55% for serine. Notably, all overlaps mentioned so far are below the threshold of 1%, indicating that the convergence of the calculations can be poor. To evaluate the performance of free energy simulations with SHAKE constraints on all hydrogens in the dual topology setup, we repeat the free energy simulations of Reference [57] in aqueous solution once without constraints and once with constraints.

Fig. 2 presents the free energy differences between alanine and serine, using both CHARMM and AMBER, with and without constraints. Results for the CHARMM force field are shown in the top cycle, the AMBER results are listed in the cycle at the bottom. In addition, we used BAR to compute the free energy differences resulting from switching the force fields (vertical arrows). Notably, the large free energy differences of the vertical arrows reflect the different arbitrary zero point references of the two force fields, and are not physically meaningful. The thick arrows indicate constraint corrections. Notably, the unconstrained horizontal free energy differences between alanine and serine are in excellent agreement with previously published results (4.94 and  $-11.62$  kcal/mol versus 4.93 and  $-11.69$  kcal/mol) [57,65].

Focusing on the constraint corrections between back and front plane, it is possible to see that the standard deviations are very low



**Fig. 2.** Free energy differences in aqueous solution between alanine (left) and serine (right), using both the CHARMM (top) and the AMBER (bottom) force fields, and both unconstrained (front) and constrained trajectories with SHAKE on all hydrogens (back). Bold arrows and numbers indicate the use of the constraint correction based on Eq. (29). All other free energy differences were calculated with BAR. All data reported in kcal/mol.

(less than 0.003 kcal/mol). This indicates the fluctuations of  $\Delta G_{\text{cons}}$  are low, and, therefore, the correction converges very quickly even though exponential averaging is used to obtain the free energy costs of the constraints. Due to the low fluctuations, it is probably possible to use fewer data points to correct for the constraints, thus saving computational costs.

The six sides of the cube shown in Fig. 2 represent thermodynamic cycles, and can also be used to gauge the accuracy and precision of our calculations. Since the free energy is a state function, the free energy change associated with closing each thermodynamic cycle should be zero. Any deviations are called cycle closing error and represent a lower bound for non-canceling errors or indicate lack of convergence. Summing all free energy differences around the whole cube leads to a total cycle closing error of 0.22 kcal/mol, which is smaller than the standard deviation obtained by error propagation, 0.32 kcal/mol. This indicates that there is no statistically significant non-canceling error in our calculations. Notably, the cycle closing error of the constrained trajectories is only 0.02 kcal/mol, while the cycle closing error of the unconstrained trajectories is 0.22 kcal/mol. This suggests that the use of constraints can reduce the error and improve convergence. Similarly, the cycle closing error of the alanine trajectories is 0.04 kcal/mol, while the serine trajectories exhibit a cycle closing error of 0.24 kcal/mol. This indicates that the unconstrained serine trajectories are not fully converged, which is also shown by the high standard deviations of the free energy difference between CHARMM and AMBER for serine (0.28 kcal/mol). Contrariwise, the standard deviation of the corresponding constrained free energy simulations is only 0.08 kcal/mol.

With one exception (the mutation of alanine to serine in CHARMM), all constrained free energy results exhibit lower standard deviations than the unconstrained simulations. On average, the standard deviations of the constrained simulations are about 33% lower than the unconstrained simulations. This can probably be attributed to the increased phase space overlap of the constrained simulations. While the overlaps of the unconstrained simulations range between 0.15 and 0.9%, the introduction of constraints increases the overlap up to 2.05%. On average, the overlap increases by 94% relative to the unconstrained value – effectively doubling the magnitude. In free energy simulations with very low overlap, the introduction of constraints, and employing the constraint correction afterwards, can, therefore, be an efficient way to improve convergence and increase precision.

