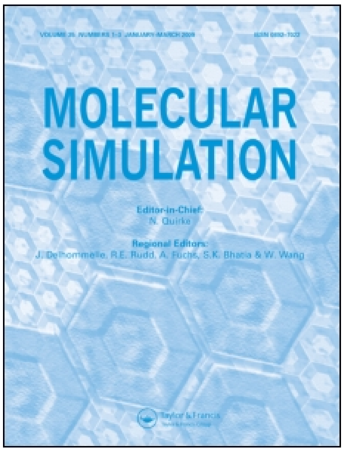

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

The Role of Bonded Energy Terms in Free Energy Simulations - Insights from Analytical Results

Stefan Boresch[a]
[a] Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Vienna, Austria

To cite this Article Boresch, Stefan(2002) 'The Role of Bonded Energy Terms in Free Energy Simulations - Insights from Analytical Results', Molecular Simulation, 28: 1, 13 — 37
To link to this Article: DOI: 10.1080/08927020211969
URL: http://dx.doi.org/10.1080/08927020211969

Molecular Simulation, 2002, Vol. 28(1–2), pp. 13 – 37

THE ROLE OF BONDED ENERGY TERMS IN FREE ENERGY SIMULATIONS – INSIGHTS FROM ANALYTICAL RESULTS

STEFAN BORESCH*

Institut für Theoretische Chemie und Molekulare Strukturbiologie,
Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria

(Received August 2000; accepted November 2000)

A statistical mechanical analysis of the role of bond stretching and angle bending terms, as well as of dummy atoms, in alchemical free energy simulations is presented. It allows one to resolve several ambiguities concerning their treatment from first principles. The relative importance of vibrational, Jacobian factor and potential-of-mean-force-like contributions from bonded energy terms to single and dual free energy differences is compared. Additional conclusions can be drawn for dummy atoms. In particular, it is shown that double free energy differences obtained in thermodynamic cycle calculations can always be made independent of the treatment of dummy atoms. Further, the equivalence of single and dual topology calculations in such applications is demonstrated. The theoretical considerations are accompanied and illustrated by simulation results for the free energy difference of solvation between iso-butane and propane.

Keywords: Jacobian factor contribution; Vibrational contribution; Potential-of-mean-force-like contribution; Dummy atoms; Thermodynamic integration

1. INTRODUCTION

The calculation of free energy differences of biomolecular processes is undoubtedly among the most interesting applications of molecular dynamics (MD) or Monte Carlo (MC) simulations. It has become a widely used tool in computational chemistry and molecular biology, as well as in pharmaceutical research [1 – 3]. While the underlying theory is straightforward and predates computer simulations [4, 5], numerous practical details

---

*Tel.: +43-1-4277-52715, Fax: +43-1-4277-52790, e-mail: stefan@mdy.univie.ac.at

ISSN: 0892-7022 © 2002 Taylor & Francis Ltd DOI: 10.1080/08927020290004359

require great care to obtain accurate and precise results. Methodological issues that often cause problems in practice include the so-called van der Waals endpoint problem [6–9], the treatment of electrostatic interactions [10–12], the correct calculation of bonded and constraint free energy contributions [13–15], and the handling of dummy atoms [16]. Overlooking even a single, apparently insignificant detail can lead to wrong results, or, at least, slow down the convergence of the calculations considerably. The latter is almost equally undesirable because free energy simulations are in any case very demanding in terms of computer time.

This paper addresses two related methodological difficulties just mentioned, the contributions from changes of bonded interactions (mainly the bond stretching and angle bending force field terms) to alchemical free energy differences, and the use of dummy atoms in such calculations. A number of existing studies on these topics, *e.g.* [17–22], were concerned with finding a solution to a practical problem encountered. Often, the results obtained were specific to a particular simulation program, and several contradictory reports in the literature regarding the importance of bonded free energy contributions were the consequence [17, 23–27]. These were resolved only recently by Boresch and Karplus [14, 15]. Starting with a detailed comparison of the so-called single and dual topology methods [28], their two-part study discussed various aspects of the treatment of bonded energy terms in free energy simulations and concluded with comprehensive methodological recommendations. The related topic of dummy atoms was studied as well. More recently, Shobana *et al.* also addressed issues concerning the use and handling of dummy atoms [16]. Since Refs. [14–16] focused on practical problems, the connection between the methodological aspects investigated and the underlying statistical mechanics was not shown very clearly. A primarily theoretical analysis, however, should be insightful for two reasons. First, it may permit one to derive earlier results in a more straightforward fashion. Second, the conclusions obtained have to be independent of methodological details, *e.g.*, they should be valid for single and dual topology simulations alike.

Such a statistical mechanical analysis forms the core of the present work, which is organized as follows. In the main part of the paper (Sect. 2), the physical origin of the contributions from changes in bonded interactions to free energy differences is derived from first principles, the configurational partition function of a polyatomic molecule in solution. In addition, a number of important consequences concerning the correct treatment of dummy atoms are deduced. The validity and the practical utility of the theoretical results are illustrated by calculations of the solvation free energy difference between iso-butane and propane (Sect. 3). The paper ends with a discussion that relates and compares the insights obtained from the theoretical considerations to current simulation practice (Sect. 4). In an appendix it is outlined how to compute analytically the free energy difference for the removal of dummy atoms from the system of interest.

Several aspects of bonded free energy contributions, which have already been covered exhaustively in Refs. [14] and [15] are omitted here: The practical differences between using thermodynamic integration and thermodynamic perturbation (the exponential formula) are not re-analyzed. Boresch and Karplus demonstrated very clearly (both from a methodological, as well as from a theoretical point of view) why the free energy difference of breaking (or of the formation) of bonds cannot be obtained from simulations. (While useful for illustrative purposes, the one alternative outlined in Ref. [14] is not practical for more complicated systems.) Finally, the subtleties concerning constrained degrees of freedom are not discussed; aside from the brief overview given in Ref. [14] we refer the reader to two very detailed treatments of the connection between constraints and free energy simulations [13, 29].

2. THEORY

No review of the theory of free energy simulations and, in particular, of the use of thermodynamic cycles is given since several presentations are widely available, *e.g*. [3]. The goal of this section is to make clear the physical origin of the contributions to free energy differences calculated by computer simulations that result from changes in bonded energy terms. To accomplish this, we first bring the configurational partition function of a polyatomic molecule in solution into a form which can be related to the special situation found in free energy simulations (Sect. 2.1). Based on this expression, the types of contributions arising in alchemical free energy simulations from bonded terms are analyzed (Sect. 2.2). Additional conclusions can be drawn for the special, but important case of dummy atoms (Sect. 2.3).

2.1. The Configurational Partition Function of a Solvated Polyatomic Molecule

The starting point for the derivations of this subsection is the classical mechanical partition function [30]

$$Q_{\text{class}} = \int d\mathbf{r} d\mathbf{p} \exp(-H(\mathbf{r}, \mathbf{p})/k_B T), \tag{1}$$

where H is the Hamiltonian of the system (polyatomic molecule in solvent), k_B is the Boltzmann constant, and T is the temperature. In Cartesian coordinates $\mathbf{r}$, the integration over the conjugate momenta $\mathbf{p}$ for each of the N atoms can be carried out immediately, and one obtains the familiar expression [30]

$$Q_{\text{class}} = Z \prod_{i=1}^{N} \Lambda_i^{-3} \tag{2}$$

where Z is the configurational partition function, and $\Lambda_i = h/(2\pi m_i k_B T)^{1/2}$ is the de Broglie wavelength (h is Planck's constant, and m_i is the mass of atom i).

We now focus on the configurational partition function [30]

$$Z = \int_V d\mathbf{r} \exp(-U(\mathbf{r})/k_B T) \tag{3}$$

of the solvated polyatomic molecule (as usual, U denotes the potential energy). Our goal is to rewrite it so that the role of bonded energy terms becomes clear. As the first step, we distinguish between the intra-solute interactions $U_{PP}(\mathbf{r_P})$, the solute–solvent interactions $U_{PW}(\mathbf{r_P}, \mathbf{r_W})$, and the solvent–solvent interactions $U_{WW}(\mathbf{r_W})$, *i.e.*,

$$U(\mathbf{r}) = U_{PP}(\mathbf{r_P}) + U_{PW}(\mathbf{r_P}, \mathbf{r_W}) + U_{WW}(\mathbf{r_W}). \tag{4}$$

The subscripts P and W denote solute and solvent, respectively; the Cartesian coordinate space $\mathbf{r}$ has been divided into solute coordinates $\mathbf{r_P}$ and solvent coordinates $\mathbf{r_W}$. A separation of the potential energy function U into three terms as in Eq. (4) is always possible in practice because of the additivity of the force fields typically used in free energy simulations [31–34].

