Theor Chem Acc (2007) 117:535–540
DOI 10.1007/s00214-007-0264-z

REGULAR ARTICLE

Do all pieces make a whole? Thiele cumulants and the free energy decomposition

Matevž Bren · Jan Florián · Janez Mavri · Urban Bren

Received: 24 October 2006 / Accepted: 17 January 2007 / Published online: 21 February 2007

Abstract The partitioning of the free energy into additive contributions originating from different groups of atoms or force field terms has the potential to provide relationship between structure and biological activity of molecules. In this article, the theoretical foundation for the free energy decomposition in the free energy perturbation (FEP) methodology is formulated using Thiele cumulants, a powerful tool from the arsenal of probability theory and mathematical statistics. We establish that rigorous decomposition of the free energy into its components is precluded by the presence of mixed potential energy terms in Thiele cumulants of second and higher orders. However, we also show that the resultant non-additivity error can be reduced to an arbitrary value by increasing the number of FEP steps. Consequently, the whole system can be in the limit of small perturbation steps adequately represented by the sum of its constituent parts.

Keywords Free energy decomposition · Free energy perturbation method · Thiele cumulants

1 Introduction

Gibbs free energy G represents the thermodynamic potential of the $\Delta(N, P, T)$ ensemble. Therefore, it dictates the direction of processes at typical laboratory conditions of constant pressure P, constant temperature T, and constant number of particles N. Since living cells also operate at such conditions, G governs the direction of biochemical processes as well [1].

The free energy perturbation (FEP) method represents a robust methodology for calculating the free energy differences. It is based on a well known equation attributed to Zwanzig [2]:

$$\Delta G_{k\to k+1} = -\beta^{-1} \ln \left\langle \mathrm{e}^{-\beta \Delta E} \right\rangle_k \tag{1}$$

where $\Delta G_{k\to k+1}$ denotes the free energy difference between states k and $k+1$ that may correspond to different chemical entities. For example, DNA duplex containing thymine (state k) could be destabilized ($\Delta G_{k\to k+1}$) by substituting oxygen atoms of thymine by fluorine atoms to obtain difluorotoluene (state $k+1$) [3]. $\Delta E = E_{k+1} - E_k$ represents the corresponding potential energy difference. $\beta^{-1} = k_B T$, where k_B stands for the Boltzmann constant, and T is the thermodynamic temperature. Notation $\langle \cdots \rangle_k$ indicates averaging over the ensemble of configurations generated by

---

Electronic supplementary material The online version of this article (doi:10.1007/s00214-007-0264-z) contains supplementary material, which is available to authorized users.

M. Bren
Institute of Mathematics, Physics and Mechanics, Jadranska 19, 1000 Ljubljana, Slovenia

M. Bren
FOV, University of Maribor, Kidričeva 55a,
4000 Kranj, Slovenia

J. Florián
Department of Chemistry, Loyola University Chicago, Chicago, IL 60660, USA

J. Florián
Institute of Physics, Charles University, Ke Karlovu 5,
12116 Prague 2, Czech Republic

J. Mavri · U. Bren (✉)
National Institute of Chemistry, Hajdrihova 19,
1001 Ljubljana, Slovenia
e-mail: bren@cmm.ki.si

a molecular dynamics (MD) simulation on the potential energy surface of state k. Equation (1) rigorously gives excess Helmholz free energy, but the difference between Helmholz and Gibbs free energy is negligible in condensed systems.

The partitioning of the free energy into additive contributions originating from groups of atoms or force field terms has been applied as a foundation for many established methods of computer-aided drug design such as scoring functions for molecular docking [4], linear interaction energy [5,6], or linear response [7,8] approximations. Using classical empirical force fields such an additive decomposition is, while always valid for the potential energy, not generally attainable for the free energy [9]. The knowledge of the conditions required for the additive dissecting of the free energy is, however, indispensable for the understanding of the relationship between structure and biological activity of molecules, since all fragmental methods like structure-based de novo ligand design [10], three-dimensional pharmacophores [11], or coarse-grained models for polymer chains [12] owe their success to its validity.