The improvements in precision incur additional computational costs. Evaluating the Hessians at each point of the trajectory is expensive, so the analysis takes about 14 times longer than a standard evaluation of the energies (e.g., 3 h instead of 12.8 min for 60,000 frames of serine in water). However, since only every 100th or 1000th frame of a simulation is saved for analysis, the post-processing costs are dwarfed by the computational expenses of actually performing the simulation itself. Moreover, in contrast to running the simulation without constraints, where each step has to be computed sequentially, these costs of the correction step are trivially parallelized, since each frame is independent of each other. Thus, using constraints actually allows us to transfer some of the computational burden from a sequential scheme (the simulation), to an embarrassingly parallel process (the post-processing). Given modern multi-core architectures, it is more efficient to run the simulation with constraints (achieving a speedup by a factor of three compared to unconstrained simulations) and then use many cores for the post-processing. Furthermore, if computer time is limited, the Hessians can be approximated by simply using the force constants from the parameter file (see Supplementary material). This approach neglects the change of vibrational entropy (which is about one or two orders of magnitude smaller than the enthalpic effect), but is only 40% slower than a normal energy evaluation (17.7 instead of 12.8 min for 60,000 frames of serine in water). These additional costs are negligible considering that four times longer trajectories are required to achieve an equivalent 50% reduction of the standard deviations with unconstrained simulations, since the precision is related to the square root of the number of frames of the trajectory. If the constraints are also used to perform the simulation at a larger time step (e.g., 2 fs instead of 1 fs), the computational costs relative to an equivalent unconstrained free energy simulation are reduced by a factor of eight. Thus, using constraints in free energy simulations, followed by an additional correction step, can lead to a substantial increase of efficiency.

## 5. Conclusions

The present work provides the means to calculate the free energy difference of releasing constraints during post-processing by using harmonic analysis. The approach relies on three assumptions a) that the constrained degrees of freedom can be approximated by a harmonic potential (which is true for bond and angle terms in molecular mechanics) and b) the constrained structure is close to the energy minimum (i.e., the Hessian of the constrained structure is very similar to the Hessian at the energy minimum). This aspect requires the user to select constraints that are close to the equilibrium bond length in the parameter set. For polar groups, such as hydroxyls, it might be necessary to account for bond stretching due to nonbonded interactions with the solvent. c) The coupling of the constrained degrees of freedom with the unconstrained degrees of freedom is small. As shown by van Gunsteren [9,10], angle constraints can change the transition rates of connected dihedral angles, so the angle constraints have to be selected obtuse enough to ensure sufficient sampling of the associated dihedral angles. Furthermore, it is necessary to ensure that the frequencies of the constrained angles are sufficiently separated from the dihedral frequencies.

To demonstrate the usefulness of the algorithm, we tested it with benchmark systems where analytical reference results are available. The maximum free energy error arising from anharmonicity in the anharmonic oscillators was 0.0007 kcal/mol, while accuracy up to machine precision (i.e., errors between  $3.4 \cdot 10^{-16}$  and  $4.1 \cdot 10^{-19}$  kcal/mol) was achieved for the purely harmonic four-atomic benchmarks.

To test the applicability to larger systems, we used simulation boxes with up to 3290 water molecules (9870 constrained degrees of freedom). While the algorithm is able to reduce the deviation from the potential energy minimum induced by constraints by more than 90% in all cases, the computation of the Hessians can become very expensive in terms of CPU time and memory. This problem can be alleviated by directly using the force constant from the parameter set (thus neglecting

changes of the vibrational entropy, which decreases the error reduction only by about 0.1%). This subject is further discussed in the Supplementary material.

The functionality was also evaluated in free energy simulations of the alchemical transformation from ethane to methanol. For the calculation of the solvation free energy difference between ethane and methanol, the correction reduced the deviation from the unconstrained free energy results from 0.15 kcal/mol (when using SHAKE constraints on all hydrogens in the free energy simulations) to 0.04 kcal/mol by accounting for the stretching of the hydroxyl group.

The mutation of alanine to serine highlights that it is possible to use the correction step to calculate free energy differences between different force fields. Notably, the phase space overlaps of the unconstrained simulations ranged between 0.15 and 0.9%, which is critically low and introduces imprecision and systematic errors. The introduction of bond constraints alleviates this problem by practically doubling the overlap to up to 2.05%. As a result, the standard deviations of the constrained simulations are, on average, 50% lower than the unconstrained simulations.

The proposed corrections are particularly useful for increasing the accuracy and efficiency of free energy calculations. By using constraints, the simulations can use larger time steps and the phase space overlap between the states involved in the free energy calculation is increased by the confinement. This allows the user to use fewer  $\lambda$  intermediate steps and shorter simulations. In the context of multiscale free energy calculations [15,66–70], the use of constraints also allows to include more atoms in the quantum region or dealing with large differences between the QM and the MM energy surface. Since the phase space overlap is related to the variance of the free energy estimate [26], the efficiency of free energy simulations can be significantly increased by using constraints.

