PCCP

The Formulation of Chemical Potentials and Free Energy Changes in Biochemical Reactions

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-05-2021-002045
Article Type:	Paper
Date Submitted by the Author:	09-May-2021
Complete List of Authors:	Cannon, William; Pacific Northwest National Laboratory, ; University of California Riverside, Mathematics Raff, Lionel; Oklahoma State University, Chemistry

ARTICLE

The Formulation of Chemical Potentials and Free Energy Changes in Biochemical Reactions

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

William R. Cannon, [a]* and Lionel M. Raff [b]

In 1994, an IUBMB-IUPAC joint committee recommended a revised formulation for standard chemical potentials and reaction free energies motivated by the fact that, in biochemistry, the reactants and products often exist in multiple charge states depending on the pH and pMg of the solution environment. The recommendation involved both the use of (1) a mathematical transform with the intent to hold the pH constant, and (2) the formulation of reference chemical potentials of ionized isomeric species based on the log sum of the individual standard chemical potentials of each isomeric species. Recently, several reports including a 2020 IUPAC report have appeared that challenged the need for such summary formulations, arguing that the standard chemical potentials were sufficient with full accounting of each of the different charge state isomers involved in a biochemical reaction. This work critically evaluates both the use of thermodynamic transforms and the different chemical potential formulations. It is shown that (1) transforms are not necessary to hold the pH constant and (2) demonstrates that the two chemical potential formulations are not equivalent. Which formulation is appropriate depends on what species are measured experimentally or whether an assumption of equilibrium among the charge state isomers is reasonable and desirable.

Introduction

The laws of thermodynamics are arguably the most important physical laws needed to understand biology, from the operation of metabolism [1-4] to natural selection [5-7]. While thermodynamics provides the principles of how biological systems work as dissipative systems, detailed experimentation uncovers the mechanisms that implement these principles and constrain the solution space. Mechanisms constrain the solution space to only those processes that are physically feasible, not necessarily thermodynamically optimal. This constraint means that, while organisms seek to optimize entropy production rates, they can never attain a true optimum in the thermodynamic sense. Instead, they compete to find better mechanisms that increase their entropy production rate and fitness. Understanding mechanisms that are both widely used in nature and that limit growth are important areas of research. Yet, technical knowledge of a mechanism without understanding the thermodynamic principles as to why the process is needed does not lead to a predictive understanding.

As such, determining the thermodynamics of biochemical reactions is essential for understanding how and why biological systems operate and why natural selection has chosen particular solutions. Not only does thermodynamics tell us which reactions are probable and which are not, but thermodynamics is also an inferential tool that provides a reliable way to estimate parameters needed for modeling systems when measurements are too time consuming, expensive or simply not technically possible [4, 8, 9]. Moreover, statistical thermodynamics is the foundation for information theory [10] and formulations of both human [11] and machine learning (aka, energy-based models).

As a result of the growing recognition of the central importance of thermodynamics in biology, there has been considerable focus on appropriate methods to calculate free energies of reaction for biochemical systems [12-18]. This work has led to significant biological insight [19-22].

Nevertheless, a fundamental, controversial issue and point of contention is how to calculate the standard free energy of reaction when (1) the system is constrained to a specified pH and (2) the reactions involve one or more reactants or products that exists in several ionic states. The first issue involves the spontaneous ionization of bulk water that results in low concentrations of hydronium and hydroxide ions, and in a pH greater than zero. In cells and in laboratory experiments the pH is regulated either by homeostasis or added buffer, respectfully, and as such is controlled such that the pH is effectively constant. When analyzing experimental data or simulating *in vivo* conditions, it is important for these conditions to be reflected in formulations of free energies and entropy. To hold the pH constant in these formulations, it has

[a] *William.cannon@pnnl.gov.*
[b] *Lionel.raff@okstate.edu.*
Electronic Supplementary Information (ESI) available: Code for the calculations used in generating the tables and figure are available as computational notebooks at https://github.com/wrcannon/CompositeReactionFreeEnergies. See DOI: 10.1039/x0xx00000x

been proposed, and accepted by IUPAC and IUBMB, that a mathematical transform is necessary [12, 23, 24].

The second issue is that at a specified pH, chemical species such as metabolites spontaneously ionize to different charge states. For instance, ATP potentially forms at least six species, neutral ATP, ATP^{-1}, ATP^{-2}, ATP^{-3}, $ATP^{-2}Mg^{+2}$, $ATP^{-3}Mg^{+2}$. These different ionic forms can be considered charge state isomers and have been referred to as pseudo-isomers by Alberty *et al.* [23]. When one considers a reaction involving multiple species such as this, the challenge is to then use the appropriate thermodynamic formulation for the change in free energy [12, 23, 25].

For example, a chemical equation for the elementary reaction catalyzed by pyruvate decarboxylase is,

$$CH_3COCO_2^- + H^+ \leftrightharpoons CH_3CHO + CO_2. \quad 1$$

The equilibrium constant for the reaction is given by the ratio of ratios of the equilibrium concentrations, $[\cdots]_{eq}$, to a reference concentration, $[\cdots]^\circ$, of the products to the reactants,

$$\mathrm{K} = \frac{\frac{[CH_3CHO]_{eq}}{[CH_3CHO]^\circ} \cdot \frac{[CO_2]_{eq}}{[CO_2]^\circ}}{\frac{[CH_3COCO_2^-]_{eq}}{[CH_3COCO_2^-]^\circ} \cdot \frac{[H^+]_{eq}}{[H^+]^\circ}}. \quad 2$$

In this description, each reference concentration, $[\cdots]^\circ$ is typically the standard reference concentration of 1 M. Exceptions occur frequently, however. Solvent reference concentrations are typically the concentration of the bulk solvent, such as 55.5 M for water. Reference concentrations for protons may be 10^{-7} M (pH = 7.0) or whatever pH at which the system of interest is typically studied. Because a chemical equation such as Eqn. *1* is written with respect to specific ionic states, the determination of the equilibrium constant, such as Eqn *2*, is straight-forward.

The analogous biochemical equation is typically written as,

$$\mathrm{pyruvate} \rightleftharpoons \mathrm{acetaldehyde} + CO_2. \quad 3$$

At any specified pH, two ionic forms of pyruvate coexist, $CH_3COCO_2^-$ and CH_3COCO_2H. By analogy to the equilibrium constant for elementary reactions, an equilibrium constant for the composite reaction of Eqn *3* can be written as,

$$K = \frac{\frac{[\mathrm{acetaldehyde}]_{eq}}{[\mathrm{acetaldehyde}]^\circ} \cdot \frac{[CO_2]_{eq}}{[CO_2]^\circ}}{\frac{[\mathrm{pyruvate}]_{eq}}{[\mathrm{pyruvate}]^\circ}}, \quad 4$$

where $[\mathrm{pyruvate}] = [CH_3COCO_2^-] + [CH_3COCO_2H]$. In the language introduced by Alberty, *et al*., [12, 24, 25] these different ionic species of pyruvate are known as pseudo-isomers. In the case of carbon dioxide, we will assume for the sake of demonstration that the only species is CO_2 and not any of the ionic states of carbonates or carbonic acid.

Questions have arisen as to whether an ionic state-specific chemical formulation, *e.g.*, Eqns *1* and *2*, should be used to determine the thermodynamics of reactions in metabolism or whether the composite approach, *e.g.* Eqns. *3* and *4*, advocated in a joint IUBMB-IUPAC recommendation in 1994 [12, 23, 24] should be used, and whether the two approaches give numerically the same result, as claimed in a 2020 IUPAC report [26] and elsewhere [16, 17]. Potentially adding confusion to both issues, when thermodynamic properties for use in biological systems were discussed in the studies [12, 23, 24], the issues were often addressed all together: the use (1) composite species (that is, pyruvate vs. CH_3COCO_2H *or* $CH_3COCO_2^-$), (2) the use of constant pH (and ionic strength) and (3) the use of mathematical transforms. As such, one might not realize that they are independent issues.

This report critically evaluates both the use of transforms which have been proposed to be necessary to hold the system at a specific pH and the two different formulations for calculating biochemical reaction free energies, specifically the method endorsed in 1994 [12, 25, 27], referred to as the Alberty method since it was Alberty who first developed it, and the method proposed by Iotti, Sabatini, *et. al.* [16, 17], referred to as the Balanced Biochemical Reaction or BBR method. Regarding the first issue of constant pH and transforms, it is shown that a Legendre transform is not necessary, nor does it make holding the pH constant more convenient. Instead, this is achieved by the use of a thermodynamic bath that is external to the reacting system of interest.

Regarding the second issue of formulation of chemical potentials for ionic species, it is shown from the perspective of statistical thermodynamics that the Alberty and BBR methods do not have the same intent and give different results. In most cases the differences are small (less than 2.0 kJ/mol) and are due to whether equilibrium is assumed among the pseudo-isomers, or not. While both methods are correct, which method is appropriate to use depends on the nature of the experimental or *in vivo* conditions that are being observed.