Introduction of the potential of mean force (PMF) [35]

$$\exp(-\beta W) = \int d\mathbf{r_W} \exp[-\beta(U_{PW} + U_{WW})] \tag{5}$$

permits to formally dispose of the solvent degrees of freedom (DOF). (The symbol β has the standard meaning of $1/k_B T$.) Simultaneously switching to internal coordinates $\tilde{\boldsymbol{q}}$ for the solute, the configurational partition function becomes

$$Z = \int d\tilde{\boldsymbol{q}} \, |\mathbf{J}| \exp[-\beta(U_{PP}(\tilde{\boldsymbol{q}}) + W(\tilde{\boldsymbol{q}}))], \tag{6}$$

where $|\mathbf{J}|$ is the Jacobian of the transformation from Cartesian to $3N-6$ internal coordinates (bonds, bond angles, dihedral angles *etc.*) [13, 36, 37]. Herschbach, Johnston and Rapp (HJR) were the first to study its functional form in great detail; Table I of Ref. [36] makes it possible to write down $|\mathbf{J}|$ immediately for all coordinate transformations of practical importance. The following two properties of the Jacobian are of particular relevance: (i) $|\mathbf{J}|$ can be written as a product of factors, one for each atom *i*, *i.e.*, $|\mathbf{J}| = \prod_i J_i$, and (ii) the J_i depend only on the bond lengths and bond angles (and *not* on dihedrals), *i.e.*, the DOF of special interest in the present context.

We, therefore, distinguish between L hard $\mathbf{q}$ and $3N-6-L$ soft $\mathbf{Q}$ DOF, *i.e.* $\{\tilde{q}\} = \{\mathbf{q}, \mathbf{Q}\}$ (*cf.* Ref. [37]), and always include the bond stretching and angle bending terms in the hard DOF $\mathbf{q}$. Equation (6) then can be rearranged and becomes

$$Z = \int d\mathbf{q}d\mathbf{Q}\, |\mathbf{J}(\mathbf{q})| \exp[-\beta U_{PP}(\mathbf{q})] \exp[-\beta(U_{PP}(\mathbf{Q}) + W(\mathbf{q};\mathbf{Q}))] \tag{7}$$

Equation (7) simplifies considerably if one replaces the hard DOF $\mathbf{q}$ in $\mathbf{J}(\mathbf{q})$ and $W(\mathbf{q};\mathbf{Q})$ by their respective equilibrium values $\mathbf{q_0}$. Setting $\mathbf{q} = \mathbf{q_0}$ implies the validity of the rigid rotator (RR) approximation [36, 38] and independence of the $\mathbf{q}$ from nonbonded interactions; the justification of these two assumptions is discussed shortly. With this approximation one can separate terms depending on $\mathbf{q}$ and $\mathbf{Q}$

$$\begin{aligned} Z = {} & |\mathbf{J}(\mathbf{q} = \mathbf{q_0})| \int d\mathbf{q} \exp(-\beta U_{PP}(\mathbf{q})) \\ & \times \int d\mathbf{Q} \exp[-\beta(U_{PP}(\mathbf{Q}) + W(\mathbf{q} = \mathbf{q_0};\mathbf{Q}))] \end{aligned} \tag{8}$$

The configurational partition function has become a product of three factors: the Jacobian $|\mathbf{J}(\mathbf{q} = \mathbf{q_0})|$, an intra-solute contribution arising solely from the hard DOF (*i.e.*, bond stretching and angle bending terms), and a configurational integral over the soft DOF (*i.e.*, the remaining dihedral angles).

Instead of discussing Eq. (8), we consider a somewhat more special case, which is slightly more insightful, and which makes the present derivations consistent with Refs. [36, 37] and [39]. If $U_{PP}(\mathbf{q})$ is described by a (generalized) harmonic potential function (this is, in fact, often the criterion to define a hard DOF!)

$$U_{PP}(\mathbf{q}) = \frac{1}{2}\sum_{i,j}^{L} K_{ij}(q_i - q_{i,0})(q_j - q_{j,0}) \tag{9}$$

the first integral in Eq. (8) can be evaluated analytically to give [36–38]

$$\int d\mathbf{q}\exp(-\beta U_{PP}(\mathbf{q})) = \left[\frac{(2\pi k_B T)^L}{\det\mathbf{K}}\right]^{1/2}. \tag{10}$$

Here det$\mathbf{K}$ is the determinant of the force constant (second derivative) matrix of $U_{PP}(\mathbf{q})$ with elements $K_{ij} = \partial^2 U_{PP}/\partial q_i \partial q_j$, *cf.* Eq. (9). Combining Eqs. (8) and (10) and using $|\mathbf{J}(\mathbf{q} = \mathbf{q_0})| = \prod_i J_i$ (*cf.* above and Refs. [36, 37] and [39]), we finally obtain

$$\begin{aligned} Z = {} & \left[\prod_{i=1}^{N} J_i\right]\left[\frac{(2\pi k_B T)^L}{\det\mathbf{K}}\right]^{1/2} \\ & \times \int d\mathbf{Q}\exp[-\beta(U_{PP}(\mathbf{Q}) + W(\mathbf{q} = \mathbf{q_0}; \mathbf{Q}))] = Z_J Z_{vib} Z_{PMF}. \end{aligned} \tag{11}$$

Equation (11) is the main result of this subsection. It shows that the hard DOF $\mathbf{q}$ (*i.e.*, mainly the bond stretching and angle bending terms) contribute to Z in a threefold manner. They appear in the Jacobian factor contribution to the configurational partition function $Z_J = \prod_i J_i$, in the determinant of the force constant matrix $\mathbf{K}$ in Z_{vib} (Eq. (10)), and as constants in the remaining configurational integral Z_{PMF} over the soft DOF $\mathbf{Q}$. The first term, Z_J, is a product of bond lengths and bond angles [36, 37, 39] and reflects the geometry of the solute. Z_{vib} is a function of the force constants of the harmonic force field terms; it seems reasonable to label this term the vibrational contribution of the solute to the configurational partition function of the system. The third term, Z_{PMF}, contains the contributions from the soft DOF, as well as all contributions from solvent (*i.e.*, from solute–solvent and solute–solvent interactions). The contributions involving the solvent (which are contained in the PMF $W(\mathbf{q_0}, \mathbf{Q})$, hence the label Z_{PMF}) depend on the geometry (size, shape *etc.*) of the solute, and, thus, are, in part, determined by the (equilibrium values of the) bond lengths and bond angles $\mathbf{q_0}$.

The separation of the configurational partition function achieved in going from Eq. (7) to Eq. (8) or, alternatively, to Eq. (11) relies on the validity of the RR approximation and the independence of the hard DOF $\mathbf{q}$ from interactions with the solvent. For bond stretching and angle bending terms described by harmonic force field terms, both assumptions are reasonable in the majority of cases. Furthermore, it should be kept in mind that we are primarily interested in qualitative insights. Thus, even in instances when it may not be exact to interpret the role of the hard DOF in the configurational partition function in terms of three separable factors, Eq. (11) still serves to emphasize the three physical origins of contributions from $\mathbf{q}$ to Z.

2.2. Consequences for (Alchemical) Free Energy Differences

Having suitably rearranged the configurational partition function of a solvated, polyatomic molecule (Eq. (11)), we can now address the role of bonded energy terms in free energy differences. We first focus on alchemical, single free energy differences; then we discuss the in practice much more important double free energy differences resulting from thermodynamic cycle calculations [3, 40]. The subsection is concluded by a comparison of the so-called single and dual topology methodology [28].