In order to obtain accurate free energy differences, the states k and $k+1$ of (1) must overlap in phase space. These consecutive states should resemble enough to ensure adequate sampling of the phase space belonging to $k+1$ when simulating k. Thus, the overall change from the initial to the final state is usually attained in a series of L small steps—FEP windows [13]. In this case, states k and $k+1$ may represent hybrid structures between two different chemical entities. Such structures could, for example, contain hybrid atoms with charges and van der Waals parameters somewhere between oxygen atoms of thymine and fluorine atoms of difluorotoluene [3]. Since free energy is a state function, its difference depends only on the choice of the initial and final states and not on the selection of a path connecting them. Therefore, the overall free energy difference can be expressed as a sum of L partial free energy differences:

$$\Delta G = \sum_{k=1}^{L} \Delta G_{k\to k+1} \tag{2}$$

By increasing the number of steps we not only improve the overlap in the phase space of the consecutive states, but also decrease $|\Delta E|$ of an individual step. The main goal of this paper is to present the reflection of this fact in the expression for the free energy difference (1) and its effect on the additivity of the free energy. We are going to address these questions with the help of Thiele cumulants, a powerful tool from the arsenal of probability theory and mathematical statistics. These cumulants and related formulations have been successfully applied to investigate proton and electron transfers [14], solvation free energies [15], and infrared spectra of liquids [16,17], to determine temperature dependence of thermodynamic functions [18], and to analyze the LRA method [19]. The history and mathematical significance of Thiele cumulants are briefly described in supplementary materials. Generalized Thiele cumulants with operators as arguments of the exponential function are called Kubo cumulants and have been extensively applied to quantum-statistical mechanics [20].

2 Thiele cumulants

Formal Thiele cumulants are the coefficients of a double power series expansion [21]. In brief, we begin with a function f that can be expanded in the power series

$$f(x) = \sum_{n=0}^{\infty} \frac{\mu_n}{n!} x^n, \quad \text{with} \quad \mu_0 = 1. \tag{3}$$

We recall that the logarithm power series

$$\ln(1+u) = u - \frac{1}{2}u^2 + \frac{1}{3}u^3 - \cdots + (-1)^{n+1}\frac{1}{n}u^n + \cdots \tag{4}$$

converges for $-1 < u \leq 1$. Therefore, if we rewrite (3) in the form

$$f(x) = 1 + \sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n \tag{5}$$

and denote

$$u(x) \equiv \sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n \tag{6}$$

we can, for $-1 < \sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n \leq 1$ evaluate the power series of $\ln f$

$$\ln f(x) = u(x) - \frac{1}{2}u^2(x) + \frac{1}{3}u^3(x) - \cdots + (-1)^{n+1}\frac{1}{n}u^n(x) + \cdots \tag{7}$$

that can be expanded in powers of x

$$\ln f(x) = \sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n - \frac{1}{2}\left(\sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n\right)^2 + \frac{1}{3}\left(\sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n\right)^3 - \frac{1}{4}\left(\sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n\right)^4 + \frac{1}{5}\left(\sum_{n=1}^{\infty} \frac{\mu_n}{n!} x^n\right)^5 - \cdots \tag{8}$$

By reordering (8) in increasing powers of x we get the power series of the form

$$\ln f(x) = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} x^n \tag{9}$$

with coefficients κ_n called the Thiele cumulants.

To calculate the Thiele cumulants we will need coefficients of the powers of power series. Let us consider the power series S,

$$S = a\,x + b\,x^2 + c\,x^3 + d\,x^4 + e\,x^5 + \cdots \tag{10}$$

and calculate the powers of S

$$S^2 = a^2\,x^2 + 2ab\,x^3 + (b^2 + 2ac)x^4 + 2(ad + bc)x^5 + \cdots \tag{11}$$

$$S^3 = a^3\,x^3 + 3a^2b\,x^4 + 3(ab^2 + a^2c)x^5 + \cdots \tag{12}$$

$$S^4 = a^4\,x^4 + 4a^3b\,x^5 + \cdots \tag{13}$$

$$S^5 = a^5\,x^5 + \cdots \tag{14}$$

Applying (10) till (14) we expand (8) and get the first five Thiele cumulants

$$\kappa_1 = \mu_1 \tag{15}$$

$$\kappa_2 = \mu_2 - \mu_1^2 \tag{16}$$

$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \tag{17}$$

$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4 \tag{18}$$

$$\kappa_5 = \mu_5 - 5\mu_1\mu_4 - 10\mu_2\mu_3 + 30\mu_1\mu_2^2 + 20\mu_1^2\mu_3 - 60\mu_1^3\mu_2 + 24\mu_1^5 \tag{19}$$

Verifications for e^{-x} (S1) and e^{x} (S2) can be found in supplementary materials.