Theory

Transforms of Free energy and constant pH Systems.

In the discussion that follows, only transforms involving pH are considered, but transforms involving pMg or any other species are treated analogously. To address the issue of formulating free energies at constant pH, it has become common practice to employ a procedure developed by Alberty [28]. In this procedure, a transformation of the Gibbs free G energy to a new function, F_A, is defined. This transformation involves the chemical potentials and

amounts of both the free and bound protons. The number of free and bound protons for a species j is (using Alberty's notation) $N_H(j)$. If the number of each species j is n_j and the chemical potential of a proton is μ_{H^+}, then the Alberty transform is,

$$F_A = G(T, P, n_1, \dots, n_{N_S}) - \sum_{j=1}^{N_S} N_H(j) n_j \mu_{H^+} . \quad 5$$

First, a transform similar to Alberty's is described but in which only the protons free in solution are employed in the transform. It is shown that this transform is a Legendre transform of the Gibbs energy of an open system. Next, the proposed IUPAC-IUBMB transform (aka, Alberty's transform) is investigated using the additional requirement that the transform include both free and bound protons. It is clear that the contribution of the bound protons does not cancel out due to either taking the derivative in a Legendre transform nor necessarily in the integration of the differential free energy. In fact, as written, the Alberty transform gives incorrect results. However, regardless of whether one uses a true Legendre transform, a Legendre transform does not hold the concentration of protons free in solution constant. Instead, it is necessary to couple the system to an external bath of protons and then consider the total free energy of the system plus the bath.

A Legendre Transform of *G* for Biochemical Systems. First, it is essential to carefully define all symbols to avoid ambiguity. N_S is the total number of species in a system. In this definition, for example, PO_4^{3-}, HPO_4^{2-}, $H_2PO_4^{-}$, and H_3PO_4 are all separate species. As such, the total number of particles in the system is $N_{total} = \sum_j^{N_S} n_j$ where n_j is the count/abundance of species j. It is worth noting at this point that water and any other solvent is included as a species.

The total Gibbs free energy is given by,

$$G(T, P, n_1, \dots, n_{N_S}) = \sum_{j=1}^{N_S} \mu_j n_j. \quad 6$$

Since the system is assumed to contain free protons in solution, free protons are included as a species in N_S in Eqn 6.

An infinitesimal change in the Gibbs free energy at constant value of the independent variables T, P is,

$$dG(T,\ P, n_1, \dots, n_{N_S}) = \sum_{j=1}^{N_S} \mu_j dn_j . \quad 7$$

Let,

$$F = G - \sum_{j=1}^{N_S} N_H(j) n_j \mu_{H^+}. \quad 8$$

As defined, F is similar to Alberty's function. However, in this equation $N_H(j)$ is the number/count of protons that have dissociated from the neutral species to form the ionic species j. That is, $N_H(j)$ is the number of protons released to solution to form ionic species j. To give a concrete example, consider the dissociation of H_nA into the ionic species $H_{n-2}A^{2-}$,

$$H_nA \rightleftharpoons 2H^+ + H_{n-2}A^{2-}. \quad 9$$

In this case, species j is $H_{n-2}A^{2-}$ and $N_H(j) = 2$. Note that as a consequence of the conservation of mass, the sum over all species j similar to $H_{n-2}A^{2-}$ including water or other solvents, $\sum_{j=1}^{N_S-1} N_H(j) n_j \mu_{H^+}$, is the total number of protons free in solution (The sum only goes to $N_S - 1$ because the species H^+ has no titratable protons). In other words, in a closed system the conservation of mass dictates that the number of free protons in solution is equal to the number of ionization events due to release of the proton from the neutral species,

$$\sum_{j \neq H^+}^{N_S-1} N_H(j) n_j = n_H^+. \quad 10$$

However, we will not initially impose this constraint because we want to use conjugate variables of concentrations such that the conjugate variables are the pairs,

$$n_j, \frac{\partial G}{\partial n_j},$$

and since the derivative is a partial derivative, it is required that each chemical species j is independent of other species. In contrast, the constraint of Eqn 10 enforces a dependency.

We now address the transform,

$$\tilde{G} = G - \mu_{H^+} n_{H^+}.$$

As shown below, this leads to the Legendre transform that replaces $\mu_{H^+} dn_{H^+}$ with $n_{H^+} d\mu_{H^+}$ while all other $N_S - 1$ chemical potentials μ_i are constant is,

$$\begin{aligned} d\tilde{G} &= d(G - \mu_{H^+} n_{H^+}) \\ &= \sum_{j=1}^{N_S} \mu_j\, dn_j - \mu_{H^+} dn_{H^+} - n_{H^+} d\mu_{H^+} \\ &= \sum_{j \neq H^+}^{N_S-1} \mu_j\, dn_j + \mu_{H^+} dn_{H^+} - \mu_{H^+} dn_{H^+} \\ &\quad - n_{H^+} d\mu_{H^+}. \\ &= \sum_{j \neq H^+}^{N_S-1} \mu_j\, dn_j - n_{H^+} d\mu_{H^+}. \end{aligned} \quad 11$$

In Eqn 11, n_{H^+} is not independent of μ_{H^+} but rather is a function of μ_{H^+}, such that $n_{H^+}(\mu_{H^+}) = \exp\left(-\left(\mu_{H^+} - \mu_{H^+}^\circ\right)/RT\right)$. Thus, explicit dependence of $d\tilde{G}$ on n_{H^+} can be removed completely,

$$\mathrm{d}\tilde{G} = \sum_{j \neq H^+}^{N_S-1} \mu_j\, \mathrm{d}n_j - e^{-(\mu_{H^+} - \mu_{H^+}^\circ)/RT} d\mu_{H^+}. \quad 12$$

Eqns 11 and 12 are appropriate for a system in which changes in the pH are addressed by changes in the chemical potential μ_{H^+}. The changes in $d\tilde{G}$ that lead to a change in reaction free energy $\Delta_r\tilde{G}$ are changes $d\xi$ due to a reaction coordinate ξ,

$$\mathrm{d}\tilde{G} = \sum_{j \neq H^+}^{N_S-1} \mu_j \frac{\partial n_j}{\partial \xi} \mathrm{d}\xi - n_{H^+}(\mu_{H^+}) \frac{\partial \mu_{H^+}}{\partial \xi} \mathrm{d}\xi \quad 13$$

$$= \sum_{j \neq H^+}^{N_S-1} \mu_j\, \gamma_j \mathrm{d}\xi - n_{H^+}(\mu_{H^+}) \frac{\partial \mu_{H^+}}{\partial \xi} \mathrm{d}\xi, \quad 14$$

where $\partial n_j/\partial\xi = \gamma_j$ is the signed stoichiometric coefficient for the reaction. As such, the Legendre transform is useful in cases when (1) the chemical potential changes during a reaction and is easier to measure than the pH, (2) the pH needs to be allowed to fluctuate as in an open system, or (3) when it is more mathematically convenient to work with the chemical potential than concentration. Holding the chemical potential for free protons μ_{H^+} constant, regardless of whether the reference state for protons is the standard state or any other state, gives

$$d\tilde{G} = \sum_{j\neq H^+}^{N_S-1} \mu_j dn_j$$

and

$$\Delta_r\tilde{G} = \sum_{j\neq H^+}^{N_S-1} \mu_j \Delta n_j.$$

However, working with concentrations that are functions of chemical potentials, such as $n_{H^+}(\mu_{H^+})\mathrm{d}\mu_{H^+}$, rather than chemical potentials that are functions of concentrations, such as $\mu_{H^+}(n_{H^+})dn_{H^+}$, does not make it easier to hold the system constant at a constant reference pH.

Now, consider the differential of F of Eqn 8,

$$F = G(T,P,n_1,\dots,n_{N_S}) - \textstyle\sum_{j=1}^{N_S} N_H(j)n_j\mu_{H^+}$$

$$dF = dG(T,P,n_1,\dots,n_{N_S}) - d\left(\sum_{j=1}^{N_S} N_H(j)n_j\mu_{H^+}\right)$$

$$= \sum_{j=1}^{N_S} \mu_j dn_j \; - d\left(\sum_{j=1}^{N_S} N_H(j)n_j\mu_{H^+}\right). \quad 15$$

Expanding the first sum and taking into account that $N_{H^+}(H^+) = 0$,

$$dF = \sum_{j=1}^{N_S-1} \mu_j dn_j + \mu_{H^+}dn_{H^+} \; - d\left(\sum_{j=1}^{N_S-1} N_H(j)n_j\,\mu_{H^+}\right). \quad 16$$

Expansion of the differential of the term in parenthesis gives,

$$dF = \sum_{j=1}^{N_S-1} \mu_j dn_j + \mu_{H^+}dn_{H^+} \; - \sum_{j=1}^{N_S-1} N_H(j)\,\mu_{H^+}dn_j - \sum_{j=1}^{N_S-1} N_H(j)n_j\,d\mu_{H^+}. \quad 17$$

The middle two terms cancel by Eqn 4 such that,

$$dF = \sum_{j=1}^{N_S-1} \mu_j dn_j - \sum_{j=1}^{N_S-1} N_H(j)n_j\,d\mu_{H^+} \; = d\tilde{G} \quad 18$$

$$= \sum_{j=1}^{N_S-1} \mu_j dn_j - n_{H^+}d\mu_{H^+} \quad = d\tilde{G}. \quad 19$$

Equation 19 is identical to Eqn 11. Therefore, dF represents a Legendre transform of dG.