2.2.1. Single Free Energy Differences

Inserting Eq. (11) into

$$\Delta A = -k_B T \ln \frac{Z_f}{Z_i}, \tag{12}$$

where the subscripts i and f denote the initial and final state of the system, one obtains

$$\begin{aligned} \Delta A &= \Delta A_J + \Delta A_{vib} + \Delta A_{PMF} \\ &= -\frac{1}{\beta} \ln \prod_k \frac{J_{k,f}(\mathbf{q_{f,0}})}{J_{k,i}(\mathbf{q_{i,0}})} - \frac{1}{\beta} \ln \left(\frac{\det \mathbf{K_i}}{\det \mathbf{K_f}} \right)^{1/2} \\ &\quad - \frac{1}{\beta} \ln \left[\frac{\int d\mathbf{Q_f} \exp[-\beta \tilde{W}_f(\mathbf{q_{f,0}}; \mathbf{Q_f})]}{\int d\mathbf{Q_i} \exp[-\beta \tilde{W}_i(\mathbf{q_{i,0}}; \mathbf{Q_i})]} \right]. \end{aligned} \tag{13}$$

In the above equation use has been made of the shorthand $\tilde{W}(\mathbf{q} = \mathbf{q_0}; \mathbf{Q}) = U_{PP}(\mathbf{Q}) + W(\mathbf{q} = \mathbf{q_0}; \mathbf{Q})$. The overall (single) free energy difference ΔA is the sum of three terms to which we refer as Jacobian factor (ΔA_J), vibrational (ΔA_{vib}) and PMF-like free energy contribution (ΔA_{PMF}); this is, of course, an immediate consequence of the factorization of the configurational partition function (Eq. (11)).

To illustrate the general Eq. (13), we consider the special case of the free energy difference between two diatomic molecules in solution that differ in the parameterization of the harmonic bond term (force constant K, equilibrium bond length r_0), as well as the nonbonded parameters describing the interaction of the diatomic molecules with the solvent. (Numerical results for such a system were presented in Ref. [15].) For this simple system the Jacobian factor is $V \times 4\pi r_0^2$ (V is the volume of the system), and the force matrix $\mathbf{K}$ consists of a single term, the force constant K [14, 36, 39]. Since there is only one hard intramolecular energy term (the bond), Z_{PMF} reduces to $\exp(-\beta W(r_0))$ (*cf.* Eq. (5)). Thus, Eq. (13) becomes

$$\Delta A = -\frac{1}{\beta}\ln\left(\frac{r_{f,0}^2}{r_{i,0}^2}\right) - \frac{1}{\beta}\ln\left(\frac{K_i}{K_f}\right)^{1/2} + [W_f(r_{f,0}) - W_i(r_{i,0})] \qquad (14)$$

This simple system is particularly useful to highlight one more time the role of the bonded DOF in ΔA_{PMF}. The change of the (equilibrium) bond length from $r_{i,0}$ to $r_{f,0}$ modifies the solute–solvent and solvent–solvent interactions; this constitutes one of the contributions to ΔA_{PMF}. In this system it can be easily as important as the free energy difference resulting from the changes of the Lennard-Jones parameters and charges of the solute.

2.2.2. Double Free Energy Differences

So far, we have focused on the role of bonded energy terms in single free energy differences. However, most practical applications are based on thermodynamic cycle calculations, *i.e.*, double free energy differences [3, 40]. The two steps involve the identical alchemical mutation in a different environment. For example, to compute a solvation free energy difference, solute I is transmuted into solute II in the gas phase and in solution. Similarly, in ligand binding calculations, the same alchemical transformation is carried out for the ligand in solution and for the ligand bound to its receptor. Applying Eq. (13) to such parallel steps, one sees that identical Jacobian factor and vibrational free energy contributions are obtained, whereas ΔA_{PMF} varies depending on the environment. Thus, ΔA_J and ΔA_{vib} cancel from a thermodynamic cycle calculation (*i.e.*, from double free energy differences $\Delta\Delta A$), provided that changes in bonded terms are handled identically in both steps of the cycle. The validity of this conclusion relies, of course, on the single approximation $\mathbf{q} = \mathbf{q_0}$ underlying Eqs. (8), (11) and (13). If there are deviations from the RR approximation and the independence of bonded energy terms from solute–solvent interactions, the cancellation of the Jacobian factor and vibrational free energy contributions from $\Delta\Delta A$ is not exact and so-called self terms are obtained [14, 15]. The failure to detect such contributions in a recent study suggests that the assumption $\mathbf{q} = \mathbf{q}_0$ is reasonable for bond stretching and angle bending DOF [14,15].

The (at least approximate) cancellation of ΔA_J and ΔA_{vib} from double free energy differences partially justifies the omission of one or both of these two contributions by many (single topology) computer codes, most notably AMBER [17, 41, 42]. It should, however, be kept in mind that for single free energy differences these terms are required. Also, to ensure that possible self term contributions to double free energy differences are not omitted (even if they are likely to be small, *cf.* above), the inclusion of all three types of bonded free energy contributions, ΔA_J, ΔA_{vib} and ΔA_{PMF}, in the calculated free energy differences is necessary as well.

Finally, it should be mentioned that the techniques explored by HJR [36] permit the analytic evaluation of ΔA_J and, to some degree, also of ΔA_{vib}. Since the Jacobian is a product of factors for each solute atom [36, 37, 39], only those atoms that are different in the initial and final state need to be considered. Because of the locality of alchemical mutations (*cf.* the discussion in Ref. [39]), ΔA_J can be obtained straightforwardly even for very complex systems. Since the force matrix is often block diagonal [36, 38], many factors of the determinants cancel from the ratio, and the analytical evaluation of ΔA_{vib} may be possible as well. Examples of such calculations can be found in Refs. [14, 15, 36] and [39], as well as in Sect. 3.2 below. Their main use lies in correcting double free energy differences where bonded terms were not treated identically in the two steps of the thermodynamic cycle [39, 43].

2.2.3. Single vs. Dual Topology Free Energy Simulations

At this point we have reached from first principles one of the central results of Ref. [14], the distinction between Jacobian factor, vibrational and PMF-like free energy contributions that result from changes in bonded energy terms. In addition, their respective role in single and double free energy differences was compared. There was no need to consider technical details of the simulations, such as the choice between a single or dual topology methodology. Background on these two approaches can be found in, *e.g.* [14–16, 27, 28]. The theoretical framework applies to both cases and can, in fact, be used to make clear the differences between the two. Consider, as an example, the alchemical transformation of RCH_2Cl into RCH_2Br (where R denotes the rest of the molecule which is not affected by the mutation). The two end states have different Lennard-Jones parameters for Cl and Br, as well as different partial charges and bonded energy terms for all atoms of the $-CH_2Cl/Br$ group. In a single topology simulation, the various energy terms, including bond stretching and angle bending terms, are changed appropriately to effect the transmutation from initial to final state. This is, of course, exactly the situation described by the present derivations and the consequences of altering hard DOF $\mathbf{q}$ are made clear by Eq. (13).

In a dual topology simulation of this alchemical mutation, on the other hand, $-CH_2Cl$ and $-CH_2Br$ are present simultaneously; both groups are connected to R, but they never "see" each other. The nonbonded interactions of the $-CH_2Cl$ and $-CH_2Br$ groups with R and, possibly, solvent are turned on (off) as a function of the coupling parameter. The bonded energy terms, however, remain unchanged (this corresponds to a so-called ideal gas molecule reference state (see Sect. 2b of Ref. [14])). In other words, the bonded energy terms (usually including the soft DOF!) are the same in the initial and in the final state. The hard $\mathbf{q}$ and soft DOF $\mathbf{Q}$ of a dual topology hybrid molecule comprise the $\mathbf{q_i}$ and $\mathbf{q_f}$, as well as $\mathbf{Q_i}$ and $\mathbf{Q_f}$, of a corresponding single topology arrangement. Similarly, $\mathbf{K_i}$ and $\mathbf{K_f}$ are identical, *i.e.*, there are no changes in force constants. If one takes this into account, Eq. (13) reduces to

$$\Delta A_{d.t.} = \Delta A_{PMF,d.t.} = -\frac{1}{\beta}\ln\left[\frac{\int d\mathbf{Q}\exp[-\beta(U(\mathbf{Q}) + W_f(\mathbf{q_0};\mathbf{Q}))]}{\int d\mathbf{Q}\exp[-\beta(U(\mathbf{Q}) + W_i(\mathbf{q_0};\mathbf{Q}))]}\right] \qquad (15)$$

in the dual topology case. Since W_i and W_f depend on the same hard DOF $\mathbf{q_0}$, $\Delta A_{PMF,d.t.}$ is solely determined by the changes in nonbonded parameters (which differentiate W_f from W_i). This was already found by Boresch and Karplus, but the present analysis does not require free energy components as in Refs. [14] and [15]. Further, from Eq. (15) one sees immediately that in dual topology calculations no Jacobian factor and vibrational free energy contributions arise – after all, no bonded terms are changed! From a comparison of Eqs. (13) and (15), it is obvious that single free energy differences obtained by a single and a dual topology calculation are different (*cf.* Ref. [15] for numeric examples). This must be the case since the two methodologies involve different end states. In spite of this identical double free energy differences are obtained in thermodynamic cycle calculations; the proof follows automatically from the considerations regarding dummy atoms presented next.