A general form of Thiele cumulants can also be provided [21]. Generally, we can express any integer n, ($n = 1, 2, 3\ldots$) as a sum

$$n = \sum_i i\,n_i \tag{20}$$

in several different ways. This is a well known problem of expressing an integer n as a sum of integers i where n_i is their frequency in the sum. For example, we can express integer 3 as a sum (20) in three different ways

$$3 = 1 \cdot 3$$

$$3 = 1 \cdot 1 + 2 \cdot 1$$

$$3 = 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 1$$

and, therefore, in this particular case we get three sequences of integers n_i that satisfy (20) for $n = 3$

$$(3), \quad (1,1), \quad (0,0,1) \tag{21}$$

with sums $\sum_i n_i$ equal to

$$3, \quad 1 + 1 = 2 \quad \text{and} \quad 0 + 0 + 1 = 1, \tag{22}$$

respectively. Based on these expressions we calculate the third Thiele cumulant

$$\kappa_3 = 3! \sum_{n_i} (-1)^{-1+\sum_i n_i} \left(-1 + \sum_i n_i\right)! \prod_i \left[\frac{(\mu_i/i!)^{n_i}}{n_i!}\right] \tag{23}$$

where we summarize the products over all three sequences (21). Indeed, with this expression we get

$$\kappa_3 = 3!\left((-1)^2\,2!\,\frac{(\mu_1/1!)^3}{3!} + (-1)^1\,1!\,\frac{(\mu_1/1!)^1}{1!}\,\frac{(\mu_2/2!)^1}{1!} + (-1)^0\,0!\,\frac{(\mu_1/1!)^0}{0!}\,\frac{(\mu_2/2!)^0}{0!}\,\frac{(\mu_3/3!)^1}{1!}\right) \tag{24}$$

that we reduce to

$$\kappa_3 = 3!\left(\frac{2!}{3!}\,\mu_1^3 - \frac{1}{2!}\mu_1\mu_2 + \frac{1}{3!}\,\mu_3\right) \tag{25}$$

and obtain the Thiele cumulant κ_3 equal to

$$\kappa_3 = 2\,\mu_1^3 - 3\mu_1\mu_2 + \mu_3 \tag{26}$$

Note that this result is identical to expression (17) .

Now we will write a general form of Thiele cumulants [21]

$$\kappa_n = n! \sum_{n_i} (-1)^{-1+\sum_i n_i} \left(-1 + \sum_i n_i\right)! \prod_i \left[\frac{(\mu_i/i!)^{n_i}}{n_i!}\right] \tag{27}$$

where we summarize the products over all sequences of integers that satisfy (20). Verifications for $n = 2$ (S3) and $n = 1$ (S4) can be found in supplementary materials.

3 Free energy decomposition

If we expand the exponential function from the expression for the free energy difference (1) into a Taylor series and apply the fact that the average of a sum equals the sum of its averaged summands, we obtain

$$\Delta G_{k\to k+1} = -\beta^{-1} \ln\left[1 + \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \left\langle (\Delta E)^n \right\rangle_k\right] \tag{28}$$

If we compare the expression in the square brackets with $f(x)$ from (5) the following identities can be established:

$$x \equiv -\beta \tag{29}$$

and

$$\mu_n \equiv \left\langle (\Delta E)^n \right\rangle_k \tag{30}$$

In analogy with (9) we can write

$$\Delta G_{k\to k+1} = -\beta^{-1}\sum_{n=1}^{\infty}\frac{\kappa_n}{n!}(-\beta)^n = \sum_{n=1}^{\infty}\frac{\kappa_n}{n!}(-\beta)^{n-1} \qquad (31)$$

where κ_n is a Thiele cumulant defined in (27) with μ_n given by (30).