The 1994 IUPAC-IUBMB transform (Alberty's transform). In the 1994 IUPAC-IUBMB transform, the transformation of *G* to *F* to is defined to include both free and bound protons, whereas the discussion above only uses those protons that are now free because of the ionization of a neutral species into an ionized species *j*. This choice was justified by the use of equilibrium and conservation of mass arguments (see section 3.3 in reference [25]) to select conjugate variables rather than the usual definition of conjugate variables of a function F: a variable *x* and the partial derivative of the function with respect to the variable, $w = \partial F/\partial x$.

To specifically demonstrate Alberty's transform as indicated in Eqns. 4.1-1 and 4.1-2 in section 4.1 of reference [25] and the discussion around Eqn 3.3-2 in section 3.3, we define a parameter $N_{\tilde{H}}(j)$ to be the number of bound H atoms in species *j*. Using Eqn 10, the number of free and bound protons are then,

$$\sum_{j=1}^{N_S} N_H(j)n_j + \sum_{j=1}^{N_S} N_{\tilde{H}}(j)n_j.$$

The transform then uses both free and bound protons,

$$F_A = G(T,P,n_1,\dots,n_{N_S}) - \left(\sum_{j=1}^{N_S} N_H(j)n_j\mu_{H^+} + \sum_{j=1}^{N_S} N_{\tilde{H}}(j)n_j\mu_{H^+}\right) \quad 20$$

$$= \sum_{j=1}^{N_s} \mu_j n_j - \left(\sum_{j=1}^{N_S} N_H(j)n_j\mu_{H^+} + \sum_{j=1}^{N_S} N_{\tilde{H}}(j)n_j\mu_{H^+}\right) \quad 21$$

$$= \sum_{j=1}^{N_s} \mu_j n_j - \left(\sum_{j=1}^{N_S-1} N_H(j)n_j\mu_{H^+} + \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j)n_j\mu_{H^+}\right), \quad 22$$

where the last equality follows since $N_H(H^+) = 0$ and $N_{\tilde{H}}(H^+) = 0$. An incremental change in F_A is given by,

$$dF_A = \sum_{j=1}^{N_s} \mu_j dn_j - d\left(\sum_{j=1}^{N_S-1} N_H(j)n_j\mu_{H^+} + \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j)n_j\mu_{H^+}\right) \quad 23$$

$$= \sum_{j=1}^{N_s} \mu_j dn_j - \sum_{j=1}^{N_S-1} N_H(j)dn_j\mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j)dn_j\mu_{H^+} - \sum_{j=1}^{N_S-1} N_H(j)n_j d\mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j)n_j d\mu_{H^+}.$$

Expanding the first term such that the contribution due to protons is explicit in the sum, 24

$$dF_A = \sum_{j=1}^{N_s-1} \mu_j dn_j + \mu_{H^+}dn_{H^+} - \sum_{j=1}^{N_S-1} N_H(j)dn_j\mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j)dn_j\mu_{H^+} - \sum_{j=1}^{N_S-1} N_H(j)n_j d\mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j)n_j d\mu_{H^+}.$$

The second and third terms in Eqn 24 cancel due the identity in Eqn. 10. This produces,

$$dF_A = \sum_{j=1}^{N_s-1} \mu_j dn_j - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) dn_j \mu_{H^+} - \sum_{j=1}^{N_S-1} N_H(j) n_j d\mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) n_j d\mu_{H^+}. \quad 25$$

Collecting terms associated with $N_H(j)$ and $N_{\tilde{H}}(j)$ and finally using the identity of Eqn 19 gives,

$$\begin{aligned} dF_A &= \left(\sum_{j=1}^{N_s-1} \mu_j dn_j - \sum_{j=1}^{N_S-1} N_H(j) n_j d\mu_{H^+} \right) - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) dn_j \mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) n_j d\mu_{H^+} \\ &= \left(\sum_{j=1}^{N_s-1} \mu_j dn_j - n_{H^+} d\mu_{H^+} \right) - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) dn_j \mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) n_j d\mu_{H^+} \\ &= d\tilde{G} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) dn_j \mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) n_j d\mu_{H^+}. \end{aligned} \quad 26$$

At constant chemical potential for the protons, $\mathrm{d}F_A = \mathrm{d}\widetilde{\mathrm{G}} - \mu_{H^+} \sum N_{\tilde{\mathrm{H}}}(j) dn_j$, in contrast to Eqn 18. By including the non-ionized sites $N_{\tilde{\mathrm{H}}}(j)$ in the transform, the 1994 IUPAC-IUBMB transform fails to represent a Legendre transform of the Gibbs free energy.

Total Free Energy Change: Coupling to external bath of protons.
Although Eqn 19 represents a Legendre transform of the Gibbs free energy, a Legendre transform alone is insufficient to address the situation in which the $\mathrm{pH} \propto \log[n_{H^+}]$ is constant. This is because holding the proton chemical potential constant instead of n_{H^+} allows n_{H^+} to fluctuate around an average value $\bar{n}_{H^+}$. Consequently, the pH also fluctuates around an average value. In order to keep the pH strictly constant, the number of protons consumed or produced in a reaction must be balanced by adding or removing protons from an external bath of protons. That is, the system must be open with respect to the bath. This point has previously been discussed by Raff and Cannon [29]. In the laboratory, pH buffers play the role of the bath. To differentiate protons that are added to the system from the bath from those that are in the system, the number/concentration of protons from the bath that are added/removed from the system are indicated by $n_{H^+,bath}$. In addition, we will refer to the differential Gibbs energy from Eqn 19 as the *system* differential Gibbs energy, $d\tilde{G}_{sys}$. The chemical potential of the protons in the bath and in the system are both given by μ_{H^+}. Since the bath just consists of protons, the free energy of the bath is $G_{bath} = n_{H^+,bath}\mu_{H^+}$. For an open system, the total free energy change is that due to changes in the system plus the contribution associated with addition/removal of protons from the bath. That is,

$$dG_{Total} = d\tilde{G}_{sys} + dG_{bath} \quad 27$$

$$= \sum_{j\neq H^+}^{N_S-1} \mu_j dn_j - n_{H^+,sys} d\mu_{H^+} + \mu_{H^+} dn_{H^+,bath}. \quad 28$$

In order to keep the pH of the system constant, the number of protons consumed in a reaction $\left(N_H(j)dn_j < 0\right)$ must be replaced by an equal number of protons from the bath such that the change in the number of protons in the bath is the same as the number of protons consumed in the reaction. Therefore, we must have $dn_{H^+,bath} = \sum_{j\neq H^+}^{N_S-1} N_H(j) dn_j$. Likewise, the number of protons removed from system and added to the bath is equal to the number of free protons released from species j as it changes from its neutral form to it's ionic form. Consequently,

$$dG_{Total} = \sum_{j\neq H^+}^{N_{S-1}} \mu_j dn_j - n_{H^+,sys} d\mu_{H^+} + \mu_{H^+} \sum_{j\neq H^+}^{N_{S-1}} N_H(j) dn_j. \quad 29$$

Rearranging terms we obtain,

$$\begin{aligned} dG_{Total} &= \sum_{j\neq H^+}^{N_{S-1}} \mu_j dn_j + \mu_{H^+} \cdot \sum_{j\neq H^+}^{N_{S-1}} N_H(j) dn_j - n_{H^+,sys} d\mu_{H^+} \\ &= \sum_{j\neq H^+}^{N_{S-1}} (\mu_j + \mu_{H^+} N_H(j)) dn_j - n_{H^+,sys} d\mu_{H^+}. \end{aligned} \quad 30$$

When the proton chemical potential doesn't change, the total free energy is given by

$$dG_{Total} = \sum_{j\neq H^+}^{N_{S-1}} (\mu_j + \mu_{H^+} N_{H^+}(j)) dn_j, \quad 31$$

and

$$\Delta G_{Total} = \sum_{j\neq H^+}^{N_{S-1}} (\mu_j + \mu_{H^+} N_{H^+}(j)) \Delta n_j \,. \quad 32$$

Consequently, the Eqn 32 represents the total free energy change of the system plus bath. The equation is a composite of a Legendre transform of the system Gibbs energy plus an additional term, due to removal/addition of protons to the bath, required to keep the pH constant.