The proposed procedure corrects for about 90% of the free energy change associated with the introduction of constraints. The main limitation lies in the computational costs of the Hessians. Even though their computation is easily parallelizable, it can become intractable if too many constraints are used in the free energy calculation. However, most applications of free energy simulations are limited to ligands with molecular weights less than 500 Da, so this limitation should not pose a problem in most cases. Furthermore, the expenses can be mitigated by approximating the Hessians by the bond or angle force constants or by calculating the Hessians only for a small subset of the frames. The standard deviations of the correction factors are one or two orders of magnitude smaller than the standard deviations of the free energy differences of interest for all test cases. Thus, it is highly unlikely that the correction for the free energy costs of constraints will be the computational bottleneck when it comes to applications to larger systems. In fact, for large systems, free energy calculations are limited by a lack of overlap and large fluctuations of the potential energy difference in the exponential averaging. By increasing the overlap and lowering the energy fluctuations, the introduction of constraints allows the use of free energy calculations for significantly larger systems. This further highlights how constraints, if properly used, can be an extremely valuable tool in computational biophysics.

## Acknowledgments

The authors would like to thank Tim Miller, Stefan Boresch, Yihan Shao, Alex Sodt and Lee Woodcock for insightful discussions, as well as Frank Pickard and Noreen Gervasi for critically reading the manuscript. This work was supported by the intramural research program of the National Heart, Lung and Blood Institute of the National Institutes of Health and utilized the high-performance computational capabilities of the LoBoS and Biowulf Linux clusters at the National Institutes of Health. (<http://www.lobos.nih.gov/> and <http://biowulf.nih.gov>).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbagen.2014.09.001>.