The potential energy difference in effectively polarized force fields always originates from various groups of atoms or types of interactions:

$$\Delta E = \sum_{i=1}^{m}\Delta E_i \qquad (32)$$

However, in the case of polarizable force fields or quantum-chemical calculations such potential energy decomposition is not valid [22]. Corresponding free energy components can be defined using (28) till (31) as

$$\Delta G^{i}_{k\to k+1} = -\beta^{-1}\ln\left\langle \mathrm{e}^{-\beta\Delta E_i}\right\rangle_k = \sum_{n=1}^{\infty}\frac{\kappa_{n,i}}{n!}(-\beta)^{n-1} \qquad (33)$$

where $\kappa_{n,i}$ is a Thiele cumulant determined via (27) with $\mu_{n,i}$ given by

$$\mu_{n,i} \equiv \left\langle(\Delta E_i)^n\right\rangle_k \qquad (34)$$

The nonadditivity error of such a free energy decomposition can be determined as

$$\begin{aligned}\Delta G^{\mathrm{NAD}}_{k\to k+1} &= \Delta G_{k\to k+1} - \sum_{i=1}^{m}\Delta G^{i}_{k\to k+1}\\ &= \sum_{n=1}^{\infty}\frac{\kappa_n - \sum_{i=1}^{m}\kappa_{n,i}}{n!}(-\beta)^{n-1} = \sum_{n=1}^{\infty}\Delta G^{NAD}_{k\to k+1,n}\end{aligned} \qquad (35)$$

There is no nonadditivity error of the first order since using relations (15) , (30), (32), (34), and (35) we obtain

$$\begin{aligned}\Delta G^{\mathrm{NAD}}_{k\to k+1,1} &= \mu_1 - \sum_{i=1}^{m}\mu_{1,i} = \left\langle\sum_{i=1}^{m}\Delta E_i\right\rangle_k\\ &\quad - \sum_{i=1}^{m}\langle\Delta E_i\rangle_k = 0\end{aligned} \qquad (36)$$

Expressions (16) , (30), (32), (34), and (35) are applied to determine the nonadditivity error of the second order:

$$\begin{aligned}\Delta G^{\mathrm{NAD}}_{k\to k+1,2}\\ &= -\frac{\beta}{2}\left[\mu_2 - \mu_1^2 - \sum_{i=1}^{m}\left(\mu_{2,i}-\mu_{1,i}^2\right)\right]\\ &= -\frac{\beta}{2}\left[\left\langle\left(\sum_{i=1}^{m}\Delta E_i\right)^2\right\rangle_k - \left\langle\sum_{i=1}^{m}\Delta E_i\right\rangle_k^2\right.\\ &\quad\left. - \sum_{i=1}^{m}\left(\left\langle(\Delta E_i)^2\right\rangle_k - \langle\Delta E_i\rangle_k^2\right)\right]\\ &= -\frac{\beta}{2}\left[\sum_{i=1}^{m}\left\langle(\Delta E_i)^2\right\rangle_k + 2\sum_{i=1}^{m}\sum_{j>i}\left\langle\Delta E_i\Delta E_j\right\rangle_k\right.\\ &\quad - \sum_{i=1}^{m}\langle\Delta E_i\rangle_k^2\\ &\quad -2\sum_{i=1}^{m}\sum_{j>i}\langle\Delta E_i\rangle_k\left\langle\Delta E_j\right\rangle_k - \sum_{i=1}^{m}\left\langle(\Delta E_i)^2\right\rangle_k\\ &\quad\left. + \sum_{i=1}^{m}\langle\Delta E_i\rangle_k^2\right]\\ &= \beta\sum_{i=1}^{m}\sum_{j>i}\left(\langle\Delta E_i\rangle_k\left\langle\Delta E_j\right\rangle_k - \left\langle\Delta E_i\Delta E_j\right\rangle_k\right)\end{aligned} \qquad (37)$$

Note that the nonadditivity error of the second and higher orders arises from the appearance of mixed potential energy terms in the expression (28) for the free energy difference. On the other hand, these mixed terms can never be generated in a sum of all free energy components given by (33).

Now, let us focus on nonmixed potential energy terms. The general summand of the nth Thiele cumulant κ_n (27) is of the form

$$\mathrm{Smd}_n \equiv A\,\mu_{r_1}^{p_1}\,\mu_{r_2}^{p_2}\,\cdots\mu_{r_\ell}^{p_\ell} \qquad (38)$$

where A is a numerical constant and

$$\sum_{i=1}^{\ell} r_i p_i = n \qquad (39)$$

If we revisit the fifth Thiele cumulant (19)

$$\begin{aligned}\kappa_5 &= \mu_5 - 5\mu_1\mu_4 - 10\mu_2\mu_3 + 30\mu_1\mu_2^2 + 20\mu_1^2\mu_3\\ &\quad -60\mu_1^3\mu_2 + 24\mu_1^5\end{aligned}$$

and take the fourth summand as an example, we indeed see that

$$A = 30,\quad r_1 = 1,\quad p_1 = 1,\quad r_2 = 2,\quad p_2 = 2\quad\text{and}$$
$$\sum_{i=1}^{2} r_i p_i = 1\cdot 1 + 2\cdot 2 = 5$$