In contrast, adding the free energy of the bath to the 1994 IUPAC-IUBMB transform gives,

$$dF_{A,tot} = d\tilde{G} + \mu_{H^+} dn_{H^+,bath} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) dn_j \mu_{H^+} - \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) n_j d\mu_{H^+}. \quad 33$$

The free energy change obtained by integrating over $dF_{A,Tot}$ at constant μ_{H^+} gives,

$$\Delta F_{A,Tot} = \Delta G_{Tot} - \mu_{H^+} \cdot \int \sum_{j=1}^{N_S-1} N_{\tilde{H}}(j) dn_j, \quad 34$$

which only gives $\Delta F_{A,Tot} = \Delta G_{Total}$ when the number of non-titratable hydrogens is conserved between reactants and products. For some reactions this is not the case. For example, the pyruvate decarboxylase reaction, pyruvate $\rightarrow$ acetaldehyde $+$ CO_2, has three non-titratable hydrogens for the reactant pyruvate but has four non-titratable hydrogens for the product acetaldehyde.

Discussion for Transforms and Constant pH Systems. The 1994 IUPAC-IUBMB transform used equilibrium and mass conservation constraints to substitute for the change in proton and magnesium ion concentrations in the Gibbs energy equation,

$$dG = \sum_{j}^{N-2} \mu_j dn_j + \mu_{H^+} dn_{H+} + \mu_{Mg^{2+}} dn_{Mg^2}.$$

While it is perfectly legitimate to reduce the number degrees of freedom using constraints in such a way, one must keep in mind that the constraints are only valid under the specified conditions. It is not always obvious when those conditions are violated. These constraints were developed by considering equality relations for processes at equilibrium [25] and not through the typical procedure for a Legendre transform in which the independent variable x of the function $F(x)$ is replaced with a co-varying variable w of x using the relationship that F varies with x as $w \equiv \partial F/\partial x$. This relationship is valid under static or dynamic, equilibrium or non-equilibrium conditions, as long as the chemical potential can be defined. Even when equilibrium conditions hold, the 1994 IUPAC-IUBMB transform will fail to yield the correct values for ΔG_{Total} if the number of non-titratable hydrogens is not conserved between reactants and products.

Rather than reducing the number of degrees of freedom, the point of a Legendre transform of a function $F(x)$ is to replace one independent variable of the function with a co-varying variable of the function. The transform is convenient when one is concerned with the change in $F(x)$ as a function of w rather than x. This is the case when w is the measured or modeled variable and not x. A succinct and clear review of the use of Legendre transforms in the physical sciences has been published by Zia, Redish and McCay [30].

The astute reader will have realized that the Legendre transform is not at all necessary to derive the correct total free energy change of the system plus bath. One can see that this is the case by simply adding the respective system and bath reactions to get an overall reaction,

$$\begin{array}{r} \mathrm{H_nA} \rightleftharpoons 2\mathrm{H^+(system)} + \mathrm{H_{n\text{-}2}A^{2\text{-}}} \\ \underline{2\mathrm{H^+(system)} \rightleftharpoons 2\mathrm{H^+(bath)}} \\ \mathrm{H_nA} \rightleftharpoons 2\mathrm{H^+(bath)} + \mathrm{H_{n\text{-}2}A^{2\text{-}}} \end{array}. \quad 35$$

The free energy of the overall reaction is simply given by Eqn 6 in which N_S includes species in both the system and the bath.

Chemical Potentials in the Case of Multiple Charge States

The standard molar chemical potential for a species in aqueous solution is the free energy required to form one mole of the species from its elements *in vacuo*, followed by solvation of the species in solvent, which in the biological case is water,

$$\mu_i^\circ = \Delta_f G_i^\circ + \Delta_{solv} G_i^\circ(water). \quad 36$$

Analogously, the chemical potential of a pseudo-isomer is the value from Eqn 36 plus the free energy of the acid dissociation/base addition step that produces the pseudo-isomer of interest,

$$\mathrm{AH_n} = \mathrm{AH_{n-m}} + \mathrm{H_m}. \quad 37$$

The standard free energy of the acid dissociation process from the neutral species to the charged species is $\Delta_{ion} G^\circ$. For any species i, the pseudo-isomers will be designated by the additional index j, with the neutral pseudo-isomer specifically designated by the index $j = 1$. If the free energy of the acid dissociation reaction (Eqn 37) from the neutral pseudo-isomer $j = 1$ to pseuso-isomer $j = j'$ is symbolized by $\Delta_{ion} G_{i,1\rightarrow j'}$, then the molar chemical potential of a pseudo-isomer j' of species i is,

$$\mu_{i,j'}^\circ = \mu_i^\circ + \Delta_{ion} G_{i,1\rightarrow j'}^\circ. \quad 38$$

Starting from these definitions of the chemical potential, the Alberty and Iotti *et al*. formulations are compared.

Alberty's formulation of chemical potentials and reaction free energies. The thermodynamic treatment of isomers that are either indistinguishable or treated as a group has been addressed by Smith [31] and Straatsma and McCammon [32]. Alberty later used an identical treatment for species that differ only by their acid dissociation state (pseudo-isomers). If there are $N_{iso}(i)$ pseudo-isomers j for a chemical species i, such as pyruvate, a composite standard chemical potential $\tilde{\mu}_i^\circ$ for the species i can be calculated from the pseudo-isomer chemical potentials $\mu_{i,j}^\circ$ as,

$$\tilde{\mu}_i^\circ = -RT \log \sum_{j=1}^{N_{iso}(i)} e^{-\frac{\mu_{i,j}^\circ}{RT}}, \quad 39$$

Where R is the gas constant and T is the temperature in Kelvin. If pseudo-isomer j *of* species *i* has a molar concentration of $n_{i,j}$, then species i has a composite concentration $n_i = \sum_j n_{i,j}$. The composite chemical potential is then,

$$\tilde{\mu}_i = \tilde{\mu}_i^\circ + RT \log n_i. \quad 40$$

If species i doesn't have any pseudo-isomers, then $\tilde{\mu}_i = \mu_i$.

According to the Alberty formulation, the standard free energy change $\Delta_r G_A^\circ$, at constant temperature and pressure is,

$$\Delta_r G_A^\circ = \sum_{i=1}^{N_S} \nu_i \tilde{\mu}_i^\circ, \quad 41$$

with an equilibrium constant $K = exp\left(-\sum_{i=1}^{N_S} \nu_i \tilde{\mu}_i^\circ / RT\right)$. Likewise, the observed free energy change is,

$$\Delta_r G_A = \sum_{i=1}^{N_S} \nu_i \, \tilde{\mu}_i, \qquad 42$$

where $\tilde{\mu}_i$ is given by Eqn 40.

The contribution of H^+ to the free energy change is a constant since n_{H^+} is fixed at the reference value. Moreover, by changing the reference concentration for protons to concentration controlled at the pH of interest, the concentration of protons appearing as a reactant or product will simply be that of the reference concentration; that is, the proton concentration is at equilibrium with respect to the reference concentration. How does this appear in $\Delta G_A = \sum_i \nu_i \tilde{\mu}_i$? On one hand, since the hydrogen ion concentration is at equilibrium with respect to the reference concentration, $\mu_{H^+} = \mu_{H^+}^\circ + RT \log \tilde{n}_{H^+} = 0$. Thus, $\nu_{H^+}\tilde{\mu}_{H^+} = 0$ in Eqns 41 and 42. But more subtly, the contribution of the reference concentration of the proton is implicitly included via Eqn 38 through $\Delta_{ion} G_{i,1\to j'}$ at the specified pH.