## References

- [1] W.F. Van Gunsteren, H.J.C. Berendsen, Algorithms for macromolecular dynamics and constraint dynamics, *Mol. Phys.* 34 (1977) 1311–1327.
- [2] V. Ovchinnikov, M. Cecchini, M. Karplus, A simplified confinement method for calculating absolute free energies and free energy and entropy differences, *J. Phys. Chem. B* 117 (3) (2013) 750–762, <http://dx.doi.org/10.1021/jp3080578> doi:10.1021/jp3080578.
- [3] D.L. Mobley, J.D. Chodera, K.A. Dill, Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change, *J. Chem. Theory Comput.* 3 (4) (2007) 1231–1235, <http://dx.doi.org/10.1021/ct700032n>.
- [4] E. Barth, K. Kuczera, B. Leimkuhler, R. Skeel, Algorithms for constrained molecular dynamics, *J. Comput. Chem.* 16 (10) (1995) 1192–1209, <http://dx.doi.org/10.1002/jcc.540161003>.
- [5] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23 (1977) 327–341.
- [6] H. Andersen, Rattle: a “velocity” version of the shake algorithm for molecular dynamics calculations, *J. Comput. Phys.* 52 (1) (1983) 24–34, [http://dx.doi.org/10.1016/0021-9991\(83\)90014-1](http://dx.doi.org/10.1016/0021-9991(83)90014-1).
- [7] B. Hess, H. Bekker, H. Berendsen, J. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (12) (1997) 1463–1472, [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4-3.0.CO;2-H](http://dx.doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4-3.0.CO;2-H).
- [8] P. Tao, X. Wu, B.R. Brooks, Maintain rigid structures in Verlet based Cartesian molecular dynamics simulations, *J. Chem. Phys.* 137 (13) (2012) 134110 <http://dx.doi.org/10.1063/1.4756796>.
- [9] W. Van Gunsteren, Constrained dynamics of flexible molecules, *Mol. Phys.* 40 (4) (1980) 1015–1019, <http://dx.doi.org/10.1080/00268978000102101>.
- [10] W. van Gunsteren, M. Karplus, Effect of constraints on the dynamics of macromolecules, *Macromolecules* 15 (6) (1982) 1528–1544, <http://dx.doi.org/10.1021/ma00234a015>.
- [11] S. Toxvaerd, Comment on constrained molecular dynamics of macromolecules, *J. Chem. Phys.* 87 (10) (1987) 6140–6143, <http://dx.doi.org/10.1063/1.453488>.
- [12] D. Tobias, C. Brooks, Molecular dynamics with internal coordinate constraints, *J. Chem. Phys.* 89 (8) (1988) 5115–5127, <http://dx.doi.org/10.1063/1.455654>.
- [13] K. Hinsen, G. Kneller, Influence of constraints on the dynamics of polypeptide chains, *Phys. Rev. E* 52 (6, B) (1995) 6868–6874, <http://dx.doi.org/10.1103/PhysRevE.52.6868>.
- [14] P. Echenique, C.N. Cavasotto, P. Garcia-Risueno, The canonical equilibrium of constrained molecular models, *Eur. Phys. J.-Spec. Top.* 200 (1) (2011) 5–54, <http://dx.doi.org/10.1140/epjst/e2011-01517-9>.
- [15] G. König, P.S. Hudson, S. Boresch, H.L. Woodcock, Multiscale free energy simulations: an efficient method for connecting classical MD Simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes, *J. Chem. Theory Comput.* 10 (4) (2014) 1406–1419, <http://dx.doi.org/10.1021/ct400118k>.
- [16] M. Sprik, G. Ciccotti, Free energy from constrained molecular dynamics, *J. Chem. Phys.* 109 (18) (1998) 7737–7744, <http://dx.doi.org/10.1063/1.477419>.
- [17] W.K. den Otter, W.J. Briels, The calculation of free-energy differences by constrained molecular-dynamics simulations, *J. Chem. Phys.* 109 (1998) 4139–4146.
- [18] W. den Otter, W. Briels, Free energy from molecular dynamics with multiple constraints, *Mol. Phys.* 98 (12) (2000) 773–781.
- [19] N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, T. Yamabe, Structure optimization via free energy gradient method: application to glycine zwitterion in aqueous solution, *J. Chem. Phys.* 113 (9) (2000) 3519–3524, <http://dx.doi.org/10.1063/1.1287785>.
- [20] S. Boresch, M. Karplus, The role of bonded terms in free energy simulations: 1. Theoretical analysis, *J. Phys. Chem. A* 103 (1999) 103–118.
- [21] L. Wang, J. Hermans, Change of bond length in free-energy simulations: algorithmic improvements, but when is it necessary, *J. Chem. Phys.* 100 (1994) 9129–9139.
- [22] W.F. van Gunsteren, T.C. Beutler, F. Fraternali, P.M. King, A. Mark, P. Smith, Computation of free energy in practice: choice of approximations and accuracy limiting factors, in: W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson (Eds.), *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, vol. 2, ESCOM Science, Leiden, 1993, pp. 315–348.
- [23] D.A. Pearlman, P.A. Kollman, The overlooked bond-stretching contribution in free energy perturbation calculations, *J. Chem. Phys.* 94 (1991) 4532–4545.
- [24] D.A. Pearlman, Determining the contributions of constraints in free energy calculations: development, characterization, and recommendations, *J. Chem. Phys.* 98 (1993) 8946–8957.
- [25] J.G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* 3 (1935) 300–313.
- [26] C.H. Bennett, Efficient estimation of free energy differences from Monte Carlo data, *J. Comp. Phys.* 22 (1976) 245–268.
- [27] M.R. Shirts, J.D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states, *J. Chem. Phys.* 129 (12) (2008) 124105 <http://dx.doi.org/10.1063/1.2978177>.
- [28] G. König, S. Boresch, Non-Boltzmann sampling and Bennett’s acceptance ratio method: how to profit from bending the rules, *J. Comput. Chem.* 32 (6) (2011) 1082–1090, <http://dx.doi.org/10.1002/jcc.21687>.