2.3. Further Consequences for Dummy Atoms

In standard MC or MD simulations, atoms cannot be created or removed. If, therefore, initial and final state of an alchemical mutation have different number of atoms (as is the case in the majority of applications; consider, *e.g.*, the transmutation of iso-butane into propane), placeholders, usually called dummy atoms, are required to keep the number of atoms the same at the two endpoints, *e.g.* [3, 14–16, 27]. Although dummy atoms are usually associated with single topology simulations, they are present in dual topology simulations as well [14–16, 27].

The following theoretical analysis of dummy atoms, similar to that of Sect. 2.2, is mostly independent of methodological details (*e.g.*, single *vs.* dual topology) and requires just two premises: (i) Dummy atoms remain bound to the physical system of interest; in particular, bonds and bond angle energy terms are not turned off since it has been shown that simulations in which this is attempted do not converge [14]. (ii) They do not participate in any nonbonded interactions. We mention that somewhat similar considerations (with a strong focus on dual topology simulations) can be found in [16]. To denote internal coordinates in which a dummy atom participates, the subscript D is used; *i.e.*, $\mathbf{q_D}$ and $\mathbf{Q_D}$ denote the hard and soft DOF involving dummy atoms. (In the following, $\mathbf{q}$ and $\mathbf{Q}$ represent only those hard and soft DOF without dummy atoms). We now consider the configurational partition of a solvated polyatomic molecule with dummy atoms. Eq. (11) is, of course, still valid. Since dummy atoms have normal bonded interactions (first premise, *cf.* above), they contribute just like regular atoms to Z_J and Z_{vib}. By contrast, several simplifications are possible in $Z_{PMF} = \int d\mathbf{Q} d\mathbf{Q_D} \exp[-\beta(U_{PP}(\mathbf{Q}, \mathbf{Q_D}) + W(\mathbf{q_0}, \mathbf{q_{0,D}}; \mathbf{Q}, \mathbf{Q_D}))]$, the remaining configurational integral (*cf.* Eq. (11)), which is slightly rewritten here to make explicit the presence of dummy atoms. First, the intramolecular energy $U_{PP}(\mathbf{Q}, \mathbf{Q_D})$ can be split into two terms, $U'_{PP}(\mathbf{Q_D})$ and $U''_{PP}(\mathbf{Q})$, the first containing all soft DOF in which one or more dummy atoms are involved, the second those without any dummy atoms. Second, one can always choose the bonded energy terms in which dummy atoms participate (both hard and soft DOF) so that $W(\mathbf{q_0}, \mathbf{q_{0,D}}; \mathbf{Q}, \mathbf{Q_D}) = W(\mathbf{q_0}; \mathbf{Q})$, *i.e.*, so that the PMF is independent of dummy atoms (this will be shown shortly). Thus, Z_{PMF} becomes

$$\begin{aligned} Z_{PMF} = &\int d\mathbf{Q_D} \exp[-\beta U'_{PP}(\mathbf{Q_D})] \\ &\times \int d\mathbf{Q} \exp[-\beta(U''_{PP}(\mathbf{Q}) + W(\mathbf{q_0}; \mathbf{Q}))] = Z_D \tilde{Z}_{PMF} \end{aligned} \tag{16}$$

The contribution of the dummy atoms (Z_D) factors out of the full configurational integral (Z_{PMF}). Similarly to the Jacobian and the vibrational contributions, Z_D also involves only intramolecular, bonded energy terms and is, thus, independent of the environment (*i.e.*, Z_D cancels from double free energy differences). The full configurational partition function (Eq. (11)) of a system with dummy atoms can, thus, be written as

$$Z = Z_J Z_{vib} Z_{PMF} = Z_{J,D} \tilde{Z}_J Z_{vib,D} \tilde{Z}_{vib,D} Z_D \tilde{Z}_{PMF} \tag{17}$$

where the Jacobian ($Z_{J,D}$) and vibrational contributions ($Z_{vib,D}$) from dummy atoms have been factored as well. The tilde indicates the contributions from the physical system.

The factorization $Z_J = Z_{J,D} \times \tilde{Z}_J$ in Eq. (17) is always possible since the Jacobian is a product of terms for each atom [36]. The validity of the analogous step $Z_{vib} = Z_{vib,D} \times \tilde{Z}_{vib}$ is less clear; furthermore we still have to justify the step $W(\mathbf{q_0}, \mathbf{q_{0,\,D}}; \mathbf{Q}, \mathbf{Q_D}) = W(\mathbf{q_0}; \mathbf{Q})$ leading to Eq. (16). Although logical at first glance (dummy atoms do not participate in any nonbonded interactions (second premise, *cf.* above)), some care concerning the choice of $\mathbf{q_D}$ and $\mathbf{Q_D}$ is required, as also pointed out by Shobana *et al.* [16]. Modern force fields frequently contain many redundant bond angle and dihedral angle DOF. For example, the new CHARMM all-atom force field [34] automatically sets up the largest possible number of angle and dihedral angle terms in which an atom can participate. If one or more of these 'over-determined' atoms are changed into dummy atoms and if all bonded energy terms are left intact, a coupling between the dummy atoms and the normal atoms of the system may occur that renders W dependent on $\mathbf{q_D}$ and/or $\mathbf{Q_D}$ [16]. Similarly, the factorization $Z_{vib} = Z_{vib,D} \times \tilde{Z}_{vib}$ may become impossible.

Complications caused by such redundant DOF have been well known for a long time in the field of molecular vibrations [36, 38]. Here, a few practical recommendations have to suffice. Ideally, each dummy atom is connected to the rest of the system by exactly three, non-redundant bonded terms. For example, at the end of a chain of $N-1$ real atoms, these should be the bond between atom N (the dummy atom) and atom $N-1$, the bond angle $\theta_{N,N-1,N-2}$, and the dihedral angle $\tau_{N,N-1,N-2,N-3}$. The choice of these three internal coordinates ensures that the contribution of the dummy atom to the configurational partition function can be factored [36]. HJR also discuss several possibilities that may arise for branched configurations. In alchemical free energy simulations one can follow their conclusions by simply removing the redundant angle and/or dihedral angle terms at the endpoint where the atoms have become dummy atoms (*cf.* [16]). If a larger group is transformed into a group of dummy atoms, this may become quite tedious. This is the rule in dual topology simulations, where at each endpoint the nonbonded interactions with a complete group are turned off (*e.g.*, in the above example either the $-CH_2Cl$ or the $-CH_2Br$ group is decoupled) [14–16]. However, there is no need to avoid redundant DOF completely. The focus has to be on the junction between the physical system and the dummy group; usually, it is sufficient to make sure that the dummy atom bound to the physical system is attached by non-redundant bonded terms only; the remainder of the dummy group can contain redundant DOF. If the suggestions just given are followed, the independence of the PMF from hard and soft DOF involving dummy atoms is ensured. Further, the force matrix automatically contains (at least) two block diagonal groups, one for the physical system, the other for the dummy atoms (the dummy group). Thus, the factorization of Z_{vib} is guaranteed as well. Practical examples are described in Ref. [16] although some of their 'forbidden' configurations seem to be unnecessarily restrictive; the treatment of the dummy atoms in the iso-butane/propane model calculation presented next also provides an illustrative case.