The Balanced Biochemical Reaction formulation of reaction free energies. In the Balanced Biochemical Reaction (BBR) method of Iotti, Sabatini *et. al* [16, 17], the chemical equation is explicitly balanced for both mass and charge by treating each pseudo-isomer explicitly. This is done by accounting for the concentrations (both reference and observed) using equilibrium mole fractions. That is, the standard concentration and chemical potential of each pseudo-isomer is adjusted according to its mole fraction at equilibrium, rather than making a composite chemical potential using Eqns 39 and 40. The equilibrium mole fraction $f_{i,j}^\circ$ for a pseudo-isomer j of species i is given by [9],

$$f_{i,j}^\circ = \frac{e^{-\mu_{i,j}^\circ / RT}}{e^{-\tilde{\mu}^\circ{}_i / RT}}, \qquad 43$$

where $\mu_{i,j}^\circ$ is defined by Eqn 38 and $\tilde{\mu}_i^\circ$ is defined as in the Alberty formulation,

$$\tilde{\mu}_i^\circ = -RT \log \sum_{j=1}^{N_{iso}(i)} e^{-\frac{\mu_{i,j}^\circ}{RT}}.$$

If species i participates in a chemical reaction with an unsigned stoichiometric coefficient ν_i, then in the BBR method each of its pseudo-isomers will be distributed among the reactants or products of the reaction according to the scaled stoichiometric coefficients such that,

$$\nu_{i,j} = f_{i,j}^\circ \cdot \nu_i, \qquad 44$$

and,

$$n_{i,j} = f_{i,j}^\circ \cdot n_i. \qquad 45$$

For standard conditions in which the reference concentration $n_i^\circ = 1\text{M}$ for the composite species i, the standard chemical potential $\mu_{i,j}^{\circ,f}$ for each pseudo-isomer j is then,

$$\begin{aligned} \mu_{i,j}^{\circ,f} &= \mu_{i,j}^\circ + \text{RT}\, log\left(f_{i,j}^\circ n_i^\circ\right) \\ &= \mu_{i,j}^\circ + RT\, log\, f_{i,j}^\circ. \end{aligned} \qquad 46$$

The second term in Eqn 46 effectively adjusts the reference concentration for the pseudoisomer from the standard value of 1 M to a new reference concentration of $f_{i,j}^\circ$ M.[†] The full chemical potential of pseudo-isomer j for species i is,

$$\mu_{i,j}^f = \mu_{i,j}^\circ + RT\, log\, f_{i,j}^\circ\, n_i, \qquad 47$$

or equivalently,

$$\mu_{i,j}^f = \mu_{i,j}^{f\circ} + RT\, log\, n_i.$$

The standard free energy change according the Balanced Biochemical Reaction method, $\Delta_r G_{BBR}$, is then,

$$\Delta_r G_{BBR}^\circ = \sum_{i=1}^{N_S} \sum_{j=1}^{N_{iso}(i)} \nu_{i,j} \mu_{i,j}^{\circ,f}. \qquad 48$$

Notice that the use of $\mu_{i,j}^{f\circ}$ instead of $\mu_{i,j}^\circ$ in Eqn 48 ensures that the standard free energy change is the condition in which the concentration n_i° of the composite compound i is at the standard concentration of 1 M. Likewise, the condition-dependent free energy change of the reaction is,

$$\Delta_r G_{BBR} = \sum_{i=1}^{N_S} \sum_{j=1}^{N_{iso}(i)} \nu_{i,j} \mu_{i,j}^f. \qquad 49$$

Comparing Eqns 48 and 49 to Eqns 39 and 40, respectively, the question is, which is the appropriate method to use? Or, as suggested [16, 17], are the two approaches actually equivalent?

What does Statistical Thermodynamics say? In order to understand the difference between Alberty's combining of chemical potentials and the use of mole fractions in the BBR approach, it is helpful to understand the difference between the probability of two events each happening together and the probability of observing one or the other of the two events.

If we are considering the probabilities p_A and p_B of two independent events A and B, the probability that both A *and* B will occur is $p_A \cdot p_B$. But if we are interested in the probability that either A occurs *or* B occurs, that probability is $p_A + p_B$.

[†] Eqn 46 could have been written as $\mu_{i,j}^{\circ,f} = \mu_{i,j}^\circ + RT\, log\, n_{i,j}^\circ$ by using the notation derived from Eqn 45 that $n_{i,j}^\circ = f_{i,j}^\circ \cdot n_i^\circ$. However, we use the notation $f_{i,j}^\circ$ to maintain consistency of notation with the use of $f_{i,j}$ in reference [26] A. Sabatini, M. Borsari, G. P. Moss, and S. Iotti, "Chemical and biochemical thermodynamics reunification (IUPAC Technical Report)," *Pure Appl Chem,* 2020, doi: doi:10.1515/pac-2019-0908., however with the difference that we add the plimsoll symbol ∘ to emphasize that equilibrium mole fractions are being used in both cases.

Consider now that there are six events that could occur, A, B, C, D, E and F. For example, these could be the faces of a die. If after $\mathcal{N}$ throws of the die, A appears n_A times, B occurs n_B and so on, the overall probability (probability density) of observing $n_A, n_B, n_C, n_D, n_E, n_F$ is,

$$Pr(n_A, \dots, n_F) = \mathcal{N}! \prod_{i=A}^{F} \frac{1}{n_i!} p_i^{n_i}. \quad 50$$

However, if one face of the die has been damaged and you can't tell if it says E or F, then we combine the counts n_E and n_F and only consider their combined probability, $p_E + p_F$. In this case, all we can say about the overall probability is that it is,

$$\begin{aligned} \Pr(n_A, \dots, n_F) = \mathcal{N}! & \frac{1}{n_A!} p_A^{n_A} \frac{1}{n_B!} p_B^{n_B} \frac{1}{n_C!} p_C^{n_C} \frac{1}{n_D!} p_D^{n_D} \\ & \cdot \frac{1}{(n_E + n_F)!} (p_E + p_F)^{n_E + n_F}. \end{aligned} \quad 51$$

Mathematically Eqn 51 is the case when we have pseudo-isomers but don't explicitly measure each pseudo-isomer – in that case, we combine the chemical potentials. Eqn 50 is the case when we measure each species completely using mole fractions. Both formulations of the probability are correct; they are just used for different scenarios.

Statistical thermodynamics is just probability theory applied to chemical systems. The probabilities p_i above are replaced by the exponent of the standard chemical potential, $e^{(-u_i^\circ/k_BT)}$, where $u_i^\circ = \mu_i^\circ / N_{Avo}$ is the standard molecular chemical potential. In this expression, μ_i° is the standard molar chemical potential and N_{Avo} is Avogadro's number.

For a system with N_S distinct chemical species i, each with a standard molecular chemical potential u_i° and a count of n_i such that there are $\mathcal{N} = \sum_i n_i$ total molecules, the free energy of the system that is analogous to the probability density is given by,

$$\frac{G(n_1, \dots, n_{N_S}, T, P)}{k_B T} = -\log\left(\mathcal{N}! \prod_{i=1}^{N_S} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i} \right), \quad 52$$

where k_B is Boltzmann's constant and T is again the temperature. This equation is turned into the more common expression of classical thermodynamics where $G = \sum_i n_i \mu_i$ as follows. First, the logarithm is expanded and Sterling's approximation is used in which $\log n! \approx n \log n - n$,

$$\frac{G(n_i, T, P)}{k_B T} = \log \mathcal{N}! - \sum_{i=1}^{N_S} log\, n_i! + n_i \log\left(e^{-u_i^\circ/k_BT}\right) \quad 53$$

$$= -\mathcal{N} \log \mathcal{N} + \mathcal{N} + \left(\sum_{i=1}^{N_S} n_i \log n_i - n_i \right) + n_i \log\left(e^{-u_i^\circ/k_BT}\right), \quad 54$$

where $G(n_i, T, P) = G(n_1, \dots, n_{N_S}, T, P)$. Using the identity $\mathcal{N} = \sum_i n_i$, the second term $\mathcal{N}$ cancels the $\sum_i n_i$ term in the parentheses,

$$\frac{G(n_i, T, P)}{k_B T} = -\mathcal{N} \log \mathcal{N} + \sum_{i=1}^{N_S} \left[n_i \log n_i - n_i \log\left(e^{-u_i^\circ/k_BT}\right) \right] \quad 55$$

$$= \sum_{i=1}^{N_S} \left[n_i \log \frac{n_i}{\mathcal{N}} + n_i \frac{u_i^\circ}{k_B T} \right] \quad 56$$

As formulated in Eqn 56, the free energy is a function of the extent of the system through the counts $n_1, \dots, n_{N_S}$ and the units are units of energy, e.g., kJ or Kcal. To make the free energy an intensive function such that the units are kJ/mol or Kcal/mol, we require that $\mathcal{N} = N_{Avo}$ (Avogadro's number) and Eqn 56 is divided through by N_{Avo} to give,

$$\frac{G(n_i, T, P)}{N_{Avo} k_B T} = \sum_{i=1}^{N_S} \left[\frac{n_i}{N_{Avo}} log\, \frac{n_i}{\mathcal{N}} + \frac{n_i}{N_{Avo}} \frac{u_i^\circ}{k_B T} \right]. \quad 57$$

Eqn 57 can be expressed in molar units using: (i) moles $N_i = n_i / N_{Avo}$, (ii) $N_{Avo} k_B T = RT$, (iii) substituting in the molar chemical potential $\mu_i^\circ = N_{Avo} u_i^\circ$ and (iv) finally, multiplying through by RT,

$$G(N_i, T, P) = \sum_i^{N_S} N_i (RT \log N_i + \mu_i^\circ) \quad 58$$

$$= \sum_i^{N_S} N_i \mu_i, \quad 59$$

which is the desired relationship.