- [29] J. Wereszczynski, J.A. McCammon, Using selectively applied accelerated molecular dynamics to enhance free energy calculations, *J. Chem. Theory Comput.* 6 (11, 3285) (2010) 3285–3292, <http://dx.doi.org/10.1021/ct100322t>.
- [30] B.R. Brooks, D. Janežič, M. Karplus, Harmonic analysis of large systems. I. Methodology, *J. Comput. Chem.* 16 (1995) 1522–1542.
- [31] A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, M. Waroquier, Vibrational modes in partially optimized molecular systems, *J. Chem. Phys.* 126 (22) (2007) 224102 <http://dx.doi.org/10.1063/1.2737444>.
- [32] H.L. Woodcock, W. Zheng, A. Ghysels, Y. Shao, J. Kong, B.R. Brooks, Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit, *J. Chem. Phys.* 129 (21) (2008) 214109 <http://dx.doi.org/10.1063/1.3013558>.
- [33] A. Ghysels, D. Van Neck, B.R. Brooks, V. Van Speybroeck, M. Waroquier, Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach, *J. Chem. Phys.* 130 (8) (2009) 084107 <http://dx.doi.org/10.1063/1.3071261>.
- [34] A. Ghysels, V. Van Speybroeck, E. Pauwels, D. Van Neck, B.R. Brooks, M. Waroquier, Mobile block Hessian approach with adjoined blocks: an efficient approach for the calculation of frequencies in macromolecules, *J. Chem. Theory Comput.* 5 (5) (2009) 1203–1215, <http://dx.doi.org/10.1021/ct800489r>.
- [35] A. Ghysels, V. Van Speybroeck, E. Pauwels, S. Catac, B.R. Brooks, D. Van Neck, M. Waroquier, Comparative study of various normal mode analysis techniques based on partial Hessians, *J. Comput. Chem.* 31 (5) (2010) 994–1007, <http://dx.doi.org/10.1002/jcc.21386>.
- [36] A. Ghysels, H.L. Woodcock, J.D. Larkin III, B.T. Miller, Y. Shao, J. Kong, D. Van Neck, V. Van Speybroeck, M. Waroquier, B.R. Brooks, Efficient calculation of QM/MM frequencies with the mobile block Hessian, *J. Chem. Theory Comput.* 7 (2) (2011) 496–514, <http://dx.doi.org/10.1021/ct100473f>.
- [37] A. Ghysels, B.T. Miller, F.C. Pickard, B.R. Brooks, Comparing normal modes across different modes and scales: Hessian reduction versus coarse-graining, *J. Comput. Chem.* 33 (28) (2012) 2250–2275, <http://dx.doi.org/10.1002/jcc.23076>.
- [38] N. Gö, H.A. Scheraga, On the use of classical statistical mechanics in the treatment of polymer chain conformation, *Macromolecules* 9 (1976) 535–542.
- [39] S. Boresch, M. Karplus, The Jacobian factor in free energy simulations, *J. Chem. Phys.* 105 (1996) 5145–5154.
- [40] D.R. Herschbach, H.S. Johnston, D. Rapp, Molecular partition functions in terms of local properties, *J. Chem. Phys.* 31 (1959) 1652–1661.
- [41] T.P. Straatsma, J.A. McCammon, Treatment of rotational isomers in free energy calculations. ii. Molecular dynamics simulation study of 18-crown-6 in aqueous solution as an example of systems with large numbers of rotational isomeric states, *J. Chem. Phys.* 91 (1989) 3631–3637.
- [42] R.W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.* 22 (1954) 1420.
- [43] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell, CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *J. Comput. Chem.* 31 (4) (2010) 671–690, <http://dx.doi.org/10.1002/jcc.21367>.
- [44] S. Boresch, The role of bonded energy terms in free energy simulations — insights from analytical results, *Mol. Simul.* 28 (2002) 13–37.
- [45] G. König, B.T. Miller, S. Boresch, X. Wu, B.R. Brooks, Enhanced sampling in free energy calculations: combining SGLD with the Bennett’s acceptance ratio and enveloping distribution sampling methods, *J. Chem. Theory Comput.* 8 (10) (2012) 3650–3662, <http://dx.doi.org/10.1021/ct300116r>.
- [46] W.L. Jorgensen, H. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926.
- [47] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux III, M. Schlenkerich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wioriewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of protein, *J. Phys. Chem. B* 102 (1998) 3586–3616.
- [48] D. Wales, M. Hodges, Global minima of water clusters (H<sub>2</sub>O)<sub>n</sub>, n < = 21, described by an empirical potential, *Chem. Phys. Lett.* 286 (1–2) (1998) 65–72, [http://dx.doi.org/10.1016/S0009-2614\(98\)00065-7](http://dx.doi.org/10.1016/S0009-2614(98)00065-7).
- [49] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem.* 29 (2008) 1859–1865.
- [50] B. Brooks, C. Brooks III, A. Mackerell Jr., L. Nilsson, R. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. Pastor, C. Post, J. Pu, M. Schaefer, B. Tidor, R. Venable, H. Woodcock, X. Wu, W. Yang, D. York, M. Karplus, CHARMM: the biomolecular simulation program, *J. Comput. Chem.* 30 (10, Sp. Iss. SI) (2009) 1545–1614, <http://dx.doi.org/10.1002/jcc.21287>.
- [51] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization and dynamics calculations, *J. Comput. Chem.* 4 (1983) 187–217.
- [52] S. Boresch, M. Karplus, The role of bonded terms in free energy simulations. 2. Calculation of their influence on free energy differences of solvation, *J. Phys. Chem. A* 103 (1999) 119–136.
- [53] H.L. Woodcock, B.T. Miller, M. Hodoscek, A. Okur, J.D. Larkin, J.W. Ponder, B.R. Brooks, MSAL: a general utility for multiscale modeling, *J. Chem. Theory Comput.* 7 (4) (2011) 1208–1219, <http://dx.doi.org/10.1021/ct00738h>.
- [54] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A* 31 (1985) 1695–1697.