Having justified Eqs. (16) and (17), we return to the role of dummy atoms in single and double free energy differences. They contribute to Z_J and Z_{vib} ($Z_{J,D}$ and $Z_{vib,D}$ in Eq. (17)), and they add the separable term Z_D resulting from the soft DOF in which they participate (Eq. (16)). Consequently, whenever dummy atoms are present at one or both end states, single free energy differences will depend on the details of the force field terms involving $\mathbf{q_D}$ and $\mathbf{Q_D}$. However, this (artificial) contribution *always* cancels from thermodynamic cycles, which can be seen as follows: While the factorization leading to Eqs. (8) and (11) relies on (i) the independence of the hard DOF from nonbonded interactions and (ii) the RR approximation, it is always valid for those intramolecular energy terms involving dummy atoms. Concerning the first point (i), there are no nonbonded interactions (and, in particular, solute–solvent interactions) to consider. Furthermore (ii), the force constants of the bond stretching and angle bending terms in which dummy atoms are involved can always be chosen high enough so that the RR approximation is applicable. (It should be kept in mind that dummy atoms are a necessary artifact of the simulation methodology, so their parameterization is completely arbitrary.) This shows that the free energy contributions from dummy atoms (resulting from $Z_{J,D}$, $Z_{vib,D}$ and Z_D) must cancel exactly from thermodynamic cycle calculations (*i.e.*, double free energy differences), provided that the dummy atoms are treated identically in the two steps. Particular treatments of dummy atoms may lead to different convergence behavior of the simulations, but the final $\Delta\Delta A$ must be independent of these technical details. This requirement can serve as a rigorous test of the setup chosen for a particular mutation or of the computer code used to carry out the simulations. The independence of double free energy differences from the treatment of dummy atoms was already argued in Ref. [14]; however, the formalism presented here makes possible a cogent, more formal proof.

In addition, the considerations of this section show immediately the equivalence of single and dual topology simulations for the calculation of double free energy differences. In dual topology simulations, there are always two dummy groups, one at each end state (in the above example, these are the $-CH_2Cl$ and the $-CH_2Br$ group) [14–16, 27, 28]. However, since the analytical considerations just given apply to dummy atoms regardless of the simulation methodology, any unphysical contributions of these dummy groups cancel from the overall result. Thus, the same double free energy difference as in an analogous single topology simulation is obtained.

3. AN ILLUSTRATIVE APPLICATION

A central result of the Theory section is the independence of double free energy differences from the treatment of dummy atoms. To verify this conclusion numerically, the free energy difference of hydration between iso-butane and propane was computed in a variety of ways (see Fig. 1). The results also permit an analysis of the influence of dummy atoms on alchemical single free energy differences. The technical details of the calculations are described next (Sect. 3.1), followed by the presentation and the discussion of the results (summarized in Tab. I) in Sect. 3.2.

3.1. Technical Details

Figure 1 shows an overview of all single free energy differences that were computed to determine the free energy difference of hydration ($\Delta\Delta A_{solv}$) between iso-butane (abbreviated ibut) and propane (abbreviated prop). Several alchemical paths (horizontal arrows in Fig. 1, transmutation of iso-butane into propane in the gas phase and in solution), as well as chemical paths (vertical arrows in Fig. 1, transfer of iso-butane or propane from the gas phase into aqueous solution) were used. Since the two molecules do not have the same number of atoms, the introduction of dummy atoms is necessary at the propane endpoint along the alchemical paths (ΔA_3, ΔA_4, $\Delta A'_3$ and $\Delta A'_4$). Their presence is made clear by the labels prop-d1 and prop-d2 in Figure 1 (details of these hybrids are given below). Along the chemical

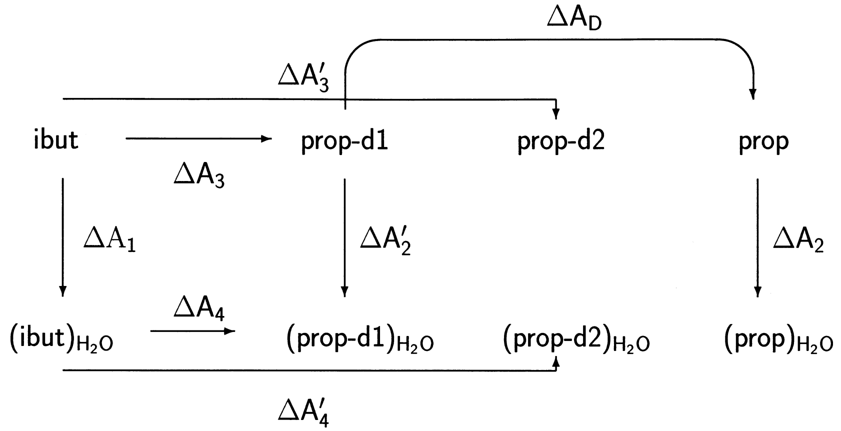

FIGURE 1 Schematic representation of chemical and alchemical (single) free energy differences calculated to determine the free energy difference of hydration $\Delta\Delta A_{solv}$ between iso-butane (ibut) and propane (prop). The labels prop-d1 and prop-d2 denote two propane-like molecules with dummy atoms attached to them (see the main text for further details).

paths (ΔA_1, ΔA_2), no dummy atoms are required. However, since neither prop-d1 nor prop-d2 are identical to propane (prop), the free energy of transferring prop-d1 from the gas phase into aqueous solution was calculated as well ($\Delta A_2'$). This permits the exact closure of the thermodynamic cycle $\Delta A_1 + \Delta A_4 - \Delta A_2' - \Delta A_3$. The transformation prop-d1 → prop, *i.e.*, the removal of dummy atoms, ΔA_D, cannot be obtained from computer simulations since it involves the breaking of bonds [14]. In the appendix it is shown how the free energy difference for this step (ΔA_D) can be calculated analytically.

All free energy simulations were carried out with the PERT module of CHARMM [44]. The locally modified version c28a2 of the program permitted the use of soft core potentials to avoid van der Waals endpoint problems [7, 8]; this enhancement is included in CHARMM versions c28a3 and higher. Only thermodynamic integration [3, 4] was used since with this method it is straightforward to include all contributions from bonded energy terms [14, 15, 45]. Parameters were taken from the CHARMM all-atom force field [34]. The solvated systems consisted of the solute (iso-butane or propane) and 211 CHARMM modified TIP3P waters [46] in a cubic box with a side length of 18.856 Å. A group-based cutoff criterion was applied to the nonbonded interactions, which were switched off smoothly between 6.5 and 7.5 Å. The nonbonded list was updated heuristically based on a 9.4 Å cutoff. Since all charge groups in this system are neutral, acceptable results can be expected from this treatment of the electrostatic interactions [10].

As SHAKE [47] was only applied to the water molecules and because the masses of the hydrogen atoms were left at 1 amu, a timestep of 0.5 fs was chosen to avoid instabilities of the integrator. It has been observed that good thermal equilibration of the system is crucial to obtain accurate and precise free energy contributions from bonded energy terms [19, 39, 48]. This was accomplished by the use of multiple Nosé-Hoover thermostats [49]. In the gas phase simulations, separate thermostats were coupled to each of the four charge groups present in iso-butane (and hence also in the hybrid propanes with dummy atoms attached to them); in the solution calculations, solute and solvent were coupled to separate thermostats.

All single free energy differences were calculated with the following protocol: At $\lambda = 0$, the system was equilibrated for 200,000 steps, followed by accumulation of $\langle \partial U / \partial \lambda \rangle_\lambda$ over 20,000 steps. The coupling parameter was then incremented in steps of 0.025 till the endpoint $\lambda = 1$ was reached; at each λ-value, 8,000 steps of equilibration were followed by 20,000 steps of data accumulation. Overall, this amounts to 1,340,000 MD steps (670 ps) for a single free energy difference. To obtain reliable error estimates for each single free energy difference, three forward and backward calculations (started from different random initial velocities) were carried out for all simulations involving solvent, and ten forward and backward calculations were carried out in the gas phase. For each mean value of the resulting six (twenty) data points, the confidence interval for the population mean of the data based on the Student t distribution with a confidence level of 0.95 is reported [50]. When free energy differences are combined (as in the calculation of $\Delta\Delta A_{solv}$), the overall error estimate was obtained by error propagation [50].