Pseudo-isomer situation. Now consider the same system as above, however with the chemical species $k, k+1, \dots, k+l$ as pseudo-isomers of each other. If we know the counts $n_k, \dots, n_{k+l}$, the free energy can then be written as,

$$\begin{aligned} \frac{G(n_i, T, P)}{k_B T} = \log \Bigg(& \mathcal{N}! \prod_{i=1}^{k-1} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i} \\ & \cdot \prod_{i=k}^{k+l} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i} \\ & \cdot \prod_{i=k+l+1}^{k-1} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i} \Bigg). \end{aligned} \quad 60$$

If we don't know the counts of the l pseudo-isomers individually but only know the total $n_{k:l} = \sum_{i=k}^{k+l} n_i$, then we can combine the pseudo-isomers such that the free energy is,

$$\begin{aligned} \frac{G(n_i, T, P)}{k_B T} = log \Bigg(& \mathcal{N}! \prod_{i=1}^{k-1} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i} \\ & \cdot \frac{1}{n_{k:l}!} \left(\sum_{i=k}^{l} e^{-u_i^\circ/k_BT} \right)^{n_{k:l}} \\ & \cdot \prod_{i=k+l+1}^{k-1} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i} \Bigg). \end{aligned} \quad 61$$

Notice the middle arguement of the logarithm which addresses the free energy contribution of the $k, k+1, \ldots, k+l$ pseudoisomers. *The multinomial expansion of the summation in parentheses is the sum over all possible configurations of the pseudo-isomers, equilibrium and non-equilibrium*. That is,

$$\left(\sum_{i=k}^{l} e^{-u_i^\circ/k_BT}\right)^{n_{k:l}} = \sum_{n_k+\cdots+n_l=n_{k:l}} \left(\frac{n_{k:l}!}{n_k!\cdots n_l!}\right) \prod_{i=k}^{l} \frac{1}{n_i!} \left(e^{-u_i^\circ/k_BT}\right)^{n_i}, \quad 62$$

where the sum $\sum_{n_k+\cdots+n_l=n_{k:l}}$ on the left-hand side indicates a multi-sum over all values of $n_k, \ldots, n_l$ such that $n_k + \cdots + n_l = n_{k:l}$. *The sum over all possible configurations of pseudoisomers is what makes this formulation of the free energy different from the Iotti, et. al BBR approach, in which only one configuration of the pseudoisomers, the equilibrium configuration, is considered.* Both formulations are correct, but they are not equivalent. The former is used when one knows nothing about the pseudoisomer state and the latter is used when one knows the exact pseudoisomer state. Just as for Eqn 52, Eqn 61 leads to,

$$G(n_1, \ldots, n_{k:l}, \ldots, n_{N_S}, T, P) = n_1\mu_1 + \cdots + n_{k:l}\mu_{k:l} + \cdots + n_{N_S}\mu_{N_S}, \quad 63$$

where the molar chemical potential for the pseudo-isomers are,

$$\mu_{k:l} = \mu_{k:l}^\circ + \log n_{k:l},$$

and,

$$\mu_{k:l}^\circ = \log \sum_{i=k}^{l} e^{-\mu_i^\circ/k_BT}. \quad 64$$

Notice that the only time that equilibria between pseudo-isomers is used is when calculating the equilibrium property $\mu_{k:l}^\circ$. The free energy change for a reaction is then,

$$\Delta_r G = \sum_i \Delta n_i \mu_i. \quad 65$$

There is no assumption of equilibrium between pseudo-isomers needed for a non-equilibrium reaction free energy change in Eqn 65 beyond that used in calculating the equilibrium property $\mu_{k:l}^\circ$ in Eqn 64. When numerically comparing the BBR free energies to the free energies obtained using Eqn 65, the values will usually be very close because the equilibrium configuration is the maximum likelihood configuration and will contribute the most to the sum in Eqn 62.

The correct version of the free energy to use, the Alberty formulation or the BBR method, depends on two conditions: (1) what is observable and what is not observable, and (2) whether the pseudo-isomers are at equilibrium with respect to one another or not. While the assumption that the pseudoisomers are in equilibrium is an excellent assumption in the case of metabolism since protonation/deprotonation occurs on the picosecond timescale [33] and enzymatic reactions generally occur on the millisecond to second timescale [34], the assumption alone is not necessarily justification for using the BBR formulation. The reason is that the values of free energies and entropies, unlike energies, depend on which degrees of freedom are measured; when comparing reaction free energies, the comparisons must use free energies based on the same degrees of freedom, and the Alberty formulation without the assumption of equilibrium between pseudo-isomers is in common use [14, 35, 36]. However, if the equilibrium assumption is well-justified and one is consistent in using the BBR approach, then the BBR approach offers a more precise value for the free energy because there is no uncertainty in the configuration of the pseudo-isomers.

In practice for measurements or modeling, whether the numerical difference is significant will depend on how far from equilibrium is the biological reaction. The calculated free energies of reaction for the pyruvate decarboxylase reaction discussed above (Eqn 3) using both the BBR and Alberty formulations are shown in Table 1. In both cases, the total concentration of each of pyruvate, acetaldehyde and CO_2 is 1 M. As can be seen, the difference between the two approaches is small compared to the magnitude of the average standard free energy change. The difference will be greatest at the pH that is the pKa of pyruvate, pH = 2.92. In this case, the chemical potential for both pseudoismers of pyruvate is $\mu_{pyr,1}^\circ = \mu_{pyr,2}^\circ = -483.6$. Consequently, the free energy contribution due to pyruvate for the Alberty formulation is,

$$\nu_{pyr} \cdot \tilde{\mu}_{pyr}^\circ = 1 \cdot [-RT\, log\left(\sum_{j=1}^{N_{i,iso}} e^{-\mu_{pyr,j}^\circ/RT}\right) + RT\, log(1)], \quad 66$$

$$= -RT\, log\left(2 \cdot e^{-\mu_{pyr,1}^\circ/RT}\right),$$

while the free energy contribution for the BBR formulation is,

$$\nu_{pyr} f_{pyr,1} \cdot \mu_{pyr,1}^\circ + \nu_{pyr} f_{pyr,2} \cdot \mu_{pyr,2}^\circ = \frac{1}{2}\mu_{pyr,1}^{f\circ} + \frac{1}{2}\mu_{pyr,2}^{f\circ} = \mu_{pyr,1}^{f\circ}, \quad 68$$

where the last equality follows since at pH = pKa, $\mu_{pyr,1}^\circ = \mu_{pyr,2}^\circ$. The factor of 2 comes into the Alberty formulation (Eqn 67) because there are $\binom{2}{1} = 2$ choices of pseudo-isomers, which reflects the uncertainty in the distribution of the pseudo-isomers. This results in a difference in chemical potential of $-RT \log 2 = -1.71$ kJ/mol, which is approximately the difference at pH = 3.0 in Table 1.

	pH								
Method	**0.0**	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**	**6.0**	**7.0**	**8.0**
BBR	-42.72	-42.83	-43.26	-42.47	-37.05	-30.96	-25.17	-19.45	-13.74
Alberty	-42.70	-42.67	-42.42	-40.76	-36.37	-30.84	-25.15	-19.45	-13.74

Table 1. Standard free energies (kJ/mol) at specified pH for the reaction catalyzed by pyruvate decarboxylase.

A more complex example is that of the pyruvate kinase reaction,

$$\text{pyruvate} + \text{ATP} \rightleftharpoons \text{phosphoenolpyruvate} + \text{ADP}. \quad 69$$

In this case the reactants pyruvate and ATP have a total of eight ionic states to consider, while the products have a total of 10 ionic states to consider, potentially exacerbating the difference in free energies calculated using the two methods. However, because a single ionic state is often dominate for each species at a given pH, the difference is again maximized near the pKa of each species. As long as each species has sufficiently different pKas, the maximum difference in $\Delta_r G$ is approximately $-RT\log 2 = -1.71$ kJ/mol, as seen near pH 6.0.

	pH								
Method	**0.0**	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**	**6.0**	**7.0**	**8.0**
BBR	-42.69	-40.92	-34.63	-32.49	-35.30	-35.69	-35.03	-29.87	-23.75
Alberty	-41.71	-39.66	-35.59	-33.63	-34.59	-34.76	-33.31	-29.11	-23.61

Table 2. Standard free energies (kJ/mol) at specified pH for the reaction catalyzed by pyruvate kinase.

An effective difference in the combinatorial nature of the two methods can be estimated using $c = e^{|\Delta\Delta_r G|/RT}$. The value of *c* is plotted against pH for the pyruvate kinase reaction in Figure 1. The maximum value of *c* is 2.00 at pH 6.0.