Inserting expressions (30) and (32) into the general summand (38), we obtain

$$
\begin{aligned}
\mathrm{Smd}_n &= \\
& A\left(\left\langle\left[\sum_{i=1}^{m}\Delta E_i\right]^{r_1}\right\rangle_k\right)^{p_1}\cdot\left(\left\langle\left[\sum_{i=1}^{m}\Delta E_i\right]^{r_2}\right\rangle_k\right)^{p_2}\cdots \\
&\times\left(\left\langle\left[\sum_{i=1}^{m}\Delta E_i\right]^{r_\ell}\right\rangle_k\right)^{p_\ell} \\
&= A\left(\left\langle\sum_{i=1}^{m}[\Delta E_i]^{r_1}+\cdots\right\rangle_k\right)^{p_1}\cdot\left(\left\langle\sum_{i=1}^{m}[\Delta E_i]^{r_2}+\cdots\right\rangle_k\right)^{p_2}\cdots \\
&\times\left(\left\langle\sum_{i=1}^{m}[\Delta E_i]^{r_\ell}+\cdots\right\rangle_k\right)^{p_\ell}
\end{aligned}
$$

Since we are interested only in the nonmixed terms, all the mixed summands are denoted by three dots. Once again, because the average of a sum equals the sum of its averaged summands, we continue:

$$
\begin{aligned}
&= A\left(\sum_{i=1}^{m}\left\langle[\Delta E_i]^{r_1}\right\rangle_k+\cdots\right)^{p_1}\cdot\left(\sum_{i=1}^{m}\left\langle[\Delta E_i]^{r_2}\right\rangle_k+\cdots\right)^{p_2}\cdots \\
&\times\left(\sum_{i=1}^{m}\left\langle[\Delta E_i]^{r_\ell}\right\rangle_k+\cdots\right)^{p_\ell}
\end{aligned}
$$

Considering only the nonmixed terms, we proceed:

$$
\begin{aligned}
&= A\left(\sum_{i=1}^{m}\left(\left\langle[\Delta E_i]^{r_1}\right\rangle_k\right)^{p_1}+\cdots\right)\cdot\left(\sum_{i=1}^{m}\left(\left\langle[\Delta E_i]^{r_2}\right\rangle_k\right)^{p_2}+\cdots\right)\cdots \\
&\times\left(\sum_{i=1}^{m}\left(\left\langle[\Delta E_i]^{r_\ell}\right\rangle_k\right)^{p_\ell}+\cdots\right)
\end{aligned}
$$

After multiplication, we get

$$
\begin{aligned}
&= A\sum_{i=1}^{m}\left(\left\langle[\Delta E_i]^{r_1}\right\rangle_k\right)^{p_1}\cdot\left(\left\langle[\Delta E_i]^{r_2}\right\rangle_k\right)^{p_2}\cdots\left(\left\langle[\Delta E_i]^{r_\ell}\right\rangle_k\right)^{p_\ell} \\
&\quad+\text{mixed terms} \qquad (40)
\end{aligned}
$$

The general summand of the nth Thiele cumulant for the ith free energy component $\kappa_{n,i}$ (33) is of the form

$$
\mathrm{Smd}_{n,i} \equiv A\,\mu_{r_1,i}^{p_1}\,\mu_{r_2,i}^{p_2}\cdots\mu_{r_\ell,i}^{p_\ell} \qquad (41)
$$

Inserting expression (34) into the general summand (41), we obtain

$$
\begin{aligned}
\mathrm{Smd}_{n,i} &= A\,\left(\left\langle[\Delta E_i]^{r_1}\right\rangle_k\right)^{p_1}\cdot\left(\left\langle[\Delta E_i]^{r_2}\right\rangle_k\right)^{p_2} \\
&\quad\times\cdots\left(\left\langle[\Delta E_i]^{r_\ell}\right\rangle_k\right)^{p_\ell} \qquad (42)
\end{aligned}
$$

Comparing results (40) and (42), we gain

$$
\mathrm{Smd}_n = \sum_{i=1}^{m}\mathrm{Smd}_{n,i} + \text{mixed terms}
$$

The same nonmixed potential energy terms are, therefore, present in the expression for the free energy difference and in the sum of its free energy components.