As already mentioned, at the propane endpoint in the alchemical calculations three dummy atoms were needed. In the transmutation one methyl (CH_3) group of iso-butane was changed into a HD_3 group in prop-d1 or prop-d2. In the prop-d1 hybrid molecule the force field parameters of all bond stretching and bond angle terms involving this group were the same as in iso-butane. The second hybrid, prop-d2, differs from prop-d1 by the equilibrium bond-length of the H—D bond, which was reduced to 0.5 Å (from 1.111 Å of the corresponding C—H bond in iso-butane). In both hybrids all Urey-Bradley-like energy terms in which dummy atoms participate were removed. In the CHARMM all-atom force field [34] the CH_3/HD_3 group is also involved in numerous dihedral angle terms (though there are no improper dihedral angle terms). While the parameters of the dihedral angles involving the C/H atom (but no dummy atoms) were changed as appropriate for the mutation of iso-butane into propane, all but one dihedral angle involving the dummy atoms were deleted at the propane endpoint. This removal of superfluous dihedrals is necessary to avoid the pitfalls mentioned in Sect. 2.3 due to redundant DOF; see also Ref. [16]. In this study only the C_1—C_2—H—D_1 dihedral was maintained (all others were deleted); however, any other dihedral angle involving one of the dummy atoms could have been used instead. The astute reader will have noticed that there is also one redundant bond angle term in the HD_3 group at the propane endpoint (there are six instead of five bond angle terms). However, this term does not prevent the factorization of Z (Eq. (17)); in fact, its presence accelerated the convergence of the calculations since the average geometry of the HD_3 group was kept close to that of the corresponding methyl group at the iso-butane endpoint.

3.2. Simulation Results

All results are summarized in Table I. It is organized in two main columns, corresponding to the chemical (vertical) and alchemical (horizontal) paths in Figure 1. In each of the two columns, the single free energy differences (arrows in Fig. 1) are listed first, followed by the free energy differences of hydration that can be calculated from them. Altogether, the simulations carried out permit to obtain $\Delta\Delta A_{solv}$ between iso-butane and propane in four different ways. These double free energy differences (last two rows in Tab. I) agree excellently with each other, and they are very close to the experimental value of -0.31 kcal/mole [51]. While the error bars of the individual results (≤ 0.33 kcal/mole) reflect their high precision, the agreement between the four $\Delta\Delta A_{solv}$ values indicates also high accuracy. In particular, the error of closing the thermodynamic cycle $\Delta A_1 + \Delta A_4 - \Delta A_2' - \Delta A_3$

TABLE I Summary of simulation results obtained for the iso-butane/propane system. All free energy differences are in kcal/mole. The labelling of the free energy differences is the same as in Figure 1

Chemical paths		*Alchemical paths*	
ΔA_1	2.10 ± 0.24	ΔA_3	20.59 ± 0.04
ΔA_2	1.74 ± 0.22	ΔA_4	19.90 ± 0.22
$\Delta A_2'$	1.70 ± 0.15	$\Delta A_3'$	23.47 ± 0.04
		$\Delta A_4'$	23.09 ± 0.19
$\Delta\Delta A_I = \Delta A_2 - \Delta A_1$	-0.36 ± 0.33	$\Delta\Delta A_{III} = \Delta A_4 - \Delta A_3$	-0.69 ± 0.22
$\Delta\Delta A_{II} = \Delta A_2' - \Delta A_1$	-0.40 ± 0.28	$\Delta\Delta A_{IV} = \Delta A_4' - \Delta A_3'$	-0.38 ± 0.19

is just -0.29 kcal/mole. Furthermore, this small value of $\approx k_BT/2$ is well within the statistical error estimate for the sum of these four steps of ± 0.36 kcal/mole (obtained from error propagation of the individual uncertainties listed in Tab. I). The high precision and accuracy of the calculations provides numeric proof that there are no problems to include free energy contributions from bonded energy terms correctly using thermodynamic integration. While in this study very elaborate protocols were used, precision and cycle closure errors on the order of $\pm k_BT$ (which should usually be sufficient) could be obtained with significantly shorter protocols (results not shown).

The first four lines in the right column of Table I list the alchemical single free energy differences. The corresponding gas phase (ΔA_3 and $\Delta A'_3$) and solution (ΔA_4 and $\Delta A'_4$) results differ, as they must since the endpoints (prop-d1 and prop-d2) are not identical because of the parameterization of the dummy atoms (*cf.* Sect. 2.3). The two hybrid molecules have unequal H—D bond lengths (*cf.* Sect. 3.1); thus, the physical origin of this difference has to be the respective Jacobian factor contributions. This can be demonstrated by a short analytical calculation. The difference between prop-d1 and prop-d2 is the equilibrium bond length of the three H—D bonds (1.111 *vs.* 0.5 Å, see (Sect. 3.1)). The corresponding free energy difference [39] is given by $\Delta A_J(d_1 \rightarrow d_2) = 3 \times k_BT \times \ln(1.111/0.5)^2 =$ 2.85 kcal/mole. This equals almost exactly the difference $\Delta A'_3 - \Delta A_3 =$ 2.88 ± 0.03 kcal/mole (gas phase), and the corresponding quantity in solution $\Delta A'_4 - \Delta A_4 = 3.19 \pm 0.34$ kcal/mole is very close. Comparing, on the other hand, $\Delta A'_2$ with ΔA_2 (left column in Tab. I), one sees that the dummy atoms do not influence chemical single free energy differences. In this step (transfer of a molecule from the gas phase into solution) only solute-solvent interactions are switched on. Since the dummy atoms of prop-d1 never participate in nonbonded interactions and their parameterization was chosen so that the properties of the physical system are not altered, their presence (compared to true propane) does not give rise to a free energy contribution.

The excellent agreement between the four $\Delta\Delta A_{solv}$ values demonstrates that double free energy differences are independent of the treatment of dummy atoms. This can be seen in three ways. First, the results obtained along the alchemical paths ($\Delta\Delta A_{III}$, $\Delta\Delta A_{IV}$) agree excellently with $\Delta\Delta A_I$. Since in the calculation of $\Delta\Delta A_I = \Delta A_2 - \Delta A_1$ no dummy atoms are involved, the independence of the overall result from dummy atoms is shown. Second, $\Delta\Delta A_I$ and $\Delta\Delta A_{II}$ are identical. While the first value was obtained with propane (ΔA_2), a hybrid propane with three dummy atoms identical to the end state of the alchemical calculations ΔA_3 and ΔA_4 was used in the calculations determining $\Delta A_2'$. The dummy atoms present in this path do not influence the result as was to be expected from the theoretical considerations of Sect. 2.3. Finally, the two free energy differences of hydration computed along two alchemical paths ($\Delta\Delta A_{III}$, $\Delta\Delta A_{IV}$) agree well. Here two parameterizations of the dummy atoms were compared (*cf.* Sect. 3.1); yet (as expected) they have no influence on the final result (the small difference of 0.3 kcal/mole is statistically not significant).

4. CONCLUDING DISCUSSION

In this paper a statistical mechanical treatment of the role of bonded interactions and of dummy atoms in free energy simulations was presented. The theoretical analysis (Sect. 2) permits one to derive many conclusions of Refs. [14–16] in a straightforward manner that is (almost) independent from methodological considerations. Three types of contributions resulting from changes of bond stretching and angle bending energy terms, Jacobian factor, vibrational and PMF-like, could be distinguished in alchemical, single free energy differences. Simultaneously, the theory demonstrates that Jacobian factor and vibrational contributions cancel to a good approximation from parallel steps in a thermodynamic cycle (*i.e.*, from double free energy differences). This underlines the need to distinguish carefully between single and double free energy differences when considering the role of bonded energy terms (as well as of dummy atoms).

There are two instances when the cancellation of Jacobian factor and vibrational contributions from double free energy differences may not be accurate: (i) Bonded terms are not treated (transmuted) identically in the two steps of the cycle. While this should never be necessary in practice, analytical techniques can be used to correct for the artificial Jacobian factor and/or vibrational contribution arising in this case (*cf.* Refs. [39] and [43], as well as the calculation of $\Delta A_J(d1 \rightarrow d2)$ in Sect. 3.2). (ii) As discussed in Sect. 2.2, the exact cancellation of Jacobian factor and vibrational free energy contribution from double free energy differences relies on the validity of the RR approximation and on the independence of the equilibrium geometry of the solute from solute–solvent and solvent–solvent interactions. Although fairly reasonable assumptions, possible problems and inaccuracies are best avoided by always including all free energy contributions from bonded energy terms in the calculations, as already advocated by Boresch and Karplus [14, 15].