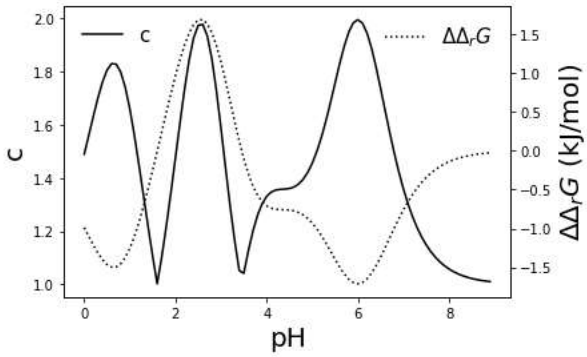

Figure 1.Dependence of effective number *c* of combinations and permutations of pseudo-isomers (left axis) and difference in free energies of reaction (right axis) between BBR and Alberty methods as a function of pH for the pyruvate kinase reaction, which has a total of 18 different pseudo-isomers.

Equivalence Between Alberty and BBR methods at Equilibrium. If in fact both methods are correct given their assumptions, it should be possible to use the equilibrium assumption to obtain the same results in the Alberty method as in the BBR method. In fact, Alberty sometimes assumed that the chemical species constituting the pseudo-isomer group were at equilibrium with H^+ and Mg^{2+} ions. He does so when he states, "At specified pH and pMg, the various forms [j] of a reactant [i] have the same $\Delta_f G'_j$ [$\Delta_f G_{i,j}$] at chemical equilibrium" [37]. As mentioned above, this is a very reasonable assumption since the timescales for protonation/deprotonation and forming ions is much faster than the timescales of the enzymatic reactions. However, when this assumption is made, the approach used by Alberty, comprised of Eqns 39 and 41 for standard conditions, is identical to the BBR approach for chemical thermodynamics, Eqn 48, as shown next.

First, in conditions other than equilibrium, the mole fraction is,

$$f_{i,j} = \frac{e^{-\mu_{i,j}/RT}}{e^{-\tilde{\mu}_i/RT}}, \quad 70$$

Next, the Alberty expression for a change in reaction free energy (Eqn 41) can be expanded using the identity $1 = \sum_j f_{i,j}$, where $f_{i,j}$ is the mole fraction under the observed conditions,

$$\Delta_r G_A^\circ = \sum_{i=1}^{N_S} \nu_i \, \tilde{\mu}_i^\circ \left(\sum_{j=1}^{N_{iso}(i)} f_{i,j} \right) \quad 71$$

$$= \sum_{i=1}^{N_S} \sum_{j=1}^{N_{iso}(i)} \nu_i \cdot f_{i,j} \tilde{\mu}_i^\circ \quad 72$$

$$= \sum_{i=1}^{N_S} \sum_{j=1}^{N_{iso}(i)} \nu_{i,j} \tilde{\mu}_i^\circ, \quad 73$$

where the last equality follows from Eqn 46 when it is assumed that $f_{i,j} = f_{i,j}^\circ$. The standard chemical potential can be further expanded using the identity $1 = e^{-\mu_{i,j}^\circ}/e^{-\mu_{i,j}^\circ}$ and the definition of the equilibrium mole fraction $f_{i,j}^\circ$,

$$\tilde{\mu}_i^\circ = -RT\log\left(e^{-\tilde{\mu}_i^\circ/RT} \frac{e^{-\mu_{i,j}^\circ/RT}}{e^{-\mu_{i,j}^\circ/RT}} \right) \quad 74$$

$$= -RT\log\left(f_{i,j}^{\circ\,-1} e^{-\mu_{i,j}^\circ/RT} \right) \quad 75$$

$$= \mu_{i,j}^\circ + RT\log f_{i,j}^\circ \quad 76$$

$$= \mu_{i,j}^{f^\circ}. \quad 77$$

Substituting into Eqn 73,

$$\Delta_r G_A^\circ = \sum_{i=1}^{N_S} \sum_{j=1}^{N_{iso}(i)} \nu_{i,j} \mu_{i,j}^{f^\circ} \quad 78$$

$$= \Delta_r G_{\text{BBR}}^\circ.$$

That is, comparing Eqn 78 to Eqn 48, it is clear that when the pseudo-isomers are assumed to be at equilibrium and one accounts for pseudo-isomers using the modified stoichiometric coefficients $\nu_{i,j}$, the Alberty formulation of Eqn 73 and BBR approaches are equivalent. Applying the equilibrium assumption to the Alberty method is how, in a recent IUPAC publication, Sabatini, *et al*, [26] came to the conclusion that the BBR formulation was equivalent to the Alberty formulation.

Conclusions for Formulations of Chemical Potentials. For enzyme catalyzed reactions the BBR approach is appropriate when the individual pseudo-isomers are at equilibrium and are explicitly measured or modeled, while the Alberty approach is appropriate

when the pseudo-isomers are not measured or modeled or are not necessarily at equilibrium with respect to one another. For both modeling and measurement, the Alberty approach is more convenient in that the number of variables needed to be accounted for is reduced significantly for biochemical reactions. However, the tradeoff is a loss in precision in the Alberty method due to uncertainty associated with the configuration of the pseudo-isomers. If the assumption of equilibrium is valid, one can get both the convenience of the Alberty approach and the precision of the BBR approach by simply applying equilibrium mole fractions to the Alberty approach as shown in Eqns 72-73. The numerical difference between the two approaches in any individual free energy of reaction is likely only ± 1.7 kJ/mol or less. These differences generally are not cumulative in a system of coupled chemical reactions, since the individual reaction free energies are constrained such that they must sum to the overall free energy change for a system, and the overall free energy change is generally determined by boundary conditions that may also have at most an error due to pseudo-isomers of ± 1.7 kJ/mol.

More generally, the approach of summing over chemical potentials has applications beyond charge state (pseudo-) isomers, rotational isomers [32], and isomeric hydrocarbons [31]. Statistical thermodynamic theories of non-equilibrium chemical reaction networks assume that the chemical species undergoing reaction will (1) relax to their equilibrium configuration between reactions and (2) after relaxation are in local equilibrium with the solvent [2, 38, 39]. This assumption can be obviated with the use of summary chemical potentials such as Eqn 39 in which the sum is over the ground state and all possible excited state configurations.

Methods

All free energies of formation in aqueous solution (reference chemical potentials) were obtained from http://equilibrator.weizmann.ac.il/, version 2.2 with source code repository commit hash f8bc4ca931f41ae08c5cf15b8945c1b1a85158d0, using the component contribution method [36].

Supplemental Material

Code for the calculations used in generating the tables and figure are available as computational notebooks at https://github.com/wrcannon/CompositeReactionFreeEnergies.

Author Contributions

Both William Cannon and Lionel Raff contributed to conceptualization, methodology, analysis and writing of the study.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

W. R. Cannon was supported by funding from the National Institute of Biomedical Imaging and Bioengineering through award U01EB022546 and by the U.S. Department of Energy's Office of Biological and Environmental Research, through projects 69513, 72228 and 74860. PNNL is operated by Battelle for the US Department of Energy under Contract DE-AC06-76RLO.

References

[1] D. A. Beard, E. Babson, E. Curtis, and H. Qian, "Thermodynamic constraints for biochemical networks," (in eng), *J Theor Biol,* vol. 228, no. 3, pp. 327-33, Jun 7 2004, doi: 10.1016/j.jtbi.2004.01.008.

[2] H. Qian and D. A. Beard, "Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium," (in eng), *Biophys Chem,* vol. 114, no. 2-3, pp. 213-20, Apr 22 2005, doi: 10.1016/j.bpc.2004.12.001.

[3] A. Bar-Even, A. Flamholz, E. Noor, and R. Milo, "Thermodynamic constraints shape the structure of carbon fixation pathways," *Biochimica et biophysica acta,* vol. 1817, no. 9, pp. 1646-59, Sep 2012, doi: 10.1016/j.bbabio.2012.05.002.

[4] W. R. Cannon, J. D. Zucker, D. J. Baxter, N. Kumar, J. M. Hurley, and J. C. Dunlap, "Prediction of metabolite concentrations, rate constants and post-translational regulation using maximum entropy-based simulations with application to central metabolism of Neurospora crassa," *Processes,* vol. 6, no. 6, May 28 2018, doi: https://doi.org/10.3390/pr6060063.

[5] E. J. Chapman, D. L. Childers, and J. J. Vallino, "How the Second Law of Thermodynamics Has Informed Ecosystem Ecology through Its History," (in English), *Bioscience,* vol. 66, no. 1, pp. 27-39, Jan 2016, doi: 10.1093/biosci/biv166.

[6] A. J. Lotka, "Natural Selection as a Physical Principle," (in eng), *P Natl Acad Sci USA,* vol. 8, no. 6, pp. 151-4, Jun 1922, doi: https://doi.org/10.1073/pnas.8.6.151.