The perturbation from the initial state I to the final state F can be performed either

- directly in a single perturbation step with the total potential energy difference ΔE consisting of components $\Delta E_1, \Delta E_2, \ldots, \Delta E_m$ and the nonadditivity error of the second order given by (37)

$$
\Delta G_{I\to F,2}^{\mathrm{NAD},1} = \beta\sum_{i=1}^{m}\sum_{j>i}\left(\langle\Delta E_i\rangle_k\left\langle\Delta E_j\right\rangle_k - \left\langle\Delta E_i\Delta E_j\right\rangle_k\right)
$$

- or step-wise in a series of L equal perturbation steps. An average step has the potential energy difference of $\Delta E/L$ constructed from the components $\Delta E_1/L, \Delta E_2/L, \ldots, \Delta E_m/L$ and the nonadditivity error of the second order defined by (37)

$$
\begin{aligned}
\Delta G_{k\to k+1,2}^{NAD} &= \beta\sum_{i=1}^{m}\sum_{j>i}\left(\left\langle\frac{\Delta E_i}{L}\right\rangle_k\left\langle\frac{\Delta E_j}{L}\right\rangle_k\right. \\
&\qquad\left.-\left\langle\frac{\Delta E_i}{L}\frac{\Delta E_j}{L}\right\rangle_k\right) \\
&= \frac{\beta}{L^2}\sum_{i=1}^{m}\sum_{j>i}\left(\langle\Delta E_i\rangle_k\left\langle\Delta E_j\right\rangle_k - \left\langle\Delta E_i\Delta E_j\right\rangle_k\right) \\
&= \frac{\Delta G_{I\to F,2}^{\mathrm{NAD},1}}{L^2} \qquad (43)
\end{aligned}
$$

The overall nonadditivity error of the second order obtained on a step-wise perturbation path $\Delta G_{I\to F,2}^{\mathrm{NAD},L}$ is, consequently, L-times the nonadditivity error of the second order for an average perturbation step:

$$
\Delta G_{I\to F,2}^{\mathrm{NAD},L} = L\times\Delta G_{k\to k+1,2}^{\mathrm{NAD}} = \frac{\Delta G_{I\to F,2}^{\mathrm{NAD},1}}{L} \qquad (44)
$$

Therefore, $\Delta G_{I\to F,2}^{\mathrm{NAD},L}$ is L-times smaller than the nonadditivity error of the second order attained on a direct perturbation path $\Delta G_{I\to F,2}^{\mathrm{NAD},1}$. Analogously, since the nonadditivity error of the nth order is given by a sum of products of n mixed ΔE_i-s, we can decrease this error by a factor of L^{n-1}, simply, by performing the perturbation $I \to F$ in a series of L equal steps rather than in a single one:

$$
\begin{aligned}
\Delta G_{I\to F,n}^{\mathrm{NAD},L} &= L\times\Delta G_{k\to k+1,n}^{\mathrm{NAD}} = L\times\frac{\Delta G_{I\to F,n}^{\mathrm{NAD},1}}{L^n} \\
&= \frac{\Delta G_{I\to F,n}^{\mathrm{NAD},1}}{L^{n-1}} \qquad (45)
\end{aligned}
$$

4 Conclusions

The partitioning of the free energy into additive contributions originating from different groups of atoms or force field terms has the potential to provide relationship between structure and biological activity of molecules. Therefore, free energy decomposition is used as a foundation for many established methods of computer-aided drug design such as scoring functions for molecular docking, linear interaction energy, or linear response approximations. All fragment-based approaches owe their success to its validity as well. Whether such a free energy partitioning is justified has been, however, vigorously debated [23–26], with thermodynamic integration [27] serving as a key theoretical tool for this discussion.