While the separation of bonded contributions (and hence the cancellation of Jacobian factor and vibrational contributions) is only approximate (*cf.* Sect. 2.2 and previous paragraph), it was shown to be exact for those bonded energy terms involving dummy atoms (Sect. 2.3). This implies that with a correctly implemented program and some care concerning the parameterization of dummy atoms (*cf.* Sects. 2.3 and 3.1, as well as Ref. [16]), double free energy differences must be independent of the treatment of the dummy atoms. The numeric results for the iso-butane/propane system (Sect. 3.2) support this theoretical conclusion. (This casts doubts, *e.g.*, on data reported by Reddy *et al.* [22], where double free energy differences were found to depend systematically on the bond length to dummy atoms.) In addition, the independence of double free energy differences from dummy atoms also shows that identical results must be obtained in single and dual topology simulations of the same process. Again, the theoretical analysis permitted a much clearer demonstration of this equivalence compared with Refs. [14] and [15].

This work also contains numeric proof for some considerations given in Ref. [14]. While Boresch and Karplus discussed conceptual differences between chemical and alchemical paths, no calculations were reported for chemical paths in Ref. [15] (with the exception of an extremely simplified, one-dimensional model system). The single free energy differences obtained for the iso-butane/propane system (in particular, those of the thermodynamic cycle $\Delta A_1 + \Delta A_4 - \Delta A'_2 - \Delta A_3$ (see Fig. 1)) can be used immediately to support the considerations given in Sect. 2e of Ref. [14]. This becomes particularly clear by comparing the free energy components obtained along the chemical and alchemical paths (data available upon request). In Ref. [14] it was also mentioned that the free energy contributions of dummy atoms could be calculated analytically when needed. A nontrivial example of the necessary steps was given here in the appendix.

Together with the methodological aspects discussed in Refs. [14], the theoretical framework presented here should suffice to resolve any questions concerning the correct treatment of bonded energy terms in free energy simulations. Unfortunately, the situation found in practice is much less clear and depends heavily on the computer program used. As already mentioned above, results by Reddy *et al.* [22] indicate problems of the MacroModel program [52]. AMBER seems not to include vibrational and Jacobian factor contributions [17, 41, 42]. Based on the results of Ref. [20], the GASP algorithm of the BOSS program [53] correctly calculates vibrational contributions, but omits the Jacobian factor term. Since vibrational and Jacobian factor contributions cancel from double free energy differences (at least to a good approximation), neither of the last two examples is critical. Nevertheless, they illustrate the practical difficulties which one may face. For an unexperienced user (and even for an experienced one) it may be very disconcerting that an alchemical single free energy difference for the identical mutation (single topology calculation, identical treatment of dummy atoms and bonded energy terms) and employing the same force field, can be very different when calculated with, *e.g.*, AMBER [42], the PERT module of CHARMM [44], or BOSS [53]. In this context it has to be remembered that much of the confusion concerning the role of bonded free energy contributions was caused by an insufficient understanding of the differences between the various computer programs used, in particular, the consequences of employing a single or a dual topology methodology (see the Discussion in Ref. [15]). It would, therefore, be extremely helpful if the behavior of the various computer programs in widespread use with respect to their treatment of bonded energy terms in free energy simulations were documented. This should be a straightforward task based on the theoretical analysis presented here.

Acknowledgements

It is a pleasure to acknowledge several stimulating discussions with Martin Karplus. I thank Othmar Steinhauser for helpful comments concerning the manuscript. Michael Brunsteiner added the first version of the soft core code to CHARMM which was used in several of the calculations reported here. A conversation with Benoît Roux helped to clarify a number of points raised in Ref. [16].

APPENDIX: CALCULATION OF ΔA_D

The thermodynamic cycle depicted in Figure 1 contains a step labelled ΔA_D, the free energy difference between prop-d1 and propane plus the three detached dummy atoms in the gas phase. Since it involves the breaking of three bonds, it cannot be obtained from a simulation [14]; however, the free energy difference corresponding to this step can be calculated analytically using the techniques described in Refs. [36] and [39]. Since the dummy atoms in prop-d1 do not affect the properties of the propane molecule (Sect. 2.3 and numeric results), all calculations can be carried out in the gas phase. Although the free energy difference between prop and prop-d1 is not needed to determine $\Delta\Delta A_{solv}$, the calculation is an interesting, non-trivial application of the analytical techniques of HJR.

Our goal is the computation of the free energy difference for the step

$$\text{prop-d1} \longrightarrow \text{prop} + 3\,\text{D} \tag{18}$$

Following HJR, this free energy difference is best computed atom by atom, attaching one dummy atom at the time to propane. The first dummy atom D_1 is held in place by the H—D_1 bond r, the C_2—H—D_1 bond angle θ_1 and the C_1—C_2—H—D_1 dihedral angle τ. The free energy difference between the bonded D_1 and the free dummy atom is then given by [36]

$$\Delta A_{D_1} = -k_B T \ln \frac{r^2 \sin\theta_1 \sqrt{(2\pi k_B T)^3 / K_r K_{\theta_1} K_\tau}}{V_0}, \tag{19}$$

where V_0 is the volume available to the free atom corresponding to the standard state (*e.g.*, 1 mole/l), and K denotes the force constants. The contribution from the dihedral angle τ is approximated harmonically.

Dummy atom D_2 is connected to propane by a bond r (H—D_2) and two bond angles, the C_2—H—D_2 bond angle θ_1, and the D_1—H—D_2 bond angle θ_2. The free energy of attaching an atom in this manner is [36]

$$\Delta A_{D_2} = -k_B T \ln \frac{(r^2/\sin\psi) \sqrt{(2\pi k_B T)^3 / K_r K_{\theta_1} K_{\theta_2}}}{V_0} \tag{20}$$

The angle ψ in the Jacobian factor term is the angle between the two planes D_2—H—D_1 and D_1—H—C_2.

For the third dummy atom D_3 an analogous free energy difference has to be added. However, our parameterization of prop-d1 used in the simulations included an additional bond angle of type θ_2, *i.e.*, an D—H—D angle. In the MD simulations, this term is required for practical reasons to keep the three dummy atoms in a geometric arrangement resembling a methyl group; without it, very bad convergence was observed in the simulations. Since it is, however, a so-called redundant degree of freedom, it complicates the analytical calculation somewhat [36]. Using the redundancy relation for YCX_3-like molecules [38] permits to express the force matrix for the system (over-determined by the redundant third θ_2 angle) in terms of the non-redundant five other bond angles (three of type θ_1, two of type θ_2), and one finds for the (vibrational) free energy contribution of this additional angle term

$$\Delta A_{D_3,red} = \frac{k_B T}{2} \ln \frac{K_{\theta_1}^3 K_{\theta_2}^2}{3K_{\theta_1}^3 K_{\theta_2}^2 + 3K_{\theta_1}^2 K_{\theta_2}^3} \tag{21}$$

Thus, the full free energy difference for the process Eq. (18) is obtained as $\Delta A_D = \Delta A_{D_1} + 2\Delta A_{D_2} + \Delta A_{D_3,red}$.

Alternatively, the free energy difference can also be obtained by computing the translational, rotational and vibrational free energy of prop-d1 and prop, respectively (where the vibrational free energy has to obtained from a normal coordinate analysis) [30, 39]. When calculating the free energy difference by both methods (using the actual parameters), the results of the two approaches differed by not more than 0.02 kcal/mole (results available upon request).

References

[1] Kollman, P. (1993). "Free energy calculations: Applications to chemical and biochemical phenomena", *Chem. Rev.*, **93**, 2395.

[2] Kollman, P. A. (1996). "Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules", *Accounts Chem. Res.*, **29**, 461.

[3] Straatsma, T. P., "Free energy by molecular simulation", In: *Reviews in Computational Chemistry*, Lipkowitz, K. B. and Boyd, D. B. Eds., VCH, New York, 1996, **9**, 81 – 127.

[4] Kirkwood, J. G. (1935). "Statistical mechanics of fluid mixtures", *J. Chem. Phys.*, **3**, 300.

[5] Zwanzig, R. W. (1954). "High-temperature equation of state by a perturbation method. I. Nonpolar gases", *J. Chem. Phys.*, **22**, 1420.

[6] Simonson, T. (1993). "Free energy of particle insertion: An exact analysis of the origin singularity for simple liquids", *Mol. Phys.*, **80**, 441.

[7] Zacharias, M., Straatsma, T. P. and McCammon, J. A. (1994). "Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration", *J. Chem. Phys.*, **100**, 9025.

[8] Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. and van Gunsteren, W. F. (1994). "Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations", *Chem. Phys. Lett.*, **222**, 529.

[9] Pomès, R., Eisenmesser, E., Post, C. B. and Roux, B. (1999). "Calculating excess chemical potentials using dynamic simulations in the fourth dimension", *J. Chem. Phys.*, **111**, 3387.