[7] A. J. Lotka, "Contribution to the Energetics of Evolution," (in eng), *P Natl Acad Sci USA,* vol. 8, no. 6, pp. 147-51, Jun 1922, doi: https://doi.org/10.1073/pnas.8.6.147.

[8] E. T. Jaynes, "Macroscopic Prediction," in *Complex Systems Operational Approaches in Neurobiology, Physics and Computers*, vol. 31, H. Haken Ed. Berlin Heidelberg: Springer-Verlag 1985, pp. 254-269.

[9] S. Britton, M. Alber, and W. R. Cannon, "Enzyme activities predicted by metabolite concentrations and solvent capacity in the cell," *Journal of The Royal Society Interface,* vol. 17, no. 171, p. 20200656, 2020, doi: doi:10.1098/rsif.2020.0656.

[10] C. E. Shannon, "A Mathematical Theory of Communication," (in English), *Bell Syst Tech J,* vol. 27, no. 3, pp. 379-423, 1948, doi: DOI 10.1002/j.1538-7305.1948.tb01338.x.

[11] S. J. Kiebel and K. J. Friston, "Free energy and dendritic self-organization," *Front Syst Neurosci,* vol. 5, p. 80, 2011, doi: 10.3389/fnsys.2011.00080.

[12] R. A. Alberty, A. Cornish-Bowden, R. N. Goldberg, G. G. Hammes, K. Tipton, and H. V. Westerhoff, "Recommendations for terminology and databases for biochemical thermodynamics," *Biophys Chem,* vol. 155, no. 2-3, pp. 89-103, May 2011, doi: 10.1016/j.bpc.2011.03.007.

[13] R. N. Goldberg, Y. B. Tewari, and T. N. Bhat, "Thermodynamics of enzyme-catalyzed reactions--a database for quantitative biochemistry," (in eng),

Bioinformatics, vol. 20, no. 16, pp. 2874-7, Nov 1 2004, doi: 10.1093/bioinformatics/bth314.

[14] R. M. Fleming and I. Thiele, "von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models," *Bioinformatics,* vol. 27, no. 1, pp. 142-3, Jan 1 2011, doi: 10.1093/bioinformatics/btq607.

[15] X. Li, F. Wu, F. Qi, and D. A. Beard, "A database of thermodynamic properties of the reactions of glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway," *Database : the journal of biological databases and curation,* vol. 2011, p. bar005, 2011, doi: 10.1093/database/bar005.

[16] S. Iotti, A. Sabatini, and A. Vacca, "Chemical and biochemical thermodynamics: from ATP hydrolysis to a general reassessment," *J Phys Chem B,* vol. 114, no. 5, pp. 1985-93, Feb 11 2010, doi: 10.1021/jp903990j.

[17] A. Sabatini, A. Vacca, and S. Iotti, "Balanced biochemical reactions: a new approach to unify chemical and biochemical thermodynamics," *PLoS One,* vol. 7, no. 1, p. e29529, 2012, doi: 10.1371/journal.pone.0029529.

[18] A. Jinich *et al.*, "Quantum chemical approach to estimating the thermodynamics of metabolic reactions," *Sci Rep,* vol. 4, p. 7022, 2014, doi: 10.1038/srep07022.

[19] R. C. Dewar, "Maximum entropy production and plant optimization theories," (in English), *Philos T R Soc B,* vol. 365, no. 1545, pp. 1429-1435, May 12 2010, doi: https://doi.org/10.1098/rstb.2009.0293.

[20] J. J. Vallino and C. K. Algar, "The Thermodynamics of Marine Biogeochemical Cycles: Lotka Revisited," (in English), *Annu Rev Mar Sci,* vol. 8, pp. 333-356, 2016, doi: https://doi.org/10.1146/annurev-marine-010814-015843.

[21] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo, "Glycolytic strategy as a tradeoff between energy yield and protein cost," *P Natl Acad Sci USA,* vol. 110, no. 24, pp. 10039-10044, doi: www.pnas.org/cgi/doi/10.1073/pnas.1215283110.

[22] S. Britton, M. Alber, and W. R. Cannon, "Enzyme activities predicted by metabolite concentrations and solvent capacity in the cell," *J R Soc Interface,* vol. 17, no. 171, p. 20200656, Oct 2020, doi: 10.1098/rsif.2020.0656.

[23] R. A. Alberty, "IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN). Recommendations for nomenclature and tables in biochemical thermodynamics. Recommendations 1994," *Eur J Biochem,* vol. 240, no. 1, pp. 1-14, Aug 15 1996, doi: 10.1111/j.1432-1033.1996.0001h.x.

[24] R. A. Alberty *et al.*, "Recommendations for Nomenclature and Tables in Biochemical Thermodynamics," (in English), *Pure Appl Chem,* vol. 66, no. 8, pp. 1641-1666, Aug 1994, doi: https://doi.org/10.1351/pac199466081641.

[25] R. A. Alberty, *Thermodynamics of Biochemical Reactions*. Hoboken, N.J.: Wiley, 2003.

[26] A. Sabatini, M. Borsari, G. P. Moss, and S. Iotti, "Chemical and biochemical thermodynamics reunification (IUPAC Technical Report)," *Pure Appl Chem,* 2020, doi: doi:10.1515/pac-2019-0908.

[27] R. A. Alberty, "Biochemical Thermodynamics," (in English), *Bba-Protein Struct M,* vol. 1207, no. 1, pp. 1-11, Jul 20 1994, doi: https://doi.org/10.1016/0167-4838(94)90045-0.

[28] R. A. Alberty, "The Fundamental Equation of Thermodynamics for Biochemical Reaction Systems," (in English), *Pure Appl Chem,* vol. 65, no. 5, pp. 883-888, May 1993, doi: DOI 10.1351/pac199365050883.

[29] L. M. Raff and W. R. Cannon, "On the Reunification of Chemical and Biochemical Thermodynamics: A Simple Example for Classroom Use," *Journal of Chemical Education,* 2019/01/08 2019, doi: 10.1021/acs.jchemed.8b00795.

[30] R. K. P. Zia, E. F. Redish, and S. R. McKay, "Making sense of the Legendre transform," *Am J Phys,* vol. 77, no. 7, pp. 614-622, 2009, doi: 10.1119/1.3119512.

[31] B. D. Smith, "Simplified Calculation of Chemical Equilibria in Hydrocarbon Systems Containing Isomers," (in English), *Aiche J,* vol. 5, no. 1, pp. 26-28, 1959, doi: DOI 10.1002/aic.690050107.

[32] T. P. Straatsma and J. A. McCammon, "Treatment of Rotational Isomers in Free Energy Evaluations. Analysis of the Evaluation of Free Energy Differences by Molecular Dynamics Simulations of Systems with Rotational Isomeric States," *Journal of Chemical Physics,* vol. 90, no. 6, pp. 3300-3304, 1989, doi: 10.1063/1.456651.

[33] A. Hassanali, F. Giberti, J. Cuny, T. D. Kuhne, and M. Parrinello, "Proton transfer through the water gossamer," (in English), *P Natl Acad Sci USA,* vol. 110, no. 34, pp. 13723-13728, Aug 20 2013, doi: 10.1073/pnas.1306642110.

[34] A. Radzicka and R. Wolfenden, "A proficient enzyme," (in eng), *Science,* Research Support, U.S. Gov't, P.H.S. vol. 267, no. 5194, pp. 90-3, Jan 6 1995, doi: 10.1126/science.7809611.

[35] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo, "eQuilibrator--the biochemical thermodynamics calculator," *Nucleic acids research,* vol. 40, no. Database issue, pp. D770-5, Jan 2012, doi: 10.1093/nar/gkr874.

[36] E. Noor, H. S. Haraldsdottir, R. Milo, and R. M. Fleming, "Consistent estimation of Gibbs energy using component contributions," *PLoS Comput Biol,* vol. 9, no. 7, p. e1003098, 2013, doi: 10.1371/journal.pcbi.1003098.

[37] R. A. Alberty, "Thermodynamics of the hydrolysis of adenosine triphosphate as a function of temperature, pH, pMg, and ionic strength," (in English), *Journal of Physical Chemistry B,* vol. 107, no. 44, pp. 12324-12330, Nov 6 2003, doi: 10.1021/jp030576l.

[38] T. Schmiedl and U. Seifert, "Stochastic thermodynamics of chemical reaction networks," (in English), *Journal of Chemical Physics,* vol. 126, no. 4, Jan 28 2007. [Online]. Available: <Go to ISI>://WOS:000243891100005.

[39] R. Rao and M. Esposito, "Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics," (in English), *Phys Rev X,* vol. 6, no. 4, Dec 22 2016, doi: 10.1103/PhysRevX.6.041064.