- [55] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [56] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins and nucleic acids, *J. Am. Chem. Soc.* 117 (1995) 5179–5197.
- [57] G. König, S. Bruckner, S. Boresch, Unorthodox uses of Bennett's acceptance ratio method, *J. Comput. Chem.* 30 (11) (2009) 1712–1718, <http://dx.doi.org/10.1002/jcc.21255>.
- [58] W. Jorgensen, C. Ravimohan, Monte-Carlo simulation of differences in free-energies of hydration, *J. Chem. Phys.* 83 (6) (1985) 3050–3054, <http://dx.doi.org/10.1063/1.449208>.
- [59] U.C. Singh, F.K. Brown, P.A. Bash, P.A. Kollman, An approach to the application of free energy perturbation methods using molecular dynamics: Applications to the transformations of  $\text{CH}_3\text{OH} \rightarrow \text{CH}_3\text{CH}_3$ ,  $\text{H}_3\text{O}^+ \rightarrow \text{NH}_4^+$ , glycine  $\rightarrow$  alanine, and alanine  $\rightarrow$  phenylalanine in aqueous solution and to  $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3 \rightarrow \text{NH}_4^+(\text{H}_2\text{O})_3$  in the gas phase, *J. Am. Chem. Soc.* 109 (1987) 1607–1614.
- [60] F. Ytreberg, D. Zuckerman, Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems, *J. Comput. Chem.* 25 (14) (2004) 1749–1759, <http://dx.doi.org/10.1002/jcc.20103>.
- [61] D. Min, W. Yang, Energy difference space random walk to achieve fast free energy calculations, *J. Chem. Phys.* 128 (19) (2008) 191102 <http://dx.doi.org/10.1063/1.2927744>.
- [62] A. Ben-Naim, Y.J. Marcus, Solvation thermodynamics of nonionic solutes, *J. Chem. Phys.* 81 (1984) 2016–2027.
- [63] S. Bruckner, S. Boresch, Efficiency of alchemical free energy simulations I: practical comparison of the exponential formula, thermodynamic integration and Bennett's acceptance ratio method, *J. Comput. Chem.* 32 (2011) 1303–1319.
- [64] S. Bruckner, S. Boresch, Efficiency of alchemical free energy simulations. II: improvements for thermodynamic integration, *J. Comput. Chem.* 32 (2011) 1320–1333.
- [65] G. König, S. Boresch, Hydration free energies of amino acids: why side chain analog data are not enough, *J. Phys. Chem. B* 113 (26) (2009) 8967–8974, <http://dx.doi.org/10.1021/jp902638y>.
- [66] G. König, F.C. Pickard, Y. Mei, B.R. Brooks, Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4, *J. Comput. Aided Mol. Des.* 28 (3, SI) (2014) 245–257, <http://dx.doi.org/10.1007/s10822-014-9708-4>.
- [67] T. Rod, U. Ryde, Quantum Mechanical Free Energy Barrier for an Enzymatic Reaction, *Phys. Rev. Lett.* 94 (13) (2005) 138302 <http://dx.doi.org/10.1103/PhysRevLett.94.138302>.
- [68] T.H. Rod, U. Ryde, Accurate QM/MM Free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase, *J. Chem. Theory Comput.* 1 (6) (2005) 1240–1251, <http://dx.doi.org/10.1021/ct0501102>.
- [69] J. Heimdal, U. Ryde, Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations, *Phys. Chem. Chem. Phys.* 14 (2012) 12592–12604, <http://dx.doi.org/10.1039/c2cp41005b>.
- [70] S.J. Fox, C. Pittcock, C.S. Tautermann, T. Fox, C. Christ, N.O.J. Malcolm, J.W. Essex, C.-K. Skylaris, Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies, *J. Phys. Chem. B* 117 (32) (2013) 9478–9485, <http://dx.doi.org/10.1021/jp404518r>.