In this paper, we set up the theoretical basis for the free energy decomposition in the free energy perturbation (FEP) formalism by using the Thiele cumulants, a powerful tool from the arsenal of probability theory and mathematical statistics. The authors find it fascinating, how two different phenomena so distant as rolling a pair of dice in a casino and the partitioning of the free energy can be described by the same mathematical laws—the Thiele cumulants. Applying the Zwanzig's perturbation formula we established that every dissection of the free energy into its components possesses an inherent error (called the 'nonadditivity error') arising solely from the appearance of mixed potential energy terms in Thiele cumulants of second and higher orders. Fortunately, it is possible to decrease this error to an arbitrary value just by increasing the number of perturbation steps. For example, by using a total of 51 FEP steps nonadditivity of less than 0.02 kcal/mol for solvation free energies of deoxyribonucleoside triphosphates was achieved [9]. Thus, the whole molecule can be adequately represented by the sum of its constituent parts in the limit of a small perturbation step. However, it should be noted that this conclusion cannot be generalized to polarizable force fields or quantum-chemical calculations, since they do not yield additive potential energy components.

Furthermore, one has to be aware that the free energy components are not state functions as only their sum is a state function. Consequently, they should be calculated on the most natural perturbation path possible. The free energy decomposition scheme presented in this article can be combined with the recent suggestion of such a perturbation path [28] to form a powerful theoretical tool for rational drug design and protein engineering [29].

Acknowledgments The authors would like to thank Dušanka Janežič from National Institute of Chemistry for careful reading of the manuscript and helpful discussions. Financial support from the Slovenian Ministry of Science and Higher Education through grant P1-0012 is gratefully acknowledged, along with the NIH grant 1U19CA105010.

References

1. Atkins PW (2000) Physical chemistry, 6th edn. Oxford University Press, Oxford
2. Zwanzig RW (1954) J Chem Phys 22:1420-1426
3. Florián J, Goodman MF, Warshel A (2000) J Phys Chem B 104:10092–10099
4. Böhm H-J, Klebe G (1996) Angew Chem Int Ed 35:2588–2614
5. Hansson T, Marelius J, Åqvist J (1998) J Comput Aided Mol Des 12:27–35
6. Åqvist J, Medina C, Samuelson JE (1994) Protein Eng 7:385–391
7. Florián J, Goodman MF, Warshel A (2002) J Phys Chem B 106:5739–5753
8. Lee FS, Chu Z-T, Bolger MB, Warshel A (1992) Protein Eng 5:215–228
9. Bren U, Martínek V, Florián J (2006) J Phys Chem B 110:12782–12788
10. Lewis RM, Leach AR (1994) J Comput-Aided Mol Des 8:467–475
11. Patani GA, La Voie EJ (1996) Chem Rev 96:3147–3176
12. Baschnagel J, Binder K, Paul W, Laso M, Suter U, Batoulis I, Jilge W, Bürger T (1991) J Chem Phys 95:6014–6025
13. Leach AR (2001) Molecular modelling: principles and applications, 2nd edn. Prentice Hall, Harlow
14. Warshel A, Hwang J-K (1986) J Chem Phys 84:4938–4957
15. Vaidehi N, Wesolowski TA, Warshel A (1992) J Chem Phys 97:4264–4271
16. Bratož S, Rios J, Guissani Y (1970) J Chem Phys 52:439–453
17. Borštnik B, Pumpernik D, Ažman A (1974) Vibrational and reorientational relaxation of hydrogen-bonded systems. In: Lascombe J (ed) Molecular motions in liquids, D. Reidel Publishing Company, Dordrecht, pp 241–246
18. Apol MEF, Amadei A, Berendsen HJC, Di Nola A (1999) J Chem Phys 111:4431–4441
19. Hummer G, Szabo A (1996) J Chem Phys 105:2004–2010
20. Kubo R (1962) J Phys Soc Jpn 17:1100–1120
21. Cramér H (1946) Mathematical methods of statistics. Princeton University Press, Princeton
22. Jordan PC, van Maaren PJ, Mavri J, van der Spoel D, Berendsen HJC (1995) J Chem Phys 103:2272–2285
23. Boresch S, Archontis G, Karplus M (1994) Proteins Struct Funct Genet 20:25–33
24. Boresch S, Karplus M (1995) J Mol Biol 254:801–807
25. Smith PE, van Gunsteren WF (1994) J Phys Chem 98:13735–13740
26. Brady GP, Sharp KA (1995) J Mol Biol 254:77–85
27. Mitchell MJ, McCammon JA (1991) J Comput Chem 12:271–275
28. Bren U, Martínek V, Florián J (2006) J Phys Chem B 110:10557–10566
29. Merzel F, Hodošček M, Janežič D, Sanson A (2006) J Comput Chem 27:446–452