[10] Brunsteiner, M. and Boresch, S. (2000). "Influence of the treatment of electrostatic interactions on the results of free energy calculations of dipolar systems", *J. Chem. Phys.*, **112**, 6953.

[11] Rozanska, X. and Chipot, C. (2000). "Modeling ion–ion interactions in proteins: A molecular dynamics free energy calculation of the guanidinum-acetate association", *J. Chem. Phys.*, **112**, 9691.

[12] Simonson, T. (2000). "Electrostatic free energy calculations for macromolecules: A hybrid molecular dynamics/continuum electrostatics approach", *J. Phys. Chem. B*, **104**, 6509.

[13] den Otter, W. K. and Briels, W. J. (1998). "The calculation of free-energy differences by constrained molecular-dynamics simulations", *J. Chem. Phys.*, **109**, 4139.

[14] Boresch, S. and Karplus, M. (1999). "The role of bonded terms in free energy simulations: 1. theoretical analysis", *J. Phys. Chem. A*, **103**, 103.

[15] Boresch, S. and Karplus, M. (1999). "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**, 119.

[16] Shobana, S., Roux, B. and Andersen, O. S. (2000). "Free energy simulations: Thermodynamic reversibility and variability", *J. Phys. Chem. B*, **104**, 5179.

[17] Pearlman, D. A. and Kollman, P. A. (1991). "The overlooked bond-stretching contribution in free energy perturbation calculations", *J. Chem. Phys.*, **94**, 4532.

[18] Straatsma, T. P., Zacharias, M. and McCammon, J. A. (1992). "Holonomic constraint contributions to free energy differences from thermodynamic integration molecular dynamics simulations", *Chem. Phys. Lett.*, **196**, 297.

[19] Wang, L. and Hermans, J. (1994). "Change of bond length in free-energy simulations: Algorithmic improvements, but when is it necessary", *J. Chem. Phys.*, **100**, 9129.

[20] Severance, D. L., Essex, J. W. and Jorgensen, W. L. (1995). "Generalized alteration of structure and parameters: A new method for free-energy perturbations in systems containing flexible degrees of freedom", *J. Comput. Chem.*, **16**, 311.

[21] Axelsen, P. H. and Li, D. (1998). "Improved convergence in dual-topology free energy calculations through use of harmonic restraints", *J. Comput. Chem.*, **19**, 1278.

[22] Reddy, M. R., Erion, M. D., Agarwal, A., Viswanadhan, V. N., McDonald, D. Q. and Still, W. C. (1998). "Solvation free energies calculated using the GB/SA model: Sensitivity of results on charge sets, protocols, and force fields", *J. Comput. Chem.*, **19**, 769.

[23] Prévost, M., Wodak, S. J., Tidor, B. and Karplus, M. (1991). "Contribution of the hydrophobic effect to protein stability: Analysis based on simulations of the Ile 96 → Ala mutation in barnase", *Proc. Natl. Acad. Sci. USA*, **88**, 10880.

[24] Elofsson, A. and Nilsson, L. (1993). "Free energy perturbations in Ribonuclease T_1 substrate binding. A study of the influence of simulation length, internal degrees of freedom and structure in free energy perturbations", *Mol. Simulation*, **10**, 255.

[25] Harris, D. and Loew, G. (1996). "Comparative study of free energies of solvation of phenylimidazole inhibitors of cytochrome P450cam by free energy simulation, AMSOL and Poisson Boltzmann methods", *J. Comput. Chem.*, **17**, 273.

[26] Rao, B. G., Kim, E. E. and Murcko, M. A. (1996). "Calculation of solvation and binding free energy differences between VX-488 and its analogs by free energy perturbation and AMSOL methods", *J. Comput.-Aided Mol. Design*, **10**, 23.

[27] Sun, Y.-C., Veenstra, D. L. and Kollman, P. A. (1996). "Free energy calculations of the mutation of Ile96 → Ala in barnase: contributions to the difference in stability", *Prot. Eng.*, **9**, 273.

[28] Pearlman, D. A. (1994). "A comparison of alternative approaches to free energy calculations", *J. Phys. Chem.*, **98**, 1487.

[29] Paci, E., Ciccotti, G., Ferrario, M. and Kapral, R. (1991). "Activation energies by molecular dynamics with constraints", *Chem. Phys. Lett.*, **176**, 581.

[30] McQuarrie, D. A., *Statistical Mechanics*, Harper and Row, New York, 1976.

[31] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995). "A second generation force field for the simulation of proteins and nucleic acids", *J. Am. Chem. Soc.*, **117**, 5179.

[32] van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P. and Tironi, I. G., Biomolecular Simulation: The GROMOS96 Manual and User Guide, vdf Hochschulverlag, Zürich, 1996.

[33] Jorgensen, W. L., Maxwell, D. S. and Tirado-Rives, J. (1996). "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids", *J. Am. Chem. Soc.*, **118**, 11225.

[34] MacKerell, A. D. Jr., Bashford, D., Bellott, M., Dunbrack, R. L. Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, III, W. E., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M. (1998). "All-atom empirical potential for molecular modeling and dynamics studies of protein", *J. Phys. Chem. B*, **102**, 3586.

[35] DeVoe, H., "Theory of the conformations of biological macromolecules in solution", In: *Structure and Stability of Biological Macromolecules*, Timascheff, S. N. and Fasman, G. D. Eds., Dekker, New York, 1969, **1**, 1–63.

[36] Herschbach, D. R., Johnston, H. S. and Rapp, D. (1959). "Molecular partition functions in terms of local properties", *J. Chem. Phys.*, **31**, 1652.

[37] Gō, N. and Scheraga, H. A. (1976). "On the use of classical statistical mechanics in the treatment of polymer chain conformation", *Macromolecules*, **9**, 535.

[38] Wilson, E. B. Jr., Decius, J. C. and Cross, P. C., Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra, McGraw-Hill, New York, 1955.

[39] Boresch, S. and Karplus, M. (1996). "The Jacobian factor in free energy simulations", *J. Chem. Phys.*, **105**, 5145.

[40] Tembe, B. L. and McCammon, J. A. (1984). "Ligand-receptor interactions", *Comput. Chem.*, **8**, 281.

[41] Pearlman, D. A. (1993). "Determining the contributions of constraints in free energy calculations: Development, characterization, and recommendations", *J. Chem. Phys.*, **98**, 8946.

[42] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K. and Kollman, P., AMBER 4.1, University of California, San Francisco, 1995.

[43] McCammon, J. A. (1998). "Theory of biomolecular recognition", *Curr. Op. Struct. Biol.*, **8**, 245.

[44] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983). "CHARMM: A program for macromolecular energy, minimization and dynamics calculations", *J. Comput. Chem.*, **4**, 187.

[45] Simonson, T. and Brünger, A. T. (1992). "Thermodynamics of protein–peptide interactions in the Ribonuclease-S system studied by molecular dynamics and free energy calculations", *Biochemistry*, **31**, 8661.

[46] Neria, E., Fischer, S. and Karplus, M. (1996). "Simulation of activation free energies in molecular systems", *J. Chem. Phys.*, **105**, 1902.

[47] Ryckaert, J. P., Ciccotti, G. and Berendsen, H. J. C. (1997). "Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes", *J. Comput. Phys.*, **23**, 327.

[48] Hermans, J. and Wang, L. (1997). "Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 Lysozyme", *J. Am. Chem. Soc.*, **119**, 2707.

[49] Hoover, W. G. (1985). "Canonical dynamics: Equlibrium phase-space distributions", *Phys. Rev.*, **A31**, 1695.

[50] Zachmann, H. G., Mathematik für Chemiker, Verlag Chemie, Weinheim, 1984.

[51] Ben-Naim, A. and Marcus, Y. J. (1984). "Solvation thermodynamics of nonionic solutes", *J. Chem. Phys.*, **81**, 2016.

[52] Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T. and Still, W. C. (1990). "MacroModel – an integrated software system for modeling organic and bioorganic molecules using molecular mechanics", *J. Comput. Chem.*, **11**, 440.

[53] Jorgensen, W. L., "BOSS – Biochemical and organic simulation system", In: *The Encyclopedia of Computational Chemistry*, Schleyer, P. v. R. Ed., Wiley, Athens, USA, 1998, **3**, 3281–